NETAJI SUBHAS OPEN UNIVERSITY

STUDY MATERIAL

EEC

PAPER 3

MODULES IX, X, XI, XII

ELECTIVE ECONOMICS HONOURS

প্রাক্‌কথন

নেতাজি সুভাষ মুক্ত বিশ্ববিদ্যালয়ের স্নাতক শ্রেণির জন্য যে পাঠক্রম প্রবর্তিত হয়েছে, তার লক্ষণীয় বৈশিষ্ট্য হ'ল প্রতিটি শিক্ষার্থীকে তাঁর পছন্দমত কোনও বিষয়ে সাম্মানিক (honours) স্তরে শিক্ষাগ্রহণের সুযোগ করে দেওয়া। এক্ষেত্রে ব্যক্তিগতভাবে তাঁদের গ্রহণ ক্ষমতা আগে থেকেই অনুমান করে না নিয়ে নিয়ত মূল্যায়নের মধ্য দিয়ে সেটা স্থির করাই যুক্তিযুক্ত। সেই অনুযায়ী একাধিক বিষয়ে সাম্মানিক মানের পাঠ-উপকরণ রচিত হয়েছে ও হচ্ছে — যার মূল কাঠামো স্থিরীকৃত হয়েছে একটি সুচিন্তিত পাঠক্রমের ভিত্তিতে। কেন্দ্র ও রাজ্যের অগ্রগণ্য বিশ্ববিদ্যালয়সমূহের পাঠক্রমগুলি অনুসরণ করে তার আদর্শ উপকরণগুলির সমন্বয়ে রচিত হয়েছে এই পাঠক্রম। সেই সঙ্গে যুক্ত হয়েছে অধ্যেতব্য বিষয়ে নতুন তথ্য, মনন ও বিশ্লেষণের সমাবেশ।

দূর-সঞ্চারী শিক্ষাদানের স্বীকৃত পদ্ধতি অনুসরণ করেই এইসব পাঠ-উপকরণ লেখার কাজ চলছে। বিভিন্ন বিষয়ের অভিজ্ঞ পণ্ডিতমণ্ডলীর সাহায্য এ কাজে অপরিহার্য এবং যাঁদের নিরলস পরিশ্রমে লেখা, সম্পাদনা তথা বিন্যাসকর্ম সুসম্পন্ন হচ্ছে তাঁরা সকলেই ধন্যবাদের পাত্র। আসলে, এঁরা সকলেই অলক্ষ্যে থেকে দূরসঞ্চারী শিক্ষাদানের কার্যক্রমে অংশ নিচ্ছেন; যখনই কোন শিক্ষার্থী এই পাঠ্যবস্তুনিচয়ের সাহায্য নেবেন, তখনই তিনি কার্যত একাধিক শিক্ষকমণ্ডলীর পরোক্ষ অধ্যাপনার তাবৎ সুবিধা পেয়ে যাচ্ছেন।

এইসব পাঠ-উপকরণের চর্চা ও অনুশীলনে যতটাই মনোনিবেশ করবেন কোনো শিক্ষার্থী, বিষয়ের গভীরে যাওয়া তাঁর পক্ষে ততই সহজ হবে। বিষয়বস্তু যাতে নিজের চেষ্টায় অধিগত হয়, পাঠ-উপকরণের ভাষা ও উপস্থাপনা তার উপযোগী করার দিকে সর্বস্তরে নজর রাখা হয়েছে। এরপর যেখানে যতটুকু অস্পষ্টতা দেখা দেবে, বিশ্ববিদ্যালয়ের বিভিন্ন পাঠকেন্দ্রে নিযুক্ত শিক্ষা-সহায়কগণের পরামর্শে তার নিরসন অবশ্যই হ'তে পারবে। তার ওপর প্রতি পর্যায়ের শেষে প্রদত্ত অনুশীলনী ও অতিরিক্ত জ্ঞান অর্জনের জন্য গ্রন্থ-নির্দেশ শিক্ষার্থীর গ্রহণ-ক্ষমতা ও চিন্তাশীলতা বৃদ্ধির সহায়ক হবে।

এই অভিনব আয়োজনের বেশ কিছু প্রয়াসই এখনও পরীক্ষামূলক — অনেক ক্ষেত্রে একেবারে প্রথম পদক্ষেপ। স্বভাবতই ত্রুটি-বিচ্যুতি কিছু কিছু থাকতে পারে, যা অবশ্যই সংশোধন ও পরিমার্জনার অপেক্ষা রাখে। সাধারণভাবে আশা করা যায়, ব্যাপকতর ব্যবহারের মধ্য দিয়ে পাঠ-উপকরণগুলি সর্বত্র সমাদৃত হবে।

ড. সুরভি বন্দ্যোপাধ্যায়

উপাচার্য

প্রথম পুনর্মুদ্রণ : মার্চ, ২০০৮

ভারত সরকারের দূর শিক্ষা পর্ষদের বিধি অনুযায়ী এবং অর্থানুকূল্যে মুদ্রিত।

Printed in accordance with the regulations and financial assistance of the Distance Education Council, Government of India.

পরিচিতি

বিষয় : ঐচ্ছিক অর্থনীতি (তৃতীয় পত্র) সাম্মানিক স্তর

ঃ বিষয় সমিতি ঃ

ঃ সদস্যবৃন্দ ঃ

১) অধ্যাপক আশিস দাশগুপ্ত
২) অধ্যাপক অমিতাভ চ্যাটার্জী
৩) অধ্যাপক আশিস ব্যানার্জী
৪) অধ্যাপক গৌতম গুপ্ত
৫) অধ্যাপক বিপ্লব দাশগুপ্ত
৬) অধ্যাপক আশিস গুহ

পাঠক্রম — পর্যায় : ই. ই. সি. - ৯

	রচনা	সম্পাদনা
একক ৩৩	ড. ব্রজেন্দ্রকুমার গুহ ঠাকুরতা	অধ্যাপক অতীন্দ্র মোহন গুণ
একক ৩৪	ঐ	ঐ
একক ৩৫	ঐ	ঐ
একক ৩৬	ঐ	ঐ

পাঠক্রম — পর্যায় : ই. ই. সি. - ১০

	রচনা	সম্পাদনা
একক ৩৭	অধ্যাপক বিশ্বনাথ দাশ	অধ্যাপক ভাগবত দাশগুপ্ত
একক ৩৮	ঐ	ঐ
একক ৩৯	ঐ	ঐ

পাঠক্রম — পর্যায় : ই. ই. সি. - ১১

	রচনা	সম্পাদনা
একক ৪০	অধ্যাপিকা নন্দিতা দাশগুপ্ত	অধ্যাপক হরেন শূর
একক ৪১	ঐ	ঐ
একক ৪২	ঐ	ঐ
একক ৪৩	ঐ	ঐ

পাঠ্যক্রম — পর্যায় : ই. ই. সি. — ১২

	রচনা	সম্পাদনা
একক ৪৪	অধ্যাপক পিপুল দত্ত	অধ্যাপক ভাগবত দাশগুপ্ত
একক ৪৫	অধ্যাপক বিশ্বজিৎ রায়	ঐ
একক ৪৬	ঐ	ঐ
একক ৪৭	অধ্যাপক অরূপ কুমার হাইত	ঐ
একক ৪৮	ঐ	ঐ

নেতাজি সুভাষ মুক্ত বিশ্ববিদ্যালয়

ই. ই. সি. — ৩

অর্থনীতির ঐচ্ছিক পাঠক্রম

পর্যায়
৯

পর্যায়
১০

পর্যায়
১১

পর্যায়

১২

একক ৩৩ ❑ বর্ণনাত্মক রাশিবিজ্ঞান

গঠন

৩৩.০ উদ্দেশ্য

এই একটি পড়বার পর আপনি জানতে পারবেন—

- রাশিবিজ্ঞানের কিছু মৌলিক ধারণাগুলি কি কি।
- রাশিতথ্যের আহরণ এবং উপস্থাপন কিভাবে করা হয়।
- পরিসংখ্যান বিভাজন পদ্ধতি কি।

সূত্রগুলি যেখানে সম্ভব বাংলায় বর্ণনা করা হয়েছে। কিন্তু বহু ক্ষেত্রে এটা সম্ভব হয়নি। সেখানে আন্তর্জাতিক প্রথা (International Norm) অনুসরণ করে আন্তর্জাতিক বর্ণমালা এবং রোমান সংখ্যা (Roman Numerals) ব্যবহার করা হয়েছে।

৩৩.১ প্রস্তাবনা

ইংরেজিতে statistics কথাটি দুরকম অর্থে ব্যবহৃত হয়। এক, সংখ্যায় প্রকাশ করা যায় এমন কিছু তথ্যের সমাহার। দুই, এই ধরনের রাশিতথ্যকে যে বিশেষ পদ্ধতিতে বিশ্লেষণ করে সিদ্ধান্তে আসা যায়, সেই পদ্ধতিগুলি নিয়ে যে শাস্ত্রে আলোচনা হয় তারও নাম statistics। দু-একটি উদাহরণ দিলে বিষয়টি পরিষ্কার হবে। ধরুন আপনি বাজারে ফুলের দোকানের পাশ দিয়ে হাঁটছেন, দুপাশে দোকানিরা ফুল বিক্রি করছেন। নানা রঙের ফুল—লাল, নীল, হলুদ, সাদা ইত্যাদি। বিক্রেতাদের বয়সেও তারতম্য রয়েছে। কারও বয়স কুড়ি বছর, কারও পঞ্চাশ, কারও বা ত্রিশ ইত্যাদি। আপনার জানতে ইচ্ছা হ'ল এ রকম দোকানের মোট সংখ্যা কত। তার মধ্যে লালফুলের দোকান কটি, হলুদ, নীল, সাদা ইত্যাদি বিভিন্ন রঙের প্রত্যেকটির দোকান সংখ্যাই বা কী। অন্য দিকে ত্রিশের কম বয়সের কজন দোকানি আছেন? আর বিভিন্ন বয়সের দোকানিদের আলাদা আলাদা সংখ্যাই বা কত? সমীক্ষা করে একটি কাগজে এইসব হিসাব লিখলেন। এ রকমভাবে কিছু রাশিতথ্যের সংগ্রহ বা সমাবেশ করা হ'ল। বাজারের ফুলের দোকানসংক্রান্ত এ ধরনের সংগ্রহকে ইংরেজিতে Collection of Statistics বা শুধু Statistics বলা হয়। বাংলায় এর পরিভাষা করা হ'য়েছে **পরিসংখ্যান**। আবার এখানেই থেমে না থেকে সংগৃহীত এই পরিসংখ্যানের বিস্তৃত বিশ্লেষণ করা যেতে পারে বিশেষ কিছু বৈজ্ঞানিক পদ্ধতি মেনে। যেমন বিভিন্ন ধরনের ফুলের গড়পড়তা দাম কত, সবচেয়ে দামী ফুলের থেকে সবচেয়ে কমদামী ফুলের দামের তফাৎ কত, বিভিন্ন দোকানির গড়পড়তা বিক্রি কত এবং আরও অনেক বিস্তৃত খবরাখবর। সংগৃহীত পরিসংখ্যানের এ ধরনের বিশ্লেষণের জন্য যে বৈজ্ঞানিক পদ্ধতি ব্যবহার করা হয় তাকেও ইংরেজিতে statistics (বা কখনও কখনও statistical methods) বলা হয়। কিন্তু বাংলাভাষায় এর নাম দেওয়া হ'য়েছে **রাশিবিজ্ঞান**।

তা হ'লে আমরা দেখতে পাচ্ছি ইংরেজিতে 'রাশিতথ্যের সমাহার' এবং 'রাশিতথ্য সংক্রান্ত বিজ্ঞান' দুটিকেই সাধারণত একই নামে (অর্থাৎ statistics নামে) উল্লেখ করা হয়ে থাকে। কিন্তু বাংলায় প্রথমটিকে বলা হয় **পরিসংখ্যান** এবং দ্বিতীয়টিকে বলা হয় **রাশিবিজ্ঞান**।

এখানে আমরা রাশিবিজ্ঞানের বিভিন্ন পদ্ধতি নিয়ে আলোচনা করবো।

৩৩.২ রাশিবিজ্ঞানের কিছু মৌলিক ধারণা

গুণ লক্ষণ (***Qualitative Character or Attribute***) *বনাম পরিমাণসূচক লক্ষণ বা চল* (***Quantitative Character or Variable***) : আমাদের দৈনন্দিন জীবনের নানা প্রয়োজনে কিংবা নানাধরনের বৈজ্ঞানিক অনুসন্ধানের প্রয়োজনে রাশিতথ্য সংগ্রহের এবং বিশ্লেষণের দরকার হয়। এই বিশ্লেষণ করা হয় নানাধরণের গাণিতিক পদ্ধতির সাহায্যে। তবে প্রথমেই যে জিনিসটি বিচার করে দেখতে হয় তা হ'ল সংগৃহীত রাশিতথ্যের প্রকৃতি। এই প্রকৃতি বিচার করা হয় রাশিতথ্যের ধরন বা লক্ষণ দেখে। এই লক্ষণ দুরকম হ'তে পারে—(ক) **গুণগত লক্ষণ** (qualitative character বা attribute) এবং (খ) **পরিমাণগত লক্ষণ বা চল** বা

চলক quantitative character বা variable)। কিছু কিছু লক্ষণ আছে যেগুলিকে সরাসরি সংখ্যার সাহায্যে প্রকাশ করা সম্ভব নয়। যেমন সূচনায় (অনুচ্ছেদ ৩৩.১-এ) উল্লেখিত ফুলের লাল, নীল, সাদা ইত্যাদি নানা রঙ, দোকানিদের পরিধেয়ের ধরন (যেমন প্যান্ট, ধুতি, পাজামা ইত্যাদি)। এগুলিকে বলা হয় গুণগত লক্ষণ (attribute)। এই গুণগত লক্ষণগুলির হিসাব করা হয় অপ্রত্যক্ষভাবে (indirectly)। যেমন, বাজারে লাল ফুলের সংখ্যা কত, হলুদ ফুলের সংখ্যা কত ইত্যাদি।

আবার কিছু কিছু রাশিতথ্য আছে যা পরিমাণসূচক। যেমন, স্কুলে পাঠরত কোনো একটি শ্রেণীর ছাত্রদের উচ্চতা। এখানে প্রতিটি ছেলের উচ্চতা প্রত্যক্ষভাবে (directly) মাপা যেতে পারে। এ রকম মাপ নিয়ে ঐ শ্রেণীতে পাঠরত ছাত্রদের উচ্চতার পরিসংখ্যান (statistics) পাওয়া যেতে পারে। এখানে ছাত্রদের উচ্চতার মাপ একটি চল (variable) যার প্রত্যেক পরিমাপ করা সম্ভব।

৩৩.৩ রাশিতথ্যের আহরণ এবং উপস্থাপন

৩৩.৩.১ প্রাথমিক সূত্রে এবং গৌণসূত্রে সংগৃহীত পরিসংখ্যান (primary and secondary data)

এর আগে আমরা যে আলোচনা করেছি সেখানে আমরা ধরে নিয়েছি যে আমাদের কাছে নানাসূত্রে সংগৃহীত কিছু পরিসংখ্যান আছে এবং এদের প্রকৃতি সম্বন্ধে বিশ্লেষণ বা আলোচনা করা হ'য়েছে। কিন্তু সর্বপ্রথম দেখা দরকার এই পরিসংখ্যান আমরা কীভাবে সংগ্রহ করতে পারি। বাজারের বিভিন্ন রঙের ফুলের হিসাবই হোক বা ভারতের বিভিন্ন প্রদেশের জনসংখ্যার হিসাবই হোক বা ক্লাসের ছেলেদের উচ্চতার হিসাবই হোক —আমাদের প্রথম প্রশ্নই হবে এইসব হিসাব আমরা কোথা থেকে এবং কীভাবে পাব।

রাশিতথ্যের হিসাব দুভাবে যোগাড় করা যেতে পারে। এক, প্রাথমিক বা মূলসূত্র থেকে। দুই, কোনো গৌণ সূত্র থেকে।

ধরা যাক, আমরা জানতে চাই কলকাতা শহরে কোনো একটি নির্দিষ্ট দিনে কত মোটরগাড়ি মালিক আছেন—এই হিসাবটা বিভিন্ন উপায়ে জোগাড় করা যেতে পারে। স্বাভাবিকভাবে প্রথমেই যে উপায়টির কথা মনে আসে তা হ'ল শহরের প্রতিটি বাড়িতে গিয়ে খোঁজ নেওয়া এবং তার থেকে একটা মোট হিসাব জোগাড় করা। কিন্তু এই পদ্ধতির অনেকগুলি অসুবিধা রয়েছে। এত বড় শহরের প্রতিটি বাড়িতে যেতে হ'লে প্রচুর লোক নিয়োগ ক'রতে হবে। অনেক সময় দরকার হবে, অনেক পরিশ্রম হবে এবং অনেক খরচ হবে। তবে সবগুলি বাড়িতে না গিয়ে, শহরের দশ শতাংশ বাড়ি যদি এমনভাবে নির্বাচন করা সম্ভব হয় যে এই বাড়িগুলিতে যত মোটরগাড়ির মালিক বাস করেন মোটামুটিভাবে তার দশগুণ মোটরগাড়ির মালিক যারা শহরে বাস করেন, তা হ'লে এই দশ শতাংশ বাড়ির থেকে পাওয়া হিসাবের দশগুণ ক'রলেই সমস্ত শহরের মোটরগাড়ির মালিকের হিসাব পাওয়া যেতে পারে। এই পদ্ধতির প্রয়োগে প্রথম পদ্ধতির চাইতে সময়, খরচ এবং পরিশ্রম অনেকটাই কমে যাবে।

উপরে যে দুটি পদ্ধতির কথা বলা হ'ল সে দুটিই প্রাথমিক সূত্র থেকে রাশিতথ্য সংগ্রহের উপায়। প্রথম পদ্ধতিটিকে বলা হয় **সম্পূর্ণ গণনা পদ্ধতি** (Complete Enumeration Method) এবং দ্বিতীয়টিকে বলা হয় **নমুনা সমীক্ষা পদ্ধতি** (Sample Survey Method)।

এবার আমরা গৌণ সূত্র থেকে রাশিতথ্য আহরণের উপায়ের কথা বলবো। এটা সুবিদিত মোটরগাড়ির প্রতিটি মালিককে গাড়ি কেনার সাথে সাথেই মোটর ভেহিকলস্ দপ্তরে (motor vehicles department) তার পঞ্জীকরণ (registration) দরকার হয়। ফলে ঐ দপ্তরে প্রতিটি গাড়ির হিসাব নথিভুক্ত করা আছে। এই খবরটা জানা থাকলে এত পরিশ্রম এবং অর্থ ব্যয় ক'রে বাড়ি বাড়ি তথ্য সংগ্রহ না ক'রে এই একটি সরকারি দপ্তর থেকে অনেক সহজে এবং কম পরিশ্রমে গাড়ির মালিকের মোট সংখ্যা জানা যেতে পারে। এই দ্বিতীয় উপায়টিকে গৌণসূত্রে রাশিতথ্য সংগ্রহের (secondary) পদ্ধতি বলা হয়।

৩৩.৩.২ প্রাথমিক এবং গৌণ সূত্র থেকে রাশিতথ্য সংগ্রহের কয়েকটি পদ্ধতি

(ক) **প্রাথমিক সূত্র :** আগেই বলা হ'য়েছে যে, প্রাথমিক সূত্র থেকে রাশিতথ্য সম্পূর্ণ গণনা পদ্ধতি (Complete Enumeration Method) বা নমুনা সমীক্ষা পদ্ধতি (Sample Survey Method) অনুসরণ করে করা যেতে পারে। এই দুটি ক্ষেত্রেই রাশিতথ্যগুলি মূল সূত্র থেকে জোগাড় করা হয়। যেমন, আগের উদাহরণে মোটরগাড়ির সংখ্যা কত জানতে মোটরগাড়ির মালিকদের বাড়িতে গিয়ে খোঁজ নেওয়ার কথা বলা হয়েছে। এই খোঁজ নানারকমভাবে নেওয়া যেতে পারে। এর মধ্যে তিনটি পদ্ধতি বহুল প্রচলিত—(1) প্রশ্নগুচ্ছ ভিত্তিক পদ্ধতি (questionnaire method), (2) সাক্ষাৎকার ভিত্তিক পদ্ধতি (interview method), এবং (3) প্রত্যক্ষ পর্যবেক্ষণ পদ্ধতি (direct observation method)।

(1) **প্রশ্নগুচ্ছ ভিত্তিক পদ্ধতি (Questionnaire Method) :** এই পদ্ধতিতে নির্দিষ্ট ব্যক্তিকে (যেমন আগের উদাহরণের গাড়ির মালিককে) একটি প্রশ্নমালা পাঠিয়ে দেওয়া হয়। এটা সাধারণত ডাকযোগে করা হয়। প্রশ্নমালার সঙ্গে অনেক সময় একটি প্রেরকের ঠিকানা লেখা প্রাক্-মূল্য-প্রদত্ত (pre-paid) খাম পাঠিয়ে দেয়া হয়, অনুরোধ করা হয় যে জবাব দেওয়া প্রশ্নমালা যেন ঐ খামে ফেরৎ দেওয়া হয়। এই পদ্ধতিটি চালু করা সহজ। কিন্তু উত্তরদাতা খুব সচেতন না হ'লে অনেক সময়ই উত্তর না পাওয়ার সম্ভাবনা থাকে।

(2) **সাক্ষাৎকার ভিত্তিক পদ্ধতি :** এখানে অনুসন্ধানকারী নির্দিষ্ট ব্যক্তি (বা প্রতিষ্ঠানের) সঙ্গে দেখা করে রাশিতথ্য সংগ্রহ করে থাকেন। এই পদ্ধতির সুবিধা হ'ল উত্তরদাতার সঙ্গে ব্যক্তিগত যোগাযোগ হওয়ায় জবাব পাওয়ার সম্ভাবনা অনেকটা বেড়ে যায়। কিন্তু এর কিছু অসুবিধাও আছে। উত্তরদাতার মর্জি এবং প্রশ্নকর্তার ব্যবহার, বুদ্ধিমত্তা এবং কাজ করার ইচ্ছার ওপর এর সাফল্য নির্ভর করে।

(3) **প্রত্যক্ষ পর্যবেক্ষণ পদ্ধতি :** কিছু কিছু ক্ষেত্রে অনুসন্ধানকারী নিজেই প্রত্যক্ষ পর্যবেক্ষণ করে রাশিতথ্যের হিসাব সংগ্রহ করতে পারেন—কোনো দ্বিতীয় ব্যক্তির উপর নির্ভর না করে। যেমন, যদি কোনো শহরের বিভিন্ন এলাকায় কটি বাড়ি আছে তা জানার দরকার হয়, তা হলে অনুসন্ধানকারী নিজেই গণনা করে বাড়ির সংখ্যার হিসাব সংগ্রহ করতে পারেন। অন্য কারও কাছ থেকে খবর নেওয়ার প্রয়োজন থাকে না।

৩৩.৩.৩ পরিসংখ্যান পরীক্ষা (Scrutiny of data)

বিভিন্ন উপায়ে সংগৃহীত পরিসংখ্যান প্রথমেই ভালভাবে পরীক্ষা (scrutiny) করা দরকার। কারণ, তথ্যসংগ্রহের সময় নানারকম ভুল থেকে যেতে পারে যা সংশোধন না করলে এগুলির ব্যবহারযোগ্যতা এবং বিশ্বাসযোগ্যতা ক্ষুণ্ণ হবে।

প্রাথমিক সূত্র সংগ্রহের সময় পরিসংখ্যানে নানারকমভাবে ভুল হতে পারে। তথ্যসংগ্রহকারী অনিচ্ছাকৃতভাবে বা ইচ্ছাকৃতভাবে ভুল করতে পারেন। অনিচ্ছাকৃত ভুল বহুবিধ হতে পারে। অনেকসময় অমনোযোগের জন্য লেখার ভুল হতে পারে। যেমন, পাঁচজন লোকের উচ্চতার মাপ (সেন্টিমিটারে) নেওয়ার সময় দেখা গেল সংগ্রহকারী লিখেছেন 158, 162, 160, 1665, 157 সেন্টিমিটার।

উপরের রাশিতথ্যের উপর চোখ বোলালেই দেখা যাবে যে, একটি ক্ষেত্রে উচ্চতা 1665 সেমি. লেখা হয়েছে। কিন্তু আমাদের অভিজ্ঞতায় 1665 সেমি. কোনো মানুষের উচ্চতা হতে পারে না। একটু চিন্তা করলেই বোঝা যাবে এটা হওয়া উচিত 165 সেমি., অনবধানতায় 6 দুবার লেখার ফলেই ভুল হয়েছে। এই জাতীয় ভুলগুলি সংগৃহীত রাশিতথ্যকে বেশ ভালো করে পর্যবেক্ষণ ক'রলেই সাধারণ বুদ্ধির প্রয়োগেই সংশোধন করা যেতে পারে।

প্রাথমিক সূত্র থেকে সংগৃহীত পরিসংখ্যান ক্ষেত্রে আর একরকম ভুল হ'ল ইচ্ছাকৃত ভুল। অনেক সময় সংগ্রহকারী নিজের কাজের ভার কমাবার জন্য সঠিক সূত্র থেকে তথ্য সংগ্রহ না করে নিজের বিচারবুদ্ধি খাটিয়ে কিছু কিছু কাল্পনিক রাশিতথ্য লিখে রাখতে পারেন। যেমন, কয়েকটি পরিবারের মাসিক আয়ের হিসাব যোগাড় করার সময় নির্দিষ্ট পরিবারগুলির সবকটিতে না গিয়ে কয়েকটি প্রতিনিধিত্বমূলক (representative) পরিবারের আয়ের হিসাব নিয়ে বাকী পরিবারগুলির আয়ের হিসাব সংগ্রহকারীর বিচারবুদ্ধি প্রয়োগ করে আন্দাজমতো লিখে রাখতে পারেন। এ ধরনের ভুল স্পষ্টতই সংগ্রহকারীর অসাধুতার জন্য হ'য়ে থাকে। এগুলি অনেক সময়ই শুধুমাত্র তথ্য পর্যবেক্ষণ করে ধরা সম্ভব হয় না। এর জন্য প্রাথমিক সংগ্রহকারীর জোগাড় করা তথ্যের কিছু নমুনা আর একদল সংগ্রহকারী বা তত্ত্বাবধায়ক (superiors) দ্বিতীয়বার মূল সূত্রে গিয়ে মিলিয়ে দেখার দরকার হয়। যেমন, আগের উদাহরণটির ক্ষেত্রে নির্দিষ্ট পরিবারগুলির (যেগুলি থেকে প্রাথমিক সংগ্রহকারী হিসাব সংগ্রহ করেছেন) মধ্য থেকে কয়েকটি পরিবারে দ্বিতীয়বার খোঁজখবর নিয়ে জানা যেতে পারে যে প্রাথমিক সংগ্রহকারী সঠিক তথ্য সংগ্রহ করেছেন কিনা।

গৌণ সূত্র থেকে সংগৃহীত রাশিতথ্যের ভুলের পরিমাণ অনেকসময় বেশ বেশি হতে পারে। কারণ, এসব ক্ষেত্রে মুখ্য (বা প্রাথমিক) সূত্রে যে তথ্য সংগৃহীত হয় তা মূলত প্রশাসনিক কাজে ব্যবহার করার জন্য হয়ে থাকে। এই উদ্দেশ্যে গৃহীত তথ্যগুলি তত নিখুঁত না হলেও প্রশাসনিক কাজের অসুবিধা হয় না। কিন্তু রাশিবিজ্ঞানীর দৃষ্টিভঙ্গি থেকে যতটা সঠিক পরিসংখ্যান দরকার সেই পরিমাণ শুদ্ধতা এসব ক্ষেত্রে নাও থাকতে পারে। সেজন্যই এসব ক্ষেত্রে সংগৃহীত পরিসংখ্যান বিশেষভাবে পরীক্ষা করা দরকার।

রাশিতথ্য পরীক্ষা করার সময় পরীক্ষক তার সাধারণ বুদ্ধি (common sense) প্রয়োগ করে তথ্যের

বিশ্বাসযোগ্যতা সম্বন্ধে একটা ধারণা করতে পারেন। একটি উদাহরণ দিলে বিষয়টি স্পষ্ট হবে। ধরা যাক, কয়েকটি মধ্যবিত্ত পরিবারের আয়-ব্যয়ের হিসাব জানার জন্য একটি সমীক্ষা করা হয়েছে। সমীক্ষায় প্রাপ্ত বিভিন্ন পরিবারের আয়ের হিসাব পরীক্ষা করে দেখা গেল কোনো ক্ষেত্রেই মাসিক আয় আট হাজার টাকার বেশি নয়। কিন্তু প্রতিটি পরিবারের ব্যয়ের হিসাব (যা খাওয়া খরচ, পোষাক-পরিচ্ছদের খরচ, বাড়িভাড়া এবং অন্যান্য প্রাসঙ্গিক বিষয়ের খরচ আলাদা আলাদা করে দেখান হয়েছে) পরীক্ষা করতে গিয়ে দেখা গেল যে একটি পরিবারের মোট মাসিক ব্যয় তিরিশ হাজার টাকারও বেশি। এই ধরনের সমীক্ষায় পরিবারগুলি মাঝে কী পরিমাণ টাকা ধার করে থাকেন তার হিসাবও নেওয়া হয়। কিন্তু এই বিশেষক্ষেত্রে দেখা গেল পরিবারটি কোনো ধার করেনি। সুতরাং এক্ষেত্রে আয়-ব্যয়ের হিসাবের এই বিরাট গরমিল কোনোক্রমেই বিশ্বাসযোগ্য নয় এবং এখানে নতুন করে সমীক্ষা করে দেখা দরকার।

উপরের উদাহরণ থেকে দেখা যাচ্ছে যে, সংগৃহীত পরিসংখ্যান খুব মনোযোগের সঙ্গে তীক্ষ্ণবুদ্ধি প্রয়োগ করে পরীক্ষা করা এবং দরকারমতো সংশোধন করা দরকার। অল্পস্বল্প ভুল হলে নিরীক্ষণকারী নিজেই বিচারবুদ্ধি প্রয়োগ করে সংশোধন করে নিতে পারেন। কিন্তু বড় ধরনের গোলমাল থাকলে (উপরের উদাহরণে যেমন দেখান হয়েছে) নতুন করে সমীক্ষার দরকার হয়।

৩৩.৩.৪ পরিসংখ্যান উপস্থাপন (Presentation of data)

পরিসংখ্যান সংগ্রহ এবং পরীক্ষা করার পর এগুলিকে পরিচ্ছন্ন এবং সুবোধ্যভাবে পরিবেশন করা দরকার। এ কাজটি এমনভাবে করতে হবে যে একজন সাধারণ বুদ্ধিসম্পন্ন, রাশিবিজ্ঞানে অনভিজ্ঞ লোক সহজেই এদের অন্তর্নিহিত অর্থ বুঝতে পারেন।

উপস্থাপিত পরিসংখ্যান নানাধরনের হতে পারে। যেমন, একটি দেশের বিভিন্ন সময়ের জনসংখ্যার হিসাব, খাদ্যশস্য উৎপাদনের হিসাব, কয়েকটি প্রয়োজনীয় জীবনধারণের উপকরণের দরের হিসাব ইত্যাদি। এসব ক্ষেত্রে সময়ের সাথে সাথে জিনিসগুলির (বা রাশিবিজ্ঞানের ভাষায় 'চলগুলির') মানের পরিবর্তন লক্ষ্য করা যায়। এই ধরনের পরিসংখ্যান সারিকে *কালীন সারি* (Time Series) বলা হয়। অর্থাৎ যে সমস্ত পরিসংখ্যানের বেলায় সময়ের পরিবর্তনের সঙ্গে নির্দিষ্ট চলের (variable) মানের (value) পরিবর্তন দেখানোই উদ্দেশ্য তাদেরই *কালীন সারি* বলা হয়।

আর এক ধরনের পরিসংখ্যান আছে যেখানে বিভিন্ন মানের বা বিভিন্ন মান সীমার মধ্যে প্রাপ্ত রাশিতথ্যের হিসাব নেওয়া হয়ে থাকে। যেমন, একটি অঞ্চলের বিভিন্ন লোকের আয়ের হিসাব এভাবে নেওয়া যেতে পারে— 500 টাকা পর্যন্ত আয়ের কতজন লোক আছে, 501 টাকা থেকে 1000 টাকা পর্যন্ত আয়ের কতজন লোক আছে, 1001 থেকে 2000 টাকা পর্যন্ত আয়ের কতজন লোক আছে ইত্যাদি। এখানে প্রতিটি ব্যক্তির আয়ের হিসাব পাওয়া যায় না, কিন্তু সমষ্টিগতভাবে একটি মানসীমার মধ্যে ঐ আয়ের মোট লোকের হিসাব পাওয়া যায়। এ ধরনের রাশিতথ্যকে বলা হয় *পরিসংখ্যা রাশিতথ্য* (Frequency Data)। অন্যদিকে কালীন সারি জাতীয় রাশিতথ্যকে (যেখানে প্রতিটি মানের হিসাব পাওয়া যায়) বলা হয় *অপরিসংখ্যা রাশিতথ্য* (Non-frequency Data)।

৩৩.৩.৫ বর্ণনাত্মক উপস্থাপন পদ্ধতি

এই পদ্ধতিতে বর্ণনার সাহায্যে, রাশিতথ্যের উপস্থাপন করা হয়ে থাকে। উদাহরণস্বরূপ, নিচে ভারত সরকারের কেন্দ্রীয় পরিসংখ্যান সংস্থা (Central Statistical Organisation বা সংক্ষেপে C.S.O.)-এর 1999 সালের "Quick Estimates of Index of Industrial Production" থেকে একটি বিবরণ দেওয়া হ'ল :

"1999-এর জুন মাসে ভারতের শিল্পের উৎপাদনের বৃদ্ধি হ'য়েছে শতকরা 5.6 হারে—যা নাকি গতবছরের হারের (শতকরা 4.5) চাইতে বেশি। এর মধ্যে যন্ত্রপাতির উৎপাদনের হার শতকরা 6.5 ভাগ এবং বিদ্যুৎ উৎপাদনের হার শতকরা 4.1 ভাগ বৃদ্ধি পেয়েছে এবং খনিজ জিনিসপত্রের হারে শতকরা 1.6 ভাগ হ্রাস হ'য়েছে।"

ওপরের উদাহরণটি একটি অত্যন্ত সংক্ষিপ্ত বর্ণনা। অনেকসময় আরও অনেক বেশি রাশিতথ্য সন্নিবেশ ক'রে বেশ কয়েক পাতা জুড়ে এ ধরনের বর্ণনাত্মক উপস্থাপন করা হয়।

এই পদ্ধতির সুবিধা হ'ল এটি ব্যবহার ক'রতে রাশিবিজ্ঞানী হওয়ার দরকার হয় না। যে কোনো সাধারণ বুদ্ধিসম্পন্ন শিক্ষিত ব্যক্তি একটু মনঃসংযোগ করলেই এর অর্থ অনুধাবন করতে পারেন। কিন্তু এই পদ্ধতি থেকে উপস্থাপিত রাশিতথ্য সম্বন্ধে সম্যক ধারনা করতে হলে পাঠককে বর্ণনাটি অত্যন্ত মনোযোগ দিয়ে (দরকার হলে একাধিকবার) পাঠ করতে হবে—বিশেষ করে বর্ণনাটি যদি দীর্ঘ হয়। এতে যথেষ্ট সময় নষ্ট হয় এবং ধৈর্যেরও দরকার হয়। একজন অভিজ্ঞ রাশিবিজ্ঞানীর কাছে তাই এই পদ্ধতি তেমন আকর্ষণীয় নয়।

৩৩.৩.৬ সারণী বিন্যাস

সারণীর সাহায্যে রাশিতথ্যকে অনেক সংক্ষেপে অনেক গ্রহণযোগ্য এবং কার্যকারীভাবে প্রকাশ করা যেতে পারে। এ ধরনের উপস্থাপনায় সাধারণত কোনোরকম বর্ণনাত্মক ব্যাখ্যার প্রয়োজন হয় না। সারণীটি সঠিকভাবে তৈরি করতে পারলে এর ভেতর সন্নিবেশিত পরিসংখ্যান অতি সহজেই পাঠকের বোধগম্য হয়। এই পদ্ধতির আর একটি বড় সুবিধা হল—রাশিতথ্য উপস্থাপনার কোথাও কোনো গলদ বা ভুল থাকলে তা সহজেই ধরা যায়। বর্ণনাত্মক পদ্ধতিতে কিন্তু এ ধরনের ছোটখাট ভুল অনেক সময় নজর এড়িয়ে যেতে পারে।

একটি সারণীকে সাধারণত নিম্নলিখিত বিভিন্ন অংশে ভাগ করা যেতে পারে :

(ক) **শিরোনাম (title) :** সারণীটির মাথায় খুব সংক্ষেপে এর ভেতরকার বিষয়বস্তুর উল্লেখ করা হ'য়ে থাকে। সারণীটির একটি নির্দেশিকা-সংখ্যা (identifying number)-ও সাধারণত উল্লেখ করা হয়ে থাকে।

(খ) **"স্টাব" (stub) বা সারণীর বাঁ-পাশের বর্ণনাত্মক স্তম্ভ :** এটি সারণীটির একদম বাঁপাশে থাকে এবং বাঁ-দিক থেকে অনুভূমিকভাবে উল্লিখিত বর্ণনা এতে দেওয়া থাকে।

(গ) **স্তম্ভ-শিরোনাম (caption) :** সারণীর মাথায় প্রতিটি স্তম্ভের (column) বর্ণনা দিয়ে এই শিরোনাম স্থাপন করা হয়। এখানে প্রতি স্তম্ভের মধ্যে উল্লিখিত একক, সংখ্যা ইত্যাদির বর্ণনাও দেওয়া থাকে।

(ঘ) **সারণীর অঙ্গ (body of the table) :** এটিই সারণীর মূল অংশ। এখানেই সংশ্লিষ্ট রাশিতথ্য প্রদর্শিত হয়।

(ঙ) **পাদটীকা (foot-note) :** অনেক ক্ষেত্রেই সারণীর নিচে একটি কিংবা একাধিক পাদটীকা (foot-note) দেওয়া থাকে। এখানে রাশিতথ্য কোথা হতে সংগৃহীত হয়েছে, অর্থাৎ তার উৎস (source), বিবরণ এবং দরকার মতো অন্যান্য তথ্যও উল্লেখ করা হয়।

কয়েকটি বিষয় মনে রাখা দরকার। প্রথমত সারণীটির দৈর্ঘ্য এবং প্রস্থ মোটামুটি আনুপাতিক হওয়া চাই। অর্থাৎ দৈর্ঘ্যের তুলনায় প্রস্থ যেন খুব বেশি বড় বা ছোট না হয়। অবশ্য সবসময় এই নিয়ম অনুসরণ করা সম্ভব হয় না। সারণীতে পরিবেশিত রাশিতথ্যের প্রকৃতির ওপর অনেক কিছু নির্ভর করে। যেমন, দেশের গত 25 বছরের খাদ্যশস্যের মূল্য এবং উৎপাদনের হিসাব যদি একটি সারণীর সাহায্যে প্রকাশ করতে হয়, তা হলে স্পষ্টতই দৈর্ঘ্যটি প্রস্থের চাইতে বেশ খানিকটা বেশি হবে। এছাড়া সারণীর ভেতরকার বিষয়বস্তু এমনভাবে পরিবেশন করতে হবে যে একজন সাধারণ পাঠকের কাছে সহজেই বোধগম্য হয়। নিচে একটি উদাহরণ দেওয়া হল :

সারণী—১

পশ্চিমবঙ্গের জলপাইগুড়ি জেলার কয়েকটি প্রধান খাদ্যশস্য ফলনের জমির পরিমাণ

একক (unit) : হাজার হেক্টর

খাদ্যশস্যের নাম	1973-74	1974-75	1975-76
তণ্ডুলজাতীয় :			
1. চাল : মোট	263.1	288.7	304.6
আউশ	68.4	84.5	94.9
আমন	194.5	204.1	209.7
অন্যান্য প্রকার চাল	0.2	0.1	—
2. গম	3.4	7.0	23.5
3. যব	2.4	3.1	4.0
4. ভুট্টা	2.2	2.2	3.2
5. অন্যান্য	0.9	0.9	1.6
তণ্ডুলজাতীয় সকল শস্য	272.0	301.9	336.9

সারণী—১ (*contd.*)

খাদ্যশস্যের নাম	1973-74	1974-75	1975-76
ডালজাতীয় :			
6. ছোলা	—	—	—
7. কলাই	0.2	0.2	0.2
৪. অন্যান্য	8.2	8.2	11.1
ডালজাতীয় সকল শস্য	8.4	8.4	11.3
সকল খাদ্যশস্য	280.4	310.3	348.2

উৎস (*Source*)—পশ্চিমবঙ্গ রাজ্য ফলিত অর্থনীতি ও পরিসংখ্যান দপ্তর (Bureau of Applied Economics and Statistics, West Bengal)।

৩৩.৩.৭ চিত্রের সাহায্যে পরিসংখ্যানের উপস্থাপনা (Diagrammatic Representation of Data)

নানা ধরনের চিত্রের সাহায্যে রাশিতথ্যের প্রকাশন এবং পরিবেশন একটি খুব জনপ্রিয় পদ্ধতি। এই পদ্ধতির সর্বাপেক্ষা বড় সুবিধা হল, এটিতে ছবি ব্যবহার করায় তা সহজেই চোখে পড়ে যায় এবং পাঠকের মনোগ্রাহী হয়। ফলে রাশিতথ্যের অন্তর্নিহিত তাৎপর্য সম্বন্ধে পাঠক একটি সাধারণ ধারণা সহজেই করে নিতে পারেন। অনভিজ্ঞ সাধারণ পাঠকের জন্য পদ্ধতিটি খুবই ফলপ্রদ। কিন্তু অন্যদিকে একথাও মনে রাখা দরকার যে, এটি একটি সরলীকৃত পদ্ধতি এবং এর সাহায্যে খুব বেশি বিশ্লেষণাত্মক বিচার করা সম্ভব নয়।

এবার আমরা কয়েকটি বিশেষ ধরনের চিত্রের ব্যবহার সম্বন্ধে আলোচনা করব।

(ক) **রেখাচিত্র (graph বা line diagram) :** রেখাচিত্রের সাহায্যে পরিসংখ্যানের উপস্থাপন একটি বহুল প্রচলিত পদ্ধতি। এই পদ্ধতিতে সমকোণে পরস্পর খণ্ডিত দুটি অক্ষরেখার (একটি উল্লম্ব, অন্যটি আনুভূমিক) সহায়তায় বিন্দু স্থাপিত করা হয়। এই পদ্ধতিটি বিশেষ করে কালীন সারির বেলায় খুবই ব্যবহৃত হয়। একটি উদাহরণ দেওয়া যাক। নিচের সারণীটিতে পশ্চিমবঙ্গে পাঁচ বছরের চালের উৎপাদনের হিসাব দেওয়া আছে :

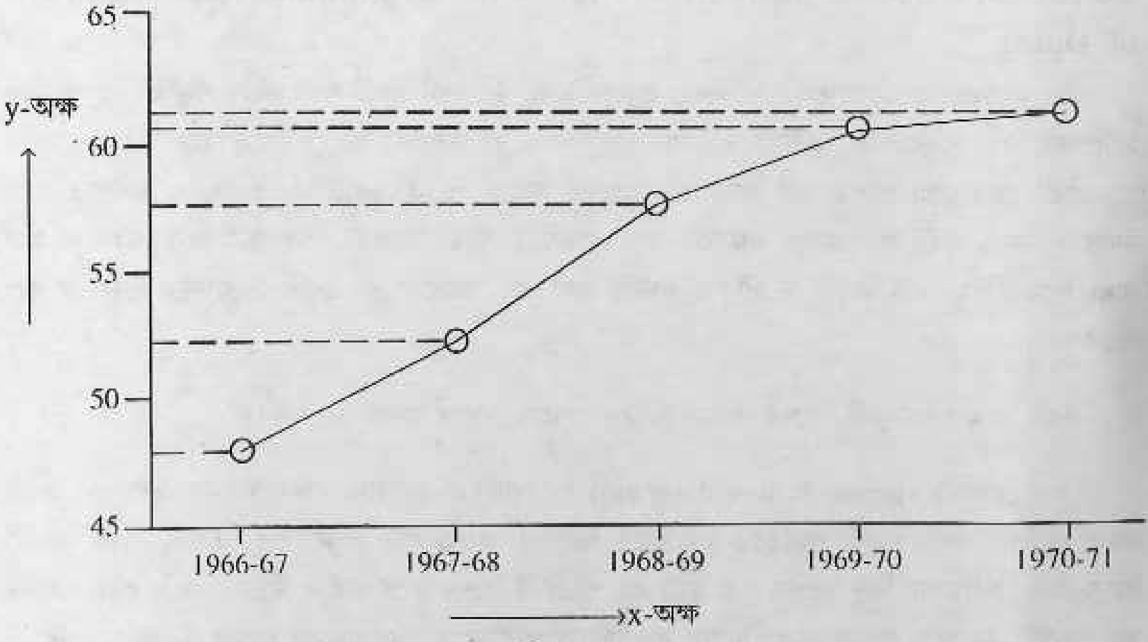
সারণী—২

পশ্চিমবঙ্গে চালের উৎপাদন

1966-67 থেকে 1970-71

বর্ষ	উৎপাদন (লক্ষ টন)
1966-67	48.24
1967-68	52.08
1968-69	57.80
1969-70	60.55
1970-71	61.40

উপরের হিসাবটি রেখাচিত্রের সাহায্যে প্রকাশ করতে হলে সময়সূচক চলটি (অর্থাৎ বর্ষ) আনুভূমিক অক্ষে দেখানো হবে এবং সেই সময়সীমার জন্য প্রাপ্ত উৎপাদনের মানগুলি উল্লম্ব অক্ষটিতে দেখান হয়। এরপর সুবিধামতো এককের স্কেল ব্যবহার করে সময়সীমার মধ্যবিন্দুতে উৎপাদনের মানগুলি রেখাপত্রে (graph paper-এ) বিন্দুর দ্বারা চিহ্নিত করা হবে। এই বিন্দুগুলিকে সরলরেখার সাহায্যে পরস্পর যুক্ত করলে কাঙ্ক্ষিত রেখাচিত্রটি পাওয়া যাবে।



চিত্র নং-১

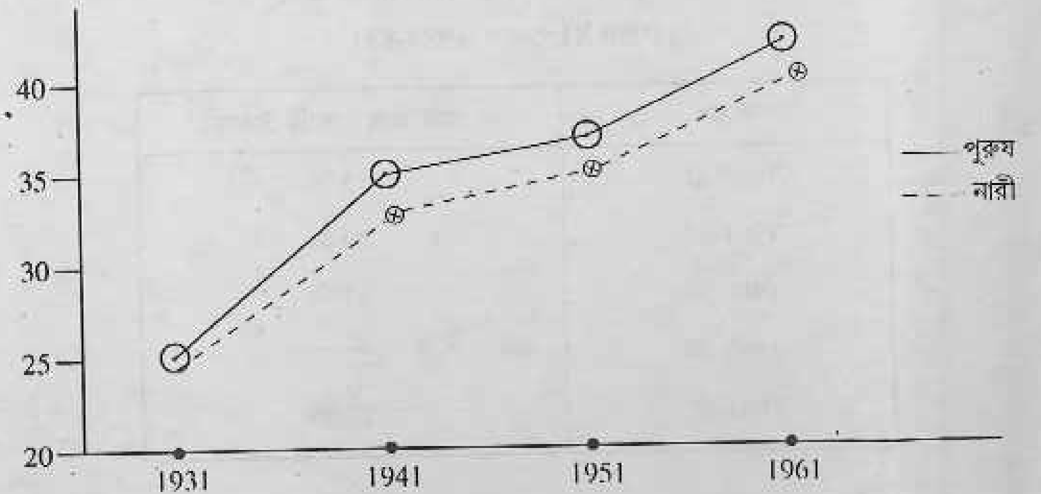
অনেক সময় একই রেখাপত্রের মধ্যে অক্ষরেখা দুটি (অর্থাৎ আনুভূমিক অক্ষ এবং উল্লম্ব অক্ষ) অভিন্ন রেখে শুধুমাত্র বিভিন্ন একক (unit) ব্যবহার করে একাধিক রেখাচিত্র অঙ্কন করা যেতে পারে। বিশেষ করে দুই বা ততোধিক তুলনামূলক পরিসংখ্যানগুচ্ছকে রেখাচিত্রের সাহায্যে প্রকাশ করতে হলে এই পদ্ধতি ব্যবহার করা হয়। যেমন নিচের রেখাচিত্রটিতে ৪—জনগণনায় ভারতবর্ষের পুরুষ ও মহিলা জনসংখ্যার প্রত্যাশিত আয়ু (expectation of life)-এর তুলনামূলক রেখাচিত্রে দেওয়া হল।

সারণী—৩

ভারতীয় নারীপুরুষের প্রত্যাশিত আয়ু **(1931-1961)**

সাল	প্রত্যাশিত আয়ু (একক-বৎসর)	
	পুরুষ	নারী
1931	26.91	26.56
1941	32.09	31.37
1951	32.45	31.66
1961	41.89	40.55
1971		
1981		
1991		

উৎস : *কেন্দ্রীয় পরিসংখ্যান সংস্থা, ভারত সরকার*
(Central Statistical Organisation, Govt. of India)



চিত্র নং-২

রেখাচিত্র অনেক সময় সাধারণ স্কেলে না দেখিয়ে 'লগ স্কেল' (logarithmic scale)-এ দেখানো হয়ে থাকে। অর্থাৎ অক্ষগুলিতে চিহ্নিত মানগুলির পরিমাণ আনুপাতিক হারে প্রকাশ করা হয়। এরকম রেখাচিত্র দুধরনের হতে পারে। যেখানে শুধুমাত্র উল্লম্ব অক্ষরেখাটি লগ-স্কেলে দেখান হয় আর অনুভূমিক অক্ষরেখাটি সাধারণ স্কেলে দেখান হয় তখন রেখাচিত্রটিকে একাক্ষ লগচিত্র (semi-logarithmic chart) বা অনুপাত চিত্র (ratio chart) বলা হয়। অন্যদিকে যেখানে উল্লম্ব অক্ষ এবং আনুভূমিক অক্ষ দুটিকেই লগস্কেলে দেখান হয় সেখানে রেখাচিত্রকে উভয়াক্ষ লগচিত্র (doubly logarithmic chart) বলা হয়। এ ধরনের চিত্রের (একাক্ষ এবং উভয়াক্ষ) বাজারে পাওয়া যায়। এদের সমীকরণগুলি নিচে দেখানো হল :

(1) একাক্ষ রেখাচিত্র—

$\log y = a + bx$, অর্থাৎ $y = AB^x$

(2) উভয়াক্ষ রেখাচিত্র—

$\log y = a + b \log x$,, অর্থাৎ $y = Ax^b$

যেখানে $\log A = a$ এবং $\log B = b$

(খ) **স্তম্ভচিত্র (bar diagram) :** স্তম্ভচিত্রের ব্যবস্থার খুবই প্রচলিত। এখানে এক-একটি শ্রেণীর রাশিতথ্যকে এক-একটি স্তম্ভ (bar)-এর সাহায্যে প্রকাশ করা হয়ে থাকে। প্রতিটি স্তম্ভ সমান প্রস্থের হয়। স্তম্ভের উচ্চতার সাহায্যে চালকের সংশ্লিষ্ট মান দেখানো হয়ে থাকে। স্তম্ভচিত্রের ব্যবহার নানাধরনের রাশিতথ্যের জন্য করা হয়ে থাকে। তবে কালীন সারি এবং ভৌগোলিক রাশিতথ্য সারি প্রদর্শনের জন্য এর ব্যবহার খুব ব্যাপক। নিচে দুটি উদাহরণ দেওয়া হল।

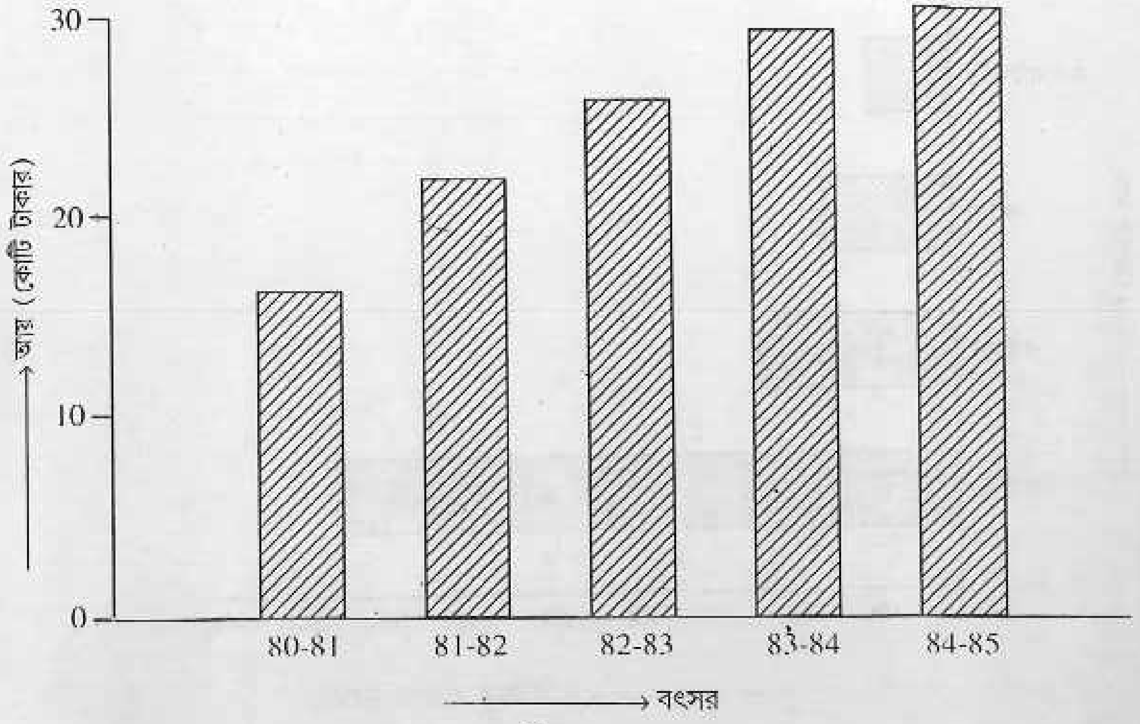
সারণী—8

অরণ্যসম্পদ থেকে পশ্চিমবঙ্গে বার্ষিক আয়

(1980-81 থেকে 1984-85)

বৎসর	মোট আয় (কোটি টাকায়)
1980-81	15.76
1981-82	20.44
1982-83	23.74
1983-84	27.77
1984-85	27.86

উৎস : *আর্থিক সমীক্ষা*, পশ্চিমবঙ্গ সরকার, 1985-86



চিত্র নং-৩

উপরের স্তম্ভচিত্রটি সারণী ৪-এ বর্ণিত কালীন সারির চিত্রায়িত রূপ। কালীন সারির ক্ষেত্রে এরকম উল্লম্ব স্তম্ভচিত্র সাধারণত ব্যবহার করা হয়ে থাকে।

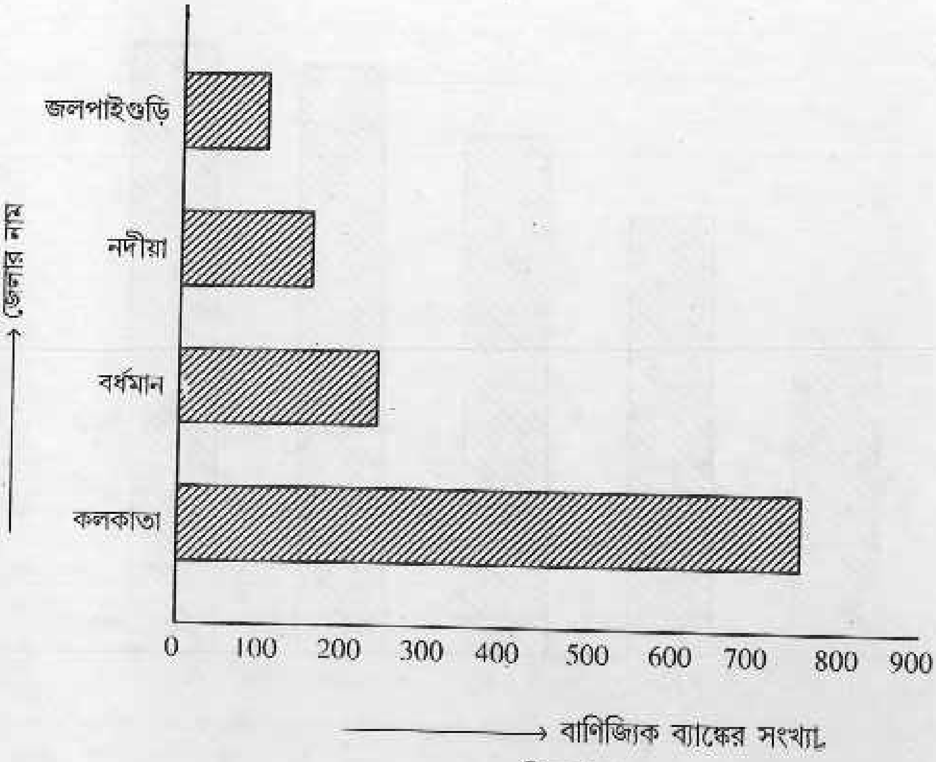
অন্যদিকে ভৌগোলিক সারির ক্ষেত্রে অনেকসময়ই আনুভূমিক স্তম্ভচিত্র ব্যবহার করা হয়ে থাকে। নিচে একটি উদাহরণ দেওয়া হল।

সারণী—৫

পশ্চিমবঙ্গে কয়েকটি জেলায় বাণিজ্যিক ব্যাঙ্কের সংখ্যা (জুন, 1984)

জেলার নাম	বাণিজ্যিক ব্যাঙ্কের সংখ্যা
কলকাতা	744
বর্ধমান	221
নদীয়া	105
জলপাইগুড়ি	81

উৎস : *আর্থিক সমীক্ষা*, পশ্চিমবঙ্গ সরকার, 1985-86।



চিত্র নং-৪

স্তম্ভচিত্রের নানারকম ধরন হতে পারে—যেমন খণ্ডিত-স্তম্ভচিত্র (divided bar diagram) কিংবা বহু স্তম্ভচিত্র (multiple bar diagram)। স্তম্ভচিত্রের প্রতিটি স্তম্ভকে বিভিন্ন গুণ (attribute) অনুযায়ী বিভিন্ন ভাগে ভাগ করে দেখিয়ে খণ্ডিত স্তম্ভচিত্র পাওয়া যায়। আবার একটি গুণেরই বিভিন্ন সময়ের রাশিতথ্যের প্রকাশ বহু স্তম্ভচিত্রের দ্বারা করা যেতে পারে।

(গ) **বৃত্ত-চিত্র (pie-diagram) :** চিত্রের সাহায্যে অনুপাতসূচক বা শতাংশসূচক রাশিতথ্য প্রকাশের জন্য কোনো কোনো সময় বৃত্তচিত্রের ব্যবহার করা হয়। বিশেষ করে যেখানে একটি চলকে কয়েকটি অংশে (সাধারণত শতাংশের হিসাবে) ভাগ করে এই অংশগুলির তুলনামূলক বিচার করা হয় সেখানে বৃত্তচিত্রের ব্যবহার খুব উপযোগী। যেমন কেন্দ্রীয় কিংবা বা রাজ্য বাজেটের জন্য বরাদ্দ টাকার কোন্ অংশ কোন্ উদ্দেশ্যে ব্যবহার করা হয়েছে, কিংবা একটি রাজ্যের মোট উৎপাদিত খাদ্যশস্যের কতটা অংশ চাল, কতটা গম, কতটা ডালজাতীয় ইত্যাদি—এই সমস্ত রাশিতথ্য বৃত্তচিত্রের সাহায্যে ভালভাবে প্রকাশ করা যায়।

একটি বিন্দুর চারপাশে কোণের মোট পরিমাণ 360° ডিগ্রী। সুতরাং কোনো রাশিতথ্যে প্রাপ্ত শতাংশগুলিকে 360° ডিগ্রীর পরিমাণমতো ভগ্নাংশে রূপান্তরিত করে একটি বৃত্তের কেন্দ্রে অনুরূপ মাপের কোণের সাহায্যে খণ্ডাংশগুলিকে প্রকাশ করা যেতে পারে। অর্থাৎ, সূত্র অনুযায়ী বৃত্তের এক-একটি খণ্ডাংশের কোণের পরিমাপ হবে :

চালের শতকরা পরিমাণ $\times \frac{360}{100}$ ডিগ্রী।

কোণের এই মাপগুলি জ্যামিতিক উপায়ে প্রোট্র্যাকটার (protractor) ও কম্পাস (compass)-এর সাহায্যে পরিমাপ করা যেতে পারে। এই মাপগুলি পাওয়া গেলে একটি বৃত্ত অঙ্কন ক'রে তার মধ্যে এগুলি মাপমতো কোণে ভাগ করে দেখানো যেতে পারে। নিচে একটি উদাহরণ দেওয়া হল।

সারণী—৬

কোনো একটি বছরের কেন্দ্রীয় সরকারের কর থেকে প্রাপ্ত আয়ের হিসাব

আয়ের উৎস	পরিমাণ (কোটি টাকায়)	শতকরা হিসাব
বহিঃশুল্ক (Customs)	10880	32
অন্তঃশুল্ক (Excise)	6800	20
আয়কর (Income Tax)	8160	24
করপোরেট কর (Corporate Tax)	3400	10
অন্যান্য কর	4760	14
মোট	34000	100

এগুলি ডিগ্রীতে প্রকাশ করে আমরা সংশ্লিষ্ট কোণের পরিমাপ পাই :

আয়ের উৎস	শতকরা হিসাব	কোণের পরিমাপ
বহিঃশুল্ক (Customs)	32	115.2°
অন্তঃশুল্ক (Excise)	20	72.0°
আয়কর (Income Tax)	24	86.4°
করপোরেট কর (Corporate Tax)	10	36.0°
অন্যান্য কর	14	50.4°
মোট	100	360.0°

সুতরাং বৃত্তের অংশে প্রকাশ করলে :

চিত্র নং—১.৭

৩৩.৪ পরিসংখ্যা বিভাজন (frequency distribution)

কালীন সারি বা ভৌগোলিক সারির বেলায় সেখানে রাশিতথ্যে প্রদর্শিত প্রতিটি মান আলাদা করে দেখে নানাধরনের সিদ্ধান্ত নিতে হয়। আমাদের আগেকার আলোচনাগুলিতে এ ধরনের বিশ্লেষণের বিষয়ে উল্লেখ করা হয়েছে। কিন্তু অন্য ধরনের পরিসংখ্যানের বেলায় বিশেষত যানসংখ্যা অধিক প্রতিটি মানের কথা মনে রেখে সম্পূর্ণ রাশিতথ্যের গতিপ্রকৃতির বিচারবিশ্লেষণ দুঃসাধ্য হয়ে পড়ে। এর প্রয়োজনও নেই। যেখানে দশজন ব্যক্তির উচ্চতার মাপ বিশ্লেষণ করা হচ্ছে সেখানে প্রতিটি উচ্চতা আলাদা করে দেখা যেতে পারে। কিন্তু যেখানে 100, 500 বা 1000 লোকের উচ্চতার মাপ সংক্রান্ত রাশিতথ্যের বিচার বিশ্লেষণ করা হবে, সেখানে প্রতিটি মাপ মনে রেখে প্রতিটি মানের বিশ্লেষণ অসম্ভব হয়ে পড়ে। এসব ক্ষেত্রে সংগৃহীত রাশিতথ্যকে অন্যধরনের পদ্ধতি ব্যবহার ক'রে সংক্ষেপিত উপায়ে প্রকাশ করা হয়। এতে রাশিতথ্যের মানের খবর পুরোপুরিভাবে না জানা গেলেও সমষ্টিগতভাবে তথ্য পাওয়া যায়। পরিসংখ্যানে এ ধরনের সংক্ষেপীকরণ পদ্ধতিকে পরিসংখ্যা বিভাজন (frequency distribution) পদ্ধতি বলা হয়। একটি উদাহরণ দেওয়া যাক্।

কলকাতার কোনো একটি অঞ্চলে অবস্থিত 1000টি বাড়ির প্রত্যেকটিতে কতগুলি করে ঘর আছে প্রত্যক্ষ অনুসন্ধানের সহায়তায় তার হিসাব নেওয়া হল। তারপর ঘরের সংখ্যা অনুযায়ী বাড়িগুলিকে 10টি শ্রেণীতে ভাগ করে একটি সারণী করা হল :

সারণী—৭

1,000 বাড়ির ঘরসংখ্যার পরিসংখ্যা বিভাজন

ঘরের সংখ্যা	বাড়ির সংখ্যা (পরিসংখ্যা)	আনুপাতিক পরিসংখ্যা
1	50	0.050
2	125	0.125
3	200	0.200
4	175	0.175
5	350	0.350
6	0	0.000
7	55	0.055
8	0	0.000
9	40	0.040
10	5	0.005
মোট	1.000	1.000

উপরের উদাহরণটিতে একটি **'চল'** বা **পরিমাণ সূচক লক্ষণ** (variable বা quantitative character)-এর পরিসংখ্যা বিভাজন দেখানো হয়েছে। এখানে ঘরের সংখ্যা হল চল (variable) এবং বাড়ির সংখ্যা হল **পরিসংখ্যা (frequency)**। ঠিক একইভাবে 'গুণলক্ষণ' (qualitative character বা attribute)-এর পরিসংখ্যা নির্ণয় করা যেতে পারে। নিচের উদাহরণটিতে দেখানো হয়েছে :

সারণী—৮

একটি বাগানের নানা রঙের ফুলের পরিসংখ্যা বিভাজন

ফুলের রঙ	পরিসংখ্যা	আনুপাতিক পরিসংখ্যা
সাদা	450	0.486
লাল	250	0.270
নীল	130	0.141
হলুদ	70	0.076
কালো	25	0.027
মোট	925	1.000

উপরের সারণীদুটির (সারণী নং ৭ ও ৮) দ্বিতীয় স্তম্ভে যে সংখ্যা দেখানো হয়েছে তা হল **আপেক্ষিক পরিসংখ্যা** (absolute frequency)। কিন্তু অনেকসময় বিশেষ করে গুণলক্ষণ সংক্রান্ত পরিসংখ্যা বিভাজনের ক্ষেত্রে লক্ষণসমূহের বিভিন্ন রূপের আপেক্ষিক গুরুত্ব সম্বন্ধে স্পষ্ট চিত্র পেতে হলে **আপেক্ষিক পরিসংখ্যা** (relative frequency) গণনা করতে হয়। আপেক্ষিক পরিসংখ্যার (কিংবা সমানুপাতিক পরিসংখ্যার) সংজ্ঞা হল :

$$\text{শ্রেণীর আনুপাতিক পরিসংখ্যা} = \frac{\text{শ্রেণীর পরিসংখ্যা}}{\text{মোট পরিসংখ্যা}}$$

উপরের সারণী দুটির তৃতীয় স্তম্ভে আপেক্ষিক পরিসংখ্যা দেখান হয়েছে।

(ক) ***বিচ্ছিন্ন চল ও অবিচ্ছিন্ন চল*** : চল (variable) দুরকম হতে পারে—(1) **বিচ্ছিন্ন চল** (discrete variable) ও (2) **অবিচ্ছিন্ন চল** বা **সন্তত চল** (continuous variable)। কিছু কিছু চল আছে যেগুলি শুধু বিচ্ছিন্ন মানই নিতে পারে। যেমন সারণী ৭-এ উল্লিখিত চলটি। বাড়ির ঘরের সংখ্যার মান সবসময় পূর্ণসংখ্যা হবে। এটির মান কোনো ভগ্নাংশ সংখ্যা হতে পারে না। 2.13টি ঘর কিংবা 6.24টি ঘর—এ রকম মান অবাস্তব। কিন্তু যখন কিছু লোকের উচ্চতার মাপ নেওয়া হচ্ছে তখন তার মান অবিচ্ছিন্ন হবে। কোনো দুই সীমার মধ্যে

মানুষের উচ্চতার মান যে কোনো সংখ্যা হতে পারে। উদাহরণস্বরূপ, 5.10, 4.32 ফুট, 6.417 ফুট ইত্যাদি। এখন তাত্ত্বিক বিচারে (theoretically) চলটির মান পুরোপুরিভাবে অবিচ্ছিন্ন–যদিও মাপযন্ত্রের সীমাবদ্ধতার জন্য বাস্তবক্ষেত্রে কিছুটা অবিচ্ছিন্নতা দেখা যেতে পারে।

বিচ্ছিন্ন চলের পরিসংখ্যা সারণী (frequency table) তৈরি করা অনেকটা সহজ। কিন্তু অবিচ্ছিন্ন চলের ক্ষেত্রে এ রকম সারণী তৈরি করতে একটি বিশেষ পদ্ধতি অবলম্বন করা হয়। অবিচ্ছিন্ন চলের মানের সংখ্যা অসীম হওয়ায় এখানে পরিসংখ্যা বিভাজনের শ্রেণীবিন্যাস বিচ্ছিন্ন চলের মতো সহজ হয় না। কৃত্রিম উপায়ে চলের প্রসারকে কয়েকটি সীমিত শ্রেণীতে ভাগ করা হয়ে থাকে। একটি উদাহরণ দেওয়া যাক। যেখানে কিছু মানুষের উচ্চতার পরিমাপ করা হচ্ছে সেখানে সর্বপ্রথম সবচাইতে লম্বা লোকটির উচ্চতা এবং সবচাইতে খাটো লোকটির উচ্চতা সম্বন্ধে ধারণা করতে হবে। ধরা যাক, সবচাইতে লম্বা লোকটির উচ্চতা 7 ফুট এবং সবচাইতে বেঁটে লোকটির উচ্চতা 3 ফুট। এবার এই দুই উচ্চতাসীমার মধ্যের মানুষগুলির উচ্চতা কয়েকটি কৃত্রিম শ্রেণীতে ভাগ করা যেতে পারে। যদি ঠিক হয় যে, উচ্চতার পরিমাপ আসন্ন (approximately) এক দশমিক অঙ্কে প্রকাশ করা হবে (যেমন 4.3 ফুট, 5.7 ফুট ইত্যাদি) তা হলে শ্রেণীবিভাগ দুই দশমিক অঙ্কের মানে প্রকাশ করলে বাস্তবক্ষেত্রে প্রায় অবিচ্ছিন্ন মানের কাছাকাছি হিসাব পাওয়া যাবে। যেমন উপরোক্ত উদাহরণের শ্রেণীবিন্যাসটি সারণী—৩-এর ধারায় করা যেতে পারে।

সারণী—৯

100 লোকের উচ্চতার পরিসংখ্যা বিভাজন

উচ্চতা (ফুট)	লোকসংখ্যা (পরিসংখ্যা)	পরিসংখ্যা ঘনত্ব (frequency density)
2.50-3.00	4	8
3.00-3.50	12	24
3.50-4.00	26	52
4.00-4.50	41	82
4.50-5.00	67	134
5.00-5.50	39	78
5.50-6.00	25	50
6.00-6.50	14	28
6.50-7.00	12	4
মোট	240	—

এই শ্রেণীবিন্যাসের সাহায্যে এক দশমিক শুদ্ধতার মানের যে কোনো উচ্চতার পরিমাপ সম্বন্ধে ধারণা করা যেতে পারে। এখানে 2.51 ফুঃ—3.51 ফুঃ বা 3.51 ফুঃ—4.51 ফুঃ হল একএকটি **শ্রেণী অন্তর (class interval)** । এক-একটি শ্রেণী অন্তরের সীমাগুলিকে বলা হয় **শ্রেণীসীমা** (class limits)। যেমন 3.51 ফুঃ হল 2.5 ফুঃ—3.5 ফুঃ শ্রেণীটির ঊর্ধ্বসীমা এবং 3.5 ফুঃ—4.5 ফুঃ শ্রেণীটির অধঃসীমা। এগুলিকে (যেমন 3.51 ফুঃ বা 2.51 ফুঃ ইত্যাদি) বলা হয় শ্রেণীসীমান্ত (class boundaries) শ্রেণী অন্তরের মধ্যবিন্দুকে বলা হয় শ্রেণীমধ্যক (class mid-point বা class mark)। যেমন, $^1/_2$ (2.5 + 3.5) = 3.0 ফুঃ হল (2.5 ফুঃ—3.5 ফুঃ) শ্রেণীটির শ্রেণীমধ্যক। কোনো শ্রেণীর ঊর্ধ্বসীমান্ত এবং অধঃসীমান্তের অন্তর বা বিয়োগফলকে বলা হয় এই শ্রেণীর শ্রেণীপ্রসার (class width) বা শ্রেণীদৈর্ঘ্য। যেমন, 3.51 ফুঃ—2.51 ফুঃ = 1.0 ফুঃ হল (2.51 ফুঃ—3.51 ফুঃ) শ্রেণীর শ্রেণীপ্রসার বা শ্রেণীদৈর্ঘ্য।

পরিসংখ্যা ঘনত্ব (*frequency density*) : কোনো শ্রেণীর পরিসংখ্যাকে শ্রেণীদৈর্ঘ্য দিয়ে ভাগ করলে ঐ শ্রেণী অন্তরের পরিসংখ্যা ঘনত্ব (frequency density) পাওয়া যায়। সূত্রে প্রকাশ করলে :

$$\text{পরিসংখ্যা ঘনত্ব} = \frac{\text{শ্রেণী-পরিসংখ্যা}}{\text{শ্রেণীপ্রসার}}$$

সারণী—৩-এ (2.51 ফুঃ — 3.51 ফুঃ) শ্রেণীটির ক্ষেত্রে :

$$\text{পরিসংখ্যা ঘনত্ব} = \frac{5 \text{ শ্রেণী-পরিসংখ্যা}}{1 \text{ শ্রেণী-প্রসার}}$$
$$= 5$$

শ্রেণীগুলির পরিসংখ্যার বিভিন্নতার জন্য পরিসংখ্যা ঘনত্বও প্রতিটি শ্রেণীর জন্য বিভিন্ন হবে (1.38) সারণীতে তৃতীয় স্তম্ভে পরিসংখ্যা শ্রেণী অনুযায়ী পরিসংখ্যা ঘনত্ব দেখানো হয়েছে।

(খ) পরিসংখ্যা বিভাজন প্রস্তুত করার ব্যবহারিক পদ্ধতি :

(1) **সিলচিহ্নের (Tally marks-এর) ব্যবহার :** –এতক্ষণ গুণলক্ষণ (attribute)-এর বিচ্ছিন্ন চল (discreet variables) চল, (continuous variable)-এর পরিসংখ্যা বিভাজনের বর্ণনা করা হয়েছে। কিন্তু বাস্তবে উৎস থেকে যে প্রণালী অবলম্বন করে এই পরিসংখ্যা বিভাজন সারণীগুলি পাওয়া যায় এখানে তার বর্ণনা করা হবে। একটা উদাহরণ দেওয়া যাক্ :

কলকাতার কোনো একটি অঞ্চলে 10টি বাড়ির মেঝের আচ্ছাদিত অংশের (covered area-র) মাপ ঐ বাড়িগুলিতে গিয়ে মাপ নিয়ে সংগ্রহ করা হয়েছে। এগুলি এ রকম :

100 বর্গ মিটার, 120 বঃ মিঃ, 160 বঃ মিঃ, 60 বঃ মিঃ, 100 বঃ মিঃ, 100 বঃ মিঃ, 140 বঃ মিঃ, 100 বঃ মিঃ, 120 বঃ মিঃ, 100 বঃ মিঃ।

ওপরের রাশিতথ্যের হিসাব নিয়ে একটি পরিসংখ্যা বিভাজন সারণী তৈরি করতে হবে। দেখা যাচ্ছে ঐ রাশিতথ্যসমূহের মানের সর্ববৃহৎটি হল 160 বঃ মিঃ এবং সবচাইতে কম মানটি হল 60 বঃ মিঃ। এই 10টি সংখ্যা নিরীক্ষণ করে দেখা যাচ্ছে কয়েকটি সংখ্যা একাধিকবার এসেছে। এই সংখ্যাগুলিকে অধঃক্রমে (descending order) সাজিয়ে তারপর এক-একটি সংখ্যা কতবার এসেছে তা সংখ্যাটির পাশে লিখতে হবে। সুবিধার জন্য সংখ্যাটির প্রতিটি আবির্ভাবের জন্য এর পাশে একটি করে চিহ্ন (বা মিলচিহ্ন) নিম্নলিখিতভাবে দেওয়া যেতে পারে।

সারণী—১০

10টি বাড়ির মেঝের আচ্ছাদিত অংশের পরিসংখ্যান

বাড়ির মেঝের আচ্ছাদিত অংশের মাপ (বর্গ মিটার)	বাড়ি সংখ্যার মিলচিহ্ন	পরিসংখ্যা
60	I	1
100	V	5
120	II	2
140	I	1
160	I	1
	10	10

সুবিধার জন্য প্রতি পাঁচটি মিলচিহ্নকে একসাথে সাজিয়ে নিয়ে একটি ব্লক (block) তৈরি করা হয় এবং পঞ্চম চিহ্নটিকে কোণাকুণিভাবে আগের চারটি চিহ্নকে খণ্ডিত করা অবস্থায় দেখান হয়। উপরে উদাহরণে এই নিয়ম অনুসরণ করা হয়েছে।

গুণগত লক্ষণের (attribute) ক্ষেত্রে প্রতিটি লক্ষণের জন্য উৎস থেকে প্রাপ্ত রাশিতথ্যগুলিকে প্রতিটি লক্ষণের জন্য মিলচিহ্নের সাহায্যে সংখ্যা নির্ণয় করে পরিসংখ্যা বিভাজন সারণী তৈরি করা হয়।

পরিমাণগত লক্ষণ বা চলের ক্ষেত্রে (variable)-এর ক্ষেত্রে পূর্বে উল্লিখিত শ্রেণীবিন্যাস পদ্ধতি অনুসরণ করে প্রতিটি শ্রেণী নির্ণয় করে সেই শ্রেণীটির অন্তর্গত মিলচিহ্নগুলির সংখ্যা নির্ণয় করা হয়ে থাকে।

(গ) **কিছু সংজ্ঞা :** এতক্ষণ পরিসংখ্যা বিভাজনের পদ্ধতি এবং কি করে প্রাথমিক রাশিতথ্য থেকে মিলচিহ্নের ব্যবহার করে এই বিভাজন ব্যবহার করা যায় এবং পরিসংখ্যা বিভাজন সারণী প্রস্তুত করা যায় তার বর্ণনা দেওয়া হয়েছে। এখানে আমরা এর সাথে সম্পর্কিত কয়েকটি সংজ্ঞার বিবরণ দেব। নিচের সারণীতে 200 লোকের দৈহিক ওজনের পরিসংখ্যা বিভাজন দেখান হয়েছে।

সারণী—১১

২০০ লোকের দৈহিক ওজনের পরিসংখ্যা বিভাজন

ওজন (কেজি)	লোকসংখ্যা	আপেক্ষিক পরিসংখ্যা (শতকরা)	ক্রমযৌগিক পরিসংখ্যা	
			ক্ষুদ্রতর সূচক (less than type)	বৃহত্তর সূচক (more than type)
50 থেকে 55 এর মধ্যে	200	10.00	200	2000
55 থেকে 60 এর মধ্যে	600	30.0	800	1800
60 থেকে 65 এর মধ্যে	800	40.0	1600	1200
65 থেকে 70 এর মধ্যে	250	12.5	18250	400
70 থেকে 75 এর মধ্যে	100	5.0	19250	150
75 এর ঊর্ধ্বে	50	2.5	2000	50
মোট	2000	100.0		

লক্ষ্য করলে দেখা যাবে ওপরের সারণীটি পরিমাণসূচক অবিচ্ছিন্ন চলের সারণী। অর্থাৎ এখানে 50 কেজির ঊর্ধ্বে কিছু মানুষের ওজনের মাপ নেওয়া হয়েছে। 3নং স্তম্ভে আপেক্ষিক পরিসংখ্যা (যা আগেই ব্যাখ্যা করা হয়েছে) শতকরা হিসাবে দেখান হয়েছে। 4নং এবং 5নং স্তম্ভে ক্রমযৌগিক পরিসংখ্যার যথাক্রমে ক্ষুদ্রতর সূচক (less than type) এবং বৃহত্তর সূচক (more than type) দেখান হয়েছে। ক্রমযৌগিক পরিসংখ্যা কম বা বেশির মোট হিসাব জানার জন্য দরকার হয়। যেমন, ১১ উল্লিখিত 2000 লোকের মধ্যে কতজনের ওজন 65 কেজির কম তা জানতে হলে 4নং স্তম্ভের তৃতীয় সারিটির সহায়তা নিয়ে বলা যেতে পারে যে এরকম 1600 ব্যক্তি আছেন। অন্যদিকে 65 কেজির বেশি ওজনের 1200 লোকের হিসাব 5নং স্তম্ভের তৃতীয় সারি থেকে পাওয়া যাচ্ছে।

৩৩.৫ লৈখিক পদ্ধতিতে পরিসংখ্যা বিভাজনের উপস্থাপনা (diagramatic representation of frequency distribution)

এর আগে আমরা অপরিসংখ্যা-রাশিতথ্যের (non-frequency data) লৈখিক পদ্ধতি সম্বন্ধে আলোচনা করেছি। এই অনুচ্ছেদে পরিসংখ্যা বিভাজনের ক্ষেত্রে লৈখিক পদ্ধতির ব্যবহার সম্পর্কে আলোচনা করবো।

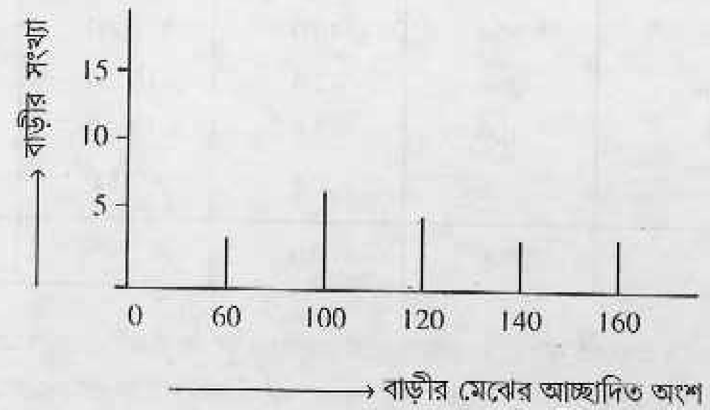
(ক) **গুণলক্ষণের পরিসংখ্যা বিভাজনের লৈখিক চিত্র :** কোনো গুণলক্ষণের পরিসংখ্যা বিভাজন বা আপেক্ষিক পরিসংখ্যা বিভাজনের উপস্থাপন পূর্ববর্ণিত (সারণী ৫ উল্লেখ্য) অপরিসংখ্যা বিভাজনের লৈখিক চিত্রেরই অনুরূপ। যেমন সারণী ৫-এ বর্ণিত জেলাভিত্তিক বাণিজ্যিক ব্যাঙ্কের সংখ্যাকে আমরা পরিসংখ্যা বলে

অভিহিত করতে পারি। কারণ নামকরণের তফাৎ ছাড়া এখানে আর কোনো তফাৎ নেই। সুতরাং ঐ ক্ষেত্রে যে স্তম্ভচিত্রটি আঁকা হয়েছে তাতে উক্ত গুণলক্ষণের লৈখিক উপস্থাপন বলেও বর্ণনা করতে পারি।

(ঘ) চলের পরিসংখ্যা বিভাজনের চিত্রভিত্তিক উপস্থাপন :

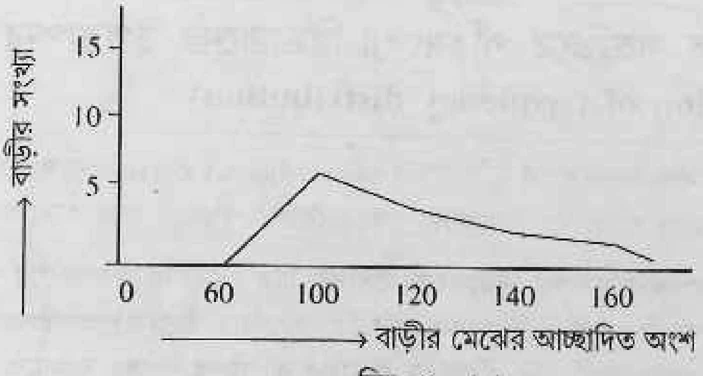
(অ) **বিচ্ছিন্ন চল** : একটি বিচ্ছিন্ন চলের ক্ষেত্রে নিম্নলিখিত পদ্ধতিগুলির কোনো একটি দ্বারা তার পরিসংখ্যান বিভাজনকে উপস্থাপন করা যেতে পারে।

[1] **পরিসংখ্যা স্তম্ভচিত্র (column diagram) :** এখানে একটি সুবিধামতো স্কেল ব্যবহার করে অনুভূমিক অক্ষরেখার বিভিন্ন বিন্দুতে চলের বিভিন্ন মানগুলি নির্দিষ্ট করে উল্লম্ব অক্ষে প্রাপ্ত পরিসংখ্যার মাপের সমান লম্ব (column) উত্তোলন করে স্তম্ভচিত্র অঙ্কন করা হয়। 1.8 চিত্রে ৯ সারণীর পরিসংখ্যা-বিভাজনের পরিসংখ্যা-স্তম্ভচিত্র দেখানো হল।



চিত্র নং—১.৮

[2] **পরিসংখ্যা বহুভুজ (frequency polygon) :** এখানে চলের বিভিন্ন মানের সাথে সংশ্লিষ্ট পরিসংখ্যাসমূহকে সমকোণী স্থানাঙ্কের (rectangular co-ordinates) সাহায্যে লেখচিত্রের বিভিন্ন বিন্দুতে সংস্থাপন করা হয়। এই বিন্দুগুলিকে সরলরেখার সাহায্যে সংযোজিত করে যে চিত্রটি পাওয়া যায় তাকে পরিসংখ্যা বহুভুজ বলা হয়। আগেকার (৯ সারণীর) পরিসংখ্যাকে নিচের চিত্রে পরিসংখ্যা বহুভুজের সাহায্যে প্রকাশ করা হল।



চিত্র নং—১.৯

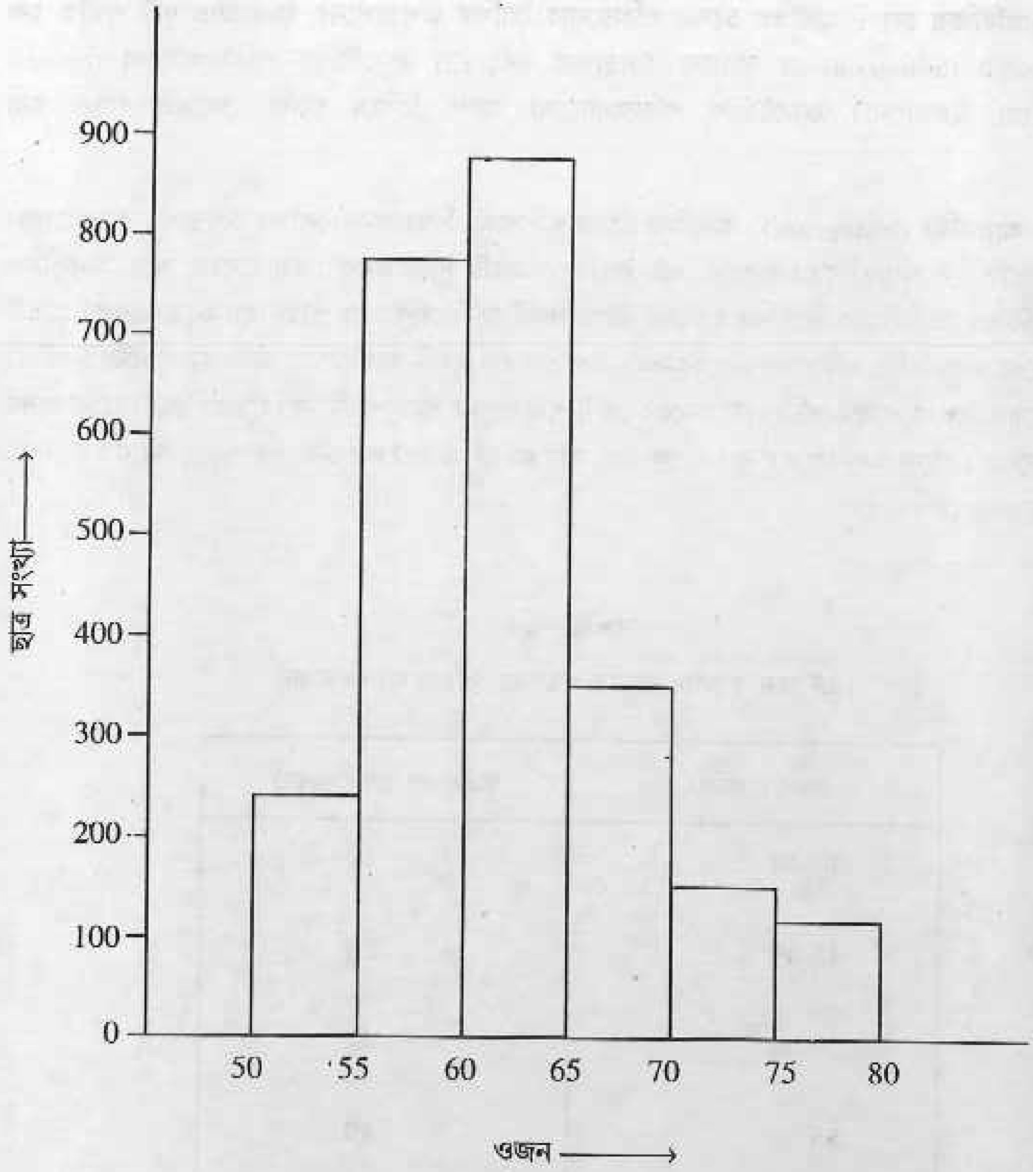
(আ) **অবিচ্ছিন্ন চল :** অবিচ্ছিন্ন চলের পরিসংখ্যার লৈখিক উপস্থাপনের বহুব্যবহৃত দুটি পদ্ধতি হল (1) আয়তচিত্র (histogram)-এর সাহায্যে উপস্থাপনা এবং (2) ক্রমযৌগিক পরিসংখ্যারেখা (ogive)-এর সাহায্যে উপস্থাপনা। ক্রমযৌগিক পরিসংখ্যারেখা অবশ্য বিচ্ছিন্ন চলের ক্ষেত্রেও অঙ্কন করা চলে।

(1) **আয়তচিত্র (histogram) :** অবিচ্ছিন্ন চলের পরিসংখ্যা বিভাজনের লৈখিক উপস্থাপনের সর্বাপেক্ষা প্রচলিত পদ্ধতি হল আয়তচিত্রের ব্যবহার। এই পদ্ধতিতে একটি সুবিধাজনক স্কেল ব্যবহার করে আনুভূমিক অক্ষটিতে বিভিন্ন শ্রেণীসীমান্ত নির্দেশিত হয় এবং উল্লম্ব অক্ষটিতে নির্দেশিত হয় পরিসংখ্যা ঘনত্ব। এরপর প্রতিটি শ্রেণীঅন্তরের ওপর নির্দিষ্ট পরিসংখ্যা ঘনত্বের সমান উচ্চতার এক একটি আয়তক্ষেত্র আঁকা হয়। স্পষ্টতই প্রতিটি আয়তক্ষেত্রের উচ্চতা সংশ্লিষ্ট পরিসংখ্যা ঘনত্বের শ্রেণী-পরিসংখ্যার সমানুপাতী হয়। সুতরাং শ্রেণী অন্তর অসম দৈর্ঘ্যের হলেও কোনো অসুবিধা হয় না। এ ক্ষেত্রেও আয়তচিত্রে আয়তক্ষেত্রগুলি পরস্পর তুলনীয় হয়। নিচে একটি উদাহরণ দেওয়া হল।

সারণী—১২

145 জন কলেজ ছাত্রের ওজনের পরিসংখ্যা বিভাজন

ওজন (কেজি)	ছাত্রসংখ্যা (পরিসংখ্যা)
40-44	10
45-49	26
50-54	47
55-59	39
60-64	16
65-69	7
মোট	145



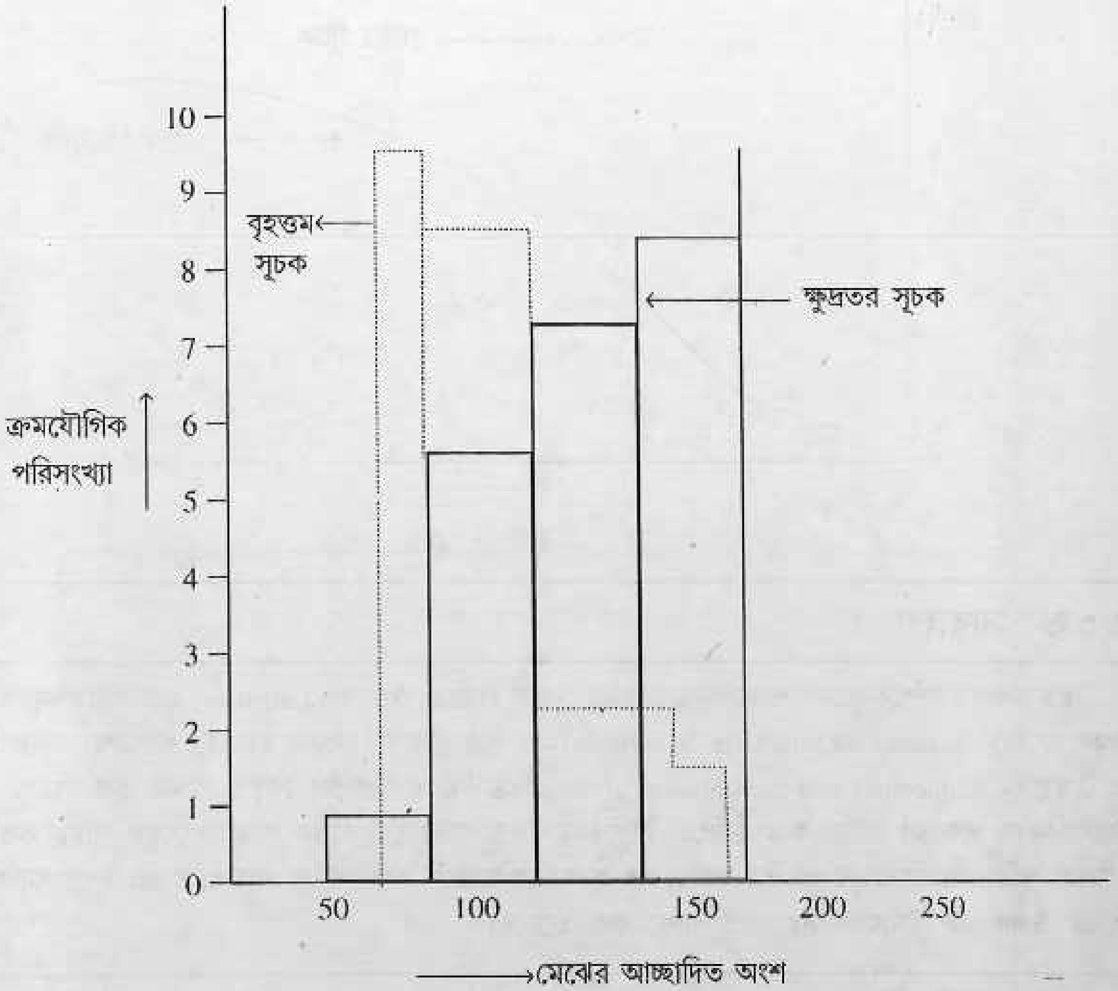
চিত্র নং—১.১০

2425 জন ছাত্রের ওজনের পরিসংখ্যা বিভাজনের আয়তচিত্র।

(গ) **ক্রমযৌগিক পরিসংখ্যারেখা (ogive) :**

(অ) **বিচ্ছিন্ন চল :** চলটির বিভিন্ন বিপরীতে উল্লম্ব অক্ষে ক্রমযৌগিক (ক্ষুদ্রতর বা বৃহত্তর) পরিসংখ্যা সূচক বিন্দুগুলি সুবিধামতো স্কেলে সংস্থাপন করা হয়। বিচ্ছিন্ন চলে যেহেতু দুটি সন্নিহিত মানের মধ্যে একটি ফাঁক থাকে (অর্থাৎ চলটির কোনো মান থাকে না) সেহেতু ঐ মধ্যবর্তী স্থানে ক্রমযৌগিক পরিসংখ্যা অপরিবর্তিত

থাকে। ফলে লেখচিত্রে ক্রমযৌগিক পরিসংখ্যা রেখা অঙ্কনের সময় সন্নিহিত বিন্দুগুলিকে সোপানচিত্রের (step diagram) মতো করে যুক্ত করা হয়। নিচের চিত্রটি দ্রষ্টব্য–

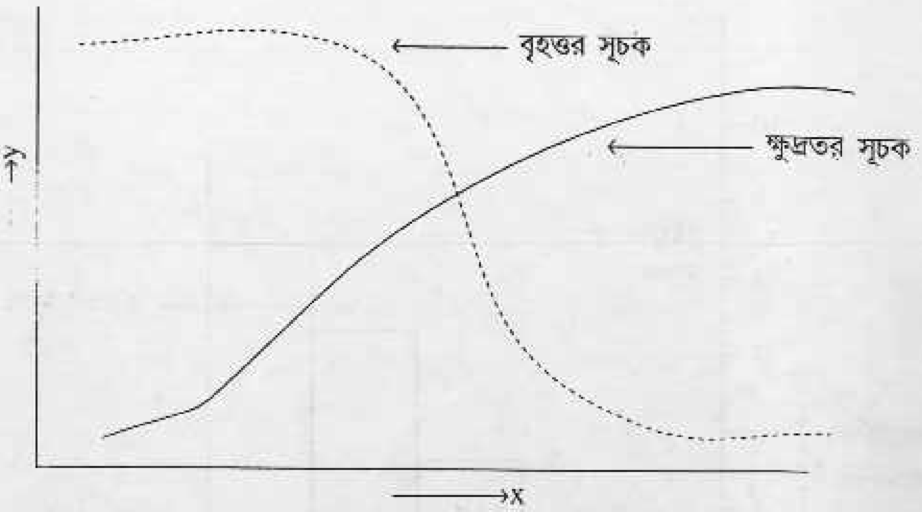


(সারণী ১০ দ্রষ্টব্য)

চিত্র নং—১.১১

(আ) **অবিচ্ছিন্ন চল :** অবিচ্ছিন্ন চলের ক্ষেত্রেও বিভিন্ন মানের বিপরীতে উল্লম্ব অক্ষে ক্রমযৌগিক (ক্ষুদ্রতর বা বৃহত্তর) বিন্দুগুলি সুবিধামতো স্কেলে সংস্থাপন করা হয়। এই ক্ষেত্রে যেহেতু চলটি অবিচ্ছিন্ন, সেজন্য আয়তচিত্রের মতো এখানেও অনুভূমিক অক্ষটিতে শ্রেণীসীমান্ত নির্দেশিত করে চলটির অবিচ্ছিন্নতা রক্ষা করা হয়। এরপর ক্ষুদ্রতর সূচক পরিসংখ্যাগুলি যথাক্রমে শ্রেণীগুলির ঊর্ধ্বসীমান্ত বা অর্ধসীমান্তের বিপরীতে সংস্থাপন করা হয়। এরপর সংস্থাপিত বিন্দুগুলি সরলরেখার সাহায্যে একটি (ক্ষুদ্রতর বা বৃহত্তর) পরিসংখ্যারেখা অঙ্কন করা হয়। শ্রেণীসীমান্তগুলি ক্ষুদ্রতর বা বৃহত্তর ক্রমযৌগিক পরিসংখ্যার ক্ষেত্রে একই থাকার জন্য একটি

লেখচিত্রেই এই দুরকম রেখাচিত্র পরিস্থাপন করা যেতে পারে। নিচের চিত্রটিকে পরিসংখ্যারেখার রূপটি দেখানো হল।



চিত্র নং—১.১২

৩৩.৬ সারাংশ

এই অধ্যায়ে রাশিবিজ্ঞানের সংজ্ঞাসমূহের ব্যাখ্যা দেওয়া হয়েছে। গুণলক্ষণ (attribute) এবং পরিমাণসূচক লক্ষণ বা চল (quantitative character or variable)-এর বিস্তৃত ব্যাখ্যা দেওয়া হয়েছে। রাশিতথ্য আহরণ ও উপস্থাপন (collection and presentation of data)-এর বিভিন্ন প্রণালীর বিস্তৃত ব্যাখ্যা করা হয়েছে। সারণীবিন্যাস পদ্ধতির ব্যাখ্যা করা হয়েছে। রাশিতথ্য উপস্থাপনের চিত্রভিত্তিক পদ্ধতিসমূহের ব্যাখ্যা করা হয়েছে। পরিসংখ্যা বিভাজন পদ্ধতির ব্যাখ্যা, এর ব্যবহারিক প্রয়োগ এবং লৈখিক পদ্ধতিতে এর উপস্থাপনার বিভিন্ন উপায়গুলি উদাহরণ সহযোগে ব্যাখ্যা করা হয়েছে।

৩৩.৭ অনুশীলনী

১। বিচ্ছিন্ন চল এবং অবিচ্ছিন্ন চলের পার্থক্য কি? উদাহরণ সহকারে ব্যাখ্যা করুন।

২। রাশিতথ্য সংক্ষেপণের প্রয়োজন হয় কেন? এই সংক্ষেপীকরণে পরিসংখ্যা বিভাজন কিভাবে সাহায্য করে?

৩। পরিসংখ্যা বিভাজনের লৈখিক উপস্থাপন পদ্ধতিগুলি বর্ণনা করুন।

৪। রাশিতথ্য উপস্থাপনের বিভিন্ন পদ্ধতি বর্ণনা করুন।

৫। সারণীবিন্যাসের দরকার হয় কেন? ছোট উদাহরণ সহকারে ব্যাখ্যা করুন।

৬। **নিম্নলিখিতগুলি ব্যাখ্যা করুন :**

(১) রেখাচিত্র, (২) স্তম্ভচিত্র, (৩) বৃত্তচিত্র।

৭। ক্রমযৌগিক পরিসংখ্যারেখা কাকে বলে? উদাহরণসহ ব্যাখ্যা করুন।

৮। আয়তচিত্র ও স্তম্ভচিত্রের মধ্যে পার্থক্য কি? ব্যাখ্যা করুন।

৯। **নিম্নের চলগুলির মধ্যে কোন্‌গুলি বিচ্ছিন্ন ও কোন্‌গুলি অবিচ্ছিন্ন তা নির্দেশ করুন :**

বাক্যপ্রতি শব্দসংখ্যা, গৃহে আবাসিকদের সংখ্যা, ছাত্রদের ওজন, কারখানায় প্রস্তুত পেন্সিলের দৈর্ঘ্য, প্রতি একশো আমের ঝুড়িতে পচা আমের অনুপাত, ব্যবসায় প্রতিষ্ঠানের কর্মীদের মাসিক আয়।

১০। কলকাতার একটি কলেজে সাম্মানিক রাশিবিজ্ঞানের প্রথম বার্ষিক শ্রেণীর ছাত্রদের পরীক্ষার ফলের ভিত্তিতে গ্রেড অনুমানে যে পরিসংখ্যা বিভাজন পাওয়া গেছে তা নিচে দেখানো হল :

গ্রেড	পরিসংখ্যা
A	5
B	8
C	5
D	3
E	1
মোট	22

(ক) এখানে যে লক্ষণটির কথা বলা হল তা গুণগত না পরিমাণগত বলুন।

(খ) প্রত্যেক শ্রেণীর জন্য আনুপাতিক পরিসংখ্যা বের করুন। অন্তত B গ্রেড পেয়েছে এমন ছাত্রদের সংখ্যা কত?

(গ) পরিসংখ্যা ও আনুপাতিক সংখ্যাগুলিকে উপযুক্ত চিত্রের মাধ্যমে দেখান।

১১। ঐ ছাত্রদের প্রত্যেকের পরিবারের আয়তন নিচে দেখানো হল :

3	4	4	7	8
4	3	5	4	5
6	3	3	6	
2	4	4	3	
4	5	7	5	

এগুলো নিয়ে একটি পরিসংখ্যা সারণী তৈরি করে পরিসংখ্যা, আনুপাতিক পরিসংখ্যা এবং উভয় ধরনের ক্রমযৌগিক পরিসংখ্যা দেখাতে হবে।

১২। (ক) পূর্ববর্তী অনুশীলনীর বিভাজনকে এবারে উপযুক্ত চিত্রের মাধ্যমে উপস্থাপন করুন।

(খ) কতগুলি পরিবারের সদস্য সংখ্যা 5 ?

অন্তত 5 সদস্য বিশিষ্ট পরিবারের সংখ্যা কত?

অনূর্ধ্ব 4 সদস্য বিশিষ্ট পরিবারের সংখ্যা কত?

শতকরা কটি পরিবারে 4 জন সদস্য রয়েছে?

১৩। একটি হাসপাতাল থেকে নবজাত 40টি শিশুর ওজন সংক্রান্ত নিচের পরিসংখ্যান পাওয়া গেছে :

নবজাত শিশুর ওজন (কেজি)

3.0	2.8	3.0	2.7	2.5	3.1	3.6	3.2	3.4	3.3
3.2	2.9	3.0	3.3	3.1	2.8	2.7	2.6	3.2	2.9
2.9	3.5	3.3	3.4	3.5	3.3	3.1	3.0	2.7	3.1
3.7	3.1	3.0	2.8	2.7	3.1	2.6	2.9	3.1	2.8

এই পরিসংখ্যার ভিত্তিতে সাতটি শ্রেণী নিয়ে পরিসংখ্যা বিভাজন তৈরি করতে হবে। সারণীতে প্রকৃত শ্রেণী সীমা, পরিসংখ্যা, আনুপাতিক পরিসংখ্যা ও উভয় প্রকারের ক্রমযৌগিক পরিসংখ্যা দেখাবেন। (2.4—2.5, 2.6—2.7, ………, 3.6—3.7 এই সাতটি শ্রেণী নিয়ে কাজ শুরু করতে পারেন।

১৪। পূর্বের অনুশীলনীতে যে পরিসংখ্যা-বিভাজন পাওয়া গেছে তাকে যথাযথ চিত্রের সাহায্যে উপস্থাপন করতে হবে। এজন্য আয়তচিত্র এবং উভয় প্রকারের 'ওজিভ' আঁকুন।

১৫। নবজাত শিশুদের ওজন-সংক্রান্ত পরিসংখ্যা সারণীর ভিত্তিতে নিচের প্রশ্নগুলির উত্তর তৈরি করুন :

(ক) শতকরা কতটি শিশুর ওজন 2.45 কেজি থেকে 3.45 কেজির মধ্যে রয়েছে?

(খ) কতজন শিশুর ওজন 2.75 কেজির নিচে রয়েছে?

(গ) কতজন শিশুর ওজন 3.25 কেজি বা ততোধিক?

একক ৩৪ ❑ মধ্যগামিতা ও মধ্যগামিতার মাপক

গঠন

৩৪.০ উদ্দেশ্য

এই এককটি পাঠ করার পর আপনি জানতে পারবেন—

- গড় কাকে বলে ও তার প্রকারভেদ।
- মধ্যমা, ভগ্নাংশ ও ভূরিষ্ঠক কাকে বলে।
- মধ্যগামিতার আরও কিছু মাপকগুলি কি কি?

৩৪.১ প্রস্তাবনা

পরিসংখ্যা বিভাজন, সারণীবিশুদ্ধ ইত্যাদির সাহায্যে প্রাথমিক সূত্রে প্রাপ্ত একগুচ্ছ অবিন্যস্ত রাশিতথ্যের সংক্ষেপীকরণ করা হয়। এর উদ্দেশ্য হল ঐ রাশিতথ্যের অন্তর্নিহিত জ্ঞাতব্য বিষয়টি অনুধাবন করা। কিন্তু এটি

হল প্রথম ধাপ। এই সংক্ষেপীকরণ প্রক্রিয়াটি এমন একটি পর্যায়ে নিয়ে যাওয়া যায় যার ফলে দুয়েকটি পরিমাপের মাধ্যমে সম্পূর্ণ পরিসংখ্যানগুচ্ছের বেশ কিছু বৈশিষ্ট্য সম্পর্কে ধারণা করা সম্ভব হয়। এ ধরনের মাপকের মধ্যে প্রথমেই উল্লেখ করতে হয় মধ্যগামিতার মাপকগুলি (Measures of Central Tendency)-র কথা।

৩৪.২ গড় (average)

কোনো চলকের অবিন্যস্ত অথবা গোষ্ঠীবদ্ধ (অর্থাৎ পরিসংখ্যা সারণীতে বিন্যস্ত) রাশিতথ্য থেকে সাধারণত দেখা যাবে যে রাশিগুলি সব সমান না হলেও তাদের মধ্যে কোনো কেন্দ্রীয় মানের কাছাকাছি থাকার প্রবণতা রয়েছে। এই প্রবণতাকে বলা হয় চলকের মধ্যগামিতা। মধ্যগামিতা নির্দেশক কেন্দ্রীয় মান পাওয়ার জন্য বিশেষ গাণিতিক পদ্ধতি ব্যবহার করা হয়। এই মধ্যগামিতা মাপকদের (measures of central tendency) আবার বলা হয় গড়মান (average)। গড় নানা প্রকারের হয়। যেমন, যৌগিক গড় (arithmetic mean), মধ্যমা (median), ভূয়িষ্ঠক (mode), গুণোত্তর গড় (geometric mean) ও বিবর্ত যৌগিক গড় (harmonic mean) । পরবর্তী অনুচ্ছেদগুলিতে এদের নিয়ে আলোচনা হবে।

৩৪.২.১ যৌগিক গড় (arithmetic mean)

প্রদত্ত মানগুলির যোগফলকে মানের সংখ্যা দিয়ে ভাগ করে যৌগিক গড় পাওয়া যায়। সূত্রে প্রকাশ করলে :—

ধরা যাক, X একটি চল যার n-সংখ্যক মান আছে। এগুলি হল $x_1, x_2, \ldots\ldots\ldots x_n$। তা হলে x-এর গাণিতিক গড় (যাকে সাধারণত $\overline{x}$ দিয়ে সূচিত করা হয়) হল :

$$\overline{x} = \frac{1}{n}(x_1 + x_2 + \ldots\ldots\ldots x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i$$

এখানে Σ-চিহ্নটি যোগফলের প্রতীক (symbol)° এটি গ্রীক অক্ষর, বড় হরফের 'সিগ্‌মা'।

বিচ্ছিন্ন চলের ক্ষেত্রে এক-একটি মানকে এক-একটি শ্রেণীর পরিচায়ক হিসাবে ধরলে প্রতিটি মানের সাথে সংশ্লিষ্ট একটি পরিসংখ্যা (fi) থাকবে। ধরা যাক, আমাদের সংখ্যা তথ্য থেকে x চলটির জন্য যে পরিসংখ্যা সারণী তৈরি করা হয়েছে, সেখানে i-তম শ্রেণীর x-এর মান x_i এবং সংশ্লিষ্ট পরিসংখ্যা হল f_i।

এক্ষেত্রে $\overline{x} = \frac{1}{n}\sum_{i=1}^{k} x_i f_i$

যেখানে $k =$ শ্রেণীসংখ্যা এবং $n = \sum_{i=1}^{k} f_i$ অর্থাৎ মোট পরিসংখ্যা।

অবিচ্ছিন্ন চলের পরিসংখ্যা সারণীর ক্ষেত্রেও আমরা একই সূত্র ব্যবহার করব। শুধু এক্ষেত্রে x_i দিয়ে আমরা i-তম শ্রেণীর মধ্যবিন্দু (class mark) সূচিত করব। তবে মনে রাখতে হবে যে এভাবে আমরা যৌগিক গড়ের শুধু আসন্ন মানই পাওয়ার আশা করতে পারি।

উদাহরণ ২.১। সাধারণ গাণিতিক গড়।

সারণী—১

৮ জন ছাত্রের প্রাপ্ত অঙ্কের নম্বর

ছাত্রের ক্রমিক সংখ্যা	অঙ্কের নম্বর (মোট নম্বর = ১০০)
1	30
2	45
3	80
4	50
5	65
6	12
7	90
8	72

এখানে $\overline{x} = \dfrac{\sum_{i=1}^{n} x_i}{n} = \dfrac{30 + 45 + \ldots\ldots\ldots\ldots + 72}{8}$

$\dfrac{444}{8} = 55.50$

অর্থাৎ এই আটজন ছাত্র গড়ে (100-র মধ্যে) 55.50 নম্বর করে পেয়েছে।

উদাহরণ ২.২। পরিসংখ্যাযুক্ত গাণিতিক গড়

সারণী—২

ঘরের সংখ্যা অনুযায়ী **100টি** বাড়ির পরিসংখ্যা বিভাজন

ঘরের সংখ্যা (চলের মান = x_i)	বাড়ীর সংখ্যা (পরিসংখ্যা = f_i)	$x_i f_i$
1	8	8
2	18	36
3	30	90
4	25	100
5	19	95
6	10	60
7	8	56
8	2	16
মোট	120	461

সুতরাং $\overline{x} = \frac{461}{120} = 3.842$ গড়ে

অর্থাৎ গড়ে প্রতিটি বাড়িতে ঘরের সংখ্যা = 3.842 (আসন্ন মানে = 4)

৩৪.২.২ যৌগিক গড়ের কিছু বৈশিষ্ট্য :

(ক) যদি চলের মানগুলি অভিন্ন হয় অর্থাৎ ধ্রুবক (constant)-এর সমান হয়, তা হলে চলটির যৌগিক গড়ও ঐ ধ্রুবকের সমান হবে।

প্রমাণ :—ধরা যাক, $x_i = a$, যেখানে a একটি ধ্রুবক। x-এর n-সংখ্যক মান আছে। $(x_1, x_2 \ldots\ldots x_n)$।
সুতরাং, $x_1 = x_2 = \ldots\ldots = x_n = a$

$$\text{এবং } \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{na}{n} = a$$

(খ) কোনো চলের যৌগিক গড় থেকে ঐ চলের প্রদত্ত মানগুলির বিচ্যুতির যোগফল শূন্য হবে।

প্রমাণ : প্রদত্ত পরিসংখ্যানে x-এর i-তম মান x_i ধরে নিলে, সংশ্লিষ্ট বিচ্যুতি হবে

$x_i - \overline{x}$ $(i = 1, 2, \ldots\ldots, n)$।

এখন $\displaystyle\sum_{i=1}^{n}(x_i - \overline{x}) = \sum_{i=1}^{n}x_i - n\overline{x} = 0$, কারণ $\displaystyle\overline{x} = \frac{1}{n}\sum_{i=1}^{n}x_i$

এবং তাই $\displaystyle\sum_{i=1}^{n}x_i = n\overline{x}$

(গ) চলের রৈখিক রূপান্তর (linear transformation) করা হলে রূপান্তরিত চলের গড় ও মূল চলের গড় অনুরূপভাবে সম্বন্ধযুক্ত হবে।

অর্থাৎ $y = a + bx$ হয়, যেখানে a ও b দুটি ধ্রুবক, তা হলে

$$\overline{y} = a + b\overline{x}$$

প্রমাণ : উপরোক্ত সমীকরণে স্পষ্টতই দেখা যাচ্ছে,

$y_i = a + bx_i$ (প্রতি i-এর জন্য)

সুতরাং, $\displaystyle\overline{y} = \frac{1}{n}\sum_{i=1}^{n}y_i = \frac{1}{n}\sum_{i=1}^{n}(a + bx_i)$

$$= \frac{1}{n}\left(na + b\sum_{i=1}^{n}x_i\right) = a + b \times \frac{1}{n}\sum_{i=1}^{n}x_i = a + b\overline{x}$$

দ্রষ্টব্য : চলকের এই ধর্ম ব্যবহার করে আমরা যৌগিকের গড়ের মান সহজে নির্ণয় করতে পারি। প্রারম্ভিক চল x-এর পরিবর্তে আমরা নিতে পারি : $u = (x - a)/c$, যেখানে a ও c-কে যথাযথভাবে বেছে নিতে হবে। আমরা প্রথমে y-এর যৌগিক গড় $\overline{y}$ বের করব। কিন্তু যেহেতু $x = a + iy$ আমরা সহজেই পাব $\overline{x} = a + c\overline{u}$ ।

এখানে রূপান্তরকের মাপনার মূলবিন্দু (origin) এবং মাত্রা (scale)-এর পরিবর্তন করা হয়েছে। চলকের মূলবিন্দু 0 থেকে a-তে এবং মাত্রা 1 থেকে c-তে পরিবর্তিত হয়েছে। ক্ষেত্রবিশেষে তবু মূলের বা শুধু মাত্রার পরিবর্তন সমীচীন মনে হতে পারে।

উদাহরণ ২.৩ :

সারণী—৩

120টি বাড়িতে ঘরের সংখ্যা

ঘরের সংখ্যা (x_i)	বাড়ির সংখ্যা (পরিসংখ্যা = f_i)	$u_i = x_i - 4$	$y_i f_i$
1	8	–3	–24
2	18	–2	–36
3	30	–1	–30
4	25	0	0
5	19	1	19
6	10	2	20
7	8	3	24
8	2	4	8
মোট	120	—	–19

উপরের গণনার ভিত্তিতে পাই , $\bar{u} = \frac{1}{120}(-19) = -0.158$

যেহেতু $x = 4 + \bar{u}$, আমরা তাই পেলাম, আগের মতোই, $\bar{x} = 4 - 0.158 = 3.842$. এখানে উল্লেখ্য যে, সাধারণভাবে গাণিতিক গড়ের যে উদাহরণ, (উদাহরণ ২.২) দেওয়া হয়েছে এবং বর্তমান উদাহরণটি একই উদাহরণ—কিন্তু বিভিন্ন উদ্দেশ্যে ব্যবহার করা হয়েছে।

উদাহরণ ২.৪ :

সারণী ৩ থেকে একটি ঘড়ির গড় বিক্রয়মূল্য নির্ণয় করুন।

সারণী—৪

৪0টি ঘড়ির বিক্রয়মূল্য

ঘড়ির দর (টাকায়) শ্রেণী-অন্তর	শ্রেণীমধ্যস্থ x_i	পরিসংখ্যা f_i	$u_i = \frac{x_i - 308}{25}$	$u_i f_i$
220.5–245.5	233	8	–3	–24
245.5–270.5	258	12	–2	–24
270.5–295.5	283	20	–1	–20
295.5–320.5	308	36	0	0
320.5–345.5	333	25	1	25
345.5–370.5	358	17	2	34
370.5–395.5	383	6	3	18
395.5–420.5	408	3	4	12
মোট	—	127	—	21

উপরের সারণী অনুসারে, $\bar{u} = \frac{1}{n} \sum_i u_i f_i = \frac{21}{127} = 0.1654.$

যেহেতু $x = a + cu = 308 + 25u$, সুতরাং

$$\bar{x} = a + c\bar{u} = 308 + 25\sqrt{0.1654} = 312.135$$

সুতরাং গড় দর = 312 টাকা 14 পয়সা।

(ঘ) ধরা যাক, n-সংখ্যক ব্যক্তির ক্ষেত্রে দুটি চল, x ও y, সম্বন্ধে তথ্য সংগ্রহ করা হয়েছে। এরূপ অবস্থায় যদি আরও একটি চল z-কে আনা হয়, যেখানে

$$z = ax + by \text{ (a ও b ধ্রুবক)},$$

তা হলে $\bar{z} = a\bar{x} + b\bar{y}$

প্রমাণ : এরূপ ক্ষেত্রে i-তম ব্যক্তির ক্ষেত্রে x_i, y_i ও z_i যথাক্রমে x, y ও z-এর মান সূচিত করলে

$z_i = ax_i + by_i$ যেখানে, $i = 1, 2 \ldots\ldots\ldots n$

সুতরাং, $$\sum_{i=1}^{n} z_i = a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} y_i$$

n-দ্বারা ভাগ করে পাই , $\overline{z} = a\overline{x} + b\overline{y}$

দ্রষ্টব্য : যদি $a = 1$ এবং $b = \pm 1$ হয়, অর্থাৎ $z = x \pm y$ হলে, $\overline{z} = \overline{x} \pm \overline{y}$ হবে।

স্পষ্টতই এই বৈশিষ্ট্যটি দুই-এর অধিক চলের ক্ষেত্রেও প্রযোজ্য।

(ঙ) যদি কোনো একটি চল x-এর n_1 ও n_2 সংখ্যক দুই গুচ্ছ মান থাকে যাদের গড় যথাক্রমে $\overline{x_1}$ ও $\overline{x_2}$, তা হলে এদের সম্মিলিত গড়ের (composite mean) মান হবে।

$$\overline{x} = \frac{n_1\overline{x_1} + n_2\overline{x_2}}{n_1 + n_2}$$

প্রমাণ : এখানে চলের প্রথম গুচ্ছে মানের যোগফল $n_1\overline{x_1}$ এবং দ্বিতীয় গুচ্ছে যোগফল $n_2\overline{x_2}$। সুতরাং এই দুটি গুচ্ছ একত্রে নিলে যোগফল হয় $n_1\overline{x_1} + n_2\overline{x_2}$ এবং এদের মোট সংখ্যা $n_1 + n_2$। সুতরাং সম্মিলিত যৌগিক গড়ের মান হবে—

$$\overline{x} = \frac{n_1\overline{x_1} + n_2\overline{x_2}}{n_1 + n_2}$$

[অনুরূপভাবে, k-সংখ্যক মানগুচ্ছ থাকলে সম্মিলিত যৌগিক গড় হবে—

$$\overline{x} = \frac{\sum_{i=1}^{k} n_i\ \overline{x_i}}{\sum_{i=1}^{k} n_i}$$]

৩৪.২.৩ মধ্যমা (Median) :

মধ্যমার সংজ্ঞা : কোনো চলের মানগুলিকে ঊর্ধ্বগ (বা নিম্নগ) ক্রম অনুসারে সমান দুভাগে ভাগ করা হলে ঠিক মধ্যস্থলে অবস্থিত মানটিকে মধ্যমা (median) বলা হয়। যদি মোট মান সংখ্যা "n" অযুগ্ম (odd) সংখ্যা হয়, তা হলে $(n + 1)/2$-তম মানটি মধ্যমা হবে। অপরপক্ষে মোট মান সংখ্যা "n" যদি যুগ্ম (even) সংখ্যা হয়, তা হলে $n/2$-তম এবং $(n/2 + 1)$-তম মানের মধ্যে অবস্থিত যে কোনো সংখ্যাকে মধ্যমা বলা যায়। এক্ষেত্রে অনেক সময় $n/2$ এবং $(n/2 + 1)$-তম মানদুটির গাণিতিক গড়কেই মধ্যমা বলে ধরা হয়।

উদাহরণ ২.৫ :

(ক) নিচে বিভিন্ন বাজার থেকে কেনা সাত কেজি আপেলের প্রতি কেজির দর দেওয়া হল। এদের মধ্যমা নির্ণয় করুন।

দর (টোকায়) 47, 49, 54, 44, 65, 51, 57। এখানে সাতটি অর্থাৎ বিযুগ্ম সংখ্যক মান আছে। তা হলে উর্ধ্বগ ক্রম হিসেবে সাজালে মানগুলি দাঁড়াবে 44, 47, 49, 51, 54, 57, 65। তাই মধ্যমা হল এই বিন্যাসের চতুর্থ মানটি, অর্থাৎ 51 টাকা।

বিচ্ছিন্ন চলের ক্ষেত্রে...............কী হত

(খ) নিচে বিভিন্ন ধরনের চালের দর কেজির মাপে দেখান আছে। দর (টাকায়) 89, 82, 75, 72, 70, 69, 52, 43, 40, 29। এখানে দশটি অর্থাৎ যুগ্ম সংখ্যক মান নিম্নগ ক্রমে দেওয়া আছে। এর মধ্যে মাঝখানের দুটি অর্থাৎ পঞ্চম এবং ষষ্ঠ মানের গাণিতিক গড়কে মধ্যমা ধরা যায়। সুতরাং $\frac{1}{2}(70 + 69) =$ 69.5 টাকা হবে মধ্যমার মান। (অবশ্য সংজ্ঞানুসারে 69 ও 70-এর মধ্যবর্তী যে কোনো মানকে মধ্যমা হিসাবে গণ্য করা যায়।) ওপরের উদাহরণ দুটিতে প্রতিটি মূল্যমান দেওয়া আছে। এদের সংখ্যাও কম। এসব ক্ষেত্রে খুব সহজেই উল্লিখিত পদ্ধতিতে মধ্যমা নির্ণয় করা যায়। কিন্তু অবিচ্ছিন্ন চলের ক্ষেত্রে, অর্থাৎ যেখানে প্রতিটি মান জানা থাকেনা এবং পরিসংখ্যা বিভাজন পদ্ধতিতে সবগুলির মানের পরিসংখ্যা কয়েকটি শ্রেণীতে ভাগ করে দেখান হয়, সেখানে উল্লিখিত পদ্ধতিতে মধ্যমা নির্ণয় করা সম্ভব হয় না। এরূপ ক্ষেত্রে অন্য পদ্ধতি অবলম্বন করা হয়। এ নিয়ে নিচে আলোচনা করা হল।

অবিচ্ছিন্ন চলের ক্ষেত্রে মধ্যমা নির্ণয় করতে হলে চলের সেই বিশেষ মানটির কথা ভাবতে হবে, যার ক্রমযৌগিক পরিসংখ্যা (cumulative frequency) ঠিক $\frac{n}{2}$। অর্থাৎ মোট পরিসংখ্যা n হলে, যে মানটির ক্রমযৌগিক পরিসংখ্যা $\frac{n}{2}$ সেটিই হবে মধ্যমা। এই মানটি দুরকমভাবে নির্ণয় করা যেতে পারে : (1) রৈখিক পদ্ধতিতে বা (2) রৈখিক অন্তঃপ্রক্ষেপণ (linear interpolation)-এর সাহায্যে।

(1) **রৈখিক পদ্ধতি :** এই পদ্ধতিতে পরিসংখ্যা বিভাজনের ক্রমযৌগিক পরিসংখ্যান চিত্র (বা ওজিভ) আঁকা হবে। $\frac{n}{2}$ উচ্চতায় অনুভূমিক অক্ষের সমান্তরাল একটি সরলরেখা টানা হলে তা যে বিন্দুতে ওজিভকে ছেদ করবে, সে-বিন্দুর (abscissa)।

(2) **রৈখিক অন্তঃপ্রক্ষেপণ পদ্ধতি :**

এই পদ্ধতিতে পরিসংখ্যা সারণীটি পরীক্ষা করে প্রথমে বিভাজনের সেই শ্রেণীটি নির্দিষ্ট করতে হবে যেখানে মধ্যমাটি অবস্থিত। এটি ক্রমযৌগিক পরিসংখ্যা সারণীটি পরীক্ষা করে (রৈখিক পদ্ধতির বর্ণনা দ্রষ্টব্য) সহজেই নির্দিষ্ট করা যায়। ধরা যাক, যে শ্রেণীতে মধ্যমাটি অন্তর্ভুক্ত তার নিম্ন এবং উর্ধ্ব শ্রেণীসীমান্ত দুটি যথাক্রমে x_l এবং x_u। আমরা ধরে নিচ্ছি যে x_l থেকে x_u পর্যন্ত পরিসংখ্যার যে বৃদ্ধিটুকু হয়েছে তা

সরলরৈখিক (linear)। তা হলে রৈখিক প্রক্ষেপণ পদ্ধতি (linear, interpolation method)-র সূত্র অনুসরণ করে যে মানের ক্রমযৌগিক পরিসংখ্যা ঠিক $\frac{n}{2}$ অর্থাৎ মধ্যমা হবে।

$$M_i = x_l + \frac{\frac{n}{2} - n_l}{n_u - n_l}(x_u - x_l)$$

$$= x_l + \frac{\frac{n}{2} - n_l}{f_0} \times C$$

যেখানে, M_i = মধ্যমা

$C = (x_u - x_l)$ = ঐ শ্রেণীঅন্তরের দৈর্ঘ্য

$f_0 = (n_u - n_l)$ = ঐ শ্রেণীঅন্তরের পরিসংখ্যা

উদাহরণ : সারণী 2.2-তে আমরা যে পরিসংখ্যা বিভাজন নিয়েছিলাম তার জন্য দু-ধরনের ক্রমযৌগিক পরিসংখ্যা নিচের সারণীতে দেখানো হল।

বাড়িতে ঘরের সংখ্যা	ক্রমযৌগিক পরিসংখ্যা	
	অনধিক ধাঁচের (less than)	অন্যূন ধাঁচের (more than)
1	8	20
2	26	112
3	56	94
4	81	64
5	100	39
6	110	20
7	118	10
8	120	2

এখানে অনধিক ধাঁচের ক্রমযৌগিক পরিসংখ্যা থেকে বোঝা যাবে চলকের সংশ্লিষ্ট মান বা তার অনধিক মান কতটি ব্যষ্টির ক্ষেত্রে প্রযোজ্য। তাই প্রারম্ভিক তথ্যকে মানের ঊর্ধ্বগ ক্রমে সাজালে প্রথম ৪টি মান হত 1, নবম থেকে 26-তম মান হত 2, 27-তম থেকে 56-তম মান হত 3 ইত্যাদি। তাই এই বিন্যাসের ঠিক মধ্যবর্তী মান দুটিই 4 হত। ফলে আমরা 4-কেই মধ্যমা বলব।

অন্যূন ধাঁচের ক্রমযৌগিক পরিসংখ্যা থেকে অনুরূপভাবে প্রারম্ভিক তথ্যের নিম্নগ ক্রমে বিন্যাসের চিত্রটি ফুটে উঠবে। এক্ষেত্রে প্রথম দুটি মান হত 8, তৃতীয় থেকে দশম পর্যন্ত মানগুলি হত 7, একাদশ থেকে বিংশ মান হত 6 ইত্যাদি। তাই এক্ষেত্রে 40-তম থেকে 64-তম মানগুলি হত 4 ; ফলে 4-কেই এদিক থেকেও মধ্যমা হিসাবে ধরতে হয়।

উদাহরণ : সারণী 2-তে ঘড়ির দর সংক্রান্ত বিভাজনের ক্রমযৌগিক পরিসংখ্যা দেখানো হল :

ঘড়ির দর (টাকায়)	ক্রমযৌগিক পরিসংখ্যা	
	অনধিক ধাঁচের (less than)	অন্যূন ধাঁচের (more than)
220.5–245.5	8	127
245.5–270.5	20	119
270.5–295.5	40	107
295.5–320.5	76	87
320.5–345.5	101	51
345.5–370.5	118	26
370.5–395.5	124	9
395.5–420.5	127	3

আমরা অনধিক ধাঁচের ক্রমযৌগিক পরিসংখ্যার উপর দৃষ্টি নিবদ্ধ করতে পারি। এক্ষেত্রে $n = 127$: যেহেতু $\frac{n}{2} = 63.5$ এবং এ-সংখ্যাটি 40 ও 76-এর মধ্যবর্তী, তাই মধ্যমা 295.5 – 320.5 এই অন্তরটিতে রয়েছে।

এখন মধ্যমার সূত্রে আমাদের বসাতে হবে—

$x_l = 295.5,$

$n_l = 40,$

$f_0 = 76 - 40 = 36$

$c = 320.5 - 295.5 = 25$

তাই নির্ণেয় মধ্যমা হবে

$$\check{M}_l = 295.5 + \frac{63.5 - 40}{36} \times 25$$

$$= 295.5 + 16.319 = 311.82 \text{ টাকা।}$$

৩৪.২.৪ ভগ্নাংশক (Quantiles বা Fractiles) :

[চলের মানগুলি (উর্ধ্বগ বা নিম্নগ) ক্রম অনুসারে সাজানো হলে যে মানটি বিভাজনকে $\frac{p}{(1-p)}$ অনুপাতে ভাগ করে, তাকে চলের p-তম ভগ্নাংশক (pth quantile বা fractile) বলা হয়। মধ্যমার ক্ষেত্রে বিভাজনটি $\frac{1}{2}$ অনুপাতে ভাগ করা হয়, মধ্যমার মতো এখানেও অবিচ্ছিন্ন চলের ক্ষেত্রে রৈখিক পদ্ধতি বা রৈখিক অন্তঃপ্রক্ষেপণ (interpolation) পদ্ধতিতে ভগ্নাংশ নির্ণয় করা যেতে পারে। শুধুমাত্র খেয়াল রাখতে হবে মধ্যমার ক্ষেত্রে $\frac{n}{2}$-র জায়গায় p-তম ভগ্নাংশকের ক্ষেত্রে $\frac{n}{p}$ মাপটি নিতে হবে। ভগ্নাংশকের মধ্যে মধ্যমার ব্যবহার সবচাইতে বেশি। এছাড়া **'প্রথম চতুর্থক ভগ্নাংশক'** (অর্থাৎ $p = \frac{1}{4}$) বা **'প্রথম চতুর্থক (1st quartile)'** এবং **'তৃতীয় চতুর্থক ভগ্নাংশক'** বা **'তৃতীয় চতুর্থক' ($p = \frac{3}{4}$, 3rd quartile)'**-এর ও কোনো কোনো বিশেষ ক্ষেত্রে প্রচলন আছে। বলাবাহুল্য, মধ্যমা হল **'দ্বিতীয় চতুর্থক ($p = \frac{1}{2}$) ভগ্নাংশক'** ।এছাড়া **দশমক (যেখানে $p = \frac{1}{10}$) এবং শততমক যেখানে ($p = \frac{1}{100}$)-এর ব্যবহারও অনেক ক্ষেত্রে করা হয়।**]

৩৪.২.৫ ভূয়িষ্ঠক (Mode) :

ভূয়িষ্ঠক বলতে চলের সেই বিশেষ মানটিকে চিহ্নিত করা হয় যার পরিসংখ্যার পরিমাণ সবচাইতে বেশি। স্পষ্টতই এই সংজ্ঞাটি বিচ্ছিন্ন চলের ক্ষেত্রে সহজে বোধগম্য। অবিচ্ছিন্ন চলের ক্ষেত্রে ভূয়িষ্ঠক বলতে আমরা বোঝাতে চাই সেই মান যার পরিসংখ্যা-ঘনত্ব বৃহত্তম। মধ্যমার মতো এখানেও রৈখিক পদ্ধতিতে বা রৈখিক অন্তঃপ্রক্ষেপণ পদ্ধতিতে ভূয়িষ্ঠকের মান নির্ণয় করা যেতে পারে। রৈখিক পদ্ধতির ক্ষেত্রে পরিসংখ্যা বিভাজনের আয়তক্ষেত্র পর্যবেক্ষণ করে মান নির্ণয় করা যায়। রৈখিক অন্তঃপ্রক্ষেপণ পদ্ধতিতে প্রথমে বিভাজনের সেই শ্রেণীটি পর্যবেক্ষণের মাধ্যমে খুঁজে বের করতে হবে যার পরিসংখ্যার পরিমাণ সর্বাধিক। তারপর ঐ শ্রেণীটির উচ্চসীমান্ত (x_u) এবং নিম্নসীমান্ত (x_l) দুটির মান নির্ণয় করতে হবে। এরূপ ক্ষেত্রে ভূয়িষ্ঠকের আসন্ন মান (M_0) হবে,

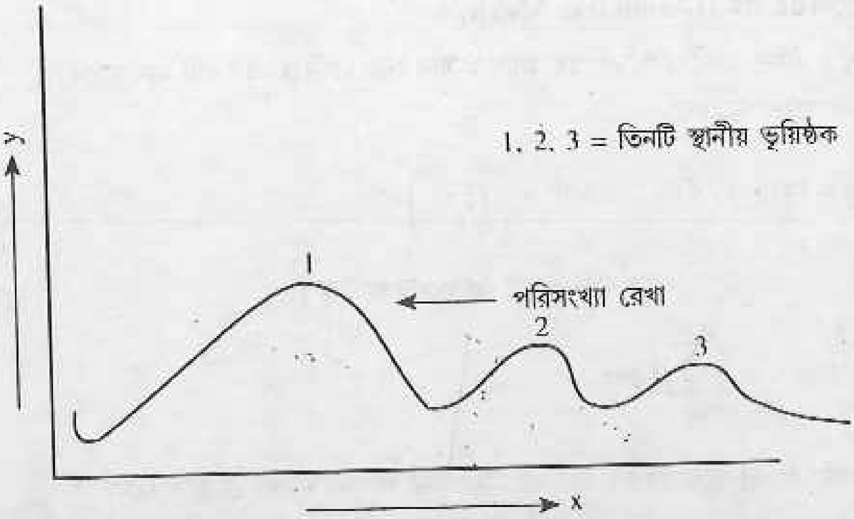
$$M_0 = \frac{X_l + X_u}{2} + \frac{C}{2}$$

যেখানে, C = শ্রেণী-অন্তরের দৈর্ঘ্য

M_0 = ভূয়িষ্ঠক

কোনো কোনো চলের ক্ষেত্রে পরিসংখ্যা রেখা স্থানীয়ভাবে (locally) একাধিক স্থানে বৃহত্তম উচ্চতা (maxima) প্রাপ্ত হয়। অবিচ্ছিন্ন চলের ক্ষেত্রে এটা খুব স্পষ্টভাবে পরিলক্ষিত হয় (নিচের চিত্র দ্রষ্টব্য)। দুই বা ততোধিক বৃহত্তম উচ্চতার পরিসংখ্যা রেখাসম্পন্ন বিভাজনকে যথাক্রমে দ্বিভূয়িষ্ঠক (bi-modal) বা

বহুভূয়িষ্ঠক (multimodal) পরিসংখ্যা বিভাজন বলা হয়। তবে এক্ষেত্রে যে মানের পরিসংখ্যা-ঘনত্ব সার্বিকভাবে (globally) বৃহত্তম (যেমন নিচের চিত্রে x_1) তাকেই আমরা ভূয়িষ্ঠক বলব।



চিত্র নং—২.১

৩৪.৩ গাণিতিক গড়, মধ্যমা এবং ভূয়িষ্ঠকের তুলনা এবং সম্পর্ক

(1) যে কোনো ধরনের মাপকের সংজ্ঞা সুনির্দিষ্ট হওয়া দরকার। উল্লিখিত মধ্যগামিতার তিনটি মাপকই এই শর্ত পূরণ করে।

(2) অধিকাংশ ক্ষেত্রেই উল্লিখিত মাপকগুলি সহজেই নির্ণয় করা সম্ভব। তবে অবিচ্ছিন্ন চলকের ভূয়িষ্ঠকের ক্ষেত্রে রাশিতথ্যের পরিমাণ প্রচুর না হলে গণনায় কিছুটা অসুবিধা হয়।

(3) একটি ভাল মাপক প্রতিটি নিরীক্ষিত মানের উপর নির্ভরশীল হওয়া উচিত। সাধারণভাবে উপরোক্ত তিনটি মাপকই এই নিয়ম অনুসরণ করে। তবে গাণিতিক গড়ের ক্ষেত্রে কোনো একটি নিরীক্ষিত মানের হেরফের হলেই নিচের মানও পরিবর্তিত হয়।

(4) একটি ভাল মাপকের কিছু সরল বীজগাণিতিক নিয়ম (algebraic properties) অনুসরণ করা উচিত। এতে পরবর্তী গণনাগুলিতে সুবিধা হয়। এদিক থেকে গাণিতিক গড়ের সুবিধা অন্যদুটি মাপকের তুলনায় অনেক বেশি। এসব কারণে যৌগিক গড়টির প্রচলনও অপেক্ষাকৃত ব্যাপক। আমরা সাধারণভাবে যৌগিক গড়ের ব্যবহারই শ্রেয় মনে করি। তবে কোনো ক্ষেত্রে যদি দেখা যায় যে পরিসংখ্যান সমষ্টিতে অল্প কয়েকটি মান আছে, যা অন্য মানের তুলনায় অত্যন্ত বড় বা অত্যন্ত ছোট, তাহলে যৌগিক গড়কে প্রতিনিধি হিসাবে গ্রহণ করা ঠিক হবে না ; বরং মধ্যমার ব্যবহার এক্ষেত্রে সঙ্গত হবে।

৩৪.৪ মধ্যগামিতার আরও কিছু মাপক

(ক) গুণোত্তর গড় (Geometric Mean) :

ধরা যাক, x এমন একটি চল, যা শুধু ধনাত্মক মান নেয়। যদি x-এর nটি মান থাকে $x_1, x_2, \ldots\ldots, x_n$, তাহলে x-এর গুণোত্তর গড় হবে :

$$x_g = (x_1 \times x_2 \times \ldots\ldots x_n)^{1/n} = \left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}}$$

[যেখানে π হল গুণফলের প্রতীক]

$$\text{বা} \quad \log x_g = \frac{1}{n}\sum_{i=1}^{n} \log x_i$$

অর্থাৎ কোনো চলের গুণোত্তর গড় হল এর মানগুলির লগারিদ্‌ম্‌-এর যৌগিক গড়।

আমরা যদি দুটি চল x এবং y-এর কথা বিবেচনা করি যেখানে উভয়েরই n-টি মান আছে যা যথাক্রমে, $(x_1, x_2, \ldots\ldots, x_n)$ ও $(y_1, y_2, \ldots\ldots, y_n)$ তা হলে সহজেই প্রমাণ করা যায় যে,

$$\left(\prod_{i=1}^{n} \frac{x_i}{y_i}\right)^{1/n} = \frac{\left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}}}{\left(\prod_{i=1}^{n} y_i\right)^{\frac{1}{n}}}$$

অর্থাৎ দুটি চলের অনুপাতের গুণোত্তর গড় এদের গুণোত্তর গড়ের অনুপাতের সমান। গুণোত্তর গড়ের এই ধর্মটি দরের সূচক সংখ্যা (index number of prices) নির্ণয়ের ক্ষেত্রে খুবই সুবিধাজনকভাবে প্রয়োগ করা হয়ে থাকে।

কোনো মূলধনের সুদ নির্ণয়ের ক্ষেত্রে কিংবা কোনো মেশিনের অবচয় (depreciation) নির্ণয়ের ক্ষেত্রে গুণোত্তর গড়ের সুষ্ঠু প্রয়োগ হয়ে থাকে। অর্থাৎ যেখানে চক্রবৃদ্ধি (compound interest)-র নিয়ম প্রযোজ্য সেসব ক্ষেত্রে গুণোত্তর গড়ের প্রয়োগ সফলভাবে করা হয়ে থাকে। নিচের উদাহরণ দুটির দ্বারা ব্যাপারটি ব্যাখ্যা করা হল।

উদাহরণ 1। এক হাজার টাকা এক ব্যক্তিকে এই শর্তে তিন বছরের জন্য ধার দেওয়া হল যে এই তিন বছরে তাকে প্রথম বছর শতকরা 4 টাকা, দ্বিতীয় বছর শতকরা 5 টাকা ও তৃতীয় বছর শতকরা 6 টাকা হারে চক্রবৃদ্ধি সুদ দিতে হবে। তার বাৎসরিক গড় সুদ কত হবে?

উত্তর। গড় সুদের হার যদি r% হয়, তা হলে

$$1000\left(1+\frac{r}{100}\right)^3 = 1000\left(1+\frac{4}{100}\right)\left(1+\frac{5}{100}\right)\left(1+\frac{6}{100}\right)$$

অর্থাৎ, $\left(1+\frac{r}{100}\right)$ হচ্ছে 1.04, 1.05 এবং 1.06 এর গুণোত্তর গড়। সুতরাং

$$\log\left(1+\frac{r}{100}\right) = (\log 1.04 + \log 1.05 + \log 1.06)/3$$

$$= 0.0211762 = \log 1.04997$$

অর্থাৎ r = গড় বাৎসরিক সুদের হার = 0.04997, বা শতকরা 4.997.

উদাহরণ 2। কোনো একটি মেশিনের অবচয়ের (depreciation) হার হল : প্রথম বছরে 20%, দ্বিতীয় ও তৃতীয় বছরে 10% করে এবং পরবর্তী তিন বছরে 5% করে। যদি অবচয়ের ক্ষেত্রে সূচক নিয়ম (Exponential Law) প্রযোজ্য হয়, তবে গড় অবচয় কত হবে?

উত্তর। ধরা যাক, মেশিনটির প্রাথমিক মূল্য হল P_0 এবং r হল এর গড় বার্ষিক অবচয়ের হার। তা হলে চক্রবৃদ্ধির নিয়মে 6 বছর পরে মেশিনটির মূল্য হবে :

$$P_0(1-r)^6 = P_0\left(1-\frac{20}{100}\right)^1\left(1-\frac{10}{100}\right)^2\left(1-\frac{5}{100}\right)^3$$

$$= P_0\frac{4}{5}\left(\frac{9}{10}\right)^2\left(\frac{19}{20}\right)^3$$

বা, $6\log(1-r) = \log\left(\frac{4}{5}\right) + 2\log\left(\frac{9}{10}\right) + 3\log\left(\frac{19}{20}\right)$

$$= -.25527$$

বা, $\log(1-r) = -.04254 = \bar{1}.95746$

$$= \log .90669$$

$$1-r = 0.90669$$

$$r = 0.09331$$

অর্থাৎ গড় বার্ষিক অবক্ষয় = $100 \times .09331 = 9.33\%$

গুণোত্তর গড় ব্যবহারের কয়েকটি অসুবিধা আছে। প্রথমত এটির নির্ণয় বেশ পরিশ্রম সাপেক্ষ। আবার চলকের মানগুলির যে কোনো একটি শূন্য হলে পুরো গড়টির মান শূন্য হবে। তাছাড়া কোনো একটি মান বা একাধিক মান ঋণাত্মক হলে গুণোত্তর গড় নির্ণয় করা সম্ভব হয় না। কারণ মানটি একটি অবাস্তব সংখ্যা হতে পারে।

(খ) প্রতি গাণিতিক গড় (**Harmonic Mean**) :

ধরা যাক, x চলটির n মান আছে —$x_1, x_2, \ldots\ldots, x_n$। তাহলে x_h এই প্রতিগাণিতিক মানটি হবে :

$$x_h = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$

$$\text{বা,} \quad \frac{1}{x_h} = \frac{1}{n}\sum_{i=1}^{h} \frac{1}{x_i}$$

অর্থাৎ প্রদত্ত মানগুলির প্রতিগাণিতিক গড়ের অনোন্যক (reciprocal) মানসমূহের অনোন্যকগুলির যৌগিক গড়ের সমান। কোনো কিছুর 'হার' (rate) নির্ণয় করতে অনেক ক্ষেত্রে প্রতিগাণিতিক গড়ের ব্যবহার সঙ্গত হবে। একটি উদাহরণ দেওয়া হল :

উদাহরণ 3। একটি গাড়ি ঘণ্টায় 60 মাইল বেগে শিয়ালদহ থেকে রাণাঘাট পৌঁছল। ফেরার সময় ঘণ্টায় 60 মাইল বেগে রাণাঘাট থেকে শিয়ালদহ পৌঁছল। গাড়ির গড় গতিবেগ কত?

উত্তর। যদি শিয়ালদহ থেকে রাণাঘাটের দূরত্ব x মাইল হয়, তা হলে গাড়িটির রাণাঘাট পৌঁছতে লেগেছে $\frac{x}{60}$ ঘণ্টা এবং ওখান থেকে কলকাতায় ফিরতে লেগেছে $\frac{x}{50}$ ঘণ্টা। অর্থাৎ মোট $\left(\frac{x}{60} + \frac{x}{50}\right)$ ঘণ্টায় ট্রেনটি 2x মাইল অতিক্রম করেছে। সুতরাং তার গড় গতিবেগ ছিল :

$$\frac{2x}{\left(\frac{x}{60} + \frac{x}{50}\right)} = \frac{2}{\frac{1}{60} + \frac{1}{50}}$$

$$= \frac{2}{0.0167 + 0.0200} = 54.5 \text{ মাইল।}$$

সংখ্যাটি 50 মাইল ও 60 মাইলের প্রতি গাণিতিক গড়।

৩৪.৫ সারাংশ

এই অধ্যায়ে মধ্যগামিতার ব্যাখ্যা এবং মধ্যগামিতার বিভিন্ন মাপক-এর সংজ্ঞা দেওয়া হয়েছে এবং উদাহরণ যোগে এদের বিস্তৃত ব্যাখ্যা করা হয়েছে। এই মাপকগুলির মধ্যে আছে যৌগিক গড়। মধ্যমা, ভূয়িষ্ঠক, গুণোত্তর গড়, প্রতিগাণিতিক গড় ও ভগ্নাংশক। এদের বিভিন্ন বৈশিষ্ট্যের বিবরণ দেওয়া হয়েছে। বেশ কিছু গাণিতিক উদাহরণ-এর সমাধান বিস্তৃত ব্যাখ্যা সহযোগে দেওয়া হয়েছে।

৩৪.৬ অনুশীলনী

১। মধ্যগামিতা বলতে কী বোঝায়? সহজ উদাহরণের সাহায্যে কয়েকটি প্রচলিত মধ্যগামিতা মাপকের সংজ্ঞা ও বিস্তৃত ব্যাখ্যা দিন।

২। কোনো একটি চলের ক্ষেত্রে যদি এই ঘটনাগুলি ঘটে :

(ক) চলের মানগুলি :— (অ) সমপরিমাণে (আ) সমানুপাতে বৃদ্ধি পায় বা হ্রাস পায় তা হলে গাণিতিক গড়ের কী ধরনের পরিবর্তন হওয়া সঙ্গত?

৩। নিম্নলিখিত মানগুলির (ক) যৌগিক গড় ও (খ) মধ্যমা নির্ণয় করুন।

(অ) রাশিবিজ্ঞানের ন'জন ছাত্রের প্রাপ্ত সংখ্যা : 87, 34, 92, 81, 57, 29, 33, 80, 90

(আ) গণিতে দশজন ছাত্রের প্রাপ্ত সংখ্যা : 58, 43, 29, 18, 75, 92, 33, 48, 9, 82

[উঃ—(অ) 65.9, 80 ; (আ) 48.7, 45.2]।

৪। ছাত্রছাত্রীদের পরিবারের আয়তন সংক্রান্ত যে পরিসংখ্যা সারণী তৈরি করেছিলেন তা থেকে পরিবারের আয়তন চলকটির যৌগিক গড়, মধ্যমা ও ভূয়িষ্ঠক নির্ণয় করুন।

৫। পূর্বের এককটিতে নবজাত শিশুদের ওজন সংক্রান্ত একটি পরিসংখ্যা সারণী তৈরি করেছিলেন। এ থেকে চলকটির যৌগিক গড়, মধ্যমা ও ভূয়িষ্ঠক নির্ণয় করুন।

৬। একটি বড় স্থানের উচ্চতম শ্রেণীর চারটি বিভাগের প্রতিটিতে ছাত্রসংখ্যা এবং মাধ্যমিক পরীক্ষায় তাদের প্রাপ্ত গড় মোট নম্বর নিচে দেওয়া হল :

বিভাগ	ছাত্রসংখ্যা	প্রাপ্ত গড় মোটসংখ্যা
ক	55	641
খ	60	597
গ	52	582
ঘ	47	539

ঐ বিভাগ চারটিকে একসঙ্গে ধরলে তাদের প্রাপ্ত মোট সংখ্যার গড় কত হবে?

৭। দুদল শ্রমিকের সাপ্তাহিক মজুরির যৌগিক গড় যথাক্রমে 650 টাকা ও 690 টাকা। তাদের একত্রে নিলে যৌগিক গড় দাঁড়ায় 675 টাকা। প্রথম দলে যদি 63 জন লোক থাকে, তবে দ্বিতীয়টিতে কতজন রয়েছে?

একক ৩৫ ❑ বিস্তৃতি এবং বিস্তৃতিমাপক

গঠন

৩৫.০ উদ্দেশ্য

এই এককটি পড়বার পর আপনি জানতে পারবেন—

- বিস্তৃতি বলতে কি বোঝায়?
- প্রসার কাকে বলে এবং গড় বিচ্যুতি ও প্রমাণ বিচ্যুতি কি?
- আপেক্ষিক বিস্তৃতিমাপক কাকে বলে?

৩৫.১ প্রস্তাবনা

একটু চিন্তা করলেই বোঝা যাবে যে শুধুমাত্র মধ্যগামিতার মাপ নিয়ে কোনো একটি চলের সমগ্র পরিসংখ্যানগুচ্ছের চরিত্র সম্পর্কে সম্যক ধারণা সম্ভব নয়। একটি বহু-উল্লেখিত উদাহরণের সাহায্যে ব্যাপারটা ব্যাখ্যা করা যাক। ধরুন আপনি একটি নদীর তীরে দাঁড়িয়ে আছেন। ওপারে যাওয়ার ইচ্ছা। ঘাটে নৌকার দেখা নেই। যদিও নদীর বিভিন্ন অংশের গভীরতার মাপ সম্বন্ধে আপনার কোনো ধারণা নেই কিন্তু আপনি জানেন

নদীটির গড় (অর্থাৎ মধ্যগামিতার মাপকের) গভীরতা হাঁটুজলের মতো। সুতরাং, সাঁতার না জানা সত্ত্বেও আপনি ঠিক করলেন "গড়ে হাঁটুজলের" এই নদী আপনি অনায়াসেই হেঁটে পার হয়ে যেতে পারবেন। এইটুক খবর জেনে আপনি নদীতে নামলেন এবং অগ্রসর হতে থাকলেন। কিন্তু কিছুদূর অগ্রসর হয়েই বিপদ বাঁধলো। নদীর গভীরতা তখন আপনার গলা পর্যন্ত হয়ে গেছে এবং আর একটু অগ্রসর হতেই বুঝতে পারলেন আরও কিছুটা অগ্রসর হলে আপনি পুরোপুরি ডুবে যাবেন। এ অনর্থ এজন্য ঘটলো যে আপনি শুধুমাত্র মধ্যগামিতার মাপকের খবরের ওপরই সম্পূর্ণ নির্ভর করেছিলেন। হয়তো নদীর অধিকাংশ স্থানের গভীরতাই হাঁটুজলের মতো বা তার চাইতেও কম। কিন্তু কিছু কিছু জায়গায় অতিরিক্ত গভীরতা থাকায় ডুবে যেতে হবে এবং হেঁটে পার হওয়া সম্ভব নয়। সুতরাং দেখা যাচ্ছে যে শুধুমাত্র মধ্যগামিতার মাপ নয়, বিভিন্ন স্থানে নদীর গভীরতা কতটা বেশি বা কম তাও জানা বিশেষ দরকার। মধ্যগামিতার মাপ গভীরতা সম্বন্ধে কিছুটা খবর দিলেও বিভিন্ন স্থানের গভীরতার মধ্যে কী পরিমাণ পার্থক্য আছে অর্থাৎ গভীরতামাপক চল (variable)-টির কতটা বিস্তৃতি (dispersion) তা না জানা থাকলে চলের বৈশিষ্ট্য সম্পর্কে জ্ঞান খুবই অসম্পূর্ণ থেকে যাবে।

৩৫.২ বিস্তৃতি বলতে কী বোঝায়?

সূচনাতে 'গড়ে হাঁটুজলের' উদাহরণে *বিস্তৃতি* (dispersion) সম্বন্ধে একটা ধারণা দেওয়া হয়েছে। এবার এর আরও একটু ব্যাখ্যা করা যাক্। নিচে দুদল ছাত্রের (প্রতি দলে ৫জন ছাত্র) ওজনের হিসাব দেওয়া আছে :

সারণী—১

দুদল ছাত্রের ওজন (কেজিতে)

	1	2	3	4	5	গড় ওজন
প্রথম দল	47	52	50	50	51	50
দ্বিতীয় দল	60	50	33	49	58	50

যদি গাণিতিক গড়ের ব্যবহার যথার্থ মনে হয়, দেখা যাবে যে দুটি ক্ষেত্রেই ছাত্রপ্রতি গড় ওজন 50 কেজি। কিন্তু একটু পর্যবেক্ষণ করলে দেখা যাবে প্রথম ক্ষেত্রে একজন ছাত্র থেকে আর একজন ছাত্রের ওজনের তফাৎ তেমন বেশি নয়। সকল ছাত্রের ওজন 47 কেজি থেকে 52 কেজির মধ্যে রয়েছে। দ্বিতীয় ক্ষেত্রে কিন্তু বিভিন্ন ছাত্রের ওজন 33 কেজি থেকে 60 কেজির মধ্যে ছড়িয়ে আছে। সুতরাং গড় ওজন (মধ্যগামিতার মাপ) এক হলেও দুদল ছাত্রের রাশিতথ্যে যথেষ্ট প্রকৃতিগত পার্থক্য আছে। এই পার্থক্যটা হল দুটি রাশিতথ্যের বিস্তৃতির মাত্রার পার্থক্য। এর থেকে বোঝা যাচ্ছে যে বিস্তৃতির মাপকের বিশেষ আবশ্যকতা আছে। এখানে আমরা এধরনের মাপকের সম্বন্ধে আলোচনা করবো।

বিস্তৃতি (dispersion) মাপার জন্য প্রধানত চার ধরনের মাপকের প্রচলন আছে। এগুলি হল (1) প্রসার (Range), (2) গড় বিচ্যুতি (mean deviation), (3) প্রমাণ বিচ্যুতি (standard deviation) ও (4) চতুর্থক-বিচ্যুতি বা আন্তঃচতুর্থক অর্ধপ্রকার (quartile deviation or semi-inter-quartile range)।

৩৫.৩ প্রসার (Range)

বিস্তৃতির সবচাইতে সহজ এবং বোধগম্য মাপক হল **প্রসার** (range)। প্রদত্ত রাশিতথ্যের সর্বাধিক এবং সর্বনিম্ন মানের পার্থক্যের পরিমাণকে বলা হয় প্রসার। পূর্বোক্ত উদাহরণে (সারণী—১) প্রথম ছাত্রদলের ওজনের প্রসার হল (52—47) = 5 কেজি এবং দ্বিতীয় ছাত্রদলের ওজনের প্রসার হল (60—33) = 27 কেজি।

বিস্তৃতির মাপক হিসাবে প্রসারের সবচাইতে সুবিধা হল এটি অত্যন্ত সহজে নির্ণয় করা যায়। তাছাড়া বিস্তৃতির ধারণা (concept)-র ক্ষেত্রে এটি অত্যন্ত সহজবোধ্য। অন্যদিকে এটি যেহেতু প্রদত্ত রাশিতথ্যের সবকটি মানের ওপর নির্ভরশীল নয়, এটি বিস্তৃতি সম্পর্কে অনেক সময় ভ্রান্ত ধারণা দিতে পারে। এমন দুটি তথ্যগুচ্ছের কথা ভাবা যায় যেখানে চলকের বৃহত্তম মান দুটি সমান এবং ক্ষুদ্রতম মান দুটিও সমান, কিন্তু একটিতে মাঝখানের মানগুলি খুবই কাছাকাছি আর অন্যটিতে অনেক বেশি ছড়ানো। এখানে দুই তথ্যগুচ্ছেই প্রসার অভিন্ন হবে, কিন্তু স্পষ্টতই প্রথম ক্ষেত্রে বিস্তৃতি স্বল্পতর।

৩৫.৩.১ গড় বিচ্যুতি (Mean Deviation) :

ধরা যাক, একটি চল (x)-এর n-টি মান $x_1, x_2, \ldots\ldots\ldots, x_n$ দেওয়া আছে। এই মানগুলিকে একগুচ্ছ রাশিতথ্য বলে ধরা হয়েছে। যদি একটি ধ্রুবক a আমাদের বলে ধরি, তা হলে সংখ্যা অনুসারে প্রদত্ত মানগুলির a-কেন্দ্রিক গড় বিচ্যুতি হবে :

$$M_{D_a} = \frac{1}{n}\sum_{i=1}^{n} | x_i - a | \qquad (৩.১.১)$$

শ্রেণীবিন্যস্ত রাশিতথ্যের ক্ষেত্রে :

$$M_{D_a} = \frac{1}{n}\sum_{i=1}^{k} f_i | x_i - a | \qquad (৩.১.২)$$

যেখানে, x_i হল i-তম শ্রেণীর জ্ঞাপক মান বা ঐ শ্রেণীর মধ্যবিন্দু, f_i হল ঐ শ্রেণীর পরিসংখ্যা (i = 1, 2.....,k)

এবং $n = \sum_{i=1}^{k} f_i$

এবং $| x_i |$ প্রতীকটির অর্থ হল x_i-এর চিহ্নবর্জিত মান বা পরমমান (absolute value)। অর্থাৎ

$| x_i | = x_i$ যদি x_i অঋণাত্মক হয় এবং $| x_i | = - x_i$ যদি x_i ঋণাত্মক হয়।

এরূপ পরমমান নেওয়ার উদ্দেশ্যটি ব্যাখ্যা করা যাক। স্পষ্টতই

$$x_1 - a,\ x_2 - a,\ x_3 - a,\ \ldots\ldots\ldots\ldots,\ x_n - a$$

এই পার্থক্যগুলির মান যত বেশি হবে বিস্তৃতির পরিমাণও তত বেশি হবে। কিন্তু যদি দেখা যায় কিছু ক্ষেত্রে $x_i > a$ অর্থাৎ $(x_i - a)$ = ধনাত্মক (+v) এবং অন্যত্র $x_i < a$ অর্থাৎ $(x_i - a)$ = ঋণাত্মক (−v) তা হলে পার্থক্যগুলির যোগফল ধনাত্মক এবং ঋণাত্মক মানের কাটাকাটিতে অনেকটা কমে যাবে। অর্থাৎ বিচ্যুতিগুলি যথেষ্ট বড় হওয়া সত্ত্বেও তাদের যোগফল (এবং গড়মান) খুব কম হবে। অর্থাৎ $\frac{1}{n} = \sum_{i=1}^{k}(x_i - a)$ একটি যথার্থ বিস্তৃতির পরিমাপক হতে পারে না। এই অসুবিধাটিকে দূর করার জন্যই পার্থক্যগুলিকে চিহ্নবর্জিত, অর্থাৎ $| x_i - a |$, বলে নেওয়া হয়েছে।

প্রমাণ করা যেতে পারে যে, মধ্যমাকেন্দ্রিক গড় বিচ্যুতি অন্য যে কোনো বিন্দুকেন্দ্রিক গড় বিচ্যুতির চাইতে কম হবে। অর্থাৎ $MD_{Mi} \leq MD_a$। এজন্য বিস্তৃতির মাপক হিসাবে সাধারণভাবে মধ্যমাকেন্দ্রিক গড় বিচ্যুতির প্রচলন দেখা যায়। নিচে একটি উদাহরণ দেওয়া হল।

উদাহরণ ৩.১। নিচে সাতজন ব্যক্তির দৈনিক আয়ের হিসাব দেওয়া আছে। (1) যৌগিক গড়-কেন্দ্রিক এবং (2) মধ্যমাকেন্দ্রিক বিচ্যুতি নির্ণয় করুন।

দৈনিক আয় (টাকায়)

215, 60, 55, 35, 28, 37, 25-এ সংখ্যাতথ্যের জন্য আমরা পাই $\frac{455}{7} = 65$ টাকা যৌগিক গড় = মধ্যমা = 37 টাকা।

গড় বিচ্যুতি নির্ণয়

	মধ্যমা কেন্দ্রিক বিচ্যুতি	গড় যৌগিক কেন্দ্রিক বিচ্যুতি
	178	150
	23	5
	18	10
	2	30
	9	37
	0	28
	12	40
মোট	242	300

তাই $MD_{\bar{x}} = \frac{300}{7} = 42.86$ টাকা এবং $MD_{M_i} = \frac{242}{7} = 34.57$ টাকা। স্পষ্টতই গড় কেন্দ্রিক গড় বিচ্যুতি মধ্যমা কেন্দ্রিক গড় বিচ্যুতির চাইতে বেশি।

উদাহরণ ৩.২। সারণী নং 2.2-এ দেওয়া পরিসংখ্যা বিভাজনের গড় বিচ্যুতি নির্ণয় (গাণিতিক গড়ভিত্তিক) করুন।

গাণিতিক গড় 3.84 [উদাহরণ 2.3 দ্রষ্টব্য]

বাড়িতে ঘরের সংখ্যা x_i	পরিসংখ্যা f_i	$\lvert x_i - \bar{x}_i \rvert$	$f_i \lvert x_i - \bar{x}_i \rvert$
1	8	2.84	22.72
2	18	1.84	33.12
3	30	0.84	25.20
4	25	0.16	4.00
5	19	1.16	22.04
6	10	2.16	21.60
7	8	3.16	25.28
8	2	4.16	8.32
মোট	120	—	162.28

সুতরাং এক্ষেত্রে গড় বিচ্যুতি দাঁড়াবে $MD_{\bar{x}} = \frac{162.28}{120} = 1.352$

গাণিতিক গড় বিচ্যুতির ধর্ম : স্পষ্টতই গাণিতিক গড়কেন্দ্রিক গড় বিচ্যুতি মূলবিন্দু (origin) নির্ভর হবে না ; কিন্তু মাত্রা নির্ভর হবে।

ধরা যাক, $u = (x - a)/c$

সুতরাং x_2 ও u_i যথাক্রমে x ও u-এর i-তম মানদ্বয় সূচিত করলে $x_i = a + cu_i$ এবং $\bar{x} = a + c\bar{u}$ ফলে, $\lvert x_i - \bar{x} \rvert = \lvert c \rvert \lvert u_i - \bar{u} \rvert$

বা, $\frac{1}{n} \sum_{i=1}^{k} \lvert x_i - \bar{x} \rvert = \lvert c \rvert \frac{1}{n} \sum_{i=1}^{k} \lvert u_i - \bar{u} \rvert$

অর্থাৎ $MD_{\bar{x}}(x) = |c|\, MD_{\bar{u}}(u)$.

মধ্যমা-কেন্দ্রিক বা ভূয়িষ্ঠক-কেন্দ্রিক গড় বিচ্যুতির বেলায়ও একই ধরনের সম্পর্ক খাটবে।

৩৫.৩.২ প্রমাণ বিচ্যুতি (Standard Deviation) :

রাশিতথ্যের মানগুলি ব্যবহার করে বিভিন্ন বিচ্যুতি (যথা $x_i - a_i$) গণনার সময় এই বিচ্যুতিগুলির চিহ্নের বিভিন্নতার (ধনাত্মক বা ঋণাত্মক) জন্য যে ধরনের অসুবিধা হয় তা দূর করার, আর একটি পদ্ধতি হল এদের বর্গ (square) নেওয়া। কারণ, বর্গ সবসময়ই অঋণাত্মক মান নেবে। এই সুবিধাটি অবলম্বন করে বিস্তৃতির যে মাপকটি উদ্ভাবিত হয়েছে তার নাম হল **প্রমাণ বিচ্যুতি** (Standard Deviation) ।

নিম্নলিখিত সূত্রটি লক্ষ্য করুন :

x-এর মানগুলি $x_i\,(i = 1, 2, \ldots\ldots, n)$ হলে, সংশ্লিষ্ট বিচ্যুতির বর্গগুলি হবে $(x_i - a)^2$ (৩.১.৫.১)

যেখানে a একটি মধ্যগামিতার মাপক। সুতরাং $\frac{1}{n}\sum_{i=1}^{n}(x_i - a)^2$ হবে বর্গ-বিচ্যুতির যৌগিক গড়। এর বর্গমূল নিলে আমরা পাই —

$$s_n = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - a)^2} \qquad (৩.১.৫.২)$$

এটিকে বলা হয় মূল-গড়-বর্গ বিচ্যুতি (root-mean-square deviation)। যদিও বর্গমূলের দুটি মান হওয়া সম্ভব (যেমন 16-র বর্গমূল 4 ও –4 দুই-ই)। আমরা এক্ষেত্রে বর্গমূলের ধনাত্মক (positive) মানটিই শুধু বিবেচনা করব। এটিকে বিস্তৃতির মাপক হিসাবে ব্যবহার করা যায়। মূল-গড়-বর্গ বিচ্যুতি নির্ণয় করতে মধ্যগামিতার মাপক হিসাবে সাধারণত $a = \bar{x}$ = গাণিতিক গড়কেই নেওয়া হয়। এরূপ মূল-গড়-বর্গ বিচ্যুতিকে বলা হয় **প্রমাণ বিচ্যুতি**। এবং এটিকে s দ্বারা সূচিত করা হয়। অর্থাৎ এখানে

$$s = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$ পরিসংখ্যা সারণীর ক্ষেত্রে, $$s = \sqrt{\frac{1}{n}\sum_{i=1}^{k}f_i(x_i - \bar{x})^2} \qquad (৩.১.৫.৩)$$

একটু বিশ্লেষণ করলে আমরা পাই :

$$\sum_{i=1}^{n}|x_i - \bar{x}|^2 = \sum_{i=1}^{n}x_i^2 - n\bar{x}^2 - 2\bar{x}\sum_{i=1}^{n}x_i$$

$$= \sum_{i=1}^{n}x_i^2 - n\bar{x}^2, \text{ কারণ } \bar{x} = \frac{1}{n}\sum_{i=1}^{n}x_i$$

সুতরাং প্রমাণ বিচ্যুতির বিকল্প সূত্র হল

$$s = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^{\,2} - \overline{x}^2} \qquad (৩.১.৫.৪)$$

অনুরূপভাবে, পরিসংখ্যা সারণীর ক্ষেত্রে,

$$s = \sqrt{\frac{1}{n}\sum_{i=1}^{k} f_i x_i^{\,2} - \overline{x}^2} \qquad (৩.১.৫.৫)$$

আবার s^2 -কে অর্থাৎ প্রমাণ বিচ্যুতির বর্গকে বলা হয় **ভেদমান** (variance)।

৩১.৫.১ প্রমাণ বিচ্যুতির ধর্মাবলী

(ক) **x চলকের প্রদত্ত মানগুলি সব সমান হলে প্রমাণ-বিচ্যুতির মান শূন্য হবে।**

প্রমাণ : ধরা যাক

$$x_1 = x_2 = \ldots\ldots\ldots = x_n = a$$

তা হলে, $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n} \times na = a$ এবং $x_i - \overline{x} = 0$

সুতরাং, প্রমাণ বিচ্যুতি = $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}$

$$= \sqrt{\frac{1}{n} \times n \times 0} = 0$$

(খ) **গড় বিচ্যুতির মতো প্রমাণ বিচ্যুতি ও মূলবিন্দু-নিরপেক্ষ (independent of the origin), কিন্তু মাত্রা-নিরপেক্ষ (independent of scale) নয়।**

প্রমাণ : যদি x-এর পরিবর্তে আমরা নতুন চলক নিই $u = \frac{x - a}{c}$,

তাহলে $x_i = a + cu_i \; (i = 1, 2, \ldots, n)$

এবং $\overline{x} = a + c\overline{u}$

সুতরাং $(x_i - \overline{x}) = c\,(u_i - \overline{u})$

এবং $\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2 = c^2\times\frac{1}{n}\sum_{i=1}^{n}(u_i-\overline{u})^2$

অর্থাৎ $\sqrt{\frac{1}{n}\Sigma(x_i-\overline{x})^2} = |c|\sqrt{\frac{1}{n}\Sigma(u_i-\overline{u})^2}$

(৩.১.৫.৬)

সুতরাং $s_x = |c|\, s_u$.

(গ) ধরা যাক x চলটির দুই গুচ্ছ মান দেওয়া আছে। যথা, $(x_{11}, x_{12}\ldots\ldots x_{1n})$ এবং $(x_{21}, x_{22}\ldots\ldots x_{2n_2})$ তা হলে x-এর এই দুই গুচ্ছ মান একত্রিত করলে সম্মিলিত প্রমাণ বিচ্যুতি **(composite standard deviation)** হবে s, যা এরূপ যে

$$s^2 = \frac{n_1s_1{}^2+n_2s_2{}^2}{n_1+n_2} + \frac{n_1(x_1-\overline{x_1})^2+n_2(x_2-\overline{x_2})^2}{n_1+n_2} \qquad (৩.১.৫.৭)$$

প্রমাণ : আমরা আগেই দেখেছি যে

$$\overline{x} = \frac{n_1\overline{x}_1+n_2\overline{x}_2}{(n_1+n_2)}$$

আবার, $\overline{x}$ থেকে প্রথম এবং দ্বিতীয় গুচ্ছের মানগুলির বিচ্যুতির যোগফল হল—

$$\sum_{i=1}^{n_1}(x_{1j}-\overline{x})^2 + \sum_{j=1}^{n_2}(x_{2j}-\overline{x})^2$$

কিন্তু যেহেতু $(x_{1j}-\overline{x})^2 = \left\{(x_{1j}-\overline{x_1})+(\overline{x_1}-\overline{x})\right\}^2$

$$= (x_{1j}-\overline{x_1})^2 + 2(\overline{x_1}-\overline{x})(x_{1j}-\overline{x_1}) + (\overline{x_1}-\overline{x})^2$$

$$= \sum_{j=1}^{n_1}(x_{1j}-\overline{x})^2 = \sum_{j=1}^{n_1}(x_{1j}-\overline{x_1})^2 + n_1(\overline{x_1}-\overline{x})^2 \left[\text{কারণ } \sum_{j=1}^{n}(x_{1j}-\overline{x}) = 0\right]$$

$$= n_1s_1{}^2 + n_1(\overline{x_1}-\overline{x})^2 \qquad \ldots\ldots\ldots\ldots(অ)$$

ঠিক অনুরূপভাবে,

$$\sum_{j=1}^{n_2}(x_{2j}-\overline{x})^2 + n_2s_2{}^2 + n_2(\overline{x_2}-\overline{x})^2 \qquad \ldots\ldots\ldots\ldots(আ)$$

সুতরাং, (অ) এবং (আ) থেকে আমরা পাই —

$$s^2 = \frac{\left\{\sum_{j=1}^{n_1}(x_{1j} - \overline{x})^2 + \sum_{i=1}^{n_2}(x_{2j} - \overline{x})^2\right\}}{(n_1 + n_2)}$$

$$= \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2} + \frac{n_1(\overline{x_1} - \overline{x})^2 + n_2(\overline{x_2} - \overline{x})^2}{n_1 + n_2} \qquad (৩.১.৫.৮)$$

অনুরূপভাবেই প্রমাণ করা যায় যে, l- সংখ্যক মানগুচ্ছের বেলায় সম্মিলিত প্রমাণ বিচ্যুতি (s)-এর সূত্র হবে—

$$s^2 = \frac{n_1 s_1^2 + n_2 s_2^2 + \ldots\ldots + n_l s_l^2}{n_1 + n_2 + \ldots\ldots + n_l} + \frac{n_1(\overline{x_1} - \overline{x})^2 + n_2(\overline{x_2} - \overline{x})^2 + \ldots\ldots + n_l(\overline{x_l} - \overline{x})^2}{n_1 + n_2 + \ldots\ldots + n_l} \qquad (৩.১.৫.৯)$$

যেখানে n_i, $\overline{x_i}$ এবং s_i যথাক্রমে i-তম গুচ্ছের মানসংখ্যা, গড়মান ও প্রমাণ বিচ্যুতি এবং

$$\overline{x} = \frac{n_1\overline{x_1} + n_2\overline{x_2} + \ldots\ldots + n_l\overline{x_l}}{n_1 + n_2 + \ldots\ldots + n_l} \qquad (৩.১.৫.১০)$$

উদাহরণ ৩.৩ :

আমরা 120টি বাড়ির জন্য ঘরসংখ্যার যে পরিসংখ্যা আগে নিয়েছিলাম তার জন্য প্রমাণ বিচ্যুতি নির্ণয় করতে পারি। এখানে প্রথমে চলকের মূল পরিবর্তন করে নেওয়া হয়েছে।

আমরা লিখতে পারি :

x_i	f_i	$u_i = x_i - 4$	$u_i f_i$	$u_i^2 f_i$
1	8	–3	–24	72
2	36	–2	–72	144
3	90	–1	–90	90
4	100	0	0	0
5	95	1	95	95
6	60	2	120	240
7	56	3	168	504
8	16	4	64	256
মোট	461	—	261	1401

উপরের সারণীর ভিত্তিতে আমরা পেলাম

$$\overline{u} = \frac{1}{n}\sum_{i=1}^{k} u_i f_i = \frac{1}{461} \times 261 = 0.566$$

$$\text{এবং } s_u^2 = \frac{1}{n}\sum_{i=1}^{k} u_i^2 f_i - \overline{u}^2$$

$$= \frac{1}{461} \times 1401 - (0.566)^2 = 3.039 - 0.320 = 2.719 \;.$$

$$\text{তাই } s_u = \sqrt{2.719} = 1.649 \text{ এবং } s_x = s_u = 1.649.$$

৩৫.৩.৩ চতুর্থক বিচ্যুতি (Quartile deviation) বা আন্তঃচতুর্থক প্রসার (Semi-interquartile range) এবং দশমিক বিচ্যুতি (Decile deviation) :

(ক) **চতুর্থক বিচ্যুতি (Quartile Deviation) বা আন্তঃচতুর্থক প্রসার (Semi-interquartile Range) :**

পূর্ববর্তী পরিচ্ছেদে আমরা মধ্যমা (median) এবং ভগ্নাংশক (fractile)-এর বর্ণনা দিয়েছি। এ বর্ণনা অনুসারে আমরা জানি যে, চলের মানগুলি ঊর্ধ্বগ বা নিম্নগ ক্রম অনুসারে সাজিয়ে নিম্নলিখিত মাপগুলি পেতে পারি :

Q_1 = প্রথম চতুর্থক। এই মানটি চলের মান বিভাজনের নিম্নগ ক্রম অনুসারে প্রথম চতুর্থাংশের বিভাজক।

Q_3 = তৃতীয় চতুর্থক। এই মানটি চলের মান বিভাজনের নিম্নগ ক্রম অনুসারে তৃতীয় চতুর্থাংশের বিভাজক।

Q_2 = মধ্যমা বা দ্বিতীয় চতুর্থক। এই মানটি চলের নিম্নগ (বা ঊর্ধ্বগ) ক্রম অনুসারে অর্ধেকের বিভাজক।

স্পষ্টতই কোনো চল কর্তৃক গৃহীত মানগুলি যত বেশি বিস্তৃত হবে এর প্রথম ও তৃতীয় চতুর্থকের তফাৎও তত বেশি হবে।

সুতরাং, $Q = \frac{1}{2}(Q_3 - Q_1)$ (৩.১.৫.১১)

এই মাপকটিকে বিস্তৃতিমাপক হিসাবে ব্যবহার করা হয়। এর নাম **চতুর্থক বিচ্যুতি** (quartile deviation) বা **অন্তঃচতুর্থক প্রসার** (semi-interquartile range)।

(খ) **দশমিক বিচ্যুতি (Decile deviation) :** আমরা ২.১.২.৪-এ বর্ণনা করেছি যে—যে মানটি চলের বিভাজনকে $\frac{p}{(1-p)}$ অনুপাতে ভাগ করে তাকে চলের p-তম ভগ্নাংশক (Quartile or Fractile) বলা হয়। সুতরাং $\frac{1}{10}/\left(1-\frac{1}{10}\right)=\frac{1}{9}$-কে 10-তম ভগ্নাংশক বলা যেতে পারে। এই দশতম ভগ্নাংশক থেকে পার্থক্যের মাপক নির্ণয় করে দশমিক বিচ্যুতি নির্ণয় করা যেতে পারে। বাস্তবে এধরনের বিচ্যুতির ব্যবহার খুবই সীমিত।

৩৫.৪ বিভিন্ন ধরনের বিস্তৃতিমাপকের তুলনা

আগেই বলা হয়েছে সবচাইতে সহজে যে বিস্তৃতিমাপক নির্ণয় করা যায় তা হল প্রসার (range)। কিন্তু এই মাপকটির সবচাইতে বড় ত্রুটি হল এটি শুধুমাত্র চলের সর্ববৃহৎ এবং সর্বনিম্ন মানের ওপর নির্ভরশীল। ফলে সর্ববৃহৎ মানটি যদি অন্যান্য মানের চাইতে অত্যধিক হয় কিংবা সর্বনিম্ন মানটি অন্যান্য মানের চাইতে অত্যন্ত কম হয় তা হলে প্রসারের মানটি আলোচ্য রাশিতথ্যের বিস্তৃতির প্রকৃত পরিচয় দেয় না। অন্যদিকে গড় বিচ্যুতি (mean deviation) এবং প্রমাণ বিচ্যুতি (standard deviation)—এই দুটি মাপকের ক্ষেত্রে এ রকম সমস্যা দেখা দেয় না। কারণ এই দুটি মাপক গণনার সময় রাশিতথ্যের (বা চলের) প্রতিটি মানকে অন্তর্ভুক্ত করা হয়। কিন্তু এদের গণনাপদ্ধতি কিছুটা দীর্ঘ। ফলে রাশিতথ্যের পরিমাণ বেশি হলে এগুলি গণনা করতে সময়ের প্রয়োজন।

উপরোক্ত তিনটি মাপকের সুবিধা-অসুবিধার বিচারে চতুর্থকের বিচ্যুতির অবস্থান মাঝামাঝি স্থানে।

তবে রাশিবিজ্ঞানের বিস্তৃত প্রয়োগে প্রমাণ-বিচ্যুতির ব্যবহার সর্বাধিক। কারণ এর এমন কিছু বৈশিষ্ট্য আছে যেগুলি বিভিন্ন বীজগাণিতিক গণনার ক্ষেত্রে সহজেই প্রয়োগ করা যায়।

তবে একটি বিশেষ ক্ষেত্রে প্রসারের ব্যবহার অত্যন্ত উপযোগী। এটি হল রাশিবিজ্ঞানসম্মত গুণ-নিয়ন্ত্রণ (Statistical Quality Control)-এর ক্ষেত্রে। এখানে একটি সীমিত মাপের মধ্যে অল্পসময়ে অনেকগুলি বিস্তৃতিমাপকের দরকার হয়। এরূপ ক্ষেত্রে নিঃসন্দেহে প্রসারের গণনা সবচাইতে সুবিধাজনক। সেজন্য এরূপ ক্ষেত্রে এটির ব্যাপক ব্যবহার করা হয়ে থাকে।

৩৫.৫ আপেক্ষিক বিস্তৃতি মাপক (Relative Measures of Dispersion)

এতক্ষণ আমরা যে মাপকগুলির বর্ণনা দিয়েছি সেগুলির একক, প্রদত্ত রাশিতথ্যের এককের অনুরূপ হয়। যেমন, 3.3 উদাহরণে পরিবার প্রতি গড় সন্তান সংখ্যা হল 2.18 এবং প্রমাণ-বিচ্যুতির পরিমাণ 1.19টি

সন্তান। এর ফলে যেখানে দুপ্রকার (কিংবা তার চাইতে বেশি সংখ্যক) রাশিতথ্যের বিস্তৃতির তুলনামূলক বিচারের প্রশ্ন আসে সেখানে অসুবিধা দেখা দেয়। যেমন ধরা যাক, কোনো একটি স্কুলের একটি শ্রেণীতে ছাত্রদের গড় উচ্চতা 140 সেন্টিমিটার এবং প্রমাণ বিচ্যুতির পরিমাণ 40.18 সেমি। আবার ঐ একই ছাত্রদলের গড় ওজন 30.2 কেজি এবং প্রমাণ-বিচ্যুতির পরিমাণ 6.12 কেজি। এখানে ছাত্রদের উচ্চতার বিস্তৃতির সাথে ওজনের বিস্তৃতির তুলনা সম্ভব নয়, কারণ এদের একক বিভিন্ন। এসব ক্ষেত্রে এমন একটি মাপকের ব্যবহার করা দরকার যেটি একক-নিরপেক্ষ (Independent of Unit of Measurement)। এরূপ একটি বহুল ব্যবহৃত মাপক হল—ভেদাঙ্ক (Coefficient of Variation বা C.V.)। এটিকে প্রতীক (symbol)-এ নিম্নরূপে দেখান হয় :

$$\text{ভেদাঙ্ক (C.V.)} = 100 \times \frac{\text{প্রমাণ বিচ্যুতি}}{\text{গাণিতিক গড়}}$$

$$= 100 \times \frac{s}{\bar{x}} \qquad (৩.১.৮.১)$$

পূর্বোক্ত উদাহরণে :

(অ) ছাত্রদের উচ্চতার ভেদাঙ্ক (C.V.) $= 100 \times \frac{140}{418} = 34.8$

(আ) ছাত্রদের ওজনের ভেদাঙ্ক (C.V.) $= 100 \times \frac{30.2}{6.12} = 41.3$

দেখা যাচ্ছে শতকরা হিসাবে ওজনের ক্ষেত্রে বিস্তৃতির পরিমাণ উচ্চতার পরিমাণের চাইতে কিছুটা বেশি।

৩৫.৬ সারাংশ

এই পরিচ্ছেদে বিস্তৃতির সংজ্ঞা এবং রাশিবিজ্ঞানে এর প্রয়োজনীয়তা সম্বন্ধে বিস্তৃত আলোচনা করা হয়েছে। প্রসার (range) গড় বিচ্যুতি (mean deviation) প্রমাণ বিচ্যুতি (standard deviation), চতুর্থক বিচ্যুতি (quartile deviation) বা আন্তঃচতুর্থক প্রসার (semi-interquartile range), দশমিক বিচ্যুতি (decile deviation) ইত্যাদির সংজ্ঞা এদের বিভিন্ন ধর্ম ও ক্ষেত্রবিশেষে প্রয়োগের উপযোগিতা সম্পর্কে আলোচনা করা হয়েছে। বিভিন্ন প্রকার বিস্তৃতি মাপকের তুলনা করা হয়েছে। পরিশেষে আপেক্ষিক বিস্তৃতি মাপকের ব্যাখ্যা এবং এর উপযোগিতা সম্বন্ধে আলোচনা করা হয়েছে। ভেদাঙ্কের (coefficient of variation)-এর সংজ্ঞা দেওয়া হয়েছে এবং উদাহরণ সহযোগে এর প্রয়োগ দেখান হয়েছে।

৩৫.৭ অনুশীলনী

(১) নিম্নলিখিত পরিসংখ্যা বিভাজনের প্রমাণ-বিচ্যুতি নির্ণয় করুন :

ওজন (কেজি)	লোকসংখ্যা
40-44	2
45-49	7
50-54	18
55-59	39
60-64	44
65-69	20
70-74	9
75-79	1
মোট	140

(২) নিম্নলিখিত সংখ্যাগুলির (ক) গাণিতিক গড়কেন্দ্রিক এবং (খ) মধ্যমাকেন্দ্রিক গড়বিচ্যুতি নির্ণয় করুন : 25, 5, 47, 103, 40

[উঃ (ক) 25.6 (খ) 24]

(৩) কোনো পরিমাপের সাহায্য ছাড়াই বলুন নিম্নের সংখ্যাগুচ্ছ দুটির কোন্‌টিতে বিস্তৃতি অধিকতর?

গুচ্ছ 1 :	43	46	47	48	50	51	55	58
গুচ্ছ 2 :	43	47	47	48	48	50	50	58

(৪) (ক) x-এর মান a ও b হলে এবং উভয়ের পরিসংখ্যা সমান হল প্রমাণ বিচ্যুতি কত হবে?

(খ) x-এর মান a, (a + b)/2 এবং b, আর এদের পরিসংখ্যা যথাক্রমে 1, n−2 ও 1। তা হলে x-এর প্রমাণ বিচ্যুতি কত হবে?

(৫) একটি চলের দুগুচ্ছ সংখ্যাতথ্যের একটিতে n_1 ও অন্যটিতে n_2 মান রয়েছে। গুচ্ছ দুটির গড়মান যথাক্রমে $\overline{x_1}$ ও $\overline{x_2}$ এবং তাদের প্রমাণ বিচ্যুতি যথাক্রমে s_1 ও s_2। দেখাও যে দুটি গুচ্ছ একসঙ্গে নিলে সামগ্রিক সমকপার্থক্য হবে s, যেখানে

$$s^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2} + \frac{n_1 + n_2}{(n_1 + n_2)^2}\left(\overline{x_1} - \overline{x_2}\right)^2$$

(৬) প্রমাণ করুন যে $x_1, x_2, \ldots\ldots x_n$ এই মানসমূহের জন্য

$$\sum_i x_i^2 \geq n\overline{x}^2$$

(৭) অনুশীলন ৮-এর সম্পর্কটি ব্যবহার করে দেখান যে, কোনো চলকের জন্য

$$MD\overline{x} \leq SD.$$

(৮) কোনো চলকের প্রদত্ত মানগুলিকে দুটি গোষ্ঠীতে ভাগ করা যাক : [1] প্রথম গোষ্ঠীতে আছে সেসব মান যেগুলি যৌগিক গড়ের অনধিক এবং [2] দ্বিতীয় গোষ্ঠীতে রয়েছে যেসব মান যেগুলি যৌগিক গড়ের অধিক। প্রথম ও দ্বিতীয় গোষ্ঠীর মানসমূহের জন্য যৌগিক গড় থেকে চিহ্নবর্জিত বিচ্যুতির যোগফল যথাক্রমে T_1 ও T_2 হলে প্রমাণ করুন যে—

$$MD_{\overline{x}} = 2T_1/n = 2T_2/n$$

[এই কারণে $MD_{\overline{x}}$ নির্ণয় করতে গেলে কোনো একটি যোগফল T_1 বা T_2, ব্যবহার করে শ্রম লাঘব করা যেতে পারে।]

একক ৩৬ ❑ কালীন সারি বিশ্লেষণ এবং সূচক সংখ্যা

গঠন

৩৬.০ উদ্দেশ্য

এই এককটি পাঠ করার পর আপনি জানতে পারবেন—

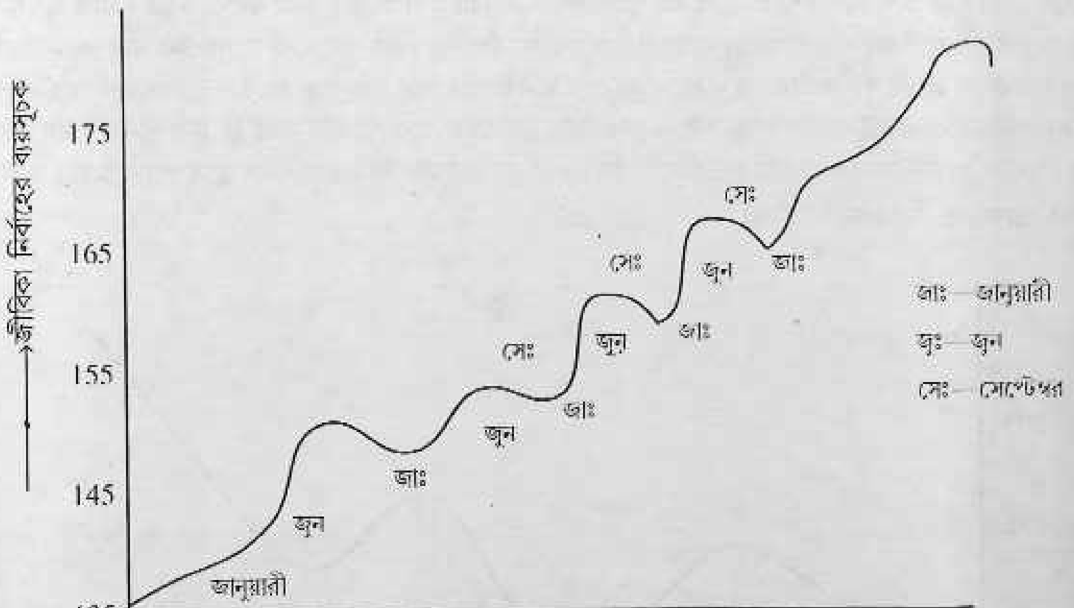
- কালীন সারির গতিধারার বিভিন্ন অংশ।
- চলমান গড় ব্যবস্থার পদ্ধতি ও তাত্ত্বিক যুক্তি।
- সুশাসিত গতিধারার পরিমাপ ও বিভাজন পদ্ধতি।
- সূচক সংখ্যা কাকে বলে।
- সূচক সংখ্যার প্রতীক, নির্ণয়ের সমস্যা, সামঞ্জস্য প্রভৃতি।

৩৬.১ প্রস্তাবনা

পরপর কয়েক বছর বা মাসের জন্য কোনো পণ্যের দর, কয়েক দিনের (বা মাসের বা বছরের তাপমাত্রা বা বৃষ্টিপাতের পরিমাণ কয়েক বছরের জনসংখ্যা বৃদ্ধির হিসাব ইত্যাদি কালীন সারির উদাহরণ। অর্থাৎ ধারাবাহিকভাবে সময়ের সাথে সম্পর্কযুক্ত রাশিতথ্যকে **কালীন সারি** (time series) বলে। অর্থনীতি সংক্রান্ত রাশিতথ্যের প্রসঙ্গে কালীন সারির ব্যাপক প্রচলন আছে। প্রকৃতপক্ষে সামাজিকবিজ্ঞানসমূহ (social sciences) কিংবা প্রাকৃতিক বিজ্ঞানসমূহে (natural sciences) যখনই অতীতের অভিজ্ঞতার ভিত্তিতে ভবিষ্যৎ সম্বন্ধে ধারণা করা প্রয়োজন হয়, তখনই নানাবিধ কালীন সারির বিশ্লেষণ প্রাসঙ্গিক হয়ে ওঠে।

৩৬.২ কালীন সারির গতিধারার বিভিন্ন অংশ (Time series movement—its different components)

কোনো একটি কালীন সারিকে লেখ বা রেখাচিত্রের (graph) সাহায্যে দেখান যেতে পারে। উদাহরণস্বরূপ, নিচের রেখাচিত্রের সাহায্যে, কলকাতার কয়েক বছরে জীবিকা নির্বাহের ব্যয়ের হ্রাস-বৃদ্ধি দেখানো হল।



কলকাতার কয়েক বছরের জীবিকা নির্বাহের ব্যয়সূচক
চিত্র নং—৪.১

কালীন সারির এ ধরনের রেখাচিত্র বিস্তৃতভাবে বিশ্লেষণ করলে দেখা যায় যে এদের কিছু বৈশিষ্ট্য আছে। খুব কম সময়ের কালীন সারির গতিবিধি অনেকটা অনিয়মিত (irregular) হয়। কিন্তু বেশ কিছু সময়ের কালীন

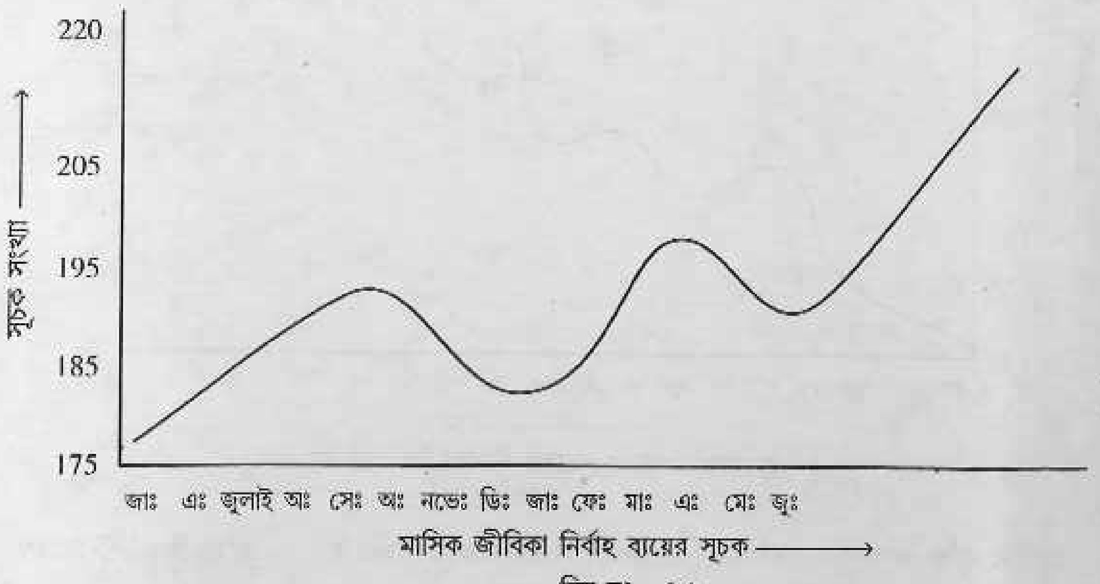
সারি নিয়ে পরীক্ষা করলে দেখা যায় কালীন সারি সময়ের সাথে সাথে কিছু নিয়মিত গতিধারা মেনে চলে। কালীন সারির নিয়মিত গতিধারাকে প্রধান তিন ভাগে ভাগ করা হয়—(ক) সুশাসিত গতিধারা (secular trend), (খ) ঋতুজ ভেদ (seasonal variation) এবং (গ) চক্রিল ভেদ (cyclical variation)। সুতরাং কোনো কালীন সারির মোট চারটি অংশ থাকে। উপরোক্ত তিনটি অংশ এবং (ঘ) অনিয়মিত গতিধারা।

(ক) সুশাসিত গতিধারা (secular trend)

কোনো একটি কালীন সারির দীর্ঘকালের রেখাচিত্র পরীক্ষা করলে অধিকাংশ ক্ষেত্রেই এর একটি দীর্ঘকালীন গতিবিধি লক্ষ্য করা যায়। দীর্ঘকালের পরিপ্রেক্ষিতে রেখাচিত্রটি সাধারণত ঊর্ধ্বমুখী বা নিম্নমুখী হয়। অনেক সময় বেশ কিছুকাল সমান্তরালভাবে চলার পর এই ঊর্ধ্বমুখী বা নিম্নমুখী প্রবণতা দেখা যায়। ৪.১নং চিত্রের রেখাচিত্রটিতে ঊর্ধ্বমুখী প্রবণতা লক্ষ্য করা যাচ্ছে। কালীন সারির এ ধরনের মসৃণ (smooth) ও দীর্ঘকালীন গতিবিধিকে সুশাসিত গতিধারা (secular trend বা শুধু trend) বলে অভিহিত করা হয়। সুশাসিত গতিধারা একটি দীর্ঘস্থায়ী ব্যাপার। কোনো ধরনের স্বল্পকালীন বা তাৎক্ষণিক পরিবর্তন এই গতিধারার দ্বারা সূচিত হয় না।

(খ) ঋতুজ ভেদ (seasonal variation)

অনেক সময় কালীন সারির রেখাচিত্র পরীক্ষা করলে দেখা যায় একই বছরের মধ্যে বিভিন্ন ঋতুতে রেখাচিত্রটির উত্থান-পতন ঘটেছে। এই উত্থান-পতনের সময়গুলি অনেকটা সুনির্দিষ্ট—অর্থাৎ প্রতিবছর নির্দিষ্ট সময়ে একই ধরনের উত্থানপতন ঘটে থাকে। যেমন ৪.২নং চিত্রে কলকাতার জীবিকা নির্বাহ ব্যয়ের সূচক লক্ষ্য করলে দেখা যায় প্রতি বছরই শীতকালে এই রেখাচিত্রটি নিম্নাভিমুখী হয়েছে এবং বর্ষাকালে এর গতি ঊর্ধ্বমুখী হয়েছে। কারণ শীতকালে ভোগ্যবস্তুসমূহের (বিশেষত খাদ্যদ্রব্যের) দর কম থাকায় সূচকের মান কমে আসে, কিন্তু বর্ষাকালে এদের দর বাড়ার সাথে সাথে সূচকের মানও বৃদ্ধি পায়। অনুরূপভাবে, ভোগ্যবস্তুসমূহের মাসিক বিক্রির পরিমাণের কালীন সারি লক্ষ্য করলে দেখা যায় যে, প্রতি বছর পূজোর সময় বিক্রির পরিমাণ খুব বেড়ে যায় (অর্থাৎ কালীন সারির রেখাচিত্র ঊর্ধ্বমুখী হয়) এবং বর্ষার সময় বিক্রির পরিমাণ হ্রাস পায় (অর্থাৎ কালীন সারির রেখাচিত্র নিম্নাভিমুখী হয়)।

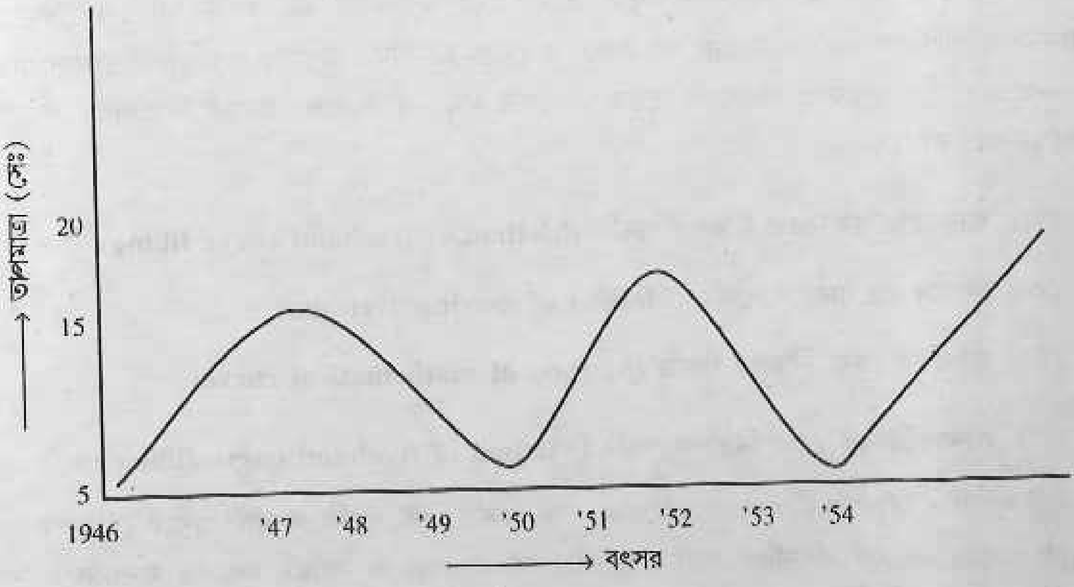


চিত্র নং—৪.২

একবছরের মধ্যে কালীন সারির রেখাচিত্রের এ রকম নিয়মিত উত্থানপতনকে ঋতুজ ভেদ (seasonal variation) বলা হয়ে থাকে। এ ধরনের উত্থান-পতন প্রধানত ঋতু পরিবর্তনের উপর (যেমন বর্ষার সময় দর বৃদ্ধি বা বিক্রি হ্রাসের বেলায়) বা সামাজিক পালা-পার্বণের উপর (যেমন পূজার সময় বিক্রির পরিমাণ বৃদ্ধির ক্ষেত্রে) নির্ভরশীল।

(গ) চক্রিল ভেদ (cyclical variation)

ঋতুজ ভেদের ক্ষেত্রে কালীন সারির উত্থান-পতন এক বৎসরের মধ্যে সীমাবদ্ধ থাকে। কিন্তু কোনো কোনো ক্ষেত্রে এক বৎসরের বেশি সময় পর পর কালীন সারির রেখাচিত্রে উত্থান-পতনের ছবি ফুটে ওঠে। সাধারণত এ রকম উত্থান-পতন ঋতুজ ভেদের উত্থান-পতনের মতো একেবারে নিয়মিতভাবে হয় না। যেমন, কোনো একবার উত্থান-পতন যদি তিন বৎসর পরে ঘটে তা হলে পরের বারও এটি ঠিক তিন বছর বাদেই হবে এমন কোনো কথা নেই। পতন-উত্থান-পতন বা উত্থান-পতন-উত্থান—এ রকম একটি পুরো সময়কে চক্র (cycle) বলে অভিহিত করা হয়ে থাকে। ব্যবসা-বাণিজ্যের ক্ষেত্রে বাজারে পর্যায়ক্রমে কিছুকাল পর পর তেজী ভাব (boom বা prosperity) এবং মন্দা ভাব (depression) দেখা যায়। অবশ্য কত সময় অন্তর এ রকম তেজী বা মন্দা ভাব দেখা দেবে তার ঠিক নেই। এ রকম ক্ষেত্রে একটি তেজী (বা মন্দা) ভাব থেকে আর একটি তেজী (বা মন্দা) ভাব পর্যন্ত সময়কে একটি চক্র (cycle) বলা হয়ে থাকে। তাই কালীন সারির এ ধরনের উত্থান-পতনকে চক্রিল ভেদ (cyclical variation) বলা হয়। নিচে (চিত্র নং—৪.৩) চক্রিল ভেদের একটি উদাহরণ দেওয়া হল।



চিত্র ন—৪.৩

কলকাতার সর্বনিম্ন তাপমাত্রার রেখাচিত্র

(ঘ) অনিয়মিত গতিধারা (irregular variation)

কালীন সারির অনিয়মিত গতিধারা দু ধরনের হতে পারে—(1) ঘটনাজাত (episodic) এবং (2) আকস্মিক (accidental)। কোনো ঘটনা (বা দুর্ঘটনা) যেমন দুর্ভিক্ষ, মহামারী, ভূমিকম্প, প্রবল ঝড়-বৃষ্টি ইত্যাদি অনেক সময় কালীন সারির গতিধারাকে বিশেষভাবে প্রভাবিত করতে পারে। উদাহরণস্বরূপ, দুর্ভিক্ষ বা মহামারীর সময় দ্রব্যমূল্যের সূচক অস্বাভাবিক বৃদ্ধি পেতে পারে। এগুলি ঘটনাজাত অনিয়মিত গতিধারার উদাহরণ। অপরপক্ষে, আকস্মিক এবং আপাতদৃষ্টিতে কারণহীনভাবেও কালীন সারির পরিবর্তন লক্ষ্য করা যায়। এ ধরনের পরিবর্তন অনিয়মিত গতিধারার উদাহরণ।

৩৬.৩ সুশাসিত গতিধারার পরিমাপ (Measurement of secular trend)

কোনো কালীন সারির সুশাসিত গতিধারার পরিমাপ করতে হলে উক্ত সারির অন্য তিনটি অংশ, যথা ঋতুজ ভেদ, চক্রিল ভেদ এবং অনিয়মিত গতিধারার প্রভাবকে সারি থেকে অপনীত (eliminate) করতে হবে। আগেই বলা হয়েছে ঋতুজ ভেদের উদ্ভব এক বৎসর সময়কালের মধ্যে হয়ে থাকে। এজন্য কোনো কালীন সারির এক বৎসরের যোগফল বা গড় নিলে তা ঋতুজ ভেদের প্রভাবমুক্ত হবে। সুতরাং সুশাসিত গতিধারা পরিমাপ করার সময় সাধারণত ঋতুজ ভেদের প্রভাব দূর করার জন্য কালীন সারির এক বৎসরের যোগফল বা গাণিতিক গড় নেওয়া হয়। এই রকম গড় থেকে অনিয়মিত গতিধারা এবং চক্রিল ভেদের প্রভাব দূর করতে পারলে সুশাসিত গতিধারার পরিমাপ পাওয়া যায়। এ উদ্দেশ্যে প্রধানত নিম্নলিখিত পন্থাগুলি অবলম্বন করা হয় :

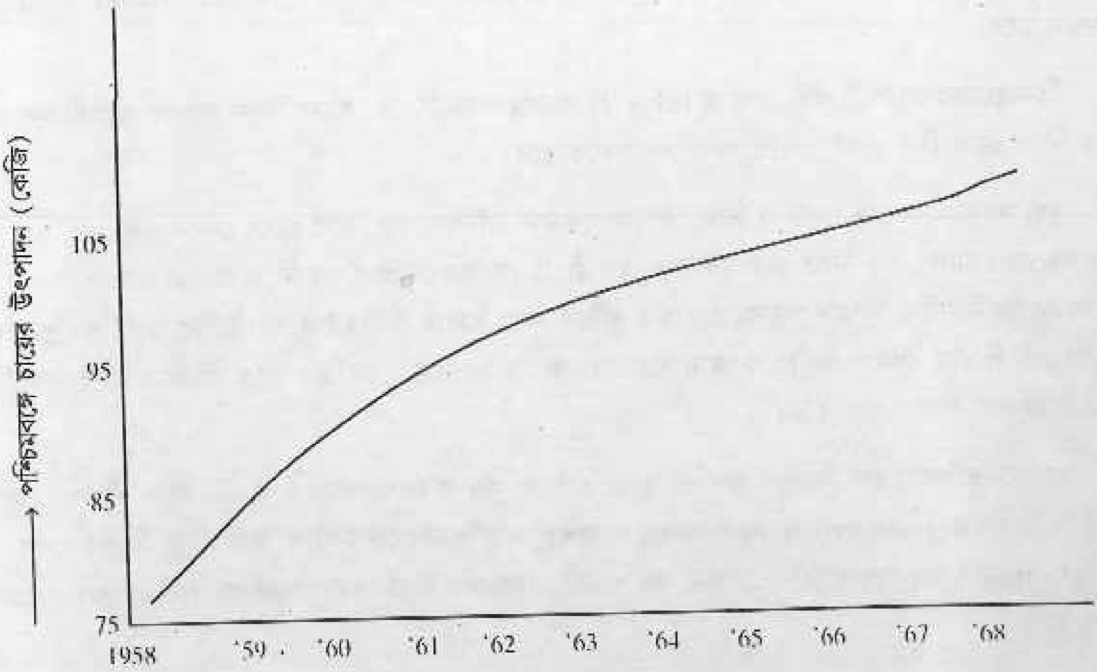
(অ) আন্দাজভিত্তিক রেখা নিরূপণ পদ্ধতি (Method of freehand curve filling)

(আ) চলমান গড় ব্যবহার পদ্ধতি (Method of moving averages)

(ই) গাণিতিক রেখা নিরূপণ পদ্ধতি (Method of mathematical curve)

(অ) আন্দাজভিত্তিক রেখা নিরূপণ পদ্ধতি (Method of freehand curve filling :

এটি সুশাসিত গতিধারা নিরূপণের সর্বাপেক্ষা সরল পদ্ধতি। এই পদ্ধতি অনুযায়ী প্রথমে লৈখিক কাগজ (graph paper)-এ মূল বাৎসরিক কালীন সারিটির বিভিন্ন সময় ও সংশ্লিষ্ট চলকের মানগুলিকে বিন্দু হিসাবে দেখানো হবে। তারপর ঐ বিন্দুগুলির ভেতর দিয়ে একটি মসৃণ রেখা (smooth curve) এমনভাবে আঁকা হয় যাতে এই রেখাটিকে ঐ বিন্দুগুলির সাধারণ ধারার সর্বাপেক্ষা ঘনিষ্ঠ প্রতিফলন (closest approximation) হিসাবে ধরা যেতে পারে। নিচে (৪নং চিত্রে) এই পদ্ধতিতে নির্ধারিত সুশাসিত গতিধারার উদাহরণ দেখান হল।



চিত্র নং— ৪.৪

পশ্চিমবঙ্গে চায়ের উৎপাদন

এই পদ্ধতির সবচেয়ে বড় সুবিধা হল এটি অত্যন্ত সহজ। সুশাসিত গতিধারা সরলরেখা হোক বা বক্ররেখা হোক এই পদ্ধতির সাহায্যে তা অতি সহজেই দেখানো যেতে পারে। অন্যদিকে এর সবচেয়ে বড় অসুবিধা হল এটি সম্পূর্ণভাবে যুক্তিনির্ভর। অর্থাৎ যে ব্যক্তি চিত্রটি আঁকবেন তার পছন্দ ও বিচারবুদ্ধির উপর এর রূপটি পুরোপুরিভাবে নির্ভরশীল।

৩৬.৪ চলমান গড় ব্যবহার পদ্ধতি

কোনো কালীন সারির তিন বছরের চলমান গড় নির্ণয় করতে হলে সর্বপ্রথম উপর্যুপরি (successive) প্রথম তিন বছরের অবেক্ষিত মানগুলির (observed values) যৌগিক গড় নিতে হবে। তারপর প্রথম বছরের অবেক্ষিত মানটি বাদ দিয়ে পরবর্তী তিন বছরের অর্থাৎ দ্বিতীয়, তৃতীয় ও চতুর্থ বছরের অবেক্ষিত মানের যৌগিক গড় নির্ণয় করতে হবে। এরকমভাবে প্রত্যেকবার সারির প্রথম দিক থেকে একটি করে

মান বাদ দিয়ে এবং নিচের দিকে একটি করে মান নিয়ে তিন বছরের যৌগিক গড় নিতে নিতে অগ্রসর হতে হবে—যতক্ষণ পর্যন্ত না কালীন সারির শেষ অবেক্ষিত মানটিও চলমান গড়ের অন্তর্ভুক্ত হচ্ছে। যে কোনো তিন বছরের অবেক্ষিত মানের গাণিতিক গড় এই তিনটি বছরের মধ্যবর্তী বছরের বিপরীতে দেখান হবে।

তিন বছরের জায়গায় পাঁচ, সাত বা $(2k + 1)$ বছরের—অর্থাৎ যে কোনো বিযুগ্ম সংখ্যক বছরের চলমান গড় নিতে হলে ঠিক একই পদ্ধতি অবলম্বন করতে হবে।

যুগ্ম সংখ্যক বছরের ক্ষেত্রেও নির্দিষ্ট বছরের চলমান যৌগিক গড় নিতে হবে। যেমন, দুই, চার, ছয় বা $2k$ বছরের যৌগিক গড় নিতে হলে দুই, চার, ছয় বা $2k$ বছরের যৌগিক গড় নিয়ে অগ্রসর হতে হবে। তবে এরকমভাবে নির্ধারিত চলমান গড়সমূহের প্রতি দুটিকে নিয়ে আবার দ্বিতীয় ক্ষেপে যৌগিক গড় নির্ণয় করতে হবে। এই দ্বিতীয় ক্ষেপে নির্ণীত চলমান গড় যে বছরের বিপরীতে অবস্থিত হবে গড়টিকে সেই বছরের প্রতিনিধিমূলক বলে ধরতে হবে।

আন্দাজভিত্তিক রেখা নিরূপণ পদ্ধতির মতো চলমান গড় ব্যবহার পদ্ধতিও একটি সরল পদ্ধতি। কিন্তু এটি ব্যক্তিনির্ভর (subjective) নয়। আবার, যেহেতু এই পদ্ধতিতে নির্ণীত সারিসমূহ বিশেষ কোনো সূত্র (formula) অনুসরণ করে না সেজন্য এই পদ্ধতি ব্যবহারের দ্বারা কোনো পূর্বাভাষ (forecast) দেওয়া সম্ভব নয়।

৩৬.৪.১ চলমান গড় পদ্ধতির ব্যবহারের তাত্ত্বিক যুক্তি (Theoretical justification for moving average method)

যেহেতু কয়েকটি বছরের গড় মান নিয়ে চলমান গড় নির্ণীত হয়, সেজন্য কালীন সারির স্বল্পকালীন উত্থান-পতন (যেমন ঋতুজ ভেদ বা অনিয়মিত গতিধারা) চলমান গড়ের সাহায্যে অপনীত করা যায়। তাছাড়া দীর্ঘকালীন উত্থান-পতন—যা নাকি চক্রিল ভেদ দ্বারা প্রদর্শিত হয়—তার প্রভাবও খানিকটা দূর করা সম্ভব হয়। বিশেষ করে যদি কোনো কালীন সারির চক্রিল ভেদ (cyclical variation)-এর প্রতিটি চক্র (cycle)-এর পরিমাপ প্রায় সমান হয় তা হলে ঐ পরিমাপের সমান (বা তার গুণিতক) চলমান গড় নিলে কালীন সারিটি বহুলাংশে চক্রিল ভেদের প্রভাবমুক্ত হয়। অর্থাৎ চলমান গড় ব্যবহার পদ্ধতি অবলম্বন করে সুশাসিত গতিধারা ছাড়া কালীন সারির অন্যান্য অংশের প্রভাব অনেকাংশে দূর করা সম্ভব হয়। ফলে এই পদ্ধতিটি সুশাসিত গতিধারা নির্ণয়ের উপযোগী।

উদাহরণ : ৪.১

নিচের সারণীটিতে 1951 থেকে 1961 পর্যন্ত ভারতের আকরিক লোহার উৎপাদনের পরিমাণ দেখান হয়েছে। লক্ষ্য করলে দেখা যাবে যে, যদিও কালীন সারিটির স্বল্পমেয়াদী উত্থান-পতন আছে, তা হলেও সুশাসিত গতিধারার ঊর্ধ্বমুখী প্রবণতা অত্যন্ত সুস্পষ্ট। স্বল্পমেয়াদী উত্থান-পতনের প্রভাব দূর করে সুশাসিত গতিধারাকে

আরও সুস্পষ্টভাবে প্রকাশ করার জন্য 5-বছরের চলমান গড় নেওয়া যেতে পারে। নিচের সারণীতে (সারণী-১) এরকম গড় নেওয়া হয়েছে।

সারণী—১

পশ্চিমবঙ্গে আকরিক লৌহ উৎপাদনের পরিমাণ (হাজার টনের এককে)

বর্ষ	আকরিক লৌহের উৎপাদন (হাজার টন)	5-বৎসরের চলমান যোগফল (5-year moving total)	5-বৎসরের চলমান গড় (5-year moving average)
1951	675.8		
1952	635.5		
1953	572.3	3332.4	666.5
1954	750.1	3363.8	672.8
1955	698.7	3457.8	691.6
1956	707.2	3815.9	763.2
1957	729.5	4105.4	821.1
1958	930.4	4879.2	975.8
1959	1039.6	6082.0	1216.4
1960	1472.5		
1961	1910.0		

5-বছরের চলমান গড়ের সারিটি পর্যবেক্ষণ করলে দেখা যাবে এটি আকরিক লৌহ উৎপাদনের প্রকৃত সারিটির থেকেও অনিয়মিত উত্থান-পতনের প্রভাবমুক্ত। তুলনামূলক রেখাচিত্র অঙ্কন করলে এটি আরও পরিষ্কারভাবে পরিলক্ষিত হয়।

৩৬.৫ গাণিতিক রেখা নিরূপণ পদ্ধতি (Method of mathematical curves)

সুশাসিত গতিধারা নির্ণয়ের জন্য গাণিতিক রেখা নিরূপণ পদ্ধতি (Method of mathematical curves)-এর ব্যাপক ব্যবহার হয়ে থাকে। এই পদ্ধতির সর্বপ্রধান সুবিধা হল, এটি সম্পূর্ণরূপে বস্তুনিষ্ঠ (objective)—

ব্যবহারকারীর পছন্দ- অপছন্দের ওপর নির্ভরশীল নয়। এই পদ্ধতির আর একটি মস্ত সুবিধা হল, এর দ্বারা সুশাসিত গতিধারার পূর্বাভাষ (forecasting of secular trend) দেয়া যায়।

গাণিতিক রেখা নিরূপণের জন্য সর্বপ্রথমে নির্ণেয় সুশাসিত গতিধারাকে কি ধরনের গাণিতিক রেখা (mathematical curve)-র দ্বারা চিহ্নিত করা যেতে পারে তা স্থির করতে হবে। এ উদ্দেশ্যে প্রদত্ত কালীন সারির রেখাচিত্র (graph)-টি ভালভাবে খালিচোখে পরীক্ষা করে দেখা হয়। এ রকম পরীক্ষার সাহায্যে সুশাসিত গতিধারার রূপ সম্বন্ধে একটি মোটামুটি ধারণা করে নেওয়া হয় এবং এক্ষেত্রে কী ধরনের গাণিতিক রেখার ব্যবহার সবচেয়ে সুবিধাজনক হবে তা স্থির করা হয়। অধিকাংশ ক্ষেত্রেই এই রেখাচিত্রের রূপটি একটি ঘাতজ অপেক্ষক (polynomial of suitable degree) হয়। এরপর একটি নির্দিষ্ট রেখা নিরূপণ পদ্ধতি অনুসরণ করে এই গাণিতিক রেখাটির ধ্রুবকসমূহে (parameters)-র প্রাক্কলন (estimation) করা হয়। ধ্রুবক প্রাক্কলনের সবচেয়ে প্রচলিত পদ্ধতি হল **লঘিষ্ঠ বর্গসমষ্টি পদ্ধতি** (method of least squares)। এছাড়া গোষ্ঠীগড় পদ্ধতিও (group average method) কোনো কোনো পরিস্থিতিতে ব্যবহার করা হয়ে থাকে।

উদাহরণ : নিচের কালীন সারির জন্য একটি ঋজুরৈখিক গতিধারা নিরূপণ করতে হবে।

বর্ষ	পশ্চিমবঙ্গে হাসপাতালের শয্যাসংখ্যা (হাজারে)	বর্ষ	পশ্চিমবঙ্গে হাসপাতালের শয্যাসংখ্যা (হাজারে)
1979	55.5	1983	61.9
1980	58.0	1984	63.7
1981	58.4	1985	64.7
1982	59.9	1986	65.3

আমরা সময়জ্ঞাপক প্রতীক নিলাম $t = 2$ (বর্ষ–1982-র শেষ) এবং সংশ্লিষ্ট সময়ে শয্যাসংখ্যাকে y_t দিয়ে সূচিত করলাম। তাহলে যদি সুসংহত গতিধারা হয়

$$T_t = a_0 + a_1 t,$$

তবে a_0 ও a_1 ধ্রুবক দুটিকে "লঘিষ্ঠ-বর্গ পদ্ধতিতে" নির্ণয় করব। এখানে

$$\sum_t (y_t - a_0 - a_1 t)^2$$

a_0 ও a_1-এর সাপেক্ষে ন্যূনতম বানাতে হবে। নর্ম্যাল সমীকরণ দুটি হবে (যেহেতু $t = -7, -5, \ldots\ldots, -1, 1, \ldots\ldots 5, 7$).

$$\left.\begin{array}{l} \sum y_t = 8a_0 \\ \sum ty_t = 168a_1 \end{array}\right\} \quad \text{বা} \quad \left.\begin{array}{r} 487.4 = 8a_0 \\ 120.0 = 168a_1 \end{array}\right\}$$

ফলে $a_0 = 60.925$. $a_1 = 0.714$

এবং ঋজুরৈখিক সুসংহত গতিধারার সমীকরণ হবে :

$$T_t = 60.925 + 0.714t.$$

উদাহরণ : ১৯৮০-৮১ সালের মূল্যস্তর অনুসারে স্থূল জাতীয় উৎপাদন (GDP) ভারতবর্ষে কীভাবে বেড়েছে তা নিচে দেখানো হল :

বর্ষ	স্থূল জাতীয় উৎপাদন (হাজার কোটি টাকায়)	বর্ষ	স্থূল জাতীয় উৎপাদন (হাজার কোটি টাকায়)
1991-92	237	1995-96	305
1992-93	249	1996-97	328
1993-94	262	1997-98	344
1994-95	283		

এই কালীন সারিটি একটি লেখচিত্রে দেখানো হলে বোঝা যাবে এক্ষেত্রে ঋজুরৈখিক গতিধারার তুলনায় এক্সপোনেনশিয়াল গতিধারা বেশি যথাযথ হবে। এর সমীকরণ ধরা যাক

$$T_t = a_0 a_1^t$$

(t = বর্ষ − 1994-95).

লগারিদম নিলে দাঁড়াবে $\log T_t = \log a_0 + (\log a_1)t$.

এক্ষেত্রে নর্মাল সমীকরণ দুটি হবে :

$$\sum_t \log y_t = 7 \log a_0$$

$$\sum_t t \log y_t = 28 \log a_1$$

এখানে y_t হল স্থূল জাতীয় উৎপাদনের পরিমাণ।

t	$\log y_t$
−3	2.3747
−2	2.3962
−1	2.4183
0	2.4518
1	2.4843
2	2.5159
3	2.5366

এই সারণী থেকে : $\sum_t \log y_t = 17.1778$

এবং $\sum_t t \log y_t = 0.7911$

সুতরাং $\log a_0 = 2.4540$,

$\log a_1 = 0.02825$.

তাই নির্ণেয় গতিধারার রূপ হল—

$$\log T_t = 2.4540 + 0.02825t$$

বা, $T_t = 284.45 \times (1.0672)^t$

৩৬.৬ ঋতুজ ভেদের পরিমাপ (Measurement of seasonal fluctuations)

সুশাসিত গতিধারার ন্যায় ঋতুজ ভেদের পরিমাপ করার প্রয়োজনীয়তাও অনেক সময় অনুভূত হয়। যেমন, পণ্যদ্রব্যাদি বাজারে রাখার জন্য বছরের বিভিন্ন ঋতুতে পণ্যের কীরকম চাহিদা হয় তা বিক্রেতার বিশেষভাবে জানা দরকার। কারণ, চাহিদা অনুযায়ী যোগান স্থির করতে হবে।

আগেই বলা হয়েছে যে, এক বৎসরের কম সময়ের কালীন সারির উত্থান-পতন ঋতুজ ভেদের দ্বারা নির্দেশিত হয়। এরকম সময়ের ব্যাপ্তি একটি ঋতু, একটি মাস বা একটি দিন হতে পারে। তবে সাধারণত মাস বা ঋতুর প্রচলনই বেশি।

৪.৪.১ ঋতুজ ভেদের কয়েকটি পরিমাপ পদ্ধতি (Certain methods of measuring seasonal fluctuations)

কয়েকটি প্রতীক (symbols) : পূর্বে কালীন সারিকে চারটি উপাংশে (components) ভাগ করা হয়েছে। যথা; (১) সুশাসিত গতিধারা (secular trend variation) (২) ঋতুজ ভেদ (seasonal variation), (৩) চক্রিল ভেদ (cyclical variation) এবং (৪) অনিয়মিত গতিধারা (irregular variation)। এগুলিকে প্রতীকে নিম্নরূপে দেখান হয় :

T_t = "t" সময়বিন্দুতে সুশাসিত গতিধারাজনিত অংশ,

S_t = "t" সময়বিন্দুতে ঋতুজ ভেদজনিত অংশ,

C_t = "t" সময়বিন্দুতে চক্রিল ভেদজনিত অংশ

I_t = "t" সময়বিন্দুতে অনিয়মিত গতিধারার অংশ

বহুল প্রচলিত একটি প্রথা অনুযায়ী কালীন সারিকে উক্ত চারটি অংশের গুণফল হিসেবে ধরা হয়। অর্থাৎ y_t যদি কালীন সারির "t" সময়বিন্দুর মান হয়, তা হলে

$$y_t = T_t \times S_t \times C_t \times I_t \qquad (৪.৪.১)$$

ধরা যাক, y_t' —এই মানটি ঋতুজ ভেদের প্রভাবশূন্য।

অর্থাৎ, $y_t' = T_t \times C_t \times I_t$ (৪.৪.২)

$$\text{তাহলে,}\quad \frac{y_t}{y_t'} = \frac{T_t \times S_t \times C_t \times I_t}{T_t \times C_t \times I_t} = S_t$$

অর্থাৎ, y_t-কে y'_t দ্বারা ভাগ করলে ঋতুজ ভেদের পরিমাপ পাওয়া যায়। বাস্তবে সুশাসিত গতিধারা (T) চক্রিল ভেদ (C) এবং অনিয়মিত গতিধারার গুণফল হিসাবে y_t'-র যথাযথ মান নির্ণয় করা সাধারণত সম্ভব হয় না। তবে অনেকটা কাছাকাছি মূল্যায়নের চেষ্টা হয়ে থাকে। নিচে বর্ণিত পদ্ধতিগুলি কোনো না কোনোভাবে উপরোক্ত তাত্ত্বিক বিচার পদ্ধতি থেকে উদ্ভূত।

৪.৪.২ মাসিক বা ত্রৈমাসিক গড় পদ্ধতি (Method of monthly or quarterly average)

এটি হল ঋতুজ ভেদ নির্ণয়ের সবচেয়ে সরল পদ্ধতি। এই পদ্ধতি অনুসরণকালে ধরে নেওয়া হয় যে, নির্দিষ্ট কালীন সারিটি সাধারণভাবে সুশাসিত গতিধারা এবং চক্রিল ভেদের প্রভাবমুক্ত। অর্থাৎ সারিটি শুধুমাত্র ঋতুজ ভেদ ও অনিয়মিত গতিধারার দ্বারা প্রভাবিত। অনিয়মিত গতিধারার প্রভাব দূর করার জন্য সারিটির মাসিক (বা ত্রৈমাসিক) মানগুলির যৌগিক গড় নেওয়া হয়ে থাকে। স্পষ্টতই পদ্ধতিটি অধিকাংশ ক্ষেত্রেই প্রযোজ্য নয়। কারণ এমন কালীন সারি খুব কমই পাওয়া যায় যা সুশাসিত গতিধারা এবং চক্রিল ভেদের প্রভাবমুক্ত।

৩৬.৬.১ চলমান গড় দ্বারা বিভাজন পদ্ধতি (Ratio to moving average method)

এ পদ্ধতি অনুযায়ী কালীন সারির মাসিক (বা ত্রৈমাসিক) মানগুলির বারোমাসের চলমান গড় নেওয়া হয়ে থাকে। আগেই বলা হয়েছে যে, এই প্রক্রিয়ায় কালীন সারিটি ঋতুজ ভেদের প্রভাবমুক্ত হবে। গড় নেওয়ার ফলে অনিয়মিত গতিধারার প্রভাবও বহুলাংশে দূরীভূত হয়। অনিয়মিত গতিধারা I-কে নিম্নোক্তরূপে প্রকাশ করা যায় : $I = I' \times I''$

যেখানে, I' = বারোমাসের চলমান গড় নেওয়ার ফলে দূরীভূত I–এর অংশ ;

I'' = বারোমাসের চলমান গড় নেওয়ার পরও I-এর যে অংশ দূরীভূত হয়নি।

তাহলে,

$$y = T \times S \times C \times I$$
$$= T \times S \times C \times I' \times I''$$

বারোমাসের চলমান গড় নেওয়ার ফলে y-র $S \times I'$ অংশ দূরীভূত হয় এবং $T \times C \times I''$ অংশ থাকে। চলমান গড়ের এই অংশ (অর্থাৎ $T \times C \times I''$) দ্বারা কালীন সারির মূল মানকে (অর্থাৎ $y = T \times S \times C \times I = T \times S \times C \times I' \times I''$) ভাগ করলে আমরা পাই :—

$$\frac{y}{T \times C \times I''} = \frac{T \times S \times C \times I' \times I''}{T \times C \times I''} = S \times I'$$

অর্থাৎ, উপরোক্ত পদ্ধতি অবলম্বন করলে কালীন সারি থেকে অনিয়মিত গতিধারার একাংশ সহ ঋতুজ ভেদ নির্ণয় করা সম্ভব। তবে যেহেতু বারোমাসের চলমান গড় নেওয়া হয়ে থাকে সেজন্য প্রথম ছয় মাসের এবং শেষের ছয় মাসের মানগুলি পাওয়া যায় না। কালীন সারির, বাকি প্রতিটি মাসকে এর বিপরীতস্থ বারোমাসের চলমান গড়ের মান দিয়ে ভাগ করে নির্ণেয় $S \times I'$-র মান নির্ণয় করা হয়ে থাকে। এরপর I-এর প্রভাব দূর করার জন্য $S \times I$ -এর মাসিক মানগুলির কয়েক বছরের গড় নেওয়া হয়ে থাকে। যদি অনিয়মিত গতিধারার প্রভাব সামান্য হয় তা হলে গাণিতিক গড় নিলেই চলে। কিন্তু মাসিক মানগুলির কয়েকটি যদি অস্বাভাবিক রকম বেশি বা কম হয় তা হলে মধ্যমা (median)-এর ব্যবহার সুবিধাজনক কিংবা অস্বাভাবিক মানগুলি বাদ দিয়ে গড় নেওয়া যেতে পারে। তুলনায় সুবিধার জন্য **যৌগিক ঋতুজ ভেদের সূচকের গড়** 100 ধরা হয়।

৩৬.৬.২ সুশাসিত গতিধারার দ্বারা বিভাজন পদ্ধতি (Ratio to trend method)

এই পদ্ধতি অনুযায়ী প্রথমে নির্দিষ্ট কালীন সারির সুশাসিত গতিধারা নির্ণয় করা হয়। তারপর মূল কালীন সারির মানগুলিকে তাদের বিপরীতস্থ সুশাসিত গতিধারার মান দিয়ে ভাগ করা হয়। অর্থাৎ $y = T \times S \times C \times I$-কে T দ্বারা ভাগ করে $C \times S \times I$ পাওয়া যায়। তারপর বিভিন্ন বছরের $C \times S \times I$-র মাসিক (বা ত্রৈমাসিক) মানগুলির গাণিতিক গড় (আ)-এ বর্ণিত পদ্ধতি অনুযায়ী নির্ণয় করা হয়। এখানে ধরে নেওয়া হয় যে, চক্রিল ভেদের প্রভাব নগণ্য এবং এ রকম গড় নেওয়ার ফলে I-এর প্রভাবও অনেকাংশে দূরীভূত হয় এবং প্রাপ্ত সারিটি শুধু S-এর প্রভাবমুক্ত হয়। এই সারিটিকে শতকরা হিসাবে প্রকাশ করে **ঋতুজ ভেদের সূচক (seasonal index)** পাওয়া যায়।

৩৬.৭ সূচক সংখ্যা (Index Numbers)

৩৬.৭.১ সূচক সংখ্যার ধারণা (Introduction concept of index numbers)

সূচক সংখ্যার দ্বারা কতকগুলি পরস্পর সম্পর্কযুক্ত চলের (related variables) পরিবর্তন পরিমাপ করা হয়। উদাহরণস্বরূপ, সময়ের পরিবর্তনের সাথে সাথে বিভিন্ন পণ্যদ্রব্যের দরের হেরফেরের পরিমাপ করা হয় দরের সূচক (price index)-এর দ্বারা। এরূপ ক্ষেত্রে পণ্যর দরসমূহ হল চল। অনুরূপভাবে, দুটি বিভিন্ন সময়ের শিল্পজাত জিনিসপত্রের উৎপাদনের পরিমাণের প্রভেদ পরিমাপ করে উৎপাদনের সূচক (index of production) নির্ণয় করা হয়।

আগে প্রধানত পণ্যের দরের পরিবর্তন পরিমাপ করার জন্যই সূচক সংখ্যার ব্যবহার হতো। এখন এই ব্যবহার আরও অনেক ব্যাপকভাবে করা হয়। এ সত্ত্বেও দরের পরিবর্তনের পরিমাপক সূচকগুলির গুরুত্বই খুব বেশি। বিভিন্ন দ্রব্যের বিশেষ করে নিত্যব্যবহার্য দ্রব্যের দরের পরিবর্তনের সাথে সাথে কর্মচারী বা শ্রমিকদের বেতন, মহার্ঘ ভাতা, বাড়িভাড়ার ভাতা ইত্যাদিরও পরিবর্তন করা দরকার হয়। এজন্য বিভিন্ন ধরনের দরের গতিপ্রকৃতির উপর লক্ষ্য রাখা শ্রমিক-মালিক, সমাজকর্মী, রাজ্য কিংবা কেন্দ্রীয় সরকারের পক্ষে প্রয়োজন হয়। এই উদ্দেশ্যে সবচেয়ে বহুল-প্রচলিত দরের সূচক হল পাইকারি দরের সূচক (wholesale price index) এবং ভোক্তাদের দরের সূচক (consumer price index বা CPI)।

৩৬.৭.২ সূচক সংখ্যায় ব্যবহৃত কয়েকটি প্রতীক (Symbols used in index numbers)

একটি সময়ের অবস্থার তুলনায় আর একটি অবস্থার (পরিমাপ সূচক সংখ্যার দ্বারা) যখন তুলনা করা হয় তখন যে সময়কে ভিত্তি করে এই তুলনা করা হয় তাকে বলা হয় ভিত্তিকাল (base period) এবং যে সময়ের তুলনা করা হয় তাকে বলা হয় চলতিকাল (current period)। এরূপ ক্ষেত্রে কোনো একটি পণ্যের সূচক সংখ্যার জন্য নিম্নলিখিত প্রতীকগুলি ব্যবহার করা হয়।

p_0 = পণ্যের ভিত্তিকালের দর (base period price of the commodity)।

p_1 = পণ্যের চলতিকালের দর (current period price of the commodity)।

q_0 = ভিত্তিকালে পণ্যটি ব্যবহারের পরিমাণ (quantity used of the commodity during the base period)।

q_1 = চলতিকালে পণ্যটি ব্যবহারের পরিমাণ (quality used of the commodity during the current period)।

এখানে (p_1-p_0) হল ভিত্তিকাল থেকে চলতিকালে দরের পরিবর্তনের গাণিতিক পরিমাপ। অপরপক্ষে $\frac{p_1}{p_0}$ হল এ রকম পরিবর্তনের আপেক্ষিক (relative) পরিমাপ। $\frac{p_1}{p_0}$-কে আপেক্ষিক দর (price relative) বলে অভিহিত করা হয়। অনুরূপভাবে, $\frac{q_1}{q_0}$-কে আপেক্ষিক পরিমাণ (quantity relative) বলে অভিহিত করা

হয়। প্রতিটি ভিন্ন ভিন্ন পণ্যের (যথা চাল, ডাল, কাপড়, লোহা, বাড়িভাড়া ইত্যাদি) জন্য আলাদা আপেক্ষিক দর ও আপেক্ষিক পরিমাণ নিরূপণ করা যেতে পারে। ভিন্ন ভিন্ন আপেক্ষিক দরগুলির গড় নির্ণয় করে তাকে দরের সূচক বলে গণ্য করা যেতে পারে। কিন্তু এরকম গড় নির্ণয়ে নানা সমস্যা দেখা দেয়। এসব সমস্যা সূচক সংখ্যা নির্ণয়ের সমস্যার অন্তর্ভুক্ত।

৩৬.৮ সূচক সংখ্যা নির্ণয়ের সমস্যাসমূহ (Problems connected with the construction of index number)

সূচক সংখ্যা নির্ণয়কালে প্রধানত নিম্নলিখিত সমস্যাগুলির সম্মুখীন হতে হয়।

(ক) সূচক সংখ্যা ব্যবহারের উদ্দেশ্য স্থির করা ;

(খ) ভিত্তিকাল নির্ণয় ;

(গ) কোন কোন পণ্যকে সূচক সংখ্যার অন্তর্ভুক্ত করা হবে তা স্থির করা ;

(ঘ). প্রয়োজনীয় রাশিতথ্য সংগ্রহ ;

(ঙ) বিভিন্ন পণ্যসংক্রান্ত রাশিতথ্যের একত্রীকরণ।

(ক) সূচক সংখ্যা ব্যবহারের উদ্দেশ্য :

কোনো সূচক সংখ্যা সঠিকভাবে নির্ণয় করার আগে তা কী উদ্দেশ্যে ব্যবহার করা হবে সে সম্পর্কে স্পষ্ট ধারণা থাকা দরকার। জীবিকা নির্বাহণ ব্যয়ের সূচক (consumer price index)-এর কথা ধরা যাক। এটি নির্ণয়ের জন্য যেসব পণ্যের দর সংগ্রহ করা হবে সেগুলি দৈনন্দিন জীবনযাত্রা নির্বাহের সাথে সম্পর্কযুক্ত হওয়া দরকার। দৈনন্দিন জীবনযাত্রা নির্বাহের জন্য লোকে প্রধানত খুচরো দরে জিনিস কিনে থাকে। সুতরাং এসব ক্ষেত্রে পণ্যের খুচরো দরই সংগ্রহ করা দরকার—পাইকারি দর নয়। অন্যদিকে পাইকারি দরের সূচক (wholesale price index) নির্ণয়কালে পণ্যের পাইকারি দর নেওয়া দরকার।

(খ) ভিত্তিকাল নির্ণয় :

আগেই বলা হয়েছে যে, ভিত্তিকালের (base period) তুলনায় চলতিকাল (current period)-এর দর অনুপাত (যা $\frac{p_1}{p_0}$ -এর দ্বারা সূচিত হয়) নির্ণয়ের দ্বারা সূচক সংখ্যা প্রস্তুত করা হয়। এই দর-অনুপাত শতকরা হিসাবে নির্ণীত হয়। অর্থাৎ ভিত্তিকালে কোনো পণ্যের দর যদি 100 টাকা হয় তবে চলতিকালে তা কত হবে—আপেক্ষিক দরের দ্বারা তা প্রকাশ করা হয়। প্রতীকে দেখাতে হলে এটা হবে $100 \times \frac{p_1}{p_0}$।

ভিত্তিকাল নির্ণয়ের সময় বিশেষ সাবধানতা অবলম্বন করা দরকার। যে সময়ে দর বিশেষ কোনো কারণে খুব বেড়ে যায় (যেমন যুদ্ধ বা প্রাকৃতিক বিপর্যয়ের সময়) বা কমে যায় (যেমন মন্দার সময়), সেরকম অস্বাভাবিক

সময়কে ভিত্তিকাল ধরা ঠিক হবে না। সাধারণভাবে একটি সুস্থিত, স্বাভাবিক (normal) সময়কেই ভিত্তিকাল ধরা উচিত।

ভিত্তিকাল ও চলতিকালের মধ্যে সময়ের পার্থক্য খুব বেশি হওয়া ঠিক নয়। কারণ, এ রকম হলে ভিত্তিকালের বাজারের অবস্থা এবং জনসাধারণের জীবনযাত্রার মানের সঙ্গে চলতিকালের বাজারের অবস্থা এবং জনসাধারণের জীবনযাত্রার মান তুলনীয় হয় না। অনেক পূর্বতন ভিত্তিকাল সূচক সংখ্যা নির্ণয়ের জন্য নিলে, নির্বাচিত পণ্যের মধ্যে বেশ কিছু পণ্য পরবর্তীকালে দুষ্প্রাপ্য বা অপ্রাপ্য হয়ে যেতে পারে, এর ফলে এগুলির সশব্দ তথ্যসংগ্রহ চলতিকালে সম্ভব হয় না। এজন্য অনেক ক্ষেত্রে কয়েক বছর (যেমন পাঁচ বছর) পর পর ভিত্তিকাল নতুন করে স্থির করা দরকার হয়।

ভিত্তিকাল খুব দীর্ঘ সময়ের (যেমন দশ বছরের গড়) বা কম সময়ের (যেমন একদিন বা একশ) হওয়া উচিত নয়। কারণ, এতে নানাপ্রকার সমস্যা দেখা দিতে পারে। প্রথম ক্ষেত্রে ভিত্তিকালের মধ্যেই দরের ওঠানামার ফলে সে-সময়ে স্থির দরস্তর ছিল এমন ভাবা অসঙ্গত হবে। অন্যদিকে, দ্বিতীয় ক্ষেত্রে ভিত্তিকালের দরস্তর আকস্মিক ঘটনাবলী দ্বারা প্রভাবিত হতে পারে।

(গ) কোন্ কোন্ পণ্যকে সূচক সংখ্যার অন্তর্ভুক্ত করা হবে তা স্থির করা :

সময়ের স্বল্পতা এবং অন্যান্য ব্যবহারিক অসুবিধার জন্য বাজারের প্রতিটি পণ্যকে সূচকসংখ্যা নির্ণয়ে অন্তর্ভুক্ত করা সম্ভব নয়। কারণ এরজন্য প্রচুর সময়ের এবং খরচের প্রয়োজন হবে এবং নির্দিষ্ট একটি সময়ের মধ্যে এবং বরাদ্দ আয়ের মধ্যে সূচকসংখ্যা নির্ণয় করা সম্ভব নয়। এজন্য প্রতিনিধিমূলক কয়েকটি পণ্যকে নমুনা হিসাবে বাছাই করা হয়ে থাকে। পণ্যগুলি এমনভাবে নির্বাচন করা হয় যাতে এদের গড় দরের গতিবিধি বাজারের সমস্ত পণ্যের গড় দরের গতিবিধির সমধর্মী হয়।

নমুনা সংগ্রহ করার উদ্দেশ্যে পণ্যগুলিকে প্রথমে কয়েকটি প্রধান গোষ্ঠী (group)-এ ভাগ করা হয়। যেমন জীবিকা নির্বাহক পণ্যগুলিকে সাধারণত এই পাঁচটি গোষ্ঠীতে ভাগ করা হয় : (1) খাদ্য, (2) জ্বালানি ও আলো, (3) পরিধেয়, (4) বাসস্থান ও (5) বিবিধ। প্রতিটি গোষ্ঠীর আবার অনেকগুলি উপগোষ্ঠী (sub-group) থাকে। যেমন খাদ্যের উপগোষ্ঠী হল চাল, গম, মাছ, মাংস, নুন, চিনি, তেল ইত্যাদি। জ্বালানির উপগোষ্ঠী হল কেরোসিন, রান্নার গ্যাস, কয়লা ইত্যাদি। বিভিন্ন উপগোষ্ঠীর কয়েকটি পণ্যকে নমুনা হিসাবে চয়ন করা হয়ে থাকে। সঠিক কতগুলি পণ্য সংগ্রহ করা দরকার তাঁর কোনো বাঁধাধরা নিয়ম নেই। তবে সাধারণভাবে বলা যায় পণ্যসংখ্যা খুব বেশি কখনও হবে না।

(ঘ) প্রয়োজনীয় রাশিতথ্য সংগ্রহ :

রাশিতথ্য সংগ্রহের সময় বিশেষ সাবধানতা অবলম্বন করা দরকার। বিভিন্ন বাজারে একই পণ্যের বিভিন্ন দর হতে পারে। তাছাড়া গুণগত মান (quality) অনুযায়ীও দরের তফাৎ হয়। সূচকসংখ্যা নির্ণয়ের জন্য পণ্যের দর সংগ্রহের সময় এসব কথা মনে রাখা দরকার এবং যাতে সব প্রকার দরই সংগ্রহ করা হয়ে থাকে সেদিকে নজর রাখা প্রয়োজন।

(ঙ) **বিভিন্ন পণ্যসংক্রান্ত রাশিতথ্যের একত্রীকরণ :**

সূচক সংখ্যা নির্ণয় প্রক্রিয়ায় প্রধানত দুটি পদ্ধতি ব্যবহার করা হয়— (1) দরের যৌগিক পদ্ধতি (aggregate price method), (2) গড় দর অনুপাত পদ্ধতি (average price relative method)।

দরের যৌগিক পদ্ধতি (Aggregate price method) :

সরল যৌগিক পদ্ধতি (Simple aggregate method) :

পণ্যসমূহের চলতিকালের দরগুলির যোগফলের সঙ্গে ভিত্তিকালের দরগুলির যোগফল তুলনা করে একটি সূচকসংখ্যা নির্ণয় করা যেতে পারে। প্রতীকে প্রকাশ করলে সূচকসংখ্যা হবে

$$I_{01} = \frac{\sum_{i=1}^{n} p_{1i}}{\sum_{i=1}^{n} p_{0i}}, \qquad (৪.৫.১)$$

যেখানে p_{0i} = i-তম পণ্যের ভিত্তিকালের দর,

এবং p_{1i} = i-তম পণ্যের চলতিকালের দর

এ রকম সূচককে সরল যৌগিক সূচক (simple aggregative index) বলে (বাস্তবক্ষেত্রে ব্যবহারের সময় I_{01} -কে 100 দ্বারা গুণ করে শতকরা হিসাবে প্রকাশ করা হয়)।

এই ধরনের সূচক ব্যবহারের সবচেয়ে বড় অসুবিধা হল এখানে সব ধরনের পণ্যের দরের উপর সমান গুরুত্ব দেওয়া হয়েছে। কিন্তু প্রকৃতপক্ষে সমান গুরুত্ব দান সঙ্গত নয়। যেমন খাদ্যদ্রব্যের ক্ষেত্রে উপরোক্ত পদ্ধতির চালের দরের যে গুরুত্ব আপেলের দরেরও সেই সমান গুরুত্ব। প্রকৃত চিত্রটি কিন্তু সম্পূর্ণ ভিন্ন। চাল সর্বাধিক ব্যবহৃত ও গুরুত্বপূর্ণ খাদ্যদ্রব্য। অন্যদিকে খাদ্যদ্রব্য হিসাবে আপেলের ব্যবহার খুবই সীমিত। এই দুটির দরকে একইভাবে গুরুত্ব দিয়ে ধরা যেতে পারে না। এই কারণে এই পদ্ধতির বিশেষ ব্যবহার দেখা যায় না।

ভারযুক্ত যৌগিক পদ্ধতি (Weighted aggregate method) :

নির্ণেয় সূচক সংখ্যাটিকে বাস্তবোচিত করতে হলে সূচকসংখ্যায় ব্যবহৃত পণ্যসমূহের দরের ভারগুলি (weight) এমন হওয়া দরকার যাতে পণ্যদরগুলির প্রকৃত গুরুত্ব সূচক সংখ্যাটিতে সঠিকভাবে প্রতিফলিত হয়। একজন ভোক্তা (consumer)-র দৃষ্টিভঙ্গী অনুযায়ী সূচকসংখ্যাটি এমন হওয়া দরকার যে, ভিত্তিকালে ভোক্তার ব্যবহৃত পণ্যগুলির দরের সঙ্গে চলতিকালে অনুরূপ দরের পার্থক্যের পরিমাপ এতে পাওয়া যায়। কিন্তু ব্যবহারের প্রশ্ন উঠলেই সেই পণ্যটি কী দরে এবং কী পরিমাণে ব্যবহার করা হয়েছে তার হিসাব নেওয়া দরকার হয়। এজন্য এরূপ ক্ষেত্রে সূচকসংখ্যা নির্ণয়ের জন্য চয়নিকৃত (selected) পণ্যগুলির দরের ভার হিসাবে এদের পরিমাণসমূহ (এখানে ধরা হচ্ছে যে পরিমাণে ভোক্তা পণ্যগুলি ব্যবহার করেছেন) নেওয়া হয়ে থাকে। ধরা যাক,

q_{0i} = ভিত্তিকালে i-তম পণ্যের ব্যবহারের পরিমাণ

q_{1i} = চলতিকালে i-তম পণ্যের ব্যবহারের পরিমাণ তা হলে ভারযুক্ত যৌগিক পদ্ধতি (weighted aggregative method) অনুসারে সূচক সংখ্যার মান হবে

$$I_{01} = \frac{\sum p_{1i} q_{1i}}{\sum p_{0i} q_{0i}} \qquad (৪.৫.২)$$

লক্ষ্যণীয় যে এখানে আমরা দর ও ব্যবহারের পরিমাণ চলতিকালে ও ভিত্তিকালে অনুরূপভাবে নিয়েছি। কিন্তু বাস্তব পরিস্থিতিতে অনেকসময় চলতিকালের ব্যবহারের পরিমাণ জানা সহজ হয় না। এরূপক্ষেত্রে ভিত্তিকালের পরিমাণটিই চলতিকালেও ব্যবহার করা হয়। অর্থাৎ সূচক সংখ্যাটির সূত্র নিম্নরূপ হয় :

$$I_{01} = \frac{\sum_{i} p_{1i} q_{0i}}{\sum p_{0i} q_{0i}} \qquad (৪.৫.৩)$$

উপরোক্ত সূচক সংখ্যা সূত্রটিকে **লাস্‌পেয়ারের সূত্র (Laspeyres' formula)** বলা হয়। বাস্তবক্ষেত্রে এই সূত্র অনুযায়ী সূচক নির্ণয় সবচেয়ে সুবিধাজনক। এজন্য এটি বহুল ব্যবহৃত।

অপরপক্ষে চলতিকালের পরিমাণকে ভার হিসাবে ব্যবহার করলে আমরা নিম্নলিখিত সূচকসংখ্যার সূত্র পাই।

$$I_{01} = \frac{\sum p_{1i} q_{1i}}{\sum_{i} p_{0i} q_{0i}} \qquad (৪.৫.৪)$$

এই সূচক সংখ্যা সূত্রটিকে **পাশের সূত্র** (Paasche's formula) বলে অভিহিত করা হয়।

বলাবাহুল্য (৪.৫.২)-এ উল্লিখিত সূত্রটি প্রকৃত পরিস্থিতির পরিচায়ক। দেখা গেছে যে, 'লাসপেয়ারের সূত্র' অনুযায়ী নির্ণীত সূচকের মান প্রকৃত সূচকের মান থেকে কিছুটা বেশি (with an upward bias) হয়। আবার পাশের সূত্র অনুযায়ী নির্ণীত সূচকের মান প্রকৃত সূচকের মান থেকে কিছুটা কম (with a downward bias) হয়।

লাসপেয়ারের সূত্রের ঊর্ধ্বমুখী ঝোঁক (upward bias) এবং পাশের সূত্রের নিম্নমুখী ঝোঁক (downward bias)-কে অপনীত করবার জন্য ভিত্তিকাল এবং চলতিকালের পরিমাণের গড়কে ভর (weight) হিসাবে নিয়ে নিম্নলিখিত সূত্রটি পাওয়া যায় :

$$I_{01} = \frac{\sum_{i} p_{1i} (q_{0i} + q_{1i})}{\sum p_{0i} (q_{0i} + q_{1i})} \qquad (৪.৫.৫)$$

এটি মারশাল-এজওয়ার্থ (**Marshall-Edgeworth**) সূত্র নামে পরিচিত। এছাড়া নিম্নলিখিত সূত্রটিও অনেক সময় ব্যবহার করা যেতে পারে।

$$I_{01} = \frac{\sum_i p_{1i}\, q_{0i}}{\sum_i p_{0i}\, q_{0i}} \times \frac{\sum_i p_{1i} \times q_{1i}}{\sum_i p_{0i} \times q_{1i}} \qquad (৪.৫.৬)$$

এটি ফিশারের আদর্শ সূত্র (**Fisher's ideal index formula**) নামে পরিচিত।

'মার্শাল এজওয়ার্থের সূত্র' এবং 'ফিশারের আদর্শ সূত্র'—এই দুটির সাহায্যেই ঊর্ধ্বমুখী এবং নিম্নমুখী ঝোঁক কমানো যায়। কিন্তু বাস্তবক্ষেত্রে এগুলি নির্ণয় করা যথেষ্ট অসুবিধাজনক। এজন্য এইসব সূচক সংখ্যার ব্যবহার বাস্তবক্ষেত্রে খুবই সীমিত।

গড়-দর অনুপাত পদ্ধতি (Average of price relatives) :

সরল গড়-আপেক্ষিক দর পদ্ধতি (Simple average of price relatives) :

আপেক্ষিক দরের সাধারণ যৌগিক গড় (simple arithmatic mean) নিলে আমরা পাই

$$I_{01} = \frac{1}{k} \sum_{i=1}^{k} \frac{p_{1i}}{p_{0i}} \qquad (৪.৫.৭)$$

যেখানে k-টি পণ্যের সমষ্টি নেওয়া হয়েছে। গুণোত্তর গড় নিলে আমরা পাই

$$I_{01} = \left\{ \prod_{i=1}^{k} \left(\frac{p_{1i}}{p_{0i}} \right) \right\}^{1/k} \qquad (৪.৫.৮)$$

অনুরূপভাবে, দর-অনুপাতের বিবর্ত যৌগিক গড় (harmonic mean), ভূয়িষ্ঠক (mode) বা মধ্যমা (**median**) ব্যবহার করে সূচক সংখ্যা নির্ণয় করা যেতে পারে।

আবার (৪.২.৭)-কে নিম্নোক্তভাবেও প্রকাশ করা যেতে পারে :

$$I_{01} = \frac{\sum_{i=1}^{k} \frac{1}{p_{0i}} \times p_{1i}}{\sum_{i=1}^{k} \frac{1}{p_{0i}} \times p_{0i}} = \frac{\sum_{i=1}^{k} wi\, p_{1i}}{\sum_{i=1}^{k} wi\, p_{0i}} \qquad (৪.৫.৯)$$

যেখানে $wi = \frac{1}{p_{0i}} = i - wj$ পণ্যের দরের ভর।

ভারযুক্ত গড় দর-অনুপাত পদ্ধতি : এই পদ্ধতিতে দর-অনুপাত $\frac{p_{1i}}{p_{0i}}$-এর সাথে একটি ভার wi-কে যুক্ত করা হয়, যাতে আমরা পাই —

$$I_{0i} = \frac{\sum_i wi \times \frac{p_{1i}}{p_{0i}}}{\sum_i wi} \qquad (৪.৫.১০)$$

অনুরূপভাবে, গুণোত্তর গড়ের ব্যবহার করে

$$I_{0i} = \left\{ \prod_i \left(\frac{p_{1i}}{p_{0i}}\right)^{wi} \right\}^{1/\Sigma w_i} \qquad (৪.৫.১১)$$

এবং বিবর্ত যৌগিক গড় ব্যবহার করে

$$I_{0i} = \frac{\sum_i wi}{\sum_i \frac{p_{0i}}{p_{1i}} wi} \qquad (৪.৫.১২)$$

আমরা যদি (৪.২.১০)-এ $wi = \sum_i p_{0i} q_{0i}$ -কে ভার হিসাবে নেই, তা হলে

$$I_{0i} = \frac{\sum wi \times \frac{p_{1i}}{p_{0i}}}{\sum wi} = \frac{\sum p_{0i} q_{0i} \times \frac{p_{1i}}{p_{0i}}}{\sum p_{0i} q_{0i}}$$

$$= \frac{\sum p_{1i} q_{0i}}{\sum p_{0i} q_{0i}} \qquad (৪.৫.১৩)$$

এটি (৪.৫.৩)-এ বর্ণিত লাস্পেয়ারের সূত্র। অনুরূপভাবে, $wi = p_{1i} q_{1i}$ নিয়ে ভারযুক্ত বিবর্ত যৌগিক গড় (weighted harmonic mean) ব্যবহার করে পাশের সূত্র পাওয়া যেতে পারে।

বাস্তব পরিস্থিতির বিচারে লাস্পেয়ারের সূত্র ব্যবহার করা সবচেয়ে সুবিধাজনক। এইজন্য অধিকাংশ ব্যবহারিক ক্ষেত্রেই এই সূত্রটির প্রয়োগ করা হয়ে থাকে।

৩৬.৯ সূচক সংখ্যার সামঞ্জস্য বিচার (Tests of consistency of index number)

দর-সূচকের সামঞ্জস্য বিচারের কয়েকটি প্রণালী আছে। এদের মধ্যে আরভিং ফিশার (Irving Fisher) কর্তৃক উদ্ভাবিত নিম্নলিখিত বিচার দুটি বিশেষভাবে উল্লেখযোগ্য :

(ক) কালবিবর্তনী বিচার (time reversal test)

(খ) উপাদানবিবর্তনী বিচার (factor reversal tect)

(ক) **কালবিবর্তনী বিচার (time reversal test) :** যেসব সূচক সংখ্যা এই বিচার পদ্ধতি মেনে চলে তাদের সূত্রগুলি সময়ের সাথে সামঞ্জস্যমূলক হয় অর্থাৎ এরূপক্ষেত্রে ভিত্তিকাল এবং চলতিকাল পরস্পর পরিবর্তনযোগ্য হয় এবং সময়ের গতিবিধির সাথে মূল্যমানের হ্রাস বা বৃদ্ধির চিত্রটি অপরিবর্তিত থাকে।

কোনো একটি বিশেষ পণ্যের দর-অনুপাতের ক্ষেত্রে এই নিয়মটি সব সময়ই অনুসৃত হয়। উদাহরণস্বরূপ 1981 সালকে ভিত্তিকাল ধরলে 1991 সালে আলুর আপেক্ষিক দর যদি দ্বিগুণ হয়, তা হলে 1991 সালের তুলনায় 1981 সালে আলুর আপেক্ষিক দর অর্ধেক হবে। সূত্রে প্রকাশ করলে ধরা যাক, p_0 = 1981 সালের আলুর দর এবং p_1 = 1991 সালের আলুর দর $= zp_0$, তা হলে $r_{01} = \frac{p_1}{p_0} = z$ এবং $r_{10} = \frac{p_0}{p_1} = \frac{1}{z}$ । সুতরাং $r_{01} \times r_{10} = 1$.

একটি বিশেষ পণ্যের ক্ষেত্রে প্রযোজ্য আপেক্ষিক দরের নিয়মটি যদি কোনো সূচকসংখ্যার ক্ষেত্রেও খাটে, তা হলে বলা হবে যে ঐ সূচক সংখ্যাটি বিবর্তনী বিচারপদ্ধতি অনুসরণ করছে। প্রতীকে প্রকাশ করলে, এরূপ ক্ষেত্রে সূচক সংখ্যাটি নিম্নলিখিত সম্পর্কটি অনুসরণ করবে :

$$I_{01} \times I_{10} = 1 \qquad (৪.৫.১৪)$$

পূর্বে উল্লিখিত সূচক সংখ্যাগুলির ভেতর (৪.৫.২), (৪.২.৫), (৪.৫.৬) এবং (৪.৫.৮)-র সূত্রগুলি এই বিচারপদ্ধতি মেনে চলে। যদি wi একটি ধ্রুবক হয়, তা হলে (৪.৫.১১) উল্লিখিত সূত্রটিও এই পদ্ধতি মেনে চলবে। তাছাড়া আপেক্ষিক দর সমূহের মধ্যমা (median) এবং ভূয়িষ্ঠক (mode)- দুটিও এই পদ্ধতি অনুসরণ করে।

(খ) **উপাদান বিবর্তনী বিচার (factor reversal test) :** কোনো পণ্যের দর যদি p হয় এবং পরিমাণ যদি q হয়, তা হলে পণ্যটির মূল্য হবে pq। দুটি বিভিন্ন সময়ে পণ্যের মূল্যের সূচক (value index) হবে

$$I_v = \frac{\sum_i p_{1i} q_{1i}}{\sum_i p_{0i} q_{0i}} \qquad (৪.৫.১৫)$$

(যেখানে v দ্বারা value বা মূল্য সূচিত হচ্ছে)।

অর্থাৎ প্রতি পণ্যের জন্য দর-অনুপাত $\left(\frac{p_{1i}}{p_{0i}}\right)$ এবং পরিমাণ অনুপাত $\left(\frac{q_{1i}}{q_{0i}}\right)$-এর গুণফল হল মূল্য অনুপাত।

$$\frac{p_{1i}}{p_{0i}} \times \frac{q_{1i}}{q_{0i}} = \frac{p_{1i}q_{1i}}{p_{0i}q_{0i}} \qquad (৪.৫.১৬)$$

ফিশারের যুক্তি অনুযায়ী এই নিয়মটি একটি সুবিধাজনক নিয়ম (যে কোনো দর-সূচকের ক্ষেত্রে)। যদি দর এবং পরিমাণ—এই দুটি উপাদানকে পাল্টাপাল্টি করা হয় তবে আমরা সংশ্লিষ্ট পরিমাণসূচক পাব। ফিশারের যুক্তিতে এ-দুটির গুণফল নিলে মূল্যসূচক পাওয়া সমীচীন। এই বিচারটির নাম দেওয়া হয়েছে **উপাদান বিবর্তনী বিচার** (factor reversal test)। প্রতীকে প্রকাশ করলে দাঁড়াবে—

$$I_p \times I_q = I_0 \qquad (৪.৫.১৭)$$

ফিশারের আদর্শসূচক (৪.৫.৬ দ্রষ্টব্য) এই নিয়মটি মেনে চলে। কারণ, এই সূচকের ক্ষেত্রে :

$$I_p = \frac{\sum_i p_{1i}\, q_{0i}}{\sum_i p_{0i}\, q_{0i}} \times \frac{\sum_i p_{1i} \times q_{1i}}{\sum_i p_{0i} \times q_{1i}}$$

$$I_q = \frac{\sum_i q_{1i}\, p_{0i}}{\sum_i q_{0i}\, p_{0i}} \times \frac{\sum_i q_{1i} \times p_{1i}}{\sum_i q_{0i} \times p_{1i}}$$

$$\text{ফলে, } I_p \times I_q = \frac{\sum_i p_{1i}\, q_{0i}}{\sum_i p_{0i}\, q_{0i}} = I_V \qquad (৪.৫.১৮)$$

৩৬.৯.১ দুটি গুরুত্বপূর্ণ সূচক সংখ্যা

(ক) **পশ্চিমবঙ্গের—ভোক্তাদের দর সূচক (Consumer price index বা CPI) :**

ভিত্তিকাল এবং চলতিকালের জীবনযাত্রার মান যদি অপরিবর্তিত রাখা হয়, তাহলে ভিত্তিকালের জীবিকা নির্বাহন ব্যয়ের তুলনায় চলতিকালে কী পরিমাণ (বেশি বা কম) অর্থব্যয় করতে হবে তা জীবিকা

নির্বাহন ব্যয়ের সূচক সংখ্যায় শতকরা হিসাবে প্রকাশ করা হয়। বিভিন্ন শ্রেণীর জনসাধারণ—যেমন উচ্চবিত্ত, মধ্যবিত্ত, নিম্নবিত্ত, শ্রমিক, কৃষক ইত্যাদি—তাঁরা জীবনযাত্রা আর্থিক আয়-ব্যয়ের দিক থেকে ভিন্ন ভিন্নভাবে যাপন করেন। এজন্য এ রকম বিভিন্ন শ্রেণীর জনসাধারণের জন্য পৃথক পৃথক জীবিকা নির্বাহন ব্যয়ের সূচক নির্ণয়ের প্রথা আছে। যেমন, ভারত সরকারের কেন্দ্রীয় শ্রম ব্যুরো (Central Labour Bureau) শ্রমিক শ্রেণীর জন্য জীবিকা নির্বাহন ব্যয়ের সূচক মাসিক ভিত্তিতে প্রকাশ করে থাকে। আবার পশ্চিমবঙ্গ সরকারও রাজ্যের বিভিন্ন আয়ের লোকের জন্য আলাদা আলাদা জীবিকা নির্বাহণ ব্যয়ের সূচক প্রকাশ করে থাকেন—কলকাতা ও রাজ্যের কয়েকটি বড় শহরের জন্য।

এই সূচক সংখ্যা নির্ণয়ের জন্য জীবিকা নির্বাহের জন্য প্রয়োজনীয় ভোগ্যপণ্যগুলিকে প্রথমে কয়েকটি প্রধান গোষ্ঠীতে (major groups-এ) ভাগ করা হয়ে থাকে। সাধারণত নিম্নলিখিত পাঁচটি প্রধান গোষ্ঠী নেওয়া হয়ে থাকে : (1) খাদ্য (food), (2) পরিধেয় (clothing), (3) আলো ও জ্বালানি (fuel and light), (4) বাসস্থান (housing) এবং (5) বিবিধ (miscellaneous)। প্রতিটি প্রধান গোষ্ঠীর জন্য এ গোষ্ঠীর প্রতিনিধিমূলক (representative) কয়েকটি ভোগ্যপণ্যের নমুনা নেওয়া হয়ে থাকে। যেমন খাদ্যের অন্তর্ভুক্ত করা হয় দানাশস্য (চাল, গম ইত্যাদি), কয়েক ধরনের সব্জী, মাছ, মাংস, ফল ইত্যাদি। এ রকম সূচকসংখ্যা নির্ণয় করার উদ্দেশ্যে প্রতিনিধিস্থানীয় ভোগ্যপণ্যগুলির দর-অনুপাতের ভারযুক্ত গড় (weighted average) নেওয়া হয়। কোনো পণ্যের ভার ভোক্তাদের কাছে ঐ পণ্যের গুরুত্ব অনুযায়ী স্থির করা হয়। এই গুরুত্ব নির্ণয় করার জন্য পারিবারিক আয়-ব্যয়কে সমীক্ষা (family budget enquiry) ক'রে বিভিন্ন ধরনের পণ্যাদির জন্য শতকরা হিসাবে কত শতাংশ খরচ করা হয় তার হিসাব সংগ্রহ করা হয়। এই শতাংশগুলির অনুপাতে ভারসমূহ নির্ণয় করা হয়ে থাকে। নিচে একটি ছোট উদাহরণের সাহায্যে বাস্তবে কিভাবে সূচক নির্ণয় করা হয় তা দেখানো হল।

উদাহরণ ৪.২

পশ্চিমবঙ্গ সরকারের ফলিত অর্থনীতি এবং পরিসংখ্যান ব্যুরো কর্তৃক সংগৃহীত রাশিতথ্যের ভিত্তিতে নভেম্বর 1950 সালকে ভিত্তিকাল ধরে 1961 সালের জীবিকা নির্বাহণ ব্যয়ের সূচকসংখ্যা নির্ণয়ের জন্য নিম্নলিখিত মানগুলি পাওয়া গেছে :

পণ্যের নাম	ভার (W_i)	দর-অনুপাত ($100 \times \frac{p_0}{p_1}$)
1. পুরুষের পরিধেয়	44.19	132
2. মহিলাদের পরিধেয়	39.06	126
3. শিশুর পরিধেয়	9.27	135
4. অন্যান্য পরিধেয়	7.48	130

ধরা যাক, পরিধেয় দ্রব্যের সূচক সংখ্যা বোঝাতে I_c প্রতীক ব্যবহার করা হবে। তা হলে এখানে

$$I_c = \frac{100 \times \sum_i \frac{p_{1i}}{p_{0i}} wi}{\sum_i wi} = \frac{12978}{100} = 129.8 \qquad (৪.৫.১৯)$$

সাধারণত ভার নির্ণয়ের জন্য যে বছরে পারিবারিক আয়-ব্যয়ক সমীক্ষা করা হয় সেই বছরকেই ভিত্তিকাল ধরা হয়। অর্থাৎ, যদি i-তম পণ্যের জন্য

p_{0i} = ভিত্তিকালের দর

q_{0i} = ভিত্তিকালে পণ্যের ব্যবহারের পরিমাণও বোঝায়, তা হলে

$p_{0i}\, q_{0i}$ = ভিত্তিকালে ব্যবহৃত i-তম পণ্যের মোট মূল্য।

সুতরাং $\sum_i p_{0i}\, q_{0i}$ = ভিত্তিকালে ব্যবহৃত সব পণ্যের মোট মূল্য।

এবং $w_i = \frac{p_{0i}\, q_{0i}}{\sum p_{0i}\, q_{0i}}$ হবে i-তম পণ্যের ভার।

সুতরাং, $$I_c = \frac{100 \times \sum_i \frac{p_{1i}}{p_{0i}} wi}{\sum_i wi}$$

অর্থাৎ, $$I_c = 100 \times \sum_i \frac{p_{1i}}{p_{0i}} \times \left(\frac{p_{0i}\, q_{0i}}{\sum p_{0i}\, q_{0i}}\right)$$

$$= 100 \times \frac{\sum_i p_{1i}\, q_0}{\sum_i p_{0i}\, q_{0i}} \qquad (৪.৫.২০)$$

অর্থাৎ I_c লাস্‌পয়ারের সূত্র অনুসরণ করে।

উদাহরণ ঃ নিচের সারণিতে পাঁচটি পণ্যের ভিত্তিবর্ষের (1980-র) দর এবং মোট ব্যবহারের পরিমাণ দেওয়া আছে এবং একই সঙ্গে 1990-এর দর এবং মোট ব্যবহারের পরিমাণ দেওয়া আছে (p দিয়ে টাকায় কিলোগ্রাম প্রতি দর এবং q দিয়ে মেট্রিক টনে মোট পরিমাণ সূচিত করা হয়েছে) :

পণ্যের ক্রমিক সংখ্যা	1980		1985	
	p	q	p	q
1	18.50	6.2	23.80	8.3
2	18.00	54.5	24.50	53.5
3	17.60	113.0	25.75	125.2
4	18.10	18.7	30.50	20.8
5	45.20	167.7	55.60	172.3

আমরা 1980-কে ভিত্তিবর্ষ ধরে 1990-এর জন্য দরসূচক নির্ণয় করতে চাই। এখানে

$$\Sigma p_0 q_0 = 11003.01 \qquad \Sigma p_0 q_1 = 11484.51$$

$$\Sigma p_1 q_0 = 14287.03 \qquad \Sigma p_1 q_1 = 14946.47$$

তাই লাসপেয়ারের সূচক $= 100 \times \frac{14287.03}{11003.01} = 129.85$

পাশে সূচক $= 100 \times \frac{14946.47}{11484.51} = 130.14$

এবং ফিশারের আদর্শ সূচক $= \sqrt{129.85 \times 130.14}$

$= 129.99$

(খ) **সর্বভারতীয় পাইকারি দরের সূচক (Index number of wholesale prices in India) :**

এই দর সূচক ভারত সরকারের অর্থনৈতিক উপদেষ্টা (Economic Adviser)-এর দপ্তর সাপ্তাহিক ভিত্তিতে প্রকাশ করে থাকেন। 1952-53-কে ভিত্তিকাল ধরে এই সূচকটি 1953 সাল থেকে প্রকাশিত হয়ে আসছিল। পাঁচটি প্রধান গোষ্ঠী (major group)-এর অন্তর্গত মোট 112টি পণ্যের জন্য 550টি ক্ষেত্র থেকে সংগৃহীত পাইকারি দরের হিসাব (quotations of wholesale prices) নিয়ে এবং সংশ্লিষ্ট দর-অনুপাতগুলি নির্ণয় করে ঐ সমস্ত আপেক্ষিক দরের ভারযুক্ত গড়কে নির্ণয় সূচক হিসাবে প্রকাশ করা হতো। পরবর্তীকালে 1970-71 সালে এই সূচকটির ভিত্তিকাল পরিবর্তন করা হয়েছে এবং মোট পণ্যের সংখ্যা এবং তথ্যসংগ্রহের ক্ষেত্রসংখ্যাও বৃদ্ধি করা হয়েছে। সম্প্রতি নব্বই-এর দশকের একটি বছরকে ভিত্তিকাল ধরে এটি আরও

পরিবর্তিত ও পরিবর্ধিত করার চেষ্টা হচ্ছে। এখানে ব্যবহৃত ভারগুলি সংশ্লিষ্ট পণ্যের যে পরিমাণ বাজারজাত (marketed) হয়েছে তার মোট মূল্যের সমানুপাতিক হয়।

৩৬.১০ সারাংশ

এই পরিচ্ছেদে (ক) কালীন সারি ও (খ) সূচক সংখ্যার বিবরণ দেওয়া হয়েছে। কালীন সারির বিভিন্ন উপাংশ (component), যথা সুশাসিত গতিধারা (secular trend), ঋতুজ ভেদ (seasonal variation), চক্রিল ভেদ (cyclical variation) এবং অনিয়মিত গতিধারা (irregular variation)-এর বিবরণ দেওয়া হয়েছে। সুশাসিত গতিধারা পরিমাপের বিভিন্ন পদ্ধতি—যেমন আন্দাজভিত্তিক রেখা নিরূপণ পদ্ধতি, চলমান গড় পদ্ধতি, গাণিতিক রেখা-নিরূপণ পদ্ধতি ইত্যাদির ব্যাখ্যা করা হয়েছে। ঋতুজ ভেদের বিভিন্ন পদ্ধতিও বিবৃত করা হয়েছে।

সূচক সংখ্যার অর্থ ব্যাখ্যা করা হয়েছে। সূচকসংখ্যা নির্ণয়ের সমস্যাসমূহ ও বিভিন্ন পদ্ধতি বর্ণিত হয়েছে। সূচক সংখ্যার সামঞ্জস্য বিচারের প্রণালীগুলির বিবরণ দেওয়া হয়েছে। এছাড়া আমাদের দেশে বহুল ব্যবহৃত দুটি সূচক সংখ্যার বিবরণ দেওয়া হয়েছে। একটি পশ্চিমবঙ্গের ও অপরটি সারা ভারতের। প্রথমটি পশ্চিমবঙ্গ সরকারের ফলিত অর্থনীতি ও পরিসংখ্যান ব্যুরো দ্বারা প্রকাশিত ভোক্তাদের দর-সূচক (বা জীবিকা নির্বাহণ ব্যয়ের সূচক) ও দ্বিতীয়টি ভারত সরকারের অর্থনৈতিক উপদেষ্টার দপ্তর কর্তৃক প্রকাশিত সাপ্তাহিক সর্বভারতীয় পাইকারি দরের সূচক।

৩৬.১১ অনুশীলনী

১। কালীন সারি কাকে বলে? কালীন সারির বিভিন্ন উপাংশ (component)-এর বর্ণনা দিন।

২। সুশাসিত গতিধারা নির্ণয়ের বিভিন্ন পদ্ধতি বর্ণনা করুন।

৩। ঋতুজ ভেদ নির্ণয়ের বিভিন্ন পদ্ধতি বর্ণনা করুন ও গুণাগুণ ব্যাখ্যা করুন।

৪। সূচকসংখ্যা কাকে বলে? সূচক সংখ্যা নির্ণয়ের সময় কি কি সমস্যার সম্মুখীন হতে হয় তার বিস্তারিত বিবরণ দিন।

৫। সূচক সংখ্যার বিভিন্ন সূত্র বর্ণনা করুন।

৬। 'লাস্‌পেয়র' ও 'পাশের' সূত্র বিবৃত করুন। এই দুটি সূত্র ব্যবহারের তুলনামূলক সুবিধা-অসুবিধা বর্ণনা করুন।

৭। সামঞ্জস্য প্রমাণের জন্য সূচক সংখ্যাসমূহকে যে সমস্ত বিচারের সম্মুখীন হতে হয় তা বর্ণনা করুন। ফিশারের 'আদর্শ' সূচককে এদিক থেকে পরীক্ষা করুন।

৮। পশ্চিমবঙ্গ ও ভারতবর্ষের দুটি বহুল ব্যবহৃত সূচক সংখ্যার বিবরণ দিন।

৯। নিম্নের কালীন সারির জন্য ঋজুরৈখিক গতিধারা নির্ণয় করুন :

বর্ষ	ভারতীয় রেলপথের মূল আয় (কোটি টাকায়)
1964-65	666
1965-66	748
1966-67	777
1967-68	823
1968-69	903
1969-70	957
1970-71	1010

১০। বেশ কয়েক বছরের জন্য ভারতবর্ষে চায়ের উৎপাদন হার (হেক্টর-প্রতি কিলোগ্রামে) নিচে দেখানো হল :

বর্ষ	উৎপাদন হার	বর্ষ	উৎপাদন হার
1955-56	902	1964-65	1102
1956-57	984	1965-66	1072
1957-58	956	1966-67	1089
1958-59	984	1967-68	1107
1959-60	985	1968-69	1146
1960-61	971	1969-70	1114
1961-62	1070	1970-71	1191
1962-63	1043	1971-72	1213
1963-64	1037		

চলমান যৌগিক গড় ব্যবহার করে সুসংহত গতিধারা নির্ণয় করুন।

১১। নিচের সারণীতে ভারতবর্ষে 1972 সাল থেকে 1975 পর্যন্ত (হাজার টনে) প্রতি ত্রৈমাসিক সময়ের ইস্পাতের উৎপাদন দেখানো হয়েছে :

বর্ষ	ত্রিমাসকাল প্রথম	দ্বিতীয়	তৃতীয়	চতুর্থ
1972	1336	1065	1215	1335
1973	1463	1039	1183	1161
1974	1306	1041	1290	1321
1975	1525	1251	1456	1408

চলমান যৌগিক গড় পদ্ধতিতে ঋতুজ ভেদের সূচকগুলি নিরূপণ করুন।

১২। 'লাসপেয়ার্স' ও 'পাশের' সূচককে দর-অনুপাতের গড় হিসাবে প্রকাশ করুন।

১৩। নিচের সারণীতে দিল্লি শহরের ক্ষেত্রে 1982 ও 1985 সালে বারোটি পণ্যের পাইকারি দর ও বিক্রয়ের পরিমাণ দেওয়া আছে :

পণ্য	1982		1985	
	দর (কুইন্টাল প্রতি টাকায়)	পরিমাণ (হাজার মেট্রিক টন)	দর (কুইন্টাল প্রতি টাকায়)	পরিমাণ (হাজার মেট্রিক টন)
চাল	277.92	1.1	366.67	6.2
গম	151.00	4.2	182.57	5.5
জোয়ার	156.75	13.1	155.75	6.1
বাজরা	176.25	106.0	186.58	116.9
যব	121.83	2.4	181.25	1.0
ভুট্টা	273.00	1.0	498.83	0.6

(ক) ল্যাসপেয়ার্সের সূত্র (খ) প্যার্লে-র সূত্র এবং (গ) ফিশারের আদর্শ সূত্র অনুসারে 1982 সালকে ভিত্তিবর্ষ ধরে 1985-এর জন্য ঐ পণ্যসমষ্টির সূচকসংখ্যা নির্ণয় করুন।

১৪। কলকাতার 201-350 টাকা এই মাসিক ব্যয়স্তরের মধ্যবিত্ত শ্রেণীর মানুষদের জন্য খাদ্য, পরিধেয়, জ্বালানি ও আলো, বাড়ি ভাড়া ও বিবিধ-এ পাঁচটি পণ্যগোষ্ঠীর জন্য 1988 সালের ভোক্তা-দর সূচক দেওয়া আছে (ভিত্তিবর্ষ 1960 = 100) :

পণ্যগোষ্ঠী	গোষ্ঠীসূচক	গোষ্ঠীভার
খাদ্য	731.4	54.31
পরিধেয়	538.1	7.36
জ্বালানি ও আলো	1025.2	4.91
বাড়িভাড়া	178.7	10.50
বিবিধ	540.1	22.92

(ক) সাধারণ ভোক্তা-দর সূচকটি নির্ণয় করুন।

(খ) একজন মধ্যবিত্ত চাকুরিজীবী 1960 সালে মাসিক মাইনে পেতেন 400 টাকা এবং 1988 সালে পেতেন 2250 টাকা। তাঁকে যদি 1960 সালের জীবনমান সংরক্ষণ করতে হত, তা হলে অতিরিক্ত ভাতা হিসাবে তার পক্ষে 1988 সালে কত টাকা পাওয়া উচিত ছিল?

(গ) এক টাকার মূল্যমান 1960 সালের তুলনায় 1988 সালে কতটা হ্রাস পেয়েছিল?

একক ৩৭ □ দ্বিচল রাশিতথ্য বিশ্লেষণ ঃ সহগতি

গঠন

৩৭.০ উদ্দেশ্য

এই এককটি পাঠ করার পর আপনারা বুঝতে পারবেন—

- দ্বিচল রাশিতথ্য বিশ্লেষণের পদ্ধতিগুলি কী কী
- সহগতি মাপক ও আদর্শ সহগতি মাপকের ধর্ম
- সহগতি গুণাঙ্ক ও তার ধর্ম
- মানক্রম সহগতি গুণাঙ্ক সম্বন্ধে স্পিয়ারম্যান ও কেন্ডালের বক্তব্য

৩৭.১ প্রস্তাবনা

সমীক্ষার প্রয়োজনে অনেক সময় প্রতিটি ব্যষ্টির জন্য যুগপৎ দুটি লক্ষণের ওপর তথ্য সংগ্রহ করা হয়। যেমন পারিবারিক আয়-ব্যয়ক সংক্রান্ত সমীক্ষায় আমরা পরিবারের মাসিক আয় (চল x) এবং মাসিক আয়ের কত শতাংশ প্রসাধন সামগ্রীর জন্য ব্যয় করা হয় (চল ৬), সে সম্বন্ধে তথ্য সংগ্রহ করতে পারি। এক্ষেত্রে চল-

দু'টির মধ্যে একটিকে স্বতন্ত্র চল (independent variable) এবং অন্যটিকে নির্ভরী চল (dependent variable) হিসাবে গণ্য করা যেতে পারে। বর্তমান উদাহরণে x ও y যথাক্রমে স্বতন্ত্র এবং নির্ভরী চল।

লক্ষণ দু'টি বর্তমান উদাহরণের মত চল না হয়ে গুণলক্ষণও হতে পারে। যেমন, একটি কারখানায় উৎপন্ন প্রতিটি দ্রব্য সম্বন্ধে আমরা সেটির গুণমান (ত্রুটিযুক্ত অথবা ত্রুটিমুক্ত) এবং সেটি কোন্ শিফটে উৎপন্ন (সকাল, দুপুর অথবা রাত্রি), এ সংক্রান্ত তথ্য সংগ্রহ করতে পারি। আবার লক্ষণ দু'টির মধ্যে একটি গুণলক্ষণ এবং অন্যটি চল হতে পারে। কয়েকটি দেশের মানব-বিকাশ উন্নয়ন সংক্রান্ত অবস্থা (উচ্চ, মধ্য এবং নিম্ন মান) এবং ক্রয়ক্ষমতা সমতুল ডলারে (purchasing power parity \$) জাতীয় আয় সংক্রান্ত তথ্য এই তৃতীয় শ্রেণীর উদাহরণ। আমাদের বর্তমান আলোচনা সীমাবদ্ধ থাকবে দু'টি লক্ষণই যখন চল সেই ক্ষেত্রে।

দ্বিচল রাশিতথ্য বিশ্লেষণের বিভিন্ন পদ্ধতি নিয়ে আলোচনা করার আগে এ ধরনের তথ্য বিশ্লেষণের উদ্দেশ্য সম্বন্ধে আমাদের স্পষ্ট একটা ধারণা থাকা দরকার। আলাদা আলাদাভাবে চল দু'টির বিভিন্ন বৈশিষ্ট্যের ওপর আলোকপাত করা ছাড়াও এক্ষেত্রে আমাদের এক বা একাধিক উদ্দেশ্য থাকা সম্ভব। যেমন,

(ক) চল দু'টি পরস্পরের ওপর আদৌ নির্ভরশীল কিনা, তা পরীক্ষা করে দেখা।

(খ) যদি প্রাথমিক বিচারে দেখা যায় চল দু'টি পরস্পর নির্ভরশীল, তাহলে এই নির্ভরশীলতার প্রকৃতি এবং পরিমাণ নিরূপণ করা।

(গ) নির্ভরশীলতার সুস্পষ্ট প্রমাণ পাওয়া গেলে ভবিষ্যতে স্বতন্ত্র চলের নতুন কোন মানের জন্য নির্ভরী চলের প্রত্যাশিত মান নিরূপণ করা। এই উদ্দেশ্যে বিশ্লেষিত দ্বিচল রাশিতথ্য ব্যবহার করে নির্ভরী চলটিকে স্বতন্ত্র চলের একটি অপেক্ষকরূপে প্রকাশ করে একটি পূর্বানুমান সূত্র দাঁড় করানো যেতে পারে।

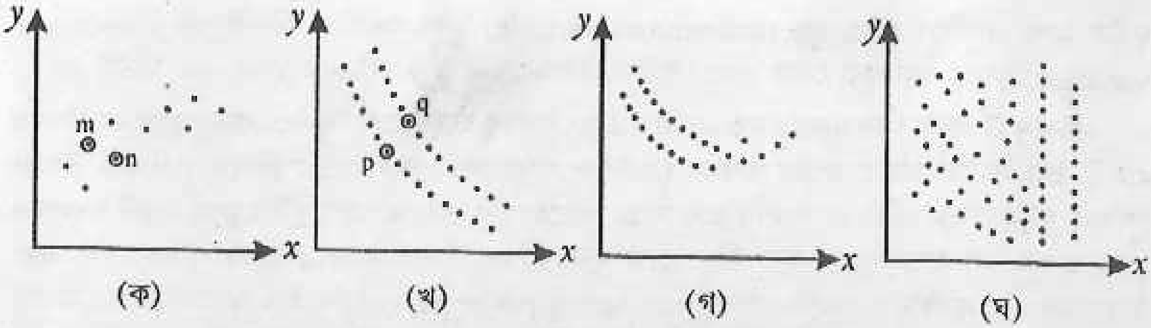
এইসব উদ্দেশ্য সামনে রেখে দ্বিচল রাশিতথ্য কীভাবে ধাপে ধাপে বিশ্লেষণ করা হবে, সে সম্বন্ধে বিস্তারিত আলোচনা করা হয়েছে পরবর্তী অনুচ্ছেদগুলিতে।

৩৭.২ বিক্ষেপণ চিত্র : সহগতি (Scatter Diagram : Correlation)

ধরা যাক, x এবং y এই দু'টি চলের ওপর n-সংখ্যক ব্যষ্টির জন্য তথ্য সংগৃহীত হয়েছে। তথ্যগুলি আমরা $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, বা সংক্ষেপে $(x_i, y_i), i = 1(1)n$ —এইভাবে নির্দেশ করে বিশ্লেষণের জন্য এগোতে পারি।

পরস্পর লম্বভাবে ছেদ করা দু'টি সরলরেখার একটিকে x-অক্ষ এবং অন্যটিকে y-অক্ষ ধরে অক্ষ দু'টিকে যথাক্রমে x-চল এবং y-চলের মাপনামাত্রা অনুযায়ী চিহ্নিত করা যাক। এই লেখচিত্রে সংগৃহীত n জোড়া মানকে nটি বিন্দুর সাহায্যে সংস্থাপন করে যে চিত্রটি পাওয়া যায় তাকে x ও y চলের বিক্ষেপন চিত্র (Scatter Diagram) বলা হয়। বিক্ষেপণ চিত্রে সংস্থাপিত বিন্দুগুলির বিন্যাস থেকে সংশ্লিষ্ট চল দু'টির মধ্যে কোনরকম সম্পর্ক বা পারস্পরিক নির্ভরশীলতা আছে কিনা, এবং থাকলে তা কী ধরনের ও কতখানি—এসব সম্বন্ধে একটা প্রাথমিক ধারণা পাওয়া যেতে পারে।

নিচের উদাহরণগুলি লক্ষ্য করা যাক।



চিত্র 1·1 বিক্ষেপণ চিত্রের বিভিন্ন রূপ

প্রথম তিনটি বিক্ষেপণ চিত্রে (ক, খ ও গ) বিন্দুগুলির বিন্যাস থেকে চল দু'টির মধ্যে একটি সুস্পষ্ট গাণিতিক সম্পর্কের আভাস পাওয়া যাচ্ছে। অর্থাৎ, দেখা যাচ্ছে, চল দু'টির গতিশীলতা পরস্পরের ওপর নির্ভরশীল। একটু লক্ষ্য করলে বোঝা যাবে, চল দু'টির মধ্যে ক ও খ চিত্রে একটি ঋজুরৈখিক সম্পর্ক এবং গ চিত্রে একটি দ্বিঘাতজ সম্পর্ক রয়েছে। কিন্তু ঘ চিত্রটিতে বিন্দুগুলির বিন্যাস পুরোপুরি এলোপাথাড়ি। এ থেকে চল দু'টির মধ্যে সুস্পষ্ট কোন গাণিতিক সম্পর্কের আভাস পাওয়া যাচ্ছে না। এ চিত্রটি থেকে সুতরাং বোঝা যাচ্ছে, চল দু'টির গতিশীলতার মধ্যে কোন সম্পর্ক নেই। অর্থাৎ, এ দু'টি পরস্পর অনধীন (independent)।

দু'টি চলের মধ্যে ঋজুরৈখিক সম্পর্ক থাকলে চল দু'টিকে সহগতিসম্পন্ন (correlated) বলা হয়। অর্থাৎ, সহগতি (correlation) হল, দু'টি চলের মধ্যে ঋজুরৈখিক সম্পর্ক। সহগতি ধনাত্মক এবং ঋণাত্মক দু'ধরনেরই হতে পারে।

যদি দেখা যায়, একটি চলের মান বাড়লে অন্যটির মধ্যে সাধারণভাবে বাড়ার, এবং একটির মান কমলে অন্যটিরও সাধারণভাবে কমার প্রবণতা রয়েছে (যেমন ক চিত্রে), তাহলে বলা হবে চল দু'টি ধনাত্মক সহগতি সম্পন্ন (positively correlated), অর্থাৎ এদের মধ্যে ধনাত্মক সহগতি রয়েছে। ধনাত্মক সহগতি সম্পন্ন চলের উদাহরণ হল পারিবারিক আয় ও বিলাসদ্রব্যের জন্য ব্যয়, দেশের জাতীয় আয় ও মানব-উন্নয়ন সূচক (Human Development Index), ছাত্রের গণিত ও রাশিবিজ্ঞানে প্রাপ্ত নম্বর, ইত্যাদি।

অন্যদিকে যদি দেখা যায় চল দু'টির মধ্যে একটির মান বাড়লে অন্যটির মান সাধারণভাবে কমার এবং একটির মান কমলে অন্যটির মান সাধারণভাবে বাড়ার প্রবণতা রয়েছে (চিত্র খ), তাহলে বলা হবে চল দু'টি ঋণাত্মক সহগতি সম্পন্ন (negatively correlated), অর্থাৎ, তাদের মধ্যে ঋণাত্মক সহগতি রয়েছে। ভোগ্যপণ্যের যোগান এবং বাজার দর, পরিবারের আয় ও খাদ্যদ্রব্যের জন্য ব্যয়িত আয়ের শতকরা অংশ—ইত্যাদি ঋণাত্মক সহগতি সম্পন্ন চলের উদাহরণ।

চল দু'টির যে কোন একটির মান বাড়লে অন্যটির মান যদি বাড়া বা কমার কোন নির্দিষ্ট প্রবণতা না থাকে, অর্থাৎ, x-এর মান বাড়লে y-এর মান বাড়া এবং কমার ঘটনা প্রায় সমান সংখ্যায় দেখা যায় (যেমন খ চিত্রে), তাহলে বলা হবে চল দু'টি সহগতিশূন্য (uncorrelated)। কৃষিজমির আয়তন এবং বিমা প্রতি ফলনের হার সহগতিশূন্য চলের উদাহরণ।

এখানে লক্ষ্যণীয়, দু'টি চল পরস্পর সম্পর্কশূন্য বা অনধীন হলে তারা সহগতিশূন্য হবে। তবে উল্টোটি

সত্য নয়, অর্থাৎ দু'টি চল সহগতিশূন্য হলে তারা পরস্পর অনধীন হতেও পারে, নাও হতে পারে—সুনির্দিষ্টভাবে কিছু বলা সম্ভব নয়। কেননা, সহগতিশূন্য কথাটির অর্থ সংশ্লিষ্ট চল দু'টির মধ্যে ঋজুরৈখিক সম্পর্ক নেই, তার অর্থ এই নয় যে, তাদের মধ্যে অন্য কোন ধরনের গাণিতিক সম্পর্ক (যথা, বৃত্তীয়, উপবৃত্তীয়, পরাবৃত্তীয় ইত্যাদি) থাকবে না। উদাহরণস্বরূপ বলা যায়, গ চিত্রটিতে স্পষ্টতই x ও y-এর মধ্যে একটি দ্বিঘাতজ সম্পর্ক রয়েছে। কিন্তু লক্ষ্য করলে দেখা যাবে, এখানে x-এর মান বাড়লে y কখনও বাড়ছে, কখনও কমছে।

সুতরাং দেখা যাচ্ছে, বিক্ষেপণ চিত্রে লব্ধ বিন্দুগুলির বিন্যাস থেকে সংশ্লিষ্ট চলদুটির সহগতি সম্বন্ধে একটা প্রাথমিক ধারণা পাওয়া যেতে পারে।

এখানে আর একটি কথা খেয়াল রাখা দরকার। দু'টি চলের মধ্যে সহগতির প্রকৃতি নির্ভর করে একটির মান বাড়ার সঙ্গে সঙ্গে অন্যটির মান পরিবর্তনের সাধারণ প্রবণতার ওপর। তার অর্থ এই নয় যে, ধনাত্মক সহগতি সম্পন্ন দু'টি চলের একটির মান বাড়লে অন্যটির মান সবসময় বাড়বে। বেশিরভাগ ক্ষেত্রে বাড়বে ঠিকই, তবে দু'চারটি ব্যতিক্রম থাকতেও পারে। যেমন, উপরের ক বিক্ষেপণ চিত্রে চল দু'টি ধনাত্মক সহগতি সম্পন্ন, কিন্তু এই চিত্রের M ও N বিন্দু দু'টি লক্ষ্য করলে দেখা যাবে, M-এর তুলনায় N-এ x-এর মান বেড়েছে বটে, কিন্তু y-এর মান কমেছে। তেমনি খ বিক্ষেপণ চিত্রটিতে চল দু'টি ঋণাত্মক সহগতি সম্পন্ন হলেও এখানে P বিন্দুর তুলনায় Q বিন্দুর x ও y উভয় চলের মানই বেড়েছে। সহগতির প্রকৃতি নির্ধারণে একটি চলের মান বাড়ার সঙ্গে অন্যটির মান বেশিরভাগ ক্ষেত্রে বাড়ছে না কমছে, অথবা গড়ে অপরিবর্তিত থাকছে, সেটিই হবে বিচার্য বিষয়।

৩৭.৩ সহগতি-মাপক (Measures of Correlation)

বিক্ষেপণ চিত্র থেকে যদি আভাস পাওয়া যায় দু'টি চলের মধ্যে আপাতগ্রাহ্য সহগতি রয়েছে, তাহলে বিশ্লেষণের পরবর্তী পর্যায়ে একটি উপযুক্ত মাপকের সাহায্যে এই সহগতির প্রকৃতি (ধনাত্মক, ঋণাত্মক, না শূন্য) এবং মাত্রা (কী পরিমাণে ধনাত্মক বা ঋণাত্মক) নির্ণয় করতে হবে।

৩৭.৩.১ আদর্শ সহগতি-মাপকের ধর্ম

একটি সহগতি-মাপকের নিম্নলিখিত ধর্মগুলি থাকলে সেটি একটি আদর্শ সহগতি-মাপক হিসাবে গণ্য হতে পারে :

(ক) একটি সংশ্লিষ্ট চল দু'টির মাপনা-একক নিরপেক্ষ হবে—নয়তো বিভিন্ন মাপনা-একক সম্পন্ন একাধিক স্বতন্ত্র চলের ওপর নির্ভরী চলের সহগতির মাত্রা পরস্পর তুলনীয় হবে না।

(খ) এটি n-এর, অর্থাৎ চল দু'টির যত জোড়া মানের ওপর ভিত্তি করে এর মান নির্ণয় করা হবে, তার ওপর নির্ভরশীল হবে না—অন্যথায়, n-এর মানে হ্রাসবৃদ্ধি ঘটিয়ে মাপকাঠির মান পরিবর্তন করা যাবে, যা আদৌ বাঞ্ছনীয় নয়।

(গ) এটি সংশ্লিষ্ট চল দু'টির মধ্যগামিতা (central tendency) এবং বিস্তৃতির (dispersion) ওপর নির্ভরশীল হবে না। কারণ হিসাবে উদাহরণ দিয়ে বলা যায়, গণিত ও রাশিবিজ্ঞানের নম্বরের মধ্যে সহগতি একটি পরীক্ষার ফল অনুসারে ভাল কলেজের ছাত্রদের প্রাপ্ত নম্বরের ভিত্তিতে যা হবে, একটি অপেক্ষাকৃত খারাপ কলেজের ছাত্রদের প্রাপ্ত নম্বরের ভিত্তিতেও মোটামুটি তাই হওয়া উচিত (ভালো কলেজটির ছাত্রদের নম্বরের গড় অন্য কলেজটির তুলনায় সাধারণত বেশি হবে, কিন্তু বিস্তৃতি হবে কম)।

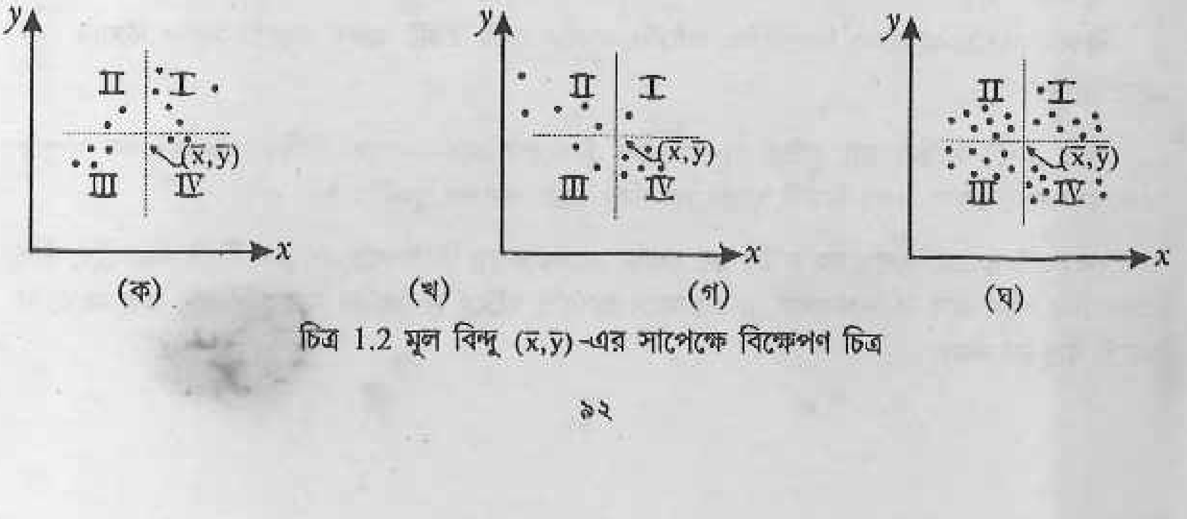
(ঘ) ধনাত্মক, ঋণাত্মক এবং শূন্য সহগতির ক্ষেত্রে মাপকাঠির মান যথাক্রমে ধনাত্মক, ঋণাত্মক এবং শূন্য হবে। এর ফলে মাপকাঠির লব্ধ মানের গাণিতিক চিহ্ন থেকেই সহগতির প্রকৃতি বোঝা যাবে সহজেই।

(ঙ) একটি সসীম প্রসারের মধ্যে এটির মান সীমাবদ্ধ থাকবে। এর ফলে লব্ধ মানটিকে সংশ্লিষ্ট সীমা মানের সঙ্গে তুলনা করে সহগতির প্রকৃত মাত্রা সম্বন্ধে একটা স্বচ্ছ ধারণা পাওয়া যাবে। যেমন মাপকাঠির মান সীমা যদি ∓ 1 হয়, লব্ধ মান 0.9 পাওয়া গেলে সহজেই বোঝা যাবে চল দু'টির মধ্যে ধনাত্মক সহগতির মাত্রা খুব বেশি। কিন্তু মানসীমা $\mp\infty$ হলে, লব্ধ মান 0.9 থেকে সহগতির মাত্রা সম্বন্ধে বিশেষ কিছু ধারণা করা যাবে না।

(চ) লব্ধ মানগুলির ক্রম অক্ষুণ্ণ রেখে যদি চল দু'টির কোন অভিন্ন রূপান্তর ঘটানো হয়, তাহলে মাপকাঠির মান অপরিবর্তিত থাকা বাঞ্ছনীয়। $u = \phi(x)$ রূপান্তরটি যদি এমন হয় যে $x_1 < x_2 \ldots < x_n$ হলে $u_1 < u_2 < \ldots < u_n$ হবে তাহলে $u = \phi(x)$ রূপান্তরটিকে x-এর একটি ক্রম-অক্ষুণ্ণ রূপান্তর (order-preserving transformation) বলে। x যদি শুধুমাত্র ধনাত্মক মান পরিগ্রহণ করে, তাহলে $u = x^2$ একটি ক্রম-অক্ষুণ্ণ রূপান্তর। কিন্তু x ধনাত্মক ও ঋণাত্মক দু'ধরনের মানই পরিগ্রহণ করলে রূপান্তরটি আর ক্রম-অক্ষুণ্ণ রূপান্তর হবে না। মাপকাঠির এই বৈশিষ্ট্যটি থাকলে যদি উভয় চলের একই ক্রম-অক্ষুণ্ণ রূপান্তর ঘটানো হয়, তাহলে মূল চল দু'টির জন্য মাপকাঠির মান এবং ক্রম-অক্ষুণ্ণ রূপান্তরিত চল দু'টির জন্য মাপকাঠির মান অভিন্ন হবে।

৩৭·৩·২ সহগতি গুণাঙ্ক (Correlation Coefficient)

1.1-এর বিক্ষেপণ চিত্রগুলিতে মূল বিন্দুটি (0, 0) থেকে $(\bar{x},\bar{y})$-এ সরিয়ে নেওয়া যাক (চিত্র 1.2)। এর ফলে দেখা যাচ্ছে, বিন্দুগুলির মোটামুটিভাবে ক চিত্রে প্রথম ও তৃতীয় পাদে, খ চিত্রে দ্বিতীয় ও চতুর্থ পাদে এবং ঘ চিত্রে চারটি পাদেই সমানভাবে ছড়িয়ে থাকার প্রবণতা রয়েছে।



(ক) (খ) (গ) (ঘ)

চিত্র 1.2 মূল বিন্দু $(\bar{x},\bar{y})$-এর সাপেক্ষে বিক্ষেপণ চিত্র

এখন $x_i' = x_i - \bar{x}$ এবং $y_i' = y_i - \bar{y}$ লেখা হলে $x_i' y_i'$-এর মান প্রথম ও তৃতীয় পাদে ধনাত্মক (যেহেতু x_i' ও y_i' দু'টিই ধনাত্মক বা দু'টিই ঋণাত্মক) এবং দ্বিতীয় ও চতুর্থ পাদে ঋণাত্মক যেহেতু x_i' ও y_i' -এর মধ্যে একটি ধনাত্মক ও অন্যটি ঋণাত্মক। সুতরাং, $\sum_{i=1}^{n} x_i' y_i'$ যোগফলটিতে ধনাত্মক সহগতির ক্ষেত্রে ধনাত্মক গুণফলগুলির সমষ্টি ঋণাত্মক গুণফলগুলির সমষ্টির তুলনায় অনেক বেশি হবে এবং সমগ্র যোগফলটির মান হবে ধনাত্মক। একইভাবে এই যোগফলটির মান ঋণাত্মক সহগতির ক্ষেত্রে ঋণাত্মক এবং সহগতিশূন্যতার ক্ষেত্রে শূন্য বা তার কাছাকাছি হবে। সুতরাং দেখা যাচ্ছে, $\sum_{i=1}^{n} x_i' y_i' = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$ এই যোগফলটির রয়েছে আদর্শ সহগতি মাপকের দু'টি গুরুত্বপূর্ণ ধর্ম। এটির গাণিতিক চিহ্ন সহগতির প্রকৃতির সঙ্গে সমঞ্জস এবং এটি চল দু'টির মধ্যগামিতা নিরপেক্ষ। এটিকে মান সংখ্যা এবং চল দু'টির মাপনা-একক ও বিস্তৃতি নিরপেক্ষ করার জন্য একে n এবং চলদুটির প্রমাণ বিচ্যুতি (standard deviation) যথাক্রমে S_x ও S_y দিয়ে ভাগ করে পাওয়া যায়।

$$r \equiv r_{xy} = \frac{\sum_{i=1}^{n} x_i' y_i'}{n S_x S_y} = \frac{\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2} \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y})^2}} \quad \ldots (1.1)$$

লবের $\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = \mathrm{Cov}(x, y)$ রাশিটিকে বলা হয় x ও y-এর সহভেদমান (covariance)। এটি ভেদমানের (variance) সঙ্গে তুলনীয় একটি প্রকাশন, যা চল দু'টির যুগবৎ চলমানতার একটি মাপক হিসাবে ব্যবহৃত হয়। তাহলে দেখা যাচ্ছে,

$$r_{xy} = \frac{\mathrm{Cov}(x, y)}{\sqrt{\mathrm{Var}(x)} \sqrt{\mathrm{Var}(y)}} \quad \ldots (1.2)$$

দ্বিচল রাশিতথ্যের জন্য (r, s)-তম কেন্দ্রীয় পরিঘাত $m_{r,s}$ দ্বারা নির্দেশ করা হলে

$$m_{r,s} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^r (y_i - \bar{y})^s \quad \ldots (1.3)$$

এবং সেক্ষেত্রে $r_{x,y} = \frac{m_{1,1}}{\sqrt{m_{2,0}}\sqrt{m_{0,2}}}$ ... (1.4)

এই কারণে r_{xy} কে অনেক সময় পূরণী পরিঘাত সহগতি গুণাঙ্ক (product moment correlation co-efficient) বলা হয়।

প্রদত্ত রাশিতথ্য থেকে r_{xy}-এর মান নির্ণয় করার জন্য এর নিচের রূপটি ব্যবহার করা সুবিধাজনক ঃ

$$r_{xy} = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}\,\bar{y}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n} y_i^2 - \bar{y}^2}}$$

$$= \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{\sqrt{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}\sqrt{n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2}} \quad \text{... (1.5)}$$

আগের আলোচনা থেকে স্পষ্ট, 1.3.1-এ বর্ণিত আদর্শ সহগতি মাপকের প্রথম চারটি ধর্মই রয়েছে r_{xy}-এর। আমরা পরে প্রমাণ করব, এর পঞ্চম ধর্মটিও রয়েছে তবে ক্রম-অক্ষুণ্ণ রূপান্তরের ক্ষেত্রে সহগতি গুণাঙ্ক অপরিবর্তনশীল নয়, অর্থাৎ আদর্শ সহগতি-মাপকের ষষ্ঠ ধর্মটি এর নেই।

মন্তব্য : দু'টি চলের মধ্যে পারস্পরিক সম্পর্কের বিচারে সহগতি গুণাঙ্কের একটা বড় সীমাবদ্ধতা হল, এই মাপকটির মান শূন্য বা শূন্যের কাছাকাছি হ'লে বড়জোর বলা যায় চল দু'টির মধ্যে ঋজুরৈখিক সম্পর্ক নেই। কিন্তু অন্য কোন ধরনের সম্পর্ক আছে কিনা, অর্থাৎ চল দু'টি পরস্পর অনধীন কিনা, সে সম্বন্ধে নিশ্চয় করে কিছু বলা সম্ভব নয়। তবে দু'টি অনধীন চলের ক্ষেত্রে সহগতি গুণাঙ্কের মান শূন্য অথবা শূন্যের কাছাকাছি হবেই।

এই প্রসঙ্গে উদাহরণ ১.২ লক্ষ্য কর।

উদা ১.১ ঃ কয়েকটি উন্নয়নশীল দেশের 1955 সালের জন্য মাথাপিছু জাতীয় উৎপাদন (আমেরিকান ডলারে) এবং মানব-বিকাশ সূচক (Human Development Index) সংক্রান্ত তথ্য নিচে দেওয়া হল। চল দু'টির মধ্যে সহগতি গুণাঙ্ক নির্ণয় কর।

দেশের ক্রমিক সংখ্যা	দেশ	মাথাপিছু জাতীয় উৎপাদন (আমেরিকান ডলারে)	মানববিকাশ সূচক
1	কেনিয়া	375	0.463
2	পাকিস্তান	381	0.453
3	ভারত	425	0.451
4	ক্যাম্বোডিয়া	133	0.422
5	নাইজিরিয়া	355	0.391
6	বাংলাদেশ	202	0.371
7	নেপাল	206	0.351
8	ইথিওপিয়া	154	0.252
9	মাদাগাস্কার	199	0.348

উৎস : Human Development Report, 1998 (UNDP)

[কোন দেশের মানব-বিকাশ সূচক হল ঐ দেশের জনসাধারণের জীবনযাত্রার সামগ্রিক মানের একটি মাপক। এটি নির্ভর করে প্রত্যাশিত আয়ু, সাক্ষরতার হার, শিক্ষার প্রসার, জাতীয় উৎপাদনের পরিমাণ এই সবের ওপর। এই সূচকের সর্বোচ্চ মান 1 হতে পারে। নোবেলজয়ী অর্থনীতিবিদ অধ্যাপক অমর্ত্য সেনের এই ধরনের মাপক উদ্ভাবনে অনেক গবেষণা কাজ রয়েছে।]

এখানে মাথাপিছু জাতীয় উৎপাদন সূচক—এই চলটিকে x এবং মানব-বিকাশ সূচক—এই চলটিকে y ধরে নিয়ে সহগতি গুণাঙ্ক নির্ণয় করার জন্য নিম্নলিখিত ছকে অঙ্কপাতন করা যেতে পারে।

সারণি ৩৭.১ : মাথাপিছু জাতীয় উৎপাদন (x) ও মানব-বিকাশ সূচক (y)-এর সহগতি গুণাঙ্ক নির্ণয়।

দেশ i	x_i	y_i	x_i^2	y_i^2	$x_i y_i$
(1)	(2)	(3)	(4)	(5)	(6)
1	375	0.463	140625	.2144	173.625
2	381	0.453	145161	.2052	172.593
3	425	0.451	180625	.2034	191.675
4	133	0.422	17689	.1781	56.126
5	355	0.391	126025	.1529	138.805
6	202	0.371	40804	.1376	74.942
7	206	0.351	42436	.1232	72.306
8	154	0.252	23716	.0635	38.806
9	199	0.348	39601	.1211	69.252
মোট	2430	3.502	756682	1.3994	988.132

সুতরাং এক্ষেত্রে

$$r_{xy} = \frac{9\times988.132-2430\times3.502}{\sqrt{9\times756682-2430^2}\sqrt{9\times1.3994-3.502^2}}$$

$$= \frac{383.328}{548.2969} = 0.6991$$

দেখা যাচ্ছে, আলোচ্য চল দু'টির মধ্যে মোটামুটি ভালরকম ধনাত্মক সহগতি রয়েছে।

উদাহরণ ১.২ : 8 খণ্ড আলু জমিতে একর প্রতি প্রদত্ত সারের পরিমাণ ও একর প্রতি ফলন সংক্রান্ত নিচের তথ্য থেকে চল দু'টির মধ্যে সহগতি গুণাঙ্ক নির্ণয় কর। বিন্দুগুলি বিক্ষেপণ চিত্রে সংস্থাপন করে চল দু'টির সম্পর্ক ও সহগতি গুণাঙ্কের মান সম্বন্ধে মন্তব্য কর।

জমির ক্রমিক সংখ্যা (i)	1	2	3	4	5	6	7	8
একর প্রতি সারের পরিমাণ (কেজি) (x_i)	0	50	100	150	200	250	300	350
একর প্রতি আলুর ফলন (কুইন্ট্যাল) (y_i)	1.8	2.3	3.5	4.2	3.6	3.2	2.5	2.0

এখানে $n = 8, \sum_{i=1}^{8} x_i = 1400, \sum_{i=1}^{8} y_i = 23.1$

$$\sum_{i=1}^{8} x_i^2 = 350000, \sum_{i=1}^{8} y_i^2 = 71.87$$ এবং

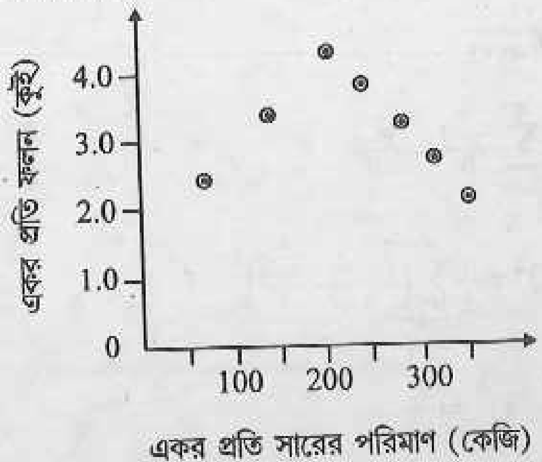
$$\sum_{i=1}^{8} x_i y_i = 4065\ ;$$

সুতরাং, সহগতি গুণাঙ্কের মান

$$r_{xy}=\frac{8\times4065-1400\times23.1}{\sqrt{8\times350000-1400^2}\sqrt{8\times75.88-23.8^2}}$$

$$=0.0306।$$

এবার বিক্ষেপণ চিত্রটি আঁকা যাক।



চিত্র ৩৭.২ আলুজমিতে একর প্রতি সারের পরিমাণ ও একর প্রতি ফলনের বিক্ষেপণ চিত্র।

দেখা যাচ্ছে, এখানে সহগতি গুণাঙ্কের মান প্রায় শূন্যের কাছাকাছি। সুতরাং, চল দু'টি সহগতি শূন্য। কিন্তু বিক্ষেপণ চিত্র থেকে বোঝা যাচ্ছে, চল দু'টির মধ্যে একটি সুস্পষ্ট দ্বিঘাতজ সম্পর্ক রয়েছে।

৩৭.৩.৩ সহগতি গুণাঙ্কের ধর্ম

সহগতি গুণাঙ্ক r-এর নিম্নলিখিত ধর্মগুলি রয়েছে।

(১) r একটি শুদ্ধ সংখ্যা, অর্থাৎ এটি সংশ্লিষ্ট চল দু'টির মাপনা-এককের ওপর নির্ভরশীল নয়। r-এর লব ও হরের মিশ্র মাপনা-একক অভিন্ন। ফলে অনুপাতটি শুদ্ধ সংখ্যা। যেমন, ওজন (কিগ্রায়) ও উচ্চতার (মিটারে) মধ্যে সহগতি গুণাঙ্ক নির্ণয়ের সময় লবের সহভেদ মানটির মাপনা-একক কিগ্রা-মিটার, এবং হরের প্রমাণ বিচ্যুতি দু'টির মাপনা-একক যথাক্রমে কিগ্রা ও মিটার। সুতরাং, গুণাঙ্কটি একটি শুদ্ধ সংখ্যা।

(২) সহগতি গুণাঙ্কের সূত্র থেকেই প্রতীয়মান, এটির মানসংখ্যা (n), সংশ্লিষ্ট চল দু'টির গাণিতিক গড় (মধ্যগামিতা-মাপক) এবং প্রমাণ বিচ্যুতি (বিস্তৃতি-মাপক) নিরপেক্ষ।

(৩) ঋজুরৈখিক রূপান্তরের ক্ষেত্রে সহগতি গুণাঙ্কের পরমমান অপরিবর্তনীয়, কিন্তু গাণিতিক চিহ্নের পরিবর্তন হতে পারে।

$u_i = \frac{x_i - a}{b}$ এবং $v_i = \frac{y_i - c}{d}, i = 1(1)n$, হলে

$$r_{xy} = \frac{bd}{|b||d|} r_{uv}, \qquad \dots (1.6)$$

অর্থাৎ, $r_{xy} = r_{uv},\ bd > 0$ হলে

$= -r_{uv},\ bd < 0$ হলে।

প্রমাণ :. $r_{uv} = \frac{Cov(u,v)}{\sqrt{Var(u)}\sqrt{Var(v)}}$

এখন $\bar{u} = \frac{1}{n}\sum_{i=1}^{n} u_i = \frac{1}{n}\sum_{i=1}^{n}\frac{x_i - a}{b} = \frac{\bar{x} - a}{b}$

এবং $S_u^2 = \frac{1}{n}\sum_{i=1}^{n}(u_i - \bar{u})^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - a}{b} - \frac{\bar{x} - a}{b}\right)^2$

$$= \frac{1}{b^2}\cdot\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{S_x^2}{b^2}$$

সুতরাং, $S_u = \frac{S_x}{|b|}$, যেহেতু প্রমাণবিচ্যুতির মান সবসময় ধনাত্মক।

একইভাবে, $\bar{v} = \frac{\bar{y} - c}{d}$ এবং $S_v = \frac{S_y}{|d|}$

আবার, $Cov(u,v) = \frac{1}{n}\sum_{i=1}^{n}(u_i - \bar{u})(v_i - \bar{v})$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - a)}{b} - \frac{\bar{x} - a}{b}\right)\left(\frac{y_i - c}{d} - \frac{\bar{y} - c}{d}\right)$$

$$= \frac{1}{bd}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

সুতরাং, $r_{uv} = \frac{1}{bd}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \,/\, \frac{S_x}{|b|}\cdot\frac{S_y}{|d|}$

$$=\frac{|b||d|}{bd}r_{xy} \Rightarrow r_{xy}=\frac{bd}{|b||d|}r_{uv}$$

লক্ষ্যণীয়, চল দু'টির এ জাতীয় ঋজু রৈখিক রূপান্তরের ক্ষেত্রে সহগতি গুণাঙ্কের ওপর a ও c'র আদৌ কোন প্রভাব পড়ে না। কেবল b ও d-'র গাণিতিক চিহ্ন আলাদা হলে গুণাঙ্কটির গাণিতিক চিহ্ন পরিবর্তিত হয়ে যায়, অন্যথায় একই থাকে। গুণাঙ্কটির পরমমান অবশ্য a, b, c, d–কোনটির ওপরই নির্ভরশীল নয়।

অনেক সময় আলোচ্য ফলটি ব্যবহার করে সহগতি গুণাঙ্কের মান অল্প আয়াসে নির্ণয় করা যায়। নিচের উদাহরণটি দেখ।

উদা. ১.৩ : নিচের সারণির প্রথম তিনটি স্তম্ভে প্রদত্ত তথ্যগুলি ব্যবহার করে বেতনভোগীদের বার্ষিক বেতন ও স্বল্প সঞ্চয়ের পরিমাণের মধ্যে সংহতি নির্ণয় কর।

ক্রমিক সংখ্যা (i)	বার্ষিক বেতন (হাজার টাকায়) (x_i)	স্বল্প-সঞ্চয়ের পরিমাণ (হাজার টাকায়) (y_i)	$u_i = 10(x_i - 200)$	$v_i = 10(y_i - 55)$	u_i^2	v_i^2	u_iv_i
1	195.8	50.2	–42	–48	1764	2304	2016
2	208.4	56.8	84	18	7056	324	1512
3	192.5	58.1	–75	31	5625	961	–2325
4	190.6	48.7	–94	–63	8836	3969	5922
5	200.0	60.0	0	50	0	2500	0
6	207.1	59.7	71	47	5041	2209	3337
7	198.2	53.5	–18	–15	324	225	270
8	191.7	55.0	–83	0	6889	0	0
9	208.8	60.0	88	50	7744	2500	4400
মোট			–69	70	43279	14992	15132

এখানে $b=\frac{1}{10}$, $d=\frac{1}{10}$, সুতরাং, $bd>0$।

অতএব, $r_{xy}=r_{uv}=\dfrac{9\times15132-(-69)(70)}{\sqrt{9\times43279-(-69)^2}\sqrt{9\times14992-(70)^2}}=0.6305$

রূপান্তরের ফলে মানগুলি অনেক ছোট হয়ে যাওয়ায় r-এর মান নির্ণয়ে খানিকটা সুবিধা হয়েছে। তবে বর্তমান যুগে যন্ত্রগণকের বহুল প্রচলনের ফলে r-এর এই ধর্মটি কাজে না লাগিয়েও চল দু'টির বড় বড় মান থেকে সহজেই গুণাঙ্কটির মান নির্ণয় করা যায়।

(৪) r-এর মান সবসময় -1 এবং $+1$-এর মধ্যে থাকে।

প্রমাণ : $r = \dfrac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{S_x S_y} = \frac{1}{n}.\sum_{i=1}^{n} u_i v_i$, যেখানে $u_i = \dfrac{x_i-\bar{x}}{S_x}$ এবং $v_i = \dfrac{y_i-\bar{y}}{S_y}$ ।

এখন $\Rightarrow \sum_{i=1}^{n}(u_i - v_i)^2 \geq 0$

$$\Rightarrow \sum_{i=1}^{n} u_i^2 + \sum_{i=1}^{n} v_i^2 \geq 2\sum_{i=1}^{n} u_i v_i$$

$$\Rightarrow \sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{S_x}\right)^2 + \sum_{i=1}^{n}\left(\frac{y_i-\bar{y}}{S_y}\right)^2 \geq 2\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{S_x}\right)\left(\frac{y_i-\bar{y}}{Y_y}\right)$$

$$\Rightarrow \frac{nSx^2}{Sx^2} + \frac{nSy^2}{Sy^2} \geq 2nr$$

$$\Rightarrow 2n \geq 2nr \Rightarrow r \leq 1 \qquad \text{... (1.7)}$$

আবার, $\sum_{i=1}^{n}(u_i + v_i)^2 \geq 0$

$$\Rightarrow \sum_{i=1}^{n} u_i^2 + \sum_{i=1}^{n} v_i^2 \geq -2\sum_{i=1}^{n} u_i v_i$$

$$\Rightarrow n + n \geq -2nr = 1 \geq -r$$

$$\Rightarrow -1 \leq r \qquad \text{... (1.8)}$$

(1.7) ও (1.8) থেকে পাওয়া যায়

$$-1 \leq r \leq 1 \text{ (প্রমাণিত)} \qquad \text{... (1.9)}$$

লক্ষ্যণীয় $r = 1$ হবে যদি প্রতিটি i-এর জন্য $u_i = v_i$ হয়, অর্থাৎ, $\dfrac{x_i-\bar{x}}{S_x} = \dfrac{y_i-\bar{y}}{S_y}$ হয়,

বা, $y_i = (\bar{y} - \frac{S_y}{S_x}\bar{x}) + \frac{S_y}{S_x}x_i$ হয়,

অর্থাৎ y এবং x-এর মধ্যে ধনাত্মক ঢাল-বিশিষ্ট একটি নিখুঁত ঋজুরৈখিক সম্পর্ক থাকে। তেমনি

$r = -1$ হবে, যদি প্রতিটি i-এর জন্য $u_i = -v_i$, বা $y_i = (\bar{y} - \frac{S_y}{S_x}\bar{x}) - \frac{S_y}{S_x}x_i$ হয়, অর্থাৎ y এবং x-এর মধ্যে ঋণাত্মক ঢালবিশিষ্ট একটি নিখুঁত ঋজুরৈখিক সম্পর্ক থাকে।

মন্তব্য : অনেক সময় r-এর নির্ণীত মান +1-এর থেকে বেশি কিংবা −1-এর থেকে কম পাওয়া গেলে সন্দেহ করা হয়, সম্ভবতঃ ব্যবহৃত (x_i, y_i), $i = 1(1)n$, এই রাশিতথ্যে কিছু ভুল ছিল। এটি ভ্রান্ত ধারণা। ব্যবহৃত রাশিতথ্য যাই হোক না কেন, r-এর মান অবশ্যই −1 এবং 1—এর মধ্যে থাকবে। এর অন্যথা হলে বুঝতে হবে অঙ্ক কষায় কোনরকম ভুল হয়েছে।

৩৭·৪ গোষ্ঠীবদ্ধ দ্বিচল রাশিতথ্য (Grouped Bivariate Data)

অর্থনৈতিক অথবা অন্য ধরনের সমীক্ষায় অনেক সময় সংগৃহীত দ্বিচল রাশিতথ্যের সংখ্যা (অর্থাৎ n-এর মান) বেশ বেশি হয়। সেক্ষেত্রে রাশিতথ্যগুলিকে প্রথমে সংক্ষেপিত করে নেওয়া হয় গোষ্ঠীবদ্ধ আকারে। ধরা যাক, x চলটির প্রসারকে $x_1' - x_1'', x_2' - x_2'', \ldots, x_k' - x_k''$ -এই মোট kটি উপপ্রসারে, এবং অনুরূপভাবে y-চলটির প্রসারকে $y_1' - y_1'', y_2' - y_2'', \ldots y_l' - y_l''$ -এই l টি উপপ্রসারে, ভাগ করে নেওয়া হল। এইবার একটি $k \times l$ দ্বিধারা ছক প্রস্তুত করে প্রতিটি (x, y) মান যে কোষে পড়ে সেখানে একটি করে মিলচিহ্ন বসিয়ে, বিভিন্ন কোষের জন্য প্রাপ্ত মিলচিহ্নগুলি গণনা করে আমরা (x, y)-চলের যৌথ পরিসংখ্যা বিভাজন (joint frequency distribution) পাব।

সারণি ৩৭.২ : x ও y চলের যৌথ পরিসংখ্যা বিভাজন

x \ y	$y_1' - y_1''$	$y_2' - y_2''$	…	$y_j' - y_j''$	…	$y_l' - y_l''$	মোট
$x_1' - x_1''$	f_{11}	f_{12}	…	f_{1j}	…	f_{1l}	f_{10}
$x_2' - x_2''$	f_{21}	f_{22}	…	f_{2j}	…	f_{2l}	f_{20}
$x_i' - x_i''$	f_{i1}	f_{i2}	…	f_{ij}	…	f_{il}	f_{i0}
$x_k' - x_k''$	f_{k1}	f_{k2}	…	f_{kj}	…	f_{kl}	f_{k0}
মোট	f_{01}	f_{02}	…	f_{0j}	…	f_{0l}	$f_{00} = n$

এখানে f_{ij} হল $k \times l$ দ্বিধারা শ্রেণীবিন্যাসে (i, j)-তম কোষের পরিসংখ্যা, অর্থাৎ (x, y)-এর n-জোড়া

মানের মধ্যে f_{ij}-টি যুগপৎ x-এর (x_i', x_i'') উপপ্রসারের এবং y-এর (y_i', y_i'') উপপ্রসারের অন্তর্গত। শ্রেণীবিন্যস্ত রাশিতথ্যে যেহেতু কিছুটা সংক্ষেপীকরণ অনিবার্য, আমরা চল দু'টির শ্রেণী মধ্যকগুলি $x_i\left(=\frac{x_i'+x_i''}{2}\right)$ এবং $y_j\left(=\frac{y_j'+y_j''}{2}\right)$ দ্বারা চিহ্নিত করে দ্বিচল পরিসংখ্যা বিভাজনটিকে সংক্ষেপে $(x_i, y_j, f_{ij}), i=1(1)k, j=1(1)l$; $n=\sum_i\sum_j f_{ij}$ —এইভাবে নির্দেশ করতে পারি। শ্রেণীমধ্যক নির্ণয়কালে অবশ্যই শ্রেণীসীমার বদলে শ্রেণীসীমান্ত নিতে হবে। এইভাবে দেওয়া তথ্য থেকে f_{ij}-সংখ্যক ব্যষ্টির প্রতিটির x ও y চলের প্রকৃত মানের পরিবর্তে আমরা যথাক্রমে x_i এবং y_j মান দু'টি রয়েছে বলে ধরে নেব। এটি স্পষ্টতই এক ধরনের সরলীকরণ এবং এর ফলে প্রকৃত তথ্য আংশিক বিসর্জন দেওয়া হল।

ধরা যাক, $\sum_{j=1}^{l} f_{ij} = f_{i0}$, $\sum_{i=1}^{k} f_{ij} = f_{0j}$ এবং

$$\sum_{i=1}^{k}\sum_{j=1}^{l} f_{ij} = \sum_{i=1}^{k} f_{i0} = \sum_{j=1}^{l} f_{0j} = f_{00} = n \quad \dots (1.10)$$

এই সংকেত চিহ্ন ব্যবহার করে $(x_i', -x_i'', f_{i0})$, বা এর সরলীকৃত রূপ $(x_i, f_{i0}), i=1(1)k$ —এই বিভাজনটিকে আমরা x-এর প্রান্তীয় বিভাজন (marginal distribution) বলব। তেমনি $(y_j' - y_j'', f_{0j})$ বা $(y_j, f_{oj}), j=1(1)l$ —এই বিভাজনটিকে বলা হবে y-এর প্রান্তীয় বিভাজন।

আবার প্রতিটি j'র জন্য $(x_i' - x_i'', f_{ij})$ বা, $(x_i, f_{ij}), i=1(1)k$, —এই বিভাজনটিকে আমরা $y \in (y_j' - y_j'')$ বা $y = y_j$ সাপেক্ষে x-এর শর্তাধীন বিভাজন (conditional distribution) বলব— এই শর্তাধীন বিভাজনের মোট পরিসংখ্যা $\sum_{i=1}^{k} f_{ij} = f_{0j}$। স্পষ্টতই j'র l-টি মানের জন্য আমরা x-এর l-টি শর্তাধীন বিভাজন পাব। তেমনি প্রতিটি i-এর জন্য $(y_i' - y_j'', f_{ij})$ বা $(y_j, f_{ij}), j=1(1)l$ —এই বিভাজনটিকে আমরা বলব $x \in (x_i' - x_i'')$ বা $x = x_i$ সাপেক্ষে y-এর মোট k-টি শর্তাধীন বিভাজন পাওয়া যাবে।

এখন, $\overline{x} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{l} x_i f_{ij} = \frac{1}{n}\sum_{i=1}^{k} x_i \sum_{j=1}^{l} f_{ij} = \frac{1}{n}\sum_{i=1}^{k} x_i f_{i0}$ এবং

$$S_x^2 = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{l}(x_i - \overline{x})^2 f_{ij} = \frac{1}{n}\sum_{i=1}^{k} f_{i0}(x_i - \overline{x})^2 \quad \text{... (1.11)}$$

তেমনি, $\overline{y} = \frac{1}{n}\sum_{j=1}^{l} y_j f_{0j}$ এবং $S_y^2 = \frac{1}{n}\sum_{j=1}^{l} f_{0j}(y_j - \overline{y})^2$ ... (1.12)

সুতরাং দেখা যাচ্ছে, গোষ্ঠীবদ্ধ দ্বিচল রাশিতথ্যের ক্ষেত্রে আমরা চল দু'টির গড় ও ভেদমান সংশ্লিষ্ট চলের প্রান্তীয় বিভাজন থেকেই পেয়ে যেতে পারি।

y-এর মান $(y_j' - y_j'')$ এই উপপ্রসারে থাকলে x-এর শর্তাধীন গড় $(\overline{x}\,|\,y_j) = \frac{1}{f_{0j}}\sum_{i=1}^{k} x_i f_{ij}$ এবং

x-এর শর্তাধীন ভেদমান $(S_x^2\,|\,y_j) = \frac{1}{f_{0j}}\sum_{i=1}^{k} x_i^2 f_{ij} - (\overline{x}^2\,|\,y_j)$ ... (1.13)

তেমনি x-এর মান $(x_i' - x_i'')$—এই উপপ্রসারে থাকলে চলটির শর্তাধীন গড় ও ভেদমান হবে

যথাক্রমে $(\overline{y}\,|\,x_i) = \frac{1}{f_{i0}}\sum_{j=1}^{l} y_j f_{ij}$, এবং $(S_y^2\,|\,x_i) = \frac{1}{f_{i0}}\sum_{j=1}^{l} y_j^2 f_{ij} - (\overline{y}^2\,|\,x_i)$ ... (1.14)

এক্ষেত্রে $\text{Cov}(x, y) = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{l} f_{ij}(x_i - \overline{x})(y_j - \overline{y})$

$$= \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{l} x_i y_j f_{ij} - \overline{x}\,\overline{y}$$

$$= \frac{1}{n}\sum_{i=1}^{k} x_i \sum_{j=1}^{l} y_j f_{ij} - \left(\frac{1}{n}\sum_{i-1}^{k}\sum_{j=1}^{l} x_i f_{ij}\right)\left(\frac{1}{n}\sum_{i-1}^{k}\sum_{j=1}^{l} y_j f_{ij}\right)$$

$$= \frac{1}{n}\sum_{i=1}^{k} x_i Y_i - \left(\frac{1}{n}\sum_{j=1}^{l} X_j\right)\left(\frac{1}{n}\sum_{i=1}^{k} Y_i\right) \quad \text{... (1.15)}$$

যেখানে $$Y_i = \sum_{j=1}^{l} y_i f_{ij}, i = 1(1)k \qquad \dots (1.16)$$

এবং $$X_j = \sum_{i=1}^{k} x_i f_{ij}, j = 1(1)l \qquad \dots (1.17)$$

এই সংকেত চিহ্ন ব্যবহার করে আমরা শ্রেণীবিন্যস্ত দ্বিচল রাশিতথ্যের ক্ষেত্রে নিম্নলিখিত সূত্রটির সাহায্যে সহজে r নির্ণয় করতে পারি :

$$r = \frac{n\sum_{i=1}^{k} x_i y_i - \left(\sum_{i=1}^{k} Y_i\right)\left(\sum_{j=1}^{l} X_j\right)}{\sqrt{n\sum_{i=1}^{k} x_i^2 f_{i0} - \left(\sum_{i=1}^{k} x_i f_{i0}\right)^2}\sqrt{n\sum_{j=1}^{l} y_j^2 f_{0j} - \left(\sum_{j=1}^{l} y_j f_{0j}\right)^2}} \qquad \dots (1.18)$$

লক্ষ্যণীয়, $$\sum_{i=1}^{k} x_i Y_i = \sum_{j=1}^{l} y_j X_j \qquad \dots (1.19)$$

$$\sum_{i=1}^{k} x_i f_{i0} = \sum_{j=1}^{l} X_j, \text{ এবং } \sum_{j=1}^{l} y_j f_{0j} = \sum_{i=1}^{k} Y_i \qquad \dots (1.20)$$

আগের মতই $u_i = \frac{x_i - a}{b}$ এবং $v_j = \frac{y_j - c}{d}$ —এই রূপান্তরের ক্ষেত্রে $bd > 0$ হলে $r_{xy} = r_{uv}$, এবং $bd < 0$ হলে $r_{xy} = -r_{uv}$ হবে। নিচের উদাহরণটি দেখ।

উদা. ১.৪ : 250টি পরিবারের মাসিক আয় এবং খাদ্য খাতে ব্যয়িত মাসিক আয়ের শতকরা অংশ সংক্রান্ত নিম্নপ্রদত্ত শ্রেণীবিন্যস্ত রাশিতথ্য থেকে চল দু'টির সহগতি গুণাঙ্ক নির্ণয় করুন।

যেসব পরিবারের মাসিক আয় 1,600 থেকে 2,400 টাকার মধ্যে, তাদের খাদ্য খাতে ব্যয়িত মাসিক আয়ের শতকরা অংশের গড় ও প্রমাণবিচ্যুতি কত? যেসব পরিবার মোট আয়ের 30 থেকে 35 শতাংশ খাদ্য খাতে ব্যয় করে, তাদের মাসিক আয়ের গড় ও প্রমাণবিচ্যুতি কত?

পরিবারের মাসিক আয়সূচক চলটিকে x এবং খাদ্য খাতে ব্যয়িত মাসিক আয়ের শতকরা অংশসূচক চলটিকে y ধরা যাক। শ্রেণীমধ্যকগুলিকে যথাক্রমে x_i ও y_j দ্বারা চিহ্নিত করে এবং $u_i = (x_i - 4.4)/0.8$ এবং $v_j = (y_j - 32.5)/5$ লিখে আমরা সারণি 1.4-এ অঙ্কপাতন করতে পারি r নির্ণয়ের জন্য।

সারণি ৩৭.৩ : 250টি পরিবারের মাসিক আয় ও খাদ্য খাতে ব্যয়িত মাসিক আয়ের শতকরা অংশের যৌথ বিভাজন

পরিবারের মাসিক আয় (হাজার টাকায়)	খাদ্য খাতে ব্যয়িত মাসিক আয়ের শতকরা অংশ						মোট
	20–25	25–30	30–35	35–40	40–45	45-50	
0.8–1.6				8	21	13	42
1.6–2.4			13	26	11	4	54
2.4–3.2			25	11	7		43
3.2–4.0		6	20	9	1		36
4.0–4.8		9	23	3			35
4.8–5.6		8	4				12
5.6–6.4	5	6					11
6.4–7.2	9	1					10
7.2–8.0	7						7
মোট	21	30	85	57	40	17	250

সারণি ৩৭.৩ থেকে দেখা যাচ্ছে,

$$r_{xy} = r_{uv} = \frac{250\times(-744)-(-360)(116)}{\{250\times1624-(-360)^2\}^{\frac{1}{2}}\{250\times484-116^2\}^{\frac{1}{2}}}$$

$$= -\frac{144240}{525.7376\times327.9390} = -0.8366$$

মাসিক আয় 1,600 – 2,400 হলে শর্তাধীন

$$\bar{v} = \frac{60}{54} = 1.1111 \Rightarrow \bar{x} = 38.94\%$$

এবং $S_v^2 = \frac{106}{54} - (1.1111)^2 = 0.7284 \Rightarrow S_v = 0.8535 \Rightarrow S_y = 4.267\%$

তেমনি মাসিক আয়ের 35% থেকে 35% খাদ্য খাতে ব্যয়িত হলে শর্তাধীন

$$\bar{u} = -\frac{105}{85} = -1.2353 \Rightarrow \bar{x} = 3.412 \text{ (হাজার টাকা)}$$

$$\text{এবং } S_u^2 = \frac{241}{85} - (-1.2353)^2 = 1.3093$$

$$\Rightarrow S_u = 1.1442 \Rightarrow S_x = 0.915 \text{ (হাজার টাকা)}$$

মন্তব্য ঃ প্রদত্ত দ্বিচল রাশিতথ্য দ্বিধারা শ্রেণীবিন্যাসের মাধ্যমে এভাবে সংক্ষেপিত করে বিভিন্ন মাপকের যেসব মান পাওয়া যায়, সেগুলি স্পষ্টতই আসন্ন মান। বর্তমান যুগে বহুল প্রচলিত এবং সহজলভ্য যন্ত্রগণকের সাহায্যে অবিন্যস্ত দ্বিচল রাশিতথ্যই বিশ্লেষণ করা যায় সহজে। এবং সেক্ষেত্রে বিভিন্ন মাপকেরও যথার্থ মানই পাওয়া যায়।

সারণি ৩৭.৪ : 250টি পরিবারের মাসিক আয় ও খাদ্য খাতে ব্যয়িত মাসিক আয়ের সহগতি গুণাঙ্ক নির্ণয়ন।

	y_j	22.5	27.5	32.5	37.5	42.5	47.5					
x_i	v_j / u_i	−2	−1	0	1	2	3	f_{i0}	$u_i f_{i0}$	$u_i^2 f_{i0}$	v_i	$u_i v_i$
1.2	−4				8	21	13	42	−168	672	89	−356
2.0	−3			13	26	11	4	54	−162	486	60	−180
2.8	−2			25	11	7		43	−86	172	25	−50
3.6	−1		6	20	9	1		36	−36	36	5	−5
4.4	0		9	23	3			35	0	0	−6	0
5.2	1		8	4				12	12	12	−8	−8
6.0	2	5	6					11	22	44	−16	−32
6.8	3	9	1					10	30	90	−19	−57
7.4	4	7						7	28	112	−14	−56
	f_{0j}	21	30	85	57	40	17	250	−360	1624	116	−744
	$v_j f_{0j}$	−42	−30	0	57	80	51	116				
	$v_j^2 f_{0j}$	84	30	0	57	160	153	484				
	u_j	65	17	−15	−141	−132	−64	−360				
	$v_j u_j$	−130	−17	0	−141	−264	−192	−744				

৩৭·৫ মানক্রম সহগতি গুণাঙ্ক (Rank Correlation Coefficient)

দু'টি লক্ষণের মধ্যে পূরণী পরিঘাত সহগতি গুণাঙ্ক নির্ণয়ের জন্য প্রয়োজন লক্ষণ দু'টির ওপর সংগৃহীত একপ্রস্থ সংখ্যামান। অর্থাৎ, লক্ষণ দু'টি মাপনযোগ্য হওয়া প্রয়োজন। কিন্তু কোন কোন গুণলক্ষণ আদৌ মাপনযোগ্য নয়—যেমন, শ্রমিকের দক্ষতা। কোন কোনটি আবার মাপনযোগ্য হলেও এগুলি মাপা সময় এবং/অথবা ব্যয়সাপেক্ষ, অথবা উপযুক্ত মাপকের দুষ্প্রাপ্যতার দরুন সহজসাধ্য নয়—যেমন, একটি দেশের মানব-সম্পদ বিকাশ অথবা ক্ষমতাবণ্টনে লিঙ্গ-বৈষম্য। অনেক সময় প্রয়োজনীয় সংখ্যামান পাওয়া গেলেও পরিঘাত পূরণী সহগতি গুণাঙ্ক নির্ণয়ের শ্রম ও সময় কাঁচামালের জন্য সহগতির স্থূলতর কোন বিকল্প মাপক উদ্ভাবনের প্রয়োজন হতে পারে। এইসব ক্ষেত্রে মানক্রম সহগতি গুণাঙ্ক (Rank correlation co-efficient) একটি সন্তোষজনক বিকল্প।

লক্ষণটি মাপনযোগ্য না হলেও অনেক সময় একগুচ্ছ ব্যষ্টিকে ঐ লক্ষণ সংক্রান্ত বৈশিষ্ট্যের মাত্রা অনুযায়ী ক্রমানুসারে সাজানো সম্ভব হয়। যেমন, শ্রমিকের দক্ষতা সহজে মাপা না গেলেও কয়েকজন শ্রমিককে তাদের সুপারভাইজার সহজেই দক্ষতার ক্রমানুসারে প্রথম (সবথেকে দক্ষ), দ্বিতীয়, তৃতীয়, ...এইভাবে সাজিয়ে দিতে পারে। n-সংখ্যক ব্যষ্টিকে এইভাবে সাজানো হলে এই বিন্যাসে ব্যষ্টিবিশেষের আপেক্ষিক অবস্থান নির্দেশক সংখ্যাটিকে তার মানক্রম (rank) বলে। গোষ্ঠীভুক্ত ব্যষ্টিদের যদি লক্ষণমাত্রার নিম্নগ ক্রম অনুসারে সাজানো হয়, তাহলে কোন ব্যষ্টির লক্ষণমাত্রা আলোচ্য ব্যষ্টির লক্ষণমাত্রা থেকে বেশি। স্পষ্টতই মানক্রম লক্ষণের প্রকৃতি, গোষ্ঠীভুক্ত ব্যষ্টির সংখ্যা ও গোষ্ঠীটির গঠনের ওপর নির্ভরশীল। 10 জন শ্রমিকের একটি দলে ক-এর দক্ষতা সংক্রান্ত মানক্রম 4 হলেও, সততা সংক্রান্ত মানক্রম 2 হতে পারে। আবার 10 জন শ্রমিকের অন্য একটি দলে এই ক-এরই দক্ষতা সংক্রান্ত মানক্রম 4-এর বদলে অন্য কিছু হতে পারে। দুই বা ততোধিক ব্যষ্টিকে একই মানক্রম দেওয়া হলে তাদের মধ্যে সমমানক্রম (tie) রয়েছে বলা হয়।

প্রকৃত সংখ্যামানের পরিবর্তে মানক্রম ব্যবহার করে কীভাবে সংশ্লিষ্ট লক্ষণ দু'টির সহগতি সম্বন্ধে একটা মোটামুটি ধারণা পাওয়া যায়, সে সম্বন্ধে নিচে আলোচনা করা হল।

৩৭·৫·১ স্পিয়ারম্যানের মানক্রম সহগতি গুণাঙ্ক (Spearman's Rank Correlation Coefficient)

n-সংখ্যক ব্যষ্টিকে দু'টি লক্ষণের বৈশিষ্ট্যমাত্রা অনুযায়ী আলাদা আলাদাভাবে ক্রমানুসারে সাজানো হলে লব্ধ দু'প্রস্থ মানক্রমের মধ্যে যে সহগতি লক্ষ্য করা যায়, তাকে বলা হয় লক্ষণ দু'টির মানক্রমিক সহগতি (rank correlation)। এই সহগতির একটি মাপক হল স্পিয়ারম্যানের মানক্রম সহগতি গুণাঙ্ক। এই মাপকটি হচ্ছে দু'প্রস্থ মানক্রমের পূরণী পরিঘাত সহগতি গুণাঙ্ক।

প্রথমে ধরা যাক, দু'প্রস্থ মানক্রমের কোনটির মধ্যেই সমমানক্রম নেই। i-তম ব্যষ্টির জন্য x-লক্ষণের মানক্রমকে u_i এবং y-লক্ষণের মানক্রমকে v_i, $i = 1(1)n$, বলা যাক। সমমানক্রম না থাকায়

$(u_1, u_2, \ldots, u_n)$এবং $(v_1, v_2, \ldots, v_n)$ এই দু'প্রস্থ মানক্রম প্রথম n-টি স্বাভাবিক সংখ্যা (1, 2, ..., n)-এর কোন না কোন বিন্যাস। সুতরাং,

$$\bar{u} = \frac{1}{n}\sum_{i=1}^{n} u_i = \frac{1}{n}\sum_{i=1}^{n} i = \frac{1}{n}.\frac{n(n+1)}{2} = \frac{n+1}{2} = \bar{v} \quad \ldots (1.21)$$

এবং $$S_u^2 = \frac{1}{n}\sum_{i=1}^{n} u_i^2 - \bar{u}^2 = \frac{1}{n}.\sum_{i=1}^{n} i^2 - \left(\frac{n+1}{2}\right)^2$$

$$= \frac{1}{n}.\frac{n(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4}$$

$$= \frac{n^2-1}{12} = S_v^2 \quad \ldots (1.22)$$

আবার $d_i = u_i - v_i$, $i = 1(1)n$, লিখে পাই

$$\frac{1}{n}\sum_{i=1}^{n} di^2 = \frac{1}{n}\sum_{i=1}^{n}(u_i - v_i)^2 = \frac{1}{n}\sum_{i=1}^{n}\{(u_i - \bar{u}) - (v_i - \bar{v})\}^2 \quad \text{যেহেতু } \bar{u} = \bar{v}$$

$$= S_u^2 + S_v^2 - 2\,\mathrm{Cov}(u, v)$$

$$\Rightarrow \mathrm{Cov}(u, v) = \left[\frac{1}{2}S_u^2 + S_v^2 - \frac{1}{n}\sum_{i=1}^{n} d_{i^2}\right]$$

$$= \frac{n^2-1}{12} - \frac{1}{2n}\sum_{i=1}^{n} d_{i^2} \quad \ldots (1.23)$$

সুতরাং মানক্রম সহগতি গুণাঙ্ক $$r_R = \frac{\mathrm{Cov}(u, v)}{S_u S_v} = \frac{\frac{n^2-1}{12} - \frac{1}{2n}\sum_{i=1}^{n} d_{i^2}}{\frac{n^2-1}{12}}$$

$$= 1 - \frac{6\sum_{i=1}^{n} d_{i^2}}{n(n^2-1)} \quad \ldots (1.24)$$

সহজেই দেখানো যায়, $-1 \le r_R \le 1$।

একটু খেয়াল করলে বোঝা যাবে, $\sum_{i=1}^{n} d_i^2$ -এর মান লঘিষ্ঠ (অর্থাৎ, 0) হবে যদি $d_i = 0$ অর্থাৎ $u_i = v_i \cdot \forall i$ হয়, অর্থাৎ লক্ষণ দু'টির মধ্যে যদি পূর্ণ ধনাত্মক সহগতি থাকে। আবার $\sum_{i=1}^{n} d_i^2$ -এর মান গরিষ্ঠ হবে যদি $v_i = n - u_i + 1$, অর্থাৎ $d_i = 2u_i - (n+1) \forall i$ হয়, অর্থাৎ লক্ষণ দু'টির মধ্যে যদি পূর্ণ ঋণাত্মক সহগতি থাকে। সুতরাং, $\sum_{i=1}^{n} d_i^2$ -এর গরিষ্ঠ মান হবে $\sum_{i=1}^{n} (2u_i - \overline{n+1})^2 = \frac{n(n^2-1)}{3}$ ।

সুতরাং, $0 \le \sum_{i=1}^{n} d_i^2 \le \frac{n(n^2-1)}{3}$

$$\Rightarrow -\frac{n(n^2-1)}{3} \le -\sum_{i=1}^{n} d_i^2 \le 0$$

$$\Rightarrow -1 \le r_R = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} \le 1 \qquad \dots (1.25)$$

স্পষ্টতই পূর্ণ ঋণাত্মক ও পূর্ণ ধনাত্মক সহগতির ক্ষেত্রে r_R – এর মান যথাক্রমে –1 ও 1.

এবার দেখা যাক, সমমানক্রম থাকলে r_R কীভাবে পাওয়া যাবে।

সমমানক্রমের ক্ষেত্রে মানক্রম দেওয়ার তিনটি পদ্ধতি আছে। ধরা যাক, r-টি মানক্রমের পর একটি k দৈর্ঘ্যের সমমানক্রম পরম্পরা আছে। প্রথম পদ্ধতিতে এই k-টি ব্যষ্টিকে অভিন্ন মানক্রম (r+1) দেওয়া হয় এবং (r+k+1)-তম ব্যষ্টির মানক্রম হয় (r+2)। দ্বিতীয় পদ্ধতিতে এই k-টি ব্যষ্টির মানক্রম প্রথম পদ্ধতির মতই প্রত্যেকের জন্য (r+1) হবে, কিন্তু (r+k+1)-তম ব্যষ্টির মানক্রম হবে (r+k+1)। তৃতীয় পদ্ধতিতে k-টি ব্যষ্টির অভিন্ন মানক্রম হবে তাদের সমমানক্রম না থাকলে যে গড় মানক্রম হত, তার সমান, অর্থাৎ $\frac{1}{k}[(r+1)+(r+2)+\ldots(r+k)] = r + \frac{(k+1)}{2}$, এবং $(r+k+1)$-তম ব্যষ্টির মানক্রম হবে $(r+k+1)$ ।

আমরা তৃতীয় পদ্ধতিটি অনুসরণ করব। এই পদ্ধতির সুবিধা, এর ফলে মোট মানক্রম $n(n+1)/2$-ই থাকে।

u-সারিতে r-টি মানক্রমের পর k-দৈর্ঘ্যের একটি সমমানক্রম পরম্পরা থাকলে r_R – এর ওপর তার কী প্রভাব পড়ে দেখা যাক।

স্পষ্টতই $\bar{u}$ অপরিবর্তিত থাকবে। এই k-টি মানক্রমের জন্য এখন $\sum_{i=1}^{k} u_i^2 = k\left(r + \frac{k+1}{2}\right)^2$

$= kr^2 + \frac{k(k+1)^2}{4} + rk(k+1)$ । কিন্তু সমমানক্রম না থাকলে এই সমষ্টি হত

$\sum_{i=1}^{k} u_i^2 = \sum_{i=1}^{k} (r+i)^2 = kr^2 + \frac{k(k+1)(2k+1)}{6} + rk(k+1)$, অর্থাৎ এই সমমানক্রম পরম্পরা থাকার দরুন S_u^2-এর মান কমে যাবে

$\frac{k(k+1)^2}{4} - \frac{k(k+1)(2k+1)}{6} = \frac{k(k^2-1)}{12}$ —এই পরিমাণে।

দেখা যাচ্ছে, এই পরিমাণটি r-এর ওপর নির্ভরশীল নয়, শুধুমাত্র পরম্পরা-দৈর্ঘ্য k-র ওপর নির্ভরশীল। এবং এটা সহজবোধ্য যে, $\sum_{i=1}^{n} u_i^2$ -এর ওপর একাধিক সমমানক্রমের প্রভাব সংযোজনশীল (additive)।

সুতরাং, u-সারিতে $k_1, k_2, \ldots, k_s$ দৈর্ঘ্যসম্পন্ন s-টি এবং v-সারিতে $k_1', k_2', \ldots, k_t'$ দৈর্ঘ্যসম্পন্ন t-টি সমমানক্রম পরম্পরা থাকলে এদের প্রভাবে S_u^2 ও S_v^2-এর মান পরিবর্তিত হয়ে দাঁড়াবে যথাক্রমে

$$S_u^2 = \frac{n^2-1}{12} - T_u, \text{ যেখানে } T_u = \frac{1}{12n}\sum_{i=1}^{s} k_i(k_i^2-1)$$

এবং $S_v^2 = \frac{n^2-1}{12} - T_v$, যেখানে $T_v = \frac{1}{12n}\sum_{i=1}^{t} k_i'(k_i'^2-1)$ ।

আবার, $\mathrm{Cov}(u,v) = \frac{S_u^2 + S_v^2}{2} - \frac{1}{2n}\sum_{i=1}^{n} d_i^2$

$$= \frac{n^2-1}{12} - \frac{T_u+T_v}{2} - \frac{1}{2n}\sum_{i=1}^{n} d_i^2$$

সুতরাং এক্ষেত্রে,

$$r_R = \frac{\frac{n^2-1}{12} - \frac{T_u+T_v}{2} - \frac{1}{2n}\sum_{i=1}^{n} d_i^2}{\left(\frac{n^2-1}{12} - T_u\right)^{1/2}\left(\frac{n^2-1}{12} - T_v\right)^{1/2}} \qquad \text{... (1.26)}$$

খেয়াল কর, লক্ষণ দু'টির মধ্যে পূর্ণ ধনাত্মক সহগতি থাকলে $u_i = v_i \; \forall i$ হবে, এবং সেক্ষেত্রে

$$r_R = \frac{\frac{n^2-1}{12} - T_u}{\frac{n^2-1}{12} - T_u} = 1 \text{ হবে।}$$

আবার লক্ষণ দু'টির মধ্যে পূর্ণ ঋণাত্মক সহগতি থাকলে আগের মতই $v_i = n - u_i + 1$ হবে, ফলে $T_u = T_v$, এবং $S_u^2 = S_v^2 = \frac{n^2-1}{12} - T_u$ হবে। এবং সেক্ষেত্রে

$$\frac{1}{n}\sum_{i=1}^{n} d_i^2 = \frac{1}{n}\sum_{i=1}^{n}[2u_i - (n+1)]^2$$

$$= 4S_u^2 = \frac{n^2-1}{3} - 4Tu \text{ হবে, এবং}$$

$$r_R = \frac{\frac{n^2-1}{12} - T_u - \left(\frac{n^2-1}{6} - 2T_u\right)}{\frac{n^2-1}{12} - T_u} = -1 \text{ হবে।}$$

মন্তব্য : দু'টি ক্ষেত্রেই r_R পাওয়ার জন্য সংশ্লিষ্ট সূত্র দু'টি [যথাক্রমে (1.25) ও (1.26)] ব্যবহার না করে আমরা সরাসরি u-সারি v-সারির মধ্যে পূরণী পরিঘাত সহগতি গুণাঙ্ক নির্ণয় করতে পারতাম।

উদা. ১.৫ : সম্মিলিত জাতিপুঞ্জের (Human Development Report, 1998) থেকে পাওয়া কয়েকটি দেশের 1995-এর জন্য মানব-বিকাশ সূচক (Human Development Index : HDI) এবং ক্ষমতা বণ্টনে লিঙ্গ বৈষম্যের মাপক (Gender Empowerment Measure : GMI) সংক্রান্ত নিচের সারণিতে দেওয়া তথ্য থেকে সংশ্লিষ্ট লক্ষণ দু'টির মধ্যে স্পিয়ারম্যানের মানক্রমিক সহগতি গুণাঙ্ক নির্ণয় কর।

দেশ (1)	HDI (2)	GEM (3)	প্রদত্ত 13টি দেশের ভিত্তিতে (4) HDI মানক্রম(u_i)	(5). GEM মানক্রম(v_i)
কানাডা	0.960	0.720	1	1
আমেরিকা	0.943	0.675	2	3
জাপান	0.940	0.472	3	8
ইংল্যান্ড	0.932	0.593	4	5
সুইজারল্যান্ড	0.930	0.654	5	4
জার্মানি	0.925	0.694	6	2
মালয়েশিয়া	0.834	0.458	7	9
কিউবা	0.729	0.523	8	6
শ্রীলঙ্কা	0.716	0.286	9	11
চিন	0.650	0.483	10	7
পাকিস্তান	0.452	0.179	11	13
ভারত	0.451	0.228	12	12
বাংলাদেশ	0.371	0.305	13	10

প্রথমে সারণটির চতুর্থ ও পঞ্চম স্তম্ভে আমরা লক্ষণ দু'টির বিচারে দেশগুলির মানক্রম বসিয়ে নিই।

খেয়াল রাখা দরকার, এগুলি গৃহীত 13টি দেশের ভিত্তিতে দেওয়া মানক্রম—Human Development Report-এ প্রদত্ত সবগুলি দেশের বিচারে মানক্রম নয়।

দেখা যাচ্ছে, n = 13 এবং

$$\sum_{i=1}^{n} d_i^2 = 0+1+25+1+1+16+4+4+4+9+4+0+9 = 78$$

সুতরাং, স্পিয়ারম্যানের মানক্রম সহগতি গুণাঙ্কের মান হবে

$$r_R = 1 - \frac{78 \times 6}{13 \times 168} = 1 - 0.2143 = 0.7857 \text{ ।}$$

উদা. ১.৬ : দক্ষতার বিচারে দু'জন সুপারভাইজারের দেওয়া 12 জন শ্রমিকের মানক্রম নিচের সারণিতে লিপিবদ্ধ করা হল। স্পিয়ারম্যানের মানক্রম সংগতি গুণাঙ্কের সাহায্যে দুই সুপারভাইজারের বিচারের সাযুজ্যতা নিরূপণ কর।

শ্রমিক	ক	খ	গ	ঘ	ঙ	চ	ছ	জ	ঝ	ঞ	ট	ঠ
১ম সুপারভাইজারের দেওয়া মানক্রম	5	6	1	2	3	8.5	8.5	4	7	11	11	11
২য় সুপারভাইজারের দেওয়া মানক্রম	6	6	2	2	2	9	6	4	8	10.5	12	10.5

খেয়াল করা দরকার, এখানে বিচার্য লক্ষণ দু'টি হল শ্রমিকদের আপেক্ষিক দক্ষতার প্রশ্নে প্রথম ও দ্বিতীয় সুপারভাইজারের মত।

প্রথম মানক্রম সারিতে 2 এবং 3 দৈর্ঘ্যের দু'টি সমমানক্রম পরম্পরা আছে। সুতরাং,

$$T_u = \frac{1}{12}\left[\frac{2^3-2}{12} + \frac{3^3-3}{12}\right] = 0.2083$$

তেমনি দ্বিতীয় মানক্রম সারিতে রয়েছে 3, 3 ও 2 দৈর্ঘ্যের তিনটি সমমানক্রম পরম্পরা। অতএব,

$$T_v = \frac{1}{12}\left[\frac{2^3-2}{12} + \frac{3^3-3}{12} + \frac{3^3-3}{12}\right] = 0.3750$$

এদিকে $\frac{1}{12}\sum_{i=1}^{12} d_i{}^2 = \frac{1}{12}[1+0+1+0+1+0.25+2.25+0+1+0.25+1+0.25] = 0.6667$

ফলে $S_u^2 = \frac{12^2-1}{12} - 0.2083 = 11.7084,$

$$S_v^2 = \frac{12^2-1}{12} - 0.3750 = 11.5417$$

এবং $r_R = \frac{11.9167 - \frac{1}{2}(0.2083+0.3750) - \frac{0.6667}{2}}{\{11.7084 \times 11.5417\}^{1/2}}$

$$= 11.2917 / 11.6248 = 0.9713.$$

৩৭.৫.২ কেন্ডালের মানক্রম সহগতি গুণাঙ্ক (Kendall's Rank Correlation Coefficient)

কেন্ডালের τ হল আর একটি মানক্রম সহগতি গুণাঙ্ক। এটির নির্ণয় পদ্ধতি নিয়ে এবার আলোচনা করা হবে।

প্রথমে ধরা যাক, দু'টি মানক্রম সারির কোনটিতেই সমমানক্রম নেই।

n-সংখ্যক ব্যষ্টি থেকে আমরা মোট $\binom{n}{2}$-টি জোড়া পেতে পারি। প্রতিটি জোড়ার জন্য দু'টি লক্ষণের মানক্রম যদি স্বাভাবিক বিন্যাসে থাকে যেমন, [(2,5), (3,7); অথবা (6,4), (11,8)], তাহলে ঐ জোড়ার জন্য +1 স্কোর এবং যদি বিপরীত বিন্যাসে থাকে [যেমন (2,5), (7,3); অথবা (6,4), (5,7)], তাহলে –1 স্কোর ধার্য করা হবে। স্পষ্টতই,সম্ভাব্য সর্বোচ্চে মোট স্কোর হবে $\binom{n}{2}$। কেন্ডালের মানক্রম সহগতি গুণাঙ্ক τ হল

$$\tau = \frac{\text{মোট স্কোর}}{\text{সম্ভাব্য সর্বোচ্চ মোট স্কোর}} = \frac{\text{মোট স্কোর}}{n(n-1)/2} \qquad \ldots (1.27)$$

লক্ষণ দু'টির মধ্যে পূর্ণ ধনাত্মক সহগতি থাকলে প্রতিটি জোড়ার জন্য স্কোর হবে +1 এবং সেক্ষেত্রে $\tau=1$; আবার লক্ষণ দু'টির মধ্যে পূর্ণ ঋণাত্মক সহগতি থাকলে প্রতিটি জোড়ার জন্য –1 স্কোর ধার্য হবে এবং সেক্ষেত্রে $\tau = -1$।

τ-নির্ণয়ের একটি সহজতর পদ্ধতি হল, ব্যষ্টিগুলিকে এমনভাবে পর পর নেওয়া যাতে প্রথম লক্ষণের মানক্রমগুলি (1, 2, ..., n)–এই স্বাভাবিক বিন্যাসে থাকে। এখন ধরা যাক, ব্যষ্টিগুলির এই বিন্যাসে দ্বিতীয় লক্ষণ দু'টির মানক্রমগুলির P-জোড়া স্বাভাবিক বিন্যাসে এবং বাকি $Q = \binom{n}{2} - P$ জোড়া বিপরীত বিন্যাসে রয়েছে। সেক্ষেত্রে,

$$\tau = \frac{P-Q}{\binom{n}{2}} = \frac{2P}{\binom{n}{2}} - 1 = 1 - \frac{2Q}{\binom{n}{2}} \qquad \ldots (1.28)$$

আমরা অন্য একভাবেও τ পেতে পারি। লক্ষণ দু'টির জন্য মানক্রমগুলিকে যদি আগের মত $(u_1, u_2, \ldots, u_n)$ এবং $(v_1, v_2, \ldots, v_n)$ লেখা হয়, তাহলে (i, j)—এই জোড়াটির জন্য ধরা যাক,

$a_{ij} = 1$, যদি $u_i < u_j$ হয়

$= -1$, যদি $u_i > u_j$ হয়;

এবং $b_{ij} = 1$, যদি $v_i < v_j$ হয়

$= -1$, যদি $v_i > v_j$ হয়

তাহলে স্পষ্টতই $$\tau = \frac{\sum a_{ij} b_{ij}}{\sqrt{\sum_c a_{ij}^2}\sqrt{\sum b_{ij}^2}} = \frac{\sum a_{ij} b_{ij}}{\binom{n}{2}} \qquad \ldots (1.29)$$

[এই সূত্রে যোগক্রিয়াটি সমস্ত (i, j)-জোড়ার জন্য, যেখানে $i < j$] ।

τ -এর জন্য (1.29) সূত্রটি সমমানক্রমের ক্ষেত্রে সহজে ব্যবহার করা যায়। এক্ষেত্রে a_{ij} ও b_{ij}-'র মান নেওয়া হবে

$$a_{ij} = 1,\ u_i < u_j \text{ হলে,}$$
$$= 0,\ u_i = u_j \text{ হলে}$$
$$= -1,\ u_i > u_j \text{ হলে;}$$

এবং $$b_{ij} = 1, v_i < v_j \text{ হলে,}$$
$$= 0, v_i = v_j \text{ হলে}$$
$$= -1, v_i > v_j \text{ হলে।}$$

স্পষ্টতই (i, j) জোড়াটির জন্য $a_{ij}\, b_{ij} = 0$ হবে, যদি a_{ij} অথবা b_{ij} অথবা উভয়ের মানই 0 হয়।

এখন u-সারিতে যদি k-দৈর্ঘ্যের একটি সমমানক্রম পরম্পরা থাকে তাহলে $\binom{k}{2} = k(k-1)/2$ টি a_{ij}'র মান শূন্য হবে। অর্থাৎ, $\sum a_{ij}^2$-এর মান এর প্রভাবে $k(k-1)/2$ পরিমাণে কমে যাবে। সুতরাং, u সারিতে $k_1, k_2, \ldots, k_s$ দৈর্ঘ্যের s টি এবং v-সারিতে $k_1', k_2', \ldots, k_t'$ দৈর্ঘ্যের t-টি সমমানক্রম পরম্পরা থাকলে স্পষ্টতই

$$\sum a_{ij}^2 = \frac{n(n-1)}{2} - T_u, \text{ যেখানে } T_u = \frac{1}{2}\sum_{i=1}^{s} k_i(k_i - 1)$$

এবং $\sum b_{ij}^2 = \frac{n(n-1)}{2} - T_v$, যেখানে $T_v = \frac{1}{2}\sum_{i=1}^{t} k_i'(k_i' - 1)$ হবে, এবং এর ফলে τ-এর মান হবে

$$\tau = \frac{\text{মোট স্কোর}}{\left\{\frac{n(n-1)}{2} - T_u\right\}^{1/2}\left\{\frac{n(n-1)}{2} - T_v\right\}^{1/2}} \qquad \ldots (1.30)$$

সহগতি সম্বন্ধে একটা মোটামুটি ধারণা পাওয়ার জন্য r_R এবং τ দু'টি গুণাঙ্কই খুব উপযোগী—বিশেষ করে যেসব ক্ষেত্রে লক্ষণ দু'টি মাপনযোগ্য নয়, বা সহজে মাপা যায় না। পূরণী পরিঘাত সহগতি গুণাঙ্ক r-এর তুলনায় r_R এবং τ অল্প আয়াসে নির্ণয় করা যায়। এছাড়া এই গুণাঙ্ক দু'টি চলের ক্রম-অক্ষুণ্ণ রূপান্তরের ক্ষেত্রে অপরিবর্তিত থাকে— সহগতি গুণাঙ্ক r-এর যে ধর্মটি নেই।

উদা. ১.৭ : ১.৫ উদাহরণে প্রদত্ত রাশিতথ্যের জন্য τ নির্ণয় করা যাক।

খেয়াল করো, এখানে HDI মানক্রমগুলি স্বাভাবিক বিন্যাসে [(1, 2, ...1, 3)-এইভাবে] নেওয়া হয়েছে। GEM মানক্রম 1-এর স্বাভাবিক ও বিপরীত বিন্যাসে অবস্থিত মানক্রমের সংখ্যা যথাক্রমে 12 এবং 0, 3-এর জন্য যথাক্রমে 10 এবং 1, ...এইভাবে এগোলে আমরা দেখব মোট স্কোর =(12–0) + (10–1) + (5–5) + (7–2) + (7–1) + (7–0) + (4–2) + (5–0) + (2–2) + (3–0) + (0–2) + (0–1)=46। সুতরাং, $\tau = 46/\binom{13}{2} = \frac{46}{78} = 0.5897$

উদা. ১.৮ : ২.৬ উদাহরণ প্রদত্ত রাশিতথ্যের ক্ষেত্রে τ-এর মান নির্ণয় করার জন্য প্রথম সুপারভাইজারের দেওয়া মানক্রম সাধারণ বিন্যাসে রেখে প্রথমে সারণিটি নতুন করে সাজানো যাক :

প্রথম সুপারভাইজারের দেওয়া মানক্রম	1	2	3	4	5	6	7	8.5	8.5	11	11	11
দ্বিতীয় সুপারভাইজারের দেওয়া মানক্রম	2	2	2	4	6	6	8	9	6	10.5	12	10.5

এখানে মোট স্কোর = (9 – 0) + (8 – 0) + (9 – 0) + (8 – 0) + (5 – 0) + (5 – 0) + (4 – 1) + (3 – 1) + (3 – 0) + (1 – 0) + (0 – 1) = 53

$$T_u = \frac{1}{2}[2\times1+3\times2] = 4$$

$$T_v = \frac{1}{2}[3\times2+3\times2+2\times1] = 7$$

অতএব, $\tau = \frac{53}{\sqrt{66-4}\sqrt{66-7}} = 0.8763$

৩৭.৬ সারাংশ

আলোচ্য পরিচ্ছেদে প্রথমে দ্বিচল রাশিতথ্যের প্রকৃতি এবং এই ধরনের রাশিতথ্য বিশ্লেষণের উদ্দেশ্য বর্ণনা করা হয়েছে। বিক্ষেপণ চিত্রের সাহায্যে দু'টি চলের পারস্পরিক সহগতির ধারণা দেওয়া হয়েছে এবং আদর্শ সহগতি মাপকের বিভিন্ন ধর্মের কথা বলা হয়েছে। এই প্রসঙ্গে সহগতি গুণাঙ্কের সংজ্ঞা, প্রকৃতি ও বিভিন্ন ধর্ম নিয়ে আলোচনা এসেছে। এরপর বলা হয়েছে দু'টি চলের পারস্পরিক সম্পর্ক নির্ধারণে এই গুণাঙ্কের সীমাবদ্ধতার কথাও। গোষ্ঠীবদ্ধ রাশিতথ্যের ক্ষেত্রে কীভাবে এই গুণাঙ্ক নির্ণয় করা হবে সে সম্বন্ধে আলোচনা করা হয়েছে এরপর। সবশেষে মানক্রম সহগতির প্রসঙ্গ এসেছে এবং এই ধরনের সহগতি মাপনার জন্য স্পিয়ারম্যান এবং কেন্ডালের গুণাঙ্ক নিয়ে আলোচনা করা হয়েছে।

৩৭.৭ অনুশীলনী

(ক) বড় প্রশ্ন :

(১) দ্বিচল রাশিতথ্যের সংজ্ঞা দিন। দ্বিচল রাশিতথ্য বিশ্লেষণের উদ্দেশ্য বর্ণনা করুন।

(২) বিক্ষেপণ চিত্র কী? দু'টি চলের মধ্যে বিভিন্ন ধরনের সম্পর্ক বিচারে বিক্ষেপণ চিত্র কীভাবে সাহায্য করে?

(৩) সহগতি-র সংজ্ঞা দিন। উদাহরণ সহযোগে বিভিন্ন ধরনের সহগতি-র পরিচয় দিন।

(৪) একটি উপযুক্ত সহগতি-মাপকের কোন্ কোন্ ধর্ম থাকা বাঞ্ছনীয়?' এগুলির মধ্যে সহগতি গুণাঙ্কের কোন্ কোন্ ধর্ম রয়েছে, প্রমাণ সহযোগে তা বিবৃত করুন।

(৫) মানক্রমিক সহগতি কাকে বলে? মানক্রমিক সহগতি গুণাঙ্কের উপযোগিতা বর্ণনা করুন।

সমমানক্রম থাকলে স্পিয়ারম্যান এবং কেন্ডালের মানক্রমিক সহগতি গুণাঙ্কের প্রকাশন দু'টি নির্ণয় করুন।

(৬) $[u_1, u_2, \dots u_n]$এবং $[v_1, v_2, \dots v_n]$ যদি দু'প্রস্থ বাস্তব রাশি হয়, তবে দেখান যে

$$\sum_{i=1}^{n} u_i^2 \sum_{i=1}^{n} v_i^2 \geq \left(\sum_{i=1}^{n} u_i v_i \right)^2$$

[এই সম্পর্কটি কোশি-স্কোয়ার্জ অসমতা (Cauchy-Scwartz Inequality) বলে]

এই ফলটি ব্যবহার করে প্রমাণ করুন $-1 \leq r \leq 1$.

(৭) যদি $u_i = a_0 + a_1 x_i + a_2 y_i$

এবং $v_i = b_0 + b_1 x_i + b_2 y_i, \ i = 1(1)n,$ হয়, তাহলে r_{uv} কে $a_0, a_1, a_2, b_0, b_1, b_2,$ S_x, S_y এবং r_{xy} -এর সাহায্যে প্রকাশ করুন।

(৮) ৫০ জন ছাত্রের উচ্চতা (সেমি-তে) ও ওজনের (কেজি-তে) ভিত্তিতে চল দু'টির সহগতি গুণাঙ্ক দাঁড়াল 0.6। পরে ধরা পড়ল যে, দু'জন ছাত্রের ওজন ও উচ্চতা (45, 168) ও (48, 152)-এর বদলে ভুলক্রমে (50, 160) ও (43, 160) লেখা হয়েছিল। এই ভুল সংশোধন করে r-এর সংশোধিত মান নির্ণয় করুন।

(৯) স্নাতক পরীক্ষায় কয়েকজন ছাত্রের সাম্মানিক অর্থনীতি ও গণিতের নম্বর দেওয়া হল :

ছাত্রের ক্রমিক সংখ্যা	1	2	3	4	5	6	7	8
সাম্মানিক অর্থনীতির নম্বর (%)	52	43	34	58	61	52	58	6
গণিতের নম্বর (%)	56	52	35	56	56	35	65	7

সাম্মানিক অর্থনীতির নম্বরকে y এবং গণিতের নম্বরকে x দিয়ে নির্দেশ করে এই রাশিতথ্যের ভিত্তিতে :

(১) x ও y-এর সহগতির গুণাঙ্ক নির্ণয় করুন।

(২) x-এর ওপর y-এর এবং y-এর ওপর x-এর নির্ভরণ রেখা নির্ণয় করুন।

যে ছাত্র গণিতে 65% পেয়েছে তার সাম্মানিক অর্থনীতির নম্বরের এবং যে ছাত্র সাম্মানিক অর্থনীতিতে 70% পেয়েছে তার গণিতের নম্বরের অনুমিত মান বের করুন।

(৩) চল দু'টির মধ্যে স্পিয়ারম্যান ও কেন্ডালের মানক্রমিক সহগতি গুণাঙ্ক নির্ণয় করুন।

(খ) **ছোট প্রশ্ন :**

(১০) দুটি চলের মধ্যে ধনাত্মক সহগতি, ঋণাত্মক সহগতি ও সহগতি শূন্যতার সংজ্ঞা দাও। এই তিন ধরনের সহগতির একটি করে বাস্তবসম্মত উদাহরণ দাও।

(১১) নিচের একটি ক্ষেত্রে বিক্ষেপণ চিত্রটি অঙ্কন করুন :

(1) $r = 1$ (2) $r = -1$ (3) $r = 0$

(4) $r = 0.8$ (5) $r = -0.6$

(১২) আদর্শ সহগতি-মাপকের কোন্ কোন্ ধর্ম স্পিয়ারম্যানের মানক্রমিক সহগতি গুণাঙ্কের রয়েছে?

(গ) **বস্তুমুখী প্রশ্ন**

(১৩) নিচের প্রশ্নগুলির প্রত্যেকটির জন্য সঠিক উত্তরটিতে দাগ দাও।

(i) x ও y-এর দু'জোড়া মান (x_1, y_1) এবং (x_2, y_2)'র ভিত্তিতে চল দু'টির সহগতি গুণাঙ্ক r নির্ণয় করা হল। x_1, x_2, y_1, y_2 – র সংখ্যা মান জানা না থাকলেও r-এর মান সম্বন্ধে কতখানি বলা যায়?

A : কিছুই বলা যাবে না। B : $r = -1$ অথবা $+1$

C : $r = 0$ । D : $r = -1$, অথবা $+1$, অথবা অনির্ণেয়।

(ii) কলকাতার কয়েকটি কলেজ থেকে 25 জন ছাত্রছাত্রীর একটি করে নমুনা নেওয়া হল। x ও y যদি এই নমুনা যথাক্রমে ছাত্র ও ছাত্রীর সংখ্যা হয়, তাহলে x ও y-এর সহগতি গুণাঙ্ক কত হবে?

A : 0 B : +1 c : –1

D : প্রত্যেক নমুনার জন্য x ও y-এর মান জানা না থাকলে বলা সম্ভব নয়।

(iii) $u = 2x - 3$, $v = 2y + 3$, $r_{xy} = 0.4$, R_{xy} (মানক্রমিক সহগতি গুণাঙ্ক) $= 0.5$। তাহলে নিচের কোন্ কোন্ উক্তি সত্য জানাতে সঠিক বিকল্পে দাগ দাও।

I $r_{uv} = 0.8$

II $R_{uv} = 0.5$

III $r_{ux} = 1$

A : কেবল I ও II সত্য

B : কেবল I ও III সত্য

C : কেবল II ও III সত্য

D : I, II, III — সবগুলিই সত্য।

(iv) (x ও y)-এর একপ্রস্থ মানের ভিত্তিতে পাওয়া নিচের কোন্ কোন্ মাপকগুলির মান পরস্পর অসমঞ্জস?

A : $v(x) = 36$, $v(y) = 64$, $Cov(x, y) = 40$

B : $v(x) = 64$, $v(y) = 36$, $Cov(x, y) = 30$

C : $v(x) = 100$, $v(y) = 49$, $Cov(x, y) = 75$

D : $v(x) = 81$, $v(y) = 25$, $Cov(x, y) = 36$

৩৭.৮ গ্রন্থপঞ্জী

(1) শৈলেশ ভূষণ চৌধুরী, অরিজিৎ চৌধুরী ও বিশ্বনাথ দাস, রাশি বিজ্ঞানের মূলতত্ত্ব, প্রথম ও দ্বিতীয় খণ্ড (দশম ও দ্বাদশ পরিচ্ছেদ)। পঃ বঙ্গ রাজ্য পুস্তক পর্ষদ, ১৯৭৬।

(2) A. M. Goon, M. K. Gupta & B. Dasgupta : Fundamentals of Statistics, Vol 1 (Chapter 12), World Press, 1998.

(3) S. C. Gupta & V. K. Kapoor : Fundamentals of Mathematical Statistics (Ch 9 & 10), Sultan Chand & Sons, 1995.

(4) Croxton, Cowden & Klien : Applied General Statistics (Ch 19), Prentice Hall of India, 1973.

(5) Ezekiel, M, Fox, K. A. & John Wiley, Methods of Correlation & Regression Analysis (Chs 5-9), 1959

একক ৩৮ □ দ্বিচল রাশিতথ্য বিশ্লেষণ : নির্ভরণ

গঠন

৩৮·০ উদ্দেশ্য

এই এককটি পড়বার পর আপনি বুঝতে পারবেন—

- রাশিতথ্য বিশ্লেষণে নির্ভরণ কাকে বলে ও তার বিভিন্ন আলোচনা এবং প্রয়োগ পদ্ধতি
- লগ রূপান্তরের ও ব্যস্ত রূপান্তরের সাহায্যে সরলরেখার প্রকৃতি
- মডিফায়েড এক্সপোনেনশিয়াল কাকে বলে, ইত্যাদি।

৩৮·১ প্রস্তাবনা

বিক্ষেপণ-চিত্র থেকে যদি আভাস পাওয়া যায় যে গৃহীত চল দু'টির গতিশীলতার মধ্যে একটা সুস্পষ্ট সম্পর্ক আছে, তাহলে দ্বিচল রাশিতথ্য বিশ্লেষণের পরবর্তী ধাপে আমাদের চেষ্টা হবে স্বতন্ত্র চলটির ভিত্তিতে

নির্ভরী চলের জন্য একটি পূর্বানুমান সূত্র (forecasting formula) হিসাবে ব্যবহার করার জন্য প্রদত্ত রাশিতথ্য কাজে লাগিয়ে এই সম্পর্কটি নির্ধারণ করা। সাধারণভাবে স্বতন্ত্র চলের ওপর নির্ভরী চলের নির্ভরতার প্রবণতাকে বলে প্রথম চলটির ওপর দ্বিতীয় চলের নির্ভরণ (regression)। নির্ভরণ সম্পর্কটি যে কোন ধরনের হতে পারে—যেমন ঋজুরৈখিক, বহুঘাতজক (polynomial), সূচক অপেক্ষক (exponential function), সংশোধিত সূচক (modified exponential) অপেক্ষক ইত্যাদি।

প্রদত্ত একপ্রস্থ রাশিতথ্যের সাযুজ্যে এই ধরনের সম্পর্ক নিরূপণ করাকে সাযুজ্য রেখা নিরূপণ (curve fitting) বলে। নির্ভরণ সম্পর্কটি ঋজুরৈখিক বা বহুঘাতজক হিসাবে প্রকাশ করা সম্ভব হলে লঘিষ্ঠ বর্গসমষ্টি (method of least squares) পদ্ধতিতে সংশ্লিষ্ট সাযুজ্যরেখাটি নিরূপণ করা যায়। পরবর্তী অনুচ্ছেদে এই পদ্ধতিটি নিয়ে আলোচনা করা হয়েছে। সাযুজ্যরেখা নিরূপণের আর একটি পদ্ধতি হল গোষ্ঠী-গড় পদ্ধতি (method of group averages)। এটি নিয়েও আলোচনা করা হয়েছে পরবর্তী আর একটি অনুচ্ছেদে।

নির্ভরণ সংক্রান্ত বর্তমান আলোচনা মোটামুটিভাবে ঋজুরৈখিক নির্ভরণের (linear regression) মধ্যেই সীমাবদ্ধ রাখা হবে। বক্ররৈখিক নির্ভরণ (curvilinear regression) নিয়ে বিশেষ কিছু বলার সুযোগ এখানে থাকবে না।

৩৮·২ লঘিষ্ঠ বর্গ-সমষ্টি পদ্ধতি (Method of Least Squares)

নির্ভরী চল y ও স্বতন্ত্র চল x-এর সম্পর্কটির জন্য ধরা যাক নিম্নলিখিত প্রতিরূপ (model)-টি প্রযোজ্য :

$$y = \phi\,(x;\ \theta_1, \theta_2, \ldots, \theta_k) + e \qquad \ldots (1.31)$$

যেখানে $\phi(x; \theta_0, \theta_1, \ldots, \theta_k) = \theta_0 + \theta_1 x + \theta_2 x^2 + \ldots + \theta_k x^k$ হল $\theta_0, \theta_1, \ldots, \theta_k$ এই $(k+1)$ টি বিশেষকের (parameter) ওপর নির্ভরশীল। x-এর একটি বহুঘাতজক, এবং e হল ভ্রান্তি-উপাদান। y-এর যে অংশটি x-এর ওপর নির্ভরশীল নয়, সেটিকেই e বলা যেতে পারে। অথবা e-কে শুধুই y-এর অবেক্ষণ ভ্রান্তি (error of observation) বলা যায়।

ধরা যাক, চল দু'টির ওপর (x_i, y_i), $i = 1(1)n$—এই n-জোড়া মান জানা আছে। গৃহীত প্রতিরূপ অনুযায়ী

$$y_i = \theta_0 + \theta_1 x_i + \theta_2 x_i^2 + \ldots + \theta_k x_i^k + e_i \qquad \ldots (1.32)$$
$$i = 1(1)n\ldots\ldots.$$

সুতরাং, $$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \theta_0 - \theta_1 x_i - \theta_k x_i^k)^2 \qquad \ldots (1.33)$$

হল বর্গ ভ্রান্তি সমষ্টি, যা (x_i, y_i), $i = 1(1)n$, জানা থাকলে শুধুমাত্র $\theta_0, \theta_1, \ldots, \theta_k$-র ওপর নির্ভর করে। লঘিষ্ঠ বর্গসমষ্টি পদ্ধতিতে $\theta_0, \theta_1, \ldots, \theta_k$ এমনভাবে নিরূপণ করা হবে যাতে এই বর্গসমষ্টির মান লঘিষ্ঠ হয়। অর্থাৎ, $\theta_0, \theta_1, \ldots, \theta_k$ নিরূপণের জন্য আমরা

$$\frac{\partial}{\partial\theta_j}\sum_{i=1}^{n} e_i^2 = 0$$

$$\text{বা, } 2\sum_{i=1}^{n}(y_i-\theta_0-\theta_1 x_i-\ldots-\theta_k x_i^k)(-x_i^j)=0$$

$$\text{বা, } \sum_{i=1}^{n} e_i x_i^j = 0 \qquad \ldots (1.34)$$

$$\text{বা, } \sum_{i=1}^{n} y_i x_i^j = \theta_0 \sum_{i=1}^{n} x_i^j + \theta_1 \sum_{i=1}^{n} x_i^{j+1} + \ldots + \theta_k \sum_{i=1}^{n} x_i^{j+k},\ j = 0(1)k \ \ldots (1.35)$$

—এই $(k+1)$ টি সমীকরণ যুগপৎ সমাধান করব। এই সমীকরণগুলিকে বলা হয় $\theta_0, \theta_1, \ldots, \theta_k$ নিরূপণের জন্য নর্মাল সমীকরণসমূহ (normal equations)।

$\theta_0, \theta_1, \ldots, \theta_k$ নিরূপণের এই পদ্ধতিটির সপক্ষে যুক্তি হল, বহুঘাতজক সম্পর্কটি এইভাবে নিরূপিত হওয়ার ফলে প্রতিটি e_i 'র মান যে যথাসম্ভব কম হবে, অর্থাৎ প্রদত্ত বিন্দুগুলি যে নিরূপিত সাযুজ্যরেখাটির যথাসম্ভব কাছাকাছি থাকবে, সে সম্বন্ধে নিশ্চিত হওয়া যায়। $\sum_{i=1}^{n} e_i$ কে $\theta_0, \theta_1, \ldots, \theta_k$-এর সম্পর্কে লঘিষ্ঠকরণ করলে এই উদ্দেশ্য সফল হত না, কেননা e_i-এর মানগুলি কোনটি ধনাত্মক, কোনটি ঋণাত্মক হওয়ায় ভ্রান্তিসমষ্টি $\sum_{i=1}^{n} e_i$ -এর মান শূন্য বা শূন্যের কাছাকাছি হলেও আলাদা আলাদা ভাবে e_i-গুলি যে নগণ্য হবে এমন কোন নিশ্চয়তা নেই। ভ্রান্তিগুলি চিহ্ন-নিরপেক্ষ করে $\sum_{i=1}^{n} |e_i|$-কে অবশ্য লঘিষ্ঠকরণ করা হলে এই সমস্যা থাকত না। তবে $\sum_{i=1}^{n}|e_i|$ অপেক্ষা $\sum_{i=1}^{n}|e_i^2|$-কে লঘিষ্ঠকরণ করা গাণিতিকভাব সহজতর বলে লঘিষ্ঠ বর্গসমষ্টি পদ্ধতিটিই ব্যবহার করা হয়।

মন্তব্য (১) : এইভাবে প্রদত্ত রাশিতথ্য কাজে লাগিয়ে নর্ম্যাল সমীকরণসমূহ থেকে পাওয়া $\theta_0, \theta_1, \ldots, \theta_k$-র সংখ্যামানগুলি প্রকৃতপক্ষে এই বিশেষকগুলির নমুনালব্ধ অনুমিত মান। অন্য আর এক প্রস্থ রাশিতথ্য থেকে নর্মাল সমীকরণ গঠন করে সেগুলির সমাধান থেকে বিশেষকগুলির অন্য সংখ্যামান পাওয়া সম্ভব। সুতরাং, নমুনালব্ধ এইসব অনুমিত মান $\theta_0, \theta_1, \ldots, \theta_k$ দ্বারা নির্দেশ করাই যুক্তিযুক্ত। কিন্তু এই পর্যায়ে বিভ্রান্তি এড়ানোর জন্য এইসব নমুনালব্ধ মানকেও $\theta_0, \theta_1, \ldots, \theta_k$ বলা হবে।

মন্তব্য (২) : (x_i, y_i), $i = 1(1)n$, দেওয়া থাকলে $y = \phi(x) = \theta_0 + \theta_1 x + \ldots + \theta_n x^{n-1}$ —এই বহুঘাতজকটি নির্ণয় করার আর একটি পদ্ধতি হল অন্তঃপ্রক্ষেপন পদ্ধতি (method of interpolation)।

পদ্ধতি দু'টির মধ্যে তফাৎ এইরকম : অন্তঃপ্রক্ষেপন পদ্ধতিতে $\theta_0, \theta_1, \ldots, \theta_{n-1}$ এমনভাবে নিরূপণ করা হয় যাতে $y_i = \phi(x_i), i = 1(1)n$, হয়—অর্থাৎ নিরূপিত বহুঘাতজক রেখাটি প্রদত্ত সবকটি বিন্দুর $(x_i, y_i), i = 1(1)n$-এর মধ্য দিয়ে যায়। আর লঘিষ্ঠ বর্গ পদ্ধতিতে $\theta_0, \theta_1, \ldots, \theta_{n-1}$ এমনভাবে নিরূপণ করা হয় যাতে $\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \theta_0 - \theta_1 x_i - \ldots - \theta_{n-1} x_i^{n-1})^2$-এর মান লঘিষ্ঠ হয়। এক্ষেত্রে বলা যায় নিরূপিত বহুঘাতজক রেখাটি প্রদত্ত বিন্দুগুলির কাছাকাছি থাকবে।

৩৮.৩ ঋজুরৈখিক নির্ভরণ (Linear Regression)

বিক্ষেপণ চিত্র থেকে এবং পরবর্তী পর্যায়ে সহগতি গুণাঙ্ক r-এর লব্ধ মান থেকে, যদি দেখা যায় আলোচ্য চল দু'টির মধ্যে একটি সুস্পষ্ট ঋজুরৈখিক সম্পর্ক রয়েছে, সেক্ষেত্রে আমাদের এই নির্ভরণ রেখা নিরূপণ করার ব্যাপারে অগ্রসর হতে হবে।

নির্ভরণ রেখা নিরূপণের বিস্তারিত পদ্ধতি এবং এর বিভিন্ন ধর্ম নিয়ে আলোচনা করা হয়েছে পরবর্তী অনুচ্ছেদগুলিতে।

৩৮.৩.১ y-এর ওপর x-এর নির্ভরণ রেখা

ঋজুরৈখিক সম্পর্কের ক্ষেত্রে স্বতন্ত্র চল x-এর ওপর নির্ভরী চল y-এর নির্ভরণ সমীকরণের প্রতিরূপ হবে

$$y_i = Y_i + e_i, \ i = 1(1)n; \qquad \ldots (৮.36)$$

যেখানে $Y_i = b_0 + b_1 x_i$ হল x-এর ওপর y_i-এর নির্ভরণ অংশ, এবং e_i হল এর ভ্রান্তি অংশ।

$$Y = b_0 + b_1 x \qquad \ldots (1.37)$$

—এই নির্ভরণ সরলরেখাটি পাওয়ার জন্য নর্মাল সমীকরণ (1.34) থেকে

$$\sum_{i=1}^{n} e_i = 0, \sum_{i=1}^{n} e_i x_i = 0 \qquad \ldots (1.38)$$

অথবা (1.35) থেকে

$$\sum_{i=1}^{n} y_i x_i^0 = b_0 \sum_{i=1}^{n} x_i^0 + b_1 \sum_{i=1}^{n} x_i$$

ও $$\sum_{i=1}^{n} y_i x_i = b_0 \sum_{i=1}^{n} x_i + b_1 \sum_{i=1}^{n} x_i^2$$

বা, $$\sum_{i=1}^{n} y_i = nb_0 + b_1 \sum_{i=1}^{n} x_i \qquad \dots (1.39)$$

ও $$\sum_{i=1}^{n} y_i x_i = b_0 \sum_{i=1}^{n} x_i + b_1 \sum_{i=1}^{n} x_i^2 \qquad \dots (1.40)$$

$(1.39) \div n$ থেকে পাওয়া যায়

$$\overline{y} = b_0 + b_1 \overline{x} \qquad \dots (1.41)$$

এবং $(1.40) \div n - \overline{x} \times (1.41)$ থেকে

$$\frac{1}{n} \sum_{i=1}^{n} x_i y_i - \overline{x}\,\overline{y} = b_1 \left[\frac{1}{n} \sum_{i=1}^{n} x_i^2 - \overline{x}^2 \right]$$

বা, $$\frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y}) = b_1 . \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2$$

বা, $$b_1 = \frac{\frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2} = \frac{\text{Cov}(x, y)}{v(x)}$$

$$= \frac{r S_y S_x}{S_x^2} = r . \frac{S_y}{S_x} = b_{yx} \qquad \dots (1.42)$$

যেখানে r হল x ও y-এর সহগতি গুণাঙ্ক এবং S_x^2 ও S_y^2 যথাক্রমে x ও y-এর ভেদমান।

সুতরাং (1.41) থেকে,

$$b_0 = \overline{y} - r \frac{S_y}{S_x} \overline{x} \qquad \dots (1.43)$$

অর্থাৎ, x-এর ওপর y-এর লঘিষ্ঠ বর্গসমষ্টি নির্ভরণ রেখা, বা সংক্ষেপে নির্ভরণ রেখাটি হবে

$$Y = \overline{y} + r\frac{S_y}{S_x}(x - \overline{x}) \qquad \ldots (1.44)$$

এই রেখাটিকে x-এর প্রদত্ত মানসাপেক্ষে y-এর প্রত্যাশিত মান পাওয়ার জন্য, অর্থাৎ y-এর পূর্বানুমান সূত্র হিসাবে ব্যবহার করা যেতে পারে। এই সূত্রে ব্যবহৃত স্বতন্ত্র চল x-কে আমরা নির্ভরক (regressor) চল বলব।

এখানে $b_1 \equiv b_{yx} = r\frac{S_y}{S_x}$ গুণাঙ্কটিকে বলা হয় x-এর ওপর y-এর নির্ভরণ গুণাঙ্ক (regression coefficient), কেননা x-এর মান 1 একক বৃদ্ধি পেলে y-এর প্রত্যাশিত মানে বৃদ্ধির পরিমাণ b_{yx}। b_{yx}-এর লব ও হর উভয় পক্ষকে n^2 দিয়ে গুণ করে পাওয়া যায়

$$b_{xy} = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \qquad \ldots (1.45)$$

প্রদত্ত রাশিতথ্য থেকে b_{yx} নির্ণয় করার জন্য (1.45) রূপটি খুব সুবিধাজনক।

উদা. ২.১ : উদাহরণ 1.1-এ প্রদত্ত তথ্য ব্যবহার করে উন্নয়নশীল দেশের মাথাপিছু জাতীয় উৎপাদন (1995)-এর ওপর সংশ্লিষ্ট দেশের ঐ বছরের মানববিকাশ সূচকের একটি নির্ভরণ রেখা নিরূপণ কর।

এই রেখাটি ব্যবহার করে 1995-এ যে দেশের মাথাপিছু জাতীয় উৎপাদন ছিল 400 আমেরিকান ডলার, সেই দেশটির ঐ বছরের মানববিকাশ সূচক সম্বন্ধে আভাস দাও।

এখানে আগের মত মাথাপিছু জাতীয় উৎপাদন চলটিকে x এবং মানববিকাশ সূচক চলটিকে y বলা হলে (1.45) থেকে পাওয়া যাবে

$$b_1 = b_{yx} = \frac{9 \times 988.132 - (2430)(3.502)}{9 \times 756682 - (2430)^2} = 0.00042;$$

তাছাড়া $\overline{y} = 0.3891$, $\overline{x} = 270$; সুতরাং,

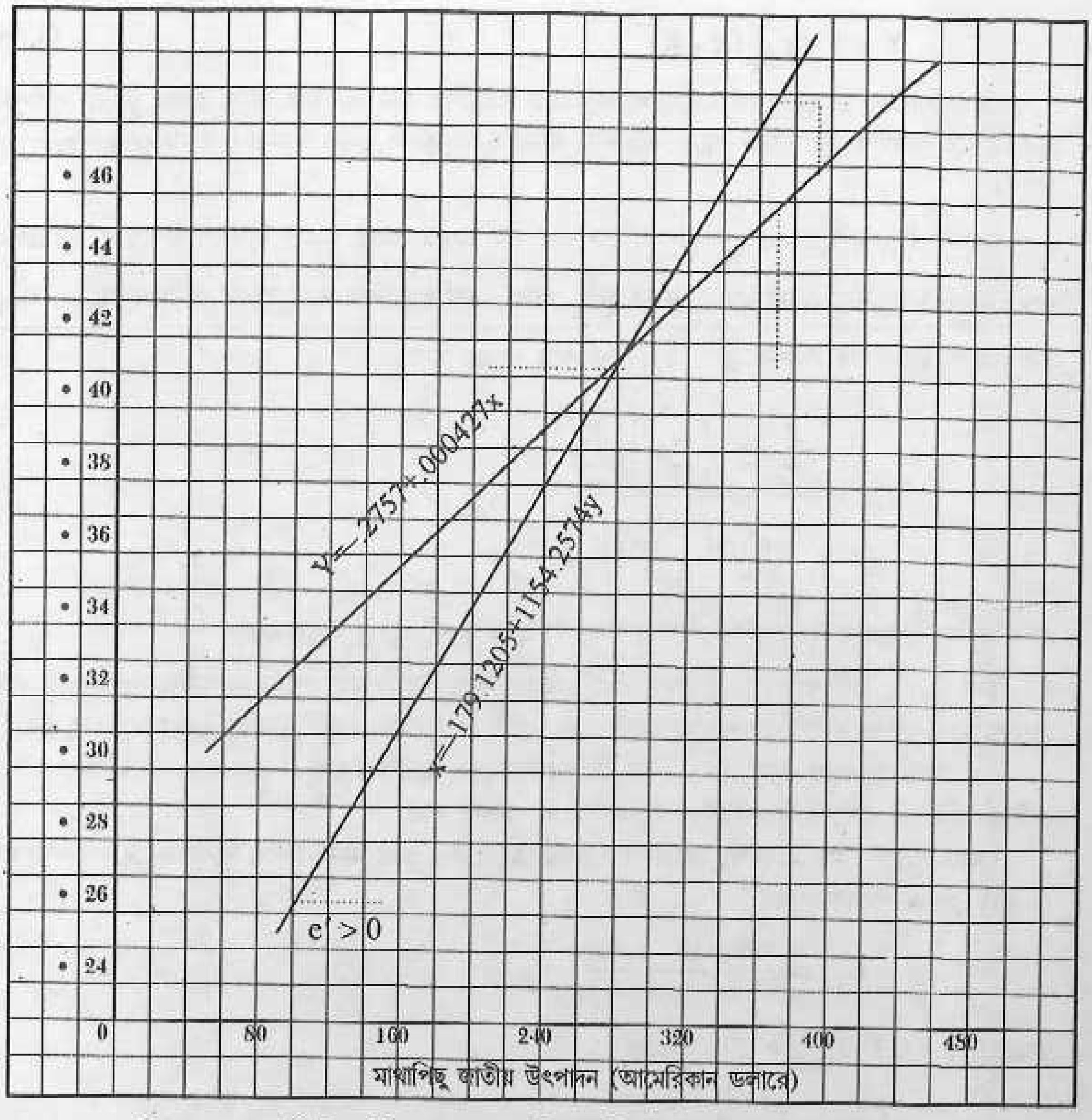
$$b_0 = 0.3891 - 0.00042 \times 270 = 0.2757 \text{ ।}$$

সুতরাং, উদ্দিষ্ট নির্ভরণ রেখাটি হবে

$$Y = 0.2757 + 0.00042x \text{ ।}$$

ফলে যে দেশের মাথাপিছু জাতীয় উৎপাদন 1995 সালে 400 ডলার ছিল, তার ঐ বছরের আনুমানিক মানববিকাশ সূচক হবে $Y = 0.2757 + 0.00042 \times 400 = 0.4437$।

চিত্র 2.3-এ আলোচ্য উদাহরণের বিক্ষেপণ চিত্র এবং x-এর ওপর y-এর নির্ভরণ রেখাটি দেখানো হয়েছে।



চিত্র ৩৮.১ কয়েকটি উন্নয়নশীল দেশের মাথাপিছু জাতীয় উৎপাদন (x) ও মানববিকাশ সূচক (y) সংক্রান্ত রাশিতথ্যের বিক্ষেপণ চিত্র ও দু'টি নির্ভরণ রেখা।

৩৮·৩·২ x-এর ওপর y-এর নির্ভরণ রেখা

লঘিষ্ঠ বর্গপদ্ধতিতে নিরূপিত $Y = b_0 + b_1 x$ নির্ভরণ সমীকরণটি যেহেতু চল দু'টির মধ্যে একটি ঋজুরৈখিক সম্পর্ক নির্দেশ করে। সুতরাং প্রশ্ন উঠতে পারে, এই রেখাটি শুধুমাত্র x দেওয়া থাকলে y-এর অনুমিত মান নির্ণয়ের জন্য কেন ব্যবহৃত হবে? উল্টোটিও তো হতে পারে, অর্থাৎ y দেওয়া থাকলে x-এর অনুমিত মান পাওয়ার জন্যও তো এই একই নির্ভরণ রেখা ব্যবহার করা যেতে পারে।

এই প্রশ্নের উত্তর হল, সাধারণত তা করা যাবে না—দু'টি চলের জন্য দু'টি পৃথক নির্ভরণ রেখা প্রয়োজন। y যদি নির্ভরী চল হয়, তাহলে আমরা ধরে নেব, y-মাপনায় খানিকটা ভ্রান্তি থাকবে, কিন্তু x মাপা যাবে ভ্রান্তিশূন্যভাবে। এখন x-কে y-এর ওপর নির্ভরী চল ভাবা হলে উল্টোটি ধরে নিতে হবে, অর্থাৎ x-মাপনায় ভ্রান্তি থাকবে, কিন্তু y মাপা যাবে ভ্রান্তিশূন্যভাবে। সেক্ষেত্রে গৃহীত প্রতিরূপটি হবে

$$x_i = X_i + e_i' \qquad \text{... (1.36')}$$

যেখানে $X_i = b_0' + b_1' y_i$ হলে y-এর ওপর x_i-এর নির্ভরণ অংশ এবং e_i' হল ভ্রান্তি অংশ।

$$X = b_0' + b_1' y \qquad \text{... (1.37')}$$

—এই নির্ভরণ সরলরেখাটি পাওয়ার জন্য নর্মাল সমীকরণ হবে

$$\frac{\partial}{\partial b_0'} \sum_{i=1}^{n} e_i'^2 = 0, \text{ বা, } \frac{\partial}{\partial b_0'} \sum_{i=1}^{n} (x_2 - b_0' - b_1' y_i)^2$$

এবং $\frac{\partial}{\partial b_1'} \sum_{i=1}^{n} e_i'^2 = 0$, বা, $\frac{\partial}{\partial b_1'} \sum_{i=1}^{n} (x_i - b_0' - b_1' y_i)^2$ ।

অর্থাৎ নর্মাল সমীকরণ দু'টি দাঁড়াবে

$$\sum_{i=1}^{n} e_i' = 0, \sum_{i=1}^{n} e_i' y_i = 0 \qquad \text{... (1.38')}$$

অথবা আরও বিস্তারিতভাবে

$$\sum_{i=1}^{n} x_i = nb_0' + b_1' \sum_{i=1}^{n} y^i \qquad \text{... (1.39')}$$

এবং $$\sum_{i=1}^{n} x_i y_i = b_0' + \sum_{i=1}^{n} y_i + b_1' \sum_{i=1}^{n} y_i^2 \qquad \text{.... (1.40')}$$

এ দু'টি সমীকরণ আগের মতই সমাধান করে পাওয়া যায়

$$b_1' = b_{xy} = \frac{\text{Cov}(x, y)}{\text{ব}(y)} = r \frac{S_x}{S_y} \qquad \text{... (1.42')}$$

($b_1' = b_{xy}$ হল y-এর ওপর x-এর নির্ভরণ গুণাঙ্ক)

এবং $(b_0' = \overline{x} - b_1'\overline{y} = \overline{x} - r\frac{S_x}{S_y}\overline{y}$ ... (1.43')

অর্থাৎ y-এর ওপর x-এর লঘিষ্ঠ বর্গ নির্ভরণ রেখাটি হবে

$$X = \overline{x} + r\frac{S_x}{S_y}(y - \overline{y}) \quad \text{... (1.44')}$$

(1.45)-এর অনুরূপ b_{yx} -এর সহজে ব্যবহারযোগ্য প্রকাশনটি হল

$$b_1' = b_{xy} = \frac{n\sum_{i=1}^{n} x_i y_i - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2} \quad \text{... (1.45')}$$

মন্তব্য : উপরের আলোচনা থেকে পরিষ্কার, যে চলটি ভ্রান্তিশূন্যভাবে মাপা যায়, তার ওপর নির্ভরী চলের নির্ভরণ রেখা নির্ণয় করা যুক্তিযুক্ত। কিন্তু বাস্তবে দেখা যায়, সাধারণত নির্ভরী, অনপেক্ষ কোন চলই ভ্রান্তিশূন্যভাবে মাপা যায় না। ব্যতিক্রম অবশ্য আছে, যেমন সময়-নির্দেশী চল (যেমন, সাল) অথবা যেসব চলের ওপর নিয়ন্ত্রণ রাখা যায় (যেমন কৃষিজমিতে প্রদত্ত একর প্রতি সারের পরিমাণ)। দু'টি চলের মাপনাই ভ্রান্তিসাপেক্ষ হলে নির্ভরণরেখা নির্ণয় করার পদ্ধতি বেশ জটিল।

বাস্তব প্রয়োগের ক্ষেত্রে দু'টি চলের মধ্যে যেটি সহজে অথবা আগেভাগে নির্ণয় করা যায়, সেটিকে অনপেক্ষ চল এবং অন্যটিকে নির্ভরী চল ধরে নিয়ে সংশ্লিষ্ট নির্ভরণ রেখাটি নিরূপণ করা হয়।

উদা. ২.২ : উদাহরণ ১.১-এ মানববিকাশ সূচক চলটিকে অনপেক্ষ এবং মাথাপিছু জাতীয় উৎপাদন চলটিকে নির্ভরী ধরা হলে আমরা পাব

$$b_1' = b_{xy} = \frac{9 \times 988.132 - 2430 \times 3.502}{9 \times 1.3994 - (3.502)^2} = 1154.2547$$

এবং $b_0' = 270 - 1154.2547 \times 0.3891 = -179.1205$

অর্থাৎ, নির্ভরণ রেখাটি হবে $X = -179.1205 + 1154.2547y$।

যে দেশের মানববিকাশ সূচক 0.400, তার মাথাপিছু জাতীয় উৎপাদনের অনুমিত মান এই নির্ভরণ রেখা থেকে পাওয়া যাবে $-179.1205 + 1154.2547 \times 0.400 = 282.6$ (ডলার)।

এই নির্ভরণ রেখাটিও বিক্ষেপণ চিত্রের ওপর দেখানো হয়েছে চিত্র 1.3-এ।

৩৮.১ চিত্রে লক্ষ্য করলে দেখা যাবে, x-এর ওপর y-এর নির্ভরণ রেখায় 5টি বিন্দুর জন্য ভ্রান্তি ধনাত্মক এবং 4টি বিন্দুর জন্য ঋণাত্মক। আর y-এর ওপর x-এর নির্ভরণ রেখাতেও 4টি বিন্দুর জন্য ভ্রান্তি ধনাত্মক এবং বাকি 5টির জন্য ঋণাত্মক।

৩৮·৩·৩ নির্ভরণ রেখা প্রদত্ত আভাসিত মানের স্বরূপ

নির্ভরণ রেখা থেকে পাওয়া নির্ভরী চলের অনুমিত মান সম্পর্কে একটা কথা বলা দরকার। ধরা যাক, আমরা $Y = b_0 + b_1 x$ নির্ভরণ রেখাটি ব্যবহার করে যে ব্যষ্টির জন্য x-এর মান x_0, তার জন্য y-এর অনুমিত মান পেয়েছি $Y_0 = b_0 + b_1 x_0$। প্রশ্ন হল, ঐ ব্যষ্টির জন্য y-এর প্রকৃত মান কি Y_0 হবেই? এই প্রশ্নের উত্তর হল—না, অভ্রান্তভাবে Y_0 নাও হতে পারে, তবে এর কাছাকাছি থাকবে। আসলে নির্ভরণ রেখা যা দেয় তা হল, অনপেক্ষ চলের একটি প্রদত্ত মানের জন্য নির্ভরী চলের শর্তাধীন প্রত্যাশিত মান বা গড়। অর্থাৎ, বেশকিছু ব্যষ্টির ক্ষেত্রে x-এর মান x_0 হলে, তাদের জন্য y-এর মান হবে গড়ে Y_0।

পরে আমরা দেখব, পূর্বানুমান সূত্র হিসাবে নির্ভরণ রেখার কার্যকারিতা নির্ভর করবে দুটি চ'লের মধ্যে সহগতি গুণাঙ্কের সংখ্যামান $|r_{xy}|$-এর ওপর।

৩৮·৩·৪ দু'টি নির্ভরণ রেখার অন্তর্বর্তী কোণ

তাহলে দেখা যাচ্ছে, এক প্রস্থ রাশিতথ্য $(x_i, y_i), i = 1(1)n$ ব্যবহার করে লঘিষ্ঠ বর্গসমষ্টি পদ্ধতিতে পাওয়া নির্ভরণ রেখা দু'টি হবে :

x-এর ওপর y-এর : $Y = \overline{y} + r\frac{S_y}{S_x}(x - \overline{x})$ ... (1.44)

এবং y-এর ওপর x-এর : $X = \overline{x} + r\frac{Sx}{S_y}(y - \overline{y})$ ... (1.44')

(1.38) এবং (1.38') থেকে পাওয়া যাচ্ছে

$$\sum_{i=1}^{n} e_i = 0 \text{ এবং } \sum_{i=1}^{n} e'_i = 0, \text{ অর্থাৎ } \overline{e} = 0 \text{ এবং } \overline{e'} = 0 \quad \text{... (1.46)}$$

সুতরাং, (1.36) ও (1.36') থেকে দেখা যাবে

$$\overline{y} = \overline{Y} \text{ এবং } \overline{x} = \overline{X} \quad \text{... (1.47)}$$

অর্থাৎ, যে কোন চলের অবেক্ষিত মানগুলির গড়, ঐ চলের নির্ভরণ রেখা থেকে পাওয়া সংশ্লিষ্ট আভাসিত মানগুলির গড়ের সমান।

স্পষ্টতই $(\bar{x}, \bar{y})$–বিন্দুটি দু'টি নির্ভরণ রেখাকেই সিদ্ধ করে, অর্থাৎ নির্ভরণ রেখা দু'টি $(\bar{x}, \bar{y})$-বিন্দুতে পরস্পর ছেদ করে (চিত্র P.3 দ্রষ্টব্য)।

এখন নির্ভরণ রেখা দু'টিকে $y = mx + c$—এই আকারে লেখা হলে

(1.44) থেকে $Y = m_{yx}x + c_{yx}$

বা, $Y = r\frac{S_y}{S_x}.x + (\bar{y} - r\frac{S_y}{S_x}.\bar{x})$।

এবং (1.44′) থেকে $y = m_{xy}x + c_{xy}$

বা, $y = \frac{S_y}{rS_x}x + (\bar{y} - r\frac{S_y}{rS_x}\bar{x})$।

$$\text{অর্থাৎ, } m_{yx} = r\frac{S_y}{S_x} = \tan\theta \qquad \dots (1.48)$$

$$\text{এবং } m_{xy} = \frac{S_y}{rS_x} = \tan\theta' \qquad \dots (1.48')$$

যেখানে θ এবং θ' হল (1.44) ও (1.44′) রেখা দু'টির সঙ্গে x-অক্ষের কোণের পরিমাণ। নির্ভরণরেখা দু'টির মধ্যবর্তী কোণটি θ হলে স্পষ্টতই

$$\tan\varphi = \tan(\theta \sim \theta') = \frac{\tan\theta \sim \tan\theta'}{1 + \tan\theta\,\tan\theta'}$$

$$= \frac{r\frac{S_y}{S_x} \sim \frac{S_y}{rS_x}}{1 + r\frac{S_y}{S_x}.\frac{S_y}{rS_x}} = \frac{(1 \sim r^2)}{r}.\frac{S_yS_x}{S_y^2 + S_x^2} \qquad \dots (1.49)$$

দু'টি রেখা পরস্পর ছেদ করলে দু'টি কোণ উৎপন্ন হয়, এদের মধ্যে একটি সূক্ষ্মকোণ, অন্যটি স্থূলকোণ; অথবা, দু'টিই সমকোণ। সূক্ষ্মকোণটিকে φ ধরলে

$$\tan\varphi = \frac{1 - r^2}{r}.\frac{S_yS_x}{S_y^2 + S_x^2},\ r > 0$$

এবং স্থূলকোণটিকে φ ধরলে

$$\tan\varphi = \frac{r^2 - 1}{r}.\frac{S_yS_x}{S_y^2 + S_x^2},\ r > 0\text{।}$$

তবে $r < 0$ হলে $\tan\varphi$-এর এই প্রকাশন দু'টিকে বদলাবদলি করে নিতে হবে।

(1.49) থেকে দেখা যাচ্ছে, $r = 0$ হলে $\tan\varphi = \mp\infty$, বা $\varphi = \mp\frac{\pi}{2}$, যার অর্থ, নির্ভরণ রেখা দু'টি পরস্পরকে সমকোণে ছেদ করবে। লক্ষ্যণীয়, সেক্ষেত্রে নির্ভরণ রেখা দু'টি হবে $Y = \bar{y}$ এবং $X = \bar{x}$, অর্থাৎ এগুলি অনপেক্ষ চলের ওপর আদৌ নির্ভরশীল হবে না এবং তার ফলে এগুলি পূর্বানুমান সূত্র হিসাবে পুরোপুরি ব্যর্থ বলে গণ্য হবে।

অন্যদিকে $r = \mp 1$ হলে $\tan\varphi = 0$, বা $\varphi = 0$। এর অর্থ, রেখাটি হয় অভিন্ন, নয়তো পরস্পর সমান্তরাল হবে। তবে যেহেতু $(\bar{x}, \bar{y})$ -বিন্দুটি দু'টি রেখাকেই সিদ্ধ করে, এরা সমান্তরাল হতে পারে না—সুতরাং অভিন্ন হবে। অর্থাৎ $r = \mp 1$ হলে চল দু'টির মধ্যে থাকে নিখুঁত ঋজুরৈখিক সম্পর্ক, এবং সেক্ষেত্রে অভিন্ন নির্ভরণ রেখাটি ব্যবহার করে যে কোন একটি চলের মান জানা থাকলে অন্য চলটির নির্ভুল মান পাওয়া যায়।

সুতরাং দেখা যাচ্ছে, পূর্বানুমান সূত্র হিসাবে নির্ভরণ রেখার কার্যকারিতা $|r|$-এর ওপর নির্ভরশীল। এবং $r = \mp 1$ হলে যে কোন একটি চলের নির্ভুল পূর্বানুমান সূত্র হিসাবে একটি নির্ভরণরেখাই যথেষ্ট। r-এর অন্য মানের জন্য আমাদের অবশ্য দু'টি পৃথক নির্ভরণ রেখা নিরূপণ করতে হবে।

৩৮·৩·৫ গোষ্ঠীবদ্ধ রাশিতথ্যের ক্ষেত্রে নির্ভরণ রেখা

১.৪ অনুচ্ছেদে আলোচিত গোষ্ঠীবদ্ধ রাশিতথ্য $(x_i, y_i, f_{ij}), i = 1(1)k, j = 1(1)l$ – এর ক্ষেত্রে (1.10), (1.11), (1.12), (1.15), (1.16) ও (1.17)-এ প্রদত্ত সংকেত চিহ্ন ব্যবহার করে আমরা নির্ভরণ গুণাঙ্ক দু'টির অন্য সূত্র পাব।

$$b_1 = b_{yx} = \frac{\mathrm{Cov}(x, y)}{v(x)} = \frac{n\sum_{i=1}^{k} x_i Y_i - \left(\sum_{i=1}^{k} Y_i\right)\left(\sum_{j=1}^{l} X_j\right)}{n\sum_{i=1}^{k} x_i^2 f_{i0} - \left(\sum_{i=1}^{k} x_i f_{i0}\right)^2} \quad \dots (1.50)$$

$$\text{এবং } b_1' = b_{xy} = \frac{\mathrm{Cov}(x, y)}{v(y)} = \frac{n\sum_{i=1}^{k} x_i Y_i - \left(\sum_{i=1}^{k} Y_i\right)\left(\sum_{j=1}^{l} X_j\right)}{n\sum_{j=1}^{l} y_j^2 f_{0j} - \left(\sum_{j=1}^{l} y_j f_{0j}\right)^2} \quad \dots (1.50')$$

আগের মতই $b_0 = \bar{y} - b_1\bar{x}$, $b_0' = \bar{x} - b_1'\bar{y}$।

$u_i = \frac{x_i - a}{b}$, $v_j \frac{y_j - c}{d}$—এই রূপান্তরের ক্ষেত্রে $b_{yx} = \frac{d}{b} b_{vu}$ এবং $b_{xy} = \frac{b}{d} b_{uv}$ হবে (অনুচ্ছেদ 1.8.6 দেখ)।

উদা. ২.৩ : ১.৭ উদাহরণে প্রদত্ত গোষ্ঠীবদ্ধ রাশিতথ্যের ক্ষেত্রে নির্ভরণ রেখা দু'টি নির্ণয় কর।

যে পরিবারের মাসিক আয় 5,686 টাকা, সেখানে মাসিক আয়ের শতকরা আনুমানিক কত অংশ খাদ্যের জন্য ব্যয়িত হয়? যে পরিবারে মাসিক আয়ের 38% খাদ্য খাতে ব্যয়িত হয়, তার আনুমানিক মাসিক আয় কত?

মাসিক আয় চলটিকে x এবং খাদ্য খাতে ব্যয়িত মাসিক আয়ের শতকরা অংশ চলটিকে y বলা হলে আমরা উদা ১.৪ থেকে পাচ্ছি $\left(u_i = \frac{x_i - 4.4}{0.8},\ v_j = \frac{y_j - 32.5}{5}\right)$

$$b_{vu} = \frac{250\times(-744)-(-360)(116)}{250\times1624-(-360)^2} = -0.5219$$

$$b_{uv} = \frac{250\times(-744)-(-360)(116)}{250\times484-116^2} = -1.3412$$

সুতরাং, $b_{yx} = \frac{5}{0.8}\times(-0.5219) = -3.2619$ এবং $b_{xy} = \frac{0.8}{5}\times(-1.3412) = -0.2146$

তাছাড়া, $\bar{x} = 4.4 + 0.8\times(-\frac{360}{250}) = 3.248$ এবং $\bar{y} = 32.5 + 5\times\frac{116}{250} = 34.82$ ।

অতএব, নির্ভরণ রেখা দু'টি হল

$$Y = (34.82 + 3.2619\times3.248) - 3.2619x$$
$$= 45.4147 - 3.2619x$$

এবং $X = (3.248 + 0.2146\times34.82) - 0.2146y$

$$= 10.7204 - 0.2146y$$ ।

সুতরাং, যে পরিবারের মাসিক আয় 5,685 টাকা, সেখানে খাদ্য খাতে ব্যয়িত মাসিক আয়ের শতকরা অনুমিত অংশ হবে 45.4147 − 3.2619 × 5.685 = 26.87; আবার যে পরিবারে খাদ্য খাতে মাসিক আয়ের 38% ব্যয়িত হয়, তার অনুমিত মাসিক আয় =10.7204 − 0.2146 × 38 = 2.5656 বা 2,565.60 টাকা।

৩৮.৩.৬ নির্ভরণ সংক্রান্ত কয়েকটি ফল

যে কোন একটি লঘিষ্ঠ-বর্গ সরল নির্ভরণ রেখার—ধরা যাক x-এর উপর y-এর কয়েকটি ধর্ম নিচে আলোচনা করা হল। অন্য নির্ভরণ রেখাটির ধর্মগুলিও একইরকম হবে।

(১) আগেই দেখানো হয়েছে $\bar{e} = 0$, এবং সেই কারণে $\bar{Y} = \bar{y}$ ।

(২) x ও y-এর মূলবিন্দু এবং মাপনা-একক পরিবর্তন করে যদি লেখা হয়

$$u_i = \frac{x_i - a}{b} \text{ এবং } v_i = \frac{y_i - c}{d},\ i = 1(1)n$$

তাহলে, $b_{yx} = \frac{Cov(x,y)}{v(x)} = \frac{bd\,Cov(u,v)}{b^2 v(u)} = \frac{d}{b}.b_{vu}$

অর্থাৎ, $b_{yx} = \frac{d}{b}.\frac{n\sum_{i=1}^{n} u_i v_i - \left(\sum_{1}^{n} u_i\right)\left(\sum_{1}^{n} v_i\right)}{n\sum_{i=1}^{n} {u_i}^2 - \left(\sum_{i=1}^{n} u_i\right)^2}$

এবং $\overline{y} = c + d\overline{v},\ \overline{x} = a + b\overline{u}$ ।

(৩) $v(Y) = S_y^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \overline{Y})^2$

$$= \frac{1}{n}\sum_{i=1}^{n}(Y_i - \overline{y})^2, \text{ যেহেতু } \overline{y} = \overline{Y}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left\{r\frac{S_y}{S_x}(x_i - \overline{x})\right\}^2$$

$$= r^2.\frac{S_y^2}{S_x^2}.S_x^2 = r^2 S_y^2$$

$$\Rightarrow \frac{S_Y^2}{S_y^2} = r^2 \Rightarrow |r| = \frac{S_Y}{S_y} \quad \text{... (1.51)}$$

(৪) $Cov(Y,e) = \frac{1}{n}\sum_{i-1}^{n} e_i(Y_i - \overline{Y}) = \frac{1}{n}.\sum_{i=1}^{n} e_i(Y_i - \overline{y})$

$$= \frac{1}{n}\sum_{i=1}^{n} e_i\{r\frac{S_y}{S_x}(x_i - \overline{x})\}$$

$$=\frac{1}{n}\cdot\frac{rS_y}{S_x}\left[\sum_{i=1}^{n}e_ix_i-\bar{x}\sum_{i=1}^{n}e_i\right]$$

=0, নর্মাল সমীকরণ দু'টি থেকে।

সুতরাং দেখা যাচ্ছে, নির্ভরী চলের নির্ভরণ-অংশ এবং ভ্রান্তি-অংশ পরস্পর সহগতি শূন্য।

(৫) $y_i = Y_i + e_i,\ i = 1(1)n$

$\Rightarrow S_y^2 = S_Y^2 + S_e^2$, যেহেতু Y ও e পরস্পর সহগতিশূন্য।

$\Rightarrow S_y^2 = r^2S_y^2 + S_e^2$

$\Rightarrow S_e^2 = (1-r^2)S_y^2 =$ ভ্রান্তি ভেদমান। ... (1.52)

এখন $S_e^2 \geq 0 \Rightarrow 1 \geq r^2 \Rightarrow -1 \leq r \leq 1$, যে ফলটি আমরা আগেই অন্যভাবে প্রমাণ করেছি।

আবার $r = \mp 1$ হলে $S_e^2 = 0$, অর্থাৎ প্রতিটি e_i-এর মান ধ্রুবক হবে। এই সঙ্গে $\bar{e} = 0$ ফলটি মনে রাখলে দেখা যাবে, এই ধ্রুবকের মান হবে শূন্য। এর অর্থ, $r = \mp 1$ হলে, প্রতিটি e_i-মান শূন্য হবে, অর্থাৎ নির্ভরণ রেখাটি হবে একটি নিখুঁত পূর্বানুমান সূত্র। পক্ষান্তরে, $r = 0$ হলে $S_e^2 = S_y^2$ হবে, অর্থাৎ, ভ্রান্তি ভেদমান হবে নির্ভরী চলটির সামগ্রিক ভেদমানের সমান, যার অর্থ নির্ভরণ রেখাটি নির্ভরী চলের চলমানতার সামান্যতম অংশও ব্যাখ্যা করতে সমর্থ হবে না। এটি অন্যভাবেও প্রতীয়মান, কেননা $r = 0$ হলে $Y_i = \bar{y}\ \forall i$, অর্থাৎ নির্ভরণ রেখাটি x-এর প্রদত্ত মানের জন্য y-এর ওপর আদৌ কোন আলোকপাত করতে সক্ষম হবে না, এবং এজন্য এটি পূর্বানুমান সূত্র হিসাবে হবে পুরোপুরি ব্যর্থ।

সুতরাং দেখা যাচ্ছে, মূল ভেদমানের কতখানি অংশ নির্ভরণ রেখা দ্বারা ব্যাখাত হচ্ছে, অর্থাৎ পূর্বানুমান সূত্র হিসাবে নির্ভরণ রেখা কতখানি কার্যকরী সে ধারণা পাওয়া যায় r^2-এর মান থেকে। r^2-কে তাই নির্ধারণ গুণাঙ্ক (Coefficient of determination) বলা হয়।

(৬) $\mathrm{Cov}(y, Y) = \frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})(Y_i-\bar{Y})$

$$=\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})\left\{r\frac{S_y}{S_x}(x_i-\bar{x})\right\},\ \text{যেহেতু}\ \bar{Y}=\bar{y}$$

$$= r.\frac{S_y}{S_x}.\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x}) = r\frac{S_y}{S_x}.r.S_yS_x = r^2S_y^2 \qquad \dots (1.53)$$

অর্থাৎ, $r_{yY} = \frac{Cov(y, Y)}{\sqrt{v(y)}\sqrt{v(Y)}} = \frac{r^2S_y^2}{S_y|r|S_y} = |r| \qquad \dots (1.54)$

সুতরাং দেখা যাচ্ছে, y-এর অবেক্ষিত মান এবং নির্ভরণরেখা লব্ধ মানের মধ্যে সহগতি গুণাঙ্ক সবসময় ধনাত্মক, এবং এটি x ও y-এর সহগতি গুণাঙ্কের পরমমানের সমান। অর্থাৎ, অবেক্ষিত মান এবং নির্ভরণ রেখা লব্ধ মানের মধ্যে কতখানি সাযুজ্য আছে, তা বোঝা যাবে r-এর পরমমান থেকে। $|r|$-এর মান 1-এর যত কাছাকাছি হবে, এই দু'সারি মানের মধ্যে সাযুজ্য হবে তত ভাল, অর্থাৎ পূর্বানুমান সূত্র হিসাবে নির্ভরণ রেখা তত কার্যকরী হবে। এ ব্যাপারটি আমরা ভ্রান্তি ভেদমানের প্রকৃতি থেকেও লক্ষ্য করেছি।

(৭) $b_{yx} = r\frac{Sy}{S_x}$ এবং $b_{xy} = r\frac{S_x}{S_y}$

যেহেতু S_x এবং S_y ধনাত্মক, b_{yx} ও b_{xy}-এর গাণিতিক চিহ্ন r-এর গাণিতিক চিহ্নের সমান হবে এবং ফলে নির্ভরণ গুণাঙ্কটি অভিন্ন গাণিতিক চিহ্ন সম্পন্ন হবে।

দেখা যাচ্ছে, $b_{yx}.b_{xy} = r^2$

$$\Rightarrow r = \pm\sqrt{b_{yx}}b_{xy} \qquad \dots (1.55)$$

অর্থাৎ, সহগতি গুণাঙ্কের সাংখ্যমান হবে দু'টি নির্ভরণ গুণাঙ্কের জ্যামিতিক গড়ের সমান, এবং দু'টি নির্ভরণ গুণাঙ্কের অভিন্ন গাণিতিক চিহ্নই হবে r-এর গাণিতিক চিহ্ন।

৩৮·4 বিভিন্ন ধরনের সাযুজ্যরেখা নিরূপণ

আমরা 1.7 অনুচ্ছেদে লঘিষ্ঠ বর্গসমষ্টি পদ্ধতি নিয়ে আলোচনা করেছি। নির্ভরী চলটিকে অনপেক্ষ চলের একটি ঘাতজক রূপে প্রকাশ করা হলে এই লঘিষ্ঠ বর্গসমষ্টি পদ্ধতির সাহায্যে সংশ্লিষ্ট সাযুজ্যরেখাটি নিরূপণ করা যায়।

চল দু'টির মধ্যে কোন ঘাতজক সম্পর্ক না থাকলেও অনেক সময় কিছু রূপান্তরের সাহায্যে সম্পর্কটিকে ঘাতজকে পরিণত করা যায়, এবং তারপর সেটি নিরূপণের জন্য সাহায্য নেওয়া যায় লঘিষ্ঠ বর্গসমষ্টি পদ্ধতির।

এইরকম কয়েকটি ক্ষেত্রে সাযুজ্যরেখা নিরূপণের পদ্ধতি নিয়ে আলোচনা করা হল নিচে।

সাধারণভাবে x-এর ওপর y-এর সম্পর্কসূচক সাযুজ্যরেখাটিকে $y=\phi(x;\theta_1,\theta_2,\ldots,\theta_k)$ দ্বারা নির্দেশ করে আমরা আলোচনা করব। প্রদত্ত n-জোড়া মান (x_i, y_i), $i=1(1)n$, কাজে লাগিয়ে কীভাবে $\theta_1,\theta_2,\ldots,\theta_k$ —এই ধ্রুবকগুলির যুক্তিসঙ্গত মান নির্ণয় করে সাযুজ্য রেখাটি নিরূপণ করা যাবে।

৩৮·৪·১ লগ-রূপান্তরের সাহায্যে সরলরেখা

ধরা যাক, $y=\phi(x;a,b)=a.b^x$ ... (1.56)

এখানে $\log y=\log a+x\log b$ (লগ-এর নিদান 10 বা e যা ইচ্ছা হতে পারে)

বা, $y'=a'+b'x$

অর্থাৎ দেখা যাচ্ছে, এখানে $y'=\log y$ এবং x-এর মধ্যে একটি ঋজুরৈখিক সম্পর্ক রয়েছে। সুতরাং, (1.56)-এ প্রদত্ত সাযুজ্যরেখাটি নিরূপণের জন্য আমরা লঘিষ্ঠ বর্গসমষ্টি পদ্ধতিতে $(x_i, y_i'=\log y_i)$, $i=1(1)n$, এই মানগুলি কাজে লাগিয়ে প্রথমে $a'=\log a$, ও $b'=\log b$, এবং তা থেকে পরে a ও b নির্ণয় করব।

তেমনি সম্পর্কটি যদি হয় $y=\phi(x;a,b)=a.x^b$ ... (1.57)

তাহলে $\log y=\log a+b\log x$

বা, $y'=a'+bx'$, $a'=\log a$,

যার অর্থ, এখানে $y'=\log y$ এবং $x'=\log x$ —এই রূপান্তরিত চল দু'টির মধ্যে ঋজুরৈখিক সম্পর্ক রয়েছে। সুতরাং, আগের মত এক্ষেত্রেও $(x_i'=\log x_i,\ y_i'=\log y_i')$, $i=1(1)n$ – এই মানগুলির ভিত্তিতে লঘিষ্ঠ বর্গসমষ্টি পদ্ধতিতে (1.57)-এ দেওয়া সাযুজ্যরেখাটি নিরূপণ করা যাবে।

এই প্রসঙ্গে উদাহরণ 1.12 দেখা যেতে পারে।

৩৮·৪·২ লগ-লগ রূপান্তরের সাহায্যে সরলরেখা

চল দু'টির সম্পর্ক

$$y=\varphi(x;a,b)=a^{x^b} \quad \ldots (1.58)$$

হলে, $\log y=x^b\log a=a'x^b$

বা, $\log(\log y)=\log a'+b\log x$

বা, $y''=a''+bx'$

অর্থাৎ, $y'' = \log(\log y)$ এবং $x' = \log x$-এর মধ্যে একটি ঋজুরৈখিক সম্পর্ক পাওয়া যাচ্ছে। সুতরাং, এক্ষেত্রে $(y_i'' = \log(\log y_i), (x_i' = \log x_i), i = 1(1)n$-এর ভিত্তিতে লঘিষ্ঠ বর্গ সমষ্টি পদ্ধতিতে (1.58)-এ দেওয়া সাযুজ্যরেখাটি নিরূপণ করা যায়।

৩৮·৪·৩ ব্যস্ত (inverse) রূপান্তরের সাহায্যে সরলরেখা

$$y = \phi(x; a, b) = \frac{1}{a+bx} \qquad \dots (1.59)$$

হলে $\frac{1}{y} = a + bx$, বা $y' = a + bx$ হবে।

তেমনি $y = \phi(x; a, b) = \frac{1}{a+\frac{b}{x}} = \frac{x}{ax+b}$ ... (1.60)

হলে, $\frac{1}{y} = a + \frac{b}{x}$, বা $y' = a + bx'$ হবে।

সুতরাং দেখা যাচ্ছে, প্রথম ক্ষেত্রে $y' = \frac{1}{y}$ এবং x-এর মধ্যে, এবং দ্বিতীয় ক্ষেত্রে $y' = \frac{1}{y}$ এবং $x' = \frac{1}{x}$-এর মধ্যে ঋজুরৈখিক সম্পর্ক রয়েছে। অর্থাৎ, এ দু'টি ক্ষেত্রেও সাযুজ্যরেখা নিরূপণে লঘিষ্ঠ বর্গ সমষ্টি পদ্ধতি ব্যবহার করা যাবে।

৩৮·৪·৪ মডিফায়েড এক্সপোনেনশিয়াল (modified exponential) রেখা

$$y = \phi(x; a, b) = a + b^x \qquad \dots (1.61)$$

সম্পর্কটিকে বলা হয় মডিফায়েড এক্সপোনেনশিয়াল রেখা। সরাসরি অথবা কোন রকম রূপান্তরের সাহায্যে এটিকে ঘাতজক আকারে প্রকাশ করা যায় না। এই ধরনের সাযুজ্যরেখা নিরূপণ করার একটি পদ্ধতি হল গোষ্ঠী গড় পদ্ধতি (method of group averages)। সাযুজ্যরেখাটিতে যদি kটি ধ্রুবক থাকে, তাহলে গোষ্ঠীগড় পদ্ধতিতে প্রদত্ত মানগুলিকে k-টি সম আয়তনের গোষ্ঠীতে ভাগ করে এই গোষ্ঠীগুলির সমষ্টি বা গড়মানের ভিত্তিতে k-টি সমীকরণ গঠন করে ধ্রুবকগুলি নির্ণয় করা হয়। k=2 হলে পদ্ধতিটিকে অর্ধ-গড় পদ্ধতি (method of semi-average) বলা হয়।

অর্ধ-গড় পদ্ধতিতে (1.61)-এ প্রদত্ত সাযুজ্যরেখাটি কীভাবে নিরূপণ করা হবে, সে সম্বন্ধে আলোচনা করা যাক।

এই ধরনের সম্পর্কের ক্ষেত্রে অনপেক্ষ চলটি সাধারণত কতকগুলি পরস্পর সমদূরত্ববিশিষ্ট মান পরিগ্রহণ করে, যেমন সময়-সূচক চলের ক্ষেত্রে। এর ফলে রৈখিক রূপান্তরের সাহায্যে অনপেক্ষ চলের

মানগুলিকে 0, 1, 2, ..., n–1 করে নেওয়া যায়। উদাহরণস্বরূপ, t=1980, 1982, ..., 1998–এই বছরগুলির জন্য y একটি কালীন সারি হলে, x = (t–1980)/2–এই রূপান্তরের সাহায্যে x-এর মানগুলি হবে 0, 1, ..., 9। সুতরাং, এক্ষেত্রে সম্পর্কটিকে লেখা যায়

$$y_i = a + b^i, \; i = 0(1)n-1 \qquad \ldots (1.62)$$

হিসাবে। এখানে দু'টি ধ্রুবক a ও b নিরূপণ করা প্রয়োজন। এজন্য আমাদের প্রদত্ত n-জোড়া মানকে এমনভাবে সমান আয়তনে দু'টি ভাগে ভাগ করতে হবে যাতে প্রতিটি ভাগে i-এর মানগুলি পরপর থাকে। $n = 2m$ (যুগ্ম) হলে প্রথম ভাগে $(0, 1, ..., m-1)$ এবং দ্বিতীয় ভাগে $(m, m+1, ..., 2m-1)$ নিতে হবে। $n = 2m+1$ (অযুগ্ম) হলে একটি মান (প্রথমটি অথবা শেষটি—বাঞ্ছনীয়ত প্রথমটি) বাদ দিতে হবে। সেক্ষেত্রে ভাগ দু'টি হবে $(1, 2, ..., m)$ এবং $(m+1, m+2, ..., 2m)$।

ধরা যাক, $n = 2m$ এবং $S_1 = \sum_{i=0}^{m-1} y_i \;;\; S_2 = \sum_{i=m}^{2m-1} y_i$

$$\text{সুতরাং, } S_1 = \sum_{i=0}^{m-1}(a + b^i) = ma + \frac{1-b^m}{1-b} \qquad \ldots (1.63)$$

$$\text{এবং } S_2 = \sum_{i=m}^{2m-1}(a + b^i) = ma + b^m \frac{1-b^m}{1-b} \qquad \ldots (1.64)$$

$$\text{অতএব, } S_1 - S_2 = \frac{(1-b^m)^2}{1-b} \qquad \ldots (1.65)$$

(1.65)-এ একমাত্র অজ্ঞাত রাশি b(S_1, S_2 এবং m-এর সংখ্যামান আমাদের জানা)। সুতরাং এটিকে যে কোন পুনরাবৃত্ত পদ্ধতিতে (iterative method) সমাধান করে b-এর মান পাওয়া যাবে। ঐ মান (1.63) ও (1.64)-এ বসিয়ে a-র যে দু'টি মান পাওয়া যাবে সে দু'টির গড়ই হবে a-র চূড়ান্ত মান।

এই জাতীয় আরো কয়েকটি সাযুজ্যরেখা হল :

$y = a + b.c^x$ (তিনটি ধ্রুবক বিশিষ্ট মডিফায়েড এক্সপোনেনসিয়াল) ... (1.66)

$y = a.b^{c^x}$ (গম্‌পার্জ) ... (1.67)

$y = a.b^x c^{d^x}$ (ম্যাকেহাম্‌) ... (1.68)

$y = \frac{a}{1+e^{b(c-x)}}$ (লজিস্টিক্‌) ... (1.69)

লক্ষ্য করলেই বোঝা যাবে (1.67), (1.68) ও (1.69) রেখাগুলিকে রূপান্তরের সাহায্যে মডিফায়েড এক্সপোনেনশিয়াল রেখার আকারে আনা যাবে।

৩৮·৪·৫ সাযুজ্যতার উৎকর্ষ বিচার (Texting of Goodness of Fit)

(x, y)-চল দু'টির একপ্রস্থ মানের ভিত্তিতে $y=\phi(x;\theta_1,\theta_2,\ldots,\theta_k)$-এই সাযুজ্যরেখাটি নিরূপণ করার পর যে প্রশ্নটির বিচার করার প্রয়োজন দেখা দেবে তা হল, সাযুজ্যরখাটি প্রদত্ত রাশিতথ্যের পক্ষে কতখানি উপযুক্ত হল, বা এর সঙ্গে কতখানি সমঞ্জস হল। নির্ভরী চল y-এর প্রদত্ত মানগুলি, সাযুজ্যরেখা থেকে পাওয়া সংশ্লিষ্ট মানগুলির কত কাছাকাছি আছে তা পরীক্ষা করে এই বিচার করা হয়।

x-এর প্রদত্ত মান x_i-এর জন্য y-এর অবেক্ষিত মান y_i এবং সাযুজ্যরেখা লব্ধ মান y_i হলে (i = 1, 2, ..., n), r_{xy}-এই দু' প্রস্থ মানের নৈকট্যের একটা মাপক হতে পারে। 38.3.1 অনুচ্ছেদে আমরা দেখেছি লঘিষ্ঠ বর্গ সমষ্টিতে পাওয়া সাযুজ্যরেখার ক্ষেত্রে $r_{yY}=|r_{xy}|$; r_{yY}-এর মান 1-এর যত কাছাকাছি হবে, প্রদত্ত রাশিতথ্যের পক্ষে সাযুজ্যরেখাটি তত উৎকৃষ্ট হয়েছে বলা যাবে। অন্যদিকে r_{yY}-এর মান শূন্যের কাছাকাছি হলে বুঝতে হবে এই সাযুজ্যরেখাটি প্রদত্ত রাশিতথ্যের পক্ষে আদৌ উপযুক্ত হয়নি।

সাযুজ্যরেখাটিকে রূপান্তরের সাহায্যে ঋজুরৈখিক আকারে আনা হলে y ও Y-এর পরিবর্তে সংশ্লিষ্ট রূপান্তরিত চল দু'টির মধ্যে সহগতি গুণাঙ্ক নির্ণয় করে এই উৎকর্ষ বিচার করা হবে। সম্পর্কটি ঘাতজকে প্রকাশ করা সম্ভব না হলেও অবেক্ষিত y এবং সাযুজ্যরেখা লব্ধ Y-এর সহগতি গুণাঙ্ক সাযুজ্যতার উৎকর্ষ বিচারে সাহায্য করবে।

সাযুজ্যতার উৎকর্ষ বিচারে অনুমিতিতত্ত্ব সম্মত কাই-বর্গ বিচার (χ^2 – test) পদ্ধতি প্রয়োগ করা হয়। এটি যথাস্থানে আলোচিত হবে।

উদা. ২.৫ : নিচের রাশিতথ্যের জন্য $y=ax^b$ সাযুজ্যরেখাটি নিরূপণ কর এবং একটি উপযুক্ত মাপকের সাহায্যে রেখাটির সাযুজ্যতার উৎকর্ষ বিচার কর।

x	2.5	3.0	3.5	4.0	4.5	5.0	5.5	6.0
y	15.06	18.14	18.95	19.73	20.73	23.14	23.96	24.61

এখানে $y=a.x^b \Rightarrow \log y=\log a+b\log x$

$$\Rightarrow y'=a'+bx'$$

এক্ষেত্রে a' ও b নির্ণয়ের জন্য নর্মাল সমীকরণ দু'টি হল

$$\sum_{i=1}^{8} y'_i = 8a' + b\sum_{i=1}^{8} x'_i$$

$$\sum_{i=1}^{8} y'_i x'_i = a'\sum_{i=1}^{8} x'_i + b\sum_{i=1}^{8} x'^2_i$$

a' ও b নির্ণয়ের জন্য আমরা নিম্নলিখিত ছকে অঙ্কপাতন করতে পারি :

i	x_i	y_i	$x'_i = \log_{10} x_i$	$y'_i = \log_{10} y_i$	x'^2_i	y'^2_i	$x'_i y'_i$	Y'_i
1	2.5	15.06	.3979	1.1778	.1583	1.3872	.4686	1.1908
2	3.0	18.14	.4771	1.2586	.2276	1.5841	.6005	1.2330
3	3.5	18.95	.5441	1.2776	.2960	1.6323	.6951	1.2687
4	4.0	19.73	.6021	1.2951	.3625	1.6773	.7798	1.2996
5	4.5	20.78	.6532	1.3176	.4267	1.7361	.8607	1.3269
6	5.0	23.14	.6990	1.3644	.4886	1.8616	.9537	1.3513
7	5.5	23.96	.7404	1.3795	.5482	1.9030	1.0214	1.3734
8	6.0	24.61	.7789	1.3911	.6056	1.9352	1.0826	1.3935
মোট	—	—	4.8920	10.4617	3.1135	13.7168	6.4624	

এখন $b = \frac{8\times6.4624-4.8920\times10.4617}{8\times3.1135-(4.8920)^2} = 0.5332$

এবং $a' = \bar{y}' - b\bar{x}' = \frac{10.4617}{8} - 0.5332 \times \frac{4.8920}{8} = 0.9786$

অর্থাৎ, $a = \text{anti}\log a' = 9.5192$;

সুতরাং, নিরূপিত সাযুজ্যরেখাটি হল

$$\log y = 0.9786 + 0.5332 \log x$$

বা, $y = 9.5192x^{0.5332}$

এই সাযুজ্যরেখা থেকে বিভিন্ন x_i – এর জন্য পাওয়া $\log y$ – এর মান Y'_i ওপরের ছকের শেষ স্তম্ভে দেওয়া হয়েছে।

এখানে $S_{y'} = 0.0669$, এবং $S_{Y'} = 0.0659 = r_{y'Y'}S_y$

সুতরাং, $r_{y'Y'} = 0.9845$ এই মানটি যেহেতু 1-এর বেশ কাছাকাছি, এখানে বলা যায় লব্ধ রেখাটির সঙ্গে প্রদত্ত রাশিতথ্যের সাযুজ্যতা খুবই চমৎকার।

৩৮·৫ সারাংশ

বর্তমান পরিচ্ছেদে লঘিষ্ঠ বর্গ পদ্ধতিতে নির্ভরণ রেখা নির্ণয় সম্বন্ধে বিস্তারিত আলোচনা করা হয়েছে। বলা হয়েছে, এই নির্ভরণ রেখাকে কীভাবে পূর্বানুমান সূত্র হিসাবে ব্যবহার করা যাবে। নির্ভরণ রেখার বিভিন্ন ধর্ম নিয়েও আলোচনা করা হয়েছে। প্রয়োজনীয় রূপান্তরের সাহায্যে কীভাবে লঘিষ্ঠ বর্গ পদ্ধতি প্রয়োগ করে বিভিন্ন ধরনের সাযুজ্যরেখা নিরূপণ করা হয়, সে আলোচনা এসেছে এরপর। গোষ্ঠীগড় পদ্ধতিতে মডি ফায়েড এক্সপোনেনশিয়াল রেখা নিরূপণের পদ্ধতিও বলা হয়েছে এই সঙ্গে। সবিশেষে বিভিন্ন ধরনের সাযুজ্যরেখা নিরূপণের উৎকর্ষ বিচারের বিভিন্ন পদ্ধতি নিয়ে আলোচনা করা হয়েছে।

৩৮·৬ অনুশীলনী

(ক) বড় প্রশ্ন :

১. নির্ভরণ-এর সংজ্ঞা দাও। লঘিষ্ঠ বর্গ সমষ্টি নির্ভরণ রেখার ধর্মগুলি প্রমাণসহ উল্লেখ করুন।

২. নির্ভরণ রেখা দু'টির অন্তর্বর্তী কোণটির পরিমাপ চল দু'টির ভেদমান ও সহগতি গুণাঙ্কের সাহায্যে প্রকাশ কর। এই প্রকাশন থেকে রেখা দু'টির (i) অভিন্ন, এবং (ii) পরস্পরের লম্ব হওয়ার শর্ত নির্ণয় করুন।

৩. r-কে কেন পূর্বানুমান সূত্র হিসাবে নির্ভরণ রেখার উপযুক্ততার মাপক বলা হয়, প্রমাণ সহযোগে লিখুন।

৪. (x, y) – এর 10-জোড়া মানের ভিত্তিতে পাওয়া দু'টি নির্ভরণ রেখা হল

$$3x + 5y = 13$$
$$x + 3y = 7$$

[এখানে $x3X$ এবং $y3Y$-আলাদাভাবে দেখানো হয়নি।]

এই 10 জোড়া মানের জন্য x ও y-এর গড় ও r-এর মান নির্ণয় করুন।

যদি $S_x = 0.8S_y$ হয়, তবে S_x ও S_y-এর মান নির্ণয় করুন।

উপযুক্ত নির্ভরণ রেখা ব্যবহার করে x যখন 1.5 তখন y-এর, এবং y যখন 1.8 তখন x-এর অনুমিত মান নির্ণয় করুন।

(খ) ছোট প্রশ্ন :

৫. x-এর ভিত্তিতে y-এর ভিত্তিতে x-এর পূর্বানুমান সূত্র হিসাবে ব্যবহারের জন্য একটিমাত্র নির্ভরণ রেখা কেন যথেষ্ট নয়?

৬. n-জোড়া মানের ভিত্তিতে একটি বহুঘাতজক সাযুজ্যরেখা নিরূপণের জন্য লঘিষ্ঠ বর্গ সমষ্টি পদ্ধতি এবং অন্তঃপ্রক্ষেপণ পদ্ধতির মধ্যে পার্থক্য কী?

(গ) বস্তুমুখী প্রশ্ন :

৭. নিচের প্রশ্নগুলির প্রত্যেকটির জন্য সঠিক উত্তরটিতে দাগ দাও :

(i) $x = 4.2 - 3.6y$ যদি y-এর ওপর x-এর নির্ভরণ রেখা হয়, তাহলে নিচের কোন্ সমীকরণটি x-এর ওপর y-এর নির্ভরণ সমীকরণ হতে পারবে না?

A : $Y = -2.8 + 0.6x$ B : $Y = 2.8 - 0.6x$

C : $Y = -2.8 - 0.6x$ D : $Y = 1.4 - 0.3x$

(ii) $b_{xy} = 1.3$ হলে নিচের কোন্ কোন্ উক্তি সত্য সঙ্গে দেওয়া যথার্থ বিকল্পটির সাহায্যে জানাও।

I $b_{yx} = 0.6$ হতে পারে।

II $r = -0.7$ হতে পারে।

III y-এর ওপর x-এর নির্ভরণ রেখাটি মূল বিন্দুর মধ্যে দিয়ে যেতে পারে।

A : কেবল I ও II সত্য।

B : কেবল I ও III সত্য।

C : কেবল II ও III সত্য।

D : I, II, III–সবগুলিই সত্য।

(গ) বড় প্রশ্ন (অতিরিক্ত) :

নিচের রাশিতথ্যের সাযুজ্যে $y = \frac{1}{a+bx}$ রেখাটি নিরূপণ কর :

x	5	10	15	20	25	30	35	40	45
y	0.570	0.562	0.435	0.476	0.435	0.407	0.389	0.392	0.370

৩৮.৬ গ্রন্থপঞ্জী

(1) শৈলেশ ভূষণ চৌধুরী, অরিজিৎ চৌধুরী ও বিশ্বনাথ দাস, রাশি বিজ্ঞানের মূলতত্ত্ব, প্রথম ও দ্বিতীয় খণ্ড (দশম ও দ্বাদশ পরিচ্ছেদ)। পঃ বঙ্গ রাজ্য পুস্তক পর্ষদ, ১৯৭৬।

(2) A. M. Goon, M. K. Gupta & B. Dasgupta : Fundamentals of Statistics, Vol 1 (Chapter 12), World Press, 1998.

(3) S. C. Gupta & V. K. Kapoor : Fundamentals of Mathematical Statistics (Ch 9 & 10), Sultan Chand & Sons, 1995.

(4) Croxton, Cowden & Klien : Applied General Statistics (Ch 19), Prentice Hall of India, 1973.

(5) Ezekiel, M, Fox, K. A. & John Wiley, Methods of Correlation & Regression Analysis (Chs 5-9), 1959

একক ৩৯ □ বহুচল রাশিতথ্য বিশ্লেষণ, বহুল নির্ভরণ এবং বহুল ও আংশিক সহগতি

গঠন

৩৯.০ উদ্দেশ্য

এই এককটি পাঠ করার পর আপনি

- বহুল নির্ভরণ সমীকরণ
- বহুল সহগতি গুণাঙ্ক
- ভ্রান্তি ভেদমান
- আংশিক সহগতি
- আংশিক নির্ভরণ সহগাঙ্ক
- বহুল ও আংশিক সহগতি সংক্রান্ত কয়েকটি ফল, ও
- ক্রমিক নির্ভরণ সংক্রান্ত বিষয়ে অভিজ্ঞতা লাভ করবেন

৩৯.১ প্রস্তাবনা

আমরা আগের পরিচ্ছেদে দ্বিচল রাশিতথ্য বিশ্লেষণের নানান পদ্ধতির কথা জেনেছি। কিন্তু অনেক সময় নির্ভরী চলের চলিষ্ণুতা একটিমাত্র অনপেক্ষ চলের সাহায্যে ব্যাখ্যা করা যায় না, এটি প্রকৃতপক্ষে নির্ভর করে

একাধিক অনপেক্ষ চলের ওপর। যেমন একটি পরিবারে একটি ভোগ্যপণ্যের জন্য চাহিদা (x_1)ঐ পণ্যের বাজারদর (x_2), ঐ পণ্যটির বিকল্প কোন পণ্যের বাজারদর (x_3), পরিবারের আয় (x_4), পরিবারের সদস্যসংখ্যা (x_5)—ইত্যাদির ওপর নির্ভর করতে পারে। এই কারণে অনেক সময় আমাদের বহুচল রাশিতথ্য বিশ্লেষণের প্রয়োজন দেখা যায়।

ধরা যাক, আমাদের বিচার্য $x_1, x_2, \ldots, x_k$—এই p-টি চল। n-টির ব্যষ্টির জন্য এই চলগুলি পরিমাপ করে যে তথ্য পাবে, সেগুলি আমরা একটি ম্যাট্রিক্স আকারে প্রকাশ করতে পারি $\underset{\sim}{X}$ দিয়ে, যেখানে

$\underset{\sim}{X}_{p \times n} = (x_{i\alpha})$, $x_{i\alpha} = \alpha$ তম ব্যষ্টির জন্য x_i-এর মান, $i = 1(1)p, \alpha = 1(1)n$

$$\begin{pmatrix} x_{11} & x_{12} \ldots x_{1n} \\ x_{21} & x_{22} \ldots x_{2n} \\ x_{p1} & x_{p2} \ldots x_{pn} \end{pmatrix} \qquad \ldots (2.1)$$

আলোচনার সুবিধার জন্য আমরা নির্ভরী চলটিকে x_1 বলব এবং $x_2, x_3, \ldots, x_p$ হবে অনপেক্ষ চল। যেসব প্রশ্নের উত্তর খোঁজার জন্য আমরা বহুচল রাশিতথ্য বিশ্লেষণ করব সেগুলি হল :

(১) নির্বাচিত $x_2, x_3, \ldots, x_p$ -এই চলগুলির মধ্যে কোন্‌গুলি x_1-কে প্রভাবিত করে?

(২) যেগুলি x_1-কে প্রভাবিত করে বলে প্রাথমিক বিচারে মনে হবে, সেগুলির ভিত্তিতে কীভাবে x_1-এর একটি পূর্বানুমান সূত্র নির্ণয় করা যাবে?

(৩) নির্ণীত সূত্রটি x_1-এর জন্য পূর্বানুমান দেওয়ার ব্যাপারে কতখানি কার্যকরী?

(৪) সূত্রটি কার্যকরী হয়েছে মনে হলে সূত্রটিতে অন্তর্ভুক্ত কোন কোন চল কি পূর্বানুমান সূত্রের গুণমান অক্ষুণ্ণ রেখে বাদ দেওয়া যায় এবং সূত্রটি কার্যকরী হয়নি মনে হলে আরো কোন্ কোন্ নতুন অনপেক্ষ চল যুক্ত করা যেতে পারে?

প্রশ্নগুলির উত্তর কীভাবে আমরা এক এক করে পেতে পারি, সে সম্বন্ধে আলোচনা করা হয়েছে পরবর্তী অনুচ্ছেদগুলিতে।

৩৯·২ বহুল নির্ভরণ সমীকরণ (Multiple Regression Equation)

৩৯.১ অনুচ্ছেদে উত্থাপিত প্রথম প্রশ্নটি বিচারের উদ্দেশ্যে প্রতিটি সম্ভাব্য অনপেক্ষ চলের সঙ্গে নির্ভরী চলের সম্পর্কটি খতিয়ে দেখার জন্য এক একটি বিক্ষেপণ চিত্রে সংশ্লিষ্ট বিন্দুগুলি সংস্থাপন করে ঐ চিত্রে বিন্দুগুলির বিন্যাস পর্যালোচনা করতে হবে।

ধরা যাক, বিক্ষেপণ চিত্র বিশ্লেষণের পর, অথবা বাস্তব অভিজ্ঞতার ভিত্তিতে, প্রাথমিক বিচারে মনে হল $x_2, x_3, \ldots, x_p$-এই $(p-1)$-টি অনপেক্ষ চল নির্ভরী চল x_1-কে প্রভাবিত করে।

লব্ধ রাশিতথ্য $\underset{\sim}{X}_{pxn}$-কে কাজে লাগিয়ে প্রথমেই আমরা $x_2, x_3, \ldots, x_p$-এর ভিত্তিতে লঘিষ্ঠ বর্গ-সমষ্টি পদ্ধতিতে x_1-এর জন্য একটি পূর্বানুমান সূত্র নিরূপণ করব। আগের মতই এখানে গৃহীত প্রতিরূপটি হবে

$$x_{1\alpha} = X_{1.23 \cdots p_\alpha} + x_{1.23 \cdots p_\alpha} \qquad \ldots (2.2)$$

যেখানে $X_{1.23 \cdots p_\alpha} = b_0 + b_2 x_{2\alpha} + \ldots + b_p x_{p\alpha}$ ... ...(2.3) হল x_1-এর নির্ভরণ অংশ ($x_2, x_3, \ldots, x_p$ 'র ওপর) এবং $x_{1.23 \cdots p_\alpha}$ হল ভ্রান্তি অংশ।

$b_0, b_2, \ldots, b_p$ নির্ণয়ের জন্য আমরা আগের মতই লঘিষ্ঠ বর্গ সমষ্টি পদ্ধতিতে

$$\sum_{\alpha=1}^{n} x_{1.23 \cdots p\alpha}^2 = \sum_{\alpha=1}^{n} (x_{1\alpha} - b_0 - b_2 x_{2\alpha} - \ldots - b_p x_{p\alpha})^2 \qquad \ldots (2.4)$$

এই প্রকাশনটিকে $b_0, b_2, \ldots, b_p$-র সাপেক্ষে লঘিষ্ঠকরণে সচেষ্ট হব। এক্ষেত্রে সংশ্লিষ্ট নর্মাল সমীকরণগুলি হবে

$$\frac{\partial}{\partial b_j} \sum_{\alpha=1}^{n} x_{1.23 \cdots p_\alpha}^2 = 0, \; j = 0, 2, 3, \ldots, p$$

বা, $$\sum_{\alpha=1}^{n} x_{1.23 \cdots p_\alpha} = 0 \qquad \ldots (2.5)$$

এবং $$\sum_{\alpha=1}^{n} x_{1.23 \cdots p_\alpha} \, x_{j\alpha} = 0, \; j = 2(1)p \qquad \ldots (2.6)$$

এই সমীকরণগুলিকে $b_0, b_2, \ldots, b_p$-র সাহায্যে প্রকাশ করলে দাঁড়াবে

$$\sum_{\alpha=1}^{n} x_{1\alpha} = nb_0 + b_2 \sum_{\alpha} x_{2\alpha} + \ldots + b_p \sum_{\alpha} x_{p\alpha} \qquad \ldots (2.7)$$

এবং $$\sum_{\alpha} x_{1\alpha} x_{j\alpha} = b_0 \sum_{\alpha} x_{j\alpha} + b_2 \sum x_{2\alpha} x_{j\alpha} + \ldots + b_p \sum_{\alpha} x_{p\alpha} x_{j\alpha} \qquad \ldots (2.8)$$

$$j = 2(1)p$$

(2.7) ÷ n থেকে পাওয়া যায়

$$\overline{x}_1 = b_0 + b_2\overline{x}_2 + \ldots + b_p\overline{x}_p \qquad \ldots (2.9)$$

এবং $(2.8) \div n - \overline{x}_j \times (2.9)$ থেকে পাওয়া যায়

$$S_{j1} = b_2 S_{j2} + b_3 S_{j3} + \ldots + b_p S_{jp} \qquad \ldots (2.10)$$

যেখানে $S_{ij} = \frac{1}{n}\sum_{\alpha}(x_{i\alpha} - \overline{x}_i)(x_{j\alpha} - \overline{x}_j)$

$= Cov(x_i, x_j), i \neq j$ হলে

এবং $= v(x_i), i = j$ হলে,

(2.10) সমীকরণটি বিস্তারিতভাবে লিখলে দাঁড়ায়

$$\begin{aligned}
s_{21} &= b_2 S_{22} + b_3 S_{23} + \ldots + b_p S_{2p} \\
s_{31} &= b_2 S_{32} + b_3 S_{33} + \ldots + b_p S_{3p} \\
\ldots & \ldots \\
s_{p1} &= b_2 S_{p2} + b_3 S_{p3} + \ldots + b_p S_{pp}
\end{aligned} \qquad \ldots (2.11)$$

বহুচল সমীকরণ সমাধানের জন্য ক্র্যামারের নিয়ম অনুসরণ করে পাওয়া যাবে

$$b_j = \frac{\begin{vmatrix} S_{22} \ldots S_{2j-1} S_{21} S_{2j+1} \ldots S_{2p} \\ S_{32} \ldots S_{3j-1} S_{31} S_{3j+1} \ldots S_{3p} \\ \ldots \ \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \\ S_{p2} \ldots S_{pj-1} S_{p1} S_{pj+1} \ldots S_{pp} \end{vmatrix}}{\begin{vmatrix} S_{22} S_{23} \ldots S_{2p} \\ S_{32} S_{33} \ldots S_{3p} \\ \ldots \\ S_{p2} S_{p3} \ldots S_{pp} \end{vmatrix}} \qquad j = 2(1)p \qquad \ldots (2.12)$$

এখন $r_{ij} = r_{x_i, x_j} = \frac{S_{ij}}{S_i S_j} = r_{ji}$ (যেখানে $S_i = \sqrt{S_{ij}}$),

$$R_{p\times p} = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ r_{p1} & r_{p2} & \cdots & r_{pp} \end{pmatrix}$$ = সহগতি ম্যাট্রিক্স, যেখানে $r_{ii} = 1, i = 1(1)p$ এবং $r_{ij} = r_{ji}$

এবং $R_{ij} = R$-ম্যাট্রিক্সে r_{ij}'র কো-ফ্যাক্টর,—এই সংকেত চিহ্ন ব্যবহার করে পাওয়া যাবে

$$b_j = \frac{\begin{vmatrix} r_{22}S_2^2 & \cdots r_{2j-1}S_2S_{j-1} & r_{21}S_2S_1 & r_{2j+1}S_2S_{j+1} & \cdots r_{2p}S_2S_p \\ r_{32}S_3S_2 & \cdots r_{3j-1}S_3S_{j-1} & r_{31}S_3S_1 & r_{3j+1}S_3S_{j+1} & \cdots r_{3p}S_3S_p \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ r_{p2}S_pS_2 & \cdots r_{pj-1}S_pS_{j-1} & r_{p1}S_pS_1 & r_{pj+1}S_pS_{j+1} & \cdots r_{pp}S_p^2 \end{vmatrix}}{\begin{vmatrix} r_{22}S_2^2 & r_{23}S_2S_3 & \cdots & r_{2p}S_2S_p \\ r_{32}S_3S_2 & r_{33}S_3^2 & \cdots & r_{3p}S_3S_p \\ \cdots & \cdots & \cdots & \cdots \\ r_{p2}S_pS_2 & r_{p3}S_pS_3 & \cdots & r_{pp}S_p^2 \end{vmatrix}}$$

$$= (-1)^{j-2}\frac{S_1}{S_j}\frac{\begin{vmatrix} r_{21} & r_{22} & \cdots r_{2j-1} & r_{2j+1} & \cdots r_{2p} \\ r_{31} & r_{32} & \cdots r_{3j-1} & r_{3j+1} & \cdots r_{3p} \\ r_{p1} & r_{p2} & \cdots r_{pj-1} & r_{pj+1} & \cdots r_{pp} \end{vmatrix}}{\begin{vmatrix} r_{22} & r_{23} & \cdots & r_{2p} \\ r_{32} & r_{33} & \cdots & r_{3p} \\ \cdots & \cdots & \cdots & \cdots \\ r_{p2} & r_{p3} & \cdots & r_{pp} \end{vmatrix}} \quad \ldots(2.13)$$

একটু লক্ষ্য করলে দেখা যাবে (2.13) সমীকরণের লবটি R-ম্যাট্রিক্সে r_{1j}-র মাইনর, অর্থাৎ $(-1)^{1+j}R_{1j}$; এবং হরটি R-ম্যাট্রিক্সে r_{11}-এর মাইনর, অর্থাৎ $(-1)^{1+1}R_{11}$ ।

সুতরাং, $b_j = (-1)^{j-2+1+j}\ \frac{S_1}{S_j}\frac{R_{1j}}{R_{11}} = -\frac{R_{1j}}{R_{11}}.\frac{S_1}{S_j}$ ... (2.14)

$$j = 2(1)p$$

বিকল্পভাবে (2.11) সমীকরণগুলি ম্যাট্রিক্স আকারে প্রকাশ করলে দাঁড়াবে

$$\begin{pmatrix} S_{21} \\ S_{31} \\ \vdots \\ S_{P1} \end{pmatrix} = \begin{pmatrix} b_2 \\ b_3 \\ \vdots \\ b_p \end{pmatrix} \begin{pmatrix} S_{22}\ S_{23}\ \dots\ S_{2p} \\ S_{32}\ S_{33}\ \dots\ S_{3p} \\ \dots\ \dots\ \dots\ \dots \\ S_{p2}\ S_{p3}\ \dots\ S_{pp} \end{pmatrix} \quad \dots (2.15)$$

এখানে $S_{\overline{p-1} \times \overline{p-1}} = (S_{ij})$-ম্যাট্রিক্সটি অ-বিশিষ্ট (non-singular) ধরে নিলে স্পষ্টতই

$$\begin{pmatrix} b_2 \\ b_3 \\ \vdots \\ b_P \end{pmatrix} = \begin{pmatrix} S_{22}\ S_{23}\ \dots\ S_{2p} \\ S_{32}\ S_{33}\ \dots\ S_{3p} \\ \dots\ \dots\ \dots\ \dots \\ S_{p2}\ S_{p3}\ \dots\ S_{pp} \end{pmatrix}^{-1} \begin{pmatrix} S_{21} \\ S_{31} \\ \vdots \\ S_{P1} \end{pmatrix} \quad \dots (2.16)$$

বা, $b_j = S^{j2}S_{21} + S^{j3}S_{31} + \dots + S^{jp}S_{p1}$ ... (2.17)

$$j = 2(1)p$$

যেখানে $(S^{ij}) = (S_{ij})^{-1}$।

$j = 2(1)p$-র জন্য bj-র মান (2.14) বা (2.17) থেকে নির্ণয় করে (2.9)-এ বসালে পাওয়া যাবে

$$b_0 = \overline{x}_1 - b_2\overline{x}_2 - \dots - b_p\overline{x}_p \quad \dots (2.18)$$

এখন (2.3)-এ b_0 এবং $b_2, b_3, \dots, b_p$-র এই মানগুলি বসালে পূর্বানুমান সূত্রটি দাঁড়াবে

$$x_{1.23\dots p} = \overline{x}_1 - \frac{R_{12}}{R_{11}}\frac{S_1}{S_2}(x_2 - \overline{x}_2) - \dots - \frac{R_{1p}}{R_{11}} \cdot \frac{S_1}{S_p}(x_p - \overline{x}_p) \quad \dots (2.19)$$

(2.19) সমীকরণটিকে বলা হয় $x_2, x_3, \dots x_p$'র উপর x_1-এর বহুল নির্ভরণ সমীকরণ (multiple regression equation)। কোন একটি ব্যষ্টির জন্য $x_2, x_3, \dots x_p$-এর মান জানা থাকলে ব্যষ্টিটির x_1-এর মান সম্বন্ধে পূর্বাভাষ পাওয়ার উদ্দেশ্যে এই সমীকরণটি ব্যবহার করা যায়। এই সমীকরণে

$$b_j \equiv b_{ij\cdot 23\dots j-1\ j+1\dots p} = -\frac{R_{ij}}{R_{11}} \cdot \frac{S_1}{S_j}, \quad j = 2(1)p, \quad \dots (2.20)$$

—এই গুণাঙ্কটিকে বলা হয় $x_2, x_3, \ldots, x_{j-1}, x_{j+1}, \ldots, x_p$ চলগুলি ধ্রুবক থাকা সাপেক্ষে x_j-র ওপর x_1-এর আংশিক নির্ভরণাঙ্ক (partial regression coefficient)। এইরকম নামকরণের কারণ, $x_2, x_3, \ldots, x_{j-1}, x_{j+1}, \ldots, x_p$'র মান অপরিবর্তিত রেখে x_j চলটিতে একক পরিমাণে বৃদ্ধি ঘটানো হলে নির্ভরণ সমীকরণ থেকে পাওয়া x_1-এর মানে বৃদ্ধির পরিমাণ দাঁড়াবে এই গুণাঙ্কটির সমান।

$p = 3$ হলে, অর্থাৎ ত্রিচল রাশিতথ্যের ক্ষেত্রে।

$$R = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{13} \\ r_{21} & r_{22} & \cdots & r_{23} \\ r_{31} & r_{32} & \cdots & r_{33} \end{pmatrix} = \begin{pmatrix} 1 & r_{12} & \cdots & r_{13} \\ r_{21} & 1 & \cdots & r_{23} \\ r_{31} & r_{32} & \cdots & 1 \end{pmatrix}$$

সুতরাং, বিভিন্ন আংশিক নির্ভরণাঙ্কের সূত্র হবে

$$b_{12.3} = -\frac{R_{12}}{R_{11}} \cdot \frac{S_1}{S_2} = \frac{r_{12} - r_{13}r_{23}}{1 - r_{23}^2} \cdot \frac{S_1}{S_3} \qquad \ldots (2.21)$$

$$\text{এবং } b_{13.2} = -\frac{R_{13}}{R_{11}} \cdot \frac{S_1}{S_3} = \frac{r_{13} - r_{12}r_{23}}{1 - r_{23}^2} \cdot \frac{S_1}{S_3} \qquad \ldots (2.22)$$

উদা. ৩.১ : কলেজে রাশিবিজ্ঞানে স্নাতক শ্রেণীতে ভর্তির জন্য যোগ্যতার নির্ধারণের উদ্দেশ্যে উচ্চমাধ্যমিকে গণিতে প্রাপ্ত শতকরা নম্বর (x_2) ও কলেজ ভর্তি পরীক্ষায় প্রাপ্ত শতকরা নম্বর (x_3)-এর ভিত্তিতে রাশিবিজ্ঞানে প্রবণতা (x_1)-নির্ণায়ক একটি পূর্বাভাস সূত্র গঠন করার জন্য কিছু রাশিতথ্য সংগৃহীত হয়েছে। বিশ্ববিদ্যালয় স্নাতক পরীক্ষায় রাশিবিজ্ঞানে প্রাপ্ত শতকরা নম্বর ঐ বিষয়ে ছাত্র/ছাত্রীর প্রবণতার মাপক ধরে নিয়ে প্রদত্ত রাশিতথ্যের ভিত্তিতে x_2 ও x_3-র ওপর x_1-এর বহুল নির্ভরণ সমীকরণ গঠন কর। এই সমীকরণটি ব্যবহার করে নমুনালব্ধ ছাত্রদের বিশ্ববিদ্যালয় পরীক্ষায় রাশিবিজ্ঞানের নম্বরের আভাসিত মানগুলিও নির্ণয় কর।

ছাত্রের ক্রমিক সংখ্যা	ছাত্রের প্রাপ্ত নম্বর (%)			
	উচ্চমাধ্যঃ গণিত	ভর্তি পরীক্ষা	বি.বি. রাশিবিজ্ঞান	বি.বি. রাশিবিজ্ঞান (আভাসিত)
(α)	$(x_{2\alpha})$	$(x_{3\alpha})$	$(x_{1\alpha})$	$x_{1.23\alpha}$
1	81	45	51	47.4
2	76	52	56	53.0
3	93	65	72	67.7
4	71	54	60	54.0
5	88	59	62	61.4
6	95	61	58	64.3
7	80	56	52	57.3
8	74	48	48	49.0
9	96	60	68	63.6
10	78	53	45	54.3
মোট	832	553	572	572.0

প্রদত্ত রাশিতথ্য বিশ্লেষণ করে পাওয়া যাচ্ছে

$$S_3 = 5.8318 \qquad S_2 = 8.6579 \qquad S_1 = 8.1707$$

$$\sum_{\alpha} x_{2\alpha}x_{3\alpha} = 46389 \qquad \sum_{\alpha} x_{3\alpha}x_{1\alpha} = 32004 \qquad \sum_{\alpha} x_{1\alpha}x_{2\alpha} = 48059$$

$$r_{23} = 0.7514 \qquad r_{13} = 0.7815 \qquad r_{12} = 0.6624$$

সুতরাং, (2.21) ও (2.22) ব্যবহার করে পাওয়া যাচ্ছে

$$b_{12.3} = \frac{0.6624 - 0.7815 \times 0.7514}{1 - 0.7514^2} \cdot \frac{8.1707}{8.6579} = 0.1630$$

এবং $$b_{13.2} = \frac{0.7815 - 0.6624 \times 0.7514}{1 - 0.7514^2} \cdot \frac{8.1707}{5.8318} = 0.9131;$$

এছাড়া (2.18) থেকে $b_0 = 57.2 - 0.1630 \times 83.2 - 0.9130 \times 55.3 = -6.8560$

সুতরাং, নির্ভরণ সমীকরণটি দাঁড়াচ্ছে

$$x_{1.23} = -6.8560 + 0.1630x_2 + 0.9131x_3$$

এই সমীকরণ থেকে পাওয়া বিশ্ববিদ্যালয় পরীক্ষায় রাশিবিজ্ঞানের নম্বরের আভাসিত মানগুলিও দেখানো হয়েছে উপরের সারণির শেষ স্তম্ভে।

৩৯·৩ বহুল সহগতি গুণাঙ্ক (Multiple Correlation Coefficient)

নির্ভরণ সমীকরণটি নিরূপণ করার পর স্বভাবতই প্রশ্ন উঠবে, এই সমীকরণটি পূর্বানুমান দেওয়ার ব্যাপারে কতখানি কার্যকরী হয়েছে! এ ব্যাপারে যা প্রয়োজন, তা হল একটি উপযুক্ত মাপক, যার সাহায্যে এই কার্যকারিতা পরিমাপ করা যাবে।

আগেই বলা হয়েছে নির্ভরী চলের লব্ধ মানকে দু'টি ভাগে ভাগ করা যায় : একটি হল নির্ভরণ-অংশ, যা পাওয়া যায় একগুচ্ছ অনপেক্ষ চলের ওপর নির্ভরী চলটির নির্ভরণ সমীকরণ থেকে, অন্যটি এর ভ্রান্তি-অংশ। স্পষ্টতই নির্ভরণ সমীকরণে কতগুলি এবং কোন্ কোন্ অনপেক্ষ চল অন্তর্ভুক্ত করা হল, তার ওপর নির্ভর করবে এই দুটি অংশের পরিমাপ। নির্ভরী চল x_1-এর α-তম ব্যষ্টির জন্য লব্ধ মানকে $x_{1\alpha}$ এবং $x_2, x_3, \ldots x_p$'র ওপর x_1-এর নির্ভরণ সমীকরণ থেকে পাওয়া মানকে $x_{1.23.....p\alpha}$ বলা হলে, $\alpha = 1(1)n$ এর জন্য এই দু' প্রস্থ মানের মধ্যে কতখানি সাযুজ্য রয়েছে তা থেকে বোঝা যাবে পূর্বানুমান সূত্র হিসাবে নির্ভরণ রেখাটি কতখানি সার্থক। এই কারণে x_1 এবং $x_{1.23........p}$ এর এই n-জোড়া মানের ভিত্তিতে পাওয়া পূরণী পরিঘাত সহগতি গুণাঙ্ক $r_{1.23......p}$-কে নির্ভরণ রেখাটির পূর্বানুমান দেওয়ার ব্যাপারে কার্যকারিতার মাপক হিসাবে গ্রহণ করা হয়েছে।

তাহলে, $r_{1.23......p} = \dfrac{Cov(x_1, x_{1.23......p})}{\sqrt{v(x_1)}\sqrt{v(x_{1.23......p})}}$

$$= \frac{\frac{1}{n}\sum_{\alpha=1}^{n}(x_{1\alpha}-\bar{x}_1)(x_{1.23....p.\alpha}-\bar{x}_{1.23....p})}{\sqrt{\frac{1}{n}\sum_{\alpha=1}^{n}(x_{1\alpha}-\bar{x}_1)^2}\sqrt{\frac{1}{n}\sum_{\alpha=1}^{n}(x_{1.23....p\alpha}-\bar{x}_{1.23....p})^2}} \qquad \ldots (2.23)$$

x_1-এর সঙ্গে $x_{1.23......p}$-র সহগতিকে বলে যৌথভাবে $x_2, x_3, \ldots, x_p$-র ওপর এর বহুল সহগতি। এবং এইভাবে n-জোড়া মানের ভিত্তিতে নির্ণীত এই সহগতি গুণাঙ্কটিকে বলে যৌথভাবে $x_2, x_3, \ldots, x_p$-র উপর x_1-র বহুল সহগতি গুণাঙ্ক।

$r_{1.23....p}$-কে বিভিন্ন r_{ij}, i, $j = 1(1)p$ —এর অপেক্ষক হিসাবে কীভাবে প্রকাশ করা যায়, এবার দেখা যাক।

(2.5) থেকে পাই $\sum_{\alpha=1}^{n} x_{1.23....p\alpha} = 0$

$$\Rightarrow \bar{x}_{1.23....p} = 0 \qquad \ldots (2.24)$$

আবার, $x_1 = X_{1.23......p} + x_{1.23......p}$

$$\Rightarrow \bar{x}_1 = \bar{X}_{1.23......p} \qquad \ldots (2.25)$$

আবার (2.6) থেকে পাই

$(Cov\, x_i, x_{1.23......p})$

$$= \frac{1}{n}\sum_{\alpha=1}^{n}(x_{i\alpha}-\bar{x}_i)x_{1.23......p\alpha}$$

$$= \frac{1}{n}\sum_{\alpha=1}^{n} x_{i\alpha}x_{1.23.....p\alpha} = 0,\ i = 2(1)p \qquad \ldots (2.26)$$

সুতরাং, $Cov(x_{1.23......p}, x_{1.23......p})$

$$= Cov(x_{1.23......p}, b_0 + b_{12}x_2 + \dots b_{1p}x_p) \qquad \dots (2.27)$$

$$= 0$$

ফলে, $Cov(x_1, x_{1.23......p})$

$$= Cov(x_{1.23.....p} + x_{1.23.....p}, x_{1.23.....p})$$

$$= V(x_{1.23......p}) + Cov(x_{1.23.....p}, x_{1.23......p})$$

$$= v(x_{1.23......p}) \qquad \dots (2.28)$$

কিন্তু, $Cov(x_1, x_{1.23......p})$

$$= \frac{1}{n}\sum_{\alpha=1}^{n}(x_{1\alpha} - \bar{x}_1)(x_{1.23......p\alpha} - \bar{x}_{1.23......p})$$

$$= \frac{1}{n}\sum_{\alpha=1}^{n}x_{1\alpha}(x_{1.23....p\alpha} - \bar{x}_1) \quad [\text{যেহেতু } \sum_{\alpha}(x_{1.23......p\alpha} - \bar{x}_{1.23......p} = 0]$$

$$= \frac{1}{n}\sum_{\alpha=1}^{n} x_{1\alpha}\left\{-\frac{R_{12}}{R_{11}}\frac{S_1}{S_2}(x_{2\alpha} - \bar{x}_2) - \dots\dots - \frac{R_{1p}}{R_{11}}\frac{S_1}{S_p}(x_{p\alpha} - \bar{x}p)\right\}$$

$$= -\frac{R_{12}}{R_{11}}.\frac{S_1}{S_2}.S_{12}\dots\dots - \frac{R_{1p}}{R_{11}}.\frac{S_1}{S_p}.S_{1p}$$

$$= -\frac{R_{12}}{R_{11}}.\frac{S_1}{S_2}.r_{12}S_1S_2 - \frac{R_{13}}{R_{11}}.\frac{S_1}{S_3}r_{13}.S_1S_3 - \dots - \frac{R_{1p}}{R_{11}}.\frac{S_1}{S_p}r_{1p}S_1S_p$$

$$= -\frac{S_1^2}{R_{11}}(r_{12}R_{12} + r_{13}R_{13} + \dots + r_{1p}R_{1p})$$

$$= -\frac{S_1^2}{R_{11}}(R - r_{11}R_{11})$$

$$= \frac{S_1^2}{R_{11}}(R_{11} - R), \quad \because r_{11} = 1$$

$$=(1-\frac{R}{R_{11}})S_1^2, \text{ যেখানে } R=|R| \qquad \text{... (2.29)}$$

অতএব,

$$r_{1.23....p}=\frac{(1-\frac{R}{R_{11}})S_1^2}{\sqrt{S_1^2}\sqrt{(1-\frac{R}{R_{11}})S_1^2}}$$

$$=(1-\frac{R}{R_{11}})^{1/2} \qquad \text{... (2.30)}$$

এখানে লক্ষ্যণীয়, $r_{1.23....p}$ যদিও একটি পূরণী পরিঘাত সহগতি গুণাঙ্ক, এর লব $Cov(x_1, x_{1.23.....p})$ যেহেতু একটি ভেদমানের অর্থাৎ, $V(X_{1.23.....p})$-এর সমান, সেজন্য এটি সবসময় ধনাত্মক হবে অর্থাৎ,

$$0 \le r_{1.23.....p} \le 1 \qquad \text{... (2.31)}$$

$$P=3 \text{ হলে, } |R|=\begin{vmatrix} 1 & r_{12} & r_{13} \\ r_{12} & 1 & r_{23} \\ r_{13} & r_{23} & 1 \end{vmatrix}$$

$$=(r-r_{23}^2)-r_{12}(r_{12}-r_{13}.r_{23})+r_{13}(r_{12}r_{23}-r_{13})$$

$$=1-r_{12}^2-r_{13}^2-r_{23}^2+2r_{12}r_{13}r_{23}$$

এবং $R_{11}=1-r_{23}^2$

$$\text{সুতরাং, } r_{1.23}=\left(1-\frac{1-r_{12}^2-r_{13}^2-r_{23}^2+2r_{12}r_{13}r_{23}}{1-r_{23}^2}\right)^{1/2}$$

$$=\left(\frac{r_{12}^2-r_{13}^2-2r_{12}r_{13}r_{23}}{1-r_{23}^2}\right)^{1/2} \qquad \text{... (2.32)}$$

$$\text{যেহেতু } r_{1.23}^2 \le 1, \text{ স্পষ্টতই } r_{12}^2+r_{13}^2+r_{23}^2 \le 1+2r_{12}r_{13}r_{23} \qquad \text{... (2.32a)}$$

উদাহরণ ৩.২ : উদাহরণ ২.১-এ প্রদত্ত রাশিতথ্যের জন্য বহুল সহগতি গুণাঙ্ক $r_{1.23}$ নির্ণয় কর এবং নির্ণীত মান থেকে বহুল নির্ভরণ সমীকরণটির পূর্বানুমান দেওয়ার ব্যাপারে কার্যকারিতার উপর মন্তব্য কর।

$$\text{এখানে, } r_{1.23} = \left[\frac{0.6624^2 + 0.7815^2 - 2\times 0.6624\times .7815\times 0.7514}{1-0.7514^2}\right]^{1/2}$$

$$= \sqrt{0.6213}$$

$$= 0.7883$$

সুতরাং দেখা যাচ্ছে, নির্ভরণ সমীকরণটি x_2 ও x_3-র ভিত্তিতে x_1-এর পূর্বানুমান দেওয়ার ব্যাপারে যথেষ্ট কার্যকরী।

৩৯·৪ ভ্রান্তি ভেদমান (Error variance)

নির্ভরণ সমীকরণটি পূর্বানুমান দেওয়ার ব্যাপারে কতখানি কার্যকারী হল তার আর একটি মাপক হল ভ্রান্তি ভেদমান।

$$\text{যেহেতু, } x_1 = X_{1.23\ldots\ldots p} + x_{1.23\ldots\ldots p}$$

$$\text{এবং } Cov(x_{1.23\ldots\ldots p}, x_{1.23\ldots\ldots p}) = 0 \quad [(2.27) \text{ থেকে}]$$

$$\text{সুতরাং, } V(x_1) = V(X_{1.23\ldots\ldots p}) + V(x_{1.23\ldots\ldots p})$$

$$\Rightarrow \text{ভ্রান্তি ভেদমান } S^2_{1.23\ldots\ldots p} = V(x_{1.23\ldots\ldots p})$$

$$= V(x_1) - V(X_{1.23\ldots\ldots p})$$

$$= S_1^2 - \left(1 - \frac{R}{R_{11}}\right)S_1^2$$

$$= \frac{R}{R_{11}} S_1^2$$

$$= S_1^2 - r^2_{1.23\ldots\ldots p} S_1^2$$

$$= (1 - r^2_{1.23\ldots\ldots p})S_1^2 \qquad \ldots (2.33)$$

$$\text{অর্থাৎ, } r^2_{1.23\ldots\ldots p} = \frac{V(X_{1.23\ldots p})}{V(x_1)} \qquad \ldots (2.34)$$

$$= 1 - \frac{V(x_{1.23...p})}{V(x_1)} \qquad \dots (2.35)$$

এখন $X_{1.23....p}$ এবং $x_{1.23.....p}$ যেহেতু যথাক্রমে x_1-এর যে অংশটি $x_2, x_3, \dots, x_p$-র ওপর এর নির্ভরণ সমীকরণ দ্বারা ব্যাখ্যাত এবং যে অংশটি অব্যাখ্যাত, আমরা বলতে পারি $r_{1.23....p}$ হল নির্ভরী চলের ভেদমানের যে অংশটি নির্ভরণ সমীকরণের দ্বারা ব্যাখ্যাত, তার পরিমাণ। এই কারণে $r^2_{1.23.......p}$-কে নির্ধারণ সহগাঙ্ক (Coefficient of determination) বলা হয়।

(2.33) থেকে দেখা যাচ্ছে $S^2_{1.23.......p}$ একান্তভাবে $r_{1.23......p}$-র সংখ্যামানের ওপর নির্ভরশীল। $r_{1.23.......p} = 0$ হলে $S^2_{1.23.......p} = S^2_1$, হবে, অর্থাৎ নির্ভরী চলের ভেদমান এবং ভেদমান হবে অভিন্ন, অর্থাৎ নির্ভরণ সমীকরণ নির্ভরী চলের ভেদমানের কিছুমাত্র অংশও ব্যাখ্যা করতে অসমর্থ, সেই কারণে এই সমীকরণ পূর্বানুমান দেওয়ার ব্যাপারে আদৌ কোন কাজে লাগবে না। অন্যদিকে, $r_{1.23.......p} = 1$, হলে $S^2_{1.23.......p} = 0$ হবে, অর্থাৎ প্রতিটি $x_{1.23.......p\alpha}$ হবে ধ্রুবক। আবার যেহেতু $\bar{x}_{1.23.......p} = 0$, এর অর্থ প্রত্যেক α-র জন্য $x_{1.23.......p\alpha} = 0$, ফলে $x_{1\alpha} = X_{1.2.......p\alpha}$, $\alpha = 1(1)n$,, অর্থাৎ পূর্বানুমান সূত্র হিসাবে নির্ভরণ সমীকরণটি সম্পূর্ণ সার্থক। এই কারণেই $r_{1.23......p}$-কে পূর্বানুমান সূত্র হিসাবে নির্ভরণ সমীকরণের কার্যকারিতার মাপক ধরা হয়।

$r_{1.23......p}$-এর মান খুব কম হলে ভ্রান্তি ভেদমান খুব বড় হবে। এর অর্থ, নির্ভরণ সমীকরণ গঠন করার ব্যাপারে যথাযথ এবং পর্যাপ্ত সংখ্যক অনপেক্ষ চল অন্তর্ভুক্ত করা সম্ভব হয়নি। সেক্ষেত্রে নির্ভরী চলের প্রকৃতি অনুযায়ী আরো কিছু অনপেক্ষ চলের সন্ধান করতে হবে যেগুলি নির্ভরী চলটিকে প্রভাবিত করে, এবং এইসব অনপেক্ষ চলকে নির্ভরণ সমীকরণে অন্তর্ভুক্ত করতে হবে।

৩৯.৫ আংশিক সহগতি (Partial Correlation)

$r_{1.23.......p}$-র মান 1-এর খুব কাছাকাছি হলে প্রশ্ন উঠতে পারে, যে সব অনপেক্ষ চল আমরা নির্ভরণ সমীকরণে অন্তর্ভুক্ত করেছি, তাদের সবগুলি নেওয়া প্রয়োজন আছে কিনা। নাকি আমরা পূর্বানুমাত্র সূত্রের গুণগত মানের তেমন হেরফের না ঘটিয়ে এইসব অনপেক্ষ চলের দু'একটিকে সমীকরণ থেকে বাদ দিতে পারি।

দু'টি চলের মধ্যে আপাত সহগতি অনেক সময় বেশ জোরালো মনে হয় সহগতি গুণাঙ্কের মান থেকে। ফলে মনে হবে চল দু'টির মধ্যে কার্যকারণ সম্পর্ক রয়েছে। কিন্তু চল দু'টির ওপর থেকে অপর এক বা একাধিক অনপেক্ষ চলের প্রভাব অপনয়ন করলে দেখা যাবে, অবশিষ্ট অংশ দু'টির মধ্যে তেমন সহগতি নেই।

প্রতি বছরে দেশের স্কুটার উৎপাদনের সংখ্যা (x) এবং ব্যাঙ্ক ডাকাতির সংখ্যা (y) সংক্রান্ত কয়েক বছরের তথ্য নিয়ে যদি দু'টি চলের মধ্যে সহগতি গুণাঙ্ক নির্ণয় করা হয়, দেখা যাবে এই গুণাঙ্কের মান ধনাত্মক এবং বেশ বড়। তা থেকে কেউ বলতে পারেন, এ দু'টি চলের মধ্যে কার্যকারণ সম্পর্ক রয়েছে, এবং দেশে ব্যাঙ্ক ডাকাতি কমাতে হলে স্কুটার উৎপাদন বন্ধ করে দিতে হবে। এটি কৃত্রিম (spurious) অথবা অর্থহীন (nonsense) সহগতির একটা চমৎকার উদাহরণ। আসল ব্যাপার হল, এই দুটি চলের সঙ্গে তৃতীয় আর একটি চল 'সময়'-এর ধনাত্মক সহগতি রয়েছে (সময়ের অগ্রগতির সাথে দুটোই ক্রমাগত বাড়ছে, এই অর্থে)। দু'টি চল থেকে যদি সময়সূচক চলটির প্রভাব অপনয়ন করা যায় তাহলে দেখা যাবে, অবশিষ্ট অংশ দু'টির মধ্যে আদৌ কোন সহগতি নেই—যা অত্যন্ত স্বাভাবিক। সুতরাং, বিভ্রান্তি দূর করার জন্য অনেক সময় দু'টি চলকে অন্য এক বা একাধিক চলের প্রভাবমুক্ত করে তাদের মধ্যে সহগতির পরিমাণ নির্ধারণের প্রশ্ন উঠতে পারে।

এই দু'টি প্রশ্নই সমাধানের জন্য আমরা আংশিক সহগতির (Partial Correlation) ধারণা কাজে লাগাব। দু'টি চলকে k-টি অন্য চলের প্রভাবমুক্ত করার পর তাদের অবশিষ্ট অংশ দু'টির মধ্যে যে সহগতি থাকে, তাকে বলা হয় চল দু'টির মধ্যে k-মাত্রিক আংশিক সহগতি। এই আংশিক সহগতি শুধু k-এর মানের ওপর নয়, চল দু'টিকে কোন্ কোন্ k-টি চলের প্রভাবমুক্ত করা হল, তার ওপরও নির্ভর করে।

এরপর আমরা আংশিক সহগতি পরিমাপ করার পদ্ধতি নিয়ে আলোচনা করব।

ধরা যাক, আমরা x_1 ও x_2 এই চলদুটিকে $x_3, x_4, \cdots\cdots, x_p$ এই চলগুলির প্রভাবমুক্ত করে এদের মধ্যে আংশিক সহগতির পরিমাপ করতে চাই। $x_{1\cdot 34\cdots\cdots p}$ এবং $x_{2\cdot 34\cdots\cdots p}$ যদি $x_3, x_4, \cdots\cdots, x_p$-র ওপর চল দু'টির নির্ভরণ সমীকরণ হয়, আমরা লিখতে পারি,

$$x_i = X_{i\cdot 34\cdots\cdots p} + x_{i\cdot 34\cdots\cdots p} \qquad \cdots \ (2.36)$$

$$i = 1, 2$$

যেখানে $x_{i\cdot 34\cdots\cdots p}$, $i = 1, 2$ হল চল দু'টির ভ্রান্তি অংশ। আমরা (2·27) -এ দেখেছি x_i এর $x_{i\cdot 34\cdots\cdots p}$ অংশটি $x_3, x_4, \cdots\cdots, x_p$ -র ওপর আদৌ নির্ভরশীল নয়, $i = 1, 2$। সুতরাং, $x_{1\cdot 34\cdots\cdots p}$ এবং $x_{2\cdot 34\cdots\cdots p}$-কে যথাক্রমে x_1 ও x_2'র $x_3, x_4, \cdots\cdots, x_p$ এই চলগুলির প্রভাবমুক্ত অংশ বলা হয়। সেই কারণে,

$$r_{12\cdot 34\cdots\cdots p} = r_{1\cdot 34\cdots\cdots p,\ x_{2\cdot 34\cdots\cdots p}}$$

$$= \frac{\mathrm{Cov}\left(x_{1\cdot 34\cdots\cdots p},\ x_{2\cdot 34\cdots\cdots p}\right)}{\sqrt{V\left(x_{1\cdot 34\cdots\cdots p}\right)}\ \sqrt{V\left(x_{2\cdot 34\cdots\cdots p}\right)}} \qquad \cdots \ (2.37)$$

—এই পূরণী পরিঘাত সহগতি গুণাঙ্কটিকে বলা হয় x_1 ও x_2-কে $(x_3, x_4, \cdots\cdots, x_p)$ এই $k-2$ টি চলের প্রভাবমুক্ত করার পর চল দু'টির মধ্যে $(k-2)$ মাত্রিক আংশিক সহগতি গুণাঙ্ক (Partial Correlation

Coefficient of Order, k – 2)। গুণাঙ্কটিতে মুখ্য নিম্নলগ্নক (Primary suffix) সংখ্যা 1 ও 2 যে দু'টি চলের মধ্যে আংশিক সহগতি নির্ণয় করা হচ্ছে সে দু'টিকে, এবং গৌণ নিম্নলগ্নক (Secondary suffix) সংখ্যা 3, 4, ..., p চল দু'টিকে যে সব অনপেক্ষ চলের প্রভাবমুক্ত করা হয়েছে সেগুলিকে নির্দেশ করছে।

লেখার সুবিধার জন্য আমরা $(x_1, x_2, \cdots\cdots, x_p)$ চলগুলি থেকে x_i-কে বাদ দিলে বাকি $p-1$ টি চলের সহগতি ম্যাট্রিক্সকে $R^{(i)}$, $i = 1, 2$ বলব। অর্থাৎ, $p \times p$ মাত্রার মূল R ম্যাট্রিক্সটি থেকে i-তম সারি এবং (i)-তম স্তম্ভ বাদ দিলে যা পড়ে থাকে, তাই হল $R^{(i)}$, যার মাত্রা $(p-1)\times(p-1)$ ।

এই প্রতীক ব্যবহার করে $X_{i\cdot 34\cdots\cdots p}$ নিরূপণ করার জন্য নর্মাল সমীকরণ হবে,

$$\sum_{\alpha} x_{i\cdot 34\cdots\cdots p\alpha} = 0 \qquad \ldots (2.38)$$

$$\sum_{\alpha} x_{i\cdot 34\cdots\cdots p\alpha} \cdot x_{j\alpha} = 0 \qquad \ldots (2.39)$$

$i = 1, 2\ ;\ j = 3(1)p$

যেখানে $x_{i\cdot 34\cdots\cdots p\alpha} = x_{i\alpha} - X_{i\cdot 34\cdots\cdots p\alpha}$ ... (2.40)

এখন, $x_{i\alpha} - \overline{x}_i = u_{i\alpha}$ এবং $x_{i\cdot 34\cdots\cdots p\alpha} - \overline{x}_{i\cdot 34\cdots\cdots p}$

$$= x_{i\cdot 34\cdots\cdots p\alpha} = u_{i\cdot 34\cdots\cdots p\alpha}$$

লিখে আমরা পাই,

$$x_{1\cdot 34\ldots\ldots p\alpha} = u_{1\alpha} - \frac{R_{13}^{(2)}}{R_{11}^{(2)}} \cdot \frac{S_1}{S_3} u_{3\alpha} - \cdots - \frac{R_{1p}^{(2)}}{R_{11}^{(2)}} \cdot \frac{S_1}{S_p} u_{p\alpha} \qquad \ldots (2.41)$$

এবং $$x_{2\cdot 34\ldots\ldots p\alpha} = u_{2\alpha} - \frac{R_{23}^{(2)}}{R_{22}^{(1)}} \cdot \frac{S_2}{S_3} u_{3\alpha} - \cdots - \frac{R_{2p}^{(2)}}{R_{22}^{(1)}} \cdot \frac{S_2}{S_p} u_{P\alpha} \qquad \ldots (2.42)$$

(2·33) থেকে

$$V\left(x_{1\cdot 34\ldots\ldots p}\right) = \frac{R^{(2)}}{R_{11}^{(2)}}\ S_1^2 \qquad \ldots (2.43)$$

এবং $$V\left(x_{2\cdot 34\ldots\ldots p}\right) = \frac{R^{(1)}}{R_{22}^{(1)}}\ S_2^2 \qquad \ldots (2.44)$$

এখন $n\,\mathrm{Cov}\left(x_{1\cdot34\ldots p}, x_{2\cdot34\ldots p}\right)=\sum\limits_{\alpha} x_{1\cdot34\ldots p\alpha}\; x_{2\cdot34\ldots p\alpha}$

$$=\sum_{\alpha} u_{1\alpha}\cdot u_{2\cdot34\ldots p\alpha} \quad \ldots\ldots(2.39) \text{ থেকে]}$$

$$=\sum_{\alpha} u_{1\alpha}\left\{u_{2\alpha}-\frac{R_{23}^{(1)}}{R_{22}^{(1)}}\cdot\frac{S_2}{S_3}u_{3\alpha}-\ldots-\frac{R_{2p}^{(1)}}{R_{22}^{(1)}}\cdot\frac{S_2}{S_p}\cdot u_{p\alpha}\right\}$$

$$\Rightarrow \mathrm{Cov}\left(x_{1\cdot34\ldots p}, x_{2\cdot34\ldots p}\right)$$

$$=\frac{1}{n}\sum u_{1\alpha}u_{2\alpha}-\frac{R_{23}^{(1)}}{R_{22}^{(1)}}\cdot\frac{S_2}{S_3}\cdot\frac{1}{n}\sum u_{1\alpha}u_{3\alpha}-\ldots-\frac{R_{2p}^{(1)}}{R_{22}^{(1)}}\cdot\frac{S_2}{S_p}\cdot\frac{1}{n}\sum u_{1\alpha}u_{p\alpha}$$

$$=r_{12}S_1S_2+\frac{R_{23}^{(1)}}{R_{22}^{(1)}}\cdot\frac{S_2}{S_3}\cdot r_{13}S_1S_3+\cdots+\frac{R_{2p}^{(1)}}{R_{22}^{(1)}}\cdot\frac{S_2}{S_p}\cdot r_{1p}S_1S_p$$

$$=\frac{S_1S_2}{R_{22}^{(1)}}\left[r_{12}R_{22}^{(1)}+r_{13}R_{23}^{(1)}+\cdots+r_{1p}\cdots R_{2p}^{(1)}\right]$$

কিন্তু $r_{12}R_{22}^{(1)}+r_{13}R_{23}^{(1)}+\cdots+r_{1p}\cdots R_{2p}^{(1)}$

$=\left|R^{(1)}\right|$-এর প্রথম সারি $\left(r_{22}\;\; r_{23}\ldots r_{2p}\right)$-র বদলে $\left(r_{12}\;\; r_{13}\ldots r_{1p}\right)$ বসিয়ে যা পাওয়া যায়,

$$=\begin{vmatrix} r_{12} & r_{13} & \cdots & r_{1p} \\ r_{32} & r_{33} & \cdots & r_{3p} \\ \vdots & \vdots & \vdots & \vdots \\ r_{p2} & r_{p3} & \cdots & r_{pp} \end{vmatrix}$$

$=|R|$ -এ r_{21} -এর মাইনর

$=|R|$ -এ r_{12} -এর মাইনর

$=-R_{12}$

সুতরাং, $Cov\left(x_{1\cdot34....p}, x_{2\cdot34....p}\right)$

$$= -\frac{R_{12}\cdot S_1S_2}{R_{22}^{(1)}} \qquad \text{... (2.45)}$$

এবং $r_{12\cdot34.....p} = -\frac{R_{12}\cdot S_1S_2}{R_{22}^{(1)}} \Bigg/ \left[\frac{R^{(2)}}{R_{11}^{(2)}}\, S_1^2 \cdots \frac{R^{(1)}}{R_{22}^{(1)}}\cdot S_2^2\right]^{\frac{1}{2}}$

$$= -\frac{R_{12}}{\sqrt{R_{11}R_{22}}} \qquad \text{... (2.46)}$$

যেহেতু $R^{(1)} = R_{11}$, $R^{(2)} = R_{22}$ এবং $R_{22}^{(1)} = R_{11}^{(2)}$, কেননা এই দুইটি রাশিই হল $|R|$ থেকে প্রথম দু'টি সারি এবং প্রথম দু'টি স্তম্ভ বাদ দিলে যে নির্ধারকটি পড়ে থাকে তাই।

স্পষ্টতই $-1 \le r_{12\cdot34....p} \le 1$ ... (2.47)

এখন $r_{12\cdot34....p}$-র সংখ্যামান যদি শূন্য অপেক্ষা বেশ খানিকটা বড় হয়, তাহলে বুঝতে হবে x_1-এর নির্ভরণ সমীকরণে x_3, x_4, ..., x_p-র অন্তর্ভুক্তির পরও x_2-কে নতুন করে নির্ভরক হিসাবে নেওয়া যুক্তিযুক্ত হবে। অন্যদিকে, $r_{12\cdot34....p}$'র সংখ্যামান শূন্যের কাছাকাছি হলে বুঝতে হবে নির্ভরণ সমীকরণে x_3, x_4, ..., x_p নেওয়ার পর নতুন করে x_2-কে নিলে বিশেষ লাভ হবে না, অর্থাৎ নির্ভরণ সমীকরণ থেকে x_2-কে বাদ দিলেও পূর্বানুমান সূত্র হিসাবে নির্ভরণ সমীকরণের গুণমান তেমন কমবে না।

এ সম্বন্ধে পরে আরো আলোচনা করা হবে।

(2·46) সূত্রটি লক্ষ্য করলে বোঝা যাবে, $r_{12\cdot34.....p}$ তে মুখ্য নিম্নলগ্নকগুলি এবং গৌণ নিম্নলগ্নকগুলি পরস্পরের মধ্যে পরিবর্তনযোগ্য—অর্থাৎ $r_{12\cdot34} = r_{21\cdot34} = r_{12\cdot43} = r_{21\cdot43}$ ... ...(2.46) -এ নিম্নলগ্নকগুলি পরস্পরের মধ্যে বদল করে সহজেই আমরা একই ধরনের আংশিক সহগতি গুণাংক পেতে পারি, যেমন—

$$r_{13\cdot24.....p} = -\frac{R_{13}}{\sqrt{R_{11}R_{33}}}$$

অথবা, $r_{23\cdot14....p} = -\frac{R_{23}}{\sqrt{R_{22}R_{33}}}$ ইত্যাদি।

$$p = 3 \text{ হলে } |R| = \begin{vmatrix} 1 & r_{12} & r_{13} \\ r_{12} & 1 & r_{23} \\ r_{13} & r_{23} & 1 \end{vmatrix}$$

$$-R_{12} = r_{12} - r_{13}r_{23},\ R_{11} = 1 - r_{23}^2,\ R_{22} = 1 - r_{13}^2$$

$$-R_{13} = r_{13} - r_{12}r_{23}$$

সুতরাং, $$r_{12\cdot3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{1-r_{13}^2}\sqrt{1-r_{23}^2}} \quad \ldots (2.48)$$

এবং $$r_{13\cdot2} = \frac{r_{13} - r_{12}r_{23}}{\sqrt{1-r_{12}^2}\sqrt{1-r_{23}^2}} \quad \ldots (2.49)$$

উদাহরণ ৩.৩ :

উদাহরণ ২.১-এ দেওয়া রাশিতথ্য থেকে আমরা $r_{12\cdot3}$ এবং $r_{13\cdot2}$ নির্ণয় করতে পারি।

$$r_{12\cdot3} = \frac{0\cdot6624 - 0\cdot7514 \times 0\cdot7815}{\sqrt{1-0\cdot7514^2}\sqrt{1-0\cdot7815^2}}$$

$$= 0\cdot1826$$

$$r_{13\cdot2} = \frac{0\cdot7815 - 0\cdot7514 \times 0\cdot6624}{\sqrt{1-0\cdot7514^2}\sqrt{1-0\cdot6624^2}}$$

$$= 0\cdot5741$$

সুতরাং দেখা যাচ্ছে, x_1 সম্বন্ধে পূর্বানুমান দেওয়ার জন্য নির্ভরণ সমীকরণে শুধু x_3 নেওয়ার পর x_2—কেও অন্তর্ভুক্ত করার প্রয়োজন অনেকখানি। আবার দেখা গেছে, এই সমীকরণে শুধু x_2-কে নেওয়ার পর x_3-কেও অন্তর্ভুক্ত করার প্রয়োজন অতখানি না হলেও খানিকটা আছে।

৩৯.৬ আংশিক নির্ভরণ সহগাঙ্ক এবং বহুল ও আংশিক সহগতি গুণাঙ্ক সংক্রান্ত কয়েকটি ফল :

(১) (2·20) থেকে পাওয়া যায়

$$b_{12\cdot34\cdots\cdots p} = -\frac{R_{12}}{R_{11}}\cdot\frac{S_1}{S_2}$$

আবার $S^2_{1\cdot 23\cdots\cdots p} = \frac{R}{R_{11}} \cdot S_1^2$

$$S^2_{2\cdot 13\cdots\cdots p} = \frac{R}{R_{22}} \cdot S_2^2$$

এবং $r_{12\cdot 34\cdots\cdots p} = -\frac{R_{12}}{\sqrt{R_{11}R_{22}}}$

সুতরাং, $b_{12\cdot 34\cdots\cdots p} = r_{12\cdot 34\cdots\cdots p} \cdot \frac{S_{1\cdot 23\cdots\cdots p}}{S_{2\cdot 13\cdots\cdots p}}$ ... (2.50)

(২) আবার $S^2_{1\cdot 34\cdots\cdots p} = \frac{R^{(2)}}{R^{(2)}_{11}} \cdot S_1^2$

$$S^2_{2\cdot 34\cdots\cdots p} = \frac{R^{(1)}}{R^{(1)}_{22}} \cdot S_2^2$$

এবং $R^{(1)} = R_{11},\ R^{(2)} = R_{22},\ R^{(2)}_{11} = R^{(1)}_{22}.$

সুতরাং, $b_{12\cdot 34\cdots\cdots p} = r_{12\cdot 34\cdots\cdots p} \cdot \frac{S_{1\cdot 34\cdots\cdots p}}{S_{2\cdot 34\cdots\cdots p}}$ ... (2.51)

লক্ষ্যণীয়, (2·51) ফলটি দ্বিচল রাশিতথ্যের ক্ষেত্রে পাওয়া $b_{12} = r_{12}\frac{S_1}{S_2}$ ফলটির অনুরূপ। কেবল প্রতিটি পদে গৌণ নিম্নলগ্নক হিসাবে 3, 4, ..., p যোগ করা হয়েছে মাত্র।

একইভাবে দ্বিচল রাশিতথ্যের ক্ষেত্রে পাওয়া ফলে

$b_{12} = \frac{Cov(x_1, x_2)}{V(x_2)}$ -এর অনুরূপে পাই

$$b_{12\cdot 34\cdots\cdots p} = \frac{Cov\left(x_{1\cdot 34\cdots\cdots p}, x_{2\cdot 34\cdots\cdots p}\right)}{V\left(x_{2\cdot 34\cdots\cdots p}\right)} \quad \ldots (2.52)$$

সুতরাং,

$$r^2_{12\cdot34\cdots\cdots p} = b_{12\cdot34\cdots\cdots p} \times b_{21\cdot34\cdots\cdots p} \qquad \dots (2.53)$$

($r_{12.34\dots\dots p}$ -এর গাণিতিক চিহ্ন হবে আংশিক নির্ভরণ সহগাঙ্ক দু'টির সাধারণ গাণিতিক চিহ্ন) যা দ্বিচল রাশিতথ্যের ক্ষেত্রে $r^2 = b_{12} \times b_{21}$-এর সঙ্গে তুলনীয়।

(৩) নর্মাল সমীকরণ (2·38) ও (2·39) থেকে দেখা যাবে, অভিন্ন নিম্নলগ্নক যুক্ত দু'টি ভ্রান্তি অংশের গুণফলের সমষ্টির দু'টি চমৎকার ধর্ম রয়েছে।

$$\sum_\alpha x_{1\cdot34\cdots\cdots p\alpha} \; x_{2\cdot34\cdots\cdots p\alpha}$$

$$= \sum_\alpha x_{1\cdot34\cdots\cdots p\alpha}\left\{u_{2\alpha} - b_{23}\, u_{3\alpha} - \cdots - b_{2p}\, u_{p\alpha}\right\}$$

$$= \sum_\alpha x_{1\cdot34\cdots\cdots p\alpha}\left\{u_{2\alpha} - b_{23}\, u_{3\alpha} - \cdots - b_{2p-1}\, u_{p-1\cdot\alpha}\right\},$$

যেহেতু, (2·38) ও (2·39) থেকে $\sum_\alpha x_{1\cdot34\cdots\cdots p\alpha} u_{p\alpha} = 0$

$$= \sum_\alpha x_{1\cdot34\cdots\cdots p\alpha} \cdot x_{2\cdot34\cdots\cdots p-1\alpha} \qquad \dots (2.54)$$

$$= \dots\dots\dots\dots$$

$$= \sum_\alpha x_{1\cdot34\cdots\cdots p\alpha} \; x_{2\alpha}$$

একইভাবে, $\sum_\alpha x_{1\cdot34\cdots\cdots p\alpha} \; x_{2\cdot34\cdots\cdots p\alpha}$

$$= \sum_\alpha x_{1\cdot34\cdots\cdots p-1\alpha} \; x_{2\cdot34\cdots\cdots p\alpha}$$

$$= \dots\dots\dots\dots$$

$$= \sum_\alpha x_{1\alpha} \; x_{2\cdot34\cdots\cdots p\alpha} \qquad \dots (2.55)$$

তেমনি এই দু'টি নর্মাল সমীকরণের সাহায্যে আমরা পাই,

$$\sum_\alpha x_{1\alpha} \; x_{2\cdot34\cdots\cdots p\alpha}$$

$$= \sum_{\alpha} x_{1\cdot 3\alpha}\ x_{2\cdot 34\cdots\cdots p\alpha}$$

$$= \sum_{\alpha} x_{1\cdot 34\cdots\cdots p\alpha}\ x_{2\cdot 34\cdots\cdots p\alpha}$$

অথবা, $\sum_{\alpha} x_{1\cdot 34\cdots\cdots p\alpha}\ x_{2\alpha}$

$$= \sum_{\alpha} x_{1\cdot 34\cdots\cdots p\alpha}\ x_{2\cdot 3\alpha}$$

$$= \sum_{\alpha} x_{1\cdot 34\cdots\cdots p\alpha}\ x_{2\cdot 34\cdots\cdots p\alpha} \qquad \cdots\ (2.56)$$

অর্থাৎ দেখা যাচ্ছে, $\sum_{\alpha} x_{1\cdot 34\cdots\cdots p\alpha}\ x_{2\cdot 34\cdots\cdots p\alpha}$ -এর মত দু'টি ভ্রান্তি অংশের গুণফলের সমষ্টিতে যদি গৌণ নিম্নলগ্নকগুলি অভিন্ন থাকে, তাহলে ভ্রান্তি অংশ দু'টির যেকোন একটি থেকে এক বা একাধিক, বা এমনকি সবকটি গৌণলগ্নক বাদ দিলেও ফলের কোন পরিবর্তন হয় না।

তেমনই একই যুক্তিতে দেখানো যায়, $\sum_{\alpha} x_{1\alpha}\ x_{2\cdot 34\cdots\cdots p\alpha}$ বা $\sum_{\alpha} x_{1\cdot 34\cdots\cdots p\alpha}\ x_{2\alpha}'$-র মত প্রকাশনে প্রথম ভ্রান্তি অংশটিতে দ্বিতীয় অংশের এক বা একাধিক নিম্নলগ্নক যোগ করলে ফল অপরিবর্তিত থাকে।

(৪) 2·6 (1) অনুচ্ছেদ থেকে পাওয়া যায়,

$$\sum_{\alpha} u^2_{1\cdot 23\cdots\cdots p\alpha} = \sum_{\alpha} u_{1\cdot 23\cdots\cdots p-1\alpha}\ u_{1\cdot 23\cdots\cdots p\alpha}$$

$$= \sum_{\alpha} u_{1\cdot 23\ldots p-1\alpha} \left\{u_{1\alpha} - b_{12\cdot 34\ldots p} u_{2\alpha} - b_{1p\cdot 23\ldots p-1} \cdot u_{p\alpha}\right\}$$

$$= \sum_{\alpha} u_{1\cdot 23\ldots p-1\alpha}\, u_{1\alpha} - b_{1p\cdot 23\ldots p-1} \sum_{\alpha} u_{p\alpha}\, u_{1\cdot 23\ldots p-1\alpha}$$

$$= \sum_{\alpha} u^2_{1\cdot 23\ldots p-1\alpha} - b_{1p\cdot 23\ldots p-1} \sum_{\alpha} u_{1\cdot 23\ldots p-1}\, u_{p\cdot 23\ldots p-1\alpha}$$

$$= \sum_{\alpha} u^2_{1\cdot 23\ldots p-1\alpha} - b_{1p\cdot 23\ldots p-1} \cdot b_{p1\cdot 23\ldots p-1} \sum_{\alpha} u^2_{1\cdot 23\ldots p-1\alpha}$$

$$= \left(1 - b_{1p\cdot 23\ldots p-1} \cdot b_{p1\cdot 23\ldots p-1}\right) \sum_{\alpha} u^2_{1\cdot 23\ldots p-1\alpha} \quad [(2.52)\text{ থেকে}]$$

$$\Rightarrow S^2_{1\cdot23.....p} = \left(1 - r^2_{1p\cdot23....p-1}\right) S^2_{1\cdot23....p-1} \quad ......[(2.53) \text{ থেকে}]$$

এখন যেহেতু $r^2_{1p\cdot23...p-1} \le 1,$ ... (2.57)

$$S^2_{1\cdot23...p} \le S^2_{1\cdot23...p-1} \quad ... (2.58)$$

অর্থাৎ, $\left(1 - r^2_{1\cdot23....p-1}\right) S^2_1 \le \left(1 - r^2_{1\cdot23.....p-1}\right) S^2_1$

$$\text{বা, } r^2_{1\cdot23...p} \ge r^2_{1\cdot23...p-1} \quad ... (2.59)$$

(2·58) বা (2·59) থেকে বোঝা যাচ্ছে, নির্ভরণ সমীকরণে নতুন একটি নির্ভরক চল যোগ করলে ভ্রান্তি ভেদমান সব সময় আগের থেকে কম বা সমান হবে অর্থাৎ, পূর্বানুমান সূত্র হিসাবে নির্ভরণ সমীকরণের গুণমান আরও উন্নত হবে বা একই থাকবে।

(2·57) সূত্রটি ক্রমান্বয়ে $S^2_{1\cdot23...p-1}$, $S^2_{1\cdot23....p-2}$, $S^2_{1.2}$ -এর ওপর প্রয়োগ করে পাওয়া যায়

$$S^2_{1\cdot23....p} = \left(1 - r^2_{1p\cdot23.....p-1}\right)\left(1 - r^2_{1p-1\cdot23....p-2}\right) S^2_{1\cdot23....p-2}$$

$$= \left(1 - r^2_{1p.23.....p-1}\right)\left(1 - r^2_{1p-1\cdot23.....p-2}\right)\cdots\left(1 - r^2_{13.2}\right)\left(1 - r^2_{12}\right) S^2_1$$

$$\text{বা, } \left(1 - r^2_{1.23.....p}\right) S^2_1 = \left(1 - r^2_{1p\cdot23.....p-1}\right)\left(\quad\right)\cdots\left(1 - r^2_{12}\right) S^2_1$$

অথবা, $\left(1 - r^2_{1\cdot23...p}\right) = \left(1 - r^2_{12}\right)\left(1 - r^2_{13\cdot2}\right)\cdots\left(1 - r^2_{1p\cdot23...p-1}\right)$ ... (2.60)

স্পষ্টতই (2·60) সূত্রটি নানাভাবে লেখা যেতে পারে।

যেমন $p = 4$ হলে,

$$\left(1 - r^2_{1\cdot234}\right) = \left(1 - r^2_{12}\right)\left(1 - r^2_{13\cdot2}\right)\left(1 - r^2_{14\cdot23}\right)$$

$$= \left(1 - r^2_{13}\right)\left(1 - r^2_{12\cdot3}\right)\left(1 - r^2_{14\cdot23}\right)$$

$$= \left(1 - r^2_{14}\right)\left(1 - r^2_{13\cdot 4}\right)\left(1 - r^2_{14\cdot 23}\right)$$

$$= \left(1 - r^2_{14}\right)\left(1 - r^2_{12\cdot 4}\right)\left(1 - r^2_{13\cdot 24}\right) \text{....... ইত্যাদি।}$$

এই ফলটি হল বিভিন্ন ক্রমের আংশিক সহগতি গুণাঙ্কের (সাধারণ সহগতি গুণাঙ্ক r_{12}-কে x_1 ও x_2-র মধ্যে শূন্য ক্রমের আংশিক সহগতি গুণাঙ্ক বলা যেতে পারে) সাহায্যে বহুল সহগতি গুণাঙ্কের একটি সূত্র। এই সূত্রটি নির্ভরণ সমীকরণে নতুন একটি নির্ভরক চল সংযুক্তির যৌক্তিকতা আছে কিনা, সে ব্যাপারে সিদ্ধান্ত নিতে সাহায্য করে।

আলোচনার সুবিধার জন্য ধরা যাক $p = 3$, এবং $|r_{12}| > |r_{13}|$ । x_1-এর পূর্বানুমান সূত্র যদি শুধুমাত্র x_2 নেওয়া হয়, তাহলে ভ্রান্তি ভেদমান দাঁড়াবে

$$S^2_{1\cdot 23} = \left(1 - r^2_{1\cdot 23}\right) S^2_1$$

$$= \left(1 - r^2_{12}\right)\left(1 - r^2_{13\cdot 2}\right) S^2_1$$

স্পষ্টতই এই পূর্বানুমান সূত্রে x_2 নেওয়ার পর x_3-কেও অন্তর্ভুক্ত করা যুক্তিযুক্ত হবে যদি $S^2_{1\cdot 23}$ -এর মান $S^2_{1\cdot 2}$ -র থেকে লক্ষ্যণীয় ভাবে কম হয়, অর্থাৎ $r^2_{13\cdot 2}$ -র মান শূন্য থেকে লক্ষ্যণীয়ভাবে বেশি হয়। এই কারণেই আংশিক সহগতি গুণাঙ্ক $r_{13\cdot 2}$-এর সংখ্যামানের ওপর নির্ভর করে এই পূর্বানুমান সূত্রে x_2 নেওয়ার পর অতিরিক্ত নির্ভরক হিসাবে x_3 কেও অন্তর্ভুক্ত করা যুক্তিযুক্ত হবে কিনা। আংশিক সহগতি গুণাঙ্কের এটি একটি বড় উপযোগিতা।

আবার যেহেতু $\left(1 - r^2_{12}\right), \left(1 - r^2_{13 2}\right), \cdots, \left(1 - r^2_{1p 23 \ldots p-1}\right)$ প্রত্যেকের মানই 1 বা 1-এর থেকে কম, (2·60) থেকে সহজেই বলা যায়, $r_{1\cdot 23\ldots\ldots p}$-র সংখ্যামান x_1-এর সঙ্গে অন্য যে কোন চলের যে কোন মাত্রার আংশিক সহগতি গুণাঙ্কের সংখ্যামান থেকে বড় বা সমান হবে।

৩৯·৭ ক্রমিক নির্ভরণ (Step-by-Step Regression)

কোন নির্ভরী চলের জন্য পূর্বানুমাত্র সূত্র নির্ণয় করার বাস্তব সমস্যা হল, সূত্রটিতে কতগুলি এবং কোন্ কোন্ নির্ভরক চল (regressor variable) অন্তর্ভুক্ত করা হবে, সে সম্বন্ধে সিদ্ধান্ত নেওয়া। নির্ভরক চলের সংখ্যা যত কম হবে, বলা বাহুল্য যে, সূত্রটি ব্যবহারের পক্ষে ততই উপযোগী হবে। আবার নির্ভরক চলের সংখ্যা যথাসম্ভব কম রাখার পাশাপাশি পূর্বানুমান সূত্রের গুণমানের দিকেও লক্ষ্য রাখতে হবে, অর্থাৎ দেখতে হবে যেন

সংশ্লিষ্ট বহুল সহগতি গুণাঙ্কের মান 1 -এর যথেষ্ট কাছাকাছি হয়। নির্ভরণ সমীকরণে এক-একটি নতুন নির্ভরক যোগ করা, অথবা ইতিমধ্যে গৃহীত একটি নির্ভরককে বাদ দেওয়ার অর্থ, সমস্ত নির্ভরণ গুণাঙ্কের মান আবার নতুন করে নির্ণয় করা, যা অত্যন্ত পরিশ্রম সাপেক্ষ। সেই কারণে এই ব্যাপারে যথেষ্ট সতর্কতার প্রয়োজন।

বর্তমান যুগে যন্ত্রগণকের বহুল প্রচলনের ফলে অবশ্য এই কাজটি তুলনামূলকভাবে অনেক সহজ হয়ে গেছে। নির্ভরণ সমীকরণকে কীভাবে একটি কার্যকরী পূর্বানুমান সূত্র হিসাবে ধাপে ধাপে গঠন করা হয় সে সম্পর্কে সংক্ষেপে আলোচনা করা হল।

প্রথমে নির্ভরী চল x_1-কে প্রভাবিত করতে পারে, এমন যতগুলি সম্ভব নির্ভরক চল নির্বাচন করা হয় এবং n- সংখ্যক ব্যষ্টির ওপর এই সবগুলি চল (নির্ভরী চল সমেত) সংক্রান্ত তথ্য সংগ্রহ করা হয়। সংগৃহীত তথ্য থেকে সংশ্লিষ্ট সহগতি ম্যাট্রিক্স নির্ণয় করার পর দেখা হয় এই ম্যাট্রিক্স কোন্ j -র জন্য r_{1j} -র সংখ্যামান সবথেকে বেশি। এই j-টিকে 2 দ্বারা নির্দেশ করে প্রথমে x_2-র ওপর x_1-এর নির্ভরণ সমীকরণ গঠন করা হয়। এর পর অন্য চলগুলিকে x_2-র প্রভাবমুক্ত করে $r_{1j\cdot2}$, $j = 3, 4, \cdots$ এই আংশিক সহগতি গুণাঙ্কগুলির মান নির্ণয় করা হয়। এবার দেখা হয় কোন $r_{1j\cdot2}$-র সংখ্যামান সবথেকে বেশি। ধরা যাক, $|r_{13\cdot2}| = \mathrm{M}_j\mathrm{ax}\,|r_{1j\cdot2}|$। তখন রাশিবিজ্ঞান সম্মত পদ্ধতি প্রয়োগ করে বিচার করা হয় সংশ্লিষ্ট পূর্ণক সহগতি গুণাঙ্কটির $(\rho_{13\cdot2})$ মান শূন্য থেকে আলাদা কিনা। আলাদা বলে বিবেচিত হলে x_3-কেনির্ভরণ সমীকরণে অন্তর্ভুক্ত করা হয়। এরপর আবার বাকি নির্ভরণ চলগুলিকে x_2 ও x_3-র প্রভাবমুক্ত করার পর $r_{1j\cdot23}$, $j = 2, 3, \cdots$ এই আংশিক সহগতি গুণাঙ্কগুলির মান নির্ণয় করে এদের মধ্যে গরিষ্ঠ সংখ্যামান বিশিষ্ট গুণাঙ্কটি বাছা হয় এবং আবার দেখা হয় সংশ্লিষ্ট পূর্ণক গুণাঙ্কটি শূন্য থেকে আলাদা কিনা। আলাদা হলে সংশ্লিষ্ট নির্ভরকটিকে নির্ভরণ সমীকরণে অন্তর্ভুক্ত করে আবার আগের মত এগোন হয়। এইভাবে ধাপে ধাপে এগোতে এগোতে যে পর্যায়ে নমুনালব্ধ গরিষ্ঠ সংখ্যামান বিশিষ্ট আংশিক সহগতি গুণাঙ্কর সংশ্লিষ্ট পূর্ণক গুণাঙ্কটির মান রাশিবিজ্ঞান সম্মত বিচারে শূন্য থেকে অভিন্ন মনে হবে, সেইখানেই নির্ভরণ সমীকরণে আর নতুন কোন নির্ভরক যোগ করা অপ্রয়োজনীয় হয়ে পড়বে এবং পূর্বানুমান সূত্রটির গঠন সম্পূর্ণ হয়েছে বুঝতে হবে।

যন্ত্রগণকের ব্যবহারের ফলে এইভাবে ধাপে ধাপে অগ্রসর হয়ে একটি কার্যকরী পূর্বানুমান সূত্র গঠনের কাজটি সহজেই করা যায়।

৩৯·৮ সারাংশ

বর্তমান পরিচ্ছেদে প্রথমে বহুচল রাশিতথ্যের প্রকৃতি এবং এই ধরনের রাশিতথ্য বিশ্লেষণের উদ্দেশ্য নিয়ে আলোচনা করা হয়েছে। এর পর পর্যায়ক্রমে লঘিষ্ঠ বর্গ সমষ্টি পদ্ধতিতে বহুল নির্ভরণ সমীকরণ, বহুল সহগতি গুণাঙ্ক এবং আংশিক সহগতি গুণাঙ্কের প্রসঙ্গ আনা হয়েছে এবং এদের বিভিন্ন ধর্ম নিয়ে বিস্তারিত আলোচনা করা হয়েছে। সবশেষে ক্রমিক নির্ভরণ পদ্ধতি নিয়ে বলা হয়েছে। বহুল নির্ভরণ রেখার সমীকরণটি কীভাবে পূর্বানুমান সূত্র হিসাবে ব্যবহার করা যায় সে সম্বন্ধেও আলোকপাত করা হয়েছে বর্তমান অনুচ্ছেদে।

৩৯·৯ অনুশীলনী

(ক) বড় প্রশ্ন :

১. বহুচল রাশিতথ্য বিশ্লেষণের উপযোগিতা বর্ণনা কর। লঘিষ্ঠ বর্গ সমষ্টি পদ্ধতিকে কীভাবে একপ্রস্থ রাশিতথ্যের ভিত্তিতে বহুল নির্ভরণ সমীকরণ গঠন করা হয়, বর্ণনা করুন।

২. বহুল সহগতি গুণাঙ্কের সংজ্ঞা দাও। একপ্রস্থ বহুচল রাশিতথ্য থেকে কীভাবে গুণাঙ্কটি নির্ধারণ করা হয়, বর্ণনা কর। এই গুণাঙ্কটির উপযোগিতা কী ?

৩. আংশিক সহগতি গুণাঙ্কের সংজ্ঞা দাও। একটি পূরণী পরিঘাত সহগতি গুণাঙ্ক হিসাবে আংশিক সহগতি গুণাঙ্কটি নিরূপণ কর। বহুচল রাশিতথ্য বিশ্লেষণে এই গুণাঙ্কটি কীভাবে কাজে লাগে ?

৪. বহুল সহগতি গুণাঙ্ককে বিভিন্ন ক্রমের আংশিক সহগতি গুণাঙ্কের সাহায্যে প্রকাশ কর। বহুচল রাশিতথ্য বিশ্লেষণে এই ফলটির গুরুত্ব আলোচনা করুন।

(a) $r_{ij} = r \ \forall \ i \neq j$ হলে $r_{1\cdot23.....p}$ এবং $r_{12\cdot34.....p}$ এই গুণাঙ্ক দু'টিকে r-এর সাহায্যে প্রকাশ কর।

(b) k-ক্রমের একটি আংশিক সহগতি গুণাঙ্ককে (k – 1) ক্রমের বিভিন্ন আংশিক সহগতি গুণাঙ্কের সাহায্যে প্রকাশ করুন।

৫. কয়েকজন নবজাতকের ওজন (কেজিতে), প্রসূতির ওজন (কেজিতে), উচ্চতা (মিটারে) এবং গর্ভকালীন ওজন বৃদ্ধি (কেজিতে) সংক্রান্ত তথ্য দেওয়া হল।

ক্রমিক সংখ্যা	নবজাতকের ওজন (কেজিতে)	প্রসূতির ওজন (কেজিতে)	প্রসূতির উচ্চতা (মিটারে)	গর্ভকালীন ওজন বৃদ্ধি (কেজিতে)
ɹɔɥqɒ□□	$x_{1\alpha}$	$x_{2\alpha}$	$x_{3\alpha}$	$x_{4\alpha}$
1	2·28	40	1·60	2·5
2	2·91	42	1·81	3·6
3	3·25	48	1·70	3·8
4	3·50	47	1·72	3·2
5	2·65	42	1·65	2·2
6	2·85	40	1·68	3·4
7	3·65	43	1·61	3·8
8	3·10	46	1·72	3·2
9	2·15	38	1·58	2·2
10	2·60	49	1·72	2·9

এই রাশিতথ্যের ভিত্তিতে নবজাতকের ওজন সম্বন্ধে পূর্বানুমান দেওয়ার জন্য অন্যচলগুলির ওপর এই চলটির নির্ভরণ সমীকরণটি গঠন করুন।

$r_{1\cdot 234}$-এর মান নির্ণয় করে এই পূর্বানুমান সূত্রটির কার্যকারিতার ওপর মন্তব্য করুন।

$r_{12\cdot 4}$ এবং $r_{13\cdot 4}$-এর মান নির্ণয় করে পূর্বানুমান সূত্রে বিভিন্ন নির্ভরক চলের অন্তর্ভুক্তি সম্বন্ধে মন্তব্য করুন।

$r_{12\cdot 34}$-এর মান নির্ণয় করুন।

(খ) সংক্ষিপ্ত প্রশ্ন ঃ

৬. $r_{1\cdot 23.....p}$-কে কেন বহুল নির্ভরণ সমীকরণের পূর্বানুমান সূত্র হিসাবে কার্যকারিতার মাপক বলা হয়?

৭. $b_{12\cdot 34}$-কে আংশিক নির্ভরণ গুণাঙ্ক কেন বলা হয়?

৮. $b_{12\cdot 3} = 1\cdot 8$ হলে $b_{21\cdot 3} = 0\cdot 6$ হতে পারে কি? ব্যাখ্যা করুন।

৯. $r_{1\cdot 234} = -0\cdot 8$ হতে পারে কি? ব্যাখ্যা করুন।

১০. x_1-এর জন্য পূর্বানুমান সূত্র গঠনের উদ্দেশ্যে x_2-কে অন্তর্ভুক্ত করার পর 1-কেও অন্তর্ভুক্ত করা দরকার কিনা, কীভাবে বোঝা যাবে?

১১. $r_{1\cdot 234}$-কে r_{13}, $r_{12\cdot 3}$ এবং $r_{14\cdot 23}$-র সাহায্যে প্রকাশ করুন।

১২. অবান্তর সহগতি কী?

১৩. দেখাও যে, $r_{12} + r_{13} + r_{23} \geq -\frac{3}{2}$

(গ) বস্তুমুখী প্রশ্ন :

১৪. $r_{12} = 0\cdot 8$, $r_{13} = 0\cdot 6$ হলে নিচের কোন্ সংখ্যাটি r_{23}-র মান হিসাবে সমঞ্জস?

(a) $-0\cdot 12$ (b) $-0\cdot 5$ (c) $0\cdot 5$ (d) $0\cdot 6$

১৫. $r_{1\cdot 234} = -0\cdot 6$ হলে নিচের কোন্ মানটি অসমঞ্জস?

(a) $r_{12} = 0\cdot 58$

(b) $r_{13\cdot 2} = 0\cdot 62$

(c) $r_{14\cdot 2} = 0\cdot 53$

(d) $r_{13\cdot 24} = 0\cdot 48$

১৬. পরস্পরের অসমঞ্জস তথ্যগুচ্ছটি চিহ্নিত করুন।

(a) $r_{12} = \cdot 8$, $r_{12\cdot 3} = 0\cdot 85$

(b) $r_{13} = -0\cdot 6$, $r_{12\cdot 3} = -0\cdot 12$

(c) $r_{12} = \cdot6,\ r_{13} = -0\cdot4$

(d) $r_{13} = \cdot7,\ r_{1\cdot23} = 0\cdot21$

১৭. x_1-এর জন্য পূর্বানুমান সূত্রে x_2 ও x_3 অন্তর্ভুক্ত করা হয়েছে। এর পর x_4-এর অন্তর্ভুক্তি যুক্তিযুক্ত কিনা নির্ভর করবে নিচের কোন্‌টির ওপর?

(a) r_{14} (b) $r_{14\cdot2}$ (c) $r_{14\cdot23}$ (d) $r_{23\cdot4}$

১৮. নিচের কোন্ প্রকাশনটি $\left(1-r_{1\cdot234}^2\right)$-এর জন্য যথার্থ?

(a) $\left(1-r_{12}^2\right)\left(1-r_{13\cdot4}^2\right)\left(1-r_{12\cdot34}^2\right)$

(b) $\left(1-r_{13}^2\right)\left(1-r_{12\cdot3}^2\right)\left(1-r_{24\cdot13}^2\right)$

(c) $\left(1-r_{14}^2\right)\left(1-r_{14\cdot3}^2\right)\left(1-r_{12\cdot34}^2\right)$

(d) $\left(1-r_{12}^2\right)\left(1-r_{12\cdot3}^2\right)\left(1-r_{12\cdot34}^2\right)$

১৯. $2x_1 + 3x_2 + 4x_3 = 18$ হলে $r_{12\cdot3} =$

(a) $\frac{4}{2\times3}$ (b) $-\frac{4}{2\times3}$ (c) 1 (d) –1

৩৯·১০ গ্রন্থপঞ্জী

(1) শৈলেন ভূষণ চৌধুরী, অরিজিৎ চৌধুরী ও বিশ্বনাথ দাস প্রণীত, রাশিবিজ্ঞানের মূলতত্ত্ব, প্রথম খণ্ড (পরিচ্ছেদ II), পঃ বঙ্গ রাজ্য পুস্তক পর্ষদ, ১৯৭৬।

(2) Goon, Gupta & Das Gupta : Fundamentals of Statistics, Vol 1 (Ch. 13) ; World Press, 1998.

(3) Gupta & Kapur : Fundamentals of Mathematical Statistics (Ch. 10) ; Sultan Chand & Sons, 1995.

(4) Yule & Kendall : An Introduction to the Theory of Statistics (Ch. 12) : Charles Griffin, 1950.

(5) Ezekiel M & Fox, K. A. : Methods of Correlation & Regression Analysis (Ch: 10–15) : John Wiley ; 1959.

একক ৪০ ❑ সম্ভাবনাতত্ত্বের উপস্থাপনা (Introduction to the Theory of Probability)

গঠন

৪০.০ উদ্দেশ্য

এই এককটি থেকে জানা যাবে—

- সম্ভাবনাতত্ত্ব কাকে বলে
- সম্ভাবনাতত্ত্বের সঙ্গে সংশ্লিষ্ট কিছু ধারণা, যেমন 'সমসম্ভব' পারিসংখ্যানিক পরীক্ষা, ঘটনা, নমুনা দেশ, সম্ভাবনাশ্রয়ী চল, সম্ভাবনা অপেক্ষক, ইত্যাদি
- গুচ্ছতত্ত্বের বিবরণ এবং সম্ভাবনাতত্ত্বের আলোচনায় গুচ্ছতত্ত্বের অবস্থান।

৪০.১ প্রস্তাবনা

'সম্ভাবনা' (probability) শব্দটির সঙ্গে আমরা দৈনন্দিন জীবনে বিশেষভাবে পরিচিত। কোন ঘটনা (event) সম্পর্কে আমাদের ধারণায় অনিশ্চয়তা (uncertainty) থাকলেই সম্ভাবনার প্রশ্ন আসে। কোন একটি বিশেষ ঘটনা (event) ঘটবে কি ঘটবে না, সে বিষয়ে আমরা যখন সুনিশ্চিত হতে পারি না, তখনই অনিশ্চয়তা কাজ করে। অনিশ্চয়তা থাকলে সম্ভাবনার মাধ্যমে সেই অনিশ্চয়তার মাত্রা নির্ণয় করার চেষ্টা করা হয়। সম্ভাবনার এই প্রচলিত ধারণায় অনিশ্চয়তার মাত্রা নির্ণয়ের ক্ষেত্রে কোন ব্যক্তিবিশেষের বিশ্বাসের মাত্রাকেই বোঝায়। এই সম্ভাবনাকে আমরা ব্যক্তিনির্ভর (subjective) সম্ভাবনা বলতে পারি।

প্রচলিত ধারণার বাইরে সংখ্যাতত্ত্বের 'সম্ভাবনা' শব্দটির একটি সুনির্দিষ্ট অর্থ আছে। সংখ্যাতত্ত্বে যে সম্ভাবনার সমস্যা নিয়ে আলোচনা করা হয় তাকে আমরা বস্তুনির্ভর (objective) সম্ভাবনা বলতে পারি। এখানে সম্ভাবনাকে এমন কোন পরীক্ষার (experiment) ফল (outcome) সম্পর্কে বিচার করা হয়, যে পরীক্ষাকে আমরা মোটামুটি একই অবস্থার মধ্যে অসংখ্যবার পুনরাবৃত্তি করতে পারি।

সম্ভাবনাতত্ত্বের সঙ্গে পরিচিত হওয়ার পূর্বশর্ত হিসাবে আমাদের কাছে কিছু প্রযুক্তিগত ধারণা সুস্পষ্ট হওয়া প্রয়োজন। এই কারণেই আমরা আমাদের আলোচ্য বিষয়টিকে কয়েকটি উপবিভাজনের মাধ্যমে অনুধাবনের চেষ্টা করব। সেটি সম্পন্ন হলে সম্ভাবনাতত্ত্বের অবয়ব আমাদের কাছে অনেক সহজ হয়ে উঠবে—এই আশা করা যায়।

৪০.২ সম্ভাবনাতত্ত্ব সংশ্লিষ্ট কিছু ধারণা

৪০.২.১ সমসম্ভব পারিসংখ্যানিক পরীক্ষা (random statistical experiment)

বিজ্ঞান ও কারিগরি চর্চার ক্ষেত্রে 'পরীক্ষা'র গুরুত্ব বিষয়ে আমরা অবগত। এই ধরনের পরীক্ষার মৌলিক বিশেষত্ব হল এই যে, বহুলাংশে একই অবস্থার মধ্যে যদি পরীক্ষাটির অসংখ্যবার পুনরাবৃত্তি করা হয়, তাহলে পরীক্ষার ফল সবসময়ে একই হবে।

কিন্তু কিছু পরীক্ষা আছে যেগুলিকে বহুলাংশে একই অবস্থার মধ্যে অসংখ্যবার পুনরাবৃত্তি করা হলে, প্রতি পুনরাবৃত্তিতে একই ফল নাও পাওয়া যেতে পারে। এই ধরনের পরীক্ষাকে 'সমসম্ভব' পারিসংখ্যানিক পরীক্ষা বলা হয়। যে কোন সমসম্ভব পারিসংখ্যানিক পরীক্ষার মৌলিক বিশেষত্ব হল এই যে, এই পরীক্ষাতে সর্বদাই অনিশ্চয়তার অবকাশ থাকে। এর পর থেকে পারিসংখ্যানিক পরীক্ষা বা পরীক্ষা বলতে আমরা 'সমসম্ভব' পারিসংখ্যানিক পরীক্ষাই বোঝাব।

উদাহরণ :

(১) একটি নিখুঁত (unbiased) ছক্কা (die) চালা হলে পরীক্ষার ফল হবে এই যে, নিম্নলিখিত সংখ্যাগুচ্ছের (set) মধ্যে যে কোন একটি সংখ্যা ছক্কার উপরিভাগে দেখা যাবে। সংখ্যাগুচ্ছটি হল {1,2,3,4,5,6}।

(২) একটি নিখুঁত মুদ্রা (coin) ক্ষেপণ করা হলে, মুদ্রাটির সোজা দিকে হয় 'Head' (H) নয় 'Tail' (T) দেখা যাবে। অর্থাৎ, পরীক্ষার ফল হল নিম্নলিখিত সংখ্যাগুচ্ছের যে কোন একটি। সংখ্যাগুচ্ছটি হল {H,T}।

(৩) একটি নিখুঁত মুদ্রাকে দুবার ক্ষেপণ করা হলে পরীক্ষার ফল হবে {HH, HT, TH, TT}। অর্থাৎ, দুই ক্ষেপণেই H, প্রথম ক্ষেপণে H, পরের ক্ষেপণে T, প্রথম ক্ষেপণে T, পরের ক্ষেপণে H এবং দুই ক্ষেপণেই T।

(৪) আমরা একটি যন্ত্র (machine) দ্বারা বৈদ্যুতিক বাতি (electric bulb) তৈরির কথা বলছি। এই পরীক্ষার ফল হবে এই যে, কিছু বাতি ত্রুটিযুক্ত (defective) হতে পারে, আবার কিছু বাতি ত্রুটিহীন (non-defective) হতে পারে। অতএব, একটি বাতি নির্মিত হলে এটি নিম্নলিখিত গুচ্ছের (set) সদস্য (member) হবে। গুচ্ছটি হল (defective, non-defective)।

স্মর্তব্য : অতঃপর সম্ভাবনাতত্ত্বের আলোচনায় মুদ্রা এবং লুডোর ছক্কার উদাহরণ বারবার উল্লেখিত হবে। আমরা ধরে নেব যে, মুদ্রা বা ছক্কাটি নিখুঁত বা সুনির্মিত। অর্থাৎ, মুদ্রার দুই দিক ও ছক্কার ছ'টি দিক পরস্পর সম্পূর্ণ অনুরূপ।

৪০.২.২ সমসম্ভব/সম-আশংসিত ঘটনা ও নমুনা দেশ (random/equally likely event and sample space)

কোন সমসম্ভব পারিসংখ্যানিক পরীক্ষার সম্ভাব্য ফলকে আমরা সমসম্ভব ঘটনা বলি। সমসম্ভব ঘটনা বলা হয় কারণ, সমস্ত প্রাসঙ্গিক তথ্যাদি হিসাবের মধ্যে নেওয়ার পর, প্রদত্ত অবস্থায় সমস্ত ঘটনাগুলিরই ঘটার সমান সম্ভাবনা থাকে। একে সম-আশংসিত ঘটনাও বলা চলে এইজন্য যে, প্রদত্ত অবস্থায় এই ঘটনাগুলির কোনটিরই ঘটবার ক্ষেত্রে অন্যগুলির থেকে বেশি প্রত্যাশা করা যায় না।

উদাহরণ :

(১) একটি ছক্কা চালা হলে ছক্কাটির ছয়দিকের যে কোন একটি দিকের উপরিভাগে দৃষ্ট হওয়ার উপায়গুলিকে আমরা সম-আশংসিত উপায় বলতে পারি।

(২) একটি মুদ্রা ক্ষেপণ করা হলে, মুদ্রাটির সোজা দিক বা উল্টো দিক পড়ার উপায়গুলিকে আমরা সম-আশংসিত উপায় বলতে পারি।

সমসম্ভব বা সম-আশংসিত ঘটনা মূলত দুই প্রকার হতে পারে। কোন কোন ঘটনা অন্যদের থেকে অপেক্ষাকৃত জটিল ধরনের হয় এবং সেগুলিকে আরও বিভাজনের সাহায্যে সরলীকরণ করা যেতে পারে। যে ঘটনা আর বিভাজন করা যায় না, তাকে আমরা মৌলিক (elementary) ঘটনা বলি। বিভিন্ন মৌলিক ঘটনা নিয়ে একটি ঘটনা গুচ্ছ বা গোষ্ঠী (group বা set of events) তৈরি হয়, যাকে আমরা মিশ্র বা যৌগিক (compound) ঘটনা বলি। কোন পারিসংখ্যানিক পরীক্ষার ফলে যে সমস্ত মৌলিক ঘটনা পাওয়া যায়, তাদের একত্রে নমুনা দেশ (sample space) বলা হয়। পরীক্ষার প্রতি ফলকে/মৌলিক ঘটনাকে নমুনা বিন্দু (sample point) বলা হয়। নমুনা দেশ মৌলিক ঘটনার সংখ্যা অনুসারে সসীম (finite) ও অসীম (infinite) দু'ই হতে পারে। যদি সব স্বাভাবিক রাশিমালার (natural numbers) 1, 2, 3, ... সমাহারে কোন নমুনা দেশ বর্ণিত হয়, তাহলে সেই নমুনা দেশকে আমরা গণনাযোগ্য অসীম (countably infinite) নমুনা দেশ বলব। যদি কোন নমুনা দেশে ঠিক ততগুলি বিন্দু থাকে যতগুলি X-অক্ষ (axis)-এর কোন অন্তরে (interval), যেমন ধরা যাক, $0 \leq x \leq 1$-তে আছে, তাহলে সেই নমুনা দেশকে আমরা গণনাতীত অসীম (non-countably infinite) নমুনা দেশ বলব। যে নমুনা দেশ সসীম বা গণনাযোগ্য অসীম, তাকে আমরা বিচ্ছিন্ন (discrete) নমুনা দেশ বলব। অন্যদিকে গণনাতীত অসীম নমুনা দেশকে আমরা অবিচ্ছিন্ন (continuous) নমুনা দেশ বলব।

স্মর্তব্য : অতঃপর ঘটনা বলতে আমরা মৌলিক ঘটনার গুচ্ছ বা গোষ্ঠীই নিয়েই আলোচনা করব।

৪০.২.৩ পরস্পর বর্জনকারী/ব্যতিরেকী (mutually exclusive) ঘটনা ও পরিপূর্ণ গুচ্ছ গঠনকারী (exhaustive) ঘটনা

দুটি ঘটনাকে পরস্পর বর্জনকারী বা পরস্পর ব্যতিরেকী বলা হয় যদি তাদের মধ্যে কোন একটি ঘটলে অপরটি ঘটতে না পারে। যেমন, একটি মুদ্রা ক্ষেপণ করা হলে, মুদ্রাটির সোজা দিক যদি পড়ে তাহলে উল্টো দিক কখনই পড়তে পারে না। অর্থাৎ, মুদ্রার সোজা দিক পড়া এবং উল্টো দিক পড়া পরস্পর ব্যতিরেকী ঘটনা।

কতকগুলি ঘটনাকে একত্রে একটি পরিপূর্ণ গুচ্ছ গঠনকারী ঘটনা বলা হয় যদি ঐ ঘটনাগুলির মধ্যে অন্তত একটি অবশ্যই ঘটে।

উদাহরণ :

(১) একটি ছক্কা চালা হলে 1, 2, 3, 4, 5, 6 পড়ার ঘটনাগুলি একত্রে একটি পরিপূর্ণ গুচ্ছ গঠন করে। কারণ এই ছয়টি ঘটনার মধ্যে একটি ঘটনা অবশ্যই ঘটবে।

(২) একটি মুদ্রা ক্ষেপণ করা হলে, মুদ্রাটির সোজা দিক পড়া বা উল্টো দিক পড়ার ঘটনা দুটি একত্রে একটি পরিপূর্ণ গুচ্ছ গঠন করে।

উপরিউক্ত উদাহরণ দুটির ঘটনাগুলি পরস্পর বর্জনকারী। পরস্পর বর্জনকারী নয় এমন ঘটনাগুচ্ছ নিয়েও আমরা পরিপূর্ণ গুচ্ছ গঠন করতে পারি। যেমন, কোন ছক্কা ফেললে 2 বা তার বেশি মানের সংখ্যা পাওয়ার ঘটনাগুলিও পরিপূর্ণ গুচ্ছের উদাহরণ।

৪০.২.৪ নিশ্চিত (certain/sure) ঘটনা বনাম অসম্ভব (impossible) ঘটনা

কোন কোন ঘটনা নিশ্চিতভাবে অবশ্যই ঘটে। যে ঘটনার নমুনা দেশে কোন মৌলিক ঘটনা বা ফল থাকে, তাহলে তাকে একটি নিশ্চিত ঘটনা বলা হয়। অপরদিকে, অসম্ভব ঘটনা বলতে বোঝায় এমন একটি ঘটনা, যার নমুনা দেশে বা ফলের গুচ্ছে (outcome set) কোন মৌলিক ঘটনা বা ফল থাকে না, অর্থাৎ গুচ্ছটি শূন্য গুচ্ছ (empty set) হয়। কোন ছক্কা দুবার ফেললে মোট সংখ্যার পরিমাণ 12-র বেশি পাওয়া একটি অসম্ভব ঘটনা।

৪০.৩ সম্ভাবনাতত্ত্বের আলোচনা

৪০.৩.১ সম্ভাবনাতত্ত্বের সংজ্ঞা

সম্ভাবনাতত্ত্ব হল গণিতশাস্ত্রের একটি শাখা যা সমসম্ভব (random) বিষয়সমূহ নিয়ে আলোচনা করে। আগেই বলা হয়েছে যে, একটি সমসম্ভব পরীক্ষায় কোন একটি ঘটনা বাস্তবে ঘটবে কি না, এ বিষয়ে সর্বদাই অনিশ্চয়তা কাজ করে। কোন ঘটনার সম্পর্কে আমাদের অনিশ্চয়তা থাকলে সম্ভাবনার মাধ্যমে সেই অনিশ্চয়তার মাত্রা নির্ণয় করার চেষ্টা করা হয়। কোন একটি ঘটনা ঘটার সম্পর্কে আমরা যে প্রত্যাশা রাখতে পারি, তাকে সেই ঘটনাটির সম্ভাবনা বলা হয়।

চলতি কথায় সম্ভাবনার কোন মান না থাকলেও সংখ্যাতত্ত্বে 'সম্ভাবনা' একটি সুনির্দিষ্ট (definite), এককবিহীন (pure), ধনাত্মক (positive), গাণিতিক (numerical) সংখ্যা। সম্ভাবনার চিহ্ন হল p (probability)।

৪০.৩.২ 'সম্ভাবনা'র মান

সম্ভাবনার সাধারণ মান 0 ও 1 এই দুই সীমার মধ্যে অবস্থান করে অর্থাৎ, p-এর দুই সীমা হল 0 এবং 1। যদি আমরা নিশ্চিত হই যে একটি ঘটনা ঘটবে, তাহলে আমরা বলব যে ঘটনার সম্ভাবনা 1। অপরদিকে যদি আমরা নিশ্চিত হই যে, সমসম্ভাব্য উপায়গুলির কোনটিই ঘটনাটি ঘটার পক্ষে অনুকূল নয়, তাহলে ঘটনাটির সম্ভাবনা হবে 0। উদাহরণ হিসাবে বলা যায় যে, কোন ঘটনার সম্ভাবনা যদি হয় 1/4, তাহলে ঘটনাটি ঘটার সম্ভাবনা থাকে 25% এবং ঘটনাটি না ঘটার সম্ভাবনা থাকে 75%।

যে ঘটনা ঘটার সম্ভাবনা হবে 0 তাকে আমরা বলি অসম্ভব ঘটনা। অপরপক্ষে, সম্ভাবনার মান 1 হওয়ার অর্থ হল যে, সেই ঘটনা সুনিশ্চিতভাবে অবশ্যই ঘটবে। এক কথায় বলা চলে যে, সম্ভাবনার মান 0 বা 1 হলে প্রকৃতপক্ষে সম্ভাবনারই অবলুপ্তি ঘটে, কারণ এখানে কোন অনিশ্চয়তার অবকাশ নেই। ঘটনাটি ঘটা বা না ঘটা সম্পর্কে আমরা সম্পূর্ণ সুনিশ্চিত। এজন্য বাস্তবে সম্ভাবনার মান সাধারণত শূন্য অপেক্ষা বেশি কিন্তু এক অপেক্ষা কম একটি ভগ্নাংশ (fraction) হবে, অর্থাৎ $0 < p < 1$।

৪০.৩.৩ সম্ভাবনা নির্ণয়ের তত্ত্ব

কোন ঘটনার সম্ভাবনার মান নির্ণয় করার তিনটি গুরুত্বপূর্ণ পদ্ধতি আছে। প্রথমটি হল সম্ভাবনার পুরাতন (classical) বা প্রাথমিক (a priori) পদ্ধতি। এটিকে লা প্লাসে'র (La Place) সংজ্ঞাও বলা হয়। দ্বিতীয়টি হল পরিসংখ্যাভিত্তিক (frequency-based) পদ্ধতি। সম্ভাবনার মান নির্ণয়কারী তৃতীয় পদ্ধতিটি হল সম্ভাবনাতত্ত্বের স্বীকার্যভিত্তিক (axiomatic) সংজ্ঞা।

১. পুরাতন বা প্রাথমিক পদ্ধতি :

এই পদ্ধতি অনুসারে সম্ভাবনা হল কোন পরীক্ষার সমসম্ভাব্য বা সম-আশংসিত ফলের মোট সংখ্যার মধ্যে একটি বিশেষ ঘটনা ঘটার অনুকূল (favourable) উপায়গুলির মোট সংখ্যার আনুপাতিক মান। অনুকূল উপায় বলতে আমরা বুঝব যে, এই উপায়গুলির যে কোন একটি ঘটলে ঘটনাটি ঘটবে। আবার ঘটনাটি ঘটলেও এগুলির যে কোন একটি ঘটবে। ধরা যাক, কোন পরীক্ষার মোট সমসম্ভাব্য ফলের সংখ্যা হল N। এর ভেতর, কোন ঘটনা A_1 ঘটার সম্ভাব্য উপায়সমূহ যদি n-সংখ্যক হয়, তাহলে A_1 ঘটার সম্ভাবনা হল n/N। সাঙ্কেতিক অর্থে আমরা লিখি, $P(A_1) = n/N$। এবং, A_1 না ঘটার সম্ভাবনা হল $P(\text{not } A_1) = (N - n)/N = 1 - n/N = 1 - P(A_1)$।

যদি কোন পরীক্ষায় $n = 0$ হয়, তাহলে $P(A_1) = 0/N = 0$; অর্থাৎ A_1 একটি অসম্ভব ঘটনা। অন্যদিকে, যদি $n = N$ হয়, তাহলে $P(A_1) = N/N = 1$; অর্থাৎ A_1 একটি নিশ্চিত ঘটনা।

উদাহরণ :

ধরা যাক, একটি নিখুঁত মুদ্রাকে একবার ক্ষেপণ করলে মুদ্রার উপরিভাগে 'H' দৃষ্ট হওয়ার সম্ভাবনা কত, তা আমরা নির্ণয় করতে চাই। আমরা জানি যে, মুদ্রাটি নিখুঁত হলে দুটি সমসম্ভাব্য ফল হল মুদ্রার উপরিভাগে হয় H নয় T দৃষ্ট হওয়া। এই দুই সম্ভাবনার মধ্যে 'H' কেবল একভাবেই দৃষ্ট হবে। অর্থাৎ একটি নিখুঁত মুদ্রাকে একবার ক্ষেপণ করলে মুদ্রার উপরিভাগে 'H' দৃষ্ট হওয়ার সম্ভাবনা হল $P(H) = 1/2$। এবং, $P(\text{not } H) = (2 - 1)/2 = 1/2$।

ধরা যাক, একটি নিখুঁত ছক্কাকে একবার চালনা করলে ছক্কার উপরিভাগে 3 বা 4 দৃষ্ট হওয়ার ঘটনা হল A_1। ছক্কাটি ছয়টি সম-আশংসিত উপায়ে চালনা করা যায়, এবং উপরিভাগে দৃষ্ট সংখ্যাগুলি হতে পারে 1, 2, 3, 4, 5 অথবা 6। যেহেতু, A_1 দুইটি উপায়ে ঘটতে পারে $P(A_1) = 2/6 = 1/3$। এবং, $P(\text{not } A_1) = (6 - 2)/6 = 4/6 = 2/3$।

২. *পরিসংখ্যাভিত্তিক পদ্ধতি :*

প্রাথমিক পদ্ধতির অসুবিধে হল যে, 'সম-সম্ভাব্য' শব্দটির অর্থ যথেষ্ট স্পষ্ট নয়। এই কারণে কিছু সংখ্যাতত্ত্ববিদ সম্ভাবনার পরিসংখ্যাভিত্তিক (frequency-based) সংজ্ঞা নির্দেশ করেছেন। তাহলে, আমাদের জানতে হবে পরিসংখ্যা (frequency) কাকে বলে।

সাধারণত সংগৃহীত সংখ্যাতাত্ত্বিক রাশিতথ্যের পরিমাণ এত বিপুল হয় যে, সেগুলির সংক্ষেপিকরণ ব্যতিরেকে তথ্যের তাৎপর্য অনুধাবন করা সম্ভব হয় না। পরিসংখ্যা নিবেশনের (frequency distribution) মাধ্যমে সেই সংক্ষেপিকরণ সম্ভব হয়। পরিসংখ্যা (frequency) হল কোন নির্দিষ্ট সময়ে কোন বিষয় বা ঘটনা যত সংখ্যক বার সংঘটিত হয় তার সংখ্যা। সংখ্যাতাত্ত্বিক বিশ্লেষণে পরিসংখ্যাযুক্ত রাশিতথ্য (frequency data) বিশেষভাবে সহায়তা করে। এই প্রকার রাশিতথ্যে প্রতিটি ব্যক্তি বা বস্তুর (individual or item) চরিত্র আলোচনার পরিবর্তে ব্যক্তি বা বস্তু গোষ্ঠীর কোন বিশেষ লক্ষণ আলোচনা করা হয়।

একটি পরীক্ষাকে N বার পুনরাবৃত্তি করা হল, যেখানে N একটি বড় সংখ্যা। যদি এই পুনরাবৃত্তিতে দেখা যায় যে, কোন একটি ঘটনা A_1, n বার ঘটেছে (অর্থাৎ A_1-এর পরিসংখ্যা হল n), তাহলে $P(A_1) = n/N$। অর্থাৎ, n/N হল আপেক্ষিক পরিসংখ্যা। পরিসংখ্যাভিত্তিক পদ্ধতিতে এই আপেক্ষিক পরিসংখ্যাকেই আমরা পরীক্ষামূলক সম্ভাবনা (empirical probability) বলি।

উদাহরণ :

একটি মুদ্রাকে 1000 বার ক্ষেপণ করলে যদি 532 বার 'H' মুদ্রার উপরিভাগে দৃষ্ট হয়, তাহলে বলা যায় যে 'H'-এর সম্ভাবনা 532/1000 ; অর্থাৎ $P(H) = 532/1000 = 0.532$।

৩. *সম্ভাবনাতত্ত্বের স্বীকার্যভিত্তিক পদ্ধতি :*

পরিসংখ্যাভিত্তিক পদ্ধতির অসুবিধে হল যে, 'N'-এর মান কত হলে 'N' একটি বড় সংখ্যা হিসাবে পরিগণিত হবে তা আমাদের কাছে অস্পষ্ট। পুরাতন সংজ্ঞার সীমিত প্রয়োগ ও বিবিধ ত্রুটির জন্য সম্ভাবনাতত্ত্বের আরও উৎকৃষ্টতর ব্যাখ্যা বা সংজ্ঞা দেওয়ার জন্য বহু গবেষণা হয়েছে। এদের মধ্যে আধুনিক সংখ্যাতত্ত্ববিদেরা গুচ্ছতত্ত্বের ভিত্তিতে সম্ভাবনাতত্ত্ব আলোচনা করছেন, যে পদ্ধতিকে বলা হয় সম্ভাবনাতত্ত্বের স্বীকার্যভিত্তিক সংজ্ঞা। স্বীকার্যভিত্তিক সংজ্ঞাতে সম্ভাবনার ক্ষেত্রে প্রথমেই কয়েকটি স্বীকার্য গ্রহণ করা হয়েছে। এগুলি হল :

স্বীকার্য ১ : যে কোন পরীক্ষার ক্ষেত্রে কোন ঘটনা A-এর জন্য একটি নির্দিষ্ট সংখ্যা P(A)পাওয়া যায়, যে সংখ্যাটিকে A ঘটনার সম্ভাবনা বলা হবে।

স্বীকার্য ২ : যে কোন ঘটনা A-এর ক্ষেত্রে, $P(A) \geq 0$।

স্বীকার্য ৩ : A যদি কোন নিশ্চিত ঘটনা হয়, তবে, $P(A) = 1$।

স্বীকার্য ৪ : $A_1, A_2, A_3, \ldots\ldots\ldots\ldots$ ঘটনাগুলি যদি যৌথভাবে পরস্পর বর্জনকারী হয়, তবে, $P(A_1 \cup A_2 \cup A_3 \cup \ldots\ldots) = P(A_1)+P(A_2)+P(A_3)+ \ldots\ldots\ldots\ldots\ldots\ldots = 1$।

বাস্তবক্ষেত্রে বহুল ব্যবহারের উপযোগী ও স্ববিরোধিতামুক্ত একটি সম্ভাবনাতত্ত্বে উপনীত হতে গেলে যে সমস্ত স্বীকার্যের প্রয়োজন হয় সেগুলির প্রকৃতি যথাসম্ভব সরল রাখা ও তাদের ব্যাখ্যা যথাসম্ভব কম রাখার উদ্দেশ্যেই এই বিশেষ কয়েকটি লক্ষণের প্রতি গুরুত্ব দেওয়া হয়েছে।

এই স্বীকার্যগুলিকে স্বতঃসিদ্ধ হিসাবে গ্রহণ করে সম্ভাবনাতত্ত্বের ক্ষেত্রে বিভিন্ন উপপাদ্য ও অনুসিদ্ধান্ত প্রমাণ করা যায়। মনে রাখতে হবে যে, এখানেও পুরাতন সংজ্ঞার মতোই সম্ভাবনাকে দীর্ঘকালীন আনুপাতিক পরিসংখ্যা হিসারে ব্যাখ্যা করা হয়। প্রকৃতপক্ষে গৃহীত স্বীকার্যগুলি এই পরিসংখ্যার কয়েকটি প্রধান লক্ষণের দিকে দৃষ্টি রেখেই প্রণীত হয়েছে। একথা বলা চলে যে, সম্ভাবনাতত্ত্বের স্বীকার্যভিত্তিক ব্যাখ্যার ফলে বাস্তবক্ষেত্রে এর বহুল প্রয়োগ করা চলে এবং সেজন্য পুরাতন সংজ্ঞার চেয়ে এই ব্যাখ্যাকে উৎকৃষ্টতর বলে বিবেচনা করা হয়।

আগেই উল্লেখ করা হয়েছে যে, ঘটনা বলতে আমরা মৌলিক ঘটনার গুচ্ছ বা গোষ্ঠী বোঝাব। 'ঘটনা'র ধারণা সুস্পষ্ট করতে গেলে এবং তার মাধ্যমে সম্ভাবনাতত্ত্বের স্বীকার্যভিত্তিক সংজ্ঞা অনুধাবন করতে হলে প্রাথমিক শর্ত হল গুচ্ছতত্ত্ব (set theory) অনুশীলন করা। নিম্নলিখিত বিভাগে গুচ্ছতত্ত্বের আলোচনা করা হল।

৪০.৪ গুচ্ছতত্ত্বের মৌলিক কিছু ধারণা

৪০.৪.১ 'গুচ্ছ' শব্দটির ধারণা

বস্তুত সম্ভাবনাতত্ত্ব এবং সংখ্যাতত্ত্ব তথা গণিতশাস্ত্রের ভিত্তিতে গুচ্ছতত্ত্বের অবস্থান। সম্ভাবনাতত্ত্বের ভাষায় গুচ্ছকে বলা যায় কিছু ঘটনার সমাহার। এই ঘটনাগুলিকে বলব গুচ্ছের সদস্য বা উপাদান (element)। সাধারণভাবে, গুচ্ছকে নির্দেশ করা হয় বড় হাতের অক্ষরে, যেমন A, B, C, D ইত্যাদি। তেমনি গুচ্ছের উপাদানকে নির্দেশ করা হয় ছোট হাতের অক্ষরে, যেমন a, b, c, d ইত্যাদি।

যদি কোন উপাদান a কোন গুচ্ছ C-এর অন্তর্গত হয়, আমরা লিখি $a \in C$। যদি a, C-এর অন্তর্গত না হয়, আমরা লিখি $a \notin C$। যদি a এবং b, C-এর অন্তর্গত হয়, আমরা লিখি $a, b \in C$। আমরা গুচ্ছতত্ত্বের আলোচনায় সর্বদাই ধরে নেব (make an assumption) যে, গুচ্ছটি সুনির্দিষ্ট (well-defined) অর্থাৎ, কোন বিশেষ উপাদান একটি গুচ্ছের অন্তর্গত কি না, তা আমরা সঠিকভাবে নিরূপণ করতে পারব।

একটি গুচ্ছের সংজ্ঞা নির্দেশিত হতে পারে তার সমস্ত উপাদানগুলির নথিকরণ করে। যদি তা সম্ভব না হয়, তাহলে গুচ্ছটির সব সদস্যগুলির কোন সাধারণ বৈশিষ্ট্য নির্দেশ করেও একটি গুচ্ছের সংজ্ঞা নির্দিষ্ট করা যায়। প্রথমটিকে বলা হয় নথিকরণ পদ্ধতি (roster method) এবং দ্বিতীয়টিকে বলা হয় বৈশিষ্ট্য পদ্ধতি (property method)।

উদাহরণ :

(১) একটি ছক্কা চালা হলে যে সম্ভাব্য সংখ্যাগুলি ছক্কার উপরিভাগে দৃষ্ট হতে পারে, সেগুলি হল {1,2,3,4,5,6}। এক্ষেত্রে গুচ্ছটির সংজ্ঞা নির্দেশ করতে আমরা নথিকরণ পদ্ধতির আশ্রয় নিয়েছি।

(২) আমরা একটি যন্ত্র (machine) দ্বারা বৈদ্যুতিক বাতি (electric bulb) তৈরির কথা বলছি। এই নির্মাণ প্রক্রিয়ায় কিছু বাতি ত্রুটিযুক্ত (defective) হতে পারে, আবার কিছু বাতি ত্রুটিহীন (non-defective) হতে পারে। অতএব, একটি বাতি নির্মিত হলে এটি নিম্নলিখিত গুচ্ছের (set) সদস্য (member) হবে। গুচ্ছটি হল (defective, non-defective)। এক্ষেত্রে গুচ্ছটির সংজ্ঞা নির্দেশ করতে আমরা বৈশিষ্ট্য পদ্ধতির আশ্রয় নিয়েছি।

(৩) ইংরাজী বর্ণমালার সবকটি স্বরবর্ণের (vowel) গুচ্ছকে দুইটি পদ্ধতির মাধ্যমেই নির্দেশ করা যায়।

(ক) নথিকরণ পদ্ধতির মাধ্যমে—{a, e, i, o, u}।

(খ) বৈশিষ্ট্য পদ্ধতির মাধ্যমে—{x |x is a vowel}, অর্থাৎ, x উপাদানের গুচ্ছ যেখানে x একটি স্বরবর্ণ।

৪০.৪.২ উপগুচ্ছ (subset)

যদি কোন গুচ্ছ A-এর প্রতিটি উপাদান, গুচ্ছ B-এরও অন্তর্গত হয়, তাহলে বলা হয় যে, A, B-এর উপগুচ্ছ। এটিকে আমরা লিখি A   B (A, B-এর অন্তর্গত) অথবা $B \supset A$ (B, A-কে ধারণ করে)। এর ভিত্তিতে বলা যায়, যে কোন গুচ্ছ A-এর ক্ষেত্রে A   A।

যদি $A \subset B$ এবং $B \subset A$, আমরা বলি A এবং B সমান এবং লিখি $A = B$। এক্ষেত্রে A এবং B গুচ্ছ দুটির অন্তর্গত উপাদানও সম্পূর্ণ এক।

যদি A এবং B সমান না হয়, অর্থাৎ যদি A এবং B-এর অন্তর্গত উপাদান সম্পূর্ণ এক না হয়, আমরা লিখি $A \neq B$।

যদি $A \subset B$ কিন্তু $A \neq B$, আমরা বলি যে, A হল B-এর একটি উপযুক্ত (proper) উপগুচ্ছ।

উদাহরণ :

(১) {a, i, u} হল {a, e, i, o, u}-এর একটি উপযুক্ত (proper) উপগুচ্ছ।

(২) {i, o, a, u, e} হল {a, e, i, o, u}-এর একটি উপগুচ্ছ কিন্তু উপযুক্ত উপগুচ্ছ নয়, কারণ দুটি গুচ্ছ সমান।

স্মর্তব্য : একটি গুচ্ছের উপাদানগুলিকে এক রেখে যদি কেবল উপাদানের পুনর্বিন্যাস (rearrangement) করা হয়, তাহলে সেই গুচ্ছের কোন পরিবর্তন হয় না, অর্থাৎ গুচ্ছ দুটি সমান থাকে।

(২) একটি ছক্কা চালা হলে যে সম্ভাব্য 'জোড়' সংখ্যাগুলি ছক্কার উপরিভাগে দৃষ্ট হতে পারে, সেগুলির গুচ্ছ হল {2,4,6}। এই গুচ্ছটি হল একটি ছক্কা চালনার সব সাম্ভাব্য ফলের গুচ্ছের, অর্থাৎ {1,2,3,4,5,6}-এর একটি উপযুক্ত উপগুচ্ছ।

৪০.৪.৩ সামগ্রিক গুচ্ছ (universal set) এবং শূন্য গুচ্ছ (empty set)

নানান কারণে আমরা আমাদের আলোচনাকে একটি সমগ্রক (universe) বা সামগ্রিক গুচ্ছের অন্তর্গত বিভিন্ন উপগুচ্ছতে সীমাবদ্ধ রাখি। এই সমগ্রক বা সামগ্রিক গুচ্ছকে 'দেশ' (space)-ও বলা হয় এবং তার উপাদানগুলিকে দেশের বিন্দু (points) বলা হয়। Ω চিহ্নটি দিয়ে একটি দেশকে চিহ্নিত করা হয় এবং তার অন্তর্ভুক্ত বিন্দুরাশিকে ω চিহ্নটি দিয়ে চিহ্নিত করা হয়।

যে গুচ্ছের কোন উপাদানই থাকে না, তাকে শূন্য গুচ্ছ বলা হয়। Φ চিহ্নটি দিয়ে শূন্য গুচ্ছকে চিহ্নিত করা হয়।

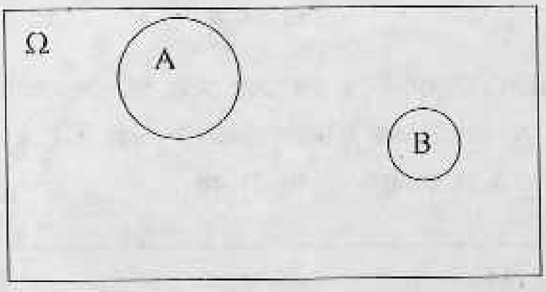
স্মর্তব্য : শূন্য গুচ্ছ Φ হল যে কোন গুচ্ছের একটি উপগুচ্ছ।

উদাহরণ :

(১) একটি ছক্কা চালনা করলে সব সম্ভাব্য ফলের গুচ্ছ—{1,2,3,4,5,6}-কেই আমরা বলি সমগ্রক বা সামগ্রিক গুচ্ছ। একটি নির্দিষ্ট ছক্কা চালনায় 7 বা 11 সংখ্যা উপরিভাগে দৃষ্ট হতে পারে—এই উপাদান সম্বলিত গুচ্ছটি অর্থাৎ {7, 11} হল একটি শূন্য গুচ্ছ।

৪০.৪.৪ ভেণ চিত্র (Venn Diagram)

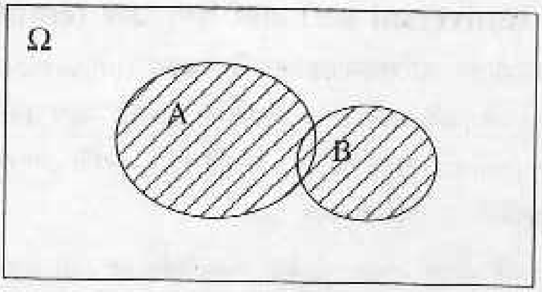
একটি সমগ্রক (universe) বা সামগ্রিক গুচ্ছ বা দেশ Ω-কে আমরা জ্যামিতিক উপায়ে একটি বর্গক্ষেত্রের অন্তর্গত বিন্দুরাশির দ্বারা চিত্রায়িত করতে পারি। এই ক্ষেত্রে Ω-এর উপগুচ্ছ A বা B-কে আমরা বর্গক্ষেত্রের অন্তর্গত বৃত্তের অন্তর্ভুক্ত বিন্দুরাশির দ্বারা প্রকাশ করে থাকি। এইরূপ চিত্রকে বলা হয় ভেণ চিত্র। এর মাধ্যমে আমরা বিভিন্ন গুচ্ছের মধ্যে সম্ভাব্য সম্পর্কের জ্যামিতিক রূপ বিষয়ে অবহিত হতে পারি। নিম্নগঠিত চিত্র ৪০.১ এইরূপ একটি ভেণ চিত্র।



চিত্র ৪০.১

৪০.৪.৫ গুচ্ছের কার্যকলাপ (Set Operations)

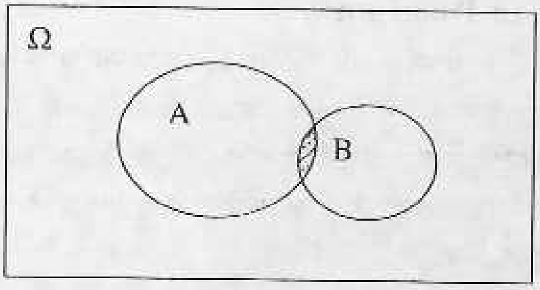
১. **সংযোগ (Union) :** যে উপাদান বা বিন্দুরাশির গুচ্ছটি হয় গুচ্ছ A বা গুচ্ছ B বা দুটিরই অন্তর্গত, সেই গুচ্ছটিকে আমরা A এবং B গুচ্ছ দুটির সংযোগ বলি এবং $A \cup B$ দ্বারা চিহ্নিত করি (চিত্র ৪০.২-এর ধূসরীকৃত এলাকা)।



চিত্র ৪০.২

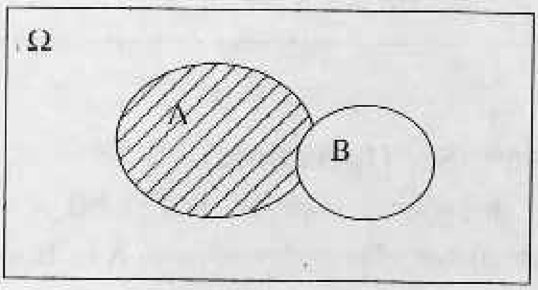
২. ছেদ (Intersection] : যে উপাদান বা বিন্দুর গুচ্ছটি, গুচ্ছ A এবং গুচ্ছ B, দুটিরই অন্তর্গত, সেই গুচ্ছটিকে আমরা A এবং B গুচ্ছ দুটির ছেদ বলি এবং $A \cap B$ দ্বারা চিহ্নিত করি (চিত্র ৪০.৩-এর ধূসরীকৃত এলাকা)।

স্মর্তব্য : দুটি গুচ্ছ A এবং B-কে পরস্পর বিচ্ছিন্ন গুচ্ছ বলা হয় যদি তাদের মধ্যে এমন কোন উপাদান না থাকে, যেটি দুটি গুচ্ছেই বিদ্যমান। A এবং B পরস্পর বিচ্ছিন্ন হলে, আমরা লিখি $A \cap B = \emptyset$। চিত্র ৪০.১ ভেণ চিত্রে A এবং B পরস্পর বিচ্ছিন্ন দুটি গুচ্ছ।



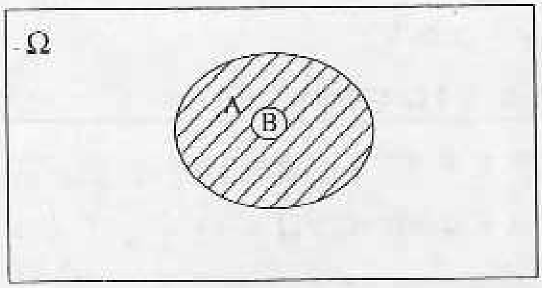
চিত্র ৪০.৩

৩. পার্থক্য (Difference) : যে গুচ্ছটিতে A গুচ্ছের সেইসব উপাদান আছে যেগুলি B গুচ্ছটির অন্তর্গত নয়, সেই গুচ্ছটিকে আমরা A এবং B গুচ্ছ দুটির পার্থক্য বলি এবং A-B দ্বারা চিহ্নিত করি (চিত্র ৪০.৪-এর ধূসরীকৃত এলাকা)। গুচ্ছ A-B বোঝায় A কিন্তু B নয়।

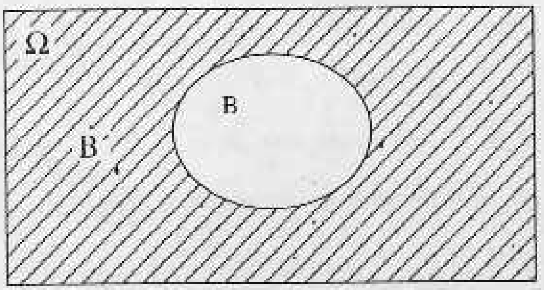


চিত্র ৪০.৪

৪. **সম্পূরক (Complement) :** যদি $B \subset A$ হয়, অর্থাৎ গুচ্ছ B, গুচ্ছ A-এর অন্তর্গত হয়, তাহলে A-B গুচ্ছটিকে বলা হয় (A গুচ্ছটির অপেক্ষায়) B গুচ্ছটির সম্পূরক ; এই গুচ্ছটিকে আমরা B'_A দ্বারা চিহ্নিত করি (চিত্র ৪০.৫-এর ধূসরীকৃত এলাকা)। যদি $A = \Omega$ হয়, আমরা $\Omega - B$ গুচ্ছটিকে বলি B গুচ্ছটির সম্পূরক এবং B' দ্বারা চিহ্নিত করি (চিত্র ৪০.৬-এর ধূসরীকৃত এলাকা)।



চিত্র ৪০.৫



চিত্র ৪০.৬

৪০.৪.৬ গুচ্ছসংক্রান্ত কয়েকটি উপপাদ্য (Theorem)

উপপাদ্য ১.১ : যে কোন গুচ্ছরাজি A, B, C-এর জন্য এই উপপাদ্যটি স্বতঃসিদ্ধ। যদি $A \subset B$ হয় এবং $B \subset C$ হয়, তাহলে $A \subset C$ হতে হবে।

উপপাদ্য ১.২ : সংযোগের স্থানান্তরযোগ্যতা সূত্র (Commutative law for unions)

$A \cup B = B \cup A$।

উপপাদ্য ১.৩ : সংযোগের সংস্রবযোগ্যতা সূত্র (Associative law for unions)

$A \cup (B \cup C) = (A \cup B) \cup C = A \cup B \cup C$।

উপপাদ্য ১.৪ : ছেদের স্থানান্তরযোগ্যতা সূত্র (Commutative law for intersections)

$A \cap B = B \cap A$।

উপপাদ্য ১.৫ : ছেদের সংস্রবযোগ্যতা সূত্র (Associative law for intersections)

$A \cap (B \cap C) = (A \cap B) \cap C = A \cap B \cap C$।

উপপাদ্য ১.৬ : প্রথম নিবেশনযোগ্যতা সূত্র (First Distributive law)

$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$।

উপপাদ্য ১.৭ : দ্বিতীয় নিবেশনযোগ্যতা সূত্র

$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$।

উপপাদ্য ১.৮ : $A - B = A \cap B'$।

উপপাদ্য ১.৯ : যদি $A \subset B$ হয়, তাহলে $A' \supset B'$ এবং $B' \subset A'$ হতে হবে।

উপপাদ্য ১.১০ : $A \cup \Phi = A$ এবং $A \cap \Phi = \Phi$।

উপপাদ্য ১.১১ : $A \cup \Omega = \Omega$ এবং $A \cap \Omega = A$।

উপপাদ্য ১.১২ক : ডি মর্গানের প্রথম সূত্র

$(A \cup B)' = A' \cap B'$।

উপপাদ্য ১.১২খ : ডি মর্গানের দ্বিতীয় সূত্র

$(A \cap B)' = A' \cup B'$।

উপপাদ্য ১.১৩ : কোন গুচ্ছ A এবং B-এর জন্য $A = (A \cap B) \cup (A \cap B)'$।

স্মর্তব্য : উপপাদ্য ১.১২ক, ১.১২খ এবং ১.১৩-কে বিস্তৃত (generalise) করা যায়।

৪০.৪.৭ দ্বৈতবাদের নীতি

গুচ্ছসংক্রান্ত যে কোন স্বতঃসিদ্ধ সিদ্ধান্ত অপরিবর্তিত থাকে যদি আমরা সংযোগ চিহ্নের স্থানে ছেদ চিহ্ন, ছেদের স্থানে সংযোগ চিহ্ন, গুচ্ছের স্থানে সম্পূরক চিহ্ন বসাই বা অন্তর্ভুক্তির প্রতীক $\subset$, $\supset$-কে পারস্পরিকভাবে পরিবর্তিত করি।

৪০.৫ সম্ভাবনাতত্ত্বের স্বীকার্যভিত্তিক সংজ্ঞাতে গুচ্ছতত্ত্বের প্রয়োগ

যেহেতু ঘটনা বলতে আমরা মৌলিক ঘটনার 'গুচ্ছ' বা গোষ্ঠী বোঝাই, সেই হেতু ঘটনাসংক্রান্ত কোন বক্তব্যকে আমরা গুচ্ছতত্ত্বের ভাষায় পরিবর্তিত করে নিতে পারি। অতএব, গুচ্ছতত্ত্বের ক্ষেত্রে উল্লেখিত সূত্রগুলি অনুসারে আমরা ঘটনা বিষয়ক কিছু সূত্র নির্ধারণ করতে পারি।

বিভিন্ন ঘটনাকে A_1, A_2, A_3, ... ইত্যাদি প্রতীকের দ্বারা চিহ্নিত করলে বিভিন্ন ঘটনার মধ্যে কতকগুলি প্রক্রিয়ার সাহায্যে নতুন ঘটনা পাওয়া যায় । এই প্রক্রিয়াগুলি আমরা সংযোগ চিহ্ন ($\cup$), ছেদ চিহ্ন ($\cap$), পার্থক্য চিহ্ন (–), সম্পূরক চিহ্ন (A ইত্যাদি কোন ঘটনাবোধক চিহ্নের উপর '), প্রভৃতি ব্যবহার করে নির্দেশ করতে পারি ।

৪০.৫.১ ঘটনাকেন্দ্রিক প্রক্রিয়াসমূহ (Event Operations)

সংযোগ : $A_1 \cup A_2$-এর অর্থ হল, যদি A_1 এবং A_2 ঘটনা দুটি পরস্পর বিচ্ছিন্ন বা বর্জনকারী না হয়, তাহলে হয় A_1, নয় A_2, নয় A_1 এবং A_2, এই দুটি ঘটনা একসঙ্গে ঘটবে। তেমনি k-টি ঘটনা $A_1, A_2, \ldots\ldots\ldots\ldots, A_k$-এর ক্ষেত্রে $A_1 \cup A_2 \cup \ldots\ldots\ldots\ldots \cup A_k$-এর অর্থ হল, $A_1, A_2, \ldots\ldots\ldots\ldots, A_k$ প্রভৃতি ঘটনাগুলির মধ্যে অন্তত একটি ঘটা।

আবার, A_1 এবং A_2 ঘটনা দুটি যদি পরস্পর বিচ্ছিন্ন বা বর্জনকারী হয়, তাহলে $A_1 \cup A_2$-এর অর্থ হবে হয় A_1, নয় A_2 ঘটা। একইভাবে $A_1 \cup A_2 \cup \ldots\ldots\ldots \cup A_k$-এর অর্থ হল, $A_1, A_2, \ldots\ldots\ldots\ldots A_k$ প্রভৃতি পরস্পর বর্জনকারী ঘটনাগুচ্ছের মধ্যে যে কোন একটি ঘটা।

উদাহরণ :

(১) একটি নিখুঁত মুদ্রা ক্ষেপণ করা হলে, মুদ্রাটির সোজা দিকে পড়ার ঘটনাকে যদি বলে A_1 এবং উল্টো দিকে পড়ার ঘটনাকে বলি A_2 তাহলে $A_1 \cup A_2$ হল হয় মুদ্রাটির সোজা দিকে পড়া নয় উল্টো দিকে পড়ার ঘটনা (যেহেতু ঘটনা দুটি এখানে পরস্পর বর্জনকারী)।

(২) একটি নিখুঁত ছক্কা চালা হলে বেজোড় সংখ্যা পাওয়ার ঘটনাকে যদি বলি A_1 এবং 3-এর গুণনীয়ক পাওয়ার ঘটনাকে যদি বলি A_2, তাহলে $A_1 \cup A_2$ হল হয় বেজোড় সংখ্যা এবং/বা 3-এর গুণনীয়ক পাওয়ার ঘটনা।

ছেদ : $A_1 \cap A_2$-এর অর্থ হল A_1 এবং A_2 ঘটনাদুটির একই সঙ্গে ঘটা। তেমনি k-টি ঘটনা $A_1, A_2, \ldots\ldots\ldots\ldots, A_k$-এর ক্ষেত্রে $A_1 \cap A_2 \cap \ldots\ldots\ldots\ldots \cap A_k$-এর অর্থ হল $A_1, A_2, \ldots\ldots\ldots\ldots A_k$ প্রভৃতি ঘটনাগুলি একই সঙ্গে ঘটা।

উদাহরণ :

(১) একটি নিখুঁত ছক্কা চালা হলে বেজোড় সংখ্যা পাওয়ার ঘটনাকে যদি বলি A_1 এবং 3-এর গুণনীয়ক পাওয়ার ঘটনাকে যদি বলি A_2, তাহলে $A_1 \cap A_2$ ঘটনাটি বোঝায় একইসঙ্গে বেজোড় সংখ্যা এবং 3-এর গুণনীয়ক পাওয়ার ঘটনা। অর্থাৎ, এক্ষেত্রে $A_1 \cap A_2$-এর একমাত্র ফল হল 3।

পার্থক্য : $A_1 - A_2$ হল A_2-এর থেকে A_1-এর পার্থক্য। কেবল A_1 ঘটবে, সেইসঙ্গে A_2 ঘটবে না। তেমনি $A_2 - A_1$ হল A_1-এর থেকে A_2-এর পার্থক্য। কেবল A_2 ঘটবে, সেইসঙ্গে A_1 ঘটবে না।

উদাহরণ :

(১) একটি নিখুঁত ছক্কা চালা হলে বেজোড় সংখ্যা পাওয়ার ঘটনাকে বলি A_1 এবং 3-এর গুণনীয়ক পাওয়ার ঘটনাকে বলি A_2। এখানে A_1 ঘটনার সম্ভাব্য ফল হয় 1, 3, 5। A_2 ঘটনার সম্ভাব্য ফল হয় 3, 6। এক্ষেত্রে $A_1 - A_1$ হল 1, 5। অর্থাৎ, এই উদাহরণে, বেজোড় সংখ্যা পাওয়া গেলেও (A_1 ঘটলেও), 3-এর গুণনীয়ক পাওয়া যাবে না (A_2 পাওয়া যাবে না)।

সম্পূরক : A_1' প্রতীক দ্বারা আমরা A_1-এর না ঘটা বুঝব এবং A_1'-কে A_1-এর সম্পূরক বা বিপরীত ঘটনা বলে অভিহিত করব।

উদাহরণ :

(১) একটি নিখুঁত ছক্কা চালা হলে বেজোড় সংখ্যা পাওয়ার ঘটনাকে যদি বলি A_1, তাহলে A_1' ঘটনাটি বেজোড় সংখ্যা না পাওয়া অর্থাৎ, জোড় সংখ্যা পাওয়া বোঝাবে।

৪০.৫.২ ঘটনাসংক্রান্ত কয়েকটি বৈশিষ্ট্য

স্থানান্তরযোগ্যতা বৈশিষ্ট্য (Commutativity characteristic)

- $A_1 \cup A_2 = A_2 \cup A_1$
- $A_1 \cap A_2 = A_2 \cap A_1$

সংস্রবযোগ্যতা বৈশিষ্ট্য (Associativity characteristic)

- $A_1 \cup (A_2 \cup A_3) = (A_1 \cup A_2) \cup A_3$
- $A_1 \cap (A_2 \cap A_3) = (A_1 \cap A_2) \cap A_3$

নিবেশনযোগ্যতা বৈশিষ্ট্য (Distributivity characteristic)

- $A_1 \cap (A_2 \cup A_3) = (A_1 \cap A_2) \cup (A_1 \cap A_3)$

অপরিবর্তনীয়তা বৈশিষ্ট্য (Idempotency characteristic)

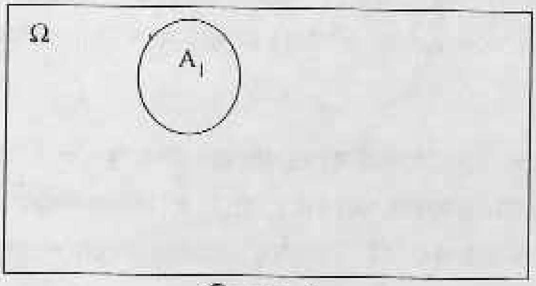
- $A_1 \cup A_1 = A_1$
- $A_1 \cap A_1 = A_1$

উপরিউক্ত চারটি বৈশিষ্ট্য ছাড়াও সম্পূরণ প্রক্রিয়ার ওপর ভিত্তি করে ঘটনার আরও তিনটি বৈশিষ্ট্য লক্ষ করা যায়।

- $(A_1')' = A_1$
- $(A_1 \cup A_2)' = A_1' \cap A_2'$
- $(A_1 \cap A_2)' = A_1' \cup A_2'$

৪০.৫.৩ সম্ভাবনাতত্ত্বে ভেণচিত্রের ব্যবহার

ধরে নিই যে, নিম্নের ভেণচিত্রে বর্গক্ষেত্র Ω দেখায় কোন 'সমসম্ভব' পারিসংখ্যানিক পরীক্ষার মোট পুনরাবৃত্তির সংখ্যা এবং তার ভিতরে চিত্রিত বৃত্ত A_1 সেই পরীক্ষার সম্ভাব্য ফল বা সমসম্ভব ঘটনা দেখায়।



চিত্র ৪০.৭

A_1' -কে বলা হয় A_1 ঘটনার সম্পূরক। $P(A_1') = (N - n) / N = 1 - n/N = 1 - P(A_1)$। অতএব, যে কোন পরীক্ষাতে দুটি পারস্পরিকভাবে বিচ্ছিন্ন সম্ভাবনা আছে—হয় ঘটনা A_1 ঘটবে নয় ঘটনা A_1'-ঘটবে। অর্থাৎ, A_1 এবং A_1' ঘটনা দুটি পারস্পরিকভাবে বিচ্ছিন্ন/পরস্পর বর্জনকারী বা পরস্পর ব্যতিরেকী—তারা একইসাথে ঘটতে পারবে না। ভেণচিত্রে দুটি ঘটনা পারস্পরিকভাবে বিচ্ছিন্ন হলে তাদের মধ্যে কোন অংশীদারি ক্ষেত্র (common area) থাকবে না। A_1 এবং A_1' দুটি পরিপূর্ণ গুচ্ছ গঠনকারী ঘটনা। $P(A_1) + P(A_1') = 1$।

এবার ধরা যাক, কোন 'সমসম্ভব' পারিসংখ্যানিক পরীক্ষার দুটি সম্ভাব্য ফল বা সমসম্ভব ঘটনা হল A_1 এবং A_2। A_1 ঘটে n বার এবং A_2 ঘটে m বার। A_1 এবং A_2 একইসাথে ঘটছে r বার। এই দুটি ঘটনার জন্য আমরা একটি সারণি চিত্র নির্মাণ করলাম।

সারণি ৪০.১

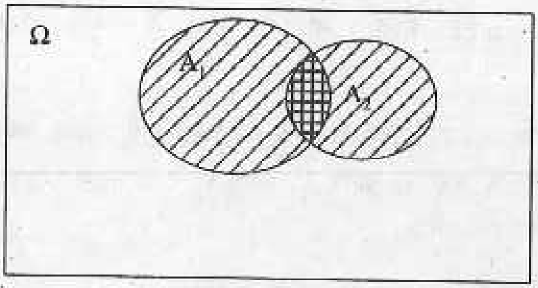
	A_2 ঘটে	A_2 ঘটে না	মোট
A_1 ঘটে	r	n–r	n
A_1 ঘটে না	m–r	N–m–n+r	N–n
মোট	m	N–m	N

ভেণচিত্র ৪০.৮-এ দুটি ঘটনা আছে A_1 এবং A_2।

$$P(A_1) = n / N$$
$$P(A_1') = (N - n) / N$$
$$P(A_2) = m / N$$
$$P(A_2') = (N - m) / N$$
$$P(A_1 \cap A_2) = r / N$$
$$P(A_1 \cap A_2') = (n - r) / N = n/N - r/N = P(A_1) - P(A_1 \cap A_2)$$
$$P(A_1' \cap A_2) = (m - r) / N = m/N - r/N = P(A_2) - P(A_1 \cap A_2)$$
$$P(A_1' \cap A_2') = (N-n-m+r) / N = 1 - n/N - m/N + r/N = 1 - P(A_1) - P(A_2) + P(A_1 \cap A_2)$$

চিত্র ৪০.৮-এর ধূসরীকৃত এলাকা $A_1 \cup A_2$ ঘটনাটি বোঝায় এবং খোপকাটা এলাকাটি $A_1 \cap A_2$ এলাকাটি বোঝায়। $A_1 \cup A_2$ ঘটনাটি তিনটি এলাকার যোগফল—A_1-এর সেই অংশ যা A_2-এর বহির্ভূত + A_2-এর সেই অংশ যা A_1-এর বহির্ভূত + A_1 এবং A_2-এর যৌথ অংশ। অতএব, $P(A_1 \cup A_2) = P(A_1 \cap A_2') + P(A_1' \cap A_2) + P(A_1 \cap A_2) = (n - r) / N + (m - r) / N + r/N = (n + m - r) / N = n/N + m/N - r/N = P(A_1) + P(A_2) - P(A_1 \cap A_2)$।

আমরা দেখেছি যে, $P(A_1' \cap A_2') = 1 - P(A_1) - P(A_2) + P(A_1 \cap A_2) = 1 - P(A_1 \cup A_2)$ । অর্থাৎ, $P(A_1 \cup A_2) + P(A_1' \cap A_2') = 1$ । অতএব, $(A_1 \cup A_2)$ এবং $(A_1' \cap A_2')$ দুটি পরস্পর ব্যতিরেকী ঘটনা বোঝায়।



চিত্র ৪০.৮

৪০.৬ সম্ভাবনাশ্রয়ী চল (random/stochastic variable)

কোন রাশিতথ্যের সংখ্যাগত লক্ষণকে চল বা চলক (variable) বলা হয়। চল আবার মূলত দুই প্রকার— গাণিতিক (numeric) চল এবং সম্ভাবনাশ্রয়ী চল। সম্ভাবনাতত্ত্বের আলোচনায় যে চলের ব্যবহার হয় তা হল সম্ভাবনাশ্রয়ী চল। সম্ভাবনাশ্রয়ী চল হল এমন এক শ্রেণীর চল যা নির্দিষ্ট সম্ভাবনাযুক্ত নির্দিষ্ট ও প্রকৃত মানসমূহ গ্রহণ করে। ধরা যাক, একটি নমুনা দেশের প্রতিটি বিন্দুকে আমরা একএকটি সংখ্যার দ্বারা সূচিত করলাম। যে নমুনা দেশের সম্পর্কে সম্ভাবনাশ্রয়ী চলকে ব্যাখ্যা করা হয়, সেই নমুনা দেশের সঙ্গে যে সমস্ত পরীক্ষা যুক্ত থাকে তার ফলসমূহ অনিশ্চিত ও সম্ভাবনার ওপর নির্ভরশীল। সম্ভাবনাশ্রয়ী চলকে সূচিত করি বড় হাতের অক্ষরে X, Y, Z ইত্যাদি প্রতীকের মাধ্যমে এবং তাদের গৃহীত মানগুলিকে x, y, z ইত্যাদি চিহ্ন দিয়ে।

সম্ভাবনাশ্রয়ী চল দুই ধরনের হয়—বিচ্ছিন্ন (discrete) এবং অবিচ্ছিন্ন (continuous)। কোন সম্ভাবনাশ্রয়ী চল X যদি x_i $(i = 1, 2, \ldots\ldots, r, \ldots\ldots)$ বিচ্ছিন্ন মানসমূহ গ্রহণ করতে পারে এবং প্রতিটি মানের ক্ষেত্রে $X = x_i$ হওয়ার ঘটনাটির যদি একটি নির্দিষ্ট সম্ভাবনা $P(X = x_i) = p_i\ (\geq 0)$ পাওয়া যায়, তবে আমরা X-কে বিচ্ছিন্ন সম্ভাবনাশ্রয়ী চল বলব। এই বিচ্ছিন্ন মানসকল সসীম বা অসীম, দুই হতে পারে। অপরদিকে, কোন সম্ভাবনাশ্রয়ী চল যদি কোন নির্দিষ্ট অন্তর (a, b)-এর ভেতর অবিচ্ছিন্নভাবে যে কোন মান গ্রহণ করতে পারে এবং ঐ অন্তরের (limit) অন্তর্ভুক্ত কোন অন্তর (α, β)-র মধ্যে ঐ চলের মান গ্রহণ করার অর্থাৎ $(a \leq \alpha \leq x \leq \beta \leq b)$ হওয়ার একটি নির্দিষ্ট সম্ভাবনা পাওয়া যায়, তবে আমরা X-কে অবিচ্ছিন্ন সম্ভাবনাশ্রয়ী চল বলব।

এই এককে সম্ভাবনাশ্রয়ী চলের ধারণাটির সাথে পাঠকদের কেবল পরিচিত করানো হল। তৃতীয় এককে এই ধারণার সম্প্রসারণ করা হবে।

৪০.৭ সারাংশ

১. সম্ভাবনাতত্ত্ব হল গণিতশাস্ত্রের একটি শাখা যা সমসম্ভব বিষয়সমূহ নিয়ে আলোচনা করে। কোন ঘটনার সম্পর্কে আমাদের অনিশ্চয়তা থাকলে সম্ভাবনার মাধ্যমে সেই অনিশ্চয়তার মাত্রা নির্ণয় করার চেষ্টা করা হয়। কোন একটি ঘটনা ঘটার সম্পর্কে আমরা যে প্রত্যাশা রাখতে পারি, তাকে সেই ঘটনাটির সম্ভাবনা বলা হয়।

২. সংখ্যাতত্ত্বে আমরা বস্তুনির্ভর সম্ভাবনার বিষয় নিয়ে আলোচনা করে থাকি। এক্ষেত্রে সম্ভাবনাকে এমন কোন পরীক্ষার ফল সম্পর্কে বিচার করা হয় যে পরীক্ষাকে আমরা মোটামুটি একই অবস্থার মধ্যে অসংখ্যবার পুনরাবৃত্তি করতে পারি এবং প্রতি পুনরাবৃত্তিতে একই ফল নাও পাওয়া যেতে পারে। এই ধরনের পরীক্ষাকে 'সমসম্ভব' পারিসংখ্যানিক পরীক্ষা বলা হয়। যে কোন সমসম্ভব পারিসংখ্যানিক পরীক্ষার মৌলিক বিশেষত্ব হল এই যে, এই পরীক্ষাতে সর্বদাই অনিশ্চয়তার অবকাশ থাকে।

৩. কোন সমসম্ভব পারিসংখ্যানিক পরীক্ষার সম্ভাব্য ফলকে আমরা সমসম্ভব ঘটনা বলি। সমসম্ভব ঘটনা বলা হয় কারণ সমস্ত প্রাসঙ্গিক তথ্যাদি হিসাবের মধ্যে নেওয়ার পর, প্রদত্ত অবস্থায় সমস্ত ঘটনাগুলিরই ঘটার সমান সম্ভাবনা থাকে। একে সম-আশংসিত ঘটনাও বলা চলে এইজন্য যে, প্রদত্ত অবস্থায় এই ঘটনাগুলির কোনটিরই ঘটবার ক্ষেত্রে অন্যগুলির থেকে বেশি প্রত্যাশা করা যায় না। সমসম্ভব বা সম-আশংসিত ঘটনা মূলত দুই প্রকার হতে পারে—অবিভাজ্য বা মৌলিক ঘটনা ও মিশ্র বা যৌগিক ঘটনা। সমস্ত মৌলিক ঘটনার সমাহারকে নমুনা দেশ বলা হয় এবং পরীক্ষার প্রতি ফলকে/মৌলিক ঘটনাকে নমুনা বিন্দু বলা হয়। মৌলিক ঘটনার সংখ্যা অনুসারে নমুনা দেশ সসীম ও অসীম—দু'ই হতে পারে। দুটি ঘটনাকে পরস্পর বর্জনকারী বা পরস্পর ব্যতিরেকী বলা হয়, যদি তাদের মধ্যে কোন একটি ঘটলে অপরটি ঘটতে না পারে। কতকগুলি ঘটনাকে একত্রে একটি পরিপূর্ণ গুচ্ছ গঠনকারী ঘটনা বলা হয় যদি ঐ ঘটনাগুলির মধ্যে অন্তত একটি অবশ্যই ঘটে। কোন ঘটনার নমুনা দেশে মৌলিক ঘটনা বা ফল থাকলে তাকে একটি নিশ্চিত ঘটনা বলা হয়। অপরদিকে, অসম্ভব ঘটনা বলতে বোঝায় এমন একটি ঘটনা, যার নমুনা দেশে বা ফলের গুচ্ছে কোন মৌলিক ঘটনা বা ফল থাকে না, অর্থাৎ গুচ্ছটি শূন্য গুচ্ছ হয়।

৪. সংখ্যাতত্ত্বে 'সম্ভাবনা' একটি সুনির্দিষ্ট, এককবিহীন, ধনাত্মক, গাণিতিক সংখ্যা, যার চিহ্ন হল p। সম্ভাবনার সাধারণ মান 0 ও 1 এই দুই সীমার মধ্যে অবস্থান করে অর্থাৎ, p-এর দুই সীমা হল 0 এবং 1। যদি আমরা নিশ্চিত হই যে, একটি ঘটনা ঘটবে, তাহলে আমরা বলব যে ঘটনার সম্ভাবনা 1, অর্থাৎ ঘটনাটি একটি নিশ্চিত ঘটনা। অপরদিকে যদি আমরা নিশ্চিত হই যে, সমসম্ভাব্য উপায়গুলির কোনটিই ঘটনাটি ঘটার পক্ষে অনুকূল নয়, তাহলে ঘটনাটির সম্ভাবনা হবে 0, অর্থাৎ ঘটনাটি একটি অসম্ভব ঘটনা। সম্ভাবনার মান সাধারণত শূন্য অপেক্ষা বেশি কিন্তু এক অপেক্ষা কম একটি ভগ্নাংশ হবে, অর্থাৎ $0 < p < 1$।

৫. কোন ঘটনার সম্ভাবনার মান নির্ণয় করার তিনটি গুরুত্বপূর্ণ পদ্ধতি আছে—পুরাতন বা প্রাথমিক পদ্ধতি, যেটিকে লা প্লাসে'র সংজ্ঞাও বলা হয়, পরিসংখ্যাভিত্তিক পদ্ধতি এবং সম্ভাবনাতত্ত্বের স্বতঃসিদ্ধ সংজ্ঞাভিত্তিক

পদ্ধতি। পুরাতন পদ্ধতি এবং পরিসংখ্যাভিত্তিক পদ্ধতি দুইয়েরই নিজস্ব সীমাবদ্ধতা আছে। তার জন্যেই সংখ্যাতত্ত্ববিদেরা গুচ্ছতত্ত্বের ভিত্তিতে সম্ভাবনাতত্ত্বে আলোচনা করছেন, যে পদ্ধতিকে বলা হয় সম্ভাবনাতত্ত্বের স্বতঃসিদ্ধ সংজ্ঞাভিত্তিক পদ্ধতি। সম্ভাবনাতত্ত্বে 'ঘটনা'র ধারণা সুস্পষ্ট করতে গেলে এবং তার মাধ্যমে সম্ভাবনাতত্ত্বের স্বতঃসিদ্ধ সংজ্ঞাভিত্তিক পদ্ধতি অনুধাবন করতে হলে প্রাথমিক শর্ত হল গুচ্ছতত্ত্ব অনুশীলন করা। গুচ্ছতত্ত্ব প্রসঙ্গে আমরা ভেণ চিত্র এবং গুচ্ছতত্ত্বের কয়েকটি উপপাদ্য আলোচনা করেছি। অতঃপর আমরা গুচ্ছতত্ত্বের মাধ্যমে সম্ভাবনাতত্ত্বে ঘটনাকেন্দ্রিক প্রক্রিয়া ও ঘটনাসংক্রান্ত কয়েকটি বৈশিষ্ট্য এবং সম্ভাবনাতত্ত্বে ভেণ চিত্রের প্রয়োগ বিবৃত করেছি। সবশেষে এই এককে আমরা 'সম্ভাবনাশ্রয়ী চল' ধারণাটির অবতারণা করলাম, যেটি তৃতীয় এককে বিস্তৃতরূপে আলোচিত হবে।

৪০.৮ অনুশীলনী

১. নিম্নলিখিত প্রতিটি ঘটনার জন্য সম্ভাবনা নিরূপণ করুন।

ক. একটি নিখুঁত ছক্কাকে একবার চালা হলে ছক্কার উপরিভাগে বেজোড় সংখ্যা দেখা যায়।

খ. একটি নিখুঁত মুদ্রাকে দুইবার ক্ষেপণ করলে মুদ্রার উপরিভাগে অন্তত একটি 'H দৃষ্ট হয়।

গ. যথেষ্ট মেশানো মোট 52টি তাসযুক্ত একটি তাসের প্যাকেট থেকে একটি টেক্কা রুহিতনের 10 অথবা ইস্কাবনের 2 টানা হয়।

ঘ. এক জোড়া নিখুঁত ছক্কাকে একবার চালা হলে ছক্কার উপরিভাগে দৃষ্ট দুটি সংখ্যার যোগফল হল 7।

ঙ. যদি একটি মুদ্রাকে 100 বার ক্ষেপণ করলে 56 বার H দৃষ্ট হয়; তাহলে মুদ্রার পরবর্তী ক্ষেপণে T দৃষ্ট হবে।

২. ধরা যাক, দুটি নিখুঁত ছক্কাকে N বার চালা হল। দুটি ছক্কার উপরিভাগে প্রাপ্ত সংখ্যা দুটির যোগফল হল X। নিম্নলিখিত ঘটনাগুলির সম্ভাবনা নিরূপণ করুন।

ক. X 2 দ্বারা বিভাজ্য ; খ. X 3 দ্বারা বিভাজ্য ; গ. X 4 দ্বারা বিভাজ্য ; ঘ. X 5 দ্বারা বিভাজ্য ; ঙ. X 6 দ্বারা বিভাজ্য ; চ. X 2 এবং 5 দ্বারা বিভাজ্য ; ছ. X 2 বা 5 দ্বারা বিভাজ্য।

৩. একটি বাক্সে ছয়টি লাল বল, চারটি সাদা বল এবং পাঁচটি নীল বল আছে। বাক্স থেকে একটি বল এলোপাতাড়িভাবে তোলা হল। সেই বলটি (ক) লাল, (খ) সাদা, (গ) নীল, (ঘ) লাল নয় এবং (ঙ) লাল অথবা সাদা—এই ঘটনাগুলির সম্ভাবনা নিরূপণ করুন।

৪. ধরা যাক, একটি নিখুঁত ছক্কাকে দুইবার চালা হল। ছক্কার উপরিভাগে প্রাপ্ত সংখ্যাটি অন্তত একবার 4 হবে, সেই ঘটনার সম্ভাবনা নিরূপণ করুন।

৪০.৯ উত্তরমালা

১ক. ছয়টি সম-সম্ভাব্য, সম-আশংসিত ক্ষেত্রের তিনটি ক্ষেত্র (যেখানে ছক্কার উপরিভাগে বেজোড় সংখ্যা 1, 3 অথবা 5 দেখা যায়), আলোচ্য ঘটনার অনুকূল। অতএব, সম্ভাবনার মান হল $p = 3/6 = 1/2$।

খ. একটি নিখুঁত মুদ্রাকে দুইবার ক্ষেপণ করলে সমসম্ভাব্য ফলগুলি হল HH, HT, TH ও TT। কেবল প্রথম তিনটি ফল আলোচ্য ঘটনার অনুকূল। অতএব, সম্ভাবনার মান হল $p = 3/4$।

গ. মোট 52টি তাসযুক্ত একটি তাসের প্যাকেট থেকে আলোচ্য ঘটনাটি (ইস্কাবনের টেক্কা, হরতনের টেক্কা, চিড়িতনের টেক্কা, রুহিতনের টেক্কা, রুহিতনের 10 এবং ইস্কাবনের 2) ছয়ভাবে ঘটতে পারে। অতএব, সম্ভাবনার মান হল $p = 6/52 = 3/26$।

ঘ. একটি ছক্কার ছয়টি দিক আছে। কোন একটি ছক্কার উপরিভাগের প্রতিটি অপর একটি ছক্কার ছয়টি উপরিভাগের প্রতিটির সাথে এমনভাবে সংশ্লিষ্ট হতে পারে, যাতে মোট 6 গুণ 6 = 36টি সমসম্ভাব্য ঘটনা ঘটতে পারে। এগুলি হল—

(1,1), (2,1), (3,1), (4,1), (5,1), (6,1),

(1,2), (2,2), (3,2), (4,2), (5,2), (6,2),

(1,3), (2,3), (3,3), (4,3), (5,3), (6,3),

(1,4), (2,4), (3,4), (4,4), (5,4), (6,4),

(1,5), (2,5), (3,5), (4,5), (5,5), (6,5),

(1,6), (2,6), (3,6), (4,6), (5,6), (6,6)

এই সম্ভাব্য ঘটনাগুলির মধ্যে ছক্কার উপরিভাগে দৃষ্ট দুটি সংখ্যার যোগফল 7 যে ছয়ভাবে পাওয়া যায়, সেগুলি হল (1, 6), (2, 5), (3, 4), (4, 3), (5, 2) এবং (6, 1)। অতএব, সম্ভাবনার মান হল $p = 6/36 = 1/6$।

ঙ. একটি মুদ্রাকে 100 বার ক্ষেপণ করলে যদি 56 বার H দৃষ্ট হয়, তাহলে এই 100 বার ক্ষেপণে $100 - 56 = 44$ বার T দৃষ্ট হতে পারে। অতএব, T-এর সম্ভাবনার মান হল $44/100 = 0.44$।

২. X-এর মান 2 থেকে 12-এর মধ্যে হতে পারে।

X	সম্ভাব্য ঘটনা	সংখ্যা
2	(1,1)	1
3	(1,2), (2,1)	2
4	(1,3), (3,1), (2,2)	3
5	(1,4), (4,1), (2,3), (3,2)	4
6	(1,5), (5,1), (3,3), (4,2), (2,4)	5
7	(1,6), (6,1), (2,5), (5,2), (4,3), (3,4)	6
8	(2,6), (6,2), (3,5), (5,3), (4,4)	5
9	(3,6), (6,3), (4,5), (5,4)	4
10	(4,6), (6,4), (5,5)	3
11	(5,6), (6,5)	2
12	(6,6)	1
মোট		36

অতএব, P(X=2) = 1/36, P(X=3) = 2/36, P(X=4) = 3/36, P(X=5) = 4/36, P(X=6) = 5/36, P(X=7) = 6/36, P(X=8) = 5/36, P(X=9) = 4/36, P(X=10) = 3/36, P(X=11) = 2/36, P(X=12) = 1/36

ক. P(X,2 দ্বারা বিভাজ্য) = P(X=2, 4, 6, 8, 10, 12) = P(X=2) + P(X=4) + P(X=6) + P(X=8) + P(X=10) + P(X=12) = 1/36 + 3/36 + 5/36 + 5/36 + 3/36 + 1/36 = 18/36 = 1/2

খ. P(X, 3 দ্বারা বিভাজ্য) = P(X=3, 6, 9, 12) = P(X=3) + P(X=6) + P(X=9) + P(X=12) = 2/36 + 5/36 + 4/36 + 1/36 = 12/36 = 1/3

গ. P(X, 4 দ্বারা বিভাজ্য) = P(X=4, 8, 12) = P(X=4) + P(X=8) + P(X=12) = 3/36 + 5/36 + 1/36 = 9/36 = 1/4

ঘ. P(X, 5 দ্বারা বিভাজ্য) = P(X=5, 10) = P(X=5) + P(X=10) = 4/36 + 3/36 = 7/36

ঙ. P(X, 6 দ্বারা বিভাজ্য) = P(X=6, 12) = P(X=6) + P(X=12) = 5/36 + 1/36 = 6/36 = 1/6

চ. P(X, 2 এবং 5 দ্বারা বিভাজ্য) = P(X=10) = 3/36 = 1/12

ছ. P(X, 2 বা 5 দ্বারা বিভাজ্য) = P(X, 2 দ্বারা বিভাজ্য) + P(X, 5 দ্বারা বিভাজ্য) – P(X, 2 এবং 5 দ্বারা বিভাজ্য) = P(X=2, 4, 6, 8, 10, 12) + P(X=5, 10) – P(X=10) = 1/2 + 7/36 – 1/12 = 18/36 + 7/36 – 3/36 = 22/36 = 11/18

৩. ধরে নিই, R, W এবং B যথাক্রমে লাল বল টানা, সাদা বল টানা এবং নীল বল টানার ঘটনাসমূহকে বোঝায়।

ক. $P(R)$ = (লাল বল টানার উপায়)/(যে কোন একটি বল টানার মোট উপায়) $= 6/(6+4+5) = 6/15 = 2/5$।

খ. $P(W) = 4/15$

গ. $P(B) = 5/15 = 1/3$

ঘ. $P(R') = 1 - P(R) = 1 - 2/5 = 3/5$

ঙ. $P(R \cup W) = P(R) + P(W) = 6/15 + 4/15 = 10/15 = 2/3$ (যেহেতু, লাল বল টানা এবং সাদা বল টানা, দুটি পরস্পর বর্জনকারী ঘটনা)।

৪. ধরা যাক, "প্রথম চালনায় 4" এই ঘটনাটি হল A_1, "দ্বিতীয় চালনায় 4" এই ঘটনাটি হল A_2 এবং $A_1 \cup A_2$ হল "প্রথম চালনায় 4 বা দ্বিতীয় চালনায় 4 বা দুই চালনাতেই 4, অর্থাৎ, ছক্কার উপরিভাগে অন্তত একবার 4 সংখ্যাটি প্রাপ্ত হবে, সেই ঘটনাটি। আমাদের নির্ণয় করতে হবে $P(A_1 \cup A_2)$।

যেহেতু, A_1 এবং A_2 দুটি পরস্পর বর্জনকারী ঘটনা নয়, $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$। যেহেতু, A_1 এবং A_2 দুটি পরস্পর স্বতন্ত্র ঘটনা, $P(A_1 \cap A_2) = P(A_1)\ P(A_2)$। অতএব, $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1)\ P(A_2) = 1/6 + 1/6 - 1/6.1/6 = 11/36$॥

৪০.১০ গ্রন্থপঞ্জী

- দাশগুপ্ত, ভাগবত ; চৌধুরী অরিজিৎ ; দাস, বিশ্বনাথ : *রাশিবিজ্ঞানের পরিভাষা*, বিশ্বভারতী, (১৯৭২)।
- সেন, রাজকুমার : *সংখ্যাতত্ত্ব*, পশ্চিমবঙ্গ রাজ্য পুস্তক পর্যদ, (১৯৮৬)।
- Das, N. G. : *Statistical Methods in Commerce, Accountancy and Economics*, M. Das, (1985).
- Goon, A. M., Gupta, M. K. & Dasgupta, B. : *Fundamentals of Statistics*, Vols. I, II, World Press, (1986).
- Mathai, A. M. & Rathi, P. N. : *Probability and Statistics*, Macmillan, Madras, (1977).
- Nagar, A. L. & Das, R. K. : *Basic Statistics*, Oxford University Press; Delhi, (1983).
- Spiegel, M. R. : *Theory and Problems of Statistics*, Schaum Publishing Co, (1982).

একক ৪১ ❑ সম্ভাবনাতত্ত্বের উপপাদ্যসমূহ (Theory of Probability)

গঠন

৪১.০ উদ্দেশ্য

এই এককটি থেকে জানা যাবে—

- নিঃশর্ত সম্ভাবনা ও শর্তাধীন সম্ভাবনা কাকে বলে
- অনধীন (বা স্বনির্ভর বা স্বতন্ত্র) ঘটনা এবং অধীন (বা পরনির্ভর) ঘটনা কাকে বলে
- যৌগিক বা মিশ্র বা যৌথ ঘটনা কাকে বলে
- Boole-এর অসমতা কী
- শর্তাধীন সম্ভাবনাসংক্রান্ত উপপাদ্য কোনগুলি
- যৌগিক সম্ভাবনার উপপাদ্য কী
- প্রাথমিক সম্ভাবনা ও উত্তর সম্ভাবনা কাকে বলে
- Bayes-এর উপপাদ্য বলতে কী বোঝায়
- সমবায় নিয়ম ও পৌনঃপুনিক প্রয়াসের ঘটনা কী।

৪১.১ প্রস্তাবনা

একক ১-এ আমরা সম্ভাবনাতত্ত্বে গুচ্ছতত্ত্বের প্রয়োগ সম্পর্কে অবহিত হয়েছি। এ ছাড়া ঘটনাকেন্দ্রিক প্রক্রিয়াসমূহ এবং ঘটনাসংক্রান্ত কয়েকটি বৈশিষ্ট্য বিষয়েও জেনেছি। সেই জ্ঞানের ভিত্তিতে এই এককে আমরা সম্ভাবনাতত্ত্বের মূল কয়েকটি উপপাদ্যের বর্ণনা দেব এবং সেগুলি প্রমাণ করব।

৪১.২ সম্ভাবনাতত্ত্বের মূল কয়েকটি উপপাদ্যের বর্ণনা ও প্রমাণ

সম্ভাবনাকে আমরা দুভাগে ভাগ করি। প্রথমত আমরা যে সম্ভাবনার কথা আলোচনা করব, তাকে আমরা নিঃশর্ত সম্ভাবনা (unconditional probability) বলব। এক্ষেত্রে, ঘটনাকে বলা হয় অনধীন (বা স্বনির্ভর বা স্বতন্ত্র) ঘটনা (independent events)। কতকগুলি ঘটনাকে অনধীন ঘটনা বলা হয় যখন তাদের মধ্যে কোন একটি ঘটনা ঘটবার সম্ভাবনা অন্য ঘটনাগুলি ঘটা বা না ঘটার দ্বারা প্রভাবিত হয় না।

উদাহরণ ১ :

দুটি মুদ্রা ক্ষেপণ করা হলে একটি মুদ্রার সোজা দিক বা উল্টো দিক পড়ার সম্ভাবনা অন্য মুদ্রাটির সোজা দিক বা উল্টো দিক পড়ার ঘটনাটির দ্বারা প্রভাবিত হয় না। সুতরাং, দুটি মুদ্রা ক্ষেপণের ক্ষেত্রে সোজা দিক বা উল্টো দিক পড়ার ঘটনা দুটি অনধীন ঘটনার উদাহরণ।

উদাহরণ ২ :

কোন একটি তাসের প্যাকেট থেকে একটি তাস টানা হল এবং রাজা ওঠার সম্ভাবনা নির্ণয় করা হল। এটি হল প্রথম ঘটনা। তারপর, এই তাসটি আবার প্যাকেটে ফিরিয়ে দেওয়া হল। আবার, আর একটি তাস টানা হল এবং রানী ওঠার সম্ভাবনা নির্ণয় করা হল। এটি হল দ্বিতীয় ঘটনা। প্রথম ঘটনাটি অবশ্যই অনধীন। দ্বিতীয় ঘটনাটিও অনধীন, যেহেতু প্রথমবার টানার পর তাসটি আবার প্যাকেটে ফিরিয়ে দেওয়া হয়েছিল।

নিম্নে আলোচিত প্রথম সাতটি উপপাদ্য নিঃশর্ত সম্ভাবনা সংক্রান্ত।

৪১.২.১ নিঃশর্ত সম্ভাবনা সংক্রান্ত উপপাদ্যসমূহ

উপপাদ্য ১ : যে কোন ঘটনা A_1-এর ক্ষেত্রে $0 \le P(A_1) \le 1$

প্রমাণ : কোন 'সমসম্ভব' পারিসংখ্যানিক পরীক্ষার ক্ষেত্রে যদি মোট সম্ভাব্য মৌলিক ঘটনার সংখ্যা হয় N ও তার ভেতর A_1 ঘটনা ঘটার পক্ষে অনুকূল মোট মৌলিক ঘটনার সংখ্যা যদি n হয়, তবে আমরা জানি যে, $0 \le n \le N$। সুতরাং, $0 \le n/N \le 1$। অথবা, $0 \le P(A_1) \le 1$ (কারণ $P(A_1) = n/N$)।

উপপাদ্য ২ : যদি কোন ঘটনা A_1 একটি অসম্ভব ঘটনা হয়, তবে $P(A_1) = 0$ এবং যদি একটি নিশ্চিত ঘটনা হয়, তবে $P(A_1) = 1$।

প্রমাণ : যদি A_1 একটি অসম্ভব ঘটনা হয়, তবে তা ঘটতে পারে না। সুতরাং, সেক্ষেত্রে $n = 0$ এবং $P(A_1) = 0/N = 0$। আবার, A_1 একটি নিশ্চিত ঘটনা হলে সর্বক্ষেত্রেই সেটি ঘটবে এবং $n = N$ হবে। সুতরাং, সেক্ষেত্রে $P(A_1) = N/N = 1$।

উপপাদ্য ৩ : যদি $A_1 \subset A_2$ হয়, অর্থাৎ A_1-এর সংঘটন A_2-এর সংঘটন সূচিত করে, তবে $P(A_1) \le P(A_2)$।

প্রমাণ : A_1 ও A_2 ঘটনা দুটি ঘটার পক্ষে অনুকূল মোট মৌলিক ঘটনার সংখ্যা যদি যথাক্রমে n এবং m হয় এবং ঘটনা দুটি ঘটার ক্ষেত্রে যদি মোট সম্ভাব্য মৌলিক ঘটনার সংখ্যা N হয়, তবে $n \le m$ হবে। কারণ, A_1 ঘটলে যেহেতু A_2 ঘটা আবশ্যিক হয়, সেজন্য A_1-এর পক্ষে অনুকূল প্রতিটি মৌলিক ঘটনাই A_2 ঘটার পক্ষেও অনুকূল হবে ; কিন্তু, A_2 ঘটার পক্ষে অনুকূল প্রতিটি মৌলিক ঘটনা A_1 ঘটার পক্ষে অনুকূল নাও হতে পারে। অতএব, $n/N \le m/N$, বা $P(A_1) \le P(A_2)$।

অনুসিদ্ধান্ত ৩.১ : যদি $A_1 = A_2$ হয়, তবে $P(A_1) = P(A_2)$।

প্রমাণ : এক্ষেত্রে $A_1 \subset A_2$ এবং সেইসঙ্গে $A_2 \subset A_1$। সুতরাং, উপপাদ্য ৩ অনুসারে আমরা পাই $P(A_1) \le P(A_2)$ এবং একইসঙ্গে $P(A_2) \le P(A_1)$। এই দুটি ফলকে একত্রিত করে পাওয়া যায় $P(A_1) = P(A_2)$।

উপপাদ্য ৪ : (যোগের উপপাদ্য বা মোট সম্ভাবনার উপপাদ্য) A_1 ও A_2 যদি পরস্পর বর্জনকারী ঘটনা হয়, তবে A_1 এবং A_2-এর মধ্যে যে কোন একটি ঘটনা ঘটার সম্ভাবনা $P(A_1 \cup A_2)$ হল $P(A_1) + P(A_2)$। অর্থাৎ, $P(A_1 \cup A_2) = P(A_1) + P(A_2)$।

প্রমাণ : ধরা যাক, A_1 ও A_2 ঘটনা দুটি ঘটার পক্ষে সম্ভাব্য মোট N-সংখ্যক সম-আশংসিত, পরস্পর উপায়গুলির মধ্যে n-সংখ্যক উপায় A_1 ঘটার পক্ষে এবং m-সংখ্যক উপায় A_2 ঘটার পক্ষে অনুকূল। সুতরাং, A_1 ও A_2 এই দুটি ঘটনার মধ্যে যে কোন একটি ঘটনা ঘটার পক্ষে অনুকূল উপায়ের সংখ্যা হল $n + m$। অতএব, $P(A_1 \cup A_2)$ $(n + m)/N = n/N + m/N = P(A_1) + P(A_2)$।

এই নিয়মটিকে আরোহী অনুমান পদ্ধতির (method of induction) মাধ্যমে ইচ্ছামত প্রসারিত করা যায়। নিম্নের উপপাদ্যে তার প্রমাণ দেওয়া হল।

উপপাদ্য ৪.১ : $A_1, A_2, A_3, \ldots\ldots\ldots, A_k$, যদি k-টি পরস্পর বর্জনকারী, সম-আশংসিত ঘটনা হয়, তবে $A_1, A_2, A_3, \ldots\ldots\ldots, A_k$-এর মধ্যে যে কোন একটি ঘটনা ঘটার সম্ভাবনা $P(A_1 \cup A_2 \cup A_3 \cup \ldots\ldots\ldots \cup A_k)$ হল $P(A_1), P(A_2), P(A_3), \ldots\ldots\ldots P(A_k)$-এর যোগফল। অর্থাৎ, $P(A_1 \cup A_2 \cup A_3 \cup \ldots\ldots\ldots \cup A_k) = P(A_1) + P(A_2) + P(A_3) + \ldots\ldots\ldots + P(A_k)$।

প্রমাণ : উপপাদ্য ৪-এর ফলকে বারবার ব্যবহার করে যে কোন k-সংখ্যক পরস্পর বর্জনকারী ঘটনার জন্য আমরা বলতে পারি

$$P(A_1 \cup A_2 \cup A_3 \cup \ldots\ldots\ldots\ldots\ldots\ldots \cup A_k)$$
$$= P(\{A_1 \cup A_2 \cup A_3 \cup \ldots\ldots\ldots\ldots\ldots\ldots \cup A_{k-1}\} \cup A_k)$$
$$= P(A_1 \cup A_2 \cup A_3 \cup \ldots\ldots\ldots\ldots\ldots\ldots \cup A_{k-1}) + P(A_k)$$
$$= P(\{A_1 \cup A_2 \cup A_3 \cup \ldots\ldots\ldots\ldots\ldots\ldots \cup A_{k-2}\} \cup A_{k-1}) + P(A_k)$$
$$= P(\{A_1 \cup A_2 \cup A_3 \cup \ldots\ldots\ldots\ldots\ldots\ldots \cup A_{k-2}\} \cup (A_{k-1}) + P(A_k)$$
$$\ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots$$
$$= P(A_1) + P(A_2) + P(A_3) + \ldots\ldots\ldots\ldots\ldots\ldots P(A_k)$$

অনুসিদ্ধান্ত ৪.১ : $A_1, A_2, A_3, \ldots\ldots\ldots\ldots\ldots\ldots A_k$ যদি একই সাথে পরস্পর বর্জনকারী ও একত্রে একটি পরিপূর্ণ গুচ্ছ গঠনকারী ঘটনা হয় তবে,

$$P(A_1 \cup A_2 \cup A_3 \cup \ldots\ldots\ldots\ldots\ldots\ldots \cup A_k)$$
$$= P(A_1) + P(A_2) + P(A_3) + \ldots\ldots\ldots\ldots\ldots\ldots + P(A_k) = 1$$

প্রমাণ : যেহেতু $A_1, A_2, A_3, \ldots\ldots\ldots\ldots\ldots\ldots, A_k$ ঘটনাগুলি একত্রে একটি পরিপূর্ণ গুচ্ছ গঠন করে সেজন্য $A_1 \cup A_2 \cup A_3 \cup \ldots\ldots\ldots\ldots\ldots\ldots \cup A_k$-কে আমরা নিশ্চিত ঘটনা বলতে পারি। সুতরাং, উপপাদ্য ২ অনুসারে $P(A_1 \cup A_2 \cup A_3 \cup \ldots\ldots\ldots\ldots\ldots\ldots \cup A_k) = 1$। আবার যেহেতু $A_1, A_2, A_3, \ldots\ldots\ldots\ldots\ldots\ldots, A_k$ পরস্পর বর্জনকারী, সেহেতু উপপাদ্য ৪.১ অনুসারে $P(A_1 \cup A_2 \cup A_3 \cup \ldots\ldots\ldots\ldots\ldots\ldots \cup A_k) = P(A_1) + P(A_2) + P(A_3) + \ldots\ldots\ldots\ldots\ldots\ldots + P(A_k)$। এই দুই ফলকে একত্রিত করলে আমরা পাই, $P(A_1) + P(A_2) + P(A_3) + \ldots\ldots\ldots\ldots\ldots\ldots + P(A_k) = 1$

অনুসিদ্ধান্ত ৪.২ : $P(A_1') = 1 - P(A_1)$

প্রমাণ : এখানে A_1 এবং A_1' পরস্পর বর্জনকারী ঘটনা। অতএব, $P(A_1 \cup A_1') = P(A_1) + P(A_1')$। আবার A_1 এবং A_1' ঘটনা দুটি পরস্পর সম্পূরক ঘটনা বলে তারা একটি পরিপূর্ণ গুচ্ছ গঠন করে। সেজন্য $P(A_1 \cup A_1') = 1$। সুতরাং, অনুসিদ্ধান্ত ৪.১ অনুসারে $P(A_1) + P(A_1') = 1$। অথবা, $P(A_1') = 1 - P(A_1)$।

অনুসিদ্ধান্ত ৪.৩ : $A = A_1 \cup A_2 \cup A_3 \cup \ldots\ldots\ldots\ldots\ldots\ldots \cup A_k$ হয়, অর্থাৎ, $A_1, A_2, A_3, \ldots\ldots\ldots\ldots\ldots\ldots, A_k$ পরস্পর বর্জনকারী ঘটনাগুলি A ঘটনার বিভিন্ন প্রকারের একত্রে একটি পরিপূর্ণ গুচ্ছ গঠন করে, তবে $P(A) = P(A_1) + P(A_2) + P(A_3) + \ldots\ldots\ldots\ldots\ldots\ldots + P(A_k)$।

প্রমাণ : প্রদত্ত তথ্য অনুসারে $A = A_1 + A_2 + A_3 + \ldots\ldots\ldots\ldots\ldots + A_k$। সুতরাং, $P(A) = P(A_1 + A_2 + A_3 + \ldots\ldots\ldots\ldots\ldots\ldots + A_k)$। এখন উপপাদ্য ৪.১ অনুযায়ী $P(A_1 + A_2 + A_3$

$+ \ldots\ldots\ldots\ldots + A_k) = P(A_1) + P(A_2) + P(A_3) + \ldots\ldots\ldots\ldots + P(A_k)$।

সুতরাং, $P(A) = P(A_1) + P(A_2) + P(A_3) + \ldots\ldots\ldots\ldots + P(A_k)$।

উপপাদ্য ৫ : যদি A_1-এর সংঘটন সূচিত করে, অর্থাৎ $A_1 \subset A_2$ হয়, তবে $P(A_2 - A_1) = P(A_2) - P(A_1)$।

প্রমাণ : প্রদত্ত তথ্য অনুসারে, A_2 ঘটতে পারে যদি এবং কেবলমাত্র যদি A_1 ও $(A_2 - A_1)$ এই দুটি পরস্পর বর্জনকারী ঘটনার যে কোন একটি ঘটে। সুতরাং, অনুসিদ্ধান্ত ৪.৩ অনুযায়ী আমরা বলতে পারি $P(A_2) = P(A_1) + P(A_2 - A_1)$। বা, $P(A_2 - A_1) = P(A_2) - P(A_1)$।

উপপাদ্য ৬ : A_1 ও A_2 যদি দুটি সাধারণ ঘটনা হয় অর্থাৎ, পরস্পর বর্জনকারী না হয়, তবে $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$।

এই নিয়মটিকে আরোহী অনুমান পদ্ধতির মাধ্যমে ইচ্ছামত প্রসারিত করা যায়। নিম্নের উপপাদ্যে তার প্রমাণ দেওয়া হল।

উপপাদ্য ৬.১ : $A_1, A_2, A_3 + \ldots\ldots\ldots + A_k$ যদি k-টি সাধারণ ঘটনা হয়, অর্থাৎ পরস্পর বর্জনকারী না হয়, তবে

$P(A_1 \cup A_2 \cup A_3 \cup \ldots\ldots\ldots\ldots \cup A_k)$

$= \{P(A_1) + P(A_2) + P(A_3) + \ldots\ldots\ldots + P(A_k)\} - \{P(A_1 \cap A_2) + P(A_1 \cap A_3) + \ldots\ldots + P(A_{k-1} \cap A_k)\} + \{P(A_1 \cap A_2 \cap A_3) + \ldots\ldots + P(A_{k-2} \cap A_{k-1} \cap A_k)\} + \ldots\ldots + (-1)^{k-1} P(A_1 \cap A_2 \cap \ldots\ldots \cap A_k)$।

$= \sum_{i=1}^{k} P(A_i) - \sum_{i=1}^{k}\sum_{j=1}^{k} P(A_i \cap A_j) + \sum_{\substack{i=1 \\ i<j<l}}^{k}\sum_{j=1}^{k}\sum_{l=1}^{k} P(A_i \cap A_j \cap A_l) - \ldots + (-1)^{k-1}$

$P(A_1 \cap A_2 \cap \ldots\ldots \cap A_k)$।

[এই সূত্রে মোট পদের সংখ্যা হল $2^k - 1 = k_{c1} + k_{c2} + \ldots\ldots\ldots + k_{ck}$

আমরা জানি, $(1 + a)^k = 1 + k_{c1}\, a + k_{c2}\, a^2 + \ldots\ldots\ldots + k_{ck}\, a^k$

অতএব, $(1 + 1)^k = 2^k = 1 + k_{c1} + k_{c2} + \ldots\ldots\ldots + k_{ck}$

যেখান থেকে পাওয়া যায়, $2^k - 1 = k_{c1} + k_{c2} + \ldots\ldots\ldots + k_{ck}$

উদাহরণ :

$k=2$ হলে মোট পদের সংখ্যা হল $2^2 - 1 = 3$

অতএব, $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$।

$k=3$ হলে মোট পদের সংখ্যা হল $2^3 - 1 = 7$

অতএব, $P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3) - P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_2 \cap A_3) + P(A_1 \cap A_2 \cap A_3)$।

সেইরকম, k=4 হলে পদের সংখ্যা হল $2^4 - 1 = 15$, k=15 হলে পদের সংখ্যা হল $2^5 - 1 = 31$, k=6 হলে পদের সংখ্যা হল $2^6 - 1 = 63$, ইত্যাদি]

প্রমাণ : ধরা যাক, k=m (≥2) ঘটনার জন্য এই উপপাদ্য সত্য। অর্থাৎ,

$$P(A_1 \cup A_2 \cup A_3 \cup \ldots\ldots\ldots\ldots \cup A_m)$$
$$= \{P(A_1) + P(A_2) + P(A_3) + \ldots\ldots\ldots\ldots + P(A_m)\} - \{P(A_1 \cap A_2) + P(A_1 \cap A_3) + \ldots\ldots + P(A_{m-1} \cap A_m)\} + \{P(A_1 \cap A_2 \cap A_3) + \ldots\ldots\ldots + P(A_{m-2} \cap A_{m-1} \cap A_m)\} + \ldots\ldots\ldots + (-1)^{m-1} P(A_1 \cap A_2 \cap \ldots\ldots\ldots \cap A_m)।$$

$$= \sum_{i=1}^{m} P(A_i) - \sum_{\substack{i=1\, j=1\\ i<j}}^{m\ m} P(A_i \cap A_j) + \sum_{\substack{i=1\, j=1\, l=1\\ i<j<l}}^{m\ m\ m} P(A_i \cap A_j \cap A_l) - \ldots. + (-1)^{m-1} P(A_1 \cap A_2 \cap \ldots\ldots\ldots\ldots \cap A_m)।$$

এবার, m+1 ঘটনার জন্যও যে এই উপপাদ্য সত্য, তা প্রমাণ করা যায়।

$$P(A_1 \cup A_2 \cup A_3 \cup \ldots\ldots\ldots\ldots \cup A_{m+1}) = P(A \cup A_{m+1})$$

যেখানে $A = A_1 \cup A_2 \cup A_3 \cup \ldots\ldots\ldots\ldots \cup A_m$

যেহেতু উপপাদ্যটি দুটি ঘটনার জন্য সত্য, সেহেতু

$$P(A \cup A_{m+1}) = P(A) + P(A_{m+1}) - P(A \cap A_{m+1})।$$
$$= P(\{A_1 \cup A_2 \cup A_3 \cup \ldots\ldots\ldots\ldots \cup A_m\}) + P(A_{m+1}) - P(\{A_1 \cup A_2 \cup A_3 \cup \ldots\ldots\ldots\ldots \cup A_m\} \cap A_{m+1})$$
$$= P(\{A_1 \cup A_2 \cup A_3 \cup \ldots\ldots\ldots\ldots \cup A_m\}) + P(A_{m+1}) - P\{(A_1 \cap A_{m+1}) \cup (A_2 \cap A_{m+1}) \cup (A_3 \cap A_{m+1}) \cup \ldots\ldots\ldots\ldots \cup (A_m \cap A_{m+1})\}।$$
$$= P(\{A_1 \cup A_2 \cup A_3 \cup \ldots\ldots\ldots\ldots \cup A_m\}) + P(A_{m+1}) - P(G_1 \cup G_2 \cup G_3 \cup \ldots\ldots\ldots\ldots \cup G_m)।$$

(যেখানে, $G_i = A_i \cap A_{m+1}$, $i = 1, 2, \ldots\ldots, m$)

$$= \sum_{i=1}^{m} P(A_i) - \sum_{\substack{i=1\, j=1\\ i<j}}^{m\ m} P(A_i \cap A_j) + \sum_{i=1\, j=1\, l=1}^{m\ m\ m} P(A_i \cap A_j \cap A_l) - \ldots. + (-1)^{m-1} P(A_1 \cap A_2 \cap \ldots\ldots\ldots\ldots \cap A_m) + P(A_{m+1}) - \sum_{i=1}^{m} P(G_i) + \sum_{i=1\, j=1}^{m\ m} P(G_i \cap G_j) - \sum_{i=1\, j=1\, l=1}^{m\ m\ m} P(G_i \cap G_j \cap G_l) + \ldots. + (-1)^m P(G_1 \cap G_2 \cap \ldots\ldots\ldots\ldots \cap G_m)।$$

$$= \sum_{i=1}^{m} P(A_i) + P(A_{m+1}) - \sum_{i=1\, j=1}^{m\ m} P(A_i \cap A_j) - \sum_{i=1}^{m} P(A_i \cap A_{m+1}) + \sum_{i=1\, j=1\, l=1}^{m\ m\ m} P(A_i \cap A_j \cap A_l) + \sum_{i=1\, j=1}^{m\ m} P(A_i \cap A_{m+1} \cap A_j \cap A_{m+1}) - \ldots. + (-1)^m P(A_1 \cap A_{m+1} \cap A_2 \cap A_{m+1} \cap \ldots\ldots \cap A_m \cap A_{m+1})$$

$$= \sum_{i=1}^{m} P(A_i) - \sum_{i=1\, j=1}^{m\ m} P(A_i \cap A_j) + \sum_{i=1\, j=1\, l=1}^{m\ m\ m} P(A_i \cap A_j \cap A_l) - \ldots. + (-1)^m P(A_1 \cap A_2 \cap \ldots\ldots\ldots\ldots \cap A_m \cap A_{m+1})$$

$[A_1 \cap A_{m+1} \cap A_2 \cap A_{m+1} = A_1 \cap A_2 \cap A_{m+1}]$

যেহেতু উপপাদ্যটি দুটি ঘটনার জন্য সত্য, সেহেতু এটি তিনটি ঘটনার জন্যও সত্য। আবার, যেহেতু উপপাদ্যটি তিনটি ঘটনার জন্য সত্য, সেহেতু এটি চারটি ঘটনার জন্যও সত্য। এভাবে উপপাদ্যটি m-সংখ্যক ধনাত্মক ঘটনার জন্য সত্য।

উপপাদ্য ৭ : A_1 ও A_2 যদি দুটি সাধারণ ঘটনা হয় অর্থাৎ পরস্পর বর্জনকারী না হয়, তবে $P(\cup A_i) \le \sum P(A_i)$

প্রমাণ : যে কোন দুটি ঘটনার ক্ষেত্রে $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$।

$$\le P(A_1) + P(A_2)$$

কারণ, $P(A_1 \cap A_2) \ge 0$ ॥

এই ফলটিকে যেমন ইচ্ছা প্রসারিত করলে দেখা যায়,

$$\begin{aligned} P(\cup A_i) &= P(\cup A_i) + P(A_k) - P[(\cup A_i) \cap A_k] \\ &\le P(\cup A_i) + P(A_k) \\ &\le P(\cup A_i) + P(A_{k-1}) - P[(\cup A_i) \cap A_{k-1}] + P(A_k) \\ &\le P(\cup A_i) + P(A_{k-1}) + P(A_k) \\ &\le \sum_{i=1}^{k} P(A_i) \end{aligned}$$

এই অসমতাটিকে Boole-এর অসমতা (Boole's inequality) বলা হয়। ঘটনাগুলি পরস্পর বর্জনকারী হলে তবে এই অসমতাটি সমতায় পরিণত হবে এবং উপপাদ্য ৪.১ রূপে প্রকাশিত হবে।

৪১.২.২ শর্তাধীন সম্ভাবনা সংক্রান্ত উপপাদ্যসমূহ

যদি কোন ঘটনা ঘটার সম্ভাবনা অন্য কোন ঘটনা ঘটা বা না ঘটার ওপর নির্ভর করে, তবে সেই ঘটনাকে অধীন (বা পরনির্ভর) (dependent) ঘটনা বলা হয়।

উদাহরণ :

কোন একটি তাসের প্যাকেট থেকে একটি তাস টানা হল এবং রাজা ওঠার সম্ভাবনা নির্ণয় করা হল। এটি হল প্রথম ঘটনা। তারপর, ধরা যাক, এই তাসটি আর প্যাকেটে ফিরিয়ে দেওয়া হল না। আবার, আর একটি তাস টানা হল এবং রানী ওঠার সম্ভাবনা নির্ণয় করা হল। এটি হল দ্বিতীয় ঘটনা। প্রথম ঘটনাটি অবশ্যই অনধীন। দ্বিতীয় ঘটনাটি হল অধীন ঘটনা, যেহেতু প্রথমবার টানার পর তাসটি আবার প্যাকেটে ফিরিয়ে না দেওয়ায় দ্বিতীয় তাসটি ওঠার সম্ভাবনা এর দ্বারা প্রভাবিত হবে।

কোন অধীন ঘটনা ঘটার সম্ভাবনাকে আমরা শর্তাধীন (conditional) সম্ভাবনা বলে থাকি। যদি জানা থাকে যে, কোন একটি ঘটনা ইতিমধ্যে ঘটে গিয়েছে এবং সেই ঘটনা ঘটার সম্ভাবনার মান ধনাত্মক, তবে এই দুটি শর্তসাপেক্ষে দ্বিতীয় একটি ঘটনা ঘটার সম্ভাবনাকেই আমরা শর্তাধীন সম্ভাবনা বলব। দুই বা ততোধিক ঘটনার একত্রিতভাবে সম্ভাবনা নির্ণয়ের ক্ষেত্রে আমরা সম্ভাবনার গুণের উপপাদ্য ব্যবহার করব। অধীন ও অনধীন, এই দুটি ঘটনার জন্য আমরা সম্ভাবনার গুণের দুটি উপপাদ্য পাব। একাধিক ঘটনার একত্রিতভাবে ঘটার ঘটনাকে যৌগিক বা মিশ্র (compound) বা যৌথ (joint) ঘটনা বলা হয়। সেইজন্য যৌগিক ঘটনাসংক্রান্ত উপপাদ্যকে যৌগিক সম্ভাবনার উপপাদ্য (theorem of compound probability)-ও বলা হয়ে থাকে।

ধরে নিই, কোন 'সমসম্ভব' পারিসংখ্যানিক পরীক্ষার ক্ষেত্রে মোট সম্ভাব্য মৌলিক ঘটনার সংখ্যা হল N। তার ভেতর A_1 ঘটনা ঘটার পক্ষে অনুকূল মোট মৌলিক ঘটনার সংখ্যা n, A_2 ঘটনা ঘটার পক্ষে অনুকূল মোট মৌলিক ঘটনার সংখ্যা m এবং A_1 ও A_2 যৌথভাবে ঘটার পক্ষে অনুকূল মোট মৌলিক ঘটনার সংখ্যা হল r।

সারণি ৪১.১

	A_2 ঘটে	A_2 ঘটে না	মোট
A_1 ঘটে	r	n–r	n
A_1 ঘটে না	m–r	N–m–n+r	N–n
মোট	m	N–m	N

$$P(A_1) = n/N$$
$$P(A_1') = (N-n)/N$$
$$P(A_2) = m/N$$
$$P(A_2') = (N-m)/N$$
$$P(A_2|A_1) = r/n$$
$$P(A_1|A_2) = r/m$$

$P(A_2|A_1)$ এবং $P(A_1|A_2)$-কে বলা হয় A_2 (A_1) ঘটনাটির শর্তাধীন সম্ভাবনা, যেখানে A_1 (A_2) ঘটনাটি ইতিমধ্যেই ঘটেছে।

এখন, $P(A_2|A_1) = r/n = (r/N)(n/N) = P(A_1 \cap A_2)/P(A_1)$ যদি $P(A_1) > 0$।

সেইরকম, $P(A_1|A_2) = r/m = (r/N)(m/N) = P(A_1 \cap A_2)/P(A_2)$ যদি $P(A_2) > 0$।

$P(A_1)$ এবং $P(A_2)$ হল নিঃশর্ত সম্ভাবনা।

$$\begin{aligned} P(A_2|A_1') = (m-r)/(N-n) &= (m-r')/N/(N-n)/N = P(A_1' \cap A_2)/P(A_1') \\ &= [P(A_2) - P(A_1 \cap A_2)]/1 - P(A_1) \\ &= [P(A_2) - P(A_2)\,P(A_1|A_2)/1 - P(A_1) \\ &= P(A_2)[1 - P(A_1|A_2)]/1 - P(A_1) \end{aligned}$$

$$P(A_2'|A_1) = (n-r)/n = (n-r)/N/n/N = P(A_1 \cap A_2')/P(A_1)$$
$$= [P(A_1) - P(A_1 \cap A_2)]/P(A_1) = 1 - [P(A_1 \cap A_2)]/P(A_1)$$

$$\begin{aligned} P(A_1'|A_2') &= (N-m-n-r)/N-n = [(N-m-n-r)/N]/[(N-n)/N] = P(A_1' \cap A_2')/P(A_1') \\ &= [1-P(A_1) - P(A_2) + P(A_1 \cap A_2)]/[1-P(A_1)] \\ &= 1-[P(A_2) - P(A_1 \cap A_2)]/[1-P(A_1)] \\ &= 1-[P(A_2) - P(A_2)\,P(A_1|A_2)]/[1-P(A_1)] \\ &= 1-P(A_2)[1-P(A_1|A_2)]/[1-P(A_1)] \end{aligned}$$

দুটি ঘটনাকে অনধীন ঘটনা বলা হয় যখন $P(A_1|A_2) = P(A_1)$, অর্থাৎ A_1 ঘটনাটির ঘটা A_2 ঘটনাটির ঘটার ওপর নির্ভর করে না। সেইরকম, $P(A_2|A_1) = P(A_2)$, অর্থাৎ A_2 ঘটনাটির ঘটা A_1 ঘটনাটির ঘটার ওপর নির্ভর করে না। যদি একই সাথে এই দুটি শর্ত পূর্ণ হয়, তবে A_1 ও A_2-কে বলা হয় পারস্পরিকভাবে স্বতন্ত্র (mutually independent) ঘটনা।

$$[P(A_1 \cap A_2)] / P(A_2) = P(A_1|A_2) = P(A_1)$$

অথবা, $P(A_1 \cap A_2) = P(A_1)\, P(A_2)$

সেইরকমই, $[P(A_1 \cap A_2)] / P(A_1) = P(A_2|A_1) = P(A_2)$

অথবা, $P(A_1 \cap A_2) = P(A_1)\, P(A_2)$

তাহলে, A_1 ও A_2 হল পারস্পরিকভাবে স্বতন্ত্র ঘটনা যদি A_1 ও A_2-এর যৌথভাবে ঘটার সম্ভাবনা হয় এই দুটি ঘটনা পৃথকভাবে ঘটার সম্ভাবনার গুণফলের সমান। দুটি ঘটনার ক্ষেত্রে এটি হল পারস্পরিক স্বতন্ত্রতার প্রয়োজনীয়তার (necessary) এবং পর্যাপ্ততার (sufficient) শর্ত।

যখন $n > 2$, তখন n-সংখ্যক ঘটনার পারস্পরিক স্বতন্ত্রতার প্রয়োজনীয়তার শর্ত হল,

$$P(A_1 \cap A_2 \cap A_3 \cap \ldots\ldots\ldots\ldots \cap A_k) = P(A_1)\, P(A_2)\, P(A_3) \ldots\ldots\ldots\ldots P(A_k)$$

কিন্তু, $n > 2$ হলে এটি কিন্তু পারস্পরিক স্বতন্ত্রতার পর্যাপ্ততার শর্ত হয় না। অর্থাৎ, যদি ঘটনাগুলিকে পারস্পরিকভাবে স্বতন্ত্র হতে হয়, তাহলে এই শর্তটি পূর্ণ হতে হবে, কিন্তু এই শর্তটি পূর্ণ হলেই যে ঘটনাগুলি পারস্পরিকভাবে স্বতন্ত্র হবে তার কোন আবশ্যিকতা নেই।

উদাহরণ :

ধরে নিই যে তিনটি ঘটনা আছে, A_1, A_2 ও A_3। $P(A_1) = 0.8$, $P(A_2) = 0.6$ ও $P(A_3) = 0.5$।

$P(A_1 \cap A_2) = 0.45$, $P(A_1 \cap A_3) = 0.35$, $P(A_2 \cap A_3) = 0.25$ এবং $P(A_1 \cap A_2 \cap A_3) = 0.24$।

এক্ষেত্রে, $P(A_1)\, P(A_2)\, P(A_3) = 0.24 = P(A_1 \cap A_2 \cap A_3)$।

কিন্তু, তিনটি ঘটনার ক্ষেত্রে পারস্পরিকভাবে স্বতন্ত্র হলেও, ঘটনাগুলিকে জুটি হিসাবে (pairwise) দেখলে দেখা যাবে যে, ওই অবস্থায় তারা পারস্পরিকভাবে স্বতন্ত্র নয়। যেমন, দুটি ঘটনা A_1 ও A_2-এর ক্ষেত্রে, $P(A_1)\, P(A_2) = 0.48 > P(A_1 \cap A_2)$। সেইরকম, A_1 ও A_3-এর ক্ষেত্রে, $P(A_1)\, P(A_3) = 0.4 > P(A_1 \cap A_3)$ এবং A_2 ও A_3-এর ক্ষেত্রে, $P(A_2)\, P(A_3) = 0.3 > P(A_2 \cap A_3)$।

এখান থেকেই প্রমাণিত হয় যে, দুইয়ের অধিক ঘটনার ক্ষেত্রে $P(A_1 \cap A_2 \cap A_3 \cap \ldots\ldots\ldots \cap A_k) = P(A_1)\, P(A_2)\, P(A_3) \ldots\ldots\ldots\ldots\ldots P(A_k)$ শর্তটি পূর্ণ হলেই যে ঘটনাগুলি পারস্পরিকভাবে স্বতন্ত্র হবে তার কোন আবশ্যিকতা নেই।

সাধারণত n ঘটনা $A_1, A_2, \ldots\ldots\ldots\ldots, A_n$-এর ক্ষেত্রে বলা যেতে পারে যে, এই ঘটনাগুলি তখনই পারস্পরিক স্বতন্ত্র হবে যখন প্রত্যেকটি ঘটনা অন্য যে কোন ঘটনা বা অন্য সব ঘটনার অনধীন হবে। এই ঘটনাগুলির পারস্পরিক স্বতন্ত্র হওয়ার প্রয়োজনীয়তার (necessary) এবং পর্যাপ্ততার (sufficient) শর্ত হল

$P(A_i \cap A_j) = P(A_i).\ P(A_j),\ 1 \le i \le j \le n$-এর ক্ষেত্রে

$P(A_i \cap A_j \cap A_k) = P(A_i).\ P(A_j).\ P(A_k),\ 1 \le i < j < k \le n$-এর ক্ষেত্রে

$P(A_1 \cap A_2 \cap \ldots\ldots\ldots \cap A_n) = P(A_1).\ P(A_2)\ \ldots\ldots\ldots\ P(A_n)$

উপরোক্ত ক্ষেত্রে মোট শর্তসংখ্যা হল ${}^nC_2 + {}^nC_3 + \ldots + {}^nC_n = 2^n - n - 1$

এখন, $n(>2)$-সংখ্যক ঘটনার ক্ষেত্রে কেবলমাত্র $P(A_i \cap A_j) = P(A_i).\ P(A_j),\ 1 \le i \le j \le n$-এর ক্ষেত্রে, শর্তগুলি পূরণ হলে ঘটনাগুলিকে জোড়ায়-জোড়ায় স্বতন্ত্র (pair-wise independent) বলা হয়ে থাকে। মনে রাখতে হবে যে, $n(>2)$-সংখ্যক ঘটনার ক্ষেত্রে, ঘটনাগুলি জোড়ায়-জোড়ায় স্বতন্ত্র হলেই যে তারা পারস্পরিক স্বতন্ত্র হবে এমন কোন মানে নেই। উদাহরণস্বরূপ ধরা যাক মাত্র তিনটি ঘটনা A_1, A_2 এবং A_3 আছে। কোন এক সমসম্ভব পরীক্ষা (random experiment)-এর ক্ষেত্রে মোট নমুনা দেশ (Total sample space)-এ মাত্র চারটি বিন্দু (point) আছে। এই চারটি বিন্দু (a, b, c এবং d) হল সম-আশংসিত (equally likely) অর্থাৎ $P\{a\} = P\{b\} = P\{c\} = P\{d\} = {}^1/_4$। ধরা যাক, তিনটি ঘটনা A_1, A_2 এবং A_3 হল $A_1 = \{a, b\}$, $A_2 = \{a, c\}$ এবং $A_3 = \{a, d\}$। তাহলে $P(A_1) = P(A_2) = P(A_3) = {}^1/_2$; $P(A_1 \cap A_2) = P(A_1 \cap A_3) = P(A_2 \cap A_3) = {}^1/_4$ এবং $P(A_1 \cap A_2 \cap A_3) = {}^1/_4$ । এক্ষেত্রে $P(A_i \cap A_j) = P(A_i).\ P(A_j),\ 1 \le i \le j \le 3$ । সুতরাং, A_1, A_2 এবং A_3 জোড়ায়-জোড়ায় স্বতন্ত্র। কিন্তু এই তিনটি ঘটনা পারস্পরিক স্বতন্ত্র নয়, কারণ পারস্পরিক স্বতন্ত্র হবার জন্য অন্য যে শর্তটি আছে যথা $P(A_1 \cap A_2 \cap A_3) = P(A_1).\ P(A_2).\ P(A_3)$—তা এখানে পূরণ হচ্ছে না। এখানে $P(A_1 \cap A_2 \cap A_3) = {}^1/_4$, কিন্তু $P(A_1).\ P(A_2).\ P(A_3) = {}^1/_8$ । সুতরাং, $P(A_1 \cap A_2 \cap A_3) \ne P(A_1).\ P(A_2).\ P(A_3)$ ।

ধরে নেওয়া যাক যে, যে দুটি ঘটনা আছে, A_1 ও A_2 এবং তারা সম্ভাব্য ঘটনা, অর্থাৎ $P(A_1) > 0$, $P(A_2) > 0$। এখন প্রশ্ন হল,

(১) যদি A_1 ও A_2 পরস্পর বর্জনকারী হয়, তবে তারা কি অনধীন?

(২) যদি A_1 ও A_2 অনধীন হয়, তবে তারা কি পরস্পর বর্জনকারী?

(১) A_1 ও A_2 পরস্পর বর্জনকারী হওয়ার অর্থ হল $P(A_1 \cap A_2) = 0$। তাহলে, $P(A_1 | A_2) = P(A_1 \cap A_2)/P(A_2) = 0/P(A_2) = 0$। তার অর্থ হল, $P(A_1 | A_2) \ne P(A_1)$ এবং $P(A_2 | A_1) \ne P(A_2)$ এবং $P(A_1 \cap A_2) \ne P(A_1)\ P(A_2)$। অর্থাৎ, A_1 ও A_2 পরস্পর বর্জনকারী হলে তারা কখনো পরস্পর অনধীন হয় না।

(২) A_1 ও A_2 পরস্পর অনধীন হওয়ার অর্থ হল $P(A_1 \cap A_2) = P(A_1) P(A_2)$। যেহেতু, $P(A_1) > 0$, $P(A_2) > 0$, সেই হেতু, $P(A_1 \cap A_2) > 0$। কিন্তু, পরস্পর বর্জনকারী ঘটনার ক্ষেত্রে $P(A_1 \cap A_2) = 0$। অতএব, দুটি ঘটনা পরস্পর অনধীন হলে তারা কখনো পরস্পর বর্জনকারী হয় না।

৪১.২.৩ যৌগিক সম্ভাবনার উপপাদ্য

উপপাদ্য ৮ : যদি A_1 ও A_2 ঘটনার ক্ষেত্রে $P(A_1) > 0$ হয়, তবে $P(A_1 \cap A_2) = P(A_1).P(A_2 | A_1)$। এই উপপাদ্যটিকে প্রসারিত করা যায়। ধরে নেওয়া যাক যে, kটি ঘটনা আছে $A_1, A_2, A_3, \ldots\ldots, A_k$। যদি $P(A_1) > 0$ হয়, $P(A_1 \cap A_2) > 0$ হয়, ……, $P(A_1 \cap A_2 \cap A_3 \cap \ldots\ldots \cap A_{k-1}) > 0$ হয়, তবে $P(A_1 \cap A_2 \cap A_3 \cap \ldots\ldots \cap A_{k-1} \cap A_k) = P(\cap A_i) = P(A_1).P(A_2 | A_1).P(A_3 | A_1 \cap A_2) \ldots\ldots\ldots\ldots P(A_k | A_1 \cap A_2 \cap A_3 \cap \ldots\ldots\ldots\ldots\ldots \cap A_{k-1})$

প্রমাণ : প্রথমে ধরে নিই যে, k=2। যদি মোট n সম-আশংসিত, পরস্পর বর্জনকারী ও একত্রে একটি পরিপূর্ণ গুচ্ছ গঠনকারী উপায়গুলির মধ্যে n_{12}-সংখ্যক উপায় A_1 ও A_2 একত্রে ঘটার পক্ষে এবং n_1-সংখ্যক উপায় A_1 ঘটার পক্ষে অনুকূল হয়, তবে $P(A_1 \cap A_2) = n_{12} / n$ ও $P(A_1) = n_1 / n$। এখানে $P(A_1) > 0$ হওয়ায় $n_1 > 0$। এখন, $n_{12} / n = (n_1 / n)(n_{12} / n_1)$। যেহেতু এক্ষেত্রে A_1 ও A_2 একত্রে ঘটার সম্ভাবনা নির্ণয় করতে হবে, সেজন্য আমরা দেখব যে, A_1 ঘটার পক্ষে অনুকূল উপায়গুলির (n_1) কতগুলি A_2 ঘটার পক্ষেও অনুকূল। যেহেতু n_{12}-সংখ্যক উপায় A_1 ও A_2 একত্রে ঘটার পক্ষে অনুকূল, সেহেতু এই উপায়গুলি A_2 ঘটার পক্ষেও অনুকূল হবে। সুতরাং, মোট n_1-সংখ্যক উপায়ের মধ্যে A_2 ঘটার পক্ষে অনুকূল উপায়ের সংখ্যা হল n_{12}। সুতরাং, $P(A_2 | A_1) = n_{12} / n_1$। অতএব, $P(A_1 \cap A_2) = P(A_1) P(A_2 | A_1)$।

এখন আমরা এই উপপাদ্যটি যে কোন সংখ্যক ঘটনার ক্ষেত্রে প্রমাণ করব। দুটি ঘটনার ক্ষেত্রে আমরা যে প্রমাণ পেয়েছি তার থেকে বলা যায় যে, $P(A_1 \cap A_2) > 0$, কারণ $P(A_1)$ ও $P(A_2 | A_1)$, উভয়েই > 0। সুতরাং, এইভাবে ঘটনার সংখ্যা তিন হলে আমরা পাব,

$P(A_1 \cap A_2 \cap A_3) = P(A_1 \cap A_2)P(A_3 | A_1 \cap A_2) = P(A_1)P(A_2 | A_1)P(A_3 | A_1 \cap A_2)$। আরোহী অনুমান পদ্ধতির মাধ্যমে ধরে নিই যে তত্ত্বটি m ঘটার জন্য সত্য। তাহলে,

$P(A_1 \cap A_2 \cap A_3 \cap \ldots\ldots\ldots \cap A_m)$
$= P(A_1).P(A_2/A_1).P(A_3/A_1 \cap A_2). \ldots\ldots\ldots\ldots P(A_m/A_1 \cap A_2 \cap A_3 \cap \ldots\ldots\ldots \cap A_{m-1})$।
m+1 ঘটনার জন্য, $P(A_1 \cap A_2 \cap A_3 \cap \ldots\ldots\ldots \cap A_{m+1}) = P(B \cap A_{m+1})$
যেখানে, $B = A_1 \cap A_2 \cap A_3 \cap \ldots\ldots\ldots \cap A_m$
অতএব, $P(A_1 \cap A_2 \cap A_3 \cap \ldots\ldots\ldots \cap A_{m+1}) = P(B \cap A_{m+1}) = P(B)P(A_{m+1} / B)$
$= P(A_1 \cap A_2 \cap A_3 \cap \ldots\ldots\ldots \cap A_m)$

$P(A_{m+1}/A_1 \cap A_2 \cap A_3 \cap \ldots\ldots\ldots \cap A_m)$
$=P(A_1).P(A_2/A_1).P(A_3/A_1 \cap A_2) \ldots\ldots\ldots\ldots P(A_m/A_1 \cap A_2 \cap A_3 \cap \ldots\ldots\ldots \cap A_{m-1})$
$(A_m/A_1 \cap A_2 \cap A_3 \cap \ldots\ldots\ldots\ldots\ldots \cap A_{m-1})$

সুতরাং, উপপাদ্যটি m+1 ঘটনার জন্য সত্য।

যেহেতু উপপাদ্যটি দুই, তিন, m+1টি ঘটনার জন্য সত্য, অতএব বলা যায় যে, উপপাদ্যটি যে কোন সংখ্যক ঘটনার জন্য সত্য। একইভাবে অগ্রসর হয়ে যে কোন সংখ্যক (k) ঘটনার ক্ষেত্রে আমরা পাব,

$$P(A_1 \cap A_2 \cap A_3 \cap \ldots\ldots\ldots \cap A_k)$$
$$= P(A_1).P(A_2 | A_1).P(A_3 | A_1 \cap A_2) \ldots\ldots\ldots\ldots P(A_k | A_1 \cap A_2 \cap A_3 \cap \ldots\ldots\ldots\ldots \cap A_{k-1})$$

অথবা, $P(\bigcap_{i=1}^{k} A_i) = P(A_1).P(A_2 | A_1).P(A_3 | A_1 \cap A_2) \ldots\ldots\ldots\ldots\ldots P(A_k | \cap A_i)$

৪১.২.৪ Bayes-এর উপপাদ্য

সাধারণভাবে আমরা যে সম্ভাবনার কথা আলোচনা করি তা হল প্রাথমিক (a-priori) সম্ভাবনা। উদাহরণ হিসাবে বলা যায়, কোন ঘটনা B কেবল তখনই ঘটতে পারে যদি পরস্পর বর্জনকারী ও একত্রে একটি পরিপূর্ণ গুচ্ছ গঠনকারী ঘটনাবলী A_i (i = 1,2, ……, k)-এর মধ্যে যে কোন একটি ঘটে। তাহলে $P(A_i)$-কে আমরা A_i-এর প্রাথমিক (a-priori) সম্ভাবনা বলতে পারি। এখন যদি B ঘটনা ঘটার ক্ষেত্রে A_i-গুলিকে আমরা বিবিধ কারণ হিসাবে মনে করি, তবে $P(A_i | B)$-কে আমরা উত্তর (posteriori) সম্ভাবনা বলব। $P'(A_i | B)$ দ্বারা আমরা A_i-এর শর্তাধীন সম্ভাবনা নির্দেশ করছি যখন জানা আছে যে, B ঘটনাটি (যেটি আবার কোন A_i ঘটনারই ফল) ইতিমধ্যেই ঘটেছে। এই সম্ভাবনাকে আমরা Bayes-এর উপপাদ্যের সাহায্যে নির্ণয় করতে পারি। এজন্য Bayes-এর উপপাদ্যকে অনেক সময় প্রকল্পের সম্ভাবনার (probabilities for hypotheses) উপপাদ্য বা বিপরীত সম্ভাবনার উপপাদ্যও বলা হয়। এটিকে আমরা এখন এইভাবে বলি :

মনে করি, কোন ঘটনা B কেবল তখনই ঘটে যখন পরস্পর বর্জনকারী ও একত্রে একটি পরিপূর্ণ গুচ্ছ গঠনকারী A_i (i = 1, 2, ……, k) ঘটনাবলীর মধ্যে যে কোনটি ঘটে। এখন, P(B), $P(A_i)$ ও $P(B | A_i)$ > 0 হলে $P(A_i | B) = [P(A_i)P(B | A_i) / \sum_{i=1}^{k}[P(A_i)P(B | A_i)]$

প্রমাণ : এখানে P(B), $P(A_i)$ ও $P(B | A_i)$ সম্ভাবনাগুলির মান আমরা জানি এবং এগুলির মান ধনাত্মক।

$P(B | A_i)$ শর্তাধীন সম্ভাবনাগুলিকে আমরা দুভাবে প্রকাশ করতে পারি।

$P(B | A_i) = P(A_i)\ P(B | A_i)$। আবার, $P(B | A_i) = P(B)\ P(A_i | B)$।

সুতরাং, $P(A_i)\ P(B | A_i) = P(B)\ P(A_i | B)$।

তাহলে, $P(A_i | B) = [P(A_i)\ P(B | A_i)] / P(B)$।

যেহেতু B ঘটনাটি $B \cap A_1$, $B \cap A_2$, ……, $B \cap A_k$ প্রভৃতি পরস্পর বর্জনকারী উপায়গুলির মধ্যে যে কোন একভাবে ঘটতে পারে সেজন্য উপপাদ্য ৪.১ অনুসারে

$$P(B) = \sum_{i=1}^{k} (B \cap A_i)$$
$$= \sum_{i=1}^{k}[P(A_i)P(B | A_i) \text{ (উপপাদ্য ৮ অনুসারে)}$$

অতএব, $P(A_i | B) = [P(A_i)P(B | A_i)] / \sum_{i=1}^{k}[P(A_i)P(B | A_i)]$।

৪১.৩ পৌনঃপুনিক প্রয়াস

যে সমস্ত ক্ষেত্রে কোন পরীক্ষা বারংবার করা হয়, বা কোন ঘটনা ঘটানোর জন্য বারংবার পৌনঃপুনিক প্রয়াস (repeated trials) করা হয়, সেখানে সম্ভাবনা নির্ণয়ের জন্য সমবায়ের নিয়মের বিশেষ প্রয়োজন হয়।

৪১.৩.১ সমবায় নিয়মের ব্যাখ্যা

n-সংখ্যক দ্রব্য থেকে r-সংখ্যক দ্রব্য যত উপায়ে গ্রহণ করা যায় (তাদের ক্রমের দিকে লক্ষ্য না রেখে) সেই সংখ্যাকেই সমবায়ের সংখ্যা বলা হয় ও এটিকে ^nC_r চিহ্নের দ্বারা সূচিত করা হয়। এক্ষেত্রে $^nC_r = n!/r!(n-r)!$ ।

সমবায়ের সাধারণ সূত্র থেকে সহজেই দেখা যায় যে $^nC_r = {^nC_{r-1}}$ এবং $^nC_r = 1$ (এখানে $0! = 1$ ধরা হয়)। এখানে একটি অন্তর্নিহিত স্বীকরণ আছে যে $n \geq r$, কারণ যদি $n < r$ হয়, তবে $^nC_r = 0$ হতে বাধ্য।

৪১.৩.২ পৌনঃপুনিক প্রয়াসের ঘটনা

যদি কোন ঘটনা ঘটার সম্ভাবনা হয় p, ও ফলে ঘটনা না ঘটার সম্ভাবনা হয় $1-p = q$, তবে ওই ঘটনা ঘটানোর জন্য n-সংখ্যক বার প্রয়াসের (যে প্রয়াসগুলি পরস্পর অনধীন) মধ্যে যদি r-সংখ্যক বার ঘটনাটি ঘটার সম্ভাবনা (P) হবে, $P = {^nC_r}\, p^r q^{n-r} = (n!)(p^r q^{n-r})/(r!)(n-r)!$। স্বাভাবিকভাবেই কোন ঘটনার ক্ষেত্রে $p + q = 1$।

৪১.৪ সারাংশ

সম্ভাবনাকে দুভাগে বিভক্ত করলে আমরা পাই নিঃশর্ত সম্ভাবনা ও শর্তাধীন সম্ভাবনা। নিঃশর্ত সম্ভাবনার ক্ষেত্রে, ঘটনাকে বলা হয় অনধীন (বা স্বনির্ভর বা স্বতন্ত্র) ঘটনা। অন্যদিকে শর্তাধীন সম্ভাবনার ক্ষেত্রে, ঘটনাকে বলা হয় অধীন ঘটনা। নিঃশর্ত সম্ভাবনা ও শর্তাধীন সম্ভাবনা প্রসঙ্গে আমরা কিছু উপপাদ্য উপস্থাপিত করতে পারি। এদের মধ্যে তিনটি গুরুত্বপূর্ণ উপপাদ্য হল যোগের উপপাদ্য, যৌগিক সম্ভাবনার উপপাদ্য এবং Bayes-এর উপপাদ্য (বা প্রকল্পের সম্ভাবনার উপপাদ্য বা বিপরীত সম্ভাবনার উপপাদ্য)। এই এককে আমরা এই তিনটি গুরুত্বপূর্ণ উপপাদ্য ছাড়াও আরও কয়েকটি উপপাদ্য প্রমাণ সহকারে উপস্থাপিত করেছি। এ ছাড়া পৌনঃপুনিক প্রয়াসের সংজ্ঞা ও সমবায়ের নিয়ম এই এককে প্রকাশ পেয়েছে।

৪১.৫ অনুশীলনী

১. একটি হোটেল তার ক্রেতাদের A ও B নামক দুটি বিশেষ পদ পরিবেশন করে। হোটেলটির ক্রেতাদের মধ্যে ৬০% পুরুষ এবং ৪০% মহিলা। পুরুষদের মধ্যে ৮০% পদ A ফরমাস করে এবং বাকিরা পদ B ফরমাস করে। মহিলাদের ভেতর ৭০% পদ B ফরমাস করে এবং বাকিরা পদ A ফরমাস করে। কী অনুপাতে হোটেলটি A এবং B প্রস্তুত করবে?

২. একটি বাক্সে তিনটি লাল এবং সাতটি সাদা বল আছে। একটি বলকে এলোপাতাড়িভাবে তোলা হল এবং তার জায়গায় অন্য রঙের একটি বল রাখা হল। এখন আবার একটি এলোপাতাড়িভাবে তোলা হল। এই বলটির লাল রঙ হওয়ার সম্ভাবনা নির্ণয় করুন।

৩. একটি ছক্কাকে ছয়বার চালা হলে অন্তত একবার ছয় পাওয়ার সম্ভাবনা নির্ণয় করুন।

৪. ধরা যাক, I, II এবং III হল তিনটি পাত্র। তাদের ভেতর যথাক্রমে (একটি সাদা, দুটি কালো, তিনটি লাল) ; (দুটি সাদা, একটি কালো, একটি লাল) এবং (চারটি সাদা, পাঁচটি কালো, তিনটি লাল) বল আছে। একটি পাত্র এলোপাতাড়ি চয়ন করা হল এবং দুটি বল তোলা হল। মনে করি, সেগুলি হল সাদা এবং লাল। সেগুলি যে পাত্র I, II বা III থেকে পাওয়া গেল, তার সম্ভাবনা কী?

৫. ধরে নিই, A ও B হল কোন পরীক্ষার দুটি সম্ভাব্য ঘটনা। ভাবা যায় যে, $P(A) = 0.4$, $P(B) = p$ এবং $P(A \cup B) = 0.7$।

(ক) p-এর কোন মানের জন্য A ও B পরস্পর বিচ্ছিন্ন হবে?

(খ) p-এর কোন মানের জন্য A ও B স্বতন্ত্র হবে?

৪১.৬ উত্তরমালা

১. ধরে নিই A : ক্রেতা A পদটি ফরমাস করে

B : ক্রেতা B পদটি ফরমাস করে

M : ক্রেতাটি পুরুষ

W : ক্রেতাটি মহিলা

অতএব, আমাদের দেওয়া আছে

$$P(M) = 0.6,\ P(W) = 0.4$$
$$P(A|M) = 0.8,\ P(B|M) = 0.2$$
$$P(A|W) = 0.3,\ P(B|W) = 0.7$$

যেহেতু মহিলা অথবা পুরুষ পদ A ফরমাস করে, আমরা A ঘটনাটিকে একটি বিচ্ছিন্ন সংযোগ বলে দেখতে পারি, যেটি হল,

$$A = (A \cap M) \cup (A \cap W)$$

সুতরাং, ক্রেতা পদ A ফরমাস করে, এই ঘটনাটির সম্ভাবনা হল

$$\begin{aligned} P(A) &= P[(A \cap M) \cup (A \cap W)] = P(A \cap M) + P(A \cap W) \\ &= P(M)\,P(A|M) + P(W)\,P(A|W) \\ &= 0.6 \times 0.8 + 0.4 \times 0.3 \\ &= 0.48 + 0.12 = 0.6 \end{aligned}$$

সেইরকমই, ক্রেতা পদ B ফরমাস করে, এই ঘটনাটির সম্ভাবনা হল

$$\begin{aligned}P(B) &= P[(B\cap M) \cup (B\cap W)] = P(B\cap M) + P(B\cap W)\\ &= P(M)\ P(B|M) + P(W)\ P(B|W)\\ &= 0.6 \times 0.2 + 0.4 \times 0.7\\ &= 0.12 + 0.28 = 0.4\end{aligned}$$

সুতরাং, হোটেলটি 0.6 : 0.4 = 3 : 2, এই অনুপাতে A এবং B, এই দুটি পদ প্রস্তুত করবে।

২. আমরা নিম্নলিখিত ঘটনাগুলি নির্দেশ করছি :

A : প্রথম টানে সাদা বল তোলা

B : প্রথম টানে লাল বল তোলা

C : দ্বিতীয় টানে লাল বল তোলা

তাহলে আমরা পাই : $P(A) = 7/10$ এবং $P(B) = 3/10$

দ্বিতীয় টানে লাল বল ওঠার ঘটনাটি দুটি পরস্পর বিচ্ছিন্ন উপায়ে সংঘটিত হতে পারে।

(ক) প্রথম টানে সাদা বল ওঠে এবং দ্বিতীয় টানে লাল বল ওঠে অর্থাৎ $(A\cap C)$ ঘটনাটি ঘটে

(খ) প্রথম টানে লাল বল ওঠে এবং দ্বিতীয় টানেও লাল বল ওঠে অর্থাৎ $(B\cap C)$ ঘটনাটি ঘটে

অতএব, সম্ভাবনার যোগের উপপাদ্য অনুসারে, এই ক্ষেত্রে সম্ভাবনা হল,

$$\begin{aligned}P(C) = P(A\cap C) + P(B\cap C) &= P(A)\ P(C|A) + P(B)\ P(C|B)\\ &= 7/10\ P(C|A) + 3/10\ P(C|B)\end{aligned}$$

$P(C|A)$ = P(যখন প্রথম টানে সাদা বল ওঠে তখন দ্বিতীয় টানে লাল বল ওঠার সম্ভাবনা) যদি প্রথম টানে সাদা বল ওঠে এবং তারপর সেটিকে অন্য রঙের (এখানে লাল) বল দিয়ে বদল করা হয়, তখন দ্বিতীয় টানের জন্য বাক্সে চারটি লাল এবং ছয়টি সাদা বল থাকে।

অতএব, $P(C|A) = 4/10$।

$P(C|B)$ = P(যখন প্রথম টানে লাল বল ওঠে তখন দ্বিতীয় টানে লাল বল ওঠার সম্ভাবনা) যদি প্রথম টানে লাল বল ওঠে তখন দ্বিতীয় টানের জন্য বাক্সে দুইটি লাল এবং আটটি সাদা বল থাকে।

অতএব, $P(C|B) = 2/10$

তাহলে, $P(A\cap C) + P(B\cap C) = P(A)P(C|A) + P(B)\ P(C|B)$

$$\begin{aligned}&= (7/10)(4/10) + (3/10)(2/10)\\ &= 28/100 + 6/100 = 34/100 = 0.34.\end{aligned}$$

৩. ধরা যাক, একটি ছক্কাকে চালা হলে i নম্বর চালনায় ছয় পাওয়ার ঘটনা হল A_i ($i = 1,2,3,4,5,6$)। তাহলে $P(A_i) = 1/6$ এবং $P(A_i') = 5/6$।

একটি ছক্কাকে ছয়বার চালা হলে অন্তত একবার ছয় পাওয়ার সম্ভাবনা হল

$P(A_1 \cup A_2 \cup A_3 \cup A_4 \cup A_5 \cup A_6) = 1 - P(A_1' \cap A_2' \cap A_3' \cap A_4' \cap A_5' \cap A_6')$

$= 1 - P(A_1')\, P(A_2')\, P(A_3')\, P(A_4')\, P(A_5')\, P(A_6')$

(যেহেতু ছক্কার চালনাগুলি পরস্পর স্বতন্ত্র, এবং সেই কারণে A_i এবং A_i' পরস্পর স্বতন্ত্র।)

$= 1 - (5/6)^6$ ।

৪. ধরা যাক, পাত্র I, পাত্র II এবং পাত্র III-কে চয়ন করার ঘটনা হল যথাক্রমে A_1, A_2 এবং A_3 এবং চয়ন করা পাত্রটি থেকে সাদা এবং লাল, এই দুটি বল তোলার ঘটনাটি হল A। তাহলে আমরা পাই,

	A_1	A_2	A_3
$P(A_i)$	1/3	1/3	1/3
$P(A \mid A_i)$	$(1\times 3)/({}^6C_2)$ $= 1/5$	$(2\times 1)/({}^4C_2)$ $= 1/3$	$(4\times 3)/({}^{12}C_2)$ $= 2/11$
$P(A \cap A_i) =$ $P(A_i)$ $P(A \mid A_i)$	1/15	1/9	2/33

আমরা জানি, $\sum P(A_i)P(A \mid A_i) = 1/15 + 1/9 + 2/33 = 118/495$

অতএব, Bayes-এর সূত্র অনুযায়ী পাত্র I থেকে দুটি সাদা এবং লাল বল টানার সম্ভাবনা হল

$$P(A_1 \mid A) = [P(A_1)\, P(A \mid A_1)]/\sum P(A_i)P(A \mid A_i) = (1/15)/(118/495) = 33/118$$

সেইরকম,

$$P(A_2 \mid A) = [P(A_2)\, P(A \mid A_2)]/\sum P(A_i)P(A \mid A_i) = (1/9)/(118/495) = 55/118$$

$$P(A_3 \mid A) = [P(A_3)\, P(A \mid A_3)]/\sum P(A_i)P(A \mid A_i) = (2/33)/(118/495) = 30/118$$ ।

৫. আমরা জানি যে,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

অথবা, $P(A \cap B) = P(A) + P(B) - P(A \cup B) = 0.4 + p - 0.7 = p - 0.3$

যদি A ও B পরস্পর বিচ্ছিন্ন হয়, তবে $P(A \cap B) = 0$, অর্থাৎ $p - 0.3 = 0$, অর্থাৎ, $p = 3$

A ও B স্বতন্ত্র হবে, যদি $P(A \cap B) = P(A)P(B)$, অর্থাৎ $p - 0.3 = (0.4)p$ অর্থাৎ, $p = 0.5$।

৪১.৭ গ্রন্থপঞ্জী

- দাশগুপ্ত, ভাগবত ; চৌধুরী অরিজিৎ ; দাস, বিশ্বনাথ : *রাশিবিজ্ঞানের পরিভাষা*, বিশ্বভারতী, (১৯৭২)।
- সেন, রাজকুমার : *সংখ্যাতত্ত্ব*, পশ্চিমবঙ্গ রাজ্য পুস্তক পর্ষদ, (১৯৮৬)।
- Cramer, H. : *The Elements of Probability Theory*, John Wiley, (1955)।
- Das, N. G. : *Statistical Methods in Commerce, Accountancy and Economics*, M. Das, (1985)।
- Freund, G. E. : *Modern Elementary Statistics*, Prentice Hall, New Delhi, (1979)।
- Goon, A. M., Gupta, M. K. & Dasgupta, B. : *Fundamentals of Statistics*, Vols. I, II, World Press, (1986)।
- Mathai, A. M. & Rathie, P. N. : *Probability and Statistics*, Macmillan, Madras, (1977)।
- Nagar, A. L. & Das, R. K. : *Basic Statistics*, Oxford University Press, Delhi, (1983)।
- Spiegel, M. R. : *Theory and Problems of Statistics*, Schaum Publishing Co., (1982)।

একক ৪২ ☐ সম্ভাবনা নিবেশন (Distribution Function)

গঠন

৪২.০ উদ্দেশ্য

এই এককটি থেকে জানা যাবে—

- সম্ভাবনা নিবেশন কাকে বলে
- বিচ্ছিন্ন এবং অবিচ্ছিন্ন সম্ভাবনাশ্রয়ী চলের পার্থক্য, তাদের সম্ভাবনা অপেক্ষক এবং তাদের সঙ্গে সংশ্লিষ্ট সম্ভাবনা নিবেশনের ধারণা
- বিচ্ছিন্ন এবং অবিচ্ছিন্ন সম্ভাবনাশ্রয়ী চলের সম্ভাবনা নিবেশনের বিশিষ্ট লক্ষণসমূহ
- বিচ্ছিন্ন এবং অবিচ্ছিন্ন সম্ভাবনাশ্রয়ী চলের কয়েকটি সম্ভাবনা নিবেশনের বর্ণনা।

৪২.১ প্রস্তাবনা

একক ৪০-এ আমরা সম্ভাবনাশ্রয়ী চলের ধারণাটির সাথে পরিচিত হয়েছি। সম্ভাবনাশ্রয়ী চল দুই রকম হয়—বিচ্ছিন্ন সম্ভাবনাশ্রয়ী চল এবং অবিচ্ছিন্ন সম্ভাবনাশ্রয়ী চল। এই এককে আমরা এই দুই ধরনের সম্ভাবনাশ্রয়ী চলের বিভিন্ন আঙ্গিক নিয়ে আলোচনা করব। এই প্রসঙ্গে বিচ্ছিন্ন সম্ভাবনাশ্রয়ী চলের বিবরণ বিস্তৃতভাবে এবং অবিচ্ছিন্ন সম্ভাবনাশ্রয়ী চলের বিবরণ সংক্ষিপ্তভাবে আলোচিত হবে। এই পর্যায়ে আমরা বিচ্ছিন্ন ও অবিচ্ছিন্ন সম্ভাবনাশ্রয়ী চলের প্রতিটির তত্ত্বগত নিবেশনের দুটি করে বিশেষ রূপের বৈশিষ্ট্য ও প্রয়োগ আলোচনা করব যেগুলি বাস্তবে সংখ্যাতাত্ত্বিক কাজে বিশেষ প্রয়োজনীয়। এই উদ্দেশ্যে বিচ্ছিন্ন চলের ক্ষেত্রে আমরা দ্বিপদ (binomial) এবং পোয়াসঁ (Poisson) নিবেশন দুটিকে এবং অবিচ্ছিন্ন চলের ক্ষেত্রে আয়তাকার (rectangular) নিবেশন ও সুষম (normal) নিবেশনকে নির্বাচন করেছি। এগুলি হল একচলক তত্ত্বগত নিবেশনের কয়েকটি রূপ।

৪২.২ বিচ্ছিন্ন সম্ভাবনাশ্রয়ী চল, সম্ভাবনা ভর অপেক্ষক এবং নিবেশন অপেক্ষক

৪২.২.১ বিচ্ছিন্ন সম্ভাবনাশ্রয়ী চল (discrete random variable)

কোন সম্ভাবনাশ্রয়ী চল X যদি x_i $(i = 1, 2, \ldots\ldots, r, \ldots\ldots)$ বিচ্ছিন্ন মানসমূহ গ্রহণ করতে পারে এবং প্রতিটি মানের ক্ষেত্রে $X = x_i$ হওয়ার ঘটনাটির একটি নির্দিষ্ট সম্ভাবনা $P(X = x_i) = p_i$ (≥ 0) পাওয়া যায়, তবে আমরা X-কে বিচ্ছিন্ন সম্ভাবনাশ্রয়ী চল বলব। এই বিচ্ছিন্ন মানসকল সসীম বা অসীম, দুই হতে পারে।

উদাহরণ :

একটি নিখুঁত মুদ্রাকে দুবার ক্ষেপণ করা হলে পরীক্ষার ফলসমূহ অর্থাৎ নমুনা দেশ হবে {HH, HT, TH, TT} । এই পরীক্ষা থেকে বিভিন্ন রকম বিচ্ছিন্ন চলের (X) উদ্ভব হতে পারে, যেমন মুদ্রার উপরিভাগে H দৃষ্ট হওয়া, HH, H-T প্রাপ্ত হওয়া, ইত্যাদি । ধরা যাক, এই উদাহরণে, X বলতে আমরা বোঝাই মুদ্রার উপরিভাগে H দৃষ্ট হওয়া । প্রতিটি নমুনা বিন্দুর সাথে আমরা X-এর একটি মান সংশ্লিষ্ট করতে পারি ।

সারণি ৪২.১

X	নমুনা বিন্দু
2	HH
1	HT
1	TH
0	TT

৪২.২.২ সম্ভাবনা অপেক্ষক (probability function)

ধরা যাক, X-এর সম্ভাব্য আরোহী (increasing) মানসমূহ হল x_1, x_2, x_3, ... এবং x-এর এই মানগুলির সাথে সংশ্লিষ্ট সম্ভাবনার মান হল যথাক্রমে $P(X = x_i) = p_i = f(x_i)$, $i = 1, 2, 3$ ...... । অর্থাৎ, যে কোন বিচ্ছিন্ন সম্ভাবনাশ্রয়ী চল X-এর সঙ্গে সংশ্লিষ্ট সম্ভাবনা ভর অপেক্ষক বলি (probability mass function, সংক্ষেপে pmf) । f(x) সম্ভাবনা অপেক্ষক হবে যদি

১. সমস্ত x-এর ক্ষেত্রে $f(x) \geq 0$ হয়, এবং

২. $\sum_x f(x) = 1$ হয় ।

f(x)-এর দ্বারা আমরা x-এর বিভিন্ন মানের জন্য সম্ভাবনার মান পেতে পারি । যখন $X = x_i$, এই অপেক্ষক হয় $P(X = x_i) = f(x_i)$, এবং x-এর অন্য মানের জন্য $f(x) = 0$ ।

উদাহরণ :

উপরিউল্লেখিত উদাহরণের পরিপ্রেক্ষিতে বলা যায় যে, এক্ষেত্রে P(HH) = P(HT) = P(TH) = P(TT) = 1/4 । তাহলে,

$P(X=0) = P(TT) = 1/4$ ।

$P(X=1) = P(HT \cup TH) = P(HT) + P(TH) = 1/4 + 1/4 = 1/2$ ।

$P(X=2) = P(HH) = 1/4$ ।

এই ক্ষেত্রে, ভর অপেক্ষক হল

সারণি ৪২.২

x	f(x)
0	1/4
1	1/2
2	1/4

৪২.২.৩ নিবেশন অপেক্ষক (distribution function)

যে কোন সম্ভাবনাশ্রয়ী চল X-এর জন্য আমরা একটি ক্রমযৌগিক নিবেশন অপেক্ষক (cumulative distribution function) বা সংক্ষেপে নিবেশন অপেক্ষকের সংজ্ঞা দিতে পারি। এই অপেক্ষককে আমরা $F(x)$ দিয়ে সূচিত করি। $F(x) = P(x \leq X)$, যেখানে X একটি বিচ্ছিন্ন সম্ভাবনাশ্রয়ী চল, এবং x প্রকৃত মানবাহী সংখ্যা $(-\infty \leq x \leq \infty)$।

সম্ভাবনা অপেক্ষক $f(x)$ থেকে নিম্নলিখিত উপায়ে আমরা নিবেশন অপেক্ষক পেতে পারি।

$F(x) = P(x \leq X) = \sum_{-\infty \leq u \leq u} f(u)$, u যদি বিচ্ছিন্ন চল হয়। এখানে $f(u)$ হল U সম্ভাবনাশ্রয়ী চলের সম্ভাবনা ভর অপেক্ষক এবং x একটি সুনির্দিষ্ট মান।

অন্যদিকে, নিবেশন অপেক্ষক থেকেও আমরা সম্ভাবনা অপেক্ষক পেতে পারি।

যদি X কেবল সসীম সংখ্যক মান $x_1, x_2, \ldots, x_n$ গ্রহণ করে, তবে X-এর নিবেশন অপেক্ষক হয়

$$
\begin{aligned}
F(x) &= 0 && -\infty < x < x_1 \\
F(x) &= f(x_1) && x_1 \leq x < x_2 \\
F(x) &= f(x_1) + f(x_2) && x_2 \leq x < x_3 \\
&\cdots\cdots\cdots\cdots\cdots\cdots \\
F(x) &= f(x_1) + \ldots\ldots\ldots\ldots + f(x_n) && x_n \leq x < \infty
\end{aligned}
$$

উদাহরণ :

আবার আমরা উপরিউল্লেখিত উদাহরণের শরণ নিই। এক্ষেত্রে পরিপ্রেক্ষিতে বলা যায় যে, X-এর সম্ভাবনা নিবেশন হয় এইরকম,

$$
\begin{aligned}
F(x) &= 0 && -\infty < x < 0 \\
F(x) &= 1/4 && 0 \leq x < 1 \\
F(x) &= 3/4 && 1 \leq x < 2 \\
F(x) &= 1 && 2 \leq x < \infty
\end{aligned}
$$

উপরে নির্মিত নিবেশন অপেক্ষকটির কয়েকটি বৈশিষ্ট্য আছে যেগুলি সাধারণভাবে সত্য।

x-এর মান 0, 1, 2-তে F(x)-এর লম্ফগুলি (jumps) যথাক্রমে 1/4, 1/2, 1/4। F(x)-কে সোপান অপেক্ষক (step function) বলা হয়। যখন $x = 1$, তখন $F(x) = 3/4$ (1/2 নয়)। গাণিতিকভাবে বলা যায় যে, নিবেশন অপেক্ষক রেখা ডানদিক থেকে 0, 1, 2 মানে অবিচ্ছিন্ন। যত আমরা বাম দিক থেকে ডানদিকে যাই, নিবেশন অপেক্ষক হয় একই থাকে বা 0 থেকে 1 এর দিকে বৃদ্ধি পায়। এজন্য F(x)-কে বলা হয় একগতি আরোহী (monotonically increasing) অপেক্ষক। যখন $a < b$, তখন $F(a) \leq F(b)$ এবং F(x) রেখা ডানদিকে অবিচ্ছিন্ন।

$$F(-\infty) = 0$$
$$F(\infty) = 1$$

স্মর্তব্য : এখানে যে সমস্ত নিবেশনের বিষয় আলোচিত হয়েছে সেগুলি সবই তত্ত্বগত নিবেশন (theoretical distribution), কারণ এগুলি হল নিবেশনের কিছু আদর্শ রূপ এবং বাস্তবে এদের উদাহরণ পাওয়া সম্ভব নয়। কিন্তু এই সমস্ত নিবেশনের সাহায্যে আমরা বাস্তবে কোন চলের নিবেশনকে যথাসম্ভব ঘনিষ্টভাবে অনুসরণ করে সেই নিবেশন সম্পর্কে সিদ্ধান্ত গ্রহণ করতে পারি। তত্ত্বগত নিবেশনের আলোচনার গুরুত্ব এইখানেই।

৪২.৩ বিচ্ছিন্ন সম্ভাবনাশ্রয়ী চলের নিবেশন—দ্বিপদ নিবেশন

৪২.৩.১ দ্বিপদ নিবেশনের সংজ্ঞা

সুইস্ (Swiss) গণিতবিদ্ James Bernoulli (১৬৫৪—১৭০৫) ১৭০০ খ্রিস্টাব্দে দ্বিপদ নিবেশন আবিষ্কার করেন। তাঁর মৃত্যুর আট বছর পর ১৭১৩ খ্রিস্টাব্দে এই তত্ত্বটি প্রথম প্রকাশিত হয়। দ্বিপদ নিবেশনকে তাই Bernoulli নিবেশনও বলা হয়।

দ্বিপদ নিবেশনের ক্ষেত্রে নিম্নলিখিত শর্তগুলি প্রযুক্ত হয়ে থাকে।

- কোন পৌনঃপুনিক প্রয়াসের পরীক্ষায় কোন সমগ্রকের মোট উপাদানের সংখ্যা হল n। পরস্পর স্বতন্ত্র প্রয়াসের সংখ্যা n একটি ধনাত্মক পূর্ণসংখ্যা এবং n-এর মান খুব বড়।
- পৌনঃপুনিক প্রয়াসের পরীক্ষায় পৌনঃপুনিক প্রয়াসসমূহ n হল সসীম এবং নির্দিষ্ট।
- প্রতিটি প্রয়াসে দুটি সম্ভাব্য পরস্পর বর্জনকারী ফল—সাফল্য (success) ও ব্যর্থতা (failure)।
- পৌনঃপুনিক প্রয়াসগুলি পরস্পর স্বতন্ত্র এই অর্থে যে, একটি প্রয়াসের ফল পরবর্তী অন্য প্রয়াসের ফলকে প্রভাবিত করে না বা পূর্ববর্তী কোন প্রয়াসের ফল দ্বারা প্রভাবিত হয় না।
- প্রতিটি প্রয়াসের ক্ষেত্রে সাফল্যের সম্ভাবনা হল p, ব্যর্থতার সম্ভাবনা হল q এবং p ও q সব প্রয়াসেই অপরিবর্তিত থাকে। $0 < p < 1$; $0 < q < 1$; $p + q = 1$; $q = 1 - p$।

n মোট দুই ধরনের উপাদানে বিভক্ত ও এই দুই ভাগের উপাদান সংখ্যা হল যথাক্রমে np ও nq, ($n = np + nq$, ও $p + q = 1$)।

৪২.৩.২ দ্বিপদ নিবেশনের সম্ভাবনা ভর অপেক্ষক

যদি উপরি উল্লেখিত শর্তগুলি পূরণ করে n-সংখ্যক স্বতন্ত্র প্রয়াসে X-সংখ্যক সাফল্যের সম্ভাবনা (অর্থাৎ, n-X সংখ্যক ব্যর্থতার সম্ভাবনা) নির্দেশিত হয়, তাহলে X হল এমন একটি সম্ভাবনাশ্রয়ী চল যা 0, 1, 2, …, n মানসমূহ গ্রহণ করতে পারে। অর্থাৎ, n-সংখ্যক স্বতন্ত্র প্রয়াসে আমরা কোন সাফল্য নাও পেতে পারি, অথবা একটি, দুটি বা সবকটি সাফল্য (=n) পেতে পারি।

আমরা 0, 1, 2, …, n সাফল্যের সঙ্গে সংশ্লিষ্ট সম্ভাবনাসমূহ নির্ণয় করতে চাই। rটি সাফল্যের সঙ্গে সংশ্লিষ্ট সম্ভাবনার সাধারণ রূপটি হল $P(X=r) = n_{C_r}\, p^r\, q^{n-r}$ যেখানে $r = 0, 1, 2, \ldots, n$।

প্রমাণ : কোন একটি নির্দিষ্ট ক্রম, ধরা যাক, SFSSFF …… SSF অনুসারে যেখানে r বার সাফল্য (S) ঘটে এবং n–r বার ব্যর্থতা (F) ঘটে। সেখানে n-সংখ্যক স্বতন্ত্র প্রয়াসে r-সংখ্যক সাফল্যের সম্ভাবনা (অর্থাৎ, n–r সংখ্যক ব্যর্থতার সম্ভাবনা) হল

$P(S\cap F\cap S\cap S\cap F\cap F\cap \ldots\ldots\ldots\ldots \cap S\cap S\cap F)$

$= P(S)P(F)P(S)P(S)P(F)P(F) \ldots\ldots P(S)P(S)P(F)$

(যৌগিক সম্ভাবনার উপপাদ্য অনুসারে যেহেতু প্রয়াসগুলি স্বতন্ত্র)।

$= pqppqq \ldots\ldots\ldots ppq$ = [r-সংখ্যক বার p এবং n – r-সংখ্যক বার q] $= p^r\, q^{n-r}$

কিন্তু n-সংখ্যক স্বতন্ত্র প্রয়াসে r-সংখ্যক সাফল্যের সম্ভাবনা এবং n–r সংখ্যক ব্যর্থতার সম্ভাবনা মোট যে কটি পরস্পর বর্জনকারী সম্ভাব্য উপায়ে পাওয়া যায়, তা হল $n!\,/\,r!(n-r)! = {}^nC_r$। এই পরস্পর বর্জনকারী উপায়ের সম্ভাবনা হল $p^r\, q^{n-r}$ এবং এগুলি সব সমান। অতএব, n-সংখ্যক স্বতন্ত্র প্রয়াসে যে কোন ক্রমানুসারে r-সংখ্যক সাফল্যের সম্ভাবনা এবং n–r সংখ্যক ব্যর্থতার সম্ভাবনা হল $P(X=r) = p^r\, q^{n-r} + p^r\, q^{n-r} + \ldots + p^r\, q^{n-r}$ (${}^nC_r$ সংখ্যক পদ) $= {}^nC_r\, p^r\, q^{n-r}$ যেখানে $r = 0, 1, 2, \ldots, n$।

মন্তব্য ১ : $P(X=r) = {}^nC_r\, p^r\, q^{n-r}$ সূত্রটিতে r-এর মান 0, 1, 2, …, n নিযুক্ত করলে আমরা n-সংখ্যক স্বতন্ত্র প্রয়াসে যথাক্রমে 0, 1, 2, …, n-সংখ্যক সাফল্যের সম্ভাবনা নিরূপণ করতে পারি। নিম্নলিখিত সারণিতে এটি দেখানো হল।

সারণি ৪২.৩

r	P(X = r)
0	${}^nC_0\, p^0\, q^n = q^n$
1	${}^nC_1\, p^1\, q^{n-1}$
2	${}^nC_2\, p^2\, q^{n-2}$
.	.
.	.
.	.
.	.
.	.
n	${}^nC_n\, p^n\, q^{n-n} = p^n$

যেহেতু এই সম্ভাবনাগুলি হল দ্বিপদ সম্প্রসারণ $(q + p)^n$-এর ধারাবাহিক পদসমূহ, তাই এই নিবেশনটিকে বলা হয় দ্বিপদ নিবেশন। এই নিবেশনে পূর্ণকাঙ্কের সংখ্যা দুটি, n এবং p এবং সমস্ত r-এর ক্ষেত্রে $P(X=r) \geq 0$।

মন্তব্য ২ : $\sum^{n}_{r=0} P(X=r) = P(X=0) + P(X=1) + P(X=2) + \ldots\ldots + P(X=n)$

$= q^n + {}^nC_1\, p\, q^{n-1} + {}^nC_2\, p^2 q^{n-2} + \ldots\ldots + p^n$

$= \sum^{n}_{r=0} {}^nC_r\, p^r q^{n-r} = (q + p)^n = 1$ যেহেতু $q + p = 1$

মন্তব্য ৩ : $P(X=r) = {}^nC_r\, p^r q^{n-r}$ পদটিকে দ্বিপদ নিবেশনের সম্ভাবনা ভর অপেক্ষক বলা হয়, যার পূর্ণকাঙ্কের সংখ্যা দুটি, n এবং p। সম্ভাবনাশ্রয়ী চল X-কে বলা হয় দ্বিপদ চল। যদি n এবং p জানা থাকে, তবে দ্বিপদ নিবেশনের সবকটি সম্ভাবনা নির্ণয় করা যায়। অবশ্যই p জানা থাকলে q জানা যায়, কারণ $q = 1 - p$।

মন্তব্য ৪ : সম্ভাবনাশ্রয়ী চল X যেহেতু কেবলমাত্র বিচ্ছিন্ন মানসমূহ গ্রহণ করে, তাই দ্বিপদ নিবেশন হল বিচ্ছিন্ন সম্ভাবনা নিবেশন (discrete probability distribution)।

মন্তব্য ৫ : n-সংখ্যক প্রয়াসের জন্য দ্বিপদ নিবেশনে $(n + 1)$-সংখ্যক পদ থাকে। ধারাবাহিকভাবে দ্বিপদ সহগগুলি (binomial coefficients) হল ${}^nC_0$, ${}^nC_1$, ${}^nC_2$, ${}^nC_{n-1}$, ${}^nC_n$। যেহেতু ${}^nC_0 = {}^nC_n = 1$, সেহেতু প্রথম ও শেষ সহগ সর্বদা 1-এর সমান হবে। আবার, যেহেতু, ${}^nC_r = {}^nC_{n-r}$, সেহেতু দ্বিপদ সহগগুলি প্রতিসম (symmetric distribution)। এছাড়া, x-এর বিভিন্ন মানের জন্য আমরা পাই, $(1 + x)^n = {}^nC_0 + {}^nC_1 x^1 + {}^nC_2 x^2 + \ldots\ldots + {}^nC_n x^n$। $x = 1$ হলে আমরা পাই, $(1 + 1)^n = {}^nC_0 + {}^nC_1 + {}^nC_2 + \ldots\ldots + {}^nC_n = 2^n$।

n-এর বিভিন্ন মানের জন্য দ্বিপদ সহগগুলির মান আমরা সহজভাবে পেতে পারি "Pascal-এর ত্রিকোণ" (Pascal's triangle) থেকে।

সারণি ৪২.৪

"Pascal-এর ত্রিকোণ"—$(a + b)^n$ সম্প্রসারণে পদগুলির সহগ

n-এর মান	দ্বিপদ সহগ	যোগফল (2^n)
1	1, 1	2
2	1, 2, 1	4
3	1, 3, 3, 1	8
4	1, 4, 6, 4, 1	16
5	1, 5, 10, 10, 5, 1	32
6	1, 6, 15, 20, 15, 6, 1	64
7	1, 7, 21, 35, 35, 21, 7, 1	128
8	1, 8, 28, 56, 70, 56, 28, 8, 1	256
9	1, 9, 36, 84, 126, 126, 84, 36, 9, 1	512
10	1, 10, 45, 120, 210, 252, 210, 120, 45, 10, 1	1024

৪২.৩.৩ দ্বিপদ নিবেশনের ধ্রুবকগুলি

সারণি ৪২.৫

r	$P(X = r)$	$r.P(X = r)$	$r^2.P(X = r)$
0	q^n	0	0
1	${}^nC_1\ pq^{n-1}$	${}^nC_1\ pq^{n-1}$	$1^2\ {}^nC_1\ pq^{n-1}$
2	${}^nC_2\ p^2q^{n-2}$	$2{}^nC_2\ p^2q^{n-2}$	$2^2\ {}^nC_2\ p^2q^{n-2}$
3	${}^nC_3\ p^3\ q^{n-3}$	$3{}^nC_3\ p^3q^{n-3}$	$3^2\ {}^nC_3\ p^3q^{n-3}$
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
n	p^n	np^n	n^2p^n

দ্বিপদ নিবেশনের গড় মান (mean) হল

$$\mu = \sum r.P(X = r) = {}^nC_1\ pq^{n-1} + 2{}^nC_2\ p^2q^{n-2} + 3{}^nC_3\ p^3q^{n-3} + \ldots\ldots\ldots + np^n$$
$$= npq^{n-1} + 2[n(n-1)]/2!p^2q^{n-2} + 3[n(n-1)(n-2)]/3!p^3q^{n-3} + \ldots\ldots\ldots\ldots\ldots + np^n$$
$$= np[q^{n-1} + (n-1)pq^{n-2} + (n-1)(n-2)/2!p^2q^{n-3} + \ldots\ldots\ldots\ldots + p^{n-1}]$$
$$= np[q^{n-1} + (n-1)C_1pq^{n-2} + (n-1)C_2\ p^2q^{n-3} + \ldots\ldots\ldots\ldots + p^{n-1}]$$
$$= np(p+q)^{n-1} \quad \text{(দ্বিপদ সম্প্রসারণের মাধ্যমে)}$$
$$= np \quad \text{(যেহেতু } (p+q) = 1\text{)}$$

দ্বিপদ নিবেশনের ভেদ মান (variance) হল

$$\mu_2 = \sum r^2.P(X = r) - [\sum r.P(X = r)]^2 = \sum r^2.P(X = r) - [\text{mean}]^2$$

এখন, $\sum r^2.P(X = r) = 1^2{}^nC_1\ pq^{n-1} + 2^2{}^nC_2\ p^2q^{n-2} + 3^2{}^nC_3\ p^3q^{n-3} + \ldots\ldots\ldots + n^2p^n$

$$= npq^{n-1} + 4n(n-1)/2p^2q^{n-2} + 9n(n-1)(n-2)]/3!p^3q^{n-3} + \ldots\ldots\ldots\ldots\ldots + n^2p^n$$
$$= np[q^{n-1} + 2(n-1)pq^{n-2} + 3/2(n-1)(n-2)p^2q^{n-3} + \ldots\ldots\ldots + np^{n-1}]$$
$$= np[\{q^{n-1} + (n-1)pq^{n-2} + (n-1)(n-2)/2p^2q^{n-3} + \ldots\ldots\ldots + 1.p^{n-1}\}$$
$$+ \{(n-1)pq^{n-2} + (n-1)(n-2)p^2q^{n-3} + \ldots\ldots\ldots + (n-1)p^{n-1}\}]$$
$$= np[\{(q+p)^{n-1}\} + (n-1)p\{q^{n-2} + (n-2)pq^{n-3} + \ldots\ldots\ldots + p^{n-2}\}]$$
$$= np[(q+p)^{n-1} + (n-1)p(q+p)^{n-2}]$$
$$= np[1 + (n-1)p] \quad \text{(যেহেতু } (p+q) = 1\text{)}$$

তাহলে, দ্বিপদ নিবেশনের ভেদমান হল,

$$\mu_2 = np[1 + np - p] - (np)^2 = np[1 + np - p - np] = npq \text{ ।}$$

এই ভেদমান μ_2 হল এই নিবেশনের গড়কেন্দ্রিক দ্বিতীয় পরিঘাত (second order central moment)। দ্বিপদ নিবেশনের গড়কেন্দ্রিক তৃতীয় পরিঘাত (third order central moment) হল $\mu_3 = npq(q - p)$ । সেইরকম, দ্বিপদ নিবেশনের গড়কেন্দ্রিক চতুর্থ পরিঘাত (fourth order central moment) হল $\mu_4 = npq[1 + 3pq(n - 2)]$ ।

অতএব, দ্বিপদ নিবেশনের পরিঘাত-নির্ভর প্রতিবৈষম্যের পরিমাপ (moment measure of skewness) হল $\beta_1 = \mu_3^2 / \mu_2^3 = [npq(q - p)]^2 / (npq)^3 = [(q - p)]^2 / npq$ । $\gamma_1 = + \sqrt{\beta_1} = [(q - p)]/\sqrt{(npq)}$

দ্বিপদ নিবেশনের পরিঘাত-নির্ভর তীক্ষ্ণতার পরিমাপ (moment measure of kurtosis) হল $\beta_2 = \mu_4 / \mu_2^2 = [npq(1 + 3pq(n - 2)]/(npq)^2$

$= [1 + 3pq(n - 2)]/(npq) = 3 + (1 - 6pq) / npq$ ।

$\gamma_2 = \beta_2 - 3 = + (1 - 6pq)/npq$

মন্তব্য ১ : যেহেতু q হল ব্যর্থতার সম্ভাবনা, সেহেতু সর্বদা $0 < q < 1$ । সুতরাং, $\mu_2 = npq < np$ । অর্থাৎ, দ্বিপদ নিবেশনের ক্ষেত্রে $\mu_2 < \mu$ ।

মন্তব্য ২ : যখন $n \to \infty$, $\beta_1 \to 0$, $\gamma_1 \to 0$, $\beta_2 \to 3$ এবং $\gamma_2 \to 0$ ।

মন্তব্য ৩ : গড় মানের উভয়পার্শ্বে দ্বিপদ নিবেশন প্রতিসম হবে যদি $\beta_1 = 0$ হয় $\Rightarrow q = p = {}^1/_2$ [যেহেতু $(p+q) = 1$] । গড় মানের উভয়পার্শ্বে দ্বিপদ নিবেশনের প্রতিবৈষম্য ধনাত্মক হবে যদি $\gamma_1 > 0$ হয়, $\Rightarrow (q-p) > 0 \Rightarrow (1-2p) > 0 \Rightarrow 1 > 2p \Rightarrow 1/2 > p \Rightarrow p < {}^1/_2$; অর্থাৎ $p < 1/2 < q$; এক্ষেত্রে প্রতিবৈষম্য দক্ষিণায়ত । সেইরকম, গড় মানের উভয়পার্শ্বে দ্বিপদ নিবেশনের প্রতিবৈষম্য ঋণাত্মক হবে যদি $\gamma_1 < 0$ হয়, $\Rightarrow (q-p) < 0 \Rightarrow (1-2p) < 0 \Rightarrow 1 < 2p \Rightarrow 1/2 < p \Rightarrow p > {}^1/_2$; অর্থাৎ $p > 1/2 > q$; এক্ষেত্রে প্রতিবৈষম্য বামায়ত । p এবং q-এর মান নির্বিশেষে যত n বৃদ্ধি পায়, তত দ্বিপদ নিবেশনের প্রতিবৈষম্য হ্রাস পেতে থাকে ।

মন্তব্য ৪ : যদি $\gamma_2 = 0$ হয়, তবে $pq = p(1-p) = 1/6$ হবে । অতএব, $p^2 - p + 1/6 = 0$; বা, $p = 0.21135, 0.78865$ । দ্বিপদ নিবেশনটিকে অতিতীক্ষ্ণ (leptokurtic) বলা হয় যদি $\gamma_2 > 0$ হয়, অর্থাৎ $0 < pq < 1/6$ হয়, অর্থাৎ $0.78865 < p < 0.21135$ হয় ; অবতীক্ষ্ণ (platykurtic) বলা হয় যদি $\gamma_2 < 0$ হয়, অর্থাৎ $1/6 < pq < 1/4$ হয় অর্থাৎ $0.21135 < p < 0.78865$ হয় ; এবং, সুষম (mesokurtic) বলা হয় যদি $\gamma_2 = 0$ হয়, অর্থাৎ $pq = 1/6$ হয় অর্থাৎ $p = 0.21135$ বা, $p = 0.78865$ হয় ।

৪২.৩.৪ দ্বিপদ নিবেশনের ভূয়িষ্ঠক (mode)

ভূয়িষ্ঠক হল X-এর সেই মান যা সম্ভাবনা অপেক্ষকের সর্বোচ্চ মান দেয় । সুতরাং, যদি $X = r$ ভূয়িষ্ঠক দেয়, তাহলে আমাদের পেতে হবে $P(X=r) > P(X=r - 1)$ এবং $P(X=r) > P(X=r + 1)$ ।

দ্বিপদ নিবেশনের ক্ষেত্রে ভূয়িষ্ঠক নির্ধারণের নিয়ম হল এরকম । ধরা যাক যে, X হল একটি দ্বিপদ চল যার পূর্ণকাঙ্ক (parameter) হল n এবং p ।

অবস্থা ১ : যেখানে $(n+1)p$ হল একটি পূর্ণসংখ্যা (integer)।

ধরা যাক, $(n+1)p = k$ হল একটি পূর্ণসংখ্যা। এক্ষেত্রে নিবেশনটি দ্বি-ভূয়িষ্ঠক বিশিষ্ট, যেখানে দুটি ভূয়িষ্ঠক হল $X = k$ এবং $X = k - 1$। অতএব, যদি $n = 9$ এবং $p = 0.4$ হয়, তাহলে $(n+1)p = 10 \times 0.4 = 4$ হল একটি পূর্ণসংখ্যা। সুতরাং, এক্ষেত্রে নিবেশনটি দ্বি-ভূয়িষ্ঠক বিশিষ্ট, যেখানে দুটি ভূয়িষ্ঠক হল $X = 4$ এবং $X = 4 - 1 = 3$।

অবস্থা ২ : যেখানে $(n+1)p$ কোন পূর্ণসংখ্যা নয়।

ধরা যাক, $(n+1)p = (k_1+f)$ যেখানে k_1 হল একটি পূর্ণসংখ্যা এবং f হল একটি ভগ্নাংশ। এক্ষেত্রে $X = k_1$-এ নিবেশনটির একটিমাত্র ভূয়িষ্ঠক আছে, যেটি হল $(n+1)p$-এর পূর্ণসংখ্যা সম্বলিত অংশ। উদাহরণস্বরূপ বলা যায় যে, যদি $n = 7$ এবং $p = 0.6$ হয়, তাহলে $(n+1)p = 8 \times 0.6 = 4.8$। সুতরাং, এক্ষেত্রে নিবেশনটির ভূয়িষ্ঠক হল 4, যেটি হল 4.8-এর পূর্ণসংখ্যা সম্বলিত অংশ।

মন্তব্য ১ : যদি np একটি পূর্ণসংখ্যা হয়, তাহলে নিবেশনটি এক-ভূয়িষ্ঠক বিশিষ্ট (unimodal) এবং গড় মান = ভূয়িষ্ঠক = np।

৪২.৪ বিচ্ছিন্ন সম্ভাবনাশ্রয়ী চলের নিবেশন—পোয়াসঁ নিবেশন

৪২.৪.১ পোয়াসঁ নিবেশনের সংজ্ঞা

ফরাসী গণিতবিদ্ Simeon D. Poisson (১৭৮১—১৮৪০) ১৮৩৭ খ্রিস্টাব্দে পোয়াসঁ নিবেশন নির্মাণ করেন। এটিকে দ্বিপদ নিবেশনের একটি সীমন্তিক রূপ বলা হয়। দ্বিপদ নিবেশনের ক্ষেত্রে যদি p-এর মান অনির্দিষ্টভাবে অল্প হয় (অর্থাৎ $p \to 0$ হয়) এবং n-এর মান অনির্দিষ্টভাবে বেশি হয় (অর্থাৎ $n \to \infty$ হয়, অর্থাৎ n-এর মান অসীমের দিকে ধাবিত হয়) এবং এমনভাবে p এবং n-এর মানগুলি পরিবর্তিত হয় যে np-এর মান স্থির থাকে ($np = \lambda$ একটি ধ্রুবক), তবে প্রমাণ করা যায় যে, পোয়াসঁ নিবেশন হল দ্বিপদ নিবেশনের একটি সীমন্তিক রূপ।

দ্বিপদ নিবেশনের সম্ভাবনা ভর অপেক্ষক হল

$$
\begin{aligned}
{}^{n}C_r p^r q^{n-r} &= (n)!/(r)!(n-r)!\ p^r(1-p)^{n-r} \\
&= [n(n-1)(n-2) \ldots\ldots (n-r+1)]/(r)!(\lambda/n)^r(1-\lambda/n)^{n-r} \\
&= (\lambda)^r/(r)!(n/n)((n-1)/(n-2)/n) \ldots ((n-r+1)/n)(1-\lambda/n)^{n-r} \\
&= (\lambda)^r/(r)!\ (1-1/n)(1-2/n) \ldots\ldots (1-(r-1)/n)(1-\lambda/n)^n(1-\lambda/n)^{-r}
\end{aligned}
$$

$n \to \infty$ হলে দ্বিপদ নিবেশনের সীমন্তিক রূপ হয়

$$
\begin{aligned}
\text{limit}_{n \to \infty}\ {}^{n}C_r\ p^r q^{n-r} = (\lambda)^r/(r)!\ &\text{limit}_{n \to \infty} (1-1/n)\ \text{limit}_{n \to \infty} (1-2/n) \ldots\ldots\ldots \\
&\text{limit}_{n \to \infty} (1-(r-1)/n)\ \text{limit}_{n \to \infty} (1-\lambda/n)^n\ \text{limit}_{n \to \infty} (1-\lambda/n)^{-r}
\end{aligned}
$$

$n \to \infty$ হলে $1/n$, $2/n$, $x-1/n$, λ/n ইত্যাদির মান অত্যন্ত অল্প হবে এবং তার ফলে $(1-1/n)$, $(1-2/n)$, ... $(1-(x-1)/n)$, $(1-\lambda/n)$, ইত্যাদি পদগুলির মান প্রায় একের সমান হবে। তাহলে আমরা পাই,

$(\lambda)^r/(r)!(1-0)(1-0)$ …… $(1-0)\ \text{limit}_{n\to\infty}(1-\lambda/n)^n\ \text{limit}_{n\to\infty}(1-\lambda/n)^{-r}$ (*)

কিন্তু আমরা জানি যে,

$\text{limit}_{n\to\infty}(1+a/n)^n = e^a$

এবং, $\text{limit}_{n\to\infty}(1+a/n)^A = 1$, যেখানে A হল একটি ধ্রুবক যেটি n থেকে স্বতন্ত্র।

এই দুটি মানকে (*)-এ বসালে আমরা দ্বিপদ নিবেশনের সম্ভাবনা অপেক্ষকের সীমন্তিক রূপ পাই। সেটি হল $(\lambda)^r/(r)!\text{x}1\text{x}e^{-\lambda}\ \text{x}1 = e^{-\lambda}(\lambda)^r/(r)!$

n অসীমের দিকে আসন্ন $(n\to\infty)$ হলে $\text{limit}_{n\to\infty}(1-\lambda/n)^n = e^{-\lambda}$, যেখানে e হল Napier-এর logarithm-এর নিধান এবং একটি ধ্রুবক (e = 2.71828...)। অতএব, $\text{limit}_{n\to\infty}{}^nC_r\ p^r\ q^{n-r} = (\lambda)^r e^{-\lambda}/(r)!$।

৪২.৪.২ পোয়াসঁ নিবেশনের সম্ভাবনা ভর অপেক্ষক

অতএব, পোয়াসঁ নিবেশনের সম্ভাবনা ভর অপেক্ষকের সাধারণ রূপ হল $P(X=r) = (\lambda)^r e^{-\lambda}/(r)!$; $r = 0,1,2,$ ………।

সাধারণত যে সমস্ত ক্ষেত্রে $p < 0.1$ এবং $\lambda < 10$, সেইসব ক্ষেত্রে দ্বিপদ নিবেশনের স্থানে পোয়াসঁ নিবেশন ব্যবহার করা হয়। পোয়াসঁ নিবেশনকে "ক্ষুদ্র সংখ্যার বিধি" (Law of small numbers)-ও বলা হয়।

মন্তব্য ১ : পোয়াসঁ নিবেশন একটি বিচ্ছিন্ন সম্ভাবনা নিবেশন, কারণ X কেবল 0,1,2; …… ইত্যাদি বিচ্ছিন্ন মানসকল গ্রহণ পারে।

মন্তব্য ২ : $P(X = r) = (\lambda)^r e^{-\lambda}/(r)!$-এ $r = 0,1,2,$ ……… ইত্যাদি মানসমূহ বসালে আমরা যথাক্রমে 0,1,2, ……… সাফল্যের সম্ভাবনা নির্ণয় করতে পারি। নিম্নের সারণিতে এটি দেখানো হল।

সারণি ৪২.৬

সাফলের সংখ্যা (r)	সম্ভাবনা P(X=r)
0	$(\lambda)^0 e^{-\lambda}/(0)! = e^{-\lambda}$
1	$\lambda e^{-r}/(1)!$
2	$(\lambda)^2 e^{-\lambda}/(2)!$
3	$(\lambda)^3 e^{-\lambda}/(3)!$
.	.
.	.
.	.
.	.
.	.

মন্তব্য ৩ : মোট সম্ভাবনা = 1 ।

$$\sum_{r=0} P(X=r) = e^{-\lambda} + (\lambda)e^{-\lambda} + (\lambda)^2e^{-\lambda}/(2)! + (\lambda)^3e^{-\lambda}/(3)! + \ldots\ldots\ldots\ldots$$
$$= e^{-\lambda}[1 + \lambda + (\lambda)^2/(2)! + (\lambda)^3/(3)! + \ldots\ldots\ldots\ldots]$$
$$= e^{-\lambda}e^{\lambda} \qquad [\text{যেহেতু, } e^x = 1 + x + (x)^2/(2)! + (x)^3/(3)! + \ldots\ldots\ldots\ldots]$$
$$= e^{-\lambda+\lambda} = e^0 = 1$$

মন্তব্য ৪ : λ-এর মান জানা থাকলে পোয়াসঁ নিবেশনের সব সম্ভাবনা নির্ণয় করা যায় । λ-কে তাই পোয়াসঁ নিবেশনের পূর্ণকাঙ্ক বলা হয় ।

মন্তব্য ৫ : পোয়াসঁ নিবেশন যে দ্বিপদ নিবেশনের সীমন্তিক রূপ, সেটি আমরা প্রমাণ করেছি । নিম্নে আমরা পোয়াসঁ নিবেশনের অন্তঃস্থিত সাধারণ রূপটি স্থাপন করব ।

ধরে নিই যে, একটি 'সমসম্ভব' পারিসংখ্যানিক পরীক্ষা নিম্নলিখিত শর্তগুলি বহাল আছে ।

(i) একটি ক্ষুদ্র সময়ের অন্তর (t, t+dt)-তে সাফল্যের সম্ভাবনা (অর্থাৎ, ঘটনাটির ঘটন) হল λdt, অর্থাৎ, সাফল্যের সম্ভাবনা সময়ের অন্তরের পরিমাপের সঙ্গে প্রত্যক্ষভাবে আনুপাতিক ।

(ii) এই সময়ের অন্তরের মধ্যে একাধিক সাফল্যের সম্ভাবনা অত্যন্ত কম ।

(iii) এই ক্ষুদ্র সময়ের অন্তর (t, t+dt)-র মধ্যে কোন একটি নির্দিষ্ট সাফল্যের সম্ভাবনা হল যথার্থ সময় t এবং পূর্ববর্তী সাফল্যের সম্ভাবনার থেকে স্বতন্ত্র ।

এই শর্তগুলি রক্ষিত হলে দেখানো যায় যে, সময়ের অন্তরের পরিমাপ t-তে rটি সাফল্যের সম্ভাবনা হল $P_r(t) = (\lambda t)^r e^{-\lambda t}/(r)!$; $r = 0, 1, 2, \ldots\ldots$ । এটি হল পোয়াসঁ নিবেশনের সম্ভাবনা নিবেশন যার পূর্ণকাঙ্ক হল λt ।

৪২.৪.৩ পোয়াসঁ নিবেশনের ধ্রুবকগুলি

সারণি ৪২.৬

r	P(X = r)	r.P(X = r)	r^2.P(X = r)
0	$e^{-\lambda}$	0	0
1	$\lambda e^{-\lambda}/(1)!$	$\lambda e^{-\lambda}$	$\lambda e^{-\lambda}$
2	$\lambda^2 e^{-\lambda}/(2)!$	$2\lambda^2 e^{-\lambda}/(2)!$	$2^2\lambda^2 e^{-\lambda}/(2)!$
3	$\lambda^3 e^{-\lambda}/(3)!$	$3\lambda^3 e^{-\lambda}/(3)!$	$3^2\lambda^3 e^{-\lambda}/(3)!$
4	$\lambda^4 e^{-\lambda}/(4)!$	$4\lambda^4 e^{-\lambda}/(4)!$	$4^2\lambda^4 e^{-\lambda}/(4)!$
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.

পোয়াসঁ নিবেশনের গড় মান হল

$$\mu = \sum r.P(X = r) = \lambda e^{-\lambda} + 2(\lambda^2)e^{-\lambda}/(2)! + 3(\lambda^3)e^{-\lambda}/(3)! + 4(\lambda^4)e^{-\lambda}/(4)! + \ldots\ldots\ldots\ldots]$$

$$= \lambda e^{-\lambda}\ [1 + \lambda + (\lambda^2)/2! + (\lambda^3)/3! + \ldots\ldots\ldots\ldots]$$

$$= \lambda e^{-\lambda}\ e^{\lambda} = \lambda e^{\,-\lambda+\lambda} = \lambda$$

পোয়াসঁ নিবেশনের ভেদ মান হল

$$\mu_2 = \sum r^2.P(X = r) - [\sum r.P(X = r)]^2 = \sum r^2.P(X = r) - [\text{গড় মান}]^2$$

এখন, $\sum r^2.P(X = r) = \lambda e^{-\lambda} + 2^2(\lambda^2)e^{-\lambda}/(2)! + 3^2(\lambda^3)e^{-\lambda}/(3)! + 4^2(\lambda^4)e^{-\lambda}/(4)! + \ldots\ldots\ldots]$

$$= \lambda e^{-\lambda}\ [1 + 2\lambda + 3\lambda^2/(2)! + 4\lambda^3/(3)! + \ldots\ldots\ldots\ldots]$$

$$= \lambda e^{-\lambda}\ [\{1 + \lambda + \lambda^2/(2)! + \lambda^3/(3)! + \ldots\ldots\ldots\ldots\} + \lambda\{1 + \lambda + \lambda^2/(2)! + \lambda^3/(3)! + \ldots\ldots\ldots\ldots\}]$$

$$= \lambda e^{-\lambda}\ [e^{\lambda} + \lambda e^{\lambda}] = \lambda e^{-\lambda}\ e^{\lambda}(1 + \lambda) = \lambda(1 + \lambda)$$

তাহলে, পোয়াসঁ নিবেশনের ভেদ মান হল

$$\mu_2 = \lambda(1 + \lambda) - (\lambda)^2 = \lambda + \lambda^2 - \lambda^2 = \lambda$$

অতএব, λ পূর্ণকাঙ্কবিশিষ্ট পোয়াসঁ নিবেশনের জন্য গড়মান = ভেদমান = λ

এই ভেদমান μ_2 হল পোয়াসঁ নিবেশনের গড়কেন্দ্রিক দ্বিতীয় পরিঘাত। পোয়াসঁ নিবেশনের গড়কেন্দ্রিক তৃতীয় পরিঘাত হল $\mu_3 = \lambda$। সেইরকম, পোয়াসঁ নিবেশনের গড়কেন্দ্রিক চতুর্থ পরিঘাত হল $\mu_4 = \lambda + 3\lambda^2$।

অতএব, পোয়াসঁ নিবেশনের পরিঘাত-নির্ভর প্রতিবৈষম্যের পরিমাপ হল

$$\beta_1 = \mu_3^2/\mu_2^3 = \lambda^2/\lambda^3 = 1/\lambda\ ।$$

$$\gamma_1 = +\sqrt{\beta_1} = \sqrt{1/\lambda} = 1/\sqrt{\lambda}\ ।$$

পোয়াসঁ নিবেশনের পরিঘাত-নির্ভর তীক্ষ্ণতার পরিমাপ হল

$$\beta_2 = \mu_4/\mu_2^2 = (\lambda + 3\lambda^2)/\lambda^2 = 3 + 1/\lambda\ ।$$

$$\Rightarrow \gamma_2 = \beta_2 - 3 = 1/\lambda\ ।$$

মন্তব্য ১ : যখন $\lambda \to \infty$, $\beta_1 \to 0$, $\gamma_1 \to 0$, $\beta_2 \to 3$ এবং $\gamma_2 \to 0$।

মন্তব্য ২ : যেহেতু পোয়াসঁ নিবেশন হল দ্বিপদ নিবেশনের সীমান্তিক রূপ, অতএব আমরা দ্বিপদ নিবেশনে $np = \lambda$, $p \to 0$ $q \to 1$ ধরে পোয়াসঁ নিবেশনের ধ্রুবকগুলিকে নির্ণয় করতে পারি।

মন্তব্য ৩ : যেহেতু $\mu_3 = \lambda > 0$, অতএব, $\gamma_1 > 0$। এর অর্থ হল যে, পোয়াসঁ নিবেশনের প্রতিবৈষম্য হল ধনাত্মক। পূর্ণকাঙ্ক λ-এর মান যত বৃদ্ধি পায়, γ_1-এর মান তত হ্রাস পায়, এবং এই উপায়ে λ-এর মান বৃদ্ধির সাথে সাথে পোয়াসঁ নিবেশনের প্রতিবৈষম্য হ্রাস পেতে থাকে। বিশেষ করে, যখন $\lambda \to \infty$, তখন $\gamma_1 \to 0$ এবং ফলত λ-এর মান বেশি হলে পোয়াসঁ নিবেশন হয় প্রতিসম (symmetrical)।

৪২.৪.৪ পোয়াসঁ নিবেশনের ভূয়িষ্ঠক

পোয়াসঁ নিবেশনের ভূয়িষ্ঠক পাওয়া যায় $X = r$ এই মানে, যদি সেখানে $P(X = r) > P(X = r-1)$ এবং $P(X = r) > P(X = r + 1)$ হয় ।

অবস্থা ১ : যেখানে λ হল একটি পূর্ণসংখ্যা ।

ধরা যাক, λ হল একটি পূর্ণসংখ্যা যার মান হয় k । এক্ষেত্রে নিবেশনটি দ্বি-ভূয়িষ্ঠক বিশিষ্ট, যেখানে দুটি ভূয়িষ্ঠক হল $X = k$ এবং $X = k - 1$ ।

অবস্থা ২ : যেখানে λ কোন পূর্ণসংখ্যা নয়।

ধরা যাক, $\lambda = (k_1 + f)$ যেখানে k_1 হল একটি পূর্ণসংখ্যা এবং f হল একটি ভগ্নাংশ । এক্ষেত্রে $X = k_1$ মানে নিবেশনটির একটিমাত্র ভূয়িষ্ঠক আছে, যেটি হল λ-এর পূর্ণসংখ্যা সম্বলিত অংশ । উদাহরণস্বরূপ বলা যায় যে, যদি $\lambda = 5.6$, তবে নিবেশনটির ভূয়িষ্ঠক হল 5, যেটি হল 5.6-এর পূর্ণসংখ্যা সম্বলিত অংশ ।

৪২.৫ অবিচ্ছিন্ন সম্ভাবনাশ্রয়ী চল, সম্ভাবনা ঘনত্ব অপেক্ষক ও নিবেশন অপেক্ষক

৪২.৫.১ অবিচ্ছিন্ন সম্ভাবনাশ্রয়ী চল (continuous random variable)

কোন সম্ভাবনাশ্রয়ী চল X যদি কোন নির্দিষ্ট অন্তর (a, b)-এর মধ্যে অবিচ্ছিন্নভাবে যে কোন মান গ্রহণ করতে পারে এবং ঐ অন্তরের অন্তর্ভুক্ত কোন অন্তর (α, β)-এর মধ্যে ঐ চলের মান গ্রহণ করার অর্থাৎ $a \leq \alpha \leq x \leq \beta \leq b$ হওয়ার একটি নির্দিষ্ট সম্ভাবনা পাওয়া যায়, তবে আমরা X-কে অবিচ্ছিন্ন সম্ভাবনাশ্রয়ী চল বলব ।

উদাহরণ :

একটি নির্দিষ্ট অবস্থায় ধান উৎপাদনের পরিমাণ, কোন বৈদ্যুতিক বাতির কার্যকাল ইত্যাদির ক্ষেত্রে আমরা মানের অবিচ্ছিন্ন অন্তরগুচ্ছ পেতে পারি এবং সেজন্য এগুলির সাথে সংযুক্ত সম্ভাবনাশ্রয়ী চলদের আমরা অবিচ্ছিন্ন সম্ভাবন্যশ্রয়ী চল বলব ।

৪২.৫.২ সম্ভাবনা ঘনত্ব অপেক্ষক (probability density function)

যে কোন অবিচ্ছিন্ন সম্ভাবনাশ্রয়ী চল X-এর কোন মান x পাওয়ার সম্ভাবনাকে $f(x)$-এর দ্বারা চিহ্নিত করলে $f(x)$-কে আমরা X-এর সঙ্গে সংশ্লিষ্ট সম্ভাবনা ঘনত্ব অপেক্ষক বলি (probability density function, সংক্ষেপে pdf) । $f(x)$ একটি সম্ভাবনা ঘনত্ব অপেক্ষক হবে যদি

৩. সমস্ত x-এর ক্ষেত্রে $f(x) \geq 0$ হয়, এবং
৪. $\int_{-\infty}^{\infty} f(x)\, dx = 1$ ।

অবিচ্ছিন্ন সম্ভাবনাশ্রয়ী চলের ক্ষেত্রে x-এর কোন নির্দিষ্ট মানের জন্য সম্ভাবনার মান পাওয়া যাবে না । এটি সহজেই প্রমাণ করা যায় । $P(X = c) = P(c \leq x \leq c) = \int_c^c f(x)\, dx = 0$, যেখানে X কোন অবিচ্ছিন্ন সম্ভাবনাশ্রয়ী চল ও c হল একটি ধ্রুবক । বরং আমরা x-এর মান কোন নির্দিষ্ট অন্তরের মধ্যে থাকার সম্ভাবনা পেতে পারি ।

এখানে $P[a \leq x \leq b] = \int_a^b f(x)\, dx$ ।

যদি x-এর সম্ভাব্য মানগুলির অন্তর (α, β) হয়, (যেখানে (α, β), (a, b) অন্তরের অন্তর্ভুক্ত), তা হলে f(x)-কে $\int^{\beta}_{\alpha} f(x)\, dx = 1$ এই শর্তটি পালন করতে হবে।

উদাহরণ :

ধরা যাক, একটি অপেক্ষক হল, $f(x) = x^2/9$ যেখানে $0 < x \leq 3$

$= 0$, অন্যত্র।

প্রদত্ত অপেক্ষকটি একটি অবিচ্ছিন্ন চলের সম্ভাবনা ঘনত্ব অপেক্ষক কারণ, সমস্ত x-এর ক্ষেত্রে $f(x) \geq 0$ হয়, এবং $\int^{\infty}_{-\infty} f(x)\, dx = 0 + \int^{3}_{0} x^2/9dx = (x^3/27)^3_0 = 1$।

আবার ধরা যাক, আর একটি অপেক্ষক হল

$f(x) = 3x$ যেখানে $0 < x \leq 2$

$f(x) = 3-x$ যেখানে $2 < x \leq 3$

$= 0$, অন্যত্র।

প্রদত্ত অপেক্ষকটি কোন অবিচ্ছিন্ন চলের সম্ভাবনা ঘনত্ব অপেক্ষক নয় কারণ, সমস্ত x-এর ক্ষেত্রে $f(x) \geq 0$ হলেও $\int^{\infty}_{-\infty} f(x)\, dx = 0 + \int^{2}_{0} 3xdx + \int^{3}_{2} (3-x)dx = (3x^2/2)^2_0 + (3x-x^2/2)^3_2 \neq 1$।

৪২.৫.৩ সম্ভাবনা নিবেশন অপেক্ষক (distribution function)

যে কোন অবিচ্ছিন্ন সম্ভাবনাশ্রয়ী চল X-এর জন্য ক্রমযৌগিক (cumulative) নিবেশন অপেক্ষক (cumulative distribution function) বা সংক্ষেপে নিবেশন অপেক্ষক F(x) হল

$F(x) = \int^{\infty}_{-\infty} f(y)\, dy$, যদি Y অবিচ্ছিন্ন চল হয়। এখানে f(y) হল Y সম্ভাবনাশ্রয়ী চলের সম্ভাবনা ঘনত্ব অপেক্ষক এবং x হল একটি সুনির্দিষ্ট মান।

৪২.৬ অবিচ্ছিন্ন সম্ভাবনাশ্রয়ী চলের নিবেশন অপেক্ষকের প্রকৃতি

৪২.৬.১ অবিচ্ছিন্ন চলের প্রত্যাশিত মান, ভেদমান এবং উচ্চতর পরিঘাত

x নামক অবিচ্ছিন্ন চলের ক্ষেত্রে প্রত্যাশিত মান হল $E(x) = \int^{\beta}_{\alpha} xf(x)\, dx$ যেখানে α ও β হল চলের সম্ভাব্য সমস্ত মানের দুই প্রান্তসীমা এবং f(x) হল চলের সম্ভাবনা ঘনত্ব অপেক্ষক। এই মান অর্থবহ হবে যদি $\int^{\beta}_{\alpha} |x| f(x)\, dx < \infty$।

অবিচ্ছিন্ন চলের ক্ষেত্রে ভেদমান হল $\mu_2 = \int^{\beta}_{\alpha} (x - \mu)^2 f(x)\, dx$। এখানে $\mu = E(x) = \int^{\beta}_{\alpha} xf(x)dx$

অবিচ্ছিন্ন চলের r-তম মূলকেন্দ্রিক পরিঘাত হবে $\mu'_r = \int^{\beta}_{\alpha} x^r f(x)dx$ এবং r-তম গড়কেন্দ্রিক পরিঘাত হবে $\mu_r = \int^{\beta}_{\alpha} (x - \mu)^2 f(x)\, dx$।

এই নিবেশনের পরিঘাত নির্ণয়ের ক্ষেত্রে আমরা পাই,

$\mu = \mu'_1 = \int^{\beta}_{\alpha} xf(x)dx = 1/(\beta - \alpha) \int^{\beta}_{\alpha} xdx = 1/(\beta - \alpha)[x^2/2]^{\beta}_{\alpha} = (\beta^2 - \alpha^2)/2(\beta - \alpha) = (\beta + \alpha)/2$।

সুতরাং, $\mu_r = \int^{\beta}_{\alpha} [x - (\beta+\alpha)/2]^r f(x)\,dx = 1/(\beta-\alpha) \int^{\beta}_{\alpha} [x - (\beta+\alpha)/2]^r\,dx$

$$= \{1/(\beta-\alpha)\}[[x-(\beta+\alpha)/2]^{r+1}/(r+1)]^{\beta}_{\alpha}$$

$$= \{1/(\beta-\alpha)(r+1)\}[(\beta-\alpha)^{r+1} - \{-(\beta-\alpha)\}^{r+1}]/2^{r+1}$$

$$= \frac{(\beta-\alpha)^{r+1}}{(r+1)} \quad \frac{1+(-1)^{r+1}}{(\beta-\alpha).2^{r+1}} = \frac{1+(-1)^{r+1}}{2(r+1)} \left(\frac{\beta-\alpha}{2}\right)^r$$

এখান থেকে পাওয়া যায় যে, r-এর মান বিজোড় হলে সমস্ত গড়কেন্দ্রিক পরিঘাত শূন্য হবে ও জোড় হলে এই পরিঘাতের মান হবে $(\beta-\alpha)^r/2^2(r+1)$ । সুতরাং, $\mu_2 = (\beta-\alpha)^2/2^2(2+1) = (\beta-\alpha)^2/12$, $\mu_3 = 0$, $\mu_4 = (\beta-\alpha)^4/2^4(4+1) = (\beta-\alpha)^4/80$ । ফলে, $\sqrt{\beta_1} = 0$, $\beta_2 = 12^2/80 = 1.8$, বা $\gamma_1 = 0$ এবং $\gamma_2 = -1.2$ । অতএব, সমনিবেশন খুবই অবতীক্ষ্ণ।

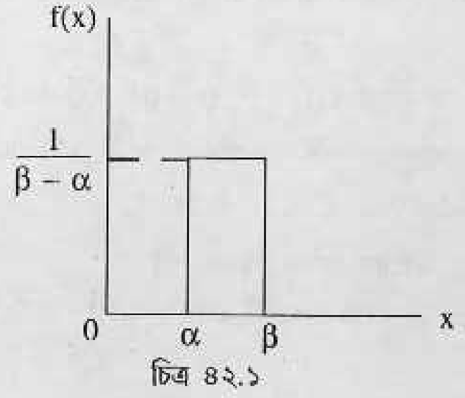
৪২.৭ অবিচ্ছিন্ন সম্ভাবনাশ্রয়ী চলের নিবেশন

৪২.৭.১ আয়তাকার বা সম নিবেশন

অবিচ্ছিন্ন ধরনের একচলক নিবেশনের সরলতম রূপ হল আয়তাকার (rectangular) বা সম (uniform) নিবেশন ।এটির সম্ভাবনা ঘনত্ব অপেক্ষকের রূপ হল $f(x) = 1/(\beta-\alpha)$, যেখানে $\alpha \le x \le \beta$ এবং $= 0$, অন্যত্র ।

যেহেতু $\alpha < \beta$, সেজন্য x-এর যে কোন মানের জন্য $f(x) \ge 0$ এবং $\int^{\beta}_{\alpha} f(x)\,dx = 1/(\beta-\alpha)\{\int^{\beta}_{\alpha} dx\} = 1/(\beta-\alpha)[x]^{\beta}_{\alpha} = (\beta-\alpha)/(\beta-\alpha) = 1$ ।

সুতরাং, এই সম্ভাবনা নিবেশন অপেক্ষকের সাহায্যে, আমরা যে কোন অবিচ্ছিন্ন চল X এর মান, (a, b) অন্তরের অন্তর্গত কোন বিশেষ অন্তর (α, β) এর মধ্যে থাকার সম্ভাবনা নির্ণয় করতে পারি । এখানে, $P[a \le X \le b] = \int^{b}_{a} f(x)\,dx = 1/(\beta-\alpha) \int^{b}_{a} dx = (b-a)/(\beta-\alpha)$, যেখানে $\alpha \le a \le b \le \beta$ এবং এই সম্ভাবনার মান কেবলমাত্র (a, b) অন্তরের দৈর্ঘ্যের ওপর নির্ভরশীল । এই নিবেশনকে আয়তাকার নিবেশন বলা হয়, কারণ f(x)-এর রেখাচিত্র অঙ্কন করলে একটি আয়তক্ষেত্র পাওয়া যায় । রেখাচিত্র ৪২.১ এই ক্ষেত্রে দ্রষ্টব্য । এখানে x অক্ষে α ও β বিন্দু দুটির পরিপ্রেক্ষিতে f(x) অক্ষ অনুযায়ী প্রাপ্ত $f(\alpha)$ ও $f(\beta)$ বিন্দুগুলিকে চারটি সরলরেখার সাহায্যে যোগ করে এই আয়তচিত্রটি গঠিত হয়েছে ।এই নিবেশনটিকে সমনিবেশন বলেও অভিহিত করা হয়, কারণ α ও β-এর মধ্যবর্তী সমদৈর্ঘ্যের যে কোন অন্তরের মধ্যে X-এর মান পাওয়ার সম্ভাবনার মান সর্বদা সমান থাকে ।

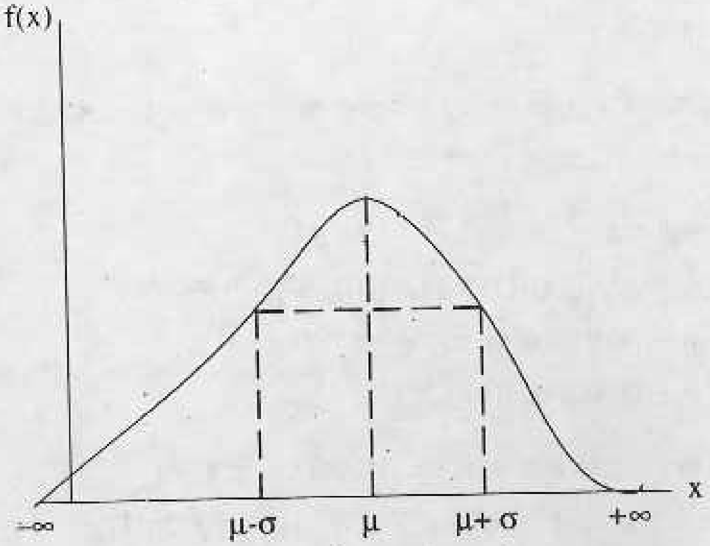


চিত্র ৪২.১

৪২.৭.২ সুষম নিবেশনের ধারণা

অবিচ্ছিন্ন ধরনের একচলক নিবেশনের সবচেয়ে গুরুত্বপূর্ণ ও বহু ব্যবহৃত রূপ হল সুষম নিবেশন। এই নিবেশনকে Gauss নিবেশনও বলা হয়। প্রচলিতভাবে, এই নিবেশনকে Normal বা স্বাভাবিক নিবেশন বলা হয় কারণ, অষ্টাদশ ও ঊনবিংশ শতাব্দীতে সব অবিচ্ছিন্ন চলের নিয়ন্ত্রণকারী সাধারণ নিয়ম হিসাবে এই নিবেশনকে প্রতিষ্ঠিত করার প্রচেষ্টা করা হয়েছিল এবং তখনই এই নিবেশনকে Normal নামে অভিহিত করা হয়।

সুষম নিবেশনের সম্ভাবনা ঘনত্ব অপেক্ষকের রূপ হল $f(x) = [1/(\sigma\sqrt{(2\pi)})]e^{-(x-\mu)^2/2\sigma^2}$, যেখানে $-\infty < x < \infty$। এখানে e ও π দুটি জ্ঞাত ধ্রুবক হওয়ায় এই নিবেশনের পূর্ণকাঙ্কের সংখ্যা দুটি, অর্থাৎ এই নিবেশনের μ ও σ। $y = f(x)$ রেখাটিকে আমরা সুষম রেখা বলি। x-এর বিভিন্ন মানের ক্ষেত্রে f(x)-এর মান লেখচিত্রে স্থাপন করে এই রেখা গঠন করা যায়। রেখাচিত্র ৪২.২ এই ক্ষেত্রে দ্রষ্টব্য।



চিত্র ৪২.২

x-এর সব মানের ক্ষেত্রে f(x) ধনাত্মক। তাছাড়া,

$$\int_{-\infty}^{\infty} f(x)\,dx = [1/(\sigma\sqrt{(2\pi)})]\int_{-\infty}^{\infty} e^{-(x-\mu)^2/2\sigma^2}\,dx$$

$$= [2/(\sigma\sqrt{(2\pi)})]\int_{\mu}^{\infty} e^{-(x-\mu)^2/2\sigma^2}\,dx\text{ ।}$$

$$= [2\sigma\sqrt{2}/(\sigma\sqrt{2\pi})]\int_{0}^{\infty} e^{-z^2}\,dz,\text{ যেখানে } z = (x-\mu)/\sigma\sqrt{2} \text{ ও } dz = dx/\sigma\sqrt{2}\text{ ।}$$

$$= [(2/\sqrt{\pi})\cdot(\sqrt{\pi}/2) = 1 \text{ (কারণ } \int_{0}^{\infty} e^{-z^2}\,dz = \sqrt{\pi}/2)$$

৪২.৭.৩ সুষম নিবেশনের প্রতিসাম্যের বৈশিষ্ট্য

সুষম নিবেশনের বৈশিষ্ট্যের মধ্যে উল্লেখযোগ্য হল এই যে, নিবেশনটি μ-এর দুপাশে প্রতিসম। কারণ, μ-এর থেকে সমদূরবর্তী যে কোন মান a-এর ক্ষেত্রে দেখা যায় যে,

$$f(\mu + a) = f(\mu - a) = [1/(\sigma\sqrt{(2\pi)})]\ e^{-a^2/2\sigma^2}$$

সুতরাং, সুষম রেখা $x = \mu$ কোটির দুপাশে প্রতিসম। এই রেখার আকৃতি অনেকটা ঘণ্টার মতো। $x = \mu$ হলে $f(x)$-এর মান সর্বাধিক $1(\sigma\sqrt{(2\pi)}$ হয়, ও x-এর পরম মান বৃদ্ধির সাথে সাথে $f(x)$-এর মান দ্রুত কমতে থাকে এবং সুষম রেখা μ-এর দুপাশে $+\infty$ ও $-\infty$ অবধি বিস্তৃত হয়। রেখাচিত্র ৪২.২ এই ক্ষেত্রে দ্রষ্টব্য।

৪২.৭.৪ সুষম নিবেশনের ধ্রুবকগুলি

সুষম নিবেশনের প্রতিসাম্যের বৈশিষ্ট্যের থেকেই দেখা যায় যে, এর গড়মান, মধ্যমা ও সংখ্যাগুরু মান, তিনটিই এক ও μ-এর সমান।

প্রমাণ : আমরা আগেই পেয়েছি যে, সুষম নিবেশনের সম্ভাবনা ঘনত্ব অপেক্ষকের রূপ হল

$$f(x) = [1/(\sigma\sqrt{(2\pi)})]\ e^{-(x-\mu)^2/2\sigma^2}\ ।$$

এখান থেকে পাওয়া যায়,

$$f'(x) = [1/(\sigma\sqrt{(2\pi)})]\ [\{-2\,(x-\mu)\}/2\sigma^2]\ e^{-(x-\mu)^2/2\sigma^2}\ ।$$
$$= -[(x-\mu)]/[\sigma^3\sqrt{(2\pi)}]\ e^{-(x-\mu)^2/2\sigma^2}\ ।$$
$$= -[(x-\mu)]/\sigma^2\ f(x)\ ।$$

সুতরাং, যেহেতু σ^2 ও $f(x) \neq 0$, $f'(x) = 0$ হলে, $x = \mu$ হবে।

আবার, $f''(x) = [(x-\mu)^2]/[\sigma^5\sqrt{(2\pi)}]\ e^{-(x-\mu)^2/2\sigma^2} - [1/\{\sigma^3\sqrt{(2\pi)}\}]\ e^{-(x-\mu)^2/2\sigma^2}$

$$= [(x-\mu)^2 - \sigma^2]/[\sigma^5\sqrt{(2\pi)}]\ e^{-(x-\mu)^2/2\sigma^2} - [1/\{\sigma^3\sqrt{(2\pi)}\}]\ e^{-(x-\mu)^2/2\sigma^2}$$
$$= [(x-\mu)^2 - \sigma^2]/[\sigma^5\sqrt{(2\pi)}]\ e^{-(x-\mu)^2/2\sigma^2} - [1/\{\sigma^3\sqrt{(2\pi)}\}]\ e^{-(x-\mu)^2/2\sigma^2}$$
$$= [(x-\mu)^2 - \sigma^2]/[\sigma^4]\ f(x)$$
$$= [f(x)/\sigma^2][\{(x-\mu)^2/\sigma^2\} - 1]$$

সুতরাং, $x = \mu$ হলে $f''(x) = -[f(x)]/\sigma^2 < 0$ অর্থাৎ ঋণাত্মক হবে।

সুষম নিবেশনের ক্ষেত্রে যেহেতু $x = \mu$ বিন্দুতে সুষম রেখার আয়তন সমান দুভাগে বিভক্ত হয়, সেজন্য মধ্যমার মান হল μ। যদি মনে করা হয় মধ্যমার মান হল a, তবে আমরা পাই,

$\int^{a}_{-\infty} f(x)\, dx = \int^{\infty}_{a} f(x)\, dx = {}^1/_2$

অর্থাৎ, $\int^{\mu}_{-\infty} f(x)\, dx + \int^{a}_{\mu} f(x)\, dx = {}^1/_2$

কিন্তু $\int^{\mu}_{-\infty} f(x)\, dx = [1/(\sigma\sqrt{(2\pi)}]\int^{\mu}_{-\infty} e^{-(x-\mu)^2/2\sigma^2}\, dx$

$= [(\sigma\sqrt{2})/\{a\sqrt{(2\pi)}\}]\int^{0}_{-\infty} e^{-z^2}\, dz$, যেখানে $z = (x-\mu)/\sigma\sqrt{2}$ ও $dz = dx/\sigma\sqrt{2}$।

$= (1/\sqrt{\pi})\}]\int^{0}_{-\infty} e^{-t^2}\, dt$, যেখানে $t = -z$ ।

$= (1/\sqrt{\pi})\,(\sqrt{\pi}/2) = {}^1/_2$ ।

অতএব, $\int^{a}_{\mu} f(x)\, dx = 0$ এবং $a = \mu$ ।

আমরা এর আগে দেখেছি যে, $f''(x) = [f(x)/\sigma^2][\{(x-\mu)^2/\sigma^2\} - 1]$ । সুতরাং, $f''(x) = 0$ হলে আমরা পাই, $(x-\mu)^2/\sigma^2 - 1 = 0$ বা, $(x-\mu)^2 = \sigma^2$ ।

অর্থাৎ, $x = \mu \pm \sigma$ বিন্দু দুটি হল সুষম রেখার দুটি নতি বিন্দু (points of inflexion) । কারণ, সুষম রেখা $x = \mu + \sigma$ বিন্দুতে অনুভূমিক অক্ষের দিকে অবতল থেকে উত্তল হয় ও $x = \mu - \sigma$ বিন্দুতে উত্তল থেকে অবতল হয় । অন্যভাবে বলা যায় যে, $x = \mu \pm \sigma$ বিন্দু দুটির মধ্যবর্তী অংশে সুষম রেখার আকৃতি নিচের দিকে অবতল থাকে ও এই অন্তরের বাইরে নিচের দিকে উত্তল হয় ।

সুষম নিবেশনের প্রতিসাম্যের বৈশিষ্ট্যের থেকে আরো দেখা যায় যে, এই নিবেশনের সমস্ত গড়কেন্দ্রিক বিজোড় সংখ্যক পরিঘাতের মান শূন্য । কারণ,

$\mu_{2r+1} = \int^{\infty}_{-\infty} (x-\mu)^{2r+1} f(x)\, dx = [1/(\sigma\sqrt{2\pi})]\int^{\infty}_{-\infty} (x-\mu)^{2r+1} e^{-(x-\mu)^2/2\sigma^2}\, dx$

$= [1/\{\sigma\sqrt{(2\pi)}\}]\,[\int^{\mu}_{-\infty} (x-\mu)^{2r+1} e^{-(x-\mu)^2/2\sigma^2}\, dx + \int^{\infty}_{\mu} (x-\mu)^{2r+1} e^{-(x-\mu)^2/2\sigma^2}\, dx]$

$= [1/(\sigma\sqrt{(2\pi)})]\,[\int^{0}_{-\infty} (z)^{2r+1} e^{-(z)^2/2\sigma^2}\, dz + \int^{\infty}_{0} (z)^{2r+1} e^{-(z)^2/2\sigma^2}\, dz]$, যেখানে $z = (x-\mu)$ ।

$= [1/(\sigma\sqrt{(2\pi)})]\,[(-1)^{2r+1}\int^{\infty}_{0} (t)^{2r+1} e^{-(t)^2/2\sigma^2}\, dt + \int^{\infty}_{0} (t)^{2r+1} e^{-(t)^2/2\sigma^2}\, dt] = 0$ (প্রথম সমাকলকে $t = -z$ ও দ্বিতীয় সমাকলকে এখানে $t = +z$ বসিয়ে ।

তবে জোড় সংখ্যক গড়কেন্দ্রিক পরিঘাতের ক্ষেত্রে পাওয়া যায়

$$\mu_{2r} = \int^{\infty}_{-\infty} (x-\mu)^{2r} f(x)\, dx = [1/(\sigma\sqrt{2\pi})]\int^{\infty}_{-\infty} (x-\mu)^{2r} e^{-(x-\mu)^2/2\sigma^2}\, dx\ ।$$

এবং সাধারণভাবে পাওয়া যায়, $\mu_{2r} = (2r-1)\sigma^2\mu_{2r-2}$ ।

এই সূত্রকে বারবার প্রয়োগ করে পাওয়া যায়, $\mu_{2r} = (2r-1)(2r-3)\ldots3.1\sigma^{2r}$ ।

সুতরাং, আমরা পাই, $\mu_4 = 3\sigma^4$ । অতএব, $\sqrt{\beta_1} = 0,\ \beta_2 = 3$ ।

এজন্য সুষম নিবেশনের ক্ষেত্রে প্রতিবৈষম্য শূন্য ও এই নিবেশন সমতীক্ষ্ণ। সুষম নিবেশনকে আদর্শ ধরে নিয়ে অন্য নিবেশনের প্রতিবৈষম্য ও তীক্ষ্ণতার পরিমাপ করা হয় ।

৪২.৭.৫ সমক মাত্রিক সুষম চল

সুষম নিবেশনের ক্ষেত্রে চলের গড় μ ও সমক বিচ্যুতি σ হলে আমরা একটি সমক মাত্রিক চল z নির্মাণ করতে পারি যেখানে $z = (x - \mu)/\sigma$ । এক্ষেত্রে z চলের গড় ও সমক বিচ্যুতি যথাক্রমে 0 ও 1 । সাধারণভাবে, x চলের নিবেশনকে $N(\mu,\sigma^2)$ এবং z চলের নিবেশনকে $N(0,1)$ প্রতীকের সাহায্যে চিহ্নিত করা হয় । z চলটিকে সমক মাত্রিক সুষম চল (standard/standardised normal variable/deviate) বলা হয় । এই চলের সম্ভাবনা ঘনত্ব অপেক্ষকের রূপ হল, $\phi(z) = 1/\sqrt{(2\pi)}\ e^{-(z)^2/2}$ ।

৪২.৮ সারাংশ

সম্ভাবনাশ্রয়ী চলকে দুভাগে বিভক্ত করা যায়—বিচ্ছিন্ন ও অবিচ্ছিন্ন সম্ভাবনাশ্রয়ী চল । যে কোন বিচ্ছিন্ন সম্ভাবনাশ্রয়ী চল X-এর কোন মান x পাওয়ার সম্ভাবনাকে $f(x)$-এর দ্বারা চিহ্নিত করলে $f(x)$-কে আমরা X-এর সঙ্গে সংশ্লিষ্ট সম্ভাবনা ভর অপেক্ষক বলি । চলটি অবিচ্ছিন্ন হলে তার সাথে সংশ্লিষ্ট সম্ভাবনা অপেক্ষককে, সম্ভাবনা ঘনত্ব অপেক্ষক বলি । যে কোন সম্ভাবনাশ্রয়ী চল X-এর জন্য আমরা একটি ক্রমযৌগিক নিবেশন অপেক্ষক বা সংক্ষেপে নিবেশন অপেক্ষকের সংজ্ঞা দিতে পারি । এই অপেক্ষককে আমরা $F(x)$ দিয়ে সূচিত করি । অন্যদিকে, নিবেশন অপেক্ষক থেকেও আমরা সম্ভাবনা অপেক্ষক পেতে পারি । চলটি বিচ্ছিন্ন হোক, বা অবিচ্ছিন্ন, তার নিবেশনের সাথে সংশ্লিষ্ট কিছু ধ্রুবক পাওয়া যায় । এগুলি হল গড়মান, ভেদমান, ভূয়িষ্ঠ ও পরিঘাত । এই এককে আমরা একচলক তত্ত্বগত নিবেশনের কয়েকটি রূপ আলোচনা করেছি । এগুলি হল বিচ্ছিন্ন চলের নিবেশনের দুটি রূপ (দ্বিপদ এবং পোয়াসঁ) ও অবিচ্ছিন্ন চলের ক্ষেত্রে আয়তাকার নিবেশন ও সুষম নিবেশন ।

৪২.৯ অনুশীলনী

১. নিম্নলিখিত অপেক্ষকগুলি সম্ভাবনা ঘনত্ব অপেক্ষক কি না, বিচার করুন ।

ক. $f(x) = 2x$ যেখানে $0 < x \le 1$
$= 4 - 2x$ যেখানে $1 < x \le 2$
$= 0$ অন্যত্র ।

খ. $f(x) = x/2$ যেখানে $0 < x \le 1$
$= 1/2$ যেখানে $1 < x \le 2$
$= (3 - x)/2$ যেখানে $2 < x \le 3$ ।
$= 0$ অন্যত্র ।

২. একটি দ্বিপদ নিবেশনের গড়মান ও ভেদমান যথাক্রমে 3 ও 2 । চলটি নিম্নলিখিত মানগুলি নেবে, এর সম্ভাবনা নির্ণয় করুন : (ক) ≤ 2 (খ) ≥ 7 ।

৩. কোন দ্বিপদ নিবেশনের সম্পর্কে নিম্নলিখিত বক্তব্যটির বিষয়ে মতামত প্রকাশ করুন :
গড়মান = 7 ও ভেদমান = 11 ।

৪. যদি কোন কোম্পানীর দ্বারা নির্মিত বৈদ্যুতিক বাতির 5% ত্রুটিযুক্ত হয় পোয়াসঁ নিবেশনের মাধ্যমে নিম্নলিখিত ঘটনাগুলির সম্ভাবনা নিরূপণ করুন । একটি নমুনায় 100টি বাতি থাকলে (ক) কোন বাতি ত্রুটিযুক্ত নয় ; (খ) 5টি বাতি ত্রুটিযুক্ত (দেওয়া আছে যে, $e^{-5} = 0.007$) ।

৫. একটি শ্রেণীর কোন একটি বিশেষ বিষয়ের পরীক্ষার গড় নম্বর হল 79 । নম্বরের সমক বিচ্যুতি হল 5 । যদি নম্বর সুষমভাবে বণ্টিত হয়, তবে 200 ছাত্রবিশিষ্ট একটি শ্রেণীর কতজন ছাত্র 75 থেকে 82 এর মধ্যে নম্বর পায়নি? দেওয়া আছে : $P[0 \le Z \le 0.7] = 0.2580$; $P[0 \le Z \le 0.8] = 0.2881$; $P[0 \le Z \le 0.6] = 0.2257$, যেখানে Z একটি সমক মাত্রিক সুষম চল ।

৪২.১০ উত্তরমালা

১ক. প্রদত্ত অপেক্ষকটি কোন অবিচ্ছিন্ন চলের সম্ভাবনা ঘনত্ব অপেক্ষক নয় কারণ, সমস্ত x-এর ক্ষেত্রে $f(x) \ge 0$ হলেও $\int_{-\infty}^{\infty} f(x)\,dx = 0 + \int_0^1 2x dx + \int_1^2 (4-2x)dx = 0 + (x^2)_0^1 + \int_1^2 (4x-x^2)_1^2\, dx \neq 1$ ।

খ. প্রদত্ত অপেক্ষকটি একটি অবিচ্ছিন্ন চলের সম্ভাবনা ঘনত্ব অপেক্ষক কারণ, সমস্ত x-এর ক্ষেত্রে $f(x) \ge 0$ হয়, এবং $\int_{-\infty}^{\infty} f(x)\,dx = 0 + \int_0^1 x/2 dx + \int_1^2 1/2 dx + \int_2^3 (3-x)/2\, dx = 0 + (x^2/4)_0^1 + (x/2)_1^2 + (3x/2-x^2/4)_2^3 = 0 + (1/4) + (1/2) + (1/4) = 1$

২. যদি n ও p দ্বিপদ নিবেশনটির ধ্রুবক হয়, তাহলে আমাদের আছে
গড়মান = np = 3, ও ভেদমান = npq = 2, যেখানে $q = 1 - p$ ।
$(npq)/np = 2/3) \Rightarrow q = 2/3 \Rightarrow p = 1/3$
এখন, $np = 3, \Rightarrow n = 9$ ।
ধরা যাক, $X \sim B(n,p)$, যেখানে $n = 9$, $p = 1/3$ ।
অতএব, $P(X=r) = {}^nC_r\, p^r q^{n-r}$ যেখানে r-এর মান 0, 1, 2, ..., n ।
$= {}^9C_r\,(1/3)^r (2/3)^{9-r}$ যেখানে r = 0, 1, 2, ..., 9 ।

ক. $P(X \le 2) = P(X = 0) + P(X = 1) + P(X = 2)$

$$= {}^9C_0 (1/3)^0 (2/3)^{9-0} + {}^9C_1 (1/3)^1 (2/3)^{9-1} + {}^9C_2 (1/3)^2 (2/3)^{9-2}$$
$$= (2/3)^9 + 9(1/3)(2/3)^8 + {}^9C_2 (1/3)^2 (2/3)^7$$
$$= (2/3)^7 [(2/3)^2 + 9(1/3)(2/3) + 36(1/3)^2]$$
$$= (2/3)^7 [(4/9) + 9(2/9) + 36(1/9)]$$
$$= (2/3)^7 [(4/9) + 2 + 4]$$
$$= (2/3)^7 [58/9] = (0.0585)(6.44) = 0.3767 ।$$

খ. $P(X \geq 7) = P(X = 7) + P(X = 8) + P(X = 9)$

$$= {}^9C_7 (1/3)^7 (2/3)^{9-7} + {}^9C_8 (1/3)^8 (2/3)^{9-8} + {}^9C_9 (1/3)^9 (2/3)^{9-9}$$
$$= {}^9C_7 (1/3)^7 (2/3)^2 + {}^9C_8 (1/3)^8 (2/3) + {}^9C_9 (1/3)^9 (2/3)^0$$
$$= (1/3)^7 [36(2/3)^2 + 9(1/3)(2/3) + (1/3)^2]$$
$$= (1/3)^7 [16 + 2 + 1/9]$$
$$= (1/3)^7 [18 + 1/9]$$
$$= (1/2187) [18 + 0.1111]$$
$$= 18.1111/2187 = 0.0083।$$

৩. n ও p ধ্রুবকবিশিষ্ট কোন দ্বিপদ নিবেশনের জন্য গড়মান = $np = 7$, ও ভেদমান = $npq = 11$; তাহলে $(npq)/np = 11/7 \Rightarrow q = 11/7 = 1.6$ ।

এটি অসম্ভব, কারণ যেহেতু q একটি সম্ভাবনা বোঝায়, সেজন্য q-কে 0 ও 1-এর মধ্যে থাকতে হবে । অতএব, বক্তব্যটি ভ্রান্ত ।

৪. আমাদের দেওয়া আছে যে, $n = 100$, ও $p = P$(বাতিগুলি ত্রুটিযুক্ত) $= 5\% = 0.05$ । যেহেতু p অত্যন্ত ছোট এবং n বড়, সেজন্য আমরা এই নিবেশনটিকে পোয়াসঁ নিবেশনের দ্বারা প্রকাশ করতে পারি । অতএব, পোয়াসঁ নিবেশনের ধ্রুবক λ-এর মান হল, $\lambda = np = 5$ ।

ধরা যাক, X হল একটি চল যা 100টি বাতির মধ্যে ত্রুটিযুক্ত বাতির সংখ্যা নির্দেশ করে । তাহলে, পোয়াসঁ নিবেশনের সূত্র ধরে, $P(X = r) = (\lambda)^r e^{-\lambda}/(r)! = (5)^r e^{-5}/(r)!$; $r = 0, 1, 2, \ldots\ldots$ ।

(ক) P(কোন বাতি ত্রুটিযুক্ত নয়) $= P(X = 0) = (5)^0 e^{-5}/(0)! = e^{-5} = 0.007$ ।

(খ) P(5টি বাতি ত্রুটিযুক্ত) $= P(X = 5) = (5)^5 e^{-5}/(5)! = [(0.007)(3125)]/120 = 4.375/24 = 0.1823$ ।

৫. ধরা যাক, X নামক সম্ভাবনাশ্রয়ী চলটি একটি পরীক্ষায় ছাত্রদের প্রাপ্ত নম্বর নির্দেশ করে । তাহলে আমাদের দেওয়া আছে যে, $X \sim N(\mu, \sigma^2)$ যেখানে $\mu = 79$ ও $\sigma = 5$ । $P[75 < X < 82]$ $= P[(75 - 79)/5 < Z < (82 - 79)/5]$ যেহেতু $Z = (X - \mu)/\sigma = (X - 79)/5$

$= P[-0.8 < Z < 0.6] = P[-0.8 < Z < 0] + P[0 < Z < 0.6]$

$= P[0 < Z < 0.8] + P[0 < Z < 0.6] = 0.2881 + 0.2257 = 0.5138$ ।

'ছাত্র 75 থেকে 82 এর মধ্যে নম্বর পায়নি' এই ঘটনাটির সম্ভাবনা, p, হল

$p = 1 - P[75 < X < 82] = 1 - 0.5138 = 0.4862$ ।

অতএব, একটি শ্রেণীতে 200 ছাত্র থাকলে, 200 গুণ p = 200 গুণ 0.4862 = 97.24 ≈ 97 জন ছাত্র 75 থেকে 82 এর মধ্যে নম্বর পায়নি।

৪২.১১ গ্রন্থপঞ্জী

- সেন, রাজকুমার : সংখ্যাতত্ত্ব, পশ্চিমবঙ্গ রাজ্য পুস্তক পর্যদ, (১৯৮৬)।
- Cramer, H : *The Elements of Probability Theory,* John Wiley, (1955)।
- Das, N. G. : *Statistical Methods in Commerce, Accountancy and Economics,* M. Das, (1985)।
- Freund, G. E. : *Modern Elementary Statistics,* Prentice Hall, New Delhi, (1979) ।
- Goon, A. M., Gupta, M. K. & Dasgupta, B. : *Fundamentals of Statistics,* Vols. I, II, World Press, (1986)।
- Kapur. J. N. & Saxena, H. C. : *Mathematical Statistics,* S. Chand, N. Delhi, (1981)।
- Mathai, A. M. & Rathie, P. N. : *Probability and Statistics,* Macmillan, Madras, (1977)।
- Nagar, A. L. & Das, R. K. : *Basic Statistics,* Oxford University Press, Delhi, (1983)।
- Spiegel, M. R. : *Theory and Problems of Statistics,* Schaum Publishing Co, (1982)।
- Yule, G. U. & Kendall, M. G. : *An Introduction to the Theory of Statistics,* Griffin, (1953)।

একক ৪৩ ❑ সম্ভাবনাশ্রয়ী চলের গাণিতিক প্রত্যাশা, পরিঘাত এবং পরিঘাত নির্দেশক অপেক্ষক (Mathematical Expectations, Moments and Moment Generating Function of a Random Variable)

গঠন

৪৩.০ উদ্দেশ্য

পরিসংখ্যা নিবেশনের আলোচনার সময়ে আমরা বিভিন্ন নিবেশনের প্রকৃতি বর্ণনা করার জন্য যেমন গড়মান, ভেদমান, ইত্যাদি কয়েকটি পরিমাপের সহায়তা নিয়ে থাকি, তত্ত্বগত সম্ভাবনা নিবেশনের প্রকৃতি বর্ণনার ক্ষেত্রেও আমরা এই ধরনের বিভিন্ন পরিমাপের বিষয় ধারণা করতে পারি। তাদের মধ্যে কয়েকটি উল্লেখযোগ্য পরিমাপ হল, সম্ভাবনাশ্রয়ী চলের গাণিতিক প্রত্যাশা, পরিঘাত এবং পরিঘাত নির্দেশক অপেক্ষক।

৪৩.১ প্রস্তাবনা

একক ৪০-এ আমরা সম্ভাবনাশ্রয়ী চলের ধারণাটির সাথে পরিচিত হয়েছি। একক ৪২-এ আমরা দেখেছি যে, সম্ভাবনাশ্রয়ী চল দুই রকম হয়—বিচ্ছিন্ন সম্ভাবনাশ্রয়ী চল এবং অবিচ্ছিন্ন সম্ভাবনাশ্রয়ী চল। এই এককে আমরা বিচ্ছিন্ন এবং অবিচ্ছিন্ন সম্ভাবনাশ্রয়ী চলের তত্ত্বগত সম্ভাবনা নিবেশনের বিভিন্ন পরিমাপের বিষয় নিয়ে আলোচনা করব।

৪৩.২ সম্ভাবনাশ্রয়ী চলের প্রত্যাশিত মান বা গাণিতিক প্রত্যাশা

৪৩.২.১ বিচ্ছিন্ন চলের প্রত্যাশিত মান বা গাণিতিক প্রত্যাশা (expected value or mathematical expectation)

সম্ভাবনা তত্ত্বের একটি গুরুত্বপূর্ণ বিষয় হল সম্ভাবনাশ্রয়ী চলের প্রত্যাশিত মান নির্ণয়। তত্ত্বগত নিবেশনের প্রকৃতি বিশ্লেষণের ক্ষেত্রে এটি সরলতম ও সর্বাধিক ব্যবহৃত পরিমাপ। প্রত্যাশিত মান সম্ভাবনা বিভাজনের অবস্থান বা চলের প্রতিনিধিস্থানীয় মান সম্পর্কে আমাদের ধারণা প্রদান করে।

X নামক বিচ্ছিন্ন চলের ক্ষেত্রে এর প্রতীক হল $E(x)$, μ_x বা μ । যেখানে x_i হল x নামক বিচ্ছিন্ন চলের সসীম সংখ্যক বিচ্ছিন্ন মানসমূহ, এবং p_i এই মানসমূহের সংশ্লিষ্ট সম্ভাবনার মান, সেখানে,

$$E(X) = x_1P(X=x_1) + x_2P(X=x_2) + \ldots\ldots\ldots\ldots + x_kP(X=x_k) = \Sigma^k_{i=1}\ x_iP(X=x_i),$$

অথবা, $E(X) = x_1p_1 + x_2p_2 + \ldots\ldots\ldots\ldots + x_kp_k = \Sigma^k_{i=1}\ x_ip_i$, (কারণ $P(X=x_i) = p_i$)

একটু গভীরভাবে দেখলে বোঝা যাবে যে, $E(X)$ হল X চলের সম্ভাবনা নিবেশনের অন্যতম মধ্যগামিতা মাপক (measure of central tendency), কারণ

$$E(X) = [x_1p_1 + x_2p_2 + \ldots\ldots\ldots\ldots + x_kp_k]/[p_1 + p_2 + \ldots\ldots\ldots\ldots + p_k]$$

$$= [\Sigma^k_{i=1}\ x_ip_i]/[\Sigma^k_{i=1}p_i],\ \text{যেখানে}\ [\Sigma^k_{i=1}p_i] = 1\ ।$$

অর্থাৎ, $E(X)$ হল X চলের যৌগিক গড়।

যদি $P(X = x_i) = p_i = f(x_i)$, তবে

$$E(x) = x_1\,f(x_1) + x_2\,f(x_2) + \ldots\ldots\ldots + x_kf(x_k) = \Sigma^k_{i=1}\ x_if(x_i)$$

এখানে যেহেতু আমরা $P(X = x_i) = p_i = f(x_i)$-কে X চলের x_i মানের দীর্ঘকালীন আনুপাতিক পরিসংখ্যা হিসাবে গণ্য করে থাকি, সেজন্য $E(X)$-কে X চলের দীর্ঘকালীন গড় মান বলেও বিবেচনা করা যায়। যখন সব সম্ভাবনার মানই সমান, তখন, $E(X) = (x_1 + x_2 + \ldots\ldots\ldots\ldots + x_k)/k$ ।

X নামক বিচ্ছিন্ন চলের সম্ভাব্য মানের সংখ্যা যেখানে অসীম, সেক্ষেত্রে গাণিতিক প্রত্যাশা নির্ণয়ে প্রাথমিক অসুবিধা দেখা দেয়। X-এর মানগুলিকে তখন আমরা কোন ক্রম অনুযায়ী সাজিয়ে নিই এবং তারপর $[\Sigma_i x_i p_i]$ নির্ণয় করি। $[\Sigma_i x_i p_i] = E(X)$ হবে কেবলমাত্র তখনই যখন এই যোগফল মানগুলির ক্রমবিন্যাসের ওপর নির্ভর করে না, অর্থাৎ ক্রম পরিবর্তিত হলেও যোগফল পরিবর্তিত হয় না বা $\Sigma|x_i|p_i < \infty$। যদি যোগফল পরিবর্তিত হয়, সেক্ষেত্রে আমরা বলি যে প্রত্যাশিত মানের কোনও অস্তিত্ব নেই।

৪৩.২.২ অবিচ্ছিন্ন চলের প্রত্যাশিত মান বা গাণিতিক প্রত্যাশা

X নামক অবিচ্ছিন্ন চলের ক্ষেত্রে প্রত্যাশিত মান হবে $E(x) = \int^{\beta}_{\alpha} xf(x)dx$, যেখানে α ও β হল চলের সম্ভাব্য সমস্ত মানের দুই প্রান্তসীমা এবং $f(x)$ হল সম্ভাবনাশ্রয়ী চলের সম্ভাবনাঘনত্ব অপেক্ষক। মানসীমা অসীম হলে বিচ্ছিন্ন চলের ক্ষেত্রে প্রত্যাশিত মানের মতো এখানে এই মান অর্থবহ হবে যদি $\int^{\beta}_{\alpha} |x| f(x)dx < \infty$ হয়। এখানে α ও β-এর মধ্যবর্তী প্রসারকে kটি সমব্যবধানযুক্ত অংশে (প্রতি অংশের অন্তর ধরা যাক, c) ভাগ করলে আমরা পাই, $x_0 = \alpha$, $x_1 = \alpha + h$, $x_2 = \alpha + 2h$, ..., $x_h = \alpha + kh = \beta$। যে কোন i-তম অন্তরের মধ্যবর্তী কোন মান x_i হলে (যেখানে $i - h/2 \leq x \leq i + h/2$) ও অন্তরের দৈর্ঘ্য h যথেষ্ট ক্ষুদ্র হলে x-এর এই অন্তরে অন্তর্ভুক্ত থাকার সম্ভাবনা হবে $\int^{i+h/2} f(x)dx \approx hf(x_i)$।

প্রত্যাশিত মানের ক্ষেত্রে $h\Sigma_{i=0}^{k-1} x_i f(x_i)$, এই যোগফলটি মোটামুটিভাবে বিচ্ছিন্ন চলের ক্ষেত্রে প্রাপ্ত যোগফল $\Sigma_x xf(x)$-এর সমান হবে। h-এর মান ক্রমশ হ্রাস ও k-এর মান ক্রমশ বৃদ্ধি পেতে থাকলে এই মানটির সীমা হবে $\int^{\beta}_{\alpha} xf(x)dx$, এটিই হবে X চলের প্রত্যাশিত মান বা গাণিতিক প্রত্যাশা।

৪৩.৩ সম্ভাবনাশ্রয়ী চলের ভেদমান ও উচ্চতর পরিঘাত

৪৩.৩.১ বিচ্ছিন্ন চলের ভেদমান বা সমক বিচ্যুতি এবং উচ্চতর পরিঘাত (variance and higher order moments)

সম্ভাবনাশ্রয়ী চলের প্রত্যাশিত মান নির্ণয় করা ছাড়াও আমরা সাধারণ পরিসংখ্যা নিবেশনের মতোই ভেদমান ও উচ্চতর পরিঘাত নির্ণয় করতে পারি। সম্ভাবনাশ্রয়ী চলের প্রত্যাশিত মান আমাদের সম্ভাবনা বিভাজনের অবস্থান বা চলের প্রতিনিধি স্থানীয় মান সম্পর্কে ধারণা দেয়। চলের সম্ভাব্য মানসমূহের বিস্তৃতি ও অন্যান্য বৈশিষ্ট্য বিষয়ে ধারণার জন্য ভেদমান ও উচ্চতর পরিঘাত নির্ণয় করার প্রয়োজন হয়।

x-এর ভেদমান হল

$V(x) = E[x - E(x)]^2 = \Sigma_x (x-\mu)^2 f(x)$। V(x)-কে আমরা $\sigma_x^{\ 2}$ বা σ^2 দিয়েও বুঝিয়ে থাকি।

$x - E(x)$, এই অন্তরটি X-এর প্রত্যাশিত মান থেকে চলের পার্থক্য বা বিচ্যুতি বোঝায়। বিচ্যুতির পরিমাপের বিষয়টিই আমাদের কাছে গুরুত্বপূর্ণ হওয়ায় প্রথমে আমরা বিচ্যুতির বর্গ করে পরে তার প্রত্যাশিত মানের ধনাত্মক বর্গমূল নির্ণয় করে চলের বিস্তৃতির পরিমাপ বা সমক বিচ্যুতি (σ_x বা σ) নির্ণয় করি, অর্থাৎ, $\sigma = \sqrt{V(x)} = \sqrt{E[x - E(x)]^2} = \sqrt{\mu_2}$।

সাধারণভাবে, সম্ভাবনাশ্রয়ী চল X-এর r-তম মূলকেন্দ্রিক পরিঘাত হবে $\mu'_r = E[x^r] = \sum_x x^r f(x)$ । X-এর প্রত্যাশিত মান হল X-এর প্রথম মূলকেন্দ্রিক পরিঘাত $\mu'_1 = E[x] = \sum_x x\ f(x)$ ।

সম্ভাবনাশ্রয়ী চল X-এর r-তম গড়কেন্দ্রিক পরিঘাত হবে $\mu_r = E[x-\mu]^r = \sum_x (x-\mu)^r f(x)$ । X-এর ভেদমান হল X-এর দ্বিতীয় গড়কেন্দ্রিক পরিঘাত $\mu_2 = E[x-\mu]^2 = \sum_x (x-\mu)^2 f(x)$

চলের একক অনুযায়ী μ'_1 এবং σ-এর একক নির্ণয় করা হবে।

বিচ্ছিন্ন সম্ভাবনাশ্রয়ী চল X-এর সম্ভাবনা নিবেশনের প্রতিবৈষম্যকে (skewness) আমরা $\delta_1 = \mu_3/\sigma^3$ এর দ্বারা পরিমাপ করি । তেমনি, তীক্ষ্ণতা (kurtosis) পরিমাপ করি $\delta_2 = (\mu_4/\sigma^4)-3$ এর দ্বারা।

৪৩.৩.২ অবিচ্ছিন্ন চলের ভেদমান বা সমক বিচ্যুতি এবং উচ্চতর পরিঘাত

সাধারণভাবে বলা যায়, r-তম মূলকেন্দ্রিক পরিঘাত হবে $\mu'_r = E[x^r] = \int^\beta_\alpha x^r f(x)dx$ এবং r-তম গড়কেন্দ্রিক পরিঘাত হবে $\mu_r = E[x-\mu]^r = \int^\beta_\alpha (x-\mu)^r f(x)dx$ । সুতরাং, ভেদমান বা μ_r হবে $E(x-\mu)^2 = E[x - E(x)]^2 = \int^\beta_\alpha (x-\mu)^2 f(x)dx$ ।

অবিচ্ছিন্ন সম্ভাবনাশ্রয়ী চল X-এর সম্ভাবনা নিবেশনের প্রতিবৈষম্যকে (skewness) আমরা $\delta_1 = \mu_3/\sigma^3$-এর দ্বারা পরিমাপ করি। তেমনি, তীক্ষ্ণতা (kurtosis) পরিমাপ করি $\delta_2 = (\mu_4/\sigma^4)-3$-এর দ্বারা।

৪৩.৪ সম্ভাবনাশ্রয়ী চলের প্রত্যাশিত মান ও ভেদমান সংক্রান্ত কিছু উপপাদ্য

সম্ভাবনাশ্রয়ী চলের গাণিতিক প্রত্যাশা ও ভেদমান সংক্রান্ত কিছু উপপাদ্য আছে। কয়েকটি প্রয়োজনীয় উপপাদ্য প্রমাণ সহযোগে উপস্থাপিত করা হল । বাকি উপপাদ্যগুলিকে কেবল উল্লেখ করা হল। এদের প্রমাণ আমাদের আলোচনা বহির্ভূত। সরলতার জন্য আমরা এখানে প্রধানত সসীম সংখ্যক মানবিশিষ্ট সম্ভাবনাশ্রয়ী চলের ক্ষেত্রেই এই উপপাদ্যগুলিকে প্রমাণ করব। তবে অন্যধরনের চলের ক্ষেত্রেও এই মানগুলি অর্থপূর্ণ হলে উপপাদ্যগুলিও সত্য হবে।

উপপাদ্য ১ : যদি $x = c$ একটি ধ্রুবক হয়, তবে $E(x) = c$ ।

প্রমাণ : $E(x) = (c + c + \ldots\ldots\ldots + c)/k = kc/k = c$ ।

উপপাদ্য ২ : যদি $x = c$ একটি ধ্রুবক হয়, তবে $V(x) = 0$ ।

প্রমাণ : $V(x) = E[x - E(x)]^2 = E[c - c]^2 = 0$ ।

উপপাদ্য ৩ : যদি $y = cx$ হয়, তবে $E(y) = cE(x)$ ।

প্রমাণ : মনে করি, X একটি সম্ভাবনাশ্রয়ী চল যেটি $x_1, x_2, \ldots\ldots\ldots, x_k$ মানসমূহ গ্রহণ করতে পারে এবং $p_1, p_2, \ldots\ldots\ldots, p_k$ হল এই মানগুলির সঙ্গে সংশ্লিষ্ট সম্ভাবনা । তাহলে আমরা পাই,

$$E(y) = E(cX) = (cx_1)\ p_1 + (cx_2)\ p_2 + \ldots\ldots\ldots + (cx_k)\ p_k = cE(X)\ ।$$

উপপাদ্য ৪ : যদি $y = cx$ হয়, তবে $V(y) = c^2V(X)$ ।

প্রমাণ : $V(y) = E[y - E(y)]^2 = E[cX - E(cX)]^2$

$$= E[c^2 \{X - E(X)\}]^2 = c^2E[X - E(X)]^2 = c^2V(X)$$

উপপাদ্য ৫ : যদি $y = aX + b$ হয়, যেখানে a এবং b দুটি ধ্রুবক, তবে

$$E(y) = E(aX + b) = aE(X) + b \text{ ।}$$

প্রমাণ : $E(y) = E(aX + b) = (ax_1 + b)\, p_1 + (ax_2 + b)\, p_2 + \ldots\ldots + (ax_k + b)\, p_k$

$$= a(p_1x_1 + p_2x_2 + \ldots\ldots + p_kx_k) + b(p_1 + p_2 + \ldots\ldots + p_k) = aE(X) + b \text{ ।}$$

উপপাদ্য ৬ : যদি $y = aX + b$ হয়, যেখানে a এবং b দুটি ধ্রুবক, তবে $V(y) = V(aX + b) = a^2V(X)$ ।

প্রমাণ : $V(y) = E[y - E(y)]^2 = E[aX + b - aE(X) - b]^2 = E[aX - aE(X)]^2$

$= a^2E[X - E(X)]^2 = a^2V(X)$ । সুতরাং, $\sqrt{V(y)} = |a|\sqrt{V(X)}$ ।

উপপাদ্য ৭ : দুটি সম্ভাবনাশ্রয়ী চল X এবং Y-এর যোগফলের গাণিতিক প্রত্যাশা হল X এবং Y-এর গাণিতিক প্রত্যাশার যোগফলের সমান। অর্থাৎ, $E(X + Y) = E(X) + E(Y)$। এই উপপাদ্যটি n-সংখ্যক সম্ভাবনাশ্রয়ী চলের জন্য প্রসারিত করা যায়। যদি $X_1, X_2, \ldots\ldots\ldots\ldots X_k$, nটি সম্ভাবনাশ্রয়ী চল হয়, তাহলে $E(X_1 + X_2 + \ldots\ldots\ldots\ldots + X_n) = E(X_1) + E(X_2) + \ldots\ldots\ldots\ldots + E(X_n)$। এই উপপাদ্যটিকে বলা হয় গাণিতিক প্রত্যাশার যোগফলের সূত্র (addition law of expectation)।

উপপাদ্য ৮ : দুটি পরস্পর স্বতন্ত্র সম্ভাবনাশ্রয়ী চল X এবং Y-এর গুণফলের গাণিতিক প্রত্যাশা হল X এবং Y-এর গাণিতিক প্রত্যাশার গুণফলের সমান। অর্থাৎ, $E(XY) = E(X)E(Y)$। এই উপপাদ্যটি n-সংখ্যক স্বতন্ত্র সম্ভাবনাশ্রয়ী চলের জন্য প্রসারিত করা যায়। যদি $X_1, X_2, \ldots\ldots\ldots\ldots X_k$, nটি স্বতন্ত্র সম্ভাবনাশ্রয়ী চল হয়, তাহলে $E(X_1, X_2, \ldots\ldots\ldots\ldots, X_n) = E(X_1)\, E(X_2)\, \ldots\ldots\ldots\ldots\, E(X_n)$। এই উপপাদ্যটিকে বলা হয় গাণিতিক প্রত্যাশার গুণফলের সূত্র (multiplication law of expectation)।

স্মর্তব্য : এখানে মনে রাখতে হবে যে, গাণিতিক প্রত্যাশার গুণফলের সূত্র কেবলমাত্র পরস্পর স্বতন্ত্র সম্ভাবনাশ্রয়ী চলের ক্ষেত্রেই প্রযোজ্য। কিন্তু গাণিতিক প্রত্যাশার যোগফলের সূত্র যে কোন সম্ভাবনাশ্রয়ী চলের ক্ষেত্রেই প্রয়োগ করা যায়।

উপপাদ্য ৯ : $V(X) = E(x^2) - \mu^2$ ।

প্রমাণ : $V(X) = E[x - E(x)]^2 = E[x - \mu]^2 = E[x^2 + \mu^2 - 2x\mu] = E(x^2) + E(\mu^2) - 2E(x\mu)$

$$= E(x^2) + E(\mu^2) - 2\mu E(x) = E(x^2) + \mu^2 - 2\mu^2 = E(x^2) - \mu^2 \text{ ।}$$

৪৩.৫ সম্ভাবনাশ্রয়ী চলের পরিঘাত নির্দেশক অপেক্ষক

সম্ভাবনাশ্রয়ী চল X-এর পরিঘাত নির্ণয়ের ক্ষেত্রে আমরা সাধারণভাবে একটি অপেক্ষক $M_x(t)$-এর সংজ্ঞা দিতে পারি যেটিকে আমরা মূলকেন্দ্রিক পরিঘাত নির্দেশক অপেক্ষক বলতে পারি। এখানে সংজ্ঞা অনুসারে

$$M_x(t) = E[e^{tx}] = \Sigma^n_{x=0} e^{tx} f(x)$$

টেইলর ক্রম সম্প্রসারণ (Taylor's series expansion) দ্বারা বলা যায়,

$$M_x(t) = E[e^{tx}]$$
$$= E\{1 + tx + t^2/2!\ x^2 + t^3/3!\ x^3 + \ldots\ldots\ldots\ldots + t^r/r!\ x^r + \ldots\ldots\ldots\ldots\}$$
$$= 1 + tE(x) + t^2/2!E(x^2) + t^3/3!E(x^3) + \ldots\ldots\ldots + t^r/r!\ E(x^r) + \ldots\ldots\ldots\ldots$$
$$= 1 + t\mu'_1 + t^2/2\mu'_2 + t^3/3!\mu'_3\ \ldots\ldots\ldots\ldots + t^r/r!\mu'_r + \ldots\ldots\ldots\ldots$$

$M_x(t)$ অপেক্ষকটি সম্প্রসারণ করলে $t^r/r!$-এর সহগ হবে r-তম মূলকেন্দ্রিক পরিঘাত।

গড়কেন্দ্রিক পরিঘাত হল $M\mu(t) = E[x^{t(x-\mu)}]$

$$= E\ \{1 + t(x-\mu) + t^2/2!\ (x-\mu)^2 + \ldots\ldots\ldots\ldots + t^r/r!\ (x-\mu)^r + \ldots\ldots\ldots\ldots\}$$
$$= 1 + tE(x-\mu) + t^2/2!\ E(x-\mu)^2 + \ldots\ldots\ldots\ldots + t^r/r!\ E(x-\mu)^r + \ldots\ldots\ldots\ldots$$
$$= 1 + t\mu_1 + t^2/2!\mu_2 + \ldots\ldots\ldots\ldots + t^r/r!\mu_r + \ldots\ldots\ldots\ldots$$

$M\mu(t)$ অপেক্ষকটি সম্প্রসারণ করলে $t^r/r!$-এর সহগ হবে r-তম গড়কেন্দ্রিক পরিঘাত। এই অপেক্ষকটি থেকে আমরা বিচ্ছিন্ন সম্ভাবনাশ্রয়ী চল X-এর নিবেশনের পরিঘাত ও তীক্ষ্ণতার পরিমাপ নির্ণয় করতে পারি।

৪৩.৬ কয়েকটি তত্ত্বগত নিবেশনের প্রকৃতি নির্ণয়

৪৩.৬.১ দ্বিপদ নিবেশনের প্রকৃতি

আমরা জানি যে, কোন বিচ্ছিন্ন সম্ভাবনাশ্রয়ী চল X-এর মূলকেন্দ্রিক পরিঘাত নির্দেশক অপেক্ষক হল

$$M_x(t) = E[e^{tx}]$$

দ্বিপদ নিবেশনের ক্ষেত্রে পরিঘাত নির্দেশক অপেক্ষক হল

$$M_x(t) = \Sigma^n_{x=0} e^{tx}\ {}^nC_x\ p^x q^{n-x} = \Sigma^n_{x=0}\ n_{Cx}\ (pe^t)^x q^{n-x} = (q + pe^t)^n$$

X-এর গড়কেন্দ্রিক পরিঘাত নির্দেশক অপেক্ষক হল $M(t) = E[x^{t(x-\mu)}] = e^{-tnp}\ (q + pe^t)^n$

$$\log M(t) = -tnp + n\log\ (q + pe^t)^n$$

t দিয়ে অবকলন (differentiate) করলে পাওয়া যায়,

$$(1/M(t))(\delta M(t)/\delta t) = -np + (npe^t)/(q+pe^t)$$
$$= (-npq - np^2e^t)/(q+pe^t)$$
$$= \{-npq - np(1-p)e^t)/(q+pe^t)$$
$$= (npq\, e^t - npq)/(q+pe^t)$$
$$= npq\,(e^t - 1)/(q+pe^t)$$

p দিয়ে অবকলন করলে পাওয়া যায়,

$$(1/M(t))(\delta M(t)/\delta p) = -nt + n(e^t - 1)/(q+pe^t)$$
$$= -nt + npq(e^t - 1)/pq(q+pe^t)$$
$$= -nt + 1/pq(1/M(t))(\delta M(t)/\delta t)$$

অথবা, $(\delta M(t)/\delta p) = -nt\, M(t)) + 1/pq\, (\delta M(t)/\delta t)$

অথবা, $1/pq\, (\delta M(t)/\delta t) = (\delta M(t)/\delta p) + nt\, M(t))$

অথবা, $(\delta M(t)/\delta t) = pq[(\delta M(t)/\delta p) + nt\, M(t))]$

সংজ্ঞা অনুসারে, $M(t) = 1 + t\mu_1 + (t^2/2!)\mu_2 + (t^3/3!)\mu_3 + (t^4/4!)\mu_4 + \ldots\ldots\ldots + (t^r/r!)\mu_r + \ldots\ldots\ldots\ldots = \sum^{\infty}_{r=0} r\mu_{r-1}.t^r/r!$

অতএব, $tM(t) = t + t^2\mu_1 + (t^3/2)\mu_2 + (t^4/6)\mu_3 + (t^5/24)\mu_4 + \ldots\ldots\ldots + (t^{r+1}/r!)\mu_r + \ldots\ldots\ldots$
$$= t + 2(t^2/2)\mu_1 + 3(t^3/6)\mu_2 + 4(t^4/24)\mu_3 + 5t^5/120\mu_4 + \ldots\ldots\ldots + (r+1)[t^{r+1}/(r+1)!].\mu_r + \ldots\ldots\ldots$$
$$= \sum^{\infty}_{r=0} r\mu_{r-1}.t^r/r!$$

অথবা, $(\delta M(t)/\delta t) = \mu_1 + t\mu_2 + (t^2/2)\mu_3 + (t^3/6)\mu_4 + \ldots\ldots\ldots$
$$= \sum^{\infty}_{r=0} \mu_{r+1}\, t^r/r!$$

এবং, $(\delta M(t)/\delta p) = 0 + td\mu_1/dp + (t^2/2!)d\mu_2/dp + (t^3/3!)d\mu_3/dp + (t^4/4!)d\mu_4/dp + \ldots\ldots\ldots + (t^r/r!)d\mu_r/dp + \ldots\ldots\ldots\ldots$
$$= \sum^{\infty}_{r=1} (t^r/r!)d\mu_r/dp$$

অতএব আমরা পাই,

$$\sum^{\infty}_{r=0} \mu_{r+1}\, t^r/r! = pq\, [\sum^{\infty}_{r=1} (t^r/r!)d\mu_r/dp + n\sum^{\infty}_{r=0} r\mu_{r-1}.t^r/r!]$$

অথবা, $\mu_{r+1} = pq\, [nr\mu_{r-1} + d\mu_r/dp]$

$\mu_0 = 1,\ \mu_1 = 0$

যখন, $r = 1$,

$$\mu_2 = pq\,[n + d\mu_1/dp] = npq$$

যখন, $r = 2$,

$$\mu_3 = pq\,[2n\mu_1 + d\mu_2/dp] = pq\,d\mu_2/dp = npq^2 = npq\,(q-p)$$

যখন, $r = 3$,

$$\mu_4 = pq\,[3n\mu_2 + d\mu_3/dp] = pq[3n^2pq + n + 6np^2 - 6np]$$
$$= npq[3npq + (1 - 6pq)]$$

x-এর সম্ভাবনা নিবেশনের প্রতিবৈষম্য হল $\delta_1 = \mu_3/\sigma^3 = \mu_3/\mu_2^{3/2}$, যেখানে μ_2-এর মান শূন্য নয়।

দ্বিপদ নিবেশনের ক্ষেত্রে $\delta_1 = [npq\,(q-p)]/[npq]^{3/2} = (q-p)/(npq)^{1/2}$।

প্রতিবৈষম্য ধনাত্মক হয় যদি $\delta_1 > 0$ হয়, অর্থাৎ $p < 1/2 < q$; এক্ষেত্রে প্রতিবৈষম্য দক্ষিণায়ত।

প্রতিবৈষম্য ঋনাত্মক হয় যদি $\delta_1 < 0$ হয়, অর্থাৎ $p > 1/2 > q$; এক্ষেত্রে প্রতিবৈষম্য বামায়ত।

নিবেশনটি প্রতিসম হয় যদি $\delta_1 = 0$ হয়, অর্থাৎ $p = 1/2 = q$ ।

তেমনি, x-এর সম্ভাবনা নিবেশনের তীক্ষ্ণতাকে বলা হয় $\delta_2 = (\mu_4/\sigma^4) - 3 = (\mu_4/\mu_2^2) - 3$, যেখানে μ_2-এর মান শূন্য নয়।

দ্বিপদ নিবেশনের ক্ষেত্রে $\delta_2 = \{npq[3npq + (1 - 6pq)]\}/(npq)^2 - 3$
$$= [3(npq)^2 + npq(1 - 6pq)]/(npq) - 3$$
$$= 3 + (1 - 6pq)]/(npq) - 3 = (1 - 6pq)]/(npq)$$

যদি $\delta_2 = 0$ হয়, তবে $pq = p(1 - p) = 1/6$ হবে।

অতএব, $p^2 - p + 1/6 = 0$

বা, $p = 0.21135, 0.78865$

দ্বিপদ নিবেশনটিকে অতিতীক্ষ্ণ (leptokurtic) বলা হয় যদি $\delta_2 > 0$ হয়, অর্থাৎ $0 < pq < 1/6$ হয়, অর্থাৎ $0.78865 < p < 0.21135$ হয় ; অবতীক্ষ্ম (platykurtic) বলা হয় যদি $\delta_2 < 0$ হয়, অর্থাৎ $1/6 < pq < 1/4$ হয় অর্থাৎ $0.21135 < p < 0.78865$ হয় ; এবং, সুষম (mesokurtic) বলা হয় যদি $\delta_2 = 0$ হয়, তবে $pq = 1/6$ হয়, অর্থাৎ $p = 0.21135$ বা, $p = 0.78865$ হয়।

৪৩.৬.২ পোয়াসঁ নিবেশনের প্রকৃতি

পোয়াসঁ নিবেশনের রূপ হল $P(X=x) = (\lambda)^x e^{-\lambda}/(x)!$; $x = 0,1,2, \ldots$ । এই বিচ্ছিন্ন নিবেশনটির ক্ষেত্রে দেখা যায় যে, $\mu = E[X]$, $\mu_2 = E[X - E(X)]^2$ ও $\mu_3 = E[X - E(X)]^3$ তিনটির মানই সমান $= (\lambda)$।

প্রমাণ : আমরা জানি যে, $\sum^{\infty}_{x=0} f(x) = \sum^{\infty}_{x=0} (\lambda)^x e^{-\lambda}/(x)! = e^{-\lambda}\sum^{\infty}_{x=0} (\lambda)^x/(x)! = e^{-\lambda}e^{\lambda} = 1$ ।

এখন, $E[(x)_r] = \sum^{\infty}_{x=0} (x)_r (\lambda)^x e^{-\lambda}/(x)! = \sum^{\infty}_{x=r} [e^{-\lambda}(\lambda)^r(\lambda)^{x-r}]/(x-r)!$ (কারণ $E[(x)_r] = 0$, যখন $x = 0, 1, \ldots\ldots, r-1$)। এখানে $(x)_r = x(x-1) \ldots (x-r+1) = x!/(x-r)!$

$= (\lambda)^r e^{-\lambda} \sum^{\infty}_{x'=0} (\lambda)^{x'}/(x')!$ যেখানে $x' = x - r$ ।

$= (\lambda)^r e^{-\lambda} e^{\lambda} = (\lambda)^r$ ।

সুতরাং, $\mu_1' = E[x] = E[(x)_1] = \lambda = \mu$ ।

$\mu_2' = E[x^2] = E[(x)_2] + E[(x)_1] = \lambda^2 + \lambda$ ।

অতএব, $\mu_2 = \mu_2' - (\mu_1')^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$ ।

অনুরূপভাবে, $\mu_3 = E[x^3] = E[(x)_3] + 3E[(x)_2] + E[(x)_1]$

(কারণ $x^3 = x(x-1)(x-2) + 3x(x-1) + x$)।

$\lambda^3 + 3\lambda^2 + \lambda$ ।

সুতরাং, $\mu_3 = \mu_3' - 3(\mu_2')(\mu_1') + 2(\mu_1')^3 = (\lambda^3 + 3\lambda^2 + \lambda) - 3(\lambda^2 + \lambda)\lambda + 2\lambda^3 = \lambda$ ।

পোয়াসঁ নিবেশনের ক্ষেত্রে, $\mu_4 = 3\lambda^2 + \lambda$ । সুতরাং, $\sqrt{\beta_1} = 1/\sqrt{\lambda}$ এবং $\beta_2 = 3 + 1/\lambda$ । অতএব, নিবেশনটি সর্বদাই দক্ষিণায়ত প্রতিবৈষম্যযুক্ত এবং অতিতীক্ষ্ণ। λ-এর মানের বৃদ্ধির সাথে সাথে এই নিবেশন ক্রমশ প্রতিসম ও সমতীক্ষ্ণতাযুক্ত হতে থাকে।

৪৩.৭ সারাংশ

তত্ত্বগত সম্ভাবনা নিবেশনের প্রকৃতি বর্ণনার ক্ষেত্রে আমরা বিভিন্ন পরিমাপের বিষয় ধারণা করতে পারি। তাদের মধ্যে কয়েকটি উল্লেখযোগ্য পরিমাপ হল সম্ভাবনাশ্রয়ী চলের গাণিতিক প্রত্যাশা, পরিঘাত এবং পরিঘাত নির্দেশক অপেক্ষক। এই এককে আমরা বিচ্ছিন্ন এবং অবিচ্ছিন্ন সম্ভাবনাশ্রয়ী চলের তত্ত্বগত সম্ভাবনা নিবেশনের প্রকৃতির কিছু নিরিখ নিয়ে আলোচনা করেছি। উদাহরণ হিসাবে, আমরা দ্বিপদ নিবেশন ও পোয়াসঁ নিবেশনের প্রকৃতির দিকে আলোকপাত করেছি।

৪৩.৮ অনুশীলনী

১. একটি সম্ভাবনাশ্রয়ী চল X-এর নিবেশনটি এইরকম :

X	–1	0	1	2
Probability(p)	1/3	1/6	1/6	1/3

X-এর গাণিতিক প্রত্যাশা নির্ণয় করুন।

২. একটি ছক্কাকে এলোপাথাড়ি ভাবে ফেলা হল। ছক্কাটির উপরিভাগে যে সংখ্যাটি দেখা যাবে, তার সম্ভাবনা নির্ণয় করুন।

৩. একটি পাত্রে 7টি সাদা ও 3টি লাল বল আছে। এই পাত্র থেকে দুটি বল একসাথে এলোপাথাড়ি ভাবে তোলা হল।

(ক) তোলা বলগুলির ভেতর কোনটিই সাদা নয়, তার সম্ভাবনা নির্ণয় করুন।

(খ) তোলা বলগুলির ভেতর একটি সাদা এবং একটি লাল বল পাওয়ার সম্ভাবনা নির্ণয় করুন।

(গ) সাদা বল তোলার গাণিতিক প্রত্যাশা নির্ণয় করুন।

৪. একটি ছক্কাকে দুইবার ক্ষেপণ করা হল। প্রতিবার ক্ষেপণে 4-এর অধিক সংখ্যা পাওয়াকে সাফল্য বলা হচ্ছে। সাফল্যের সংখ্যার সম্ভাবনা নিবেশনের প্রত্যাশা ও ভেদমান নির্ণয় করুন।

৫. কোন শীতল স্থানের বিষয়ে 25 বৎসর কালব্যাপী কোন একটি সমীক্ষায় নির্ধারিত হয়েছিল যে, এই সময়ের ভেতর 10 বৎসর কম শীত পড়েছিল, 8 বৎসর মাঝারি শীত এবং বাকি 7 বৎসর অত্যধিক শীত পড়েছিল। কোন কোম্পানী কম শীতে এক বৎসরে 1000টি গরম কোট বিক্রয় করে, মাঝারি শীতে 1300টি এবং অত্যধিক শীতে 2000টি গরম কোট বিক্রয় করে। যদি একটি গরম কোট উৎপাদনের ব্যয় হয় 173 টাকা এবং কোম্পানীটি দোকানগুলিকে 248 টাকা দরে কোট বিক্রয় করে, তাহলে কোম্পানীটির বাৎসরিক প্রত্যাশিত মুনাফা নির্ণয় করুন।

৪৩.৯ উত্তরমালা

১. $E[x] = \sum px = (1/3)(-1) + (1/6)(0) + (1/6)(1) + (1/3)(2) = -(1/3) + 1/6 + 2/3 = 1/2$ ।

২. ধরা যাক, ছক্কাটির উপরিভাগে যে সংখ্যাটি দেখা যাবে সেটি হল X। তাহলে X হল একটি সম্ভাবনাশ্রয়ী চল যেটি 1, 2, 3, 4, 5, 6 এর যে কোন মান গ্রহণ করতে পারে, এবং প্রতিটি মান গ্রহণের সম্ভাবনা হল 1/6 । তাহলে আমরা পাই,

X	1	2	3	4	5	6
P	1/6	1/6	1/6	1/6	1/6	1/6

$E[x] = \sum px = 1.1/6 + 2.1/6 + 3.1/6 + 4.1/6 + 5.1/6 + 6.1/6 = 21/6 = 7/2$ ।

৩. একটি পাত্রের 7টি সাদা ও 3টি লাল বলের থেকে দুটি বল $^{10}C_2$ উপায়ে চয়ন করা যায়। ধরা যাক, X হল যে কটি সাদা বল তোলা হয়েছে তার সংখ্যা। X-এর সম্ভাবনা নিবেশন নিম্নলিখিত উপায়ে পাওয়া যায়।

$P(X = 0) = P($তোলা বলগুলির ভেতর কোনটিই সাদা নয়$) = P($তোলা বল দুটিই লাল$) = {}^3C_2/{}^{10}C_2 = 6/90 = 1/15$ (যেহেতু, 3টি লাল বলের থেকে দুটি লাল বল $^3C_2$ উপায়ে চয়ন করা যায়)।

$P(X = 1) = P($তোলা বলগুলির ভেতর একটি সাদা$) = P($তোলা বলগুলির ভেতর একটি সাদা এবং অপরটি লাল$) = ({}^7C_1.{}^3C_1)/{}^{10}C_2 = 42/90 = 21/45$ (যেহেতু, 7টি সাদা বলের থেকে একটি সাদা বল $^7C_1$ উপায়ে চয়ন করা যায় এবং 3টি লাল বলের থেকে একটি লাল বল $^3C_1$ উপায়ে চয়ন করা যায় এবং এই উপায়গুলি পারস্পরিকভাবে সম্পর্কিত)।

অনুরূপভাবে আমরা পাই,

$P(X=2) = P($তোলা বলগুলির দুটিই সাদা$) = {}^7C_1/{}^{10}C_2 = (42/2)(2/90) = 42/90 = 21/45$ ।

অতএব, সাদা বল তোলার গাণিতিক প্রত্যাশা হল $E[x] = \Sigma xP(x) = (0).(1/15) + (1)(21/45) + (2)(21/45) = 21/45 + 42/45 = 63/45 = 1.4$ ।

৪. একটি ছক্কার একটি ক্ষেপণে 4-এর অধিক সংখ্যা পাওয়া যাবে যদি 5 বা 6 পাওয়া যায়। তাহলে P(সাফল্য (S)) = 2/6 = 1/3 । অতএব, P(ব্যর্থতা (F)) = 1–1/3 = 2/3 । ধরা যাক, সাফল্যের সংখ্যাকে আমরা নির্দেশ করি X চলটির মাধ্যমে। X চলটি তিনটি মান গ্রহণ করতে পারে —0, 1 এবং 2 ।

$P(X=0) = P(F$ এবং $F) = P(F).P(F) = 2/3.2/3 = 4/9$ ।

$P(X=1) = P(F$ এবং $S) + P(S$ এবং $F) = P(F).P(S) + P(S).P(F) = 2/3.1/3 + 1/3.2/3 = 4/9$ ।

$P(X=2) = P(S$ এবং $S) = P(S).P(S) = 1/3.1/3 = 1/9$ ।

প্রত্যাশিত মান ও ভেদমান নির্মাণ

x	P(X)	x.P(x)	x^2.P(x)
0	4/9	0	0
1	4/9	4/9	4/9
2	1/9	2/9	4/9
মোট	1	$\Sigma xP(x) = 6/9$	$\Sigma x^2P(x) = 8/9$

$E[x] = \Sigma xP(x) = 0(4/9) + 1(4/9) + 2(1/9) = 6/9 = 2/3$ ।

$E[x - E[x]]^2 = \Sigma x^2.P(x) - [\Sigma xP(x)]^2 = 8/9 - (2/3)^2 = 4/9$ ।

৫. উল্লেখিত পরিসংখ্যানের ভিত্তিতে আমরা কোন একটি বৎসরের শীতে বিক্রিত গরম কোটের সংখ্যার সম্ভাবনা নিবেশন গঠন করতে পারি।

শীতের ধরন	সম্ভাবনা (p)	গরম কোটের সংখ্যা (x)	px
কম	10/25	1000	(10/25).1000 = 400
মাঝারি	8/25	1300	(8/25).1300 = 416
অত্যধিক	7/25	2000	(7/25).2000 = 560

অতএব, কোন একটি বৎসরের শীতে বিক্রিত গরম কোটের প্রত্যাশিত সংখ্যা হল

$E[x] = \sum p.x = 400 + 416 + 560 = 1376$ । যেহেতু, একটি কোটের উৎপাদনের ব্যয় হল 173 টাকা এবং বিক্রয়মূল্য হল 248 টাকা, তাই কোম্পানীটির কোট প্রতি মুনাফা হল 75 টাকা।

সুতরাং, কোম্পানীটির বাৎসরিক প্রত্যাশিত মুনাফা হল (1376 টাকা) . (75) = 103200 টাকা।

৪৩.১০ গ্রন্থপঞ্জী

- দাশগুপ্ত, ভাগবত ; চৌধুরী অরিজিৎ ; দাস, বিশ্বনাথ : *রাশিবিজ্ঞানের পরিভাষা,* বিশ্বভারতী, (১৯৭২)।
- Cramer, H. : *The Elements of Probability Theory,* John Wiley, (1955)।
- Das, N. G. : *Statistical Methods in Commerce, Accountancy and Economics,* M. Das, (1985)।
- Freund, G. E. : *Modern Elementary Statistics,* Prentice Hall, New Delhi, (1979)।
- Goon, A. M., Gupta, M. K. & Dasgupta, B. : *Fundamentals of Statistics,* Vols. I, II, World Press, (1986)।
- Kapur, J. N. & Saxena, H. C. : *Mathematical Statistics,* S. Chand, N. Delhi, (1981)।
- Mathai, A. M. & Rathie, P. N. : *Probability and Statistics,* Macmillan, Madras (1977)।
- Nagar, A. L. & Das, R. K. : *Basic Statistics,* Oxford University Press, Delhi, (1983)।
- Spiegel, M. R. : *Theory and Problems of Statistics,* Schaum Publishing Co., (1982)।
- Yule, G. U. & Kendall, M. G. : *An Introduction to the Theory of Statistics,* Griffin, (1953)।

একক ৪৪ ❑ নমুনা-সংগ্রহ সংক্রান্ত ভূমিকা—নমুনা-সংগ্রহের পদ্ধতিসমূহ—নমুনা-সংগ্রহের প্রকারভেদ

গঠন

৪৪.০ উদ্দেশ্য

নমুনা-সমীক্ষায় সাধারণত পূর্ণকের আংশভিত্তিক নমুনার ওপর প্রাপ্ত তথ্য থেকে পূর্ণক সম্বন্ধে জানার প্রচেষ্টা করা হয়। বইটির এই অংশে মূলত সম্পূর্ণ সমীক্ষার তুলনায় নমুনা-সমীক্ষায় সুবিধাসমূহ এবং কয়েকটি বিশেষ প্রচলিত নমুনা-সংগ্রহ পদ্ধতির সম্বন্ধে আলোচনা করা হয়েছে।

৪৪.১ প্রস্তাবনা

রাশিবিজ্ঞানে পূর্ণক বা সমগ্রক (population) বলতে কিছু এককের সমষ্টি বোঝায়। নানা বৈজ্ঞানিক পরীক্ষা নিরীক্ষার প্রয়োজনে প্রায়শই আমরা পূর্ণকের কোন বৈশিষ্ট্য বা পূর্ণকাঙ্ক (parameter) সম্বন্ধে জানতে আগ্রহ প্রকাশ করে থাকি। উদাহরণস্বরূপ বলা যেতে পারে, ভারতবর্ষে এড্‌স আক্রান্ত ব্যক্তির অনুপাত কত; কোন কারখানায় উৎপাদিত বস্তুগুলোর মধ্যে ত্রুটিযুক্ত বস্তুর অনুপাত কত; জার্মানদের গড় আয়ু ভারতীয়দের থেকে বেশী কিনা; কোন জায়গার অধিবাসীদের অর্থনৈতিক অবস্থা কেমন—তাদের আয় কি মোটামুটিভাবে সমান, না তাদের আয়ের মধ্যে উল্লেখযোগ্য বৈষম্য আছে ইত্যাদি। এইসব প্রশ্নের উত্তর দিতে গেলে সবচেয়ে ভাল হত যদি সম্পূর্ণ সমীক্ষা বা পূর্ণাঙ্গ পর্যবেক্ষণ (Complete enumeration or census) করা যেত যেখানে পূর্ণকের প্রত্যেকটা এককের ওপর পরীক্ষা ও নিরীক্ষণ করে প্রাপ্তব্য রাশিতথ্যের ভিত্তিতে চূড়ান্ত সিদ্ধান্ত নেওয়া হয়। কিন্তু অধিকাংশ ক্ষেত্রেই নিম্নলিখিত অসুবিধাগুলোর জন্য সম্পূর্ণ সমীক্ষা বাস্তবে সম্ভব হয়ে ওঠে না। অসুবিধাগুলো হল—

(ক) সম্পূর্ণ সমীক্ষা সময়সাপেক্ষ এবং/বা ব্যয়সাপেক্ষ হতে পারে;

(খ) পূর্ণকের আয়তন বা সংখ্যা (population size) বড় হওয়ার জন্য পরিচালনগত অসুবিধা দেখা যেতে পারে;

(গ) সম্পূর্ণ সমীক্ষা অনেক সময় ধ্বংসাত্মক হয়ে পড়ে অর্থাৎ পূর্ণকের একক বা সদস্য নিরীক্ষণের সময় নষ্ট বা ধ্বংস হয়ে যায়, যেমন—বিজলী বাতির আয়ু নির্ধারণের জন্য উৎপাদিত সব বাতিকে জ্বালান সম্ভব নয় আবার চকের গুণাগুণ বিচারের জন্য সব চক্ ব্যবহার করে দেখা সম্ভব নয় কখন চক্ ভাঙে।

উপরোক্ত আলোচনার পরিপ্রেক্ষিতে দেখা যাচ্ছে যে ইচ্ছা বা অন্যান্য সুবিধা থাকলেও সম্পূর্ণ সমীক্ষা করা অধিকাংশ ক্ষেত্রে করা সম্ভব নয়। সেইজন্য প্রায়শই আমরা প্রতিনিধিমূলক কিছু একককে পূর্ণক থেকে নির্বাচিত করে থাকি এবং এই নির্বাচিত এককের ওপর সমীক্ষালব্ধ ফলাফলের ভিত্তিতে পূর্ণকের প্রকৃতি নির্ধারণ করে থাকি। পূর্ণক থেকে এরূপ নির্বাচিত এককের সমষ্টিকে অংশক বা নমুনা (Sample) বলা হয় এবং

নমুনার ওপর যে সমীক্ষা চালান হয় তাকে বলা হয় নমুনা-সমীক্ষা (Sample survey)। নমুনা থেকে পূর্ণক সম্পর্কে যে অনুমিতি (inference) করা হয় অর্থাৎ ছোট অংশ থেকে বৃহত্তর ক্ষেত্রের সম্পর্কে ধ্যানধারণা নেওয়ার প্রক্রিয়াকে আরোহী অনুমিতি (inductive inference) বলা হয়। প্রাত্যহিক জীবনযাত্রায় এই নমুনালব্ধ সিদ্ধান্তের ব্যবহার প্রায়শই লক্ষ্য করা যায়। যেমন, ভাত রান্নার সময় চাল ঠিকমত সেদ্ধ হল কিনা দেখার জন্য প্রতিটি চালকে নিয়ে পরীক্ষা করা হয় না, পরিবর্তে, হাঁড়ির চালগুলোকে ভালভাবে নাড়িয়ে দু-একটা চাল নিয়ে পরীক্ষা করে সিদ্ধান্ত নেওয়া হয়। এখানে বলা ভাল যে, চাল নেওয়ার আগে হাঁড়ির চালগুলোকে ভালভাবে মিশিয়ে নেওয়া হয় শুধু প্রতিনিধিমূলক (representative) নমুনা পাওয়ার জন্য। অনুরূপভাবে দেখা যেতে পারে, রেশনের দোকানে চালের গুণগত মান বিচার করার জন্য বস্তার প্রতিটি চালের দানা পরীক্ষা করা হয় না। এক্ষেত্রেও নমুনাভিত্তিক সিদ্ধান্ত নেওয়া হয়।

এই স্তরে, পূর্ণক, নমুনা, সম্পূর্ণ সমীক্ষা ও নমুনা সমীক্ষার সম্বন্ধে একটা ধারণা দেওয়া হ'ল। পরবর্তী অনুচ্ছেদগুলোতে আলোচনা করা হবে যে নমুনা-সমীক্ষার সুবিধাগুলো কী, কী করে পক্ষপাতশূন্য নমুনা গ্রহণ করা হয়, নমুনা সংগ্রহের পদ্ধতিগুলো কী ইত্যাদি।

৪৪.২ নমুনা সংক্রান্ত কয়েকটি মৌলিক ধারণা

(ক) নমুনা-একক (Sampling unit) ঃ নমুনা সমীক্ষায় পূর্ণকের যে ক্ষুদ্রতম অংশ থেকে নমুনা সংগ্রহ করা হয় সেই অংশটিকে নমুনা-একক বলা হয়। স্বভাবত নমুনা সমীক্ষার উদ্দেশ্যের ওপর এই নমুনা-একক নির্ভর করবে। যেমন, কোনও রাজ্যের কৃষক-পরিবার পিছু গড় আয় সম্বন্ধে জানতে আগ্রহী হ'লে একটি কৃষক পরিবারকে নমুনা-একক হিসাবে গণ্য করা হবে। আবার কোনও বিদ্যালয়ের ছাত্রদের সম্বন্ধে খবর সংগ্রহ করতে হলে একটি ছাত্রই হবে নমুনা-একক। কোনও বন সম্পর্কিত সমীক্ষায় একটি গাছই হবে নমুনা-একক।

(খ) নমুনা-এককের পূর্ণতালিকা (Sampling Frame) ঃ পূর্ণক থেকে নমুনা সংগ্রহ করতে হ'লে যথাযথ পরিচিতিসহ সমস্ত এককের একটি পূর্ণতালিকা তৈরী করা হয়। এই তালিকাটিকেই নমুনা-এককের কাঠামো (frame) বা পূর্ণতালিকা বলা হয় এবং নমুনা সমীক্ষার ক্ষেত্রে এটি একটি প্রয়োজনীয় ভূমিকা নেয়। দেখতে হবে যে, নমুনা-এককের পূর্ণতালিকাটি যেন নির্ভুল, সমসাময়িক, সম্পূর্ণ, উপযুক্ত এবং পুনরাবৃত্তিহীন হয়।

(গ) নমুনাজ এবং অনমুনাজ ভ্রান্তি (Sampling and non-sampling errors) ঃ উপাত্ত (Data) সংগ্রহ থেকে বিশ্লেষণের বিভিন্ন পর্যায়ে যে সকল ভ্রান্তির উদ্ভব হয়, সেগুলোকে নমুনাজ এবং অনমুনাজ ভ্রান্তি হিসাবে ভাগ করা যেতে পারে। নমুনার অন্তর্গত কয়েকটা এককের ভিত্তিতে পূর্ণক সম্বন্ধে সিদ্ধান্ত নেওয়ার সময় যে ভ্রান্তির সৃষ্টি হয়, তাকে নমুনাজ ভ্রান্তি বলা হয়। প্রসঙ্গতঃ উল্লেখ্য, সম্পূর্ণ সমীক্ষার ক্ষেত্রে কোন নমুনাজ ভ্রান্তি নেই। যেহেতু নমুনা-সমীক্ষায় পূর্ণকের প্রতিটি একককে নিয়ে কাজ না করে পূর্ণকের

কয়েকটা একক নিয়ে তৈরী একটা অংশ নিয়ে কাজ করা হয়, সেইজন্য নমুনাজ ভ্রান্তির সৃষ্টি হয়। সমগ্র পূর্ণকের বৈশিষ্ট্যের সঙ্গে যদি নমুনালব্ধ বৈশিষ্ট্যের তফাৎ দেখা যায়, তাহলে তফাৎটা নমুনাজ ভ্রান্তির জন্য হয়েছে বলে ভাবা যেতে পারে। স্বভাবতই সবচেয়ে কম নমুনাভ্রান্তিযুক্ত নমুনাকেই পূর্ণকের সবচেয়ে ভাল প্রতিনিধিমূলক নমুনা হিসাবে গণ্য করা হয়। নমুনাজ ভ্রান্তি কমানোর অন্যান্য উপায়গুলোর মধ্যে দুটি উল্লেখযোগ্য হল নমুনা সংখ্যা বৃদ্ধি করা এবং সম্ভাবনাশ্রয়ী নমুনা সংগ্রহ করা।

অন্যদিকে এমন কিছু ভ্রান্তি আছে যা নমুনা সমীক্ষা এবং সম্পূর্ণ সমীক্ষা উভয় ক্ষেত্রেই লক্ষ্য করা যায়। যেমন, আয় সম্পর্কিত সমীক্ষায় আয় কমিয়ে বলার প্রবণতা দেখা যায়। বয়সও কম করে বলার প্রবণতা লক্ষ্য করা গেছে। নানা প্রকার পক্ষপাতমূলক পদ্ধতি এই অনমুনাজ ভ্রান্তির মূল কারণ।

৪৪.৩ সম্পূর্ণ সমীক্ষা বনাম নমুনা সমীক্ষা

আগেই বলা হয়েছে যে, সম্পূর্ণ সমীক্ষার ক্ষেত্রে পূর্ণকের প্রতিটি একক সংক্রান্ত তথ্য নেওয়া হয় এবং নমুনা সমীক্ষার ক্ষেত্রে পূর্ণক থেকে নেওয়া একটি প্রতিনিধিমূলক অংশক বা নমুনার অন্তর্গত এককের কাছ থেকে তথ্য গ্রহণ করা হয়। সম্পূর্ণ সমীক্ষার তুলনায় নমুনা সমীক্ষার সুবিধাগুলো নিম্নরূপ ঃ

(ক) **অর্থের সাশ্রয় ঃ** নমুনা সমীক্ষায় নমুনার অন্তর্গত একক পিছু বিশ্লেষণের ব্যয় বেশী হলেও মোট এককের সংখ্যা কম হওয়ায়, সম্পূর্ণ সমীক্ষার তুলনায় নমুনা সমীক্ষায় মোট ব্যয়ের পরিমাণ স্বাভাবিকভাবে কম হবে।

(খ) **সময়ের সাশ্রয় ঃ** জরুরীভিত্তিক কম সময়ের মধ্যে কোন সিদ্ধান্ত নিতে হলে বা সম্পূর্ণ সমীক্ষা চালানোর মত প্রয়োজনীয় অর্থ বা সময় না থাকলে নমুনা সমীক্ষা গ্রহণ করাই বাঞ্ছনীয় কেননা নমুনাতে পূর্ণকের কয়েকটা একক অন্তর্ভুক্ত থাকায় সম্পূর্ণ সমীক্ষার তুলনায় নমুনা সমীক্ষায় সময় অনেক কম লাগে।

(গ) **অধিকতর নির্ভুল ঃ** যেহেতু সম্পূর্ণ সমীক্ষার তুলনায় নমুনা সমীক্ষা ছোট, নমুনালব্ধ তথ্যের ওপর যথাযথ সতর্কতামূলক ব্যবস্থা অবলম্বন করে ত্রুটি সংশোধন করা সম্ভব হয়ে ওঠে। এইজন্য নমুনা-লব্ধ মান উচ্চমানের বলে মনে করা করা যেতে পারে। তাছাড়া অনমুনাজ ভ্রান্তি নমুনা সমীক্ষায় তুলনামূলকভাবে অনেক কম হয়।

(ঘ) **অধিকতর পরিধি ঃ** যদি দেখা যায় যে কোন সমীক্ষায় তথ্য আহরণে ও বিশ্লেষণের জন্য উচ্চট্রেনিং-প্রাপ্ত বিশেষজ্ঞ বা ব্যয়বহুল আধুনিক যন্ত্রের প্রয়োজন, তাহলে সেই সমীক্ষার ক্ষেত্রে সম্পূর্ণ সমীক্ষা কার্যত অসম্ভব হয়ে পড়বে এবং নমুনা সমীক্ষাকেই একমাত্র বিকল্প পথ হিসাবে নেওয়া যেতে পারে। এছাড়া নমুনা সমীক্ষায় অন্তর্ভুক্ত এককের সংখ্যা কম থাকায় এবং প্রয়োজনীয় সময় কম লাগায়, অনেক অতিরিক্ত বিষয়ের ওপর তথ্য সহজেই গ্রহণ করার সুযোগ থাকে। এছাড়া পূর্ণকের পরিধিই বাড়িয়ে নেবার সুযোগ থাকে।

(ঙ) **অধিকতর প্রয়োগ ঃ** পূর্ণক যদি আয়তনের দিক থেকে বা সদস্য সংখ্যার দিক দিয়ে অস্বাভাবিক বড় হয়ে পড়ে তাহলে নমুনা সমীক্ষা ছাড়া উপায় নেই। এর আগে ৪৪.১ অনুচ্ছেদে উল্লেখ করা হয়েছে যে, ধ্বংসাত্মমূলক সমীক্ষার ক্ষেত্রে নমুনা সমীক্ষা করা হয়ে থাকে। এছাড়া পূর্ণক যদি অনির্দিষ্ট (indefinite) বা অসীম (infinite) হয় যেমন, মুদ্রা নিক্ষেপণের ক্ষেত্রে বা মহাকাশের নক্ষত্রপুঞ্জের ক্ষেত্রে, তাহলে সেই পূর্ণকের ক্ষেত্রে সম্পূর্ণ সমীক্ষা করা সম্ভব নয়।

উপরোক্ত আলোচনায় আমরা নমুনা সমীক্ষার সুবিধাগুলো দেখলাম। তাহলে প্রশ্ন উঠতে পারে, সম্পূর্ণ সমীক্ষা কি কখনই করা হয় না? যদি পূর্ণক খুব বড় না হয়, সমীক্ষা চালানোয় যদি ব্যাপক খরচ ও সময় না লাগে এবং যদি পরিচালনাগত কোন অসুবিধা না দেখা দেয়, তাহলে সম্পূর্ণ সমীক্ষাকে নমুনা সমীক্ষার তুলনায় শ্রেয় হিসাবে গণ্য করা যেতে পারে এবং সেইক্ষেত্রে সম্পূর্ণ সমীক্ষাকে অনায়াসেই সুপারিশ করা যায় অধিকতর ভাল ফলের আশায়।

৪৪.৪ নমুনা-সমীক্ষার বিভিন্ন পর্যায়

নমুনা সমীক্ষার পুরো ব্যাপারটাকে প্রধানত তিনটি পর্যায়ে ভাগ করা যেতে পারে—সুষ্ঠু পরিকল্পনা, মূলসমীক্ষা এবং বিশ্লেষণ ও বিবরণী তৈরী করা। প্রতিটি পর্যায়ের আবার বিভিন্ন অংশ বা ধাপ রয়েছে।

সুষ্ঠু পরিকল্পনার জন্য নিম্নলিখিত ধাপগুলি মেনে নিয়ে কাজ করা শ্রেয়।

(ক) **উদ্দেশ্য নির্ধারণ করা ঃ** এখানে নমুনা সমীক্ষার উদ্দেশ্যগুলো পরিষ্কার এবং সঠিকভাবে নিরূপণ করতে হবে কেননা পরবর্তী কার্যক্রমের অনেকাংশই এটার ওপর নির্ভর করে। নমুনা সমীক্ষাটা কেন করা হচ্ছে, এর উপযোগিতা, এর সঠিক পরিচালন এবং এর মাধ্যমে প্রাপ্ত ফল সুদূরপ্রসারী বৃহত্তর কোন ক্ষেত্রে কাজে আসবে কিনা—এসব ব্যাপারে সংশ্লিষ্ট ব্যক্তিদের বা এজেন্সীদের যথেষ্ট স্বচ্ছ ধারণা থাকা আবশ্যক। এছাড়া গণনাকারী এবং অন্যান্য কর্মীসংখ্যা, কর্মীদের যোগ্যতা ও মান, সময়সীমা, সম্ভাব্য ব্যয়ের হিসাবের খসড়া (estimate), নমুনাভ্রান্তি ইত্যাদি ব্যাপারে সঠিক সিদ্ধান্ত নেওয়া প্রয়োজন।

(খ) **পূর্ণকের সঠিক সংজ্ঞা নিরূপণ ঃ** সমীক্ষায় অন্তর্গত পূর্ণকের সম্বন্ধে সম্যক পরিচিতি থাকা অত্যন্ত জরুরী। উদাহরণস্বরূপ বলা যেতে পারে যে, কোন গ্রামাঞ্চলের শিশু মৃত্যুহার সম্বন্ধে জানতে হলে নিম্নলিখিত বিষয়গুলোর সম্বন্ধে স্পষ্টজ্ঞান থাকা আবশ্যক— কোন ভৌগোলিক অবস্থানের বা লোকসংখ্যার ভিত্তিতে একটি জায়গাকে গ্রাম বলা যেতে পারে, শিশু বলতে কোন্ বয়সের বাচ্চাদের বোঝাবে, শিশু মৃত্যুহার নির্ণয়ের ক্ষেত্রে মৃত্যুজনিত কোন কারণগুলো ধরে কাজ করতে হবে ইত্যাদি।

(গ) **রাশিতথ্যের ধরন নির্ণয় ঃ** যদিও উদ্দেশ্য ঠিক হয়ে যাওয়ার সাথে পরিষ্কার হয়ে যায় যে কোন্ কোন্ বিষয়ের ওপর রাশিতথ্য আহরণ করতে হবে, তবুও উচিত কাজ হল যে মূল-সমীক্ষার আগে, বিষয়ের

সাথে সঙ্গতি রেখে কিছু সহজ, সরল প্রশ্ন দিয়ে একটি প্রাথমিক বিবরণপত্র (schedule) তৈরী করা এবং তা কিছুসংখ্যক সমীক্ষায় অন্তর্ভুক্ত ব্যক্তির ওপরে পরীক্ষামূলকভাবে প্রয়োগ করা। এই পরীক্ষামূলক সমীক্ষার (pilot enquiry) ভিত্তিতে প্রাথমিক বিবরণপত্রের অসঙ্গতি সংশোধন করে মূল বিবরণপত্র প্রকাশ করা হয়। সাধারণত এই পরীক্ষামূলক সমীক্ষা থেকেই একটা স্পষ্ট ধারণা হয়ে যায় কোন্ কোন্ বিষয়ের ওপর রাশিতথ্য আহরণ করা উচিত এবং কী প্রকারের রাশিতথ্য [নৈর্ব্যক্তিক (objective) না বিষয়মুখী (subjective)] গ্রহণ করা হবে। বিবরণপত্র প্রকাশের আগে দেখতে হবে যে কত কমসংখ্যক প্রশ্নের মাধ্যমে কত বেশী তথ্য সংগ্রহ করা যায়, প্রশ্নগুলো সহজবোধ্য এবং পরস্পর সঙ্গতিপূর্ণ কিন্তু প্রশ্নগুলো ধর্মনিরপেক্ষ এবং ব্যক্তিনিরপেক্ষ কিনা।

(ঘ) **রাশিতথ্য সংগ্রহের পদ্ধতি নির্ণয় :** সাধারণত এখানে সিদ্ধান্ত নেওয়া হয় যে কোন্ পদ্ধতির মাধ্যমে রাশিতথ্য সংগ্রহ করতে হবে—প্রকাশ্য সাক্ষাতের ভিত্তিতে না ডাকযোগে বিবরণপত্র (schedule) পাঠিয়ে না সরাসরি পর্যবেক্ষণের ভিত্তিতে। তবে প্রত্যেক পদ্ধতির সুবিধা ও অসুবিধাগুলো যথাযথ বিচার করে সঠিক সিদ্ধান্ত নেওয়া উচিত।

(ঙ) **নমুনা-একক ঠিক করা :** সমীক্ষার উদ্দেশ্য অনুযায়ী নমুনা-একক সম্বন্ধে পরিষ্কার ধারণা থাকা উচিত। যেমন কৃষিসমীক্ষার ক্ষেত্রে একটি চাষযোগ্য জমির প্লটকে না কয়েকটি প্লটগুচ্ছকে না গ্রামের সমস্ত চাষযোগ্য জমিকে একক হিসাবে নেওয়া হবে, আবার কোন আর্থ-সামাজিক সমীক্ষায় একটি পরিবারকে না পরিবারভুক্ত সদস্যকে একক হিসাবে নেওয়া হবে দেখা দরকার। নমুনা-একক ঠিক করার পরে পূর্ণকের অন্তর্ভুক্ত সমস্ত নমুনা-এককের একটি সুসংবদ্ধ পূর্ণ তালিকা গঠন করা আবশ্যক এবং দেখতে হবে যে তালিকাটি ত্রুটিমুক্ত কিনা অর্থাৎ দেখা প্রয়োজন পূর্ণকের বাইরে থেকে কোন একক তালিকাভুক্ত হয়েছে কিনা বা পূর্ণকের এক বা একাধিক সদস্য একাধিকবার এসেছে কিনা। নমুনা সংগ্রহের আগে এইরকম কোনও ত্রুটিযুক্ত ঘটনা ঘটে থাকলে তা সংশোধন করে সুসম্পূর্ণ তালিকা প্রকাশ করা উচিত।

(চ) **নমুনা-সমীক্ষার পরিকল্পনা :** এখানে সমীক্ষার খরচ ও উৎকর্ষতার দিকে দৃষ্টি রেখে নমুনা সংগ্রহ পদ্ধতি ও নমুনা-সংখ্যা ঠিক করা হয়।

(ছ) **সমীক্ষা-কর্মীবৃন্দের অনুশীলন :** এই স্তরে সমীক্ষায় নিযুক্ত সর্বস্তরের কর্মীবৃন্দের সমীক্ষা-সম্বন্ধীয় যাবতীয় কাজ, বিবরণপত্রের যথার্থ ব্যবহার ও তার পূরণ সম্বন্ধে বিশদভাবে তাত্ত্বিক ও বাস্তব অনুশীলনের ব্যবস্থা থাকা প্রয়োজন।

উপরোল্লিখিত ধাপগুলো অনুসরণ করার পরে আসে পরবর্তী পর্যায় অর্থাৎ মূলসমীক্ষা। এখানে নমুনা-তালিকা থেকে কোন উপযুক্ত নমুনা সংগ্রহ পদ্ধতিতে নির্বাচিত নমুনা-এককগুলির কাছ থেকে প্রয়োজনীয় তথ্যাবলি সংগ্রহ করে বিবরণপত্রে লিপিবদ্ধ করা হয়।

এর পরের পর্যায়ে আসে সংগৃহীত রাশিতথ্যের বিশ্লেষণ ও প্রতিবেদন তৈরী করা। প্রথমে ভাল করে পরীক্ষা করে দেখা উচিত যে সংগৃহীত উপাত্তগুলির মধ্যে কোন অসঙ্গতি বা বৈসাদৃশ্য আছে কিনা। সন্দেহজনক

বিবরণপত্র পুনঃসমীক্ষার জন্য ফেরৎ পাঠান উচিত। সংশোধনী নিরীক্ষণের (scrutiny) বিচারে অতিক্রান্ত উপাত্তগুলো কতগুলো প্রয়োজনীয় সারণিতে (table) বিন্যাস করে উপযুক্ত রাশিবিজ্ঞানসম্মত পরীক্ষা-নিরীক্ষার মাধ্যমে বিশ্লেষণ করা হয়। পরিশেষে, বিশ্লেষণভিত্তিক ফলাফলসহ সমীক্ষালব্ধ সকল তথ্য কয়েকটি সারণি ও বিবরণীর মাধ্যমে সাধারণের অবগতির জন্য প্রকাশ করা হয়।

৪৪.৫ নমুনা সংগ্রহের বিভিন্ন পদ্ধতি

আগেই বলা হয়েছে যে, নমুনা যেন পূর্ণকের প্রতিনিধিত্ব করে অর্থাৎ পূর্ণকের সামগ্রিক চরিত্র যেন নমুনায় সঠিকভাবে প্রতিফলিত হয়। এই উদ্দেশ্যসিদ্ধির জন্য নমুনা সাধারণত ব্যক্তি-নির্ভর ও ব্যক্তি-নিরপেক্ষ হতে পারে। যদি কোন ব্যক্তি কোন বিশেষ উদ্দেশ্যসাধনের জন্য তাঁর ইচ্ছা-অনিচ্ছার ওপর নির্ভর করে খেয়াল-খুশীমত নমুনা সংগ্রহ করেন, তাহলে সেই নমুনা ব্যক্তি-নির্ভর হয়ে পড়ে। আবার যদি খেয়াল-খুশী মাফিক নমুনা সংগ্রহ না করে বিশেষ নিয়ম-মাফিক নমুনা সংগ্রহ করা হয়, তাহলে সেই নমুনাকে ব্যক্তি-নিরপেক্ষ হিসাবে গণ্য করা হয়। ব্যক্তি-নিরপেক্ষ নমুনা চয়ন আবার তিনপ্রকারের হয়—সম্ভাবনাশ্রয়ী, সম্ভাবনাবিহীন ও মিশ্র। যদি পূর্ণকের প্রতিটি এককের নমুনায় অন্তর্ভুক্ত হওয়ার পূর্ব নির্দিষ্ট ধনাত্মক সম্ভাবনা থাকে, তাহলে সেই নমুনাচয়নকে সম্ভাবনাশ্রয়ী (probability sampling) বলা হয়। আবার যদি দেখা যায় পূর্ণকের এককের নমুনায় আসার নির্দিষ্ট কোন সম্ভাবনা নেই তাহলে সেই নমুনাচয়ন সম্ভাবনাবিহীন হবে যেমন, কোন কারখানায় উৎপাদিত বৈদ্যুতিক বাতির ক্ষেত্রে প্রতি পঞ্চম বাতি নির্বাচন করা হয়। আবার এই উদাহরণের ক্ষেত্রে যদি প্রথম পাঁচটি বৈদ্যুতিক বাতি থেকে একটিকে কোন সম্ভাবনাশ্রয়ী পদ্ধতি দ্বারা নির্বাচন করা হয় এবং তারপর প্রতি পঞ্চম বাতিকে নির্বাচন করা হয়, তাহলে সেই নমুনাচয়নকে মিশ্র বলা হয়ে থাকে।

রাশিবিজ্ঞানে সচরাচর নিম্নলিখিত নমুনা সংগ্রহ পদ্ধতিগুলো ব্যবহার করা হয়—

(i) সরল-সমসম্ভব নমুনা-সংগ্রহ।

(ii) স্তর-বিন্যস্ত সমসম্ভব নমুনা-সংগ্রহ।

(iii) নিয়মানুগ নমুনা-সংগ্রহ।

(iv) বহুবিভাগী নমুনা-সংগ্রহ।

(v) বহুপর্যায়ী নমুনা-সংগ্রহ।

(vi) গুচ্ছ নমুনা-সংগ্রহ।

(vii) বরাদ্দনির্দিষ্ট নমুনা-সংগ্রহ।

এগুলোর মধ্যে প্রথমটি নিয়ে বিস্তারিত আলোচনা হবে এবং অন্যান্য পদ্ধতিগুলোর সম্বন্ধে সংক্ষিপ্ত বিবরণ পরে দেওয়া হবে। বিশ্লেষণভিত্তিক ফলাফলসহ খুঁটিনাটি আলোচনা করার অবকাশ এখানে নেই—

শুধুমাত্র পদ্ধতিগুলোর সঙ্গে পরিচিতির প্রয়াস হয়েছে। বিশদ বিবরণের জন্য আগ্রহী শিক্ষার্থীদের গ্রন্থপঞ্জীতে উল্লিখিত বইগুলো অনুধাবন করা উচিত।

৪৪.৫.১ সরল সমসম্ভব নমুনা-সংগ্রহ

এটি একটি সম্ভাবনাশ্রয়ী নমুনাচয়ন পদ্ধতি। এই নমুনাচয়নের ক্ষেত্রে পূর্ণকের প্রতিটি এককের নমুনায় অন্তর্ভুক্ত হওয়ার সম্ভবনা সমান অর্থাৎ পূর্ণকের কোন বিশেষ একককে অন্য কোন এককের তুলনায় পক্ষপাতিত্ব করা হয় না।

সরল সমসম্ভব নমুনা সংগ্রহ দু'ধরনের হয়—পুনঃস্থাপনাসহ সরল সমসম্ভব নমুনা সংগ্রহ (Simple random sampling with replacement বা সংক্ষেপে, SRSWR) এবং পুনঃস্থাপনাবিহীন সরল সমসম্ভব নমুনা সংগ্রহ (Simple random sampling without replacement বা সংক্ষেপে, SRSWOR)।

পুনঃস্থাপনাসহ সরল সমসম্ভব নমুনা সংগ্রহের ক্ষেত্রে পূর্ণক থেকে প্রতিবারে একটি করে একক নির্বাচন করা হয় এবং নির্বাচিত একককটিকে পরীক্ষা-নিরীক্ষার পরে পূর্ণকে ফিরিয়ে দেওয়া হয় পরবর্তী একককটিকে নির্বাচিত করার আগে। এক্ষেত্রে পূর্ণকের আয়তন (population size) অপরিবর্তিত থাকে এবং পূর্ণকের এককগুলোর নমুনায় একাধিকবার আসার সম্ভাবনা থাকে। যদি পূর্ণক এবং নমুনার আয়তন যথাক্রমে N এবং n হয়, তাহলে পুনঃস্থাপনাসহ সরল সমসম্ভব নমুনা সংগ্রহের ক্ষেত্রে সহজেই দেখা যায় যে পূর্ণক থেকে কোন একক চয়নের সময় পূর্ণকের N সংখ্যক এককের প্রত্যেকটার নমুনায় অন্তর্ভুক্ত হওয়ার সম্ভাবনা সমান এবং তা হল $1/N$।

পুনঃস্থাপনাবিহীন সরল সমসম্ভব নমুনাসংগ্রহের ক্ষেত্রে পূর্ণক থেকে একটি একটি করে একক চয়ন করলেও নির্বাচিত একককটিকে কখনই পূর্ণকে ফেরৎ দেওয়া হয় না। যে কারণে প্রতিবার একক চয়নের পরে পূর্ণকের আয়তন ক্রমশ কমতে থাকে এবং তার গঠনেরও পরিবর্তন হয়। এই ধরনের নমুনা সংগ্রহের ক্ষেত্রে পূর্ণকের কোন একক একাধিকবার আসতে পারে না। এখানে প্রতি নির্বাচনের সময় পূর্ণকের অবশিষ্ট এককগুলোর প্রতিটার নমুনাতে অন্তর্ভুক্ত হওয়ার সম্ভাবনা সমান থাকে অর্থাৎ i-তম একক নির্বাচনের সময় পূর্ণকে অবস্থিত $(N - i + 1)$ অবশিষ্ট এককগুলোর প্রত্যেকটার নির্বাচনের সম্ভাবনা সমান এবং তা হল $\frac{1}{N-i+1}$।

ওপরে বর্ণনা দেওয়া দুটো সরল সমসম্ভব নমুনা সংগ্রহের ক্ষেত্রে সহজে দেখা যায়, যে কোনও একটি নির্দিষ্ট এককের (ধরা যাক, K-তম একক) যে কোনও চয়নের ক্ষেত্রে (ধরা যাক, i-তম উত্তোলনের সময়) নমুনাতে অন্তর্ভুক্ত হওয়ার সম্ভাবনা $\frac{1}{N}$। পুনঃস্থাপনাসহ সরল সমসম্ভব নমুনাসংগ্রহের ক্ষেত্রে এটা সহজেই অনুধাবনযোগ্য। পুনঃস্থাপনাবিহীন সরল সমসম্ভব নমুনা সংগ্রহের ক্ষেত্রে, K-তম এককের i-তম উত্তোলনে আসার সম্ভবনা হল,

$$\frac{N-1}{N}\times\frac{N-2}{N-1}\times\ldots\ldots\times\frac{N-i+1}{N-i+2}\times\frac{1}{N-i+1}=\frac{1}{N}$$

কেননা, i-তম উত্তোলনের আগের উত্তোলন অবধি K-তম ছাড়া অন্য এককের আসার সম্ভাবনা ধরতে হবে এবং i-তম উত্তোলনে K-তম এককেরই আসার সম্ভাবনা ধরতে হবে।

পুনঃস্থাপনাসহ ও পুনঃস্থাপনাবিহীন সরল সমসম্ভব নমুনার সংজ্ঞা আবার অন্যভাবেও দেওয়া যায়। যদি N সংখ্যক এককবিশিষ্ট পূর্ণক থেকে n সংখ্যক একককে একের পর এক নেওয়া হয়, তাহলে পুনঃস্থাপনাসহ পদ্ধতিতে মোট N^n সংখ্যক নমুনা সংগ্রহ করা যেতে পারে এবং পুনঃস্থাপনাবিহীন পদ্ধতিতে মোট N_{P_n} নমুনা নেওয়া সম্ভব (এখানে এককগুলোর আসার ক্রমকে গুরুত্ব দেওয়া হয় অর্থাৎ নমুনাতে সংগৃহীত এককগুলোর সমস্ত সম্ভবপর বিন্যাস বিবেচনা করা হয়)। পুনঃস্থাপনাসহ সরল সমসম্ভব নমুনা সংগ্রহের ক্ষেত্রে সম্ভাব্য মোট N^n সংখ্যক নমুনার মধ্যে কোন একটাকে সমান সম্ভাবনাসহ $\left(\frac{1}{N^n}\right)$ নির্বাচিত করা হয় এবং পুনঃস্থাপনাবিহীন সরল সমসম্ভব নমুনা সংগ্রহের ক্ষেত্রে সম্ভাব্য N_{P_n} নমুনা থেকে কোন একটাকে সমান সম্ভাবনাসহ $\left(\frac{1}{N_{P_n}}\right)$ নির্বাচন করা হয়। আবার, যদি নমুনার n সংখ্যক একককে পূর্ণকের N সংখ্যক একক থেকে একসাথে নেওয়া হয় পুনঃস্থাপনাবিহীন সংগ্রহের ক্ষেত্রে, তাহলে দেখা যায় মোট এইভাবে N_{C_n} সংখ্যক নমুনা নেওয়া যেতে পারে (এখানে এককগুলোকে আসার ক্রম অনুযায়ী পৃথক করে দেখা সম্ভব হয় না অর্থাৎ নমুনাতে সংগৃহীত এককগুলোর বিন্যাস অগ্রাহ্য করা হয়)। এক্ষেত্রে পুনঃস্থাপনাবিহীন সরল সমসম্ভব নমুনাকে সম্ভাব্য N_{C_n} নমুনার মধ্যে কোন একটিকে সমান সম্ভাবনাসহ $\left(\frac{1}{N_{C_n}}\right)$ চয়ন করা হিসাবে ভাবা যায়।

পরবর্তী অনুচ্ছেদে দেখব কী করে সরল সমসম্ভব নমুনা সংগ্রহ করা হয়।

(ক) **সমসম্ভব নমুনা সংগ্রহ পদ্ধতি :** সমসম্ভব নমুনা সংগ্রহের ক্ষেত্রে বাস্তবে সাধারণত দু'প্রকার পদ্ধতি গ্রহণ করা হয়।

প্রথম পদ্ধতি হল 'লটারী'—যেটা বাস্তবে বহুল প্রচলিত। এই পদ্ধতিতে প্রথমে পূর্ণকের অনুকরণে একটা ব্যবহারযোগ্য বিকল্প পূর্ণক তৈরী করা হয়। পূর্ণকের প্রত্যেকটা একককে সমআকার ও সমআয়তনের কাগজ, কার্ড বা ধাতব সিলিন্ডারের সাহায্যে প্রকাশ করে বিকল্প পূর্ণক তৈরী করা হয়। তারপরে পূর্ণকের এককগুলোকে পরপর ক্রমিক নং দিয়ে চিহ্নিত করা হয় এবং ঐ সংখ্যার অনুরূপ সংখ্যা দিয়ে বিকল্প পূর্ণকের কাগজ, কার্ড বা সিলিন্ডারগুলোকে চিহ্নিত করা হয়। এরপর বিকল্প পূর্ণকের এককগুলোকে ভালভাবে মিশিয়ে নেওয়ার পর নমুনার জন্য প্রয়োজনীয় সংখ্যক একক নেওয়া হয়। বিকল্প পূর্ণক থেকে পাওয়া এককগুলোতে যে সংখ্যাগুলো থাকে, সেই সংখ্যাযুক্ত এককগুলোকে পূর্ণক থেকে নির্বাচন করে নমুনায় অন্তর্ভুক্ত করা হয়। মনে রাখতে হবে যে, বিকল্প পূর্ণক থেকে প্রত্যেকবার একক চয়নের আগে এককগুলোকে ভালভাবে মিশিয়ে নিতে হবে।

যতই লটারী পদ্ধতি জনপ্রিয় এবং প্রচলিত হোক, রাশিবিজ্ঞানীরা সরল সমসত্ত্ব নমুনা সংগ্রহের জন্য এটাকে আদৌ সুপারিশ করেন না কেননা এখানে ব্যক্তিগত পক্ষপাত (personal bias) থাকার সম্ভাবনা থেকেই যায়। তাছাড়া, বিকল্প পূর্ণক সঠিকভাবে গঠন নাও হতে পারে—এককগুলোর মধ্যে কমবেশী তফাৎ অধিকাংশক্ষেত্রে দেখা যায়। পূর্ণক যদি খুব বড় হয় তাহলে বিকল্প পূর্ণক গঠন বাস্তবে অসম্ভব হয়ে পড়ে এবং তখন লটারীর মাধ্যমে নমুনা সংগ্রহ কার্যত অচল হয়ে যায়।

রাশিবিজ্ঞানীরা যে পদ্ধতিতে সমসম্ভব নমুনা সংগ্রহের নির্দেশ দেন তা হল সমসম্ভব সংখ্যাসারি বা রাশিমালার (Random number series) ব্যবহার। নীচে সমসম্ভব সংখ্যাসারি বা রাশিমালার সংজ্ঞা দেওয়া হল।

(খ) **সমসম্ভব সংখ্যাসারি বা রাশিমালা ঃ** এটি 0, 1, 2, ... 9 অঙ্কগুলোকে নিয়ে তৈরী এমন একটা ঋজুরৈখিক বা আয়তাকার সংখ্যাসারি যেখানে প্রত্যেক অঙ্কের যে কোন স্থানে আসার সম্ভাবনা সমান এবং যে কোন দুটি স্থানে অবস্থিত দুটো অঙ্ক পরস্পর অনপেক্ষ (independent)। এই সংখ্যাসারি এমনভাবে তৈরী করা হয়েছে যে কোন জায়গা থেকে 0, 1, 2........, 9—এই দশটি অঙ্কের যে কোন একটা অঙ্কের নির্বাচনের সম্ভাবনা সমান এবং তা হল $\frac{1}{10}$। অনুরূপভাবে, এই সংখ্যাসারির যে কোন অংশ থেকে দুই-অঙ্ক বিশিষ্ট সংখ্যা (মোট এরূপ একশটি সম্ভাব্য সংখ্যা আছে— 00,01,03...,98,99) বা তিন-অঙ্ক বিশিষ্ট সংখ্যা (এরূপ মোট হাজারটি সম্ভাব্য সংখ্যা আছে— 000, 001, 003, ..., 998, 999) নেওয়ার সম্ভাবনা সমান এবং তা হল যথাক্রমে $\frac{1}{100}$ এবং $\frac{1}{1000}$। একভাবে তিনের বেশী অঙ্ক বিশিষ্ট সংখ্যার কথাও ভাবা যেতে পারে। পাঠের সুবিধার জন্য সচরাচর এই রাশিমালা চার অঙ্কবিশিষ্ট সংখ্যার সাহায্যে প্রকাশ করা হয়। সেইজন্য সমসম্ভব সংখ্যাসারিতে প্রতি পাঁচটা লাইনের পরে এবং প্রতি চারটে স্তম্ভের পরে একটু ফাঁক রাখা হয় (পরিশিষ্ট দ্রষ্টব্য)। বিভিন্ন সমসম্ভব সংখ্যাসারির মধ্যে Tippett, L. H. C.-এর সংখ্যাসারি (Tracts for Computers, No.-XV), Kendall, M. G. এবং Smith, B.-এর সংখ্যাসারি (Tracts for computers, No-XXIV), Fisher, R. A. ও Yates, F.-এর সংখ্যাসারি (Statistical tables in Biological, Agricultural & Medical Research), Rao, Mitra, Mathai এবং Ramam-urthy-র সংখ্যাসারি (Tables and Formulae for Statistical work) উল্লেখযোগ্য।

সমসম্ভব সংখ্যাসারির ক্ষেত্রে পূর্ণকের বিকল্পপূর্ণক তৈরীর প্রয়োজন হয় না। সংখ্যাসারি থেকে নেওয়া সংখ্যাগুলো সমসম্ভব হওয়ার ফলে, বিকল্প পূর্ণকের এককগুলোকে যে মিশ্রণ করা হত, তার প্রয়োজনও হয় না। পূর্ণক যত বড় হোক না কেন সমসম্ভব সংখ্যাসারির সাহায্যে নমুনা সংগ্রহ সর্বদাই সম্ভব। এইসব সুবিধার জন্য রাশিবিজ্ঞানে সমসম্ভব নমুনা সংগ্রহের ক্ষেত্রে সমসম্ভব সংখ্যাসারি একটা গুরুত্বপূর্ণ ভূমিকা নিয়ে থাকে।

(গ) **সমসম্ভব নমুনা সংগ্রহে সমসম্ভব সংখ্যাসারির ব্যবহার ঃ** একটা উদাহরণের সাহায্যে দেখা যাক কী করে সমসম্ভব সংখ্যাসারি ব্যবহার করে সরল সমসম্ভব নমুনা সংগ্রহ করা হয়।

উদাহরণ ঃ ধরা যাক, কোন বিদ্যালয়ের কোন শ্রেণীর 32 জন ছাত্র থেকে 10 জন ছাত্রকে সরল সমসম্ভব উপায়ে নির্বাচন করতে হবে।

প্রথমে শ্রেণীর 32 জন ছাত্রকে 01, 02......, 32 সংখ্যা দ্বারা চিহ্নিত করা হয়। যেহেতু পূর্ণকের আয়তন দুই-অঙ্ক বিশিষ্ট সংখ্যা—32, সেইজন্য সমসম্ভব সংখ্যাসারি থেকে দুই-অঙ্ক বিশিষ্ট সংখ্যা নেওয়া হবে এবং সেই সংখ্যা যদি 01 থেকে 32-র মধ্যে থাকে, তাহলে তা গ্রহণ করা হবে অন্যথায় তা বর্জন করা হবে। এই প্রক্রিয়ায় দেখা যাচ্ছে যে, বিশাল সংখ্যক দুই-অঙ্ক বিশিষ্ট সংখ্যাকে বাদ দিতে হয়। 00 এবং 33, 34,......, 99–এই 68টি সংখ্যার মধ্যে কোন রাশি যদি আসে, তাহলে প্রথমেই তাকে গ্রহণ করা হবে না। তাহলে দেখা যাচ্ছে যে 68% দুই-অঙ্ক বিশিষ্ট সংখ্যা বর্জন করা হচ্ছে বা অন্যভাবে বলতে হ'লে, দুই-অঙ্ক বিশিষ্ট সংখ্যা বাতিল বা বর্জনের সম্ভাবনা হল 0.68—যেটা খুবই বেশী। এর ফলে দেখা যাচ্ছে যে, প্রয়োজনীয় নমুনা সংগ্রহ করতে অনেক সংখ্যা বাতিল হতে পারে এবং মোট সময় অনেক লেগে যেতে পারে।

ওপরে আলোচিত সমস্যার পরিপ্রেক্ষিতে, বর্জন বা বাতিলের সম্ভাবনাকে কমানোর জন্য পূর্ণকের এককগুলোকে একটার বদলে একের বেশী সংখ্যা দিয়ে চিহ্নিত করা হয়। অতএব পূর্ণকের এককগুলো চিহ্নিতকরণের পরে হয়ে পড়ে নিম্নরূপ—

01, 02, 03,..............,32

33, 34, 35,..............,64

65, 66, 67,..............,96

অর্থাৎ 33 এবং 65 সংখ্যা দিয়ে 01 নং ছাত্রটিকে চিহ্নিত করা হচ্ছে। আবার 34 এবং 66 সংখ্যা দুটো দ্বারা 02 নং ছাত্রটিকে চিহ্নিত করা হচ্ছে ইত্যাদি। যেহেতু দুই-অঙ্কবিশিষ্ট সংখ্যা মোট 100টি আছে (00, 01, 02,.....,99), সমসম্ভব বণ্টনের জন্য সর্বাধিক 96 (32-র দুই-অঙ্কবিশিষ্ট সর্বোচ্চ গুণিতক) অবধি সংখ্যা নিতে পারা যায়। এখানে দেখা যাচ্ছে, প্রত্যেকটা ছাত্র সমসংখ্যক অর্থাৎ তিনটে করে সংখ্যা দ্বারা চিহ্নিত হচ্ছে। আগের তুলনায় সমসম্ভব সংখ্যাসারি থেকে যে সংখ্যা নেওয়া হবে তার পরিধি বাড়ানোর ফলে দেখা যাচ্ছে যে খুব কম সংখ্যক দুই-অঙ্ক বিশিষ্ট সংখ্যা বাতিল বলে গণ্য করা হবে এবং এরা হল যথাক্রমে 00 এবং 97,98,99—মোট চারটি। এবং শতকরা বাতিলের পরিমাণ দাঁড়ায় মাত্র 4% অর্থাৎ এখন দুই-অঙ্কবিশিষ্ট সংখ্যার বাতিলের সম্ভাবনা মাত্র 0.04—খুবই কম, এমনকি 0.1 (10%) নয়।

এতক্ষণ আলোচনার পরে এটা পরিষ্কার হয়ে যাওয়া উচিত যে, সমসম্ভব সংখ্যাসারি থেকে নির্বাচিত উপযুক্ত আয়তনের সংখ্যার পরিধি কী হওয়া উচিত অর্থাৎ কত থেকে কত সংখ্যা নিয়ে কাজ করতে হবে, তার সম্বন্ধে সঠিক সিদ্ধান্ত নিতে হলে বর্জনের সম্ভাবনাকে যথেষ্ট গুরুত্ব দেওয়া উচিত এবং সংখ্যার পরিধি এমন নিয়ে কাজ করা উচিত যাতে বাতিল বা বর্জনের সম্ভাবনা যথেষ্ট কম থাকে।

আবার উপরোক্ত উদাহরণে ফেরা যাক। যখন ঠিক করা হয়েছে যে 01 থেকে 96 অবধি সংখ্যাকে গ্রহণ করা হবে এবং অন্য দুই-অঙ্কের সংখ্যাকে বাতিল করা হবে, তখন কোন সমসম্ভব সংখ্যাসারির যে

কোন জায়গা থেকে দুই-অঙ্কের কোন সংখ্যাকে নির্বাচন করা হয়। সাধারণত প্রথম সংখ্যাটিকে সংখ্যাসারির যে কোন জায়গা অর্থাৎ যে কোন পঙক্তি এবং সেই পঙক্তির অন্তর্ভুক্ত যে কোন স্তম্ভ থেকে নির্বাচন করার পরে, পরবর্তী সংখ্যাগুলোকে পংক্তি বরাবর বা স্তম্ভ বরাবর বা কোণাকুনিভাবে পরপর নির্বাচন করা হয়। প্রথমে নমুনা আয়তন (sample size), বর্তমান উদাহরণে নমুনা-আয়তন হল 10, অনুযায়ী প্রয়োজনীয় সংখ্যক দুই-অঙ্কবিশিষ্ট সংখ্যা সংখ্যাসারির থেকে নেওয়া হয়। নির্বাচিত সংখ্যাগুলোর মধ্যে কোন এক বা একাধিক সংখ্যা যদি বর্জনাঞ্চলের অন্তর্ভুক্ত কোন সংখ্যা হয় (যেমন, এখানে 00, 97, 98 এবং 99), তাহলে প্রথমেই তাকে বাতিল করা হবে অন্যথায় নির্বাচিত সংখ্যাটিকে পূর্ণকের আয়তন, বর্তমান উদাহরণে পূর্ণকের আয়তন 32, দিয়ে ভাগ করা হয় এবং ভাগশেষটি কত তা দেখা হয়। স্বভাবতই ভাগশেষ 0 থেকে 31-র মধ্যে থাকবে। 1 থেকে 31 অবধি ভাগশেষ অনুরূপ সংখ্যা দ্বারা চিহ্নিত পূর্ণকের একককে নির্দেশিত করে এবং ভাগশেষ যদি শূন্য হয়, তাহলে পূর্ণকের সর্বশেষ অর্থাৎ 32-তম ছাত্রটিকে নেওয়া হবে। ধরা যাক, সমসম্ভব সংখ্যাসারি থেকে নির্বাচিত সংখ্যাটি হল 76। তাহলে 32 দিয়ে ভাগ করলে ভাগশেষ দাঁড়ায় 12। সুতরাং শ্রেণী থেকে 12 তম ছাত্রটিকে নির্বাচন করা হবে। এইভাবে একক চয়ন চালিয়ে যাওয়া হবে, যতক্ষণ না নির্দিষ্ট আয়তনের নমুনা না পাওয়া যায় অর্থাৎ যতক্ষণ না 10 জন ছাত্রকে নির্বাচন করা যাচ্ছে, ততক্ষণ এই চয়ন প্রক্রিয়া চালিয়ে যাওয়া হবে।

এই প্রসঙ্গে মনে রাখতে হবে যে, একই একক একাধিকবার এসে তা গ্রহণ করা হবে যদি পুনঃস্থাপনাসহ পদ্ধতিতে সরল সমসম্ভব নমুনা সংগ্রহ করা হয় এবং তাকে প্রথমবারের পর বর্জন করা হবে যদি পুনঃস্থাপনাবিহীন সরল সমসম্ভব নমুনা সংগ্রহ করা হয়।

নীচের সারণিটি থেকে সংক্ষিপ্তাকারে ওপরের চয়নপ্রণালী বোঝা যাবে। এখানে পরিশিষ্টে প্রদত্ত সমসম্ভব সংখ্যাসারির ষষ্ঠ পংক্তির নবম স্তম্ভ থেকে শুরু করে পরপর পংক্তি বরাবর সংখ্যা নেওয়া হয়েছে।

সারণি ৪৪.১—সমসম্ভব নমুনা চয়ন

নির্বাচিত সংখ্যা	নির্বাচিত সংখ্যাকে 32 দিয়ে ভাগ করে প্রাপ্ত ভাগশেষ	গৃহীত ক্রমিক সংখ্যা	মন্তব্য
55	23	23	গ্রহণ করা হবে
33	01	01	”
25	25	25	”
77	13	13	”
43	11	11	”
48	16	16	”

নির্বাচিত সংখ্যা	নির্বাচিত সংখ্যাকে 32 দিয়ে ভাগ করে প্রাপ্ত ভাগশেষ	গৃহীত ক্রমিক সংখ্যা	মন্তব্য
09	09	09	,,
71	07	07	,,
25	25	25	(i) (নিচে দ্রষ্টব্য)
80	16	16	(ii) (নিচে দ্রষ্টব্য)

মন্তব্য (i) এবং (ii) ঃ 25-তম এবং 16-তম ছাত্রদুটোকে আবার বিবেচনা করা হবে যদি পুনঃস্থাপনাসহ সরল সমসম্ভব নমুনা নেওয়া হয়। আর যদি সরল সমসম্ভব নমুনাটিকে পুনঃস্থাপনাবিহীন পদ্ধতিতে নেওয়া হয়, তাহলে ওই দুটো সংখ্যাকে বাতিল করে, পরে আবার সমসম্ভব সংখ্যাসারি থেকে দুটো সংখ্যা নিয়ে দেখতে হবে। তাহলে ওপরের সারণিটি থেকে দেখা যাচ্ছে যে, পুনঃস্থাপনাসহ সরল সমসম্ভব নমুনায় 23, 01, 25, 13, 11, 16, 09, 07, 25 এবং 16 ক্রমিক সংখ্যাযুক্ত ছাত্রগুলোকে অন্তর্ভুক্ত করতে হবে। পুনঃস্থাপনাবিহীন সরল সমসম্ভব নমুনার ক্ষেত্রে যদিও প্রথম আটজনকে অতি অনায়াসেই বিবেচনা করা যেতে পারে। শেষের দুটির জন্য অন্য সংখ্যা নির্বাচন করে আলাদা কোন ছাত্র চয়নের ব্যাপারে সঠিক সিদ্ধান্ত নেওয়া দরকার।

বিশেষ দ্রষ্টব্য

(১) সাধারণত পূর্ণকের আয়তনের ওপর সমসম্ভব সংখ্যার আয়তন নির্ভর করে। দৃষ্টান্তস্বরূপ, পূর্ণকের আয়তন যদি দুই অঙ্কবিশিষ্ট হয়, তাহলে সমসম্ভব সংখ্যাসারি থেকে দুই-অঙ্কের সংখ্যা নির্বাচন করা হবে, আবার পূর্ণকের আয়তন তিন অঙ্কবিশিষ্ট হ'লে তিন অঙ্কের সংখ্যা নেওয়া হয় ইত্যাদি। কিন্তু, আবার এর ভিন্নতাও ক্ষেত্রবিশেষে লক্ষ্য করা হয়—যেমন, 52টা কলেজ থেকে 5টা কলেজকে নিতে হলে, দুই-অঙ্কবিশিষ্ট সংখ্যা নিয়ে কাজ করলে বিশেষ উপকার পাওয়া যাবে না কেননা এখানে বর্জনের সম্ভাবনা খুবই বেশী এবং তা হল 48%, যেহেতু 01 থেকে 52-র বেশী সংখ্যা নিয়ে কাজ করা সম্ভব নয় সমসম্ভব নমুনা নিতে গেলে। এই কারণে এখানে দুই-অঙ্কবিশিষ্ট সংখ্যা চয়নের কথা ছেড়ে দিয়ে তিন-অঙ্কবিশিষ্ট সংখ্যা চয়ন করতে হবে অর্থাৎ 001 থেকে 988 (52-র তিন অঙ্কবিশিষ্ট সর্বোচ্চ গুণিতক) অবধি সংখ্যা থেকে নির্বাচিত সংখ্যাকে নিয়ে কাজ করতে হবে। এরফলে তিন অঙ্কবিশিষ্ট সংখ্যার বর্জনের সম্ভাবনা অত্যন্ত কম এবং তা হল মাত্র 0.012। সুতরাং বর্জনের সম্ভাবনার কথা ভেবে সমসম্ভব সারি থেকে নেওয়া সংখ্যার আয়তন ঠিক করতে হবে। যদি পূর্ণকের আয়তনবিশিষ্ট সংখ্যা নিয়ে বিশেষ উপকার না পাওয়া যায়, তাহলে পরের আয়তনবিশিষ্ট সংখ্যা নিয়ে কাজ করা উচিত।

(২) যদি পূর্ণকের আয়তন 5, 10, 20, 25, 50, 100, 1000 ইত্যাদি হয় তাহলে সবকটি এক অঙ্ক, দুই অঙ্ক বা তিন অঙ্ক বিশিষ্ট সংখ্যাকে ক্ষেত্রবিশেষে ব্যবহার করা হয়। যেহেতু সবকটি সমসম্ভব সংখ্যাকে নিয়ে কাজ করা হয় সেইজন্য এখানে বর্জনের সম্ভাবনা শূন্য।

৪৪.৫.২ স্তরবিন্যস্ত সমসম্ভব নমুনা-সংগ্রহ

সরল সমসম্ভব নমুনার ক্ষেত্রে এমন হতে পারে যে পূর্ণকের কোন বিশেষ স্তর বা গোষ্ঠীভুক্ত সদস্যরা নাও আসতে পারে। এই অসুবিধা স্তর বিন্যস্ত সমসম্ভব নমুনা সংগ্রহের মাধ্যমে এড়ান যেতে পারে। এখানে সমগ্র পূর্ণকটিকে কয়েকটি স্তরে ভাগ করা হয় যেখানে প্রতিটি স্তরের অন্তর্ভুক্ত এককগুলো কোন বৈশিষ্ট্যের ভিত্তিতে পরস্পর সদৃশ অথচ স্তরগুলোর মধ্যে যথেষ্ট পার্থক্য আছে। উদাহরণস্বরূপ বলা যেতে পারে যে কোন জায়গার জনসাধারণকে আয়ের ভিত্তিতে নিম্নবিত্ত, মধ্যবিত্ত এবং উচ্চবিত্ত স্তরে ভাগ করা যেতে পারে। এখানে দেখা যাচ্ছে যে প্রতিটি গোষ্ঠী বা স্তরভুক্ত জনগণ আয়ের ভিত্তিতে অপেক্ষাকৃত সদৃশ অথচ বিভিন্ন স্তরের লোকেদের মধ্যে যথেষ্ট বৈসাদৃশ্য আছে। উপরোক্ত তথ্যের ভিত্তিতে পূর্ণককে বিভিন্ন স্তরে ভাগ করার পরে প্রয়োজনীয় সদস্যসংখ্যা সরল সমসম্ভব নমুনার ভিত্তিতে প্রতিটি স্তর থেকে স্বতন্ত্রভাবে নেওয়া হয়। তাহলে দেখা যাচ্ছে যে, স্তরবিন্যস্ত সমসম্ভব নমুনা সংগ্রহের ক্ষেত্রে প্রতিটি স্তরের অন্তর্ভুক্ত কিছু সংখ্যক একক নমুনাতে থাকবেই। এখন প্রশ্ন ওঠে, প্রতিটি স্তর থেকে কতগুলো একক নেওয়া যায়। n_i যদি i-তম স্তর থেকে নেওয়া নমুনা-সংখ্যা হয়, প্রকৃষ্ট বণ্টনের (optimum allocation) ক্ষেত্রে n_i-র মান এমনভাবে নির্ণয় করা হয় যে, নির্দিষ্ট খরচের সাপেক্ষে প্রাককলকের মোট ভেদমান (variance) যেন কম হয় অর্থাৎ সূক্ষতা (precisioh) যেন বেশী হয়। প্রাক্কলকের সূক্ষতা হিসাবে ভেদমানের অনোন্যক বা পূরক (reciprocal)-কে ধরা হয়। নানা বণ্টনের ক্ষেত্রে n_i-র মান নিম্নরূপ ঃ

(ক) **প্রকৃষ্ট বণ্টন ঃ** $n_i \infty \dfrac{N_i \sigma_i}{\sqrt{C_i}}$

যেখানে N_i হল i-তম স্তরের অন্তর্গত পূর্ণকের একক সংখ্যা, σ_i হল i-তম স্তরের অন্তর্গত N_i সংখ্যক এককে সমক-বিচ্যুতি (standard deviation) এবং C_i হল i-তম স্তরের কোন একককে নমুনায় অন্তর্ভুক্তি করার খরচ।

(খ) **নেম্যান (Neyman)-এর বণ্টন ঃ** $n_i \propto N_i \sigma_i$ (এখানে ধরে নেওয়া হয় যে একক প্রতি খরচ প্রতিটি স্তরের ক্ষেত্রে সমান।)

(গ) **বোলে (Bowley's) সমানুপাতিক বণ্টন ঃ** $n_i \infty N_i$ (এই বণ্টন অনুযায়ী সচরাচর বড় আয়তনের স্তর থেকে বেশী সংখ্যক একক নেওয়া হয় এবং কম আয়তনের স্তর থেকে কমসংখ্যক একক নেওয়া হয় নমুনা সংগ্রহের উদ্দেশ্যে।) এক্ষেত্রে প্রতিটি স্তরের সমক বিচ্যুতি ও একক প্রতি খরচ সমান ধরা হয়। প্রথমে উল্লেখ করা সুবিধা ছাড়া সরল সমসম্ভব নমুনার তুলনায় স্তরবিন্যস্ত সমসম্ভব নমুনা সংগ্রহের আরো কিছু সুবিধা আছে যেমন বহুক্ষেত্রে স্তরবিন্যস্ত সমসম্ভব নমুনা সংগ্রহে পরিচালনাগত সুবিধা লক্ষ্য করা যায়। এছাড়া এখানে প্রাক্কলকের সূক্ষ্মতা বেশী পাওয়া যায়। যদি পূর্ণকের এককের মধ্যে ভালরকম তারতম্য বা বৈচিত্র্য দেখা যায় যেমন কোন জায়গায় স্বাক্ষর ও নিরক্ষর উভয়প্রকার লোক বাস করে অথবা কোন জায়গার লোকজন ঘর ছাড়াও অন্য জায়গায় বাস করে যেমন হস্‌পিটাল, হোটেল বা অন্য কোন প্রতিষ্ঠানে ইত্যাদি, তাহলে সেখানে স্তরবিন্যস্ত সমসম্ভব নমুনা সংগ্রহকে উপযুক্ত বলে মনে করা হয়।

যদি দেখা যায় পূর্ণকটিকে বিভিন্ন স্তরে ভাগ করা ব্যয়সাপেক্ষ এবং অসুবিধাজনক এবং পর্যাপ্ত পরিমাণে ভ্রমশূণ্যতা (accuracy) পাওয়ার মত যদি বিভিন্ন স্তরের মধ্যে যথেষ্ট তফাৎ বা পার্থক্য না থাকে, তাহলে স্তরবিন্যস্ত সমসম্ভব নমুনা সংগ্রহকে উপযোগী হিসাবে ধরা উচিত নয়।

৪৪.৫.৩ নিয়মানুগ নমুনা-সংগ্রহ

যেসব পূর্ণকের এককগুলো কোন নিয়ম অনুসারে, হয় বর্ণানুক্রমিক (alphabetical) নয় কালক্রমানুসারী (chronological), সাজান থাকে এবং প্রত্যেকটি এককের সাথে একটা ক্রমিকসংখ্যা থাকে, সেইসব পূর্ণক থেকে নমুনা সংগ্রহ করতে হলে এই বিশেষ পদ্ধতিতে নমুনা সংগ্রহ করা যেতে পারে। যদি পূর্ণকের আয়তন N এবং নমুনার আয়তন n হয় এবং $\frac{N}{n} = k$ (k, যেখানে, একটি অখণ্ড সংখ্যা), তাহলে নিয়মানুগ পদ্ধতিতে পূর্ণক-তালিকার প্রথম k-সংখ্যক একক থেকে যে কোন একটি একককে সরল সমসম্ভব পদ্ধতিতে নির্বাচন করা হয় এবং তারপর থেকে পরপর k-তম একক সংগ্রহ করে যাওয়া হয় যতক্ষণ না প্রয়োজনীয় সংখ্যক n একককে পাওয়া যায়। উদাহরণস্বরূপ, ধরা যাক কোন জেলার অন্তর্গত 5000 গ্রাম থেকে 50টা গ্রামকে নিতে হবে। এখানে N = 5000, n = 50 এবং k = 100। গ্রামগুলিকে প্রথমে 1 থেকে 1000 ক্রমিক সংখ্যা দেওয়া হল। প্রথমে 1 থেকে 100-র মধ্যে কোন একটি সংখ্যা, ধরা যাক, 35 সমসম্ভব সংখ্যাসারির মাধ্যমে নির্বাচন করা হল। তাহলে 35 ক্রমিক সংখ্যার গ্রামটিকে প্রথমেই চয়ন করা হল। এরপর তালিকা থেকে প্রতি 100-তম গ্রামকে চয়ন করা হবে অর্থাৎ তালিকাভুক্ত 135, 235, 335,........ইত্যাদি ক্রমিব নং বিশিষ্ট গ্রামগুলোকে নমুনাতে অন্তর্ভুক্ত করা হবে।

তাহলে দেখা যাচ্ছে যে, সরল সমসম্ভব নমুনা সংগ্রহ পদ্ধতির মত এখানেও নমুনা সংগ্রহের আগে পূর্ণকের এককগুলোর পূর্ণতালিকা তৈরী করা বিশেষ প্রয়োজন। সরল সমসম্ভব নমুনা সংগ্রহের ক্ষেত্রে যেখানে সমসম্ভব সংখ্যাসারির সাহায্যে নমুনা-একককে নির্বাচন করা হয়, নিয়মানুগ নমুনা সংগ্রহের ক্ষেত্রে শুধু প্রথমটিকেই সমসম্ভবভাবে নির্বাচন করা হয়। এইজন্য নিয়মানুগ নমুনা সংগ্রহ পদ্ধতিকে মিশ্র নমুনা সংগ্রহ পদ্ধতি হিসাবে ভাবা যেতে পারে—অংশত সম্ভাবনাশ্রয়ী (প্রথম নমুনা-এককটি নির্বাচনের ক্ষেত্রে) ও অংশত সম্ভাবনা-নিরপেক্ষ (প্রথম এককটি ছাড়া বাকী নমুনা-একক সংগ্রহের ক্ষেত্রে কেননা এখানে বাকী এককগুলো নির্দিষ্ট হয়ে পড়ে যেইমাত্র প্রথমটিকে নির্বাচন করা হয়)।

কখনও কখনও দেখা যায় যে নমুনাজ অন্তর (Sampling interval) $\frac{N}{n}$ অখণ্ড সংখ্যা নয়। তখন বৃত্তীয় নিয়মানুগ নমুনা সংগ্রহ পদ্ধতিতে (Circular Systematic sampling) নমুনা সংগ্রহ করা যেতে পারে। এখানে পূর্ণকের N সংখ্যক একক থেকে যে কোন একটিকে সমসম্ভব পদ্ধতিতে নির্বাচন করা হয় এবং নমুনায় অন্তর্ভুক্ত করা হয়। তারপর থেকে প্রতি K-তম একককে নির্বাচন করে যাওয়া হয় বৃত্তাকারে যতক্ষণ না প্রয়োজনীয় n সংখ্যক একককে নেওয়া হয়। এখানে K হল $\frac{N}{n}$-র সবচেয়ে কাছের অখণ্ড সংখ্যা।

ধরা যাক পূর্ণকের আকার N = 31 ও নমুনার আকার n = 4। 31টি একক থেকে যে কোন একটি সমসম্ভব পদ্ধতিতে নির্বাচন করা হ'ল। ধরা যাক, নির্বাচিত এককটির ক্রমিক সংখ্যা 13। যেহেতু $\frac{31}{4} = 7.75$; যা 8 এর আসন্ন। তাহলে অন্যান্য নির্বাচিত ক্রমিক সংখ্যা বৃত্তাকারে নিলে হবে 21, 29, ও 6।

নিয়মানুগ নমুনা সংগ্রহ পদ্ধতি ব্যবহারিক দিক থেকে অত্যন্ত সুবিধাজনক। পদ্ধতিটি খুব সহজ ও কম সময়ের মধ্যে এই পদ্ধতিতে নমুনা সংগ্রহ করা যায়। অসুবিধা বা ত্রুটির মধ্যে দেখা যায় যে এই পদ্ধতিতে প্রাপ্ত প্রাককলক বেশ ভালরকম পক্ষপাতদুষ্ট (biased) হয়, যদি k-তম এককগুলিতে বা k-র কোন গুণিতক এককগুলির ক্ষেত্রে পুনরাবৃত্ত বৈশিষ্ট্য (periodic features) দেখা যায়। যদি ক্রমানুসারে সাজান মানগুলিতে ক্রমশ বাড়ার বা কমার প্রবণতা থাকে তাহলে এই পদ্ধতির প্রাককলক সরল সম্ভব পদ্ধতির থেকে অধিকতর নিখুঁত। এইসব অসুবিধা সত্ত্বেও এই পদ্ধতি বেশ ভালভাবে ব্যবহৃত হয় উৎপন্ন দ্রব্যের নমুনা নেওয়ার সময়, কৃষিভিত্তিক সমীক্ষায়, বনাঞ্চলে ও আর্থ-সামাজিক সমীক্ষার জন্য।

৪৪.৫.৪ বহুবিভাগী নমুনা-সংগ্রহ

বিভিন্ন বিভাগ বা ধাপ অনুযায়ী নমুনা সংগ্রহ পদ্ধতিকে বহু বিভাগী নমুনা সংগ্রহ বলা হয়। এখানে প্রথমে পূর্ণককে কয়েকটা বড় অংশে ভাগ করা হয়, যেগুলোকে প্রথম বিভাগীয় একক (first stage unit) হিসাবে গণ্য করা হয়। এরপর প্রথম বিভাগীয় এককগুলোকে তুলনামূলকভাবে কয়েকটা ছোট অংশে ভাগ করা হয়, যেগুলোকে বলা হয় দ্বিতীয় বিভাগীয় একক (second stage unit)। এইভাবে দ্বিতীয় বিভাগীয় একককে তৃতীয় বিভাগীয় এককে (third stage unit), তৃতীয় বিভাগীয় একককে চতুর্থ বিভাগীয় এককে (fourth stage unit) ইত্যাদিতে ভাগ করে যাওয়া হয় যতক্ষণ না প্রয়োজনীয় গ্রহণযোগ্য নমুনা এককে পৌঁছতে পারা যায়।

এই নমুনা সংগ্রহ পদ্ধতিতে প্রথমে প্রথম বিভাগীয় এককগুলো থেকে সমসম্ভব উপায়ে কিছু একককে নেওয়া হয়। এরপর প্রত্যেক নির্বাচিত প্রথম বিভাগীয় একক থেকে কয়েকটা দ্বিতীয় বিভাগীয় একককে সমসম্ভব উপায়ে নেওয়া হয়। এইভাবে নমুনা সংগ্রহ করে যেতে হবে যতক্ষণ না সর্বশেষ নমুনা এককের বাঞ্ছিত নমুনা নেওয়া হয়। উদাহরণস্বরূপ, কোন রাজ্যের কয়েকটা পরিবার নির্বাচন করতে হলে রাজ্যের অন্তর্গত জেলাগুলোকে প্রথম বিভাগীয় একক হিসাবে ধরা যেতে পারে। আবার, প্রতিটি জেলার অন্তর্গত ব্লকগুলোকে, ব্লকের অন্তর্গত গ্রামগুলোকে এবং গ্রামের অন্তর্গত পরিবারগুলোকে যথাক্রমে দ্বিতীয়, তৃতীয় এবং চতুর্থ বিভাগীয় একক হিসাবে গণ্য করা যেতে পারে। বহুবিভাগী নমুনা সংগ্রহের মাধ্যমে নমুনা সংগ্রহ কালেও প্রথমে কয়েকটা জেলাকে সমসম্ভব উপায়ে নির্বাচন করা হয়। তারপর, প্রত্যেক নির্বাচিত জেলাগুলো থেকে কতগুলো ব্লককে প্রত্যেক নির্বাচিত মহকুমা থেকে কয়েকটা গ্রামকে এবং প্রত্যেক নির্বাচিত গ্রাম থেকে প্রয়োজনীয় সংখ্যক পরিবারগুলিকে নির্বাচন করা হয় সমসম্ভব উপায়ে।

বহুবিভাগী নমুনা সংগ্রহ পদ্ধতির ব্যবহারিক সুবিধাগুলোর মধ্যে প্রথমেই উল্লেখ করতে হয় এর বিস্তৃত প্রয়োগসীমা। এছাড়া এখানে দ্বিতীয় নমুনা এককের পূর্ণতালিকা করা প্রয়োজন শুধুমাত্র নির্বাচিত প্রথম বিভাগীয়

এককগুলোর ক্ষেত্রে। পরবর্তী বিভাগীয় নমুনা একক সম্বন্ধে একই কথা প্রযোজ্য। এইজন্য এই পদ্ধতির সাহায্যে নমুনা সংগ্রহে সর্বোপরি খরচ কম হয়। এছাড়া দুর্গমস্থানযুক্ত পূর্ণকের ক্ষেত্রে এই পদ্ধতিতে নমুনা সংগ্রহ করা বেশ সুবিধাজনক। অসুবিধার মধ্যে বলতে হয়, অন্যান্য নমুনাসংগ্রহ পদ্ধতিগুলোর তুলনায় এই পদ্ধতি যথেষ্ট নিখুঁত নয়। তাছাড়া বিভাগ বেড়ে যাওয়ার সাথে রাশিবিজ্ঞানসম্মত বিশ্লেষণও জটিল হয়ে পড়ে।

৪৪.৫.৫ বহুপর্যায়ী নমুনা-সংগ্রহ

যদি কোন সমীক্ষার ক্ষেত্রে দেখা যায় যে, যে সকল তথ্য সংগ্রহ করতে হবে তারা সমান গুরুত্বপূর্ণ নয় এবং তাদের সংগ্রহ করার খরচ এক নয়, তাহলে সেই সমীক্ষায় এই প্রক্রিয়ায় নমুনা সংগ্রহ করা যেতে পারে। এখানে প্রথম পর্যায়ে কিছু একককে নির্বাচন করে, অতি সহজেই এবং কম খরচে পাওয়া যায় এরকম কিছু গুরুত্বপূর্ণ তথ্য সংগ্রহ করা হয়। পরবর্তীকালে, দ্বিতীয় পর্যায়ে, প্রথম পর্যায়ে নির্বাচিত নমুনার একটি অংশ নির্বাচন করার পরে অন্যান্য তথ্য গ্রহণ করা হয়। প্রথম পর্যায়ে পাওয়া তথ্যাবলীকে দ্বিতীয় পর্যায়ে নমুনা সংগ্রহের ব্যাপারে ব্যবহার করা যেতে পারে। এইভাবে পর্যায় সংখ্যা প্রয়োজনমত বাড়ান যেতে পারে। উদাহরণস্বরূপ, যদি জানতে চাওয়া হয় কোন শিল্পাঞ্চলে শতকরা কতজন যক্ষ্মা রোগাক্রান্ত, তাহলে এইভাবে এগোন যাবে। প্রথম পর্যায়ে পূর্ণকের N সংখ্যক ব্যক্তির মধ্য থেকে নির্বাচিত n সংখ্যক ব্যক্তির ক্ষেত্রে প্রাথমিক ডাক্তারি পরীক্ষার মাধ্যমে জানা ভাল যে n_1 জনের মধ্যে যক্ষ্মার জীবাণু থাকতে পারে এবং অবশিষ্ট $n_2 = (n - n_1)$ জনের মধ্যে যক্ষ্মার জীবাণু হয়ত নেই। এরপর দ্বিতীয় পর্যায়ে, প্রথম পর্যায়ে পাওয়া তথ্যের ভিত্তিতে তৈরী করা দুটো ভাগ থেকে যথাক্রমে m_1 এবং m_2 জনকে নির্বাচিত করা হল এবং এক্স-রে পরীক্ষার মাধ্যমে দেখা গেল যে তাদের মধ্যে যথাক্রমে x_1 এবং x_2 জন যক্ষ্মা রোগাক্রান্ত। এইভাবে বহুপর্যায়ী নমুনাসংগ্রহ করা হয়। বর্তমান উদাহরণে পূর্ণকের মধ্যে যক্ষ্মা রোগাক্রান্ত ব্যক্তির শতকরা হিসাবের প্রাক্কলব হিসাবে নেওয়া যেতে পারে।

$$\frac{\dfrac{n_1 x_1}{m_1} + \dfrac{n_2 x_2}{m_2}}{n_1 + n_2} \times 100$$

লক্ষ্যণীয় যে, বহুবিভাগী ও বহুপর্যায়ী নমুনা সংগ্রহ পদ্ধতি দুটো কিন্তু আলাদা—বহুপর্যায়ী নমুনা সংগ্রহের ক্ষেত্রে নমুনা-একক একই থাকে, শুধুমাত্র বিভিন্ন পর্যায়ে নমুনা সংখ্যার পরিবর্তন ঘটে, অথচ বহুবিভাগী নমুনা সংগ্রহের ক্ষেত্রে বিভিন্ন বিভাগের নমুনা-একক আলাদা।

৪৪.৫.৬ গুচ্ছ নমুনা-সংগ্রহ

এই নমুনা সংগ্রহ পদ্ধতিতে প্রথমে পূর্ণককে কয়েকটা গুচ্ছ বা ব্লকে ভাগ করা হয় যেখানে কোন গুচ্ছের অন্তর্গত নমুনা-এককগুলোর মধ্যে বৈসাদৃশ্য থাকতে পারে কিন্তু বিভিন্ন গুচ্ছগুলোর মধ্যে তারতম্য খুব একটা লক্ষ্য করা যায় না। এরপর গুচ্ছগুলোর মধ্য থেকে কয়েকটাকে উপযুক্ত নমুনা সংগ্রহ পদ্ধতির মাধ্যমে

সাধারণত সরল সমসম্ভব নমুনা সংগ্রহ পদ্ধতি দ্বারা নির্বাচন করে প্রত্যেকটার ওপর সম্পূর্ণ সমীক্ষা করা হয়। যদিও এই পদ্ধতি খুব সহজ ও সাধারণ, কিন্তু এটা মোটেই বিজ্ঞানসম্মত নয় এবং এখানে ব্যক্তিগত পক্ষপাত (personal bias) থাকা অস্বাভাবিক নয়। তবুও পরিচালনগত ও প্রশাসনিক স্তরে কিছু সুবিধা থাকায় এই পদ্ধতিতে নমুনা সংগ্রহ অনেকক্ষেত্রে পছন্দ করা হয়।

মন্তব্য : গুচ্ছ নমুনা সংগ্রহর সঙ্গে স্তর বিন্যস্ত সমসম্ভব নমুনা সংগ্রহের মিল খুঁজে পেলেও গুচ্ছের সঙ্গে স্তরকে এক করলে হবে না কেননা স্তরের অন্তর্ভুক্ত নমুনা-এককগুলোর মধ্যে মোটামুটি সাদৃশ্য থাকলেও গুচ্ছের ক্ষেত্রে তা না-ও হতে পারে।

৪৪.৫.৭ বরাদ্দনির্দিষ্ট নমুনা সংগ্রহ

এই পদ্ধতিতে কোন অন্তর্সম (homogeneous) পূর্ণককে কয়েকটা অংশে (segments) ভাগ করা হয় এবং পূর্বনির্দিষ্ট অংশ পিছু বরাদ্দ করা অনুপাত অনুযায়ী গণনাকারীকে বলা হয় প্রতি অংশ থেকে নমুনা সংগ্রহ করতে। এখানে গণনাকারী তার নিজস্ব বিচারবুদ্ধি কাজে লাগিয়ে অংশ পিছু বরাদ্দ করা নমুনা সংগ্রহ করতে পারেন। উদাহরণস্বরূপ, কোন বিশ্ববিদ্যালয়ের 20টি বিভাগের মোট 3000 জন ছাত্র-ছাত্রীর মধ্যে যদি 300 জনকে নির্বাচন করতে হয়, তাহলে 20টি বিভাগের প্রত্যেকটির জন্য বরাদ্দ করা নির্দিষ্ট অনুপাত অনুযায়ী ছাত্র-ছাত্রী নির্বাচন করে এই 300 জনকে নিতে হবে। মনে রাখতে হবে, বিভাগ-পিছু বরাদ্দ করা অনুপাত, সেই বিভাগের ছাত্র-ছাত্রীর মোট সংখ্যার ওপর নির্ভর করছে। বরাদ্দনির্দিষ্ট নমুনা সংগ্রহকে একপ্রকার বিবেচনা-নির্ভর নমুনা সংগ্রহ (Judgement Sampling) হিসাবে ভাবা যেতে পারে। যদিও এই পদ্ধতি সহজ ও খরচ কম, এই পদ্ধতিতে অনেক ধরনের ভুল ও পক্ষপাত ঘটার সম্ভাবনা থাকে। আর্থ-সামাজিক সমীক্ষা, জনমত সমীক্ষা এবং বাজার-বিশ্লেষণ (Market research)-এর ক্ষেত্রে এই নমুনা সংগ্রহ পদ্ধতির বহুল প্রচলন লক্ষ্য করা যায়।

৪৪.৬ সারাংশ

রাশিবিজ্ঞানে পূর্ণক বলতে সমীক্ষার অন্তর্গত এককগুলোর সমষ্টি বোঝায়। সাধারণত পূর্ণকের বৈশিষ্ট্য প্রধানত পূর্ণকাঙ্কের প্রাক্‌কলনের উদ্দেশ্যে সম্পূর্ণ সমীক্ষার বদলে নমুনা সমীক্ষা বেশীরভাগ ক্ষেত্রে করা হয়। এই নমুনা সমীক্ষার নানা সুবিধা আছে সম্পূর্ণ সমীক্ষার তুলনায় যেমন অর্থের সাশ্রয়, সময়ের সাশ্রয়, অধিকতর নির্ভুল, অধিকতর পরিধি এবং অধিকতর প্রয়োগ।

নমুনা সমীক্ষা সংক্রান্ত কতগুলো ধারণার মধ্যে নমুনা একক বলতে সমীক্ষার অন্তর্গত সর্বশেষ ক্ষুদ্রতম একককটিকে বোঝায়। নমুনাজ তালিকা হল পূর্ণকের সমস্ত এককের একটি তালিকা, এবং নমুনাজ ভ্রান্তি বলতে নমুনা সমীক্ষা করার ফলে যে সব ভ্রান্তির সৃষ্টি হয় তাদের বোঝায়।

নমুনা সমীক্ষা সাধারণত বিভিন্ন ধাপ বা পর্যায় অনুযায়ী করা হয়।

নমুনা সমীক্ষার নানা পদ্ধতিগুলোর মধ্যে সরল সমসম্ভব নমুনায় সমসম্ভব সংখ্যাসারির সাহায্যে নমুনা গ্রহণ করা হয়। এই পদ্ধতি দু'রকম— পুনঃস্থাপনাসহ ও পুনঃস্থাপনাবিহীন। পুনঃস্থাপনাবিহীন পদ্ধতি পূর্ণকাঙ্কর সূক্ষ্ম প্রাক্‌কলক দিতে বেশী কার্যকরী। এই পদ্ধতিতে ব্যক্তিগত পক্ষপাত থাকেই না এবং পূর্ণকের এককগুলো অনধীনভাবে সমান সম্ভাবনাসহ নমুনায় অন্তর্গত হতে পারে। তবে এই পদ্ধতিতে নমুনা সংগ্রহের আগে সম্পূর্ণ এবং সমসাময়িক (up to date) নমুনা এককের পূর্ণ তালিকা থাকা প্রয়োজন।

স্তরবিন্যস্ত নমুনা সংগ্রহ বৈসাদৃশ্যমূলক পূর্ণকের ক্ষেত্রে সাধারণত ব্যবহার করা হয়। এখানে পূর্ণককে কয়েকটি স্তরে ভাগ করা হয় যাদের অন্তর্গত এককগুলো যথেষ্ট সদৃশ অথচ তারা নিজেরা পরস্পরে বেশ ভিন্ন। এই নমুনা সংগ্রহ পদ্ধতিতে প্রত্যেক স্তর থেকে কিছু একক আসেই। স্তরের অন্তর্গত নমুনা সংখ্যা ঠিক করার জন্য বেম্যান, বাউলি বা প্রকৃষ্ট বণ্টনের সূত্রের সাহায্য নেওয়া যেতে পারে।

নিয়মানুগ নমুনা সংগ্রহ পদ্ধতি খুবই উপযোগী যদি সমীক্ষার সময় অবধি নমুনা এককের সম্পূর্ণ তালিকা বর্ণানুক্রম বা কালক্রম অনুসারে পাওয়া যায়। এটি একটি মিশ্র সম্ভাবনাশ্রয়ী সংগ্রহ পদ্ধতি। সমসম্ভব সংখ্যাসারির সাহায্যে পাওয়া এক থেকে নমুনাজ অন্তরের মধ্যে কোন সংখ্যা অনুযায়ী প্রথম এককটিকে নিয়ে, বাকী এককগুলো নমুনাজ অন্তর অনুযায়ী নেওয়া হয়। এই সংগ্রহ পদ্ধতিকে মোটেই উপযুক্ত ভাবা হয় না যখন নমুনাজ-অন্তর অনুযায়ী নির্বাচিত এককগুলোর ক্ষেত্রে কোন পর্যাবৃত্ত বৈশিষ্ট্য দেখা যায়।

বহুবিভাগী নমুনা সংগ্রহের ক্ষেত্রে পূর্ণকের বৃহত্তর বিভাগগুলো থেকে সমসম্ভব সংখ্যাসারি দ্বারা নির্বাচিত কতগুলো বিভাগ থেকে কতগুলো ছোট বিভাগ নির্বাচন করা হয়। আবার ছোট বিভাগগুলো থেকে বারবার বিভাগগত নির্বাচনের মাধ্যমে সবশেষে নমুনাজ এককে পৌছান হয়। এখানে একমাত্র নির্বাচিত বিভাগগুলোর অন্তর্গত এককগুলোর তালিকা করলেই চলে। বহুবিভাগী নমুনা সংগ্রহে বিভাগ পিছু একক আলাদা।

বহুপর্যায়ী নমুনা সংগ্রহের ক্ষেত্রে কতগুলো পর্যায়ে নমুনা গ্রহণ করা হয়। প্রথম পর্যায়ে, নমুনা-এককগুলো থেকে সহজ, কম-ব্যয়সাধ্য অথচ গুরুত্বপূর্ণ এমন কিছু তথ্য সংগ্রহ করা হয়। প্রথম পর্যায়ের এককগুলো থেকে দ্বিতীয় পর্যায়ে কিছু একক নির্বাচন করে অন্যান্য তথ্য সংগ্রহ করা হয়। প্রয়োজনবোধে পর্যায় সংখ্যা বাড়ান যেতে পারে। এখানে একই একক নিয়ে বিভিন্ন পর্যায়ে কাজ করা হয়।

গুচ্ছ নমুনা সংগ্রহের ক্ষেত্রে পূর্ণককে কয়েকটা গুচ্ছে বা ব্লকে ভাগ করা হয় এবং গুচ্ছগুলোর মধ্য থেকে কয়েকটা নির্বাচন করে প্রত্যেকের ওপর সম্পূর্ণ সমীক্ষা করা হয়।

বরাদ্দনির্দিষ্ট নমুনা সংগ্রহের ক্ষেত্রে পূর্ণককে কয়েকটা অংশে ভাগ করার পর, অংশপিছু বরাদ্দ করা নির্দিষ্ট অনুপাতের ভিত্তিতে প্রত্যেক অংশ থেকে নমুনা সংগ্রহ করা হয়।

৪৪.৭ অনুশীলনী

সংজ্ঞা নির্দেশ করুন :

১। (ক) পূর্ণক, (খ) নমুনা-একক, (গ) নমুনা-এককের পূর্ণতালিকা, (ঘ) নমুনাজ এবং অনমুনাজ ভ্রান্তি, (ঙ) আরোহী-অনুমিতি।

২। নমুনা সমীক্ষা করা হয় কেন? সম্পূর্ণ সমীক্ষার তুলনায় নমুনা সমীক্ষার সুবিধাসমূহ বর্ণনা করুন।

৩। নমুনা সমীক্ষার বিভিন্ন পর্যায়গুলো সম্বন্ধে বিস্তারিত আলোচনা করুন।

৪। সরল সমসম্ভব নমুনা কাকে বলে? পুনঃস্থাপনাসহ ও পুনঃস্থাপনাবিহীন সরল সমসম্ভব নমুনার প্রভেদ কোথায়?

৫। সমসম্ভব সংখ্যাসারি কাকে বলে? সরল সমসম্ভব নমুনা চয়নের ক্ষেত্রে সমসম্ভব সংখ্যাসারির উদাহরণসহ ব্যাখ্যা কর।

নিম্নলিখিত ক্ষেত্রে সমসম্ভব সংখ্যাসারির সাহায্যে নমুনা চয়ন বর্ণনা করুন :

৬। (ক) কোন পরিবারের 10 জন সদস্য থেকে পুনঃস্থাপনাসহ পদ্ধতিতে 3 জনকে নির্বাচন করতে হবে;

(খ) কলকাতার 52টি কলেজ থেকে 5টিকে পুনঃস্থাপনাবিহীন পদ্ধতিতে নির্বাচন করতে হবে;

(গ) কোন এলাকার চারটে রাস্তায় যথাক্রমে 4টি, 12টি, 10টি এবং 6টি বাড়ী আছে। পুনঃস্থাপনাবিহীন পদ্ধতিতে সেই এলাকা থেকে 5টি বাড়ীকে নির্বাচন করতে হবে।

৭। (ক) পশ্চিমবঙ্গে কারখানার শ্রমিকদের জীবনযাত্রার মান নির্ধারণের জন্য সমীক্ষা পদ্ধতি বর্ণনা করুন।

(খ) পশ্চিমবঙ্গে শিক্ষিতদের (গ্র্যাজুয়েট) মধ্যে বেকারত্বের অবস্থা সমীক্ষা করার জন্য কীভাবে অগ্রসর হওয়া যায়, তা বর্ণনা করুন।

৮। স্তরবিন্যস্ত নমুনা সংগ্রহ পদ্ধতিকে সরল সমসম্ভব নমুনা সংগ্রহ পদ্ধতির তুলনায় কেন শ্রেয় মনে করা হয়? স্তরের অন্তর্গত নমুনা-সংখ্যা নির্ধারণপদ্ধতি উল্লেখ করে, স্তরবিন্যস্ত নমুনা সংগ্রহ পদ্ধতিটি ব্যাখ্যা করুন।

৯। বহুবিভাগী ও বহুপর্যায়ী নমুনা সংগ্রহ পদ্ধতি দুটোর সম্বন্ধে আলোচনা করুন।

সংক্ষিপ্ত টীকা লিখুন ঃ

১০। (ক) গুচ্ছ নমুনা সংগ্রহ;

(খ) বরাদ্দনির্দিষ্ট নমুনা সংগ্রহ।

১১। বহুবিভাগী ও বহুপর্যায়ী এবং স্তরবিন্যস্ত সমসম্ভব নমুনা সংগ্রহ ও বরাদ্দনির্দিষ্ট নমুনা সংগ্রহ পদ্ধতিগুলোর মধ্যে মৌলিক পার্থক্য কী?

নিম্নলিখিত মন্তব্যগুলো বিচার করে সঠিক না ভুল নিরূপণ করুন ঃ

১২। (ক) সম্পূর্ণ সমীক্ষায় প্রাপ্ত উপাত্ত বা রাশিতথ্য (data) স্বাভাবিকভাবে ভ্রান্তিশূন্য;

(খ) সমসম্ভব নমুনা নেওয়া মানে পূর্ণক থেকে এলোমেলো কিছু একক নেওয়া।

(গ) কয়েকটি বিশেষ ক্ষেত্রে সম্পূর্ণ সমীক্ষা করা সম্ভব নয়—নমুনা সমীক্ষা যেখানে অপরিহার্য।

(ঘ) 10টি একক থেকে 2টি একককে পুনঃস্থাপনাবিহীন সমসম্ভব পদ্ধতিতে নেওয়া হলে (যেখানে নির্বাচন হওয়ার ক্রমকে গুরুত্ব দেওয়া হয়) সর্বমোট 90টি নমুনা পাওয়া সম্ভব।

(ঙ) নিয়মানুগ নমুনা সংগ্রহ একটি সম্ভাবনাশ্রয়ী পদ্ধতি।

যথাযথ এবং সংক্ষিপ্ত উত্তর দিন :

১৩। (ক) কী কী অসুবিধার জন্য সম্পূর্ণ সমীক্ষা করা যায় না?

(খ) কোন পরিস্থিতিতে সম্পূর্ণ সমীক্ষা করা উচিত?

(গ) রচয়িতাসহ যে কোন দুইটি সমসম্ভব সংখ্যাসারির নাম উল্লেখ করুন।

(ঘ) 10টি একক থেকে 2টি একককে পুনঃস্থাপনাসহ ও পুনঃস্থাপনাবিহীন পদ্ধতিতে নির্বাচন করলে মোট সম্ভাব্য নমুনা সংখ্যা কত?

(ঙ) স্তরবিন্যস্ত নমুনা সংগ্রহ কালে কোন্ ভিত্তিতে পূর্ণককে স্তরে ভাগ করা হয়?

(চ) বহুবিভাগী নমুনা সংগ্রহের একটি উদাহরণ দিন।

(ছ) সরল সমসম্ভব নমুনার তুলনায় নিয়মানুগ নমুনার সুবিধাগুলো কী?

(জ) নিয়মানুগ নমুনা সংগ্রহ পদ্ধতিতে কখন নমুনা নেওয়া উচিত নয়?

(ঝ) গুচ্ছ নমুনা সংগ্রহ ব্যবহার করা যায় এমন কোন পরিস্থিতি উল্লেখ করুন।

(ঞ) গুচ্ছ নমুনা সংগ্রহে নমুনা-একক হিসাবে কাকে নেওয়া হয়?

৪৪.৮ গ্রন্থপঞ্জী

১। দাশগুপ্ত, গুহঠাকুরতা, অধিকারী—*রাশিবিজ্ঞানের প্রয়োগ পদ্ধতি* (পশ্চিমবঙ্গ রাজ্য পুস্তক পর্ষৎ), ১৯৭৬।

২। সেন, রাজকুমার—*সংখ্যাতত্ত্ব* (পশ্চিমবঙ্গ রাজ্য পুস্তক পর্ষৎ)।

৩। দাশগুপ্ত, চৌধুরী, দাস—*রাশিবিজ্ঞানের পরিভাষা* (পশ্চিমবঙ্গ রাজ্য পুস্তক পর্ষৎ)।.

৪। Goon, Gupta & Dasgupta—*Fundamentals of Statistics–Vol. II*, World Press 2001.

৫। Cochran—*Sampling Techniques*–Wiley Eastern.

৬। Murthy—*Statistical Theory and Methods*–Statistical Publishing Society.

একক ৪৫ ☐ নমুনাজ বিভাজন (Sampling Distribution)

গঠন

৪৫.০ উদ্দেশ্য

এই অনুচ্ছেদে নমুনাজ বিভাজনের মূল বিষয়গুলোর সঙ্গে পরিচিত করা হবে। নিম্নে লিখিত বিষয়গুলি এই এককে আলোচিত হবে—

১। একক, পূর্ণক ও পূর্ণকাঙ্ক-র সংজ্ঞা এবং নমুনাঙ্ক ও তার প্রয়োজন।

২। নমুনাঙ্কের বিভাজন—নমুনাজ বিভাজন।

৩। নমুনাঙ্ক বিভাজনের গড় (প্রত্যাশা) ও সমক ভ্রান্তির ধারণা।

৪৫.১ প্রস্তাবনা

ইতিমধ্যে আমরা রাশিবিজ্ঞানভিত্তিক তথ্য ও তার বিভিন্ন প্রয়োজনীয়তার কথা জানি। এও জানি কীভাবে এই তথ্য থেকে সুচারু বিশ্লেষণের মধ্য দিয়ে আরও জ্ঞানলাভ করা যায়। এই তথ্য বিশ্লেষণের প্রয়োজন প্রায় সব ক্ষেত্রেই রয়ে গেছে। বিজ্ঞানের বিশেষ শাখায়—ভৌতবিজ্ঞান, রসায়ন বিজ্ঞান, প্রাণী বিজ্ঞান, উদ্ভিদবিদ্যা, অর্থবিদ্যা ইত্যাদিতে এর প্রয়োজন অপরিসীম। অন্যদিকে কৃষি, শিল্প, ব্যবসা, যাতায়াত ব্যবস্থা প্রভৃতি ক্ষেত্রেও রাশিবিজ্ঞানের প্রয়োগ অনস্বীকার্য। তাই জীবনের এমন কোন দিক এখন আর বাদ রাখা যাবে না যেখানে রাশিতথ্য বিশ্লেষণের প্রয়োজন নেই। এই বিশ্লেষণ শুধু যে সংগৃহীত তথ্যের অন্তর্নিহিত অংশই আমাদের কাছে খুলে দেয় তাই নয়, ভবিষ্যতে তথ্য সংগ্রহ থেকে শুরু করে এর পূর্ণাঙ্গ বিশ্লেষণ কী কী ভাবে পরিবর্তন বা পরিমার্জন করা যায় যাতে আরো ভালোভাবে আলোচ্য পূর্ণক সম্বন্ধে জানতে পারি ও তার ব্যাপারে উপযুক্ত সিদ্ধান্তে পৌঁছতে পারি তাও জানতে পারি। (এখানে বলে রাখা যায় যে কোন সমীক্ষার অন্তর্ভূত ব্যষ্টিসমূহের সমষ্টিকেই আমরা পূর্ণক বলব।) সরকারি সিদ্ধান্তের জন্যও রাশিবিজ্ঞানভিত্তিক তথ্য বিশ্লেষণের আয়োজন ও প্রয়োজন দুইই আছে। এমন সব ক্ষেত্রেই সমগ্র সমষ্টিটার অবেক্ষণ সম্ভব নাও হতে পারে, বেশীরভাগ ক্ষেত্রে অসম্ভব। তাই উপযুক্তভাবে সংগৃহীত নমুনার ওপর ভিত্তি করেই সকল বিশ্লেষণ ও তার প্রয়োগ করতে হয়। এই নমুনা যেমন হবে, তাতে সমষ্টির যে যে অংশ প্রতিফলিত হবে এবং নমুনাভিত্তিক যে ফরমুলা আমরা ব্যবহার করব সবকিছুই সম্ভবনাতত্ত্ব অনুসরণ করে। সুতরাং ফরমুলা সম্বন্ধে এটুকু এখনই বলা যায় যে কোনভাবেই আমরা বিশ্লেষণের ফল সম্বন্ধে নিশ্চিত হতে পারি না। তবুও তা কেমন হতে পারে বা কেমন হতে পারে না তার সম্ভবনাতত্ত্বভিত্তিক বিচার সম্ভব।

৪৫.২ ধারণা ও সংজ্ঞা

আলোচনার সুবিধার্থে আমরা কতকগুলো ব্যবহৃত শব্দের সংজ্ঞা নির্দেশ করব।

উদাহরণ ১. ঃ একটা উদাহরণ নিয়ে শুরু করা যাক। একটা শ্রেণীতে (যেমন নেতাজি সুভাষ মুক্ত বিশ্ববিদ্যালয়ের কোন এক বছরে স্নাতক স্তরের প্রথম বর্ষ) যত ছাত্র আছে সবাই মিলে একটা সমষ্টি বা

পূর্ণক তৈরী করে। প্রত্যেক ছাত্র এক একটা একক। আমরা এই শ্রেণীর সব ছাত্রের গড় উচ্চতা (সেমি) জানতে চাই। এই গড় উচ্চতা গোটা শ্রেণীটার জন্য প্রযোজ্য, কিন্তু প্রত্যেক ছাত্রের জন্য নয়। তবে এটা মাপতে গেলে প্রত্যেক ছাত্রের উচ্চতা (সেমি) মাপতে হবে।

উদাহরণ ২. : প্রথম উদাহরণ অনুসারে আমরা ঐ শ্রেণীর ছাত্রদের পারিবারিক সদস্যপিছু গড় মাসিক আয় (টাকায়) জানতে চাই। একই সমষ্টি ও একই একক রইল। কিন্তু প্রত্যেক ছাত্রের জন্য এখন দুটো চল একই সঙ্গে জানতে হবে। এক, তার পরিবারে সদস্য সংখ্যা। দুই, পারিবারিক মাসিক মোট আয়।

মন্তব্য : উদাহরণ দুটোতে (1 ও 2) একই পূর্ণক আছে। কিন্তু পূর্ণকের পৃথক দুটো বৈশিষ্ট্য দু'জায়গায় আলোচিত। এগুলোই পূর্ণকাঙ্ক—প্রতিটি মান একই সঙ্গে গাণিতিকভাবে সংযুক্ত করলে তবেই এর মান পাওয়া যাবে।

৪৫.২.১ একক

অবেক্ষণ করা যায় এমন এক বা একাধিক সত্তাকে আমরা **মৌলিক একক** বা **একক** বলি। এককগুলো অবশ্যই সুনির্দিষ্ট এবং পরস্পর বিচ্ছিন্ন হবে। এর ফলে একের থেকে অপর একককে আলাদা করে চেনা যাবে ও পৃথকভাবেই প্রত্যেকের অবেক্ষণ সম্ভব হবে। তবে অনুসন্ধানের প্রয়োজন অনুসারেই এককের মনোনয়ন করা হয়।

উদাহরণ 3 (ক) একটা পরিবার (খ) একটা বাসগৃহ (গ) একটা গ্রাম বা শহর (ঘ) একটা গাছ ইত্যাদি যে কোনটাই একক হতে পারে। একক সমীক্ষার উদ্দেশ্যের উপর নির্ভরশীল।

৪৫.২.২ পূর্ণক

সুনির্দিষ্ট কতকগুলো এককের সমষ্টিকে পূর্ণক বলা হয়। আবার আমাদের লক্ষ্যের উপাদানসমগ্রকেও পূর্ণক বলা হয়।

পূর্ণককে সুনির্দিষ্ট করার জন্য দেশ, কাল ইত্যাদির উল্লেখ করা হয়। ভৌগলিক সীমাও এই প্রয়োজনে নেওয়া যায়। একটা পূর্ণকে যতগুলো একক থাকে তাকে **পূর্ণক সংখ্যা** বা **পূর্ণকের আয়তন** বলে। পূর্ণকসংখ্যা অসীম হলে পূর্ণকটাকে সসীম বলা হবে। পূর্ণকসংখ্যা অসীম হ'লে পূর্ণকটাকে অসীম বলা হবে।

কোন কোন সময় সম্পূর্ণভাবে একটা পূর্ণক আমাদের লক্ষ্য হয় না। তখন, তার যে অংশ আমরা লক্ষ্য হিসাবে নেব তাকে **লক্ষ্য-ক্ষেত্র** (domain of study) বা **লক্ষ্য-পূর্ণক** (study population) বলব। শহর এলাকার শ্রমিকদের জীবিকা নির্বাহণ ব্যয়ের সূচকসংখ্যা নির্ণয় করতে গেলে শহর অঞ্চল ও সেখানের শ্রমিক শ্রেণীকেই কেবল পূর্ণক হিসাবে নিতে হবে; দেশের সমগ্র শ্রমিকশ্রেণী বা সমগ্র এলাকা আমাদের আলোচনায় আসবে না।

৪৫.২.৩ পূর্ণকাঙ্ক

পূর্ণকের সমস্ত এককের জন্য নেওয়া কোন চলের মানগুলোর গড়, সমক বিচ্যুতি ইত্যাদির মত অপেক্ষককে বা পরিমাপকে **পূর্ণকাঙ্ক** বলা হয়।

একটা সসীম পূর্ণক নেওয়া যাক যাতে N সংখ্যক একক আছে এবং যাদের আমরা u_1, u_2,.....u_n বলে নির্দিষ্ট করতে পারি। u_i-এর ক্ষেত্রে চল y-এর মান Y_i দিয়ে চিহ্নিত করা যাক, $i = 1, 2, \ldots, N$। এখন

$\overline{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$: y-এর পূর্ণক গড়,

$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N} (Y_i - \overline{Y})^2$: y- এর পূর্ণক সমক বিচ্যুতি, ইত্যাদি এক একটা পূর্ণকাঙ্ক

আবার আমরা যে বাইনোমিয়াল, পোয়ার্স নর্ম্যাল ইত্যাদি তত্ত্বগত নিবেশন জানি তার প্রত্যেকটারই এক বা একাধিক এমন ধ্রুবক থাকে যার মান নির্দিষ্ট করলে সম্পূর্ণ নিবেশনটা জানা হয়ে যায়। এমন ধ্রুবককেও আমরা পূর্ণকাঙ্ক বলি।

৪৫.২.৪ নমুনাঙ্ক

নমুনার ওপর ভিত্তি করে যে পরিমাপ করা হয় যেমন নমুনাজ গড়, নমুনাজ সমক বিচ্যুতি ইত্যাদি তাকে **নমুনাঙ্ক** বলে।

যদি y_1, y_2,......, y_n n সংখ্যক নমুনা হয় তবে $\overline{y} = \frac{1}{n}\sum_{j=1}^{n} y_j$

$$s^2 = \frac{1}{n}\sum_{j=1}^{n} (y_j - \overline{y})^2$$

৪৫.২.৫ নমুনাজ বিভাজন

একটা পূর্ণক থেকে নমুনা নেওয়া হয় প্রধানত পূর্ণকটার কোন পূর্ণকাঙ্কের মানের ধারণা করতে বা কোন প্রকল্প বিচার করতে। এখন, এমন একটা নমুনার ওপর নির্ভর করে আমাদের একটা না একটা নমুনাঙ্ক নিয়েই কাজ করতে হয়, রাশিবিজ্ঞানের অনুমানতত্ত্বের প্রয়োজনে। আবার একই পূর্ণক থেকে একাধিক নমুনা সংগ্রহ করা সম্ভব। একটা নমুনাতে পূর্ণকটার যে যে সদস্য থাকবে, অন্য একটা নমুনাতে সেগুলোই যে থাকবে তার কোন কথা নেই—তারা প্রথম নমুনার সদস্যগুলো থেকে আলাদাও হতে পারে। আবার যে চলটা (y) আমাদের আলোচনার বিষয় তার মানও এক সদস্য থেকে অপরটায় আলাদা হতে পারে। তাই একটা নমুনার সদস্যগুলোর যে y-মান থাকবে অপরটায় সেগুলো অন্যরকম হতেই পারে। তা'হলে নমুনাঙ্কটার মান কেমন

হতে পারে? অবশ্যই একটা নমুনার জন্য যে মান হবে, অন্যটার জন্য তা আলাদা হতে পারে। এই যে একই নমুনাঙ্কের মান বিভিন্ন নমুনার জন্য আলাদা হতে পারে একে নমুনাজ চাঞ্চল্য (sampling fluctuation) বলে। এই আলোচনায় আমরা একটা নমুনাতে যে সংখ্যক সদস্য পূর্ণকটা থেকে নেওয়া হচ্ছে তা স্থির (একই) রাখছি, আর একে নমুনাসংখ্যা n বলছি।

যদি বেশ কয়েকটা এমন নমুনা নেওয়া হয় (অবশ্যই একই পূর্ণক থেকে) যাতে একই সংখ্যক (n) সদস্য আছে তবে নমুনাঙ্কটার এক সারি মান পাওয়া যাবে। এই মানগুলোর একটা পরিসংখ্যা বিভাজন করা যাবে। আরো এমন নমুনা নিতে থাকলে আর ঐ নমুনাঙ্কটার মান নির্ণয় করতে থাকলে পরিসংখ্যা বিভাজনটা পরিবর্তন হতে থাকবে। এইভাবে যখন অসীম সংখ্যক নমুনা নেওয়া হবে, যেগুলোতে n একই থাকবে এই পরিসংখ্যা বিভাজনটা যে সীমায় পৌছাবে তাকে আমরা নমুনাঙ্কটার নমুনাজ বিভাজন বলব।

কোন তত্ত্বগত নিবেশন (theoretical distribution) থেকে সম্ভাবনাশ্রয়ী নমুনা (Random Sample) নিলেও একটা নমুনাঙ্কের নমুনাজ বিভাজন কেমন হবে তা তত্ত্বগতভাবেই নির্ণয় করা যায়। এক্ষেত্রে সম্ভাবনাতত্ত্বের নিরিখে পূর্ণকটার তত্ত্বগত প্রকৃতি অবশ্যই জানা থাকতে হবে।

৪৫.২.৬ নমুনাঙ্কের প্রত্যাশা ও সমক ভ্রান্তি

এখন আমরা একটা নমুনাঙ্কের নমুনাজ বিভাজনকে তার সম্ভাবনা নিবেশন (probability distribution) বলতে পারি। তাই নমুনাজ বিভাজনের ও গড়, মধ্যমা, প্রমাণ বিচ্যুতি ইত্যাদি এবং সাধারণভাবে বিভিন্ন পরিঘাত থাকতে পারে। এগুলোর মধ্যে নমুনাজ বিভাজনের গড় ও প্রমাণ বিচ্যুতি আমরা খুবই কাজে লাগাই। এই গড়কে বলা হয় **নমুনাঙ্কের প্রত্যাশা**, আর প্রমাণ বিচ্যুতিকে নমুনাঙ্কের প্রমাণ ভ্রান্তি বা সমক ভ্রান্তি (standard error) বলা হয়। তাহলে T যদি একটা নমুনাঙ্ক হয়, এর প্রত্যাশা হবে E(T) ও সমকভ্রান্তি s.e.(T) বলা হবে। (s.e. লেখা হয় সমক ভ্রান্তির ইংরাজি standard error-এর সংক্ষেপ হিসাবে)

প্রথম যে সদস্যকে পূর্ণক থেকে সম্ভবনাশ্রয়ী নমুনা সংগ্রহে নেওয়া হল তার y-মানকে y_1 বলা যাক। এভাবে দ্বিতীয় সদস্য-র y-মানকে y_2 ইত্যাদি। n তম যে সদস্যকে নেওয়া হল তার y-মানকে y_n বলা যাক। তার ফলে $y_1, y_2......,y_n$ প্রত্যেকটাই এক একটা সম্ভবনাশ্রয়ী চল। একটা নমুনাঙ্ক T তাই nটা সম্ভবনাশ্রয়ী চল $y_1, y_2......,y_n$-এর একটা অপেক্ষক। $(y_1, y_2......,y_n)$ হল একটা সম্ভবনাশ্রয়ী নমুনা। চল $y_1, y_2......,y_n$ -এর বাস্তবিক মান একটা নমুনাতে নেওয়া হ'লে তাকে আমরা $(y_1, y_2......,y_n)$-এর একটা প্রকৃতমান (realization) বলব।

৪৫.৩ নমুনাজ গড়ের বিভাজন : একটি উদাহরণ

পাঁচজন ছাত্রের সমষ্টিকে একটা পূর্ণক হিসাবে নেওয়া হ'ল। এখানে N = 5। এই ছাত্রদের ক্রমিক সংখ্যা S_1, S_2, S_3, S_4 ও S_5 দিয়ে দেওয়া হল। আলোচ্য চল (y) হিসাবে আমরা এদের বয়স নিলাম। এদের বয়স যথাক্রমে 7, 5, 4, 6 এবং 3 বছর। নমুনা-সংখ্যা n = 2 নেওয়া যাক।

SRSWR-এর ক্ষেত্রে

সমস্ত সম্ভাব্য নমুনাগুলো (যার সমষ্টি হল নমুনা দেশ : Sample Space)

এইরকম : $(S_1, S_1), (S_1, S_2), (S_1, S_3), (S_1, S_4), (S_1, S_5)$

$(S_2, S_1), (S_2, S_2), (S_2, S_3), (S_2, S_4), (S_2, S_5)$

$(S_3, S_1), (S_3, S_2), (S_3, S_3), (S_3, S_4), (S_3, S_5)$

$(S_4, S_1), (S_4, S_2), (S_4, S_3), (S_4, S_4), (S_4, S_5)$

$(S_5, S_1), (S_5, S_2), (S_5, S_3), (S_5, S_4), (S_5, S_5)$,

মোট $N^n = 5^2 = 25$টা সম্ভাব্য নমুনা এখানে আছে।

প্রত্যেক-এর বাস্তবায়িত হওয়ার সম্ভবনা $\frac{1}{N^n} = \frac{1}{25}$

আমরা নমুনাঙ্ক হিসাবে $\bar{y}$, নমুনাজ গড় নিচ্ছি। দেখা যাক $\bar{y}$-এর মান কেমন পরিবর্তন হয় আর তাদের সম্ভাবনাই বা কেমন।

ক্রমিক সংখ্যা	নমুনা	সম্ভাবনা	নমুনা মান	$\bar{y}$-এর নমুনা মান
1	(S_1, S_1)	$\frac{1}{25}$	(7, 7)	7.0
2	(S_1, S_2)	$\frac{1}{25}$	(7, 5)	6.0
3	(S_1, S_3)	$\frac{1}{25}$	(7, 4)	5.5
4	(S_1, S_4)	$\frac{1}{25}$	(7, 6)	6.5
5	(S_1, S_5)	$\frac{1}{25}$	(7, 3)	5.0
6	(S_2, S_1)	$\frac{1}{25}$	(5, 7)	6.0
7	(S_2, S_2)	$\frac{1}{25}$	(5, 5)	5.0
8	(S_2, S_3)	$\frac{1}{25}$	(5, 4)	4.5
9	(S_2, S_4)	$\frac{1}{25}$	(5, 6)	5.5
10	(S_2, S_5)	$\frac{1}{25}$	(5, 3)	4.0
11	(S_3, S_1)	$\frac{1}{25}$	(4, 7)	5.5

ক্রমিক সংখ্যা	নমুনা	সম্ভাবনা	নমুনা মান	$\bar{y}$-এর নমুনা মান
12	(S_3, S_2)	$\frac{1}{25}$	(4, 5)	4.5
13	(S_3, S_3)	$\frac{1}{25}$	(4, 4)	4.0
14	(S_3, S_4)	$\frac{1}{25}$	(4, 6)	5.0
15	(S_3, S_5)	$\frac{1}{25}$	(4, 3)	3.5
16	(S_4, S_1)	$\frac{1}{25}$	(6, 7)	6.5
17	(S_4, S_2)	$\frac{1}{25}$	(6, 5)	5.5
18	(S_4, S_3)	$\frac{1}{25}$	(6, 4)	5.0
19	(S_4, S_4)	$\frac{1}{25}$	(6, 6)	6.0
20	(S_4, S_5)	$\frac{1}{25}$	(6, 3)	4.5
21	(S_5, S_1)	$\frac{1}{25}$	(3, 7)	5.0
22	(S_5, S_2)	$\frac{1}{25}$	(3, 5)	4.0
23	(S_5, S_3)	$\frac{1}{25}$	(3, 4)	3.5
24	(S_5, S_4)	$\frac{1}{25}$	(3, 6)	4.5
25	(S_5, S_5)	$\frac{1}{25}$	(3, 3)	3.0
মোট	—	1	—	—

$\bar{y}$-এর একই মান একাধিক বার আছে। তাই এই মানগুলোর পরিসংখ্যা বিভাজন করে $\bar{y}$-এর নমুনাজ বিভাজন পাওয়া যাবে।

$\bar{y}$-এর মান	ট্যালি চিহ্ন	পরিসংখ্যা
3.0	/	1
3.5	//	2
4.0	///	3

$\bar{y}$-এর মান	ট্যালি চিহ্ন	পরিসংখ্যা
4.5	////	4
5.0	~~////~~	5
5.5	////	4
6.0	///	3
6.5	//	2
7.0	/	1
মোট	—	25

সুতরাং y-এর নমুনাজ্ বিভাজন এমন দেখাবে—

$\bar{y}$-এর মান	সম্ভাবনা
3.0	$\frac{1}{25}$
3.5	$\frac{2}{25}$
4.0	$\frac{3}{25}$
4.5	$\frac{4}{25}$
5.0	$\frac{5}{25}$
5.5	$\frac{4}{25}$
6.0	$\frac{3}{25}$
6.5	$\frac{2}{25}$
7.0	$\frac{1}{25}$
মোট	1

সংকেত এক্ষেত্রে মাত্র $^5C_2 = 10$টি নমুনা সম্ভব :

(S_1, S_2), $(S_1; S_3)$, (S_1, S_4), (S_1, S_5), (S_2, S_3), (S_2, S_4), (S_2, S_5), (S_3, S_4), (S_3, S_5), (S_4, S_5)। প্রত্যেকটি নমুনার y-র মান দুটির গড় নির্ণয় করে, $\bar{y}$-এর বিভিন্ন মান ও তাদের সম্ভাবনা নির্ণয় করলেই নমুনাজ বিভাজন পাওয়া যাবে।

৪৫.৪ নমুনাজ গড়ের প্রত্যাশা ও সমক ভ্রান্তি

প্রথমতঃ আমরা একটা সসীম পূর্ণক নিই যাতে N সংখ্যক সদস্য আছে। y হল আমাদের আলোচ্য চল। সমস্ত সদস্যের y-মানগুলো

$Y_1, Y_2, \ldots\ldots\ldots, Y_N$ [পূর্ণক মান]

বলে চিহ্নিত করা যাক। এখানে $i = 1, 2, \ldots\ldots, N$

$Y_i = i$ তম সদস্যের y-মান $i = 1, 2, \ldots\ldots, N$

তাহলে পূর্ণকটার (y-এর সাপেক্ষে) গড় (μ) ও সমক বিচ্যুতি (σ^2) হল—

$$\left.\begin{aligned}\mu &= \frac{1}{N}\sum_{i=1}^{N} Y_i \\ \sigma^2 &= \frac{1}{N}\sum_{i=1}^{N}(Y_i-\mu)^2\end{aligned}\right] \text{ এই দুটো পূর্ণকাঙ্ক}$$

এই পূর্ণকটা থেকে সমসম্ভব নমুনা নেওয়া হল যাতে n সংখ্যক সদস্য থাকে। এদের আমরা $y_1, y_2, \ldots\ldots y_n$ [নমুনাজ মান] বলে চিহ্নিত করি। এখানে

y_j = সমসম্ভবভাবে নেওয়া j তম সদস্যের জন্য y-এর মান $j = 1, 2, \ldots\ldots, n$

তাহলে নমুনাজ গড় $\bar{y}$ হবে

$$\bar{y} = \frac{1}{n}\sum_{j=1}^{n} y_j$$

এক্ষেত্রে $\bar{y}$ ই আমাদের আলোচ্য নমুনাঙ্ক।

আমরা এখন $E(\bar{y})$ ও s.e. $(\bar{y})$ নির্ণয় করব।

প্রথমত : $E(\bar{y})$:

$$E(\bar{y}) = E\left(\frac{1}{n}\sum_{j=1}^{n} y_j\right)$$

$$= \frac{1}{n}\sum_{j=1}^{n} E(y_j), \text{ প্রত্যাশার যোগের নিয়ম কাজে লাগিয়ে } \ldots\ldots\ldots\ldots\ldots\text{(a)}$$

এখন, নমুনা সমীক্ষার আলোচনা থেকে আমরা y_j-র সম্ভাবনা বিভাজন পাই যা এমন হবে—

$$P(y_j = Y_i) = \frac{1}{N} \quad i = 1(1)N$$

$[\, j = 1, 2, \ldots\ldots\ldots n]$

যাকে তালিকাতে লিখলে এমন দেখাবে—

y_j-র মান	সম্ভাবনা
Y_1	$\frac{1}{N}$
Y_2	$\frac{1}{N}$
$\vdots$	$\vdots$
Y_N	$\frac{1}{N}$
সমষ্টি	1

[SRSWR ও SRSWOR উভয় ক্ষেত্রেই]

তাহলে,

$$E(y_j) = \sum_{i=1}^{N} y_i \, P(y_j = Y_i)$$

$$= \sum_{i=1}^{N} Y_i \cdot \frac{1}{N}$$

$$= \mu \quad j = 1(l)\, n \ ....(b)$$

(কারণ এই ফলটা j-র ওপর নির্ভর করে না)

(b) এর ফল (a)-তে বসিয়ে পাওয়া যায় যে

$$E(\overline{y}) = \frac{1}{n}\sum_{i=1}^{n} \mu = \frac{1}{n} n\mu = \mu$$

$$\boxed{E(\overline{y}) = \mu} \ ....(*)$$

দ্বিতীয়তঃ Var$(\overline{y})$

$$\text{Var}(\overline{y})$$

$$= \text{var}\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)$$

$$= \frac{1}{n^2} \text{Var}\left(\sum_{j=1}^{n} y_j\right)$$

$$= \frac{1}{n^2}\left[\sum_{j=1}^{n} \text{Var}(y_j) + \sum_{j=1, j\neq j'}^{n} \sum_{j'=1}^{n} \text{Cov}(y_j, y_{j'})\right] ....(c)$$

(c) -এর প্রথম পদে $Var(y_j)$ নির্ণয় করা যায় y_j-র সম্ভাবনা বিভাজন থেকে—

$Var(y_j)$

$= E\{y_j - E(y_j)\}^2$

$= E(y_j - \mu)^2$, (b)-এর ব্যবহারে পাওয়া যায়

$= \sum_{i=1}^{N} (Y_i - \mu)^2 . P(y_j = Y_i)$, প্রত্যাশার সংজ্ঞা অনুসারে

$= \sum_{i=1}^{N} (Y_i - \mu)^2 - \frac{1}{N}$

$= \sigma^2 \;\forall j = 1(1)n$..............(d)

[কারণ এই ফলটা j-র ওপর নির্ভর করে না]

আবার (c)-র দ্বিতীয় পদের জন্য $y_j, y_{j'}$ এর যৌথ বিভাজন প্রয়োজন।

$P(y_j = Y_i, y_{j'} = Y_{i'})$, যেখানে $j \neq j'$

$= P(y_j = Y_i). P(y_{j'} = Y_{i'} \mid y_j = Y_i)$, সম্ভবনার গুণের নিয়ম অনুসারে

$= \frac{1}{N}.\frac{1}{N}$, SRSWR-এর ক্ষেত্রে $\forall\; i, i'$ ....(e)

এবং

$P(y_j = Y_i, y_{j'} = Y_{i'}), j \neq j'$

$= \frac{1}{N}.\frac{1}{N-1}$, SRSWOR-এর ক্ষেত্রে এবং $i \neq i'$

O, SRSWOR-এর ক্ষেত্রে যখন $i = i'$ ....(f)

সুতরাং $Cov(y_j, y_{j'})$ SRSWR ও SRSWOR-এর জন্য আলাদাভাবে নির্ণয় করতে হবে।

এখন $Cov(y_j, y_{j'})$

$= E\{y_j - E(y_j)\}\{y_{j'}, - E(y_{j'})\}$

$= E(y_j - \mu)(y_{j'} - \mu)$, (b) অনুসারে

$= \sum_{i=1}^{N} \sum_{i'=1}^{N} (Y_i - \mu)(Y_{i'} - \mu). P(y_j = Y_i, y_{j'} = Y_{i'})$

$\therefore Cov_{SRSWR}(y_j, y_{j'}), j \neq j'$

$$= \sum_{i=1}^{N} \sum_{i'=1}^{N} (Y_i - \mu)\ (Y_{i'} - \mu)\ \frac{1}{N^2},$$ (e) অনুসারে

$$= \frac{1}{N^2} \sum_{i=1}^{N} \left\{ (Y_i - \mu) \sum_{i'=1}^{N} (Y_{i'} - \mu) \right\}$$

$$= \frac{1}{N^2} \sum_{i=1}^{N} [(Y_i - \mu).0]$$

$$= 0 \ ....(g)$$

তাই, (d) ও (g)-কে (c)-তে ব্যবহার করে পাই—

$$\mathrm{Var}_{SRSWR}\ (\overline{y}) = \frac{1}{n^2}\left[\sum_{i=1}^{n} \sigma^2 + 0\right]$$

$$= \frac{1}{n^2}.n\sigma^2$$

$$= \frac{\sigma^2}{n} ....(h)$$

$$\Rightarrow s.\ e._{SRSWR}\ (\overline{y}) = \frac{\sigma}{\sqrt{n}} ....(i)$$

SRSWOR-এর ক্ষেত্রে যখন $j \neq j'$

$$\mathrm{Cov}_{SRSWOR}\ (Y_j, Y_{j'})$$

$$= \sum_{i=1}^{N} \sum_{\substack{i'=1 \\ i \neq i'}}^{N} (Y_i - \mu)(Y_{i'} - \mu).\frac{1}{N(N-1)} + 0$$

$$= \frac{1}{N(N-1)} \sum_{i=1}^{N} \left[(Y_i - \mu) \sum_{\substack{i'=1 \\ i' \neq i}}^{N} (Y_{i'} - \mu) \right]$$

$$= \frac{1}{N(N-1)} \sum_{i=1}^{N} \left[(Y_i - \mu) \left\{ \sum_{i'=1}^{N} (Y_{i'} - \mu) - (Y_i - \mu) \right\} \right]$$

$$= \frac{1}{N(N-1)} \sum_{i=1}^{N} [(Y_i - \mu)\{0 - (Y_i - \mu)\}]$$

$$= -\frac{1}{N(N-1)} \sum_{i=1}^{N} (Y_i - \mu)^2$$

$$= -\frac{\sigma^2}{N-1} ....(j)$$

এখন (d) ও (j)-কে (c)-তে ব্যবহার করে পাই

$\text{Var}_{\text{SRSWOR}}(\bar{y})$

$$= \frac{1}{n^2}\left[\sum_{i=1}^{n} \sigma^2 + \sum_{i=1}^{n}\sum_{\substack{j'=1 \\ j\neq j'}}^{n}\left(-\frac{\sigma^2}{N-1}\right)\right]$$

$$= \frac{1}{n^2}\left[n\sigma^2 - n(n-1).\frac{\sigma^2}{N-1}\right]$$

$$= n.\sigma^2.\frac{1}{n^2}\left[1 - \frac{n-1}{N-1}\right]$$

$$= \frac{\sigma^2}{n}.\frac{N-n}{N-1} \text{..............................} (k)$$

$$\Rightarrow \boxed{\text{s.e.}_{\text{SRSWOR}}(\bar{y}) = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}} \text{.................} (l)$$

টীকা 1 : n → N

যদি নমুনার সদস্য সংখ্যা n থেকে বাড়ানো হয় তবে আমরা সাধারণভাবে এটাই মেনে নেব যে পূর্ণকটা সম্বন্ধে আরো বেশী তথ্য পাওয়া যাবে। তার ফলে যখন $n = N$ হবে সমস্ত পূর্ণকটার তথ্যই এসে যাবে আর নমুনাঙ্কটা (এক্ষেত্রে যেমন অনুরূপ পূর্ণকাঙ্কটার (এক্ষেত্রে যেমন μ) সমান হয়ে যাবে। তখন নমুনাজ চাঞ্চল্যের প্রশ্নই উঠবে না; s.e. এর মান শূন্য হতে হবে।

টীকা 1 থেকে এই বলা যায় যে সসীম পূর্ণক-এর থেকে নমুনা সংগ্রহের ক্ষেত্রে SRSWOR-কে SRSWR-এর তুলনায় বেশী পছন্দ করব অন্তত s.e.$(\bar{y})$ এর সাপেক্ষে।

টীকা 2 : n → ∞

পূর্ণকটার সদস্য সংখ্যা যদি বাড়তে থাকে তাহলে $\text{s.e.}_{\text{SRSWR}}(\bar{y})$-এর পরিবর্তন হবে না। আবার

$$\lim_{N\to\infty} \text{s.e.}_{\text{SRSWOR}}(\bar{y})$$

$$= \lim_{N\to\infty} \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

$$= \frac{\sigma}{\sqrt{n}}$$

$$\because \lim_{N\to\infty}\frac{N-n}{N-1} = \lim_{N\to\infty}\frac{1-\frac{n}{N}}{1-\frac{1}{N}} = 1$$

তাই পূর্ণকটার অসীম সংখ্যক সদস্য থাকলে s.e.$(\bar{y})$ দু'ক্ষেত্রে একই হবে।

টীকা 2 একথাই বলছে যে অসীম পূর্ণক থেকে নমুনা সংগ্রহের ক্ষেত্রে SRSWOR ও SRSWR-এর মধ্যে কোনও পার্থক্য থাকছে না।

৪৫.৫ নমুনাজ অনুপাতের বিভাজন

পূর্ণকের সদস্যগুলোকে একটা গুণলক্ষণ-র সাপেক্ষে দেখা যাক। যেমন, একটা শ্রেণীর ছাত্রদের চশমা আছে কিনা তার সাপেক্ষে দেখা, সে কী ভাষায় কথা বলে সেভাবে দেখা, ইত্যাদি। আলোচ্য গুণলক্ষণটা 'C' দিয়ে চিহ্নিত করি, N সংখ্যক সদস্যযুক্ত একটা পূর্ণকে যদি P অনুপাতের সদস্যদের 'C' গুণটা থাকে তাহলে P একটা পূর্ণকাঙ্ক যার মান $0 \le P \le 1$ এবং একে আমরা পূর্ণক-অনুপাত বলব।

[$P = 0$ মানে C একদম নেই। $P = 1$ মানে C সবারই আছে]

এই পূর্ণকের সদস্যের জন্য X চল এমনভাবে নেওয়া যাক যে i তম সদস্যের ক্ষেত্রে X-এর মান X_i যাতে

$X_i = 1$, যদি i-তম সদস্যের cg-গুন থাকে,

$= 0$, যদি না থাকে।

$i = 1, 2, \ldots\ldots, N$

এই মানগুলোর গড় (μ) ও ভেদমান (σ^2) এমন হবে—

$$\mu = \frac{1}{N}\sum_{i=1}^{N} X_i$$

$$= \frac{1}{N}.NP$$

$$= P$$

এবং
$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N} X_i^2 - \mu^2$$

$$= \frac{1}{N}.NP - P^2$$

$$= P(1 - P)$$

এমন পূর্ণক থেকে n সদস্যকে একটা নমুনা (অবশ্যই সমসম্ভব) হিসাবে নেওয়া হল যার মধ্যে f সদস্যের 'C' গুণটা আছে। তাহলে নমুনাটাতে আলোচ্য অনুপাতের প্রকাশ কেমন হবে?

$\frac{f}{n} = p$ (বলা যাক)

কে আমরা **নমুনাজ** অনুপাত বলব।

('C' গুণ সাপেক্ষে)

j তম নমুনা-সদস্যের জন্য x_j-মান এমন নেওয়া যাক যে

$x_j = 1$, যদি j-তম নমুনা-সদস্যের C-তম গুণটা থাকে 0, যদি না থাকে

$j = 1(1)n$

তাহলে, নমুনাজ গড়

$$\bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j$$

$$= \frac{1}{n}.f$$

$$= p$$

এটাই আলোচ্য নমুনাঙ্ক। এরও মান নমুনার সংগ্রহের সঙ্গে সঙ্গে পরিবর্তন হতেই পারে। তাই p-র নমুনাজ চাঞ্চল্য। সুতরাং এই p-এর নমুনাজ বিভাজন হচ্ছে p-এর মানের পরিসংখ্যা বিভাজনের সীমাস্থ অবস্থা।

৪৫.৬ p-এর নমুনাজ বিভাজনের প্রত্যাশা ও সমক ভ্রান্তি

আমরা তুলনা করেই লিখে ফেলতে পারি।

আমরা এইখানে দেখছি $\bar{x} = p$, $\mu = P$

$$\therefore\ E(\bar{x}) = \mu \Rightarrow E(p) = P \text{ ....}(i)$$

$$s.e._{SRSWR}(\bar{x}) = \frac{\sigma}{\sqrt{n}} \Rightarrow s.e._{SRSWR}(p) = \frac{\sqrt{P(1-P)}}{\sqrt{n}} \text{ ....}(ii)$$

$$s.e._{SRSWOR}(\bar{x}) = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} \Rightarrow s.e._{SRSWOR}(p) = \frac{\sqrt{P(1-P)}}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} \text{ ....}(iii)$$

৪৫.৭ সারাংশ

একটা পূর্ণক কতকগুলো সদস্য বা একক-সমষ্টি। পূর্ণকটা সসীম বা অসীম হতে পারে। সম্পূর্ণ পূর্ণকটার একটা সুনির্দিষ্ট অংশকে অনেক সময় লক্ষ্য পূর্ণক হিসাবে নিতে হয়।

রাশিবিজ্ঞানের প্রয়োজনে নমুনাভিত্তিক তত্ত্ব এসে পড়ে। এর জন্যে আসে নমুনাঙ্ক। এর মান এক নমুনা থেকে আরেকটার ক্ষেত্রে আলাদা আলাদা হয় (হতে পারে)। এই সমস্ত মানগুলোর পরিসংখ্যা বিভাজনের সীমা-অবস্থাই নমুনাঙ্কটার নমুনাজ বিভাজন।

নমুনাজ বিভাজনের গড় ও সমক বিচ্যুতি থাকতেই পারে। অন্যান্য ধর্ম ও তাদের পরিমাপও সম্ভব। এক্ষেত্রে সমক বিচ্যুতির নাম সমক ভ্রান্তি।

আমাদের আলোচনার স্বল্প পরিবারে নমুনাজ গড় ও নমুনাজ অনুপাতের যে বিভাজন হবে তার গড় ও সমক ভ্রান্তি উপস্থিত করা গেল। অন্য কোন নির্দিষ্ট নমুনাঙ্কের ক্ষেত্রে নমুনাজ বিভাজন এবং তার গড় ও সমক ভ্রান্তি নির্ণয় করা যেতে পারে।

৪৫.৮ অনুশীলনী

1. একটা প্রদেশের সমগ্র জনগোষ্ঠীকে কি একটা পূর্ণক হিসাবে দেখা যায়? এক্ষেত্রে নমুনা একক কী হতে পারে?
2. একটা শ্রেণীর সমস্ত ছাত্রকে একটা পূর্ণক হিসাবে দেখলে এমন একটা উদাহরণ তৈরী করুন যাতে—
   (i) একটা বিচ্ছিন্ন চলকের উপর তথ্য সংগ্রহ করা হবে
   (ii) দুটো পূর্ণকাঙ্ক প্রয়োজন হবে
   (iii) নমুনাঙ্ক দুটি সহজেই অনুমেয় হয়।
3. 4টি নারকেল গাছে যথাক্রমে 40, 32, 38, ও 42টি নারকেল ধরেছে। এই চারটে গাছকেই পূর্ণক হিসাবে ধরে নারকেল সংখ্যার পূর্ণক গড় ও সমক বিচ্যুতি নির্ণয় করুন। দুটো নারকেল গাছের নমুনা কতভাবে নেওয়া যাবে (i) SRSWR ও (ii) SRSWOR ক্ষেত্রে? উভয়ক্ষেত্রে নমুনাগড়ের নমুনাজ বিভাজন নির্ণয় করুন ও নমুনাজ বিভাজনের গড় ও সমক বিচ্যুতি নির্ণয় করুন। দেখান যে মানদুটি সূত্রের সাহায্যে প্রাপ্ত মানের সমান।
4. ৪৫.৩ উদাহরণে যদি SRSWOR নেওয়া হয় তাহলে নমুনাজ বিভাজন নির্ণয় করুন।
5. নমুনাজ গড়ের গাণিতিক প্রত্যাশার সাপেক্ষে SRSWR ও SRSWOR তুলনা করুন। কোনটি শ্রেয় ও কেন আলোচনা করুন।
6. ৪৫.৩ উদাহরণে SRSWR ও SRSWOR উভয় ক্ষেত্রে নমুনাজ বিভাজনের গড় ও সমক বিচ্যুতি নির্ণয় করুন ও দেখান যে তা নমুনাজ গড়ের গাণিতিক প্রত্যাশা ও সমক-ভ্রান্তির সূত্রে প্রাপ্ত মানের সমান।
7. SRSWR ও SRSWOR উভয় ক্ষেত্রে নমুনা অনুপাতের গাণিতিক প্রত্যাশা ও সমক ভ্রান্তির সূত্র নির্ণয় করুন।
8. SRSWR ও SRSWOR উভয় ক্ষেত্রে নমুনা গড়ের গাণিতিক প্রত্যাশা ও সমক ভ্রান্তির সূত্র নির্ণয় করুন।

একক ৪৬ ❑ বিন্দু প্রাককলন

গঠন

৪৬.০ উদ্দেশ্য

এই অনুচ্ছেদে আমরা প্রাককলনের প্রয়োজনীয়তার কথা আলোচনা করব। কোন নমুনাঙ্ককে প্রাককলনের কাজে লাগানো যাবে কিনা তাও আলোচিত হবে। তাছাড়া আলোচিত হবে—

(i) প্রাককলক ও প্রাককলনী মান কাকে বলে। একই প্রাককলক বিভিন্ন প্রাককলনী মান নিতে পারে। এই বিভিন্নতা এক নমুনা থেকে অন্যটার সদস্য-পার্থক্যের জন্যই হয়। এটাই নমুনাজ চাঞ্চল্য।

(ii) বিভিন্ন প্রাককলকের মধ্যে কোনটাকে গ্রহণ করব—একটা নির্দিষ্ট পূর্ণকাঙ্কের মান জানার জন্য বা এর কোন অপেক্ষকের মান জানার জন্য।

(iii) পক্ষপাতশূন্য ও লঘিষ্ঠ-ভেদমান-অপক্ষপাতী প্রাককলক কাকে বলে।

(iv) যদি কোন পূর্ণকাঙ্কের বা তার কোন অপেক্ষকের জন্য পক্ষপাতশূন্য প্রাককলক থাকে তা অদ্বিতীয় নাও হতে পারে।

(v) সমঞ্জস ও দক্ষ প্রাককলক কাকে বলে।

(vi) কয়েকটা প্রাককলন পদ্ধতির নাম।

(vii) গরিষ্ঠ আশংসামান পদ্ধতি কী এবং তা কেমনভাবে ব্যবহার করা যায়।

৪৬.১ প্রস্তাবনা

একটা পূর্ণক সম্বন্ধে ধারণা করতে আমরা তার থেকে নমুনা সংগ্রহ করি। প্রধানত সমসম্ভব নমুনা। নমুনা যেহেতু পূর্ণকের একটা অংশমাত্র তাই এর থেকে পূর্ণকটার সম্পূর্ণ ধারণা করা সম্ভব নয়। তবে যুক্তিনিষ্ঠ ও গাণিতিক পদ্ধতিতে এই নমুনার ওপর ভিত্তি করেই পূর্ণকটার ধারণা করা যায়। এই আলোচনাকে আমরা **পুরানী অনুমানতত্ত্ব (Classical Inference)** বলি।

একটা পূর্ণক যদি সম্পূর্ণভাবে পর্যবেক্ষণ করা হয়, তাহলে তো এর কোন বৈশিষ্ট্য বা প্রকৃতিই (Characteristic) অজানা থাকে না। ধারণা পূর্ণ হয়। নমুনা সমীক্ষা (Sample survey) অংশে আমরা একথা জেনেছি যে কখন এবং কেন নমুনা সংগ্রহ করার প্রয়োজন হয়। তাই এক্ষেত্রে আমরা সোজাসুজি ভেবেই নিই যে নমুনাজ তথ্য আমাদের হাতেই আছে; শুধু প্রশ্ন এই যে পূর্ণকটার সম্বন্ধে তা আমাদের কী জানাতে পারে।

অনুমানতত্ত্বর প্রধানত দুটো ভাগ :

1. পূর্ণকের বৈশিষ্ট্য বর্ণনা করতে আমরা এক বা একাধিক পূর্ণকাঙ্কর প্রয়োজন বোধ করি আর তাদের মান অজানা থাকতে পারে। নমুনার ওপর ভিত্তি করেই এই পূর্ণকাঙ্কগুলোর মান সম্বন্ধে ধারণা করা যায়। এই পদ্ধতিকে **প্রাককলন পদ্ধতি** (Estimation) বলে।

2. পূর্ণক সম্বন্ধে বা তার এক বা একাধিক পূর্ণকাঙ্ক সম্বন্ধে কোন পূর্ব ধারণা বক্তব্যকে আমরা প্রকল্প বলি। এমন কোনও প্রকল্প বিচার করার জন্যও নমুনা ব্যবহার করতে হয়। যে পদ্ধতিতে কোন প্রকল্প বিচার করা হয়—তার গ্রহণযোগ্যতার নিরিখে—তাকে আমরা প্রকল্প বিচার (testing of hypothesis) বলি।

প্রাককলন পদ্ধতিতে একটা বিশেষ মানকে পূর্ণকাঙ্কের মান হিসাবে নির্ণয় করা হয়। তাই তখন একে আমরা বিন্দু প্রাককলন বলি। যখন একটা অন্তর (interval) নির্ণয় করা হয় যার মধ্যে পূর্ণকাঙ্কটা থাকার একটা নির্দিষ্ট সম্ভাবনা থাকে তাকে আমরা অন্তর প্রাককলন বলি।

এই পরিচ্ছেদে আমরা বিন্দু প্রাককলনের তত্ত্ব আলোচনা করব। তার জন্য ধরে নেব যে পূর্ণকটার গাণিতিক রূপ জানা আছে। আলোচ্য চল X যদি গণনসাধ্য মানগ্রাহী হয় তবে p.m.f টা জানা থাকছে আর X যদি অবিচ্ছিন্ন হয় তবে p. d. f. টা জানা থাকছে। এর ক্রমযৌগিক নিবেশন অপেক্ষক (c.d.f.)কে $F_\theta(x)$ বলে নেওয়া যাক যেখানে θ হল একটা পূর্ণকাঙ্ক যার মান অজানা। θ এক্ষেত্রে আবার এক বা একাধিক স্থানাঙ্কবিশিষ্ট হতে পারে।

৪৬.২ ধারণা ও সংজ্ঞা

৪৬.২.১ প্রাককলনী মান (estimate) ও প্রাককলক (estimator)

$F_\theta(x)$ থেকে সমসম্ভব নমুনা (X_1, X_2,........, X_n) নেওয়া হল, যার মধ্যে n সংখ্যক অবেক্ষক (observations) আছে। X_1, X_2,........, X_n-এর মান যথাক্রমে X_1, X_2,........, X_n হলে, তথ্য সমষ্টি হচ্ছে (X_1, X_2,........, X_n)।

θ যদি পূর্ণকটার গড় হয় তবে এই মানগুলো আমরা কীভাবে কাজে লাগাবো? কেউ বলতে পারে X_1, X_2...........X_n-এর গড় নিয়ে, যেমন

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

কেউ বলবে X_{mode}, X_{median} বা

$$g = (X_1 . X_2 ....X_n)^{1/n}; \text{ ইত্যাদি।}$$

এর প্রত্যেকটা থেকেই এক একটা মান পাওয়া যাবে। এই মানটাকে বলে প্রাককলনী মান। আবার এদের প্রত্যেকটাই এক একটা "ফরমুলা" যাতে n সংখ্যক চল X_1, X_2,........X_n যুক্ত। তাই

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

বা অন্যগুলো এক একটা নমুনাঙ্ক (statistic)। এক্ষেত্রে নমুনাঙ্ক প্রাককলনে কাজে লাগছে। তাই এখন এর নাম প্রাককলক। এভাবে বহু সংখ্যক প্রাককলক ভাবা যেতে পারে। সুতরাং, একটা যুক্তিযুক্ত নমুনাঙ্কের সাহায্যে অজানা মানবিশিষ্ট পূর্ণকাঙ্ক θ-র প্রাককলনী মান নির্ণয় করাই বিন্দু প্রাককলন-এর মূল কাজ। এতে আমরা θ-র মান-দেশের একটা মান বা বিন্দুই পেতে পারি। তাই এই আলোচনাকে "বিন্দু" প্রাককলন বলা হয়।

X_1, X_2,.........X_n এর একটা অপেক্ষক।

$T(X_1, X_2,.........X_n)$-কে θ-র একটা প্রাককলক হিসাবে নেওয়া যাক। তাহলে, মূল মান X_1, X_2,.........X_n বসিয়ে যে মান পাই তা হ'ল θ-র একটা প্রাককলনী মান। এখন θ-র প্রাককলকটা লেখা যেতে পারে—

$$\hat{\theta} = T(X_1, X_2, ..., X_n)$$

উদাহরণ 1 :

একটা পূর্ণকাঙ্ক λ যুক্ত পোয়াসঁ নিবেশন থেকে নেওয়া সমসম্ভব নমুনা $X_1, X_2, \ldots\ldots X_n$ হ'লে

$$\bar{x} = \frac{1}{x} \sum_{i=1}^{n} X_i$$

λ-র একটা প্রাককলক। আবার যদি সবচেয়ে ছোট মানযুক্ত নমুনা অবেক্ষক ও সবচেয়ে বড় মানযুক্ত নমুনা অবেক্ষক বাদ দিয়ে বাকি $(n - 2)$ অবেক্ষকগুলোর গড় নেওয়া হয়, তাও λ-র একটা প্রাককলক হবে।

উদাহরণ 2 :

ধরা যাক একটা বাইনোমিয়াল (m, p) নিবেশনের p-এর মান জানা নেই। এই নিবেশনটা থেকে সমসম্ভব নমুনা $X_1, X_2, \ldots\ldots X_n$ নেওয়া হলো। তাহলে

$$\frac{1}{mn} \sum_{i=1}^{n} x_i, \quad \frac{1}{2m}(x_1 + x_n)$$

দুটোই p-এর প্রাককলক।

এখন এটা স্পষ্ট যে একই পূর্ণকাঙ্কের জন্য অনেক প্রাককলক পাওয়া সম্ভব। তাহলে এদের মধ্যে কোনটা ভালো বা অনেকগুলো ভালোর মধ্যে কোনটা শ্রেষ্ঠ তা ঠিক করতে হবে। এর জন্য প্রয়োজন কতকগুলো যুক্তিযুক্ত ধর্ম (logical properties) যা সঠিক প্রাককলকটা খুঁজতে সাহায্য করবে।

অনেক সময় θ-র একটা অপেক্ষক $g(\theta)$-র মান জানার প্রয়োজন হয়। তখন $T(X_1, X_2, \ldots\ldots, X_n)$ -কে $g(\theta)$-র প্রাককলক হিসাবে নিতে হবে। যেমন, $N(\mu, \sigma^2)$ নিবেশনের জন্য আমরা যদি $E(X^2) = \mu^2 + \sigma^2$ এর মান জানতে আগ্রহী হই তবে আলাদা করে μ ও σ^2-এর মান জানার বদলে $\mu^2 + \sigma^2$- যেটা μ ও σ^2-এর একটা অপেক্ষক, তার মান জানতে চাইব।

৪৬.২.২ পক্ষপাতশূন্য প্রাককলক

একটা প্রাককলক $T(X_1, X_2, \ldots\ldots, X_n)$-কে তখনই θ-র একটা অপেক্ষক $g(\theta)$-র জন্য পক্ষপাতশূন্য বলা হবে যখন $E_\theta(T)$-র অস্তিত্ব থাকবে এবং

$$E_\theta(T) = g(\theta) \;\; \forall \;\; \theta \in \Theta,$$

যেখানে Θ হল পূর্ণকাঙ্ক দেশ (parameter space)

T যদি $g(\theta)$-র জন্য পক্ষপাতশূন্য প্রাককলক হয় তার অর্থ দাঁড়াবে— T-র মান $g(\theta)$-র সঠিক মানের আশেপাশে থাকবে অর্থাৎ $g(\theta)$-র চেয়ে বড় বা ছোট হতে পারে। এর কারণ, $X_1, X_2, \ldots\ldots, X_n$ এর

বিশেষ ও নির্দিষ্ট মান যা নমুনা অবেক্ষক হিসাবে নেওয়া হয়েছে। এবং এইসব মানগুলোর গাণিতিক গড় $g(\theta)$-র সমান হবে θ-র মূল মানের ওপর নির্ভর না করেই। পূর্ণকাঙ্ক অপেক্ষক $g(\theta)$-কে প্রাককলনীয় (estimable) বলা হবে যদি এর জন্য অন্তত একটা অপক্ষপাতী প্রাককলক থাকে।

উদাহরণ **3 :**

একটা সসীম ভেদমান যুক্ত পূর্ণক গড় (μ) ও ভেদমান (σ^2) জানা নেই। μ-এর মান সম্বন্ধে ধারণা করার এই পূর্ণক থেকে সমসম্ভাবনাযুক্ত নমুনা $X_1, X_2, \ldots\ldots\ldots, X_n$ নেওয়া হল।

আমরা জানি,

$$E(X_j) = \mu, \quad j = 1(1)n$$

সুতরাং, $X_1, X_2, \ldots\ldots\ldots, X_n$ এর প্রত্যেকটাই μ-এর জন্য পক্ষপাতহীন প্রাককলক।

আবার,

$$E\left(\frac{X_1 + X_2}{2}\right) = \frac{E(X_1) + E(X_2)}{2}$$

$$= \mu$$

সুতরাং $\left(\frac{X_1 + X_2}{2}\right)$ ও μ-এর জন্য পক্ষপাতহীন. প্রাককলক।

এখন দেখা যাক μ-এর জন্য এমন ঋজুরৈখিক ও পক্ষপাতহীন প্রাককলক কতগুলো পাওয়া যেতে পারে। $X_1, X_2, \ldots\ldots\ldots, X_n$ এর একটা রৈখিক অপেক্ষক বা T নেওয়া যাক এমনভাবে যে

$$T = C_1 X_1 + C_2 X_2 + \ldots\ldots\ldots + C_n X_n,$$

যেখানে $C_1, C_2, \ldots\ldots\ldots, C_n$ এমন nটা ধ্রুবক যাদের মান যেমন ইচ্ছা তেমনই নেওয়া যায়।

এক্ষেত্রে, $E(T) = C_1E(X_1) + C_2E(X_2) + \ldots\ldots\ldots + C_nE(X_n)$

$$= C_1\mu + C_2\mu + \ldots\ldots + C_n\mu$$

$$= (C_1 + C_2 + \ldots\ldots + C_n)\mu$$

আমরা T-কে μ-এর জন্য পক্ষপাতহীন প্রাককলক করতে চাই। তার মানে E(T)-এর সঙ্গে μ-এর মান সমান হতেই হবে, μ-এর মান যাই হোক না কেন।

তাই, $E(T) = \mu$

$$\Rightarrow (C_1 + C_2 + \ldots\ldots + C_n)\mu = \mu$$

$$\Rightarrow C_1 + C_2 + \ldots\ldots + C_n = 1 \ldots\ldots\ldots\ldots\ldots\ldots\text{(i)}$$

অসীম সংখ্যক উপায়ে $C_1, C_2, \ldots\ldots, C_n$ পেতে পারি যাতে এগুলো (i)-র শর্তটা মেনে চলে। সুতরাং μ-এর ও অসীম সংখ্যক পক্ষপাতহীন প্রাককলক আছে। পক্ষপাতহীনতার নিরিখে এগুলো সবই সমান যোগ্য।

এখন প্রশ্ন, এদের মধ্যে থেকে আবার কোনটা μ-এর জন্য যোগ্যতম?

এই কাজে যে প্রাককলকটা অন্য একটার চেয়ে μ-এর বেশী কাছাকাছি থাকার প্রবণতা আছে সেটাই বেশী যোগ্য বলে নেওয়া যেতে পারে। এজন্য প্রাককলক T-এর ভেদমান বিচার করতে হবে।

৪৬.২.৩ লঘিষ্ঠ ভেদমান ও অপক্ষপাতী প্রাককলক

T_1 ও T_2 যদি উভয়েই $g(\theta)$-এর জন্য পক্ষপাতশূন্য হয় এবং

$$V_\theta(T_1) \le V_\theta(T_2).\ \forall\ \theta \in \text{Ⓗ}$$

তাহলে T_1-কে T_2-র তুলনায় বেশী গ্রহণযোগ্য বলা হবে। যখন দুটো ভেদমান সমান তখন T_1 ও T_2 সমযোগ্য হয়ে যাবে।

সংজ্ঞা 2 :

(লঘিষ্ঠ-ভেদমান-অপক্ষপাতী প্রাককলক)

Minimum-Variance-unbiased Estimator বা m. v. u. e.

n সংখ্যক নমুনা $X_1, X_2, \ldots\ldots, X_n$ এর একটা অপেক্ষক $T = T(X_1, X_2, \ldots\ldots, X_n)$-কে $g(\theta)$-র **লঘিষ্ঠ-ভেদমান-অপক্ষপাতী প্রাককলক** বলা হবে যদি (1) $E_\theta(T) = g(\theta)\ \forall\ \theta \in \text{Ⓗ}$

অর্থাৎ $g(\theta)$-র জন্য T অপেক্ষপাতী (θ-র যে কোন মানের জন্য)

এবং (2) $Var_\theta(T) \le Var_\theta(T')\ \forall\ \theta \in \text{Ⓗ}$

যেখানে T' হল অন্য যে কোন প্রাককলক যা $X_1, X_2, \ldots\ldots, X_n$-এর একটা অপেক্ষক ও (1) নং শর্ত মেনে চলে।

উদাহরণ 3-এর শেষ অংশে দেখলাম যে ঋজু রৈখিক প্রাককলক একই পূর্ণকাঙ্কের জন্য অসীম সংখ্যক হতে পারে। সুতরাং এদের মধ্যে কোনটা যে কোন এক নির্দিষ্ট অর্থে সর্বোত্তম হবে তা ঠিক করার প্রয়োজন। সবথেকে বেশী প্রয়োগ করা হয় এমন উপায়টা হল, এমন একটা প্রাককলক নেওয়া যার ভেদমান অন্য সব অপক্ষপাতী প্রাককলকগুলোর ভেদমানের মধ্যে সবচেয়ে ছোট। এই প্রাককলকেই আমরা সর্বোত্তম-ঋজু-রৈখিক-অপক্ষপাতী প্রাককলক (Best Linear-Unbiased Estimator) বা লঘিষ্ঠ-ভেদমান-ঋজু-রৈখিক প্রাককলক (Minimum-Variance-Linear Unbiased Estimator) বলব।

উদাহরণ 4 :

সসীম ভেদমানযুক্ত পূর্ণকের ক্ষেত্রে নমুনাজগড় $\overline{X}$ পূর্ণকাঙ্ক μ(পূর্ণক গড়)-এর জন্য লঘিষ্ঠ-ভেদমান-ঋজু রৈখিক প্রাককলক হবে।

উদাহরণ 5 :

(অপক্ষপাতী প্রাককলক বিষয়ক)

এমন একটা অসীম পূর্ণক যার ভেদমান σ^2 (এবং গড় μ)-এর অস্তিত্ব আছে তার থেকে অনপেক্ষ সম্ভবনাশ্রয়ী নমুনা $X_1, X_2, \ldots\ldots\ldots, X_n$ নেওয়া হল। নমুনা ভেদমান—

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

এখন $E\left[\sum_{i=1}^{n}(X_i - \overline{X})^2\right]$

$$= E\left[\sum_{i=1}^{n} X_i^2 - n\overline{X}^2\right]$$

$$= \sum_{i=1}^{n} E(X_i^2) - n\,E(\overline{X}^2) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(a)$$

আমরা জানি,

$$E(X_i^2) = V(X_i) + \{E(X_i)\}^2$$

$$= \sigma^2 + \mu^2 \ldots\ldots\ldots\ldots\ldots\ldots\ldots(b)$$

এবং $E(\overline{X}^2) = V(\overline{X}) + \{E(\overline{X})\}^2$

$$= \frac{\sigma^2}{n} + \mu^2 \ldots\ldots\ldots\ldots\ldots(c)$$

অতএব, (b) ও (c)-কে (a)-তে ব্যবহার করে পাওয়া যায় যে—

$$E\left[\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2\right]=E\left\{\sum_{i=1}^{n}X_i^2-n\overline{X}^2\right\}$$

$$=\sum_{i=1}^{n}E\left(Xi^2\right)-nE\left(\overline{X}^2\right)$$

$$=\sum_{i=1}^{n}\left(\sigma^2+\mu^2\right)-n\left(\frac{\sigma^2}{n}+\mu^2\right)$$

$$=n\sigma^2+n\mu^2-\sigma^2-n\mu^2$$

$$=(n-1)\sigma^2$$

$$\Rightarrow E\left[\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2\right]=\sigma^2\ \forall\mu\ \&\ \sigma^2$$

অর্থাৎ নমুনা ভেদমান S^2 পূর্ণক ভেদমান σ^2-র একটা পক্ষপাতহীন প্রাককলক। এক্ষেত্রে $S^2=\frac{1}{n-1}\Sigma(X_i-\overline{X})^2$-এ ভাজক n-এর বদলে $n-1$ নিতে হবে। এই একই নমুনার ক্ষেত্রে আমরা নমুনাজ খণ্ড-গড় $\overline{X}_i$ ও খণ্ড-ভেদমান s_i^2 পেতে পারি—

$$\overline{X}_i=\frac{X_1+X_2+\dots\dots\dots\dots+X_i}{i}.$$

$$S_i^2=\frac{1}{i-1}\sum_{j=1}^{i}\left(X_j-\overline{X}_i\right)^2\quad i=2(1)n$$

একই ভাবে দেখানো যায় যে পূর্ণক গড় μ-এর জন্য প্রত্যেক $\overline{X}_i$ ও পূর্ণক ভেদমান σ^2-র জন্য প্রত্যেক s_i^2 অপক্ষপাতী প্রাককলক।

৪৬.২.৪ সমঞ্জস প্রাককলক ও দক্ষ প্রাককলক

এতক্ষণ যে আলোচনা হল তাতে n (নমুনার সদস্য-সংখ্যা) স্থির (একই) ছিল। এখন আমরা এও ভাবতে পারি যে n-এর মান ধীরে ধীরে বাড়ানো হচ্ছে। তাহলে আমরা মনে করব যে পূর্ণকটা সম্বন্ধে আরো বেশী ধারণা করা যাবে। অন্ততঃ এটা ভাবতেও পারি, যে পূর্ণকাঙ্ক বা তার অপেক্ষকের মান জানতে চাইছি তা আরো ভালো বা সঠিকভাবে অনুমান করা যাবে। যখন n-এর মান বাড়ানো সম্ভব হবে তখন আমরা একটা প্রাককলকের মান কীভাবে পরিবর্তন হবে তাই দেখব। দেখব যে এই মান অজানা পূর্ণকাঙ্কটার (বা তার অপেক্ষকটার) কাছে বা আরো কাছে যাচ্ছে কিনা। সেক্ষেত্রে আমরা প্রাককলকটিকে সমঞ্জস বলব।

এক্ষেত্রে আমরা প্রাককলক $T \equiv T(X_1, X_2, \dots\dots\dots, X_n)$-কে চলগুলোর অপেক্ষক হিসাবে যেমন দেখব, n-এর অপেক্ষক হিসাবেও তেমন চিন্তা করব।

এখন সমঞ্জস-এর একটা সাধারণ সংজ্ঞা প্রয়োজন।

T_1 ও T_2 উভয়েই পূর্ণকাঙ্ক θ-র প্রাককলক যারা $X_1, X_2, \ldots\ldots, X_n$-এর ওপর ভিত্তি করে আছে। তাহলে যদি θ-এর যে কোন মানের জন্য ও যে কোন $\varepsilon(>0)$-র জন্য $P_\theta\left(|T_1-\theta|<\varepsilon\right)>P_\theta\left(|T_2-\theta|<\varepsilon\right)$ হয়, আমরা T_1-কে θ-র প্রাককলনের কাজে বেশী পছন্দ করব।

ওপরের বক্তব্যে প্রাককলক ও পূর্ণকাঙ্কের দূরত্ব কম হওয়ার বেশী সম্ভাবনা চাওয়া হয়েছে।

সংজ্ঞা 3 :

n-এর মানের ওপর নির্ভরশীল প্রাককলক T যদি এমন হয় যে, যেকোন ছোট ও ধনাত্মক মান ε ও η-র জন্য যখনই $n \geq n_0$ হবে তখনই $P_\theta(|T-\gamma(\theta)|<\varepsilon)>1-\eta$

যেখানে n_0 অবশ্যই ε ও η-এ ওপর নির্ভর করবে ও সম্ভবত θ-র ওপরও নির্ভর করবে, তাহলে T–কে পূর্ণকাঙ্ক অপেক্ষা $\gamma(\theta)$-র জন্য সমঞ্জস বলা হবে।

সমঞ্জস প্রাককলক হওয়ার জন্য প্রাককলক T এমন হতে হবে যে

(i) $E(T) \to \gamma(\theta)$, যখন $n \to \infty$

ও (ii) $\text{Var}(T) \to 0$, যখন $n \to \infty$

এটা প্রমাণ করা যায়। এখন একটা প্রাককলক T $\gamma(\theta)$-র জন্য সমঞ্জস কিনা তা বিচারের কাজে এটি ব্যবহার করব।

উদাহরণ ঃ

সসীম ভেদমান (σ^2) বিশিষ্ট পূর্ণকের গড় μ-এর জন্য নমুনাজ গড় $\overline{X}$ নেওয়া যাক।

এখন $E(\overline{X}) = \mu$ (সবসময়)

আর $\text{Var}(\overline{X}) = \dfrac{\sigma^2}{n}$ (পূর্ণকটাকে অসীম ধরে)

$\to 0$ যেহেতু $n \to \infty$

(যেহেতু $\sigma^2 < \infty$)

তাই, $\overline{X}$ μ-এর জন্যে সমঞ্জস প্রাককলক।

টীকা :

$\overline{X}$ একটা সমঞ্জস প্রাককলক একথা আসলে ছোট করে বলা হল। আসল ও সম্পূর্ণ অর্থ এই যে প্রাককলক-ক্রম $\overline{X}_n = (X_1 + X_2 + \ldots\ldots + X_n)/n$ সমঞ্জস।

টীকা :

$\gamma(\theta)$-র জন্য T একটা সমঞ্জস প্রাককলক। একটা ধ্রুবক 'a' ও n-এর অপেক্ষক $\psi(n)$ নেওয়া যাক, যাতে 'a' এবং n পরস্পর নির্ভরশীল না হয় ও $\psi(n)$ n-এর সঙ্গে ক্রমবর্ধমান।

তাহলে $T + \frac{a}{\psi(n)}$ এবং $T\left\{1 + \frac{a}{\psi(n)}\right\}$

$\gamma(\theta)$-র জন্য সমঞ্জস হবে। অর্থাৎ সমঞ্জস প্রাককলক অদ্বিতীয় নয়।

এখন প্রয়োজন এত অসংখ্য সমঞ্জস প্রাককলকের মধ্যে তুলনা করা যাতে $\gamma(\theta)$-র সবচেয়ে উপযুক্ত সমঞ্জস প্রাককলকটা পাওয়া যায়। এর জন্য আমরা দেখব কোন প্রাককলকটার $\gamma(\theta)$-র দিকে সবচেয়ে দ্রুত অভিসরণ (Convergence) ঘটছে। এই কাজ সাধারণভাবে বা সব প্রাককলকের ক্ষেত্রে সম্ভব নয়। তাই আমরা দেখব, n-এর মান অসীমের দিকে যাওয়ার ফলে কোন কোন প্রাককলকের নমুনাজ বিভাজন সীমাস্থ অবস্থায় নর্ম্যাল হচ্ছে। তখন, যে প্রাককলকের সীমাস্থ নর্ম্যাল অবস্থার ভেদমান অন্য সবগুলোর চেয়ে ছোট তাকেই আমরা দক্ষ প্রাককলক বলব।

তাই $\gamma(\theta)$-র একটা সমঞ্জস প্রাককলক T-র যদি সীমাস্থ নর্ম্যাল বিভাজন থাকে ও ভেদমান $C(\theta)$ হয় এবং অন্য যে কোন সমঞ্জস প্রাককলক T'-র ক্ষেত্রে সীমাস্থ নর্ম্যাল বিভাজনের ভেদমান $C'(\theta)$ হয় যাতে

$$C(\theta) \leq C'(\theta) \ \forall \ \theta$$

তবে T-কে দক্ষ প্রাককলক বলা হবে।

উদাহরণ : X চলের বিভাজন থেকে একটা সমসম্ভব নমুনা $(X_1, X_2, \ldots\ldots X_n)$ নেওয়া গেল। ধরা যাক X-এর বিভাজনটার μ_4-র অস্তিত্ব আছে।

তাহলে পূর্ণক-ভেদমান (σ^2)-এর একটা সমঞ্জস প্রাককলক

$$\frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

এটা অবশ্য σ^2-র জন্য পক্ষপাতশূন্য নয়।

৪৬.৩ প্রাককলন-পদ্ধতি

আগের অধ্যায়ে আমরা একটা বিন্দু-প্রাককলকের বিভিন্ন বৈশিষ্ট্য আলোচনা করেছি। এখন আমরা কী পদ্ধতিতে বিন্দু-প্রাককলক নির্ণয় করা যায় তাই দেখব। এই পদ্ধতিগুলো হচ্ছে—

(a) গরিষ্ঠ আশংসামান পদ্ধতি (Method of Maximum Likelihood)

(b) লঘিষ্ঠ কাই-বর্গ পদ্ধতি (Method of Minimum Chi-square)

(c) পরিঘাত পদ্ধতি (Method of Moments)

(d) লঘিষ্ঠ বর্গ পদ্ধতি (Method of Least Squares)

৪৬.৩.১ গরিষ্ঠ আশংসামান পদ্ধতি

নমুনাজ চল $X_1, X_2, \ldots\ldots\ldots, X_n$ গুলোকে আমরা মূল চল হিসাবে নিই। θ এমন একটা পূর্ণকাঙ্ক যার মান জানার প্রয়োজন। এখানে θ পূর্ণকাঙ্ক-ভেক্টরও হতে পারে।

যখন চলগুলো বিচ্ছিন্ন, আমরা f_θ দিয়ে $X_1, X_2, \ldots\ldots\ldots, X_n$-এর যৌথ সম্ভাবনা-ভর অপেক্ষক লিখব। সুতরাং $f_\theta(X_1, X_2, \ldots\ldots\ldots, X_n)$

$$= P_\theta[X_1 = x_1, X_2 = x_2, \ldots\ldots\ldots, X_n = x_n]$$

এটা নির্দিষ্ট নমুনা-মান $x_1, x_2, \ldots\ldots\ldots, x_n$ পাওয়ার সম্ভাবনা।

$x_1, x_2, \ldots\ldots\ldots, x_n$ জানা থাকা অবস্থায় এটা আবার θ-র একটা অপেক্ষক। তাই এই সম্ভবনাকেই এক্ষেত্রে আমরা আশংসা অপেক্ষক বলব আর তাকে $L(\theta)$ বলে চিহ্নিত করব।

যখন চলগুলো সব অবিচ্ছিন্ন তখন f_θ যদি যৌথ সম্ভবনা-ঘনত্ব অপেক্ষক হয়, f_θ-কেই আমরা আশংসা অপেক্ষক বলব এবং $L(\theta)$ দিয়ে এটাকেও চিহ্নিত করব।

সাধারণ অবস্থায় $X_1, X_2, \ldots\ldots\ldots, X_n$-এর অনপেক্ষ ও অভিন্ন বিভাজন থাকবে এবং $L(\theta)$-কে আমরা এমন লিখতে পারব যে—

$$L(\theta) = f_\theta(x_1).\ f_\theta(x_2)\ldots\ldots\ldots\ldots f_\theta(x_n).$$

$$= \prod_{i=1}^{n} f_\theta(x_i)$$

এখন θ-র একটা মান $\hat{\theta}$ (পূর্ণকাঙ্ক-দেশ Ⓗ-এর ভেতরের কোন মান)-কে আমরা **গরিষ্ঠ আশংসামান প্রাককলক** বলব যদি এই মানের জন্য আশংসা-অপেক্ষকটা যত বড় সম্ভব মান নেয়। সুতরাং, $\hat{\theta}$ গরিষ্ঠ-আশংসামান প্রাককলক হবে যখন

$$L(\hat{\theta}) \geq L(\theta)\ \forall\ \theta \in \text{Ⓗ}.$$

বা, $L(\hat{\theta}) = \max\limits_{\theta} L(\theta)$

$L(\theta)$-কে সরাসরি কাজে লাগিয়ে $\hat{\theta}$ নির্ণয় করা অনেক সময় তেমন সহজ হয় না। বরং $\text{Log}_e L(\theta)$ নিয়ে অনেক সুবিধা পাওয়া যায়।

তখন $\text{Log}_e L(\hat{\theta}) \geq \log_e L(\theta) \;\forall\; \theta \in \Theta$

বা, $\text{Log}_e L(\hat{\theta}) = \max_{\theta} \log_e L(\theta)$

অনেক ক্ষেত্রেই অন্তরকলনের সাহায্য নেওয়া যেতে পারে। তাহলে $\hat{\theta}$ হবে

$$\frac{d}{d\theta}\log_e L(\theta) = 0$$

সমীকরণটার একটা মূল যেখানে $\log_e L(\theta)$-এর একটা সামগ্রিক গরিষ্ঠ মান (global maximum) থাকবে। [যেক্ষেত্রে $\theta = \hat{\theta}$-এ ব্যবকলনের অস্তিত্ব থাকে না সেখানে এই পদ্ধতি কাজে লাগে না।] এই সমীকরণটাকে **আশংসা সমীকরণ** বলা হয়।

উদাহরণ 1 : একটা বাইনোমিয়াল (m, p) নিবেশন থেকে n সংখ্যক নমুনা $X_1, X_2, \ldots\ldots\ldots, X_n$ নেওয়া হল যেগুলো অনধীন ও সমসম্ভব। এক্ষেত্রে m-র মান জানা আছে এবং p জানা নেই। সুতরাং

$L(p) = X_1, X_2, \ldots\ldots\ldots, X_n$ এর যৌথ সম্ভবনা-ভর অপেক্ষক

$$= \prod_{i=1}^{n} \binom{m}{x_i} p^{x_i}(1-p)^{m-x_i}, \quad x_1 = 0(1)m,\; 0 < p < 1$$

$$= \left\{\prod_{i=1}^{n}\binom{m}{x_i}\right\} p^{\sum_{i=1}^{n} x_i} (1-p)^{mn - \sum_{i=1}^{n} x_i}$$

$$\Rightarrow \log_e L(p) = \log_e\left\{\prod_{i=1}^{n}\binom{m}{x_i}\right\} + \left(\sum_{i=1}^{n} x_i\right)\log_e p + \left(mn - \sum_{i=1}^{n} x_i\right)\log_e(1-p)$$

সুতরাং আশংসা সমীকরণ হচ্ছে—

$$\frac{d}{dp}\log_e L(p) = 0$$

এই সমীকরণে p-র মূলকে $\hat{p}$ লিখলে,

$$\left(\sum_{i=1}^{n} x_i\right)\cdot\frac{1}{\hat{p}} + \left(mn - \sum_{i=1}^{n} x_i\right)\cdot\frac{-1}{1-\hat{p}} = 0$$

$$\Rightarrow (1-\hat{p})\sum_{i=1}^{n} x_i - \left(mn - \sum_{i=1}^{n} x_i\right)\hat{p} = 0$$

$$\Rightarrow \hat{p}mn = \sum_{i=1}^{n} x_i$$

$$\Rightarrow \hat{p} = \frac{1}{mn}\sum_{i=1}^{n} x_i$$

তাই বাইনোমিয়াল (m, p) নিবেশনে p-এর গরিষ্ঠ আশংসামান প্রাককলক হচ্ছে

$$\Rightarrow \hat{p} = \frac{1}{mn}\sum_{i=1}^{n} x_i$$

$$\text{or, } \hat{p} = \frac{\overline{X}}{m}$$

উদাহরণ 2 : একটা পোয়াসঁ (λ) নিবেশন থেকে অনধীন ও সমসম্ভব নমুনা $X_1, X_2, \ldots\ldots\ldots, X_n$ নেওয়া হল। পূর্ণকাঙ্ক λ-র গরিষ্ঠ আশংসা মান প্রাককলক নির্ণয় করতে হবে।

এক্ষেত্রে আশংসা-অপেক্ষক

$L(\lambda) = X_1, X_2, \ldots\ldots\ldots, X_n$ এর যৌথ সম্ভাবনা ভর অপেক্ষক

$$= \prod_{i=1}^{n} \left\{ e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} \right\}$$

$$= e^{-n\lambda} . \frac{\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} (x_i!)}$$

$$\Rightarrow \log_e L(\lambda) = -n\lambda + \left(\sum_{i=1}^{n} x_i \right) \log_e \lambda - \log_e \left(\prod_{i=n}^{n} x_i! \right)$$

∴ আশংসা-সমীকরণ হচ্ছে

$$\frac{d}{d\lambda} \log_e L(\lambda) = 0$$

এর মূল $\hat{\lambda}$ হলে, আমরা লিখতে পারি

$$-n + \left(\sum_{i=1}^{n} x_i \right) \frac{1}{\hat{\lambda}} = 0$$

$$\Rightarrow \hat{\lambda} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$\text{i.e. } \boxed{\hat{\lambda} = \overline{X}}$$

টীকা : উদাহরণ 2-তে λ-র গরিষ্ঠ আশংসামান প্রাকলক আসলে নমুনাজ গড়। উদাহরণ 1 এ p-র গরিষ্ঠ আশংসামান প্রাককলকটা "নমুনাজ অনুপাত", কারণ সেখানে নমুনাজ সাফল্য সংখ্যাগুলোর গড় হচ্ছে $\overline{X}$।

উদাহরণ 3 : একটা নর্ম্যাল (μ, σ^2) নিবেশনে μ ও σ^2-র গরিষ্ঠ আশংসামান প্রাককলক নির্ণয় করতে এই নিবেশন থেকে অনধীন ও সমসম্ভব নমুনা $X_1, X_2, \ldots\ldots\ldots, X_n$ নেওয়া হল। তাহ'লে এক্ষেত্রে পূর্ণকাঙ্ক হচ্ছে একটা ভেক্টর এবং μ ও σ^2 এর এবং সদস্য। এখানে আশংসা অপেক্ষকটা হবে এমন—

$$L(\mu,\sigma^2) = \prod_{i=1}^{n} \left\{ \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2} \right\}$$

$$= \frac{1}{(\sigma^2)^{\frac{n}{2}}(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

$$\Rightarrow \log_e L(\mu,\sigma^2) = -\frac{n}{2}\log_e \sigma^2 - \frac{n}{2}\log_e(2\pi) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

$\therefore$ আশংসা-সমীকরণ দু'টো হবে :

$$\frac{\partial}{\partial\mu}\log_e L(\mu;\sigma^2) = 0 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\text{(i)}$$

$$\frac{\partial}{\partial\sigma^2}\log_e L(\mu,\sigma^2) = 0 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\text{(ii)}$$

এদের সমাধান করলে মূল দুটোকে $\hat{\mu}$ ও $\hat{\sigma}^2$ লেখা যাক।

তাই, (i) $\Rightarrow -\frac{1}{2\hat{\sigma}^2}\sum_{i=1}^{n} 2(x_i-\hat{\mu})(-1) = 0$

$$\Rightarrow \sum_{i=1}^{n}(x_i-\hat{\mu}) = 0$$

$$\Rightarrow \sum_{i=1}^{n} x_i = n\hat{\mu}$$

$$\Rightarrow \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i$$, নমুনাজ গড়

(ii) $\Rightarrow -\frac{n}{2}\cdot\frac{1}{\hat{\sigma}^2} + \frac{1}{2\hat{\sigma}^4}\sum_{i=1}^{n}(x_i-\mu)^2 = 0$

$$\Rightarrow \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2$$

এই সমাধান স্তরে $\hat{\mu}$ -এর মান ব্যবহার করলে পাই

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2,$$ নমুনাজ ভেদমান

গরিষ্ঠ-আশংসামান প্রাককলকের কয়েকটা বিশেষ বৈশিষ্ট্য বা ধর্ম আছে। এগুলো প্রাককলকের যেসব ধর্ম থাকার প্রয়োজন বলে আমরা মনে করেছি মোটামুটি তার স্বপক্ষে।

যেমন—

1. পক্ষপাতশূন্যতা : গরিষ্ঠ আশংসামান প্রাককলক পক্ষপাতশূন্য নাও হতে পারে। এখানের উদাহরণ 3-এ যে $\hat{\sigma}^2$ তা σ^2 এর জন্য পক্ষপাতশূন্য নয়। তবে একটু সহজ পরিবর্তন করে আমরা পক্ষপাতশূন্য প্রাককলক পেতে পারি। $\frac{n\hat{\sigma}^2}{n-1}$ পক্ষপাতশূন্য হবে।
2. সমঞ্জস অবস্থা: কতকগুলো সাধারণ শর্ত পূর্ণ হলেই গরিষ্ঠ আশংসামান প্রাককলক সমঞ্জস প্রাককলক হবে।
3. সীমাস্থ নর্ম্যাল অবস্থা : সাধারণ কতকগুলো শর্ত সাপেক্ষে গরিষ্ঠ-আশংসামান প্রাককলকের সীমাস্থ নিবেশন নর্ম্যাল হবে।
4. দক্ষতা : যেসব সমঞ্জস প্রাককলকের সীমাস্থ অবস্থার নিবেশন নর্ম্যাল তাদের মধ্যে গরিষ্ঠ আশংসামান প্রাককলকের ভেদমান সবচেয়ে ছোট। তাই এটা দক্ষ।
5. অপরিবর্তনীয়তা : θ-র একমান-বিশিষ্ট (Single-valued) অপেক্ষক $\psi(\theta)$ যদি এমন হয় যে এর একটা স্থির বিপরীত মান আছে এবং θ-র গরিষ্ঠ-আশংসামান প্রাককলক $\hat{\theta}$, তবে $\psi(\hat{\theta})$ ও $\psi(\hat{\theta})$-র গরিষ্ঠ-আশংসামান প্রাককলক হবে।

৪৬.৪ অনুশীলনী

1. ধরা যাক $X_1, X_2, \ldots\ldots\ldots, X_n$ অনধীন সম্ভবনাশ্রয়ী বারন্যোলি চল। এগুলো এমন যে $P(X_i = 1) = p$ ও $P(X_j = 0) = 1 - p$, $0 < p < 1$, $i = 1(1)n$.
   $S = X_1, X_2, \ldots\ldots\ldots, X_n$
   দেখাতে হবে যে $\frac{S}{n}$, p-র একটা পক্ষপাতশূন্য প্রাককলক। $\frac{S}{n}$ এর ভেদমান ও সমক ভ্রান্তি কত হবে?
2. ধরা যাক পূর্ণকাঙ্ক λ যুক্ত একটা পোয়াসঁ নিবেশন থেকে সম্ভবনাশ্রয়ী নমুনা $(X_1, X_2, \ldots\ldots\ldots, X_n)$ নেওয়া হল। λ-র জন্য একটা প্রাককলক প্রস্তাব করুন। এবং তা পক্ষপাতশূন্য কিনা বিচার করুন।
3. ধরা যাক পূর্ণকাঙ্ক λ যুক্ত একটা পোয়াসঁ নিবেশন থেকে সম্ভবনাশ্রয়ী নমুনা X নেওয়া হল। $E\{X(X-1)\}$ নির্ণয় করে λ^2-এর একটা পক্ষপাতশূন্য প্রাককলক প্রস্তাব করুন।

4. একটা বাইনোমিয়াল (m, p) নিবেশনের $E\{X(X-1)\}$ নির্ণয় করে, তা থেকে একটা মাত্র সম্ভবনাশ্রয়ী অবেক্ষণ X-এর ওপর নির্ভর করে, p^2-এর একটা পক্ষপাতশূন্য প্রাককলক নির্ণয় করুন।

5. একটি নর্ম্যাল (μ, σ^2) নিবেশনের μ-এর মান জানা থাকলে σ^2-এর গরিষ্ঠ আশংসামান প্রাককলক নির্ণয় করুন।।

6. একটি নর্ম্যাল (μ, σ^2) নিবেশনের σ^2-এর মান জানা থাকলে μ-এর গরিষ্ঠ আশংসামান প্রাককলক নির্ণয় করুন।

7. গরিষ্ঠ আশংসামান প্রাককলন পদ্ধতিটি ব্যাখ্যা করুন। গরিষ্ঠ আশংসামান প্রাককলকের প্রধান ধর্মগুলি উল্লেখ করুন।

8. পক্ষপাতহীন, সমঞ্জস ও দক্ষ প্রাককলকের সংজ্ঞা উদাহরণসহ লিখুন।

9. প্রাককলন, প্রাককলক ও প্রাককলিত মান কাকে বলে উদাহরণ দিয়ে বোঝান। বিন্দু প্রাককলন বলতে কী বোঝেন?

৪৬.৫ গ্রন্থপঞ্জী

১। রাশিবিজ্ঞানের মূলতত্ত্ব—প্রথম ও দ্বিতীয় খণ্ড (1976)
ডঃ শৈলেশভূষণ চৌধুরী, ডঃ অরিজিৎ চৌধুরী, ডঃ বিশ্বনাথ দাস (পশ্চিমবঙ্গ রাজ্য পুস্তক পর্যৎ)

২। রাশিবিজ্ঞানের প্রয়োগ পদ্ধতি— —do—

৩। Fundamentals of Statistics–Vols. I(1998) & II(2001) Goon, A. M.; Gupta, M. K & Dasgupta, (World Press)

৪। An Outline of Statistical Theory–Vols I & II (1998)
–do–

একক ৪৭ ❑ অন্তর প্রাককলন

গঠন

৪৭.০ উদ্দেশ্য

এই অধ্যায় পড়ার পর আমরা নিম্নোক্ত বিষয় সম্বন্ধে জানতে সমর্থ হব।

(ক) অন্তর প্রাককলনের প্রণালী।

(খ) নমুনা রাশিতথ্য অবলম্বন করে গড়, ভেদমান, অনুপাতের আস্থা সীমা নির্ণয়।

৪৭.১ প্রস্তাবনা

আমরা দেখেছি যে একটা বিন্দু প্রাককলক, সমসম্ভব নমুনার দ্বারা যথাযোগ্যভাবে মনোনীত বিন্দু প্রাক-কলকের মান, অজানা পূর্ণকাঙ্কের প্রাককলিত মান নির্ণয় করে। স্পষ্টতই আমরা আশা করতে পারি যে প্রাককলিত মান পূর্ণকাঙ্কের সঠিক মানের সমান হবে না। যখন পূর্ণকাঙ্ক θ-এর বিন্দু প্রাককলক $\hat{\theta}$, তখন নমুনা-ভ্রান্তি $\hat{\theta} - \theta \neq 0$ এবং বিন্দু প্রাককলক খুব সম্ভবত সঠিক হবে না। বিন্দু প্রাককলকের এই সমস্যাকে দুর করার জন্য প্রচলিতভাবে আমরা পূর্ণকাঙ্ক θç-এর জন্য একটা প্রসার দিতে পারি এবং সেই প্রসারের মধ্যে সঙ্গত কারণে পূর্ণকাঙ্কের সঠিক মান থাকার একটা নির্দিষ্ট উচ্চ সম্ভাবনা থাকবে। এই সকল প্রসারকে বলা হয় আস্থা-অন্তর।

৪৭.২ নমুনাজ বিভাজন সংক্রান্ত কয়েকটি উল্লেখযোগ্য ফলাফল

(ক) যদি $X_1, X_2, \ldots\ldots, X_n$ নর্ম্যাল নিবেশন $N(\mu, \sigma^2)$ থেকে গৃহীত একটা সমসম্ভব নমুনা হয়, তখন—

I. $\bar{X}$-এর নিবেশন নর্ম্যাল $N(\mu, \sigma^2/n)$ হবে।

II. $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$-এর নিবেশন নর্ম্যাল $N(0, 1)$ হবে।

III. $\bar{X}$-এবং s^2 এর নিবেশন পরস্পর নিরপেক্ষ হবে যেখানে $s^2 = \frac{1}{x-1} \sum_i (X_i - \bar{X})^2$

IV. $\frac{(n-1)s^2}{\sigma^2}$ এর নিবেশন $(n - 1)$ স্বাতন্ত্র্যমাত্রা যুক্ত χ^2 – নিবেশন হবে।

V. $\frac{\bar{X}-\mu}{s/\sqrt{n}}$ এর নিবেশন $(n - 1)$ স্বাতন্ত্র্যমাত্রা যুক্ত t-নিবেশন হবে।

VI. $\frac{\sum_i (X_i - \mu)^2}{\sigma^2}$ এর নিবেশন n স্বাতন্ত্র্যমাত্রা যুক্ত χ^2 নিবেশন হবে।

যদি X এবং Y-এর নিবেশন যথাক্রমে n_1 ও n_2 স্বাতন্ত্র্য মাত্রাযুক্ত χ^2-নিবেশন হয়, তখন X + Y-এর নিবেশন $n_1 + n_2$ স্বাতন্ত্র্যমাত্রা যুক্ত χ^2 নিবেশন হবে, যেখানে X এবং Y পরস্পর নিরপেক্ষ।

(গ) যদি $X_1, X_2, \ldots\ldots, X_{n_1}$ নর্ম্যাল নিবেশন $N(\mu_1, \sigma_1^2)$ থেকে গৃহীত একটি সমসম্ভব নমুনা হয় এবং $Y_1, Y_2, \ldots\ldots, Y_{n_2}$ অপর একটি নর্ম্যাল নিবেশন $N(\mu_2, \sigma_2^2)$ থেকে গৃহীত আর একটি সমসম্ভব নমুনা হয় (যেখানে নিবেদন দুটি পরস্পর নিরপেক্ষ) তখন

I. $\dfrac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}$ এর নিবেশন N(0,1) হবে।

II. যদি $\sigma_1{}^2 = \sigma_2{}^2 = \sigma^2$ হয়, তখন

$\dfrac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sigma\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}$ এর নিবেশন N(0, 1) হবে।

III. σ^2-এর মান যদি অজানা থাকে,

$\dfrac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{s\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}$ এর নিবেশন $n_1 + n_2 - 2$ স্বাতন্ত্র্যমাত্রা যুক্ত t–নিবেশন হবে যেখানে

$$s^2 = \frac{\sum_{i=1}^{n_1}(X_i-\overline{X})^2+\sum_{i=1}^{n_2}(Y_i-\overline{Y})^2}{n_1+n_2-2}$$

IV. $\dfrac{\sum_{i=1}^{n_1}(X_i-\mu_1)^2/n_1\sigma_1^2}{\sum_{i=1}^{n_2}(y_i-\mu_2)^2/n_2\sigma_2^2}$ এর নিবেশন n_1, n_2 স্বাতন্ত্র্যমাত্রা যুক্ত F নিবেশন হবে।

V. $\dfrac{\sum_{i=1}^{n_1}(X_i-\overline{X})^2/n_1\sigma_1^2}{\sum_{i=1}^{n_2}(Y_i-\overline{Y})^2/n_2\sigma_2^2}$ এর নিবেশন $\overline{n_1-1}$, $\overline{n_2-1}$ স্বাতন্ত্র্যমাত্রা যুক্ত F নিবেশন হবে।

৪৭.৩ অন্তর প্রাককলন

ধরা যাক $X_1, X_2, \ldots\ldots\ldots, X_n$ nটি সম্ভসম্ভব নমুনা একটি পূর্ণক থেকে নেওয়া হয়েছে, যার সম্ভাবনা ঘনত্ব (ভড়) অপেক্ষক $f_\theta(.)$, যেখানে θ অজানা পূর্ণকাঙ্ক। θ -এর অন্তর প্রাককলক বার করার জন্য আমরা দুটি যথোচিত নমুনাঙ্ক $T_1 = T_1(X_1 X_2 \ldots\ldots\ldots X_n)$ এবং $T_2 = T_2(X_1 X_2 \ldots\ldots\ldots X_n)$ খুঁজে বের করব।

যেখানে T_1, T_2 তাদের অন্তর্গত প্রসার (T_1, T_2) এর মধ্যে পূর্ণক θ এর সঠিক মান (তা যাই হোক না কেন) অন্তর্ভুক্ত রাখার সম্ভাবনা $1 - \alpha$, [α সাধারণত 0.01 কিংবা 0.05 নেওয়া হয়]

৪৭.৩.১ আস্থা-সীমা

আমরা নিম্নলিখিতভাবে আস্থা-সীমার বিধিবৎ সংজ্ঞা দিতে পারি :

সংজ্ঞা : কোনও পূর্ণকের পূর্ণকাঙ্ক θ এর $100(1 - \alpha)\%$ আস্থা-সীমা হল একটি সমসম্ভব অন্তর $[T_1, T_2]$ যার জন্য

$$P [T_1 \leq \theta \leq T_2) = 1 - \alpha$$

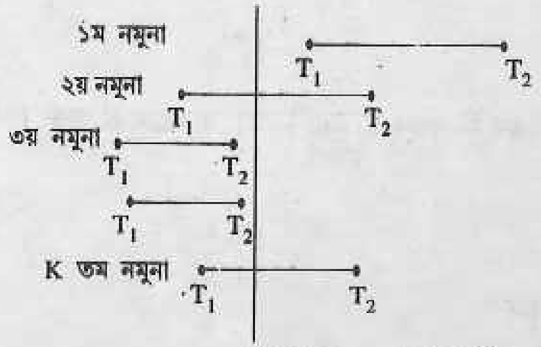
θ এর মান যাই হোক না কেন।

T_1-কে বলা হয় অধঃ আস্থা-সীমা।

T_2-কে বলা হয় ঊর্ধ্ব আস্থা-সীমা।

$(1 - \alpha)$-কে বলা হয় আস্থা-অঙ্ক।

উপরোক্ত সংজ্ঞার সম্ভাবনা বিবৃতিকে অবশ্যই সাবধানে ব্যাখ্যা করতে হবে। সংজ্ঞা থেকে এটা কখনই বোঝা যায় না যে পূর্ণকাঙ্ক θ, অন্তর $[T_1, T_2]$ এর মধ্যে থাকার সম্ভাবনা $1 - \alpha$, বাস্তবিকই এখানে পূর্ণকাঙ্ক θ একটি অজানা ধ্রুবক। কিন্তু অন্তর $[T_1, T_2]$ পরিবর্তনীয়। বিভিন্ন সমসম্ভব নমুনালব্ধ অবেক্ষণ সমুদয় $X_1, X_2, \ldots\ldots, X_n$ এর জন্য বিভিন্ন অন্তর $[T_1, T_2]$ পাওয়া যাবে। এই সমস্ত অন্তর θ–কে হয় অন্তর্ভুক্ত করবে, নয় করবে না। এই ধারণা নীচের চিত্র দ্বারা বিশদ ব্যাখ্যা করা হল।



θ (সঠিক অজানা মান) চিত্র—৪.১

সংজ্ঞার সম্ভাবনা বিবৃতি এটাই বোঝায় যে যদি বারংবার পূর্ণক থেকে নমুনা সংগ্রহ করা হয় তাহলে অবশেষে $100(1 - \alpha)\%$ অন্তর পূর্ণকাঙ্ক θ-কে অন্তর্ভূক্ত করবে এবং $100\,\alpha\%$ অন্তর θ-কে অন্তর্ভুক্ত করবে না।

৪৭.৩.২ আস্থা-সীমা নির্ণয়ের প্রচলিত পদ্ধতি

ধরা যাক কোন পূর্ণক থেকে $X_1, X_2, \ldots\ldots, X_n$ একটি সমসম্ভব নমুনা নেওয়া হল যেখানে পূর্ণকের সম্ভাবনা ঘনত্ব (ভর) অপেক্ষক $f_\theta(.)$ এবং পূর্ণকাঙ্ক θ।

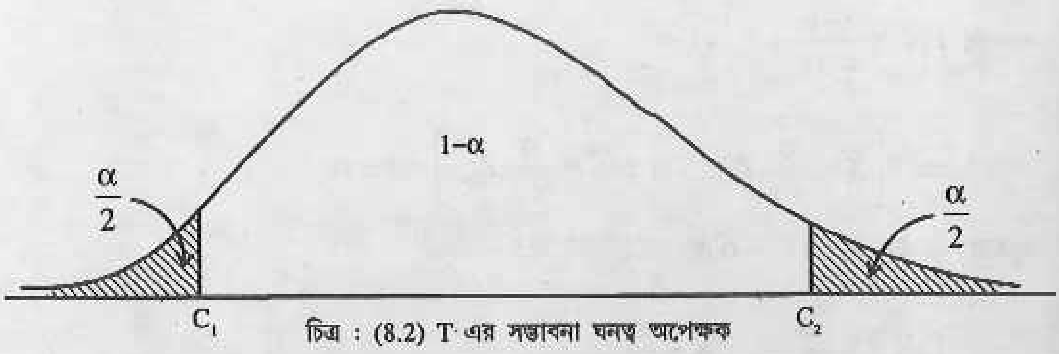
নীচের তিনটি ধাপের সাহায্যে আমরা পূর্ণকাঙ্ক θ-এর $100(1-\alpha)\%$ আস্থা-সীমা পেতে পারি।

প্রথম ধাপ : এখানে একটি সম্ভবনাশ্রয়ী চল $T(X_1 \ldots\ldots X_n)$ খুঁজে বের করতে হবে যা একটি সমসম্ভব নমুনা $X_1, X_1 \ldots X_n$ -এর অপেক্ষক এবং যার সম্ভাবনা নিবেশন θ-এর উপর নির্ভরশীল নয়। এই T-কেপর্যাপ্ত (Sufficient) নমুনা অপেক্ষক বলা হয়।

দ্বিতীয় ধাপ : দুটি ধ্রুবক C_1, C_2 খুঁজতে হবে যার জন্য

$$P(C_1 \le T \le (C_2) = 1 - \alpha$$

θ-এর মান যাই হোক না কেন এবং C_1, C_2 নিম্নোক্ত চিত্রের (8.2) মতো T-এর সম্ভাবনা নিবেশন কে বিভাজন করবে।



চিত্র : (8.2) T-এর সম্ভাবনা ঘনত্ব অপেক্ষক

তৃতীয় ধাপ ঃ যদি প্রভেদ $C_1 \le T \le C_2$-কে অবস্থান পরিবর্তন করে $T_1 \le \theta \le T_2$ করা সম্ভব হয় যেখানে T_1 এবং T_2 উভয়েই সমসম্ভব নমুনা $X_1, X_2, \ldots\ldots, X_n$ এর অপেক্ষক কিন্তু θ-এর উপর নির্ভরশীল নয়, তখন $[T_1, T_2]$ θ-এর $100(1-\alpha)\%$ কাঙ্খিত আস্থা-সীমা।

৪৭.৪ প্রামাণ্য পূর্ণকাঙ্কগুলির আস্থা-সীমা

ধরা যাক $X_1, X_2, \ldots\ldots, X_n$ n আয়তন বিশিষ্ট সমসম্ভব নমুনা নর্ম্যাল নিবেশন $N(\mu, \sigma^2)$ থেকে নেওয়া হয়েছে।

মনে করা যাক, $\bar{X} = \frac{1}{n} \sum_i X_i$ = নমুনাগড় এবং $s^2 = \frac{1}{n-1} \sum_i (X_i - \bar{X})^2$ = নমুনা ভেদমান।

৪৭.৪.১ পূর্ণক গড়ের (μ) আস্থা-সীমা

এখানে σ^2-এর মান জানা এবং অজানা এই দুটি ক্ষেত্রেই আমরা পূর্ণক গড় (μ)-এর আস্থা-সীমা বের করব।

ক্ষেত্র—I : σ^2-এর মান জানা

আমরা জানি $Z = \dfrac{\overline{X} - \mu}{\sigma / \sqrt{n}}$ এর নিবেশন N(O, 1) এবং Z-এর নিবেশন μ-এর উপর নির্ভর করছে না। সুতরাং Z-কে আমরা পর্যাপ্ত নমুনা অপেক্ষক হিসাবে নিতে পারি।

এবার ধ্রুবক C_1, C_2-কে এমনভাবে বের করতে হবে যাতে

$$P\left[C_1 \leq \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \leq C_2\right] = 1 - \alpha \text{ হয়।}$$

স্পষ্টত : $C_1 = -Z_{\alpha/2}$ এবং $C_2 = Z_{\alpha/2}$ [$Z_{\alpha/2}$, N(0,1) নিবেশনের ঊর্ধ্ব $\alpha/2$ বিন্দু এবং $-Z_{\alpha/2}$, N(0,1)-এর অধঃ $\alpha/2$ বিন্দু।]

আবার $P\left[C_1 \leq \dfrac{\overline{X} - \mu}{\sigma / \sqrt{n}} \leq C_2\right] = 1 - \alpha$

$$\Rightarrow P\left[\overline{X} - \frac{\sigma}{\sqrt{n}} Z_{\alpha/2} \leq \mu \leq \overline{X} + \frac{\sigma}{\sqrt{n}} Z_{\alpha/2}\right] = 1 - \alpha$$

সুতরাং μ-এর 100 (1 − α)% আস্থাসীমা হবে

$$I = \left[\overline{X} - \frac{\sigma}{\sqrt{n}} Z_{\alpha/2}, \overline{X} + \frac{\sigma}{\sqrt{n}} Z_{\alpha/2}\right]$$

এবং আস্থা-সীমার দৈর্ঘ্য $= \dfrac{2\sigma}{\sqrt{n}} Z_{\alpha/2}$

ক্ষেত্র-II : σ^2-এর মান অজানা

আমরা জানি $T = \dfrac{\overline{X} - \mu}{s / \sqrt{n}}$ এর নিবেশন (n − 1) স্বাতন্ত্র্য মাত্রাযুক্ত t নিবেশন এবং T-এর নিবেশন μ-এর উপর নির্ভর করছে না। সুতরাং T-কে আমরা পর্যাপ্ত অপেক্ষক হিসাবে নিতে পারি, এবার ধ্রুবক C_1, C_2-কে এমনভাবে বের করতে হবে যাতে

$$P\left[C_1 \leq \frac{\overline{X} - \mu}{s / \sqrt{n}} \leq C_2\right] = 1 - \alpha \text{ হয়}$$

স্পষ্টত : $C_1 = -t_{\alpha/2}, n-1$ এবং $C_1 = t_{\alpha/2}$ $n-1$ যেখানে $t_{\alpha/2}$, $(n-1)$ স্বাতন্ত্র্যমাত্রা যুক্ত t নিবেশনের ঊর্ধ্ব $\frac{\alpha}{2}$ —বিন্দু এবং $-t_{\alpha/2}$, $\overline{n-1}$ স্বাতন্ত্র্যমাত্রা যুক্ত t নিবেশনের অধঃ $\frac{\alpha}{2}$ —বিন্দু।

আবার $P\left[C_1 \le \frac{\overline{X}-\mu}{s/\sqrt{n}} \le C_2\right] = 1-\alpha$

$$\Rightarrow P\left[\overline{X} - \frac{s}{\sqrt{n}} t_{\frac{\alpha}{2}} \le \mu \le \overline{X} + \frac{s}{\sqrt{n}} t_{\frac{\alpha}{2}}\right] = 1-\alpha$$

সুতরাং এ ক্ষেত্রে μ-এর $100(1-\alpha)\%$ আস্থা-সীমা হবে

$$I = \left[\overline{X} - \frac{s}{\sqrt{n}} t_{\alpha/2}, \overline{X} + \frac{s}{\sqrt{n}} t_{\alpha/2}\right].$$

এবং আস্থা-সীমার দৈর্ঘ্য $= \frac{2s}{\sqrt{n}} t_{\alpha/2}$

৪৭.৪.২ পূর্ণক ভেদমান (σ^2)-এর আস্থা-সীমা

μ-এর মতো এখানেও আমরা দুটি ক্ষেত্রে σ^2-এর আস্থা-সীমা নির্ণয় করব যখন μ-এর মান জানা বা অজানা।

ক্ষেত্র-I : μ-এর মান জানা

আমরা জানি যদি μ জানা থাকে $\chi^2 = \sum_i (X_i - \mu)^2 / \sigma^2$ এর নিবেশন n স্বাতন্ত্র্যমাত্রা যুক্ত χ^2-নিবেশন হবে, এবং χ^2-এর নিবেশন σ^2-এর উপর নির্ভর করছে না। সুতরাং χ^2-কে আমরা পর্যাপ্ত অপেক্ষক হিসাবে নিতে পারি। ধ্রুবক C_1, C_2 এমনভাবে বের করতে হবে যাতে

$$P\left[C_1 \le \frac{\sum_i (X_i-\mu)^2}{\sigma^2} \le C_2\right] = 1-\alpha \text{ হয়।}$$

$$\Rightarrow P\left[\frac{\sum_i (X_i-\mu)^2}{C_2} \le \sigma^2 \le \frac{\sum_i (X_i-\mu)^2}{C_1}\right] = 1-\alpha$$

স্পষ্টত $C_1 = \chi^2_{1-\alpha/2,n}$, $C_2 = \chi^2_{\alpha/2,n}$

সুতরাং σ^2-এর $100(1-\alpha)\%$ আস্থা-সীমা হবে

$$I = \left[\frac{\sum_i (X_i-\mu)^2}{\chi^2_{\alpha/2,n}}, \frac{\sum_i (X_i-\mu)^2}{\chi^2_{1-\alpha/2,n}}\right]$$

ক্ষেত্র-II : μ-এর মান অজানা

আমরা জানি $\chi^2 = \sum_i (X_i - \overline{X})^2 / \sigma^2$ এর নিবেশন $n-1$ স্বাতন্ত্র্যমাত্রাযুক্ত χ^2 নিবেশন এবং χ^2 -এর নিবেশন σ^2-এর উপর নির্ভর করছে না। সুতরাং χ^2-কে আমরা পর্যাপ্ত নমুনা অপেক্ষক হিসাবে নিতে পারি। ধ্রুবক C_1, C_2-কে আমরা এমনভাবে নিতে পারি যাতে

$$P\left[C_1 \le \frac{\sum_i (X_i - \overline{X})^2}{\sigma^2} \le C_2\right] = 1-\alpha$$

$$\Rightarrow P\left[\frac{\sum_i (X_i - \overline{X})^2}{C_2} \le \sigma^2 \le \frac{\sum_i (X_i - \overline{X})^2}{C_1}\right] = 1-\alpha$$

স্পষ্টত : $C_1 = \chi^2_{1-\alpha/2,\, n-1}$ এবং $C_2 = \chi^2_{\alpha/2,\, n-1}$

সুতরাং σ^2 এর $100\,(1-\alpha)\%$ আস্থা-সীমা হবে

$$I = \left[\frac{\sum_i (X_i - \overline{X})^2}{\chi^2_{\alpha/2,\, n-1}}, \frac{\sum_i (X_i - \overline{X})^2}{\chi^2_{1-\alpha/2,\, n-1}}\right]$$

৪৭.৪.৩ পূর্ণক-অনুপাতের (P) আস্থা-সীমা

ধরা যাক, পূর্ণক থেকে nটি সমসম্ভব নমুনা নেওয়া হল যেখানে Xটি নমুনা একটি নির্দিষ্ট বৈশিষ্ট্যমূলক তখন নমুনাঙ্ক X-এর নিবেশন হবে দ্বিপদ নিবেশন যার পূর্ণকাঙ্কগুলি n এবং P। আমরা জানি $\frac{X}{n}$-এর নিবেশনটি প্রায় সঠিক নর্ম্যাল নিবেশন $N\left(P, \frac{P(1-P)}{n}\right)$। সুতরাং $\frac{\frac{X}{n}-P}{\sqrt{\frac{P(1-P)}{n}}}$ -এর নিবেশনও প্রায় সঠিক নর্ম্যাল $N(0,1)$ নিবেশন হবে যখন n যথেষ্ট বড়। অতএব যখন n যথেষ্ট বড়, P-এর আস্থা-সীমা বের করার জন্য $\frac{\frac{X}{n}-P}{\sqrt{\frac{P(1-P)}{n}}}$-কে মুখ্য অপেক্ষক হিসাবে নিতে পারি।

আবার,

$$P\left[-Z_{\alpha/2} \le \frac{\frac{X}{n}-P}{\sqrt{\frac{P(1-P)}{n}}} \le Z_{\alpha/2}\right] = 1-\alpha$$

যেহেতু ওপরের মন্তব্য থেকে P-এর আস্থা-সীমা বের করতে গেলে দ্বিপদ সমীকরণের সমাধান করতে হবে, যেটা সহজ হবে না। সেহেতু আমরা মুখ্য অপেক্ষকের হরে P-এর পরিবর্তে $\hat{P} = \frac{X}{n}$ ব্যবহার করব। সুতরাং উপরের মন্তব্যকে সরলীকরণ করে P-এর 100 (1 – α)% আস্থা-সীমা পাই

$$I = \left[\hat{P} - Z_{\alpha/2}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}};\ \hat{P} + Z_{\alpha/2}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}\right]$$

৪৭.৫ সারাংশ

এই এককে নিম্নলিখিত বিষয়গুলি সম্বন্ধে আমরা জানতে পারলাম—

I. অন্তর প্রাককলনের পদ্ধতি;

II. প্রামাণ্য পূর্ণকাঙ্কগুলি (পূর্ণক-গড়, পূর্ণক-ভেদমান ও পূর্ণক অনুপাত)-র আস্থা-সীমা নির্ণয়।

৪৭.৬ অনুশীলনী

১। একটি শস্যচূর্ণনের কারখানার 16টি গুঁড়ো করার বলের নির্দিষ্ট সময়ের পর দেখা গেল ওজন হ্রাস 324 গ্রাম। পূর্ব অভিজ্ঞতা থেকে জানা আছে ওজন হ্রাসের সমক বিচ্যুতি 0.68 গ্রাম। সঠিক গড় ওজন হ্রাসের 99% আস্থা-সীমা নির্ণয় করুন।

২। ভারী কাজের সময় নাড়ির স্পন্দন বৃদ্ধি বিভিন্ন শ্রমিকের ক্ষেত্রে বিভিন্ন রকম হতে পারে। 16 জন শ্রমিককে ভারী কাজ করতে দেবার পর দেখা গেল প্রতি মিনিটে গড়ে 16.5 বার স্পন্দন বেড়ে গেছে এবং স্পন্দনের ভেদমান 16.81। এই রাশিতথ্য ব্যবহার করে স্পন্দন বৃদ্ধির 95% আস্থা-সীমা বের করুন। স্পন্দনের ভেদমানেরও 95% আস্থা-সীমা নির্ণয় করুন।

৩। একটি ইঞ্জিনের 6টি বিভিন্ন মডেলকে পরীক্ষামূলকভাবে চালিয়ে দেখা গেল যে 1 গ্যালন জ্বালানীতে মেশিনগুলি যথাক্রমে 24, 28, 21, 23, 32 এবং 22 মি. চলে। ইঞ্জিনের 1 গ্যালন জ্বালানীতে গড় চলার সময়ের 95% আস্থা-সীমা বের করুন।

৪। একটি শহরের 300টি দুর্ঘটনার সমসম্ভব নমুনা থেকে দেখা গেল যে খারাপ রাস্তার কারণে 120টি দুর্ঘটনা ঘটেছে। রাস্তা খারাপ হবার কারণে দুর্ঘটনার অনুপাতের 95% আস্থা-সীমা নির্ণয় করুন।

৪৭.৭ উত্তর সংকেত

১। ধরা যাক ওজন হ্রাসের পরিমাপ নর্ম্যাল $N(\mu, \sigma^2)$ নিবেশন। এখানে σ জানা আছে। সুতরাং 99% আস্থা-সীমা হবে

$$\left[\overline{X}-Z_{.005}\frac{6}{\sqrt{n}},\ \overline{X}+Z_{.005}\frac{\sigma}{\sqrt{n}}\right]$$

দেওয়া আছে $\overline{X} = 324,\ \sigma = 0.68,\ n = 16$

সুতরাং আস্থা-সীমা

$$\left[324-Z_{0.005}\frac{0.68}{\sqrt{16}}, 324+Z_{.005}\frac{0.68}{\sqrt{16}}\right]$$

২। ধরা যাক, নাড়ির স্পন্দন বৃদ্ধির নিবেশন নর্ম্যাল $N(\mu, \sigma^2)$ নিবেশন। যেহেতু এখানে σ^2 অজানা, স্পন্দন বৃদ্ধির 95% আস্থা-সীমা।

$$\left[\overline{X}-t_{0.025;n-1}\sqrt{\frac{s^2}{n}}, \overline{X}+t_{0.025,n-1}\sqrt{\frac{s^2}{n}}\right]$$

দেওয়া আছে $\overline{X} = 16.5,\ s^2 = 16.81,\ n = 16$

সুতরাং আস্থাসীমা হবে

$$\left[16.5-t_{0.025,15}\sqrt{\frac{16.81}{16}}, 16.5+t_{0.025,15}\sqrt{\frac{16.81}{16}}\right]$$

এবং স্পন্দন বৃদ্ধির ভেদমানের 95% আস্থা-সীমা হবে

$$\left[\frac{(n-1)s^2}{\chi^2_{0.025,n-1}}, \frac{(n-1)s^2}{\chi^2_{0.975,n-1}}\right]$$

দেওয়া আছে $s^2 = 16.81,\ n = 16$

সুতরাং আস্থা-সীমা

$$\left[\frac{(15\times16.81)}{\chi^2_{0.025,15}}, \frac{(15\times16.81}{\chi^2_{0.975,15}}\right]$$

৩। ধরা যাক, মেশিনের চলার সময়ের নিবেশন নর্ম্যাল $N(\mu, \sigma^2)$ এখানে যেহেতু σ^2 অজানা, 99% আস্থা-সীমা হবে

$$\left[\overline{X}-t_{.005,n-1}\frac{s}{\sqrt{n}},\ \overline{X}+t_{.005,n-1}\frac{s}{\sqrt{n}}\right]$$

এখানে রাশিতথ্য থেকে পাই $\overline{X}=25$ এবং $s^2 = 17.6,\ n = 6$

সুতরাং 99% আস্থা-সীমা হবে

$$\left[25-t_{0.005,5}\sqrt{\frac{17.6}{6}}, 25+t_{0.005,5}\sqrt{\frac{17.6}{6}}\right]$$

৪। দুর্ঘটনা অনুপাত (P)-এর 95% আস্থা-সীমা হবে

$$\left[\hat{P}-Z_{0.025}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}, \hat{P}+Z_{0.025}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}\right]$$

দেওয়া আছে X = 120 = দুর্ঘটনার সংখ্যা

$$n = 300$$

$$\hat{P}=\frac{120}{300}=0.4$$

সুতরাং P-এর 95% আস্থা-সীমা হবে

$$\left[0.4-Z_{0.025}\sqrt{\frac{\cdot 4(1-\cdot 4)}{300}}, 0.4+Z_{0.025}\sqrt{\frac{\cdot 4(1-\cdot 4)}{300}}\right]$$

৪৭.৮ গ্রন্থপঞ্জী

১। রাশিবিজ্ঞানের মূলতত্ত্ব—প্রথম ও দ্বিতীয় খণ্ড (১৯৭৬)
ডঃ শৈলেশ ভূষণ চৌধুরী, ডঃ অরিজিৎ চৌধুরী
ডঃ বিশ্বনাথ দাস (পশ্চিমবঙ্গ রাজ্য পুস্তক পর্ষৎ)

২। রাশিবিজ্ঞানের প্রয়োগ পদ্ধতি—do–

৩। Fundamentals of Statistics–Vol I (1998) & II (2001)
Goon, A. M; Gupta, M. K & Dasgupta, B, World Press

৪। An Outline of Statistical Theory–Vol I & II (1998)
–do–

একক ৪৮ ❑ প্রকল্প বিচার

গঠন

৪৮.০ উদ্দেশ্য

এই একক পড়ার পর আমরা নিম্নোক্ত বিষয় সম্বন্ধে জানতে সমর্থ হব।

I প্রকল্প বিচার সম্বন্ধীয় মৌলিক ধারণা।

II সাধারণ পরিসংখ্যানগত বিচার নির্ণয়।

III সাধারণ পরিসংখ্যানগত প্রকল্পের উপর প্রকল্প বিচারের ধারণার প্রয়োগ।

IV নর্ম্যাল, χ^2, t, এবং F প্রভৃতি প্রচলিত পরিসংখ্যানগত বিচারের ব্যবহার।

V বিভিন্ন পরিস্থিতিতে নমুনা রাশিতথ্যের উপর প্রচলিত পরিসংখ্যানগত বিচারের প্রয়োগ।

৪৮.১ প্রস্তাবনা

আমরা ইতিপূর্বে জেনেছি যে জানা নমুনা থেকে অজানা পূর্ণকের সম্বন্ধে ধারণা করার জন্য যে বিজ্ঞানসম্মত তত্ত্ব ব্যবহার করা হয় তাকে রাশি বিজ্ঞানভিত্তিক অনুমানতত্ত্ব বলা হয়। রাশি বিজ্ঞানভিত্তিক অনুমানতত্ত্ব প্রধানত দুই প্রকার—

(ক) প্রাককলন তত্ত্ব, (পূর্ববর্তী দুটি এককে বিস্তারিত আলোচনা হয়েছে)।

(খ) প্রকল্প বিচার তত্ত্ব।

অনেক সময় রাশি বিজ্ঞানভিত্তিক সিদ্ধান্ত সংক্রান্ত সমস্যার উদ্ভব হয় যখন আমরা পূর্ণকাঙ্ক θ-এর যথাযথ প্রাককলন সম্বন্ধে আগ্রহী নই, কিন্তু নমুনা রাশিতথ্যের ওপর নির্ভর করে জানার পূর্বে ধারণা করা পূর্ণকাঙ্ক θ সম্বন্ধীয় পূর্ব ধারণাকে যাচাই করতে চাই। আমাদের নমুনা রাশি ব্যবহার করে দেখতে হবে θ সম্বন্ধে পূর্ব ধারণাকে আমরা গ্রহণ করব না বর্জন করব। যে তত্ত্বের উপর নির্ভর করে এই সিদ্ধান্ত নেওয়া হয় তাকে বলা হয় প্রকল্প বিচার তত্ত্ব। এবং এই পদ্ধতিকে বলা হয় প্রকল্প বিচার পদ্ধতি। আমরা এই এককে তত্ত্ব সম্বন্ধে খুব বেশী আলোচনা করব না কেবলমাত্র কয়েকটি প্রচলিত পরিসংখ্যানগত বিচার পদ্ধতি নিয়ে বিশদ আলোচনা করব।

৪৮.২ প্রকল্প বিচার

৪৮.২.১ পরিসংখ্যানগত প্রকল্প সম্বন্ধীয় বিবৃতি

সংজ্ঞা ঃ পূর্ণকাঙ্ক সম্বন্ধীয় কোনও পূর্ব ধারণা বা মন্তব্যকে প্রকল্প বলে। একটি উদাহরণ থেকে আমরা প্রকল্প সম্বন্ধে পরিষ্কার ধারণা পেতে পারি—

ধরা যাক একটি দুগ্ধ প্রস্তুতকারী সংস্থা 500 মিলি-এর দুগ্ধ বোতল স্বয়ংক্রিয় যন্ত্র দ্বারা ভর্তি করে। যদি যন্ত্রের কার্যকারিতা ঠিক থাকে, তাহলে সকল বোতলে দুগ্ধের পরিমাণ সঠিক 500 মিলিই হবে। এতদনুসারে, θ যদি সঠিক গড় দুগ্ধ ভর্তির পরিমাণ হয়, সংস্থাটি ধারণা করতে পারে $\theta = 500$ মিলি। সংস্থাটির এই জানার পূর্বে ধারণা করা মতকে যাচাই করার প্রয়োজন হতে পারে। প্রথমত : $\theta = 500$ মিলি, দ্বিতীয়ত : $\theta \neq 500$ মিলি। প্রথম θ সম্বন্ধীয় বিবৃতিকে বলা হয় মুখ্যপ্রকল্প এবং একে H_0 দ্বারা চিহ্নিত করা হয়। দ্বিতীয় প্রকার বিবৃতিকে বলা হয় বৈকল্পিক প্রকল্প এবং একে H_1 দ্বারা চিহ্নিত করা হয়। সমসম্ভব নমুনার ভিত্তিতে মুখ্য প্রকল্প বিচার করে দেখা হয় সেই প্রকল্প গ্রহণযোগ্য কি না। মুখ্য প্রকল্প যদি গ্রহণযোগ্য না হয়, তাহলে বৈকল্পিক প্রকল্প গ্রহণযোগ্য হবে, যে পদ্ধতির সাহায্যে নমুনার উপর নির্ভর করে মুখ্য প্রকল্প/বৈকল্পিক প্রকল্প-এর গ্রহণযোগ্যতা বিচার করা হয়, তাকেই প্রকল্প বিচার পদ্ধতি বলে।

সাধারণত পূর্ণকাঙ্ক θ-এর মান অজানা থাকে। θ একমাত্রিক কিংবা বহুমাত্রিক হতে পারে। θ-এর সমস্ত মান পূর্ণকাঙ্কদেশ Θ থেকে গৃহীত হয়। ধরা যাক Θ_0, Θ-এর একটি সাবসেট। তখন আমরা মুখ্য প্রকল্পকে $H_0 : \theta \epsilon \Theta_0$ এভাবে বর্ণনা করতে পারি, এবং বৈকল্পিক প্রকল্প $H_1 : \theta \epsilon \bar{\Theta}_0$, যেখানে $\bar{\Theta}_0 = \Theta - \Theta_0$। সাধারণত Θ সমস্ত বাস্তব সংখ্যার সেট। Θ_0 একটি বিন্দু সেট $\theta = \theta_0$ হতে পারে, কিংবা $\theta \leq \theta_0$ বা $\theta \geq \theta_0$ এই রকম অসমীকরণ হতে পারে। আবার একই রকমভাবে Θ_1 একটি বিন্দু সেট $\theta = \theta_1$ হতে পারে, কিংবা $\theta \leq \theta_0$ বা $\theta \geq \theta_0$ এই রকমও হতে পারে, কিংবা $\theta \neq \theta_0$ ও হতে পারে। যখন $\theta \leq \theta_0$ বা $\theta > \theta_0$ বা, $\theta \neq \theta_0$ তখন প্রকল্পগুলিকে যথাক্রমে বামপাক্ষিক বা দক্ষিণপাক্ষিক বা উভয়পাক্ষিক প্রকল্প বলা হয়।

৪৮.২.২ সরল ও মিশ্র প্রকল্প

প্রকল্প দুই প্রকার হতে পারে—

I. সরল প্রকল্প,

II. মিশ্র প্রকল্প।

I. যে প্রকল্প θ–কে স্বতন্ত্রভাবে উল্লেখ করে, যাতে পূর্ণকটি সম্বন্ধে সম্পূর্ণ জানা হয়ে যায় তাকে সরল প্রকল্প বলে। উদাহরণস্বরূপ নর্ম্যাল $N(\mu, \sigma^2)$ নিবেশনের, যেখানে σ^2 জানা আছে $Ho : \mu = \mu_0$ একটি সরল প্রকল্প। কিংবা $N(\mu, \sigma^2)$ নিবেশনের যেখান μ জানা আছে, $H_0 : \sigma^2 = \sigma_0^2$ একটি সরল প্রকল্প।

II. যে প্রকল্প θ–কে স্বতন্ত্রভাবে উল্লেখ করে না যার ফলে পূর্ণকটি সম্বন্ধে সম্পূর্ণ জানা হয় না, তাকে মিশ্র প্রকল্প বলে। উদাহরণস্বরূপ $N(\mu, \sigma^2)$ নিবেশনের, $H_0 : \mu \leq \mu_0$ যেখানে σ^2 মান জানা থাক বা না থাক, একটি মিশ্র প্রকল্প। আবার $N(\mu, \sigma^2)$ পূর্ণকে যদি μ ও σ^2 অজানা হয় তাহলে $H_0 : \mu = \mu_0$ একটি মিশ্র প্রকল্প।

৪৮.২.৩ পরিসংখ্যানগত বিচার

ধরা যাক X_1, X_2,.........,X_n n সংখ্যক সমসম্ভব নমুনা একটি পূর্ণক থেকে নেওয়া হল, যার গাণিতিক রূপ f_0 [θ ε Ⓗ] জানা আছে, যেখানে f_θ দ্বিপদ বিভাজনের সম্ভাবনা ভর অপেক্ষক, কিংবা পোঁয়াস বিভাজনের সম্ভাবনা ভর অপেক্ষক বা নর্ম্যাল বিভাজনের সম্ভবনা ঘনত্ব অপেক্ষক হতে পারে। সাধারণত θ–এর মান অজানা থাকে। নমুনার মধ্যে যে তথ্য আছে তাকে ব্যবহার করে আমরা দুটি সিদ্ধান্তে উপনীত হতে পারি।

1. H_0 : θ ε $Ⓗ_0$-কে সঠিক বলে গ্রহণ করতে পারি।

2. H_0-কে বর্জন করতে পারি এবং H_1 : θ ε Ⓗ-কে সঠিক বলে গ্রহণ করতে পারি।

পরিসংখ্যানগত প্রকল্প বিচার একটি নিয়ম যা আমাদের উপরিউক্ত কোনও একটি সিদ্ধান্তে উপনীত হতে সাহায্য করে। সিদ্ধান্তে উপনীত হবার জন্য পর্যবেক্ষণীয় নমুনার ওপর নির্ভর করতে হবে। পর্যবেক্ষণীয় নমুনা বিন্দু যে দেশ S-এর সৃষ্টি করে তাকে নমুনাদেশ বলে। পর্যবেক্ষণীয় নমুনা (X_1, X_2,.........,X_n) কে আমরা n-মাত্রিক কোনও দেশ (S)-এর কোনও বিন্দু হিসাবে ভাবতে পারি। প্রথমত, আমরা S-কে দু'ভাগে ভাগ করব। ধরা যাক একটি ভাগ C এবং অন্য ভাগটি $\bar{C} = S - C$। যদি পর্যবেক্ষণীয় নমুনা C-এর মধ্যে থাকে, তখন আমরা মুখ্য প্রকল্পকে বর্জন করব। আবার যদি নমুনা $\bar{C}$-এর মধ্যে থাকে, তখন আমরা মুখ্য প্রকল্পকে গ্রহণ করব এবং বৈকল্পিক প্রকল্পকে বর্জন করব। অঞ্চল C-কে বলে বর্জনাঞ্চল এবং $\bar{C}$-কে বলা হয় গ্রহণাঞ্চল। বর্জনাঞ্চলকে পর্যবেক্ষণীয় নমুনা X_1, X_2,.........,X_n-এর অপেক্ষক $T(X_1, X_2, ..., X_n)$ দ্বারা বর্ণনা করা যেতে পারে। T-কে বলা হয় বিচার নমুনাঙ্ক।

৪৮.২.৪ দুই প্রকার ভ্রান্তি

পরিসংখ্যানগত বিচার করে কোনও সিদ্ধান্তে উপনীত হতে গেলে আমরা নিম্নলিখিত দুই প্রকারের ভ্রান্তিকে এড়াতে পারি না।

1. প্রথম প্রকার ভ্রান্তি : মুখ্য প্রকল্প H_0 সত্য হলেও পর্যবেক্ষণীয় নমুনা যদি বর্জনাঞ্চল C-তে পড়ে, তখন মুখ্য প্রকল্প H_0 বর্জিত হতে পারে। সুতরাং সঠিক মুখ্য প্রকল্পকে বর্জন করার ফলে যে ভ্রান্তির সৃষ্টি হয় তাকে প্রথম প্রকার ভ্রান্তি বলে।

2. দ্বিতীয় প্রকার ভ্রান্তি : মুখ্য প্রকল্প H_0 সত্য না হলেও পর্যবেক্ষণীয় নমুনা যদি গ্রহণাঞ্চল $\bar{C}$-এ পড়ে তখন মুখ্য প্রকল্প গৃহীত হতে পারে। মুখ্য প্রকল্প সত্য না হলেও তাকে গ্রহণ করার ফলে যে ভ্রান্তির উৎপত্তি হয় তাকে দ্বিতীয় প্রকার ভ্রান্তি বলে। পরিসংখ্যানগত বিচার প্রয়োগ করার ফলে যে চার প্রকার পরিস্থিতির উদ্ভব হয় তা নীচের সারণিতে পরিকল্পনীয় ভাগ দেওয়া হল।

সঠিক অবস্থা / সিদ্ধান্ত	মুখ্য প্রকল্প (H_0) সঠিক	বৈকল্পিক প্রকল্প (H_1) সঠিক
মুখ্য প্রকল্প (H_0) গ্রহণ	কোনও ভ্রান্তি নেই	দ্বিতীয় প্রকার ভ্রান্তি
মুখ্য প্রকল্প (H_0) বর্জন	প্রথম প্রকার ভ্রান্তি	কোনও ভ্রান্তি নেই

আমাদের উদ্দেশ্য হওয়া উচিত দুই প্রকারের ভ্রান্তিকে এক সঙ্গে যথাসম্ভব হ্রাস করে একটি বর্জনাঞ্চলকে চিহ্নিত করা। প্রথম প্রকার ও দ্বিতীয় প্রকার ভ্রান্তির সম্ভাবনাকে যথাক্রমে α এবং β দ্বারা চিহ্নিত করা হয়। সুতরাং

$$\alpha(\theta) = P_\theta\,(H_0 \text{ বর্জন করা})$$

$$= P_\theta\,[(X_1, X_2, \ldots\ldots\ldots, X_n)\ \varepsilon\ c],\ \theta\ \varepsilon\ \text{Ⓗ}_0$$

এবং,

$$\beta(\theta) = P_\theta\,(H_0 \text{ গ্রহণ করা})$$

$$= P_\theta\,[(X_1, X_2, \ldots\ldots\ldots, X_n)\ \varepsilon\ \bar{c}],\ \theta\ \varepsilon\ \text{Ⓗ}_1$$

যেখানে $P_\theta(\cdot)$ বিবৃতি $(\cdot)$-র সম্ভাবনা যখন পূর্ণকাঙ্কের সঠিক মান θ। আদর্শ বিচার এমন হওয়া উচিত যেখানে দুই প্রকারের ভ্রান্তিই শূন্যর কাছাকাছি, যদি না শূন্য হয়। কিন্তু দুর্ভাগ্যবশত উভয় প্রকারের ভ্রান্তির সম্ভাবনাকে সাধারণত একসঙ্গে হ্রাস করা সম্ভব নয়। যদি কোনও একটির সম্ভাবনা কমতে কমতে শূন্যের দিকে যায়, তখন অন্যটির সম্ভাবনা বাড়তে বাড়তে 1-এর দিকে যায়। এটা দেখার জন্য আমরা একটা বিচার নেব যার জন্য $\alpha(\theta) = 0\ \forall\ \theta\ \varepsilon\ \text{Ⓗ}_0$। এটা তখনই সম্ভব যখন বর্জনাঞ্চল $C = \phi$ [নাল সেট]। অর্থাৎ পর্যবেক্ষণীয় নমুনা যাই হোক না কেন বিচার কখনই মুখ্য প্রকল্প H_0-কে বর্জন করবে না। তখন $\alpha(\theta) = P_\theta\,[(X_1, X_2, \ldots\ldots\ldots, X_n)\ \varepsilon\ \phi] = 0\ \forall\ \theta\ \varepsilon\ \text{Ⓗ}_0$।

কিন্তু যেহেতু গ্রহণাঞ্চল $\bar{C} = S - \phi = S$

$$\beta(\theta) = P_\theta\left[(X_1, X_2 \ldots\ldots\ldots\ldots X_n)\ \varepsilon\ S\right] = 1\ \forall \theta \varepsilon\ \text{Ⓗ}_1$$

একইরকমভাবে পর্যবেক্ষণীয় নমুনা যাই হোক না কেন যদি একটি বিচার মনোনীত করা হয় যা সর্বদা মুখ্য প্রকল্প (H_0)-কে বর্জন করে অর্থাৎ বর্জনাঞ্চল $C = S$, তখন $\alpha(\theta) = 1$, পক্ষান্তরে $\beta(\theta) = 0$.

$\alpha(\theta)$ এবং $\beta(\theta)$ উভয়কে যেহেতু একসঙ্গে কমানো সম্ভব নয়, প্রথমে $\alpha(\theta)$-এর ঊর্ধ্বসীমা $\alpha[\alpha(\theta) \le \alpha,\ \forall\ \theta\ \varepsilon\ \text{Ⓗ}_0]$-কে স্থির করে নেওয়া হয়। α-কে সাধারণত .01 বা .05 নেওয়া হয়। α-কে বলা হয় বিচারের অথবা বর্জনাঞ্চল C-এর সংশয়মাত্রা। সংশয়মাত্রা স্থির করে নেবার পর, α সংশয়মাত্রা যুক্ত বিচারগুলির মধ্যে 'যে বিচারের দ্বিতীয় প্রকার ভ্রান্তির সম্ভাবনা $\beta(\theta),\ \forall\ \theta\ \varepsilon\ \text{Ⓗ}_1$। সর্বনিম্ন সেই বিচারটিকেই মনোনীত করব।

টীকা : সংশয়মাত্রা α-কে শতকরাভাবে 100 α% হিসাবেও লেখা যায়।

৪৮.২.৫ বিচারের শক্তি অপেক্ষক

আমরা দেখেছি যে প্রথম ও দ্বিতীয় প্রকার ভ্রান্তির সম্ভাবনা দুটি পূর্ণকাঙ্ক θ-এর অপেক্ষক। প্রথমটির ক্ষেত্রে $\theta \in \Theta_0$ এবং দ্বিতীয়টির ক্ষেত্রে $\theta \in \Theta_1$। এখন আমরা একটি অপেক্ষক উপস্থাপিত করব যার থেকে $\alpha(\theta)$ এবং $\beta(\theta)$ উভয়কেই নির্ণয় করা যায়। বিচারের শক্তি অপেক্ষক, মুখ্য প্রকল্প H_0-কে বর্জন করার সম্ভাবনা যখন পূর্ণকাঙ্কের সঠিক মান θ, $\theta \in \Theta$। শক্তি অপেক্ষককে সাধারণত $\upsilon(\theta)$ দিয়ে চিহ্নিত করা হয়। সুতরাং

$$\upsilon(\theta) = P_\theta (\text{মুখ্য প্রকল্প } H_0 \text{ বর্জন করা})$$

$$= P_\theta [(X_1, X_2, \ldots\ldots\ldots, X_n) \in C], \ \theta \in \Theta$$

সংজ্ঞা অনুসারে $\alpha(\theta)$ হচ্ছে মুখ্য প্রকল্পকে বর্জন করার সম্ভাবনা যখন মুখ্য প্রকল্প H_0 সঠিক অর্থাৎ $\theta \in \Theta_0$ সুতরাং

$$\alpha(\theta) = \upsilon(\theta) \ \forall \ \theta \in \Theta_0$$

এবং $\beta(\theta)$ হচ্ছে মুখ্য প্রকল্প H_0-কে গ্রহণ করার সম্ভাবনা যখন মুখ্য প্রকল্প H_0 সত্যি নয় অর্থাৎ $\theta \in \Theta$। সুতরাং

$$\beta(\theta) = 1 - \gamma(\theta) \ \forall \ \theta \in \Theta_1$$

তাহলে আমরা দেখলাম

$$\gamma(\theta) = \begin{cases} \alpha(\theta) & \forall \, \theta \in \Theta_0 \\ 1-\beta(\theta) & \forall \, \theta \in \Theta_1 \end{cases}$$

$\gamma(\theta)$-এর লেখচিত্রকে বিচারের শক্তি রেখা বলে। তাহলে শক্তি রেখার আকার কেমন হবে? আদর্শ শক্তি রেখা আমরা তাকেই বলব যার জন্য $\alpha(\theta) = 0$ এবং $\beta(\theta) = 0$, অর্থাৎ

$$\gamma(\theta) = \begin{cases} 0 & \forall \ \theta \ \in \Theta_0 \\ 1 & \forall \ \theta \ \in \Theta_1 \end{cases}$$

কিন্তু আমরা দেখেছি যে এই আদর্শ পরিস্থিতি কোনও বিচারের ক্ষেত্রেই অর্জন করা সম্ভব হয় না। যেটা আমরা করি তা হল $\alpha(\theta)$-কে নিয়ন্ত্রণ করি সর্বোচ্চ α পর্যন্ত যখন $\theta \in \Theta_0$ এবং $\beta(\theta)$-কে হ্রাস করি যখন $\theta \in \Theta$।

অর্থাৎ $\gamma(\theta) \le \alpha \ \forall \ \theta \in \Theta_0$ এবং $\gamma(\theta)$, $\theta \in \Theta$, যতটা সম্ভব বেশী হওয়া উচিত।

৪৮.৩ পরিসংখ্যানগত বিচার—একটি পূর্ণকের ক্ষেত্রে

এখানে আমরা পূর্ণক গড় (μ), পূর্ণক ভেদমান (σ^2) এবং পূর্ণক অনুপাত (P)-এর কিছু প্রচলিত বিচার সম্বন্ধে আলোচনা করব। নিম্নে তিন প্রকার বৈকল্পিক প্রকল্প যথা বাম পাক্ষিক, দক্ষিণ পাক্ষিক এবং উভয় পাক্ষিক এর জন্যই বিচার আলোচনা করা হল।

৪৮.৩.১ পূর্ণক গড় (μ) সম্বন্ধীয় প্রকল্প বিচার

প্রকল্প বিবৃতকরণ : এখানে আলোচ্য তিন প্রকার প্রকল্প

প্রকল্প	উদাহরণ
$H_0 : \mu = \mu_1$ I $H_1 : \mu > \mu_0$ (দক্ষিণ পাক্ষিক)	$\mu =$ নতুন উচ্চফলনশীল বীজের গড় কৃষিজাত উৎপাদন/একর। $\mu_0 =$ প্রচলিত বীজের জানা গড় কৃষিজাত উৎপাদন/একর।
II $H_0 : \mu = \mu_0$ $H_1 : \mu < \mu_0$ (বাম পাক্ষিক)	$\mu =$ দূষণরোধকারী কর্মসূচী নেবার পর গড় দূষণমাত্রা। $\mu_0 =$ দূষণমাত্রার নিরাপদ সীমা।
III $H_0 : \mu = \mu_0$ $H_1 : \mu \neq \mu_0$ (উভয় পাক্ষিক)	$\mu =$ উৎপাদিত বলবেয়ারিং-এর গড় ব্যাসার্ধ। $\mu_0 =$ গড় ব্যাসার্ধের নির্দিষ্ট মান।

(ক) বিচার নমুনাঙ্কের মনোনয়ন

প্রথমত, উপরিউক্ত তিন প্রকার প্রকল্প থেকে একটি প্রকল্পকে মনোনয়ন করতে হবে। দ্বিতীয়ত, সমসম্ভব নমুনা $X_1, X_2, \ldots\ldots\ldots, X_n$-এর উপর নির্ভর করে একটি বিচার নমুনাঙ্ক খুঁজে বের করতে হবে। এখানে $\overline{X}$-কে আমরা সঙ্গত কারণে বিচার নমুনাঙ্ক হিসাবে নিতে পারি।

(খ) বিচারের সংশয়মাত্রা নির্ধারণ

যে বর্জনাঞ্চল $\mu = \mu_0$-কে বর্জন করে, কিন্তু $\mu > \mu_0$-কে গ্রহণ করে তা অবশ্যই $\mu < \mu_0$-এর যে কোনও মানকে বর্জন করবে। একইরকমভাবে যে বর্জনাঞ্চল $\mu = \mu_0$-কে বর্জন করে $\mu < \mu_0$-এর অনুকূলে, তা অবশ্যই $\mu > \mu_0$-এর যে কোনও মানকে বর্জন করবে। বামপাক্ষিক ও দক্ষিণপাক্ষিক প্রকল্পের জন্য H_0-এর মধ্যে μ-এর সমস্ত মানের জন্য প্রথম প্রকার ভ্রান্তির সম্ভাবনা সর্বাধিক হবে যখন $\mu = \mu_0$। উপরের তিনটি প্রকল্প (I, II, III)-এর জন্য সংশয়মাত্রা যদি α হয় তাহলে নিম্নোক্ত প্রকল্পগুলির জন্য সংশয়মাত্রা α হবে।

$$\text{I}' \quad H_0 : \mu \leq \mu_0$$
$$H_1 : \mu > \mu_0$$
$$\text{II}' \quad H_0 : \mu \geq \mu_0$$
$$H_1 : \mu < \mu_0$$

সুতরাং I এবং II প্রকল্পের যা বিচার হবে, I$'$ এবং II$'$-এর বিচারও যথাক্রমে একই হবে।

(গ) বর্জনাঞ্চল নির্ধারণ

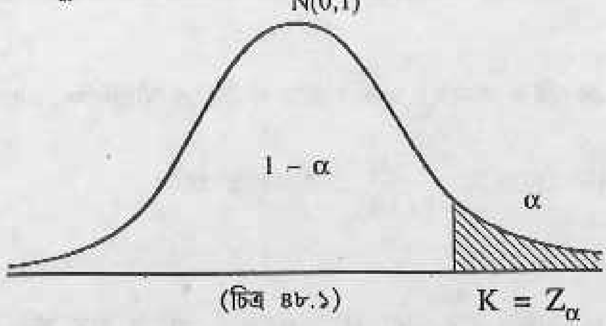
পূর্ণক নিবেশন এবং পূর্ণক ভেদমান (σ^2)-এর উপর নির্ভর করে আমরা তিনটি ক্ষেত্রে বিভিন্ন রকম বর্জনাঞ্চল নির্ধারণ করতে পারি।

ক্ষেত্র—I

ধরা যাক, পূর্ণক নিবেশনটি $N(\mu, \sigma^2)$ যেখানে σ^2 জানা আছে। যদি $X_1, X_2, \ldots\ldots\ldots, X_n$ nটি সমসম্ভব নমুনা $N(\mu, \sigma^2)$ থেকে নেওয়া হয়, তখন $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$-এর নিবেশন হবে $N(0,1)$। এখানে তিন প্রকার বৈকল্পিক প্রকল্পের জন্য বর্জনাঞ্চল বিভিন্ন প্রকার হবে। বিচার নমুনাঙ্ক হিসাবে $Z = \frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$ কে ব্যবহার করব।

(ঘ) দক্ষিণ পাক্ষিক বৈকল্পিক প্রকল্প

মুখ্য প্রকল্প H_0 কে বর্জন করা হবে যদি $Z > K$ হয়, যেখানে K একটি ধ্রুবক এবং K-কে এমনভাবে নির্ধারণ করতে হবে যাতে $\alpha(\mu_0) = P\mu_0\ (Z > K) = \alpha$ হয়। স্পষ্টত K প্রমাণ নর্ম্যাল নিবেশনের উর্ধ্ব $100\alpha\%$ বিন্দু। অর্থাৎ $K = Z_\alpha$ (চিত্র ৪৮.১)

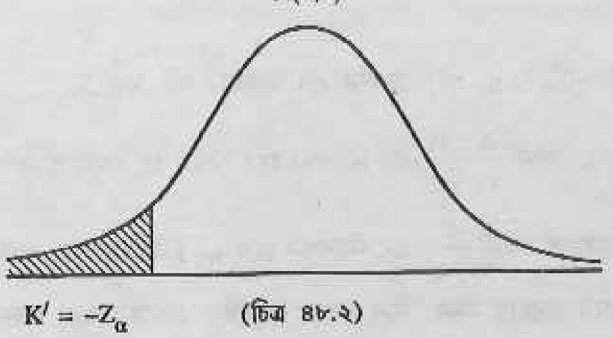


(চিত্র ৪৮.১)

(ঙ) বাম পাক্ষিক বৈকল্পিক প্রকল্প

মুখ্য প্রকল্প H_0-কে বর্জন করা হবে যদি $Z < K'$ হয়, যেখানে K' একটি ধ্রুবক এবং একে এমনভাবে নির্ধারণ করতে হবে যাতে $\alpha(\mu_0) = P\mu_0\ (Z < K') = \alpha$ হয়।

স্পষ্টত K' প্রমাণ নর্ম্যাল নিবেশনের নিম্ন $100\alpha\%$ বিন্দু। অর্থাৎ $K' = -Z_\alpha$ [চিত্র ৪৮.২]



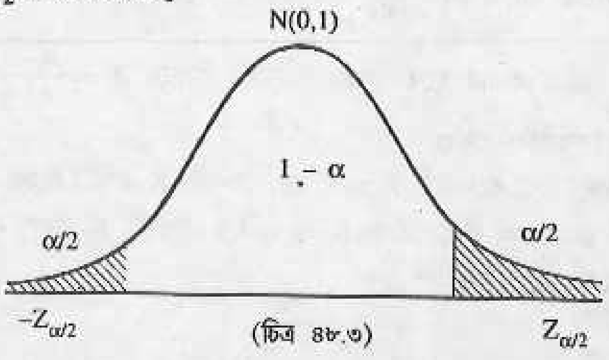
(চিত্র ৪৮.২)

(চ) উভয় পাক্ষিক বৈকল্পিক প্রকল্প

মুখ্য প্রকল্প H_0-কে বর্জন করা হবে যদি $|Z| > K^*$ হয়, যেখানে K^* একটি ধ্রবক এবং একে এমনভাবে নির্ধারণ করতে হবে যাতে $\alpha(\mu_0) = P_{\mu_0}(|Z| > K^*) = \alpha$ হয়।

স্পষ্টত K^* প্রমাণ নর্মাল নিবেশনের উর্ধ্ব $100\frac{\alpha}{2}\%$ বিন্দু।

অর্থাৎ $K^* = Z_{\alpha/2}$ [চিত্র ৪৮.৩]



(চিত্র ৪৮.৩)

উপরোক্ত তিনটি বৈকল্পিক প্রকল্পের জন্য বিচার পদ্ধতিকে নিম্নলিখিতভাবে বর্ণনা করা যেতে পারে।

1. n, সংখ্যক নমুনা থেকে $Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$ বের করতে হবে।

2. সংশয়মাত্রা $100\alpha\%$ হলে

(ক) $H_0 : \mu = \mu_0$ বর্জিত হবে এবং $H_1 : \mu > \mu_0$ গৃহীত হবে যদি $Z > Z_\alpha$ হয়।

(খ) $H_0 : \mu = \mu_0$ বর্জিত হবে এবং $H_1 : \mu < \mu_0$ গৃহীত হবে যদি $Z < -Z_\alpha$ হয়।

(গ) $H_0 : \mu = \mu_0$ বর্জিত হবে এবং $H_1 : \mu \neq \mu_0$ গৃহীত হবে যদি $|Z| > Z_{\alpha/2}$ হয়।

টীকা : প্রমাণ নর্ম্যাল নিবেশনের এই সমস্ত শতকরা বিন্দু পিয়ারসন ও হার্টলের বায়োমেট্রিক সারণির প্রথমভাগে পাওয়া যাবে।

ক্ষেত্র—II

ধরা যাক, পূর্ণক নিবেশনটি $N(\mu, \sigma^2)$ যেখানে σ^2 অজানা। যদি $X_1, X_2, \ldots\ldots, X_n$ nটি সমসম্ভব নমুনা $N(\mu, \sigma^2)$ থেকে নেওয়া হয়, তখন $\frac{\overline{X} - \mu}{s/\sqrt{n}}$ এর নিবেশন হবে $N(0, 1)$, যেখানে $s^2 = \frac{1}{n-1}\sum_i (X_i - \overline{X})^2$। মুখ্য প্রকল্প $H_0 : \mu = \mu_0$ অনুসারে $t = \frac{\overline{X} - \mu_0}{s/\sqrt{n}}$ এর নিবেশন হবে $\overline{n-1}$ স্বাতন্ত্র্যমাত্রা যুক্ত t নিবেশন। এখানে t-কে আমরা বিচার নমুনাঙ্ক হিসাবে ব্যবহার করব। তিন প্রকার বৈকল্পিক প্রকল্পের জন্য বর্জনাঞ্চল বিভিন্ন প্রকার হবে।

(ছ) দক্ষিণ পাক্ষিক বৈকল্পিক প্রকল্প

মুখ্য প্রকল্প H_0-কে বর্জন করা হবে যদি $t > K$ হয়, যেখানে K একটি ধ্রুবক এবং K-কে এমনভাবে নির্ধারণ করতে হবে যাতে $\alpha(\mu_0) = P\mu_0 (t > K) = \alpha$ হয়।

স্পষ্টত K, $\overline{n-1}$ স্বাতন্ত্র্যমাত্রা যুক্ত t নিবেশনের ঊর্ধ্ব $100\alpha\%$ বিন্দু। অর্থাৎ $K = t_{\alpha, n-1}$

(জ) বাম পাক্ষিক বৈকল্পিক প্রকল্প

মুখ্য প্রকল্প H_0-কে বর্জন করা হবে যদি $t < K'$ হয়, যেখানে K' একটি ধ্রবক এবং একে এমনভাবে নির্ধারণ করতে হবে যাতে $\alpha(\mu_0) = P\mu_0 (t < K') = \alpha$ হয়।

স্পষ্টত K' স্বাতন্ত্র্যমাত্রা যুক্ত t নিবেশনের নিম্ন $100\alpha\%$ বিন্দু। অর্থাৎ $K' = -t_\alpha$

(ঝ) উভয় পাক্ষিক বৈকল্পিক প্রকল্প

মুখ্য প্রকল্প H_0-কে বর্জন করা হবে যদি $|t| > K^*$ হয়, যেখানে K^* একটি ধ্রুবক এবং একে এমনভাবে নির্ধারণ করতে হবে যাতে $\alpha(\mu_0) = P\mu_0 (|t| > K^*) = \alpha$ হয়।

স্পষ্টত K^* $\overline{n-1}$ স্বাতন্ত্র্যমাত্রা যুক্ত t নিবেশনের ঊর্ধ্ব $100\frac{\alpha}{2}\%$ বিন্দু, অর্থাৎ $K^* = t_{\alpha/2}$

উপরোক্ত তিনটি বৈকল্পিক প্রকল্পের জন্য বিচার পদ্ধতিকে নিম্নলিখিতভাবে বর্ণনা করা যেতে পারে।

1. n সংখ্যক নমুনা থেকে $t = \frac{\overline{X} - \mu_0}{s/\sqrt{n}}$ বের করতে হবে।

2. সংশয় মাত্রা $100\alpha\%$ হলে

   (ক) H_0 বর্জিত হবে এবং $H_1 : \mu > \mu_0$ গৃহীত হবে যদি $t > t_\alpha$ হয়।

   (খ) $H_0 : \mu = \mu_0$ বর্জিত হবে এবং $H_1 : \mu < \mu_0$ গৃহীত হবে যদি $t < -t_\alpha$ হয়।

   (গ) $H_0 : \mu = \mu_0$ বর্জিত হবে এবং $H_1 : \mu \neq \mu_0$ গৃহীত হবে যদি $|t| > t_{\alpha/2}$ হয়।

টীকা : t নিবেশনের এই সকল শতকরা বিন্দু পিয়ারসন ও হার্টলের বায়োমেট্রিক সারণি থেকে পাওয়া যাবে।

ক্ষেত্র—III

ধরা যাক পূর্ণক নিবেশনটি অজানা এবং নমুনা আয়তন (n) বৃহৎ (20টির বেশী)

মুখ্য প্রকল্প অনুসারে যেহেতু n বৃহৎ, $\frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$

আবার যদি σ-এর পরিবর্তে s নিই তখন মুখ্য প্রকল্প অনুসারে $\frac{\overline{X} - \mu_0}{s/\sqrt{n}} \sim$ প্রায় সঠিক N(0, 1 ।

সুতরাং ক্ষেত্র—I -এর বিচারগুলি প্রায় সঠিকভাবে এক্ষেত্রে ব্যবহার করা যাবে। শুধুমাত্র σ^2-এর পরিবর্তে s^2 ব্যবহার করা হবে যখন σ^2 অজানা।

বিভিন্ন ক্ষেত্রে যে বিচার পদ্ধতিগুলি পেয়েছি সেগুলি সংক্ষিপ্ত আকারে নীচের সারণি (৪৮.১)-তে দেওয়া হল।

সারণির ৪৮.১ : পূর্ণক গড় (μ)-এর বিচার

পূর্ণক নিবেশন সম্বন্ধে ধারণা	H_0	H_1	H_0 বর্জিত হবে যদি	বিচার নমুনাঙ্ক	মন্তব্য
পূর্ণক নিবেশন $N(\mu, \sigma^2)$ σ^2 জানা আছে	$\mu = \mu_0$ $\mu = \mu_0$ $\mu = \mu_0$	$\mu > \mu_0$ $\mu < \mu_0$ $\mu \neq \mu_0$	$Z > Z_\alpha$ $Z < - Z_\alpha$ $\lvert Z \rvert\ Z_{\alpha/2}$	$Z = \dfrac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}$	দক্ষিণপুচ্ছ নর্ম্যাল বিচার বামপুচ্ছ নর্ম্যাল বিচার উভপুচ্ছ নর্ম্যাল বিচার
পূর্ণ নিবেশন $N(\mu, \sigma^2)$ σ^2 অজানা	$\mu = \mu_0$ $\mu = \mu_0$ $\mu = \mu_0$	$\mu > \mu_0$ $\mu < \mu_0$ $\mu \neq \mu_0$	$t > t_\alpha$ $t < -t_\alpha$ $\lvert t \rvert < t_{\alpha/2}$	$t = \dfrac{\overline{X} - \mu_0}{s / \sqrt{n}}$	দক্ষিণপুচ্ছ t বিচার বামপুচ্ছ t বিচার উভপুচ্ছ t বিচার
অজানা পূর্ণক নমুনা আয়তন বৃহৎ	$\mu = \mu_0$ $\mu = \mu_0$ $\mu = \mu_0$	$\mu > \mu_0$ $\mu < \mu_0$ $\mu \neq \mu_0$	$Z > Z_\alpha$ $Z < -Z_\alpha$ $\lvert Z \rvert < Z_{\alpha/2}$	$Z = \dfrac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}$ σ যদি অজানা থাকে σ এর পরিবর্তে s নির্তে হবে	দক্ষিণপুচ্ছ নর্ম্যাল বিচার বামপুচ্ছ নর্ম্যাল বিচার উভপুচ্ছ নর্ম্যাল বিচার

৪৮.৩.২ পূর্ণক-ভেদমান (σ^2) সম্বন্ধীয় প্রকল্প বিচার

পূর্ণক গড়ের মতো এখানেও আমরা পূর্ণক ভেদমান σ^2-এর তিন প্রকার প্রকল্পের বিষয়ে আলোচনা করব।

I $\quad H_0 : \sigma^2 = \sigma_0^2 \quad H_1 : \sigma^2 > \sigma_0^2$ দক্ষিণ পাক্ষিক বৈকল্পিক প্রকল্প

II $\quad H_0 : \sigma^2 = \sigma_0^2 \quad H_1 : \sigma^2 < \sigma_0^2$ বাম পাক্ষিক বৈকল্পিক প্রকল্প

III $\quad H_0 : \sigma^2 = \sigma_0^2 \quad H_1 : \sigma^2 \neq \sigma_0^2$ উভয় পাক্ষিক বৈকল্পিক প্রকল্প

পূর্ব ধারণা থেকে σ^2-এর মান নির্ধারিত।

যখন কোনও বস্তুর সাদৃশ্যগত নিয়ন্ত্রণের প্রয়োজন হয় সেই ক্ষেত্রে σ^2-এর বিচার দরকারী। উদাহরণস্বরূপ বলা যেতে পরে পুশপিনের দৈর্ঘ্য 170 ± 1 মিলি মি.। পিন কিনতে ইচ্ছুক সংস্থা পিনের দৈর্ঘ্য 170 মিমি বা বিভিন্ন পিনের দৈর্ঘের পার্থক্য খুব বেশী না হলে, সেই পিন ক্রয় করতে উৎসাহী হতে পারে, অর্থাৎ ভেদমান একটি নির্দিষ্ট সীমার মধ্যে আছে কিনা সেইটি যাচাই করে নেয়।

(ক) বর্জনাঞ্চল নির্ধারণ

পূর্ণক নিবেশন এবং পূর্ণক গড় (μ)-এর উপর নির্ভর করে তিনটি ক্ষেত্রে তিনটি বিভিন্ন রকম বর্জনাঞ্চল নির্ণয় করতে পারি।

ক্ষেত্র—I

ধরা যাক পূর্ণক নিবেশনটি $N(\mu, \sigma^2)$, এখানে μ জানা আছে। যদি $X_1, X_2, \ldots\ldots\ldots, X_n$ nটি সমসম্ভব নমুনা $N(\mu, \sigma^2)$ থেকে নেওয়া হয়। আমরা জানি $\frac{\sum_i (X_i - \mu)^2}{\sigma^2}$-এর নিবেশন n স্বাতন্ত্র্যমাত্রাযুক্ত χ^2-নিবেশন। সুতরাং মুখ্য প্রকল্প অনুসারে $\chi^2 = \frac{\sum_i (X_i - \mu)^2}{\sigma_0^2}$-এর নিবেশন n স্বতন্ত্র্যমাত্রা যুক্ত χ^2 নিবেশন। এক্ষেত্রে χ^2-কে আমরা বিচার নমুনাঙ্ক হিসাবে ব্যবহার করব। তিন প্রকার বৈকল্পিক প্রকল্পের জন্য বর্জনাঞ্চল বিভিন্ন প্রকার হবে।

(খ) দক্ষিণ পাক্ষিক বৈকল্পিক প্রকল্প

মুখ্য প্রকল্প H_0-কে বর্জন করা হবে যদি $\chi^2 > K$ হয়, যেখানে K একটি ধ্রুবক এবং K-কে এমনভাবে নির্ধারণ করতে হবে যাতে $\alpha(\sigma_0^2) = P_{\sigma_0^2}(\chi^2 > K) = \alpha$ হয়।

স্পষ্টত K n-স্বাতন্ত্র্যমাত্রা যুক্ত χ^2 নিবেশনের ঊর্ধ্ব $100\,\alpha\%$ বিন্দু। অর্থাৎ $K = \chi^2_{\alpha,n}$

(গ) বাম পাক্ষিক বৈকল্পিক প্রকল্প

মুখ্য প্রকল্প H_0-কে বর্জন করা হবে যদি $\chi^2 < K'$ হয়, যেখানে K' একটি ধ্রুবক এবং K'-কে এমনভাবে নির্ধারণ করতে হবে যাতে $\alpha(\sigma_0^2) = P_{\sigma_0^2}(\chi^2 < K') = \alpha$ হয়।

স্পষ্টত K' n স্বাতন্ত্র্যমাত্রা যুক্ত χ^2 নিবেশনের ঊর্ধ্ব $100\,(1-\alpha)\%$ বিন্দু। অর্থাৎ $K^1 = \chi^2_{1-\alpha,\,n}$

(ঘ) উভয় পাক্ষিক বৈকল্পিক প্রকল্প

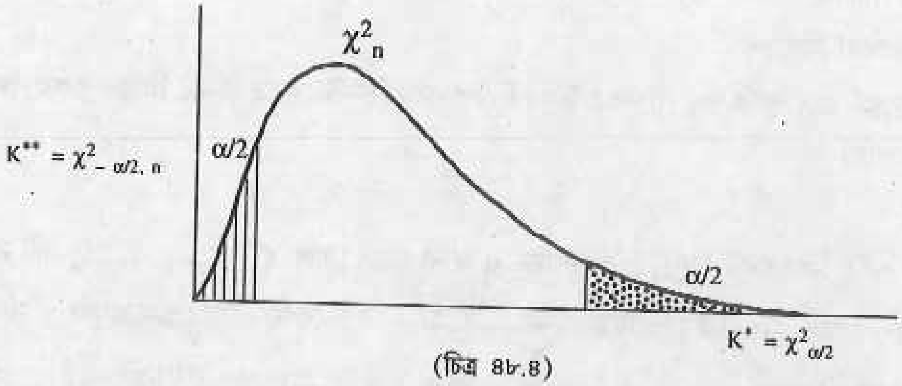
মুখ্য প্রকল্প H_0-কে বর্জন করা হবে যদি $\chi^2 > K^*$ হয়। কিংবা $\chi^2 < K^{**}$ হয়। এখানে K^* এবং K^{**} দুটি ধ্রুবক এবং এদের এমনভাবে নির্ধারণ করতে হবে যাতে

$$\alpha(\sigma_0^2) = \alpha \text{ হয় অর্থাৎ } \alpha(\sigma_0^2) = P_{\sigma_0^2}(\chi^2 > K^*) + P_{\sigma_0^2}(\chi^2 < K^{**})$$

$$= \alpha$$

সাধারণত উভয় পুচ্ছের পরিমাপ সমান $\frac{\alpha}{2}$ নেওয়া হয়। সেক্ষেত্রে

$K^* = \chi^2_{\alpha/2,\, n}$ এবং $K^{**} = \chi^2_{1-\alpha/2,\, n}$ (চিত্র ৪৮.৪)



(চিত্র ৪৮.৪)

উপরোক্ত তিনটি বৈকল্পিক প্রকল্পের জন্য বিচার পদ্ধতিকে নিম্নলিখিতভাবে বর্ণনা করা যেতে পারে।

1. n সংখ্যক নমুনা থেকে $\chi^2 = \frac{\sum_i (X_i - \mu)^2}{\sigma_0^2}$ বের করতে হবে।

2. সংশয়মাত্রা 100 α% হলে

(ক) $H_0 : \sigma^2 = \sigma_0^2$ বর্জিত হবে এবং $H_1 : \sigma^2 > \sigma_0^2$ গৃহীত হবে যদি $\chi^2 > \chi^2_{\alpha,n}$ হয়।

(খ) H_0 বর্জিত হবে এবং $H_1 : \sigma^2 < \sigma_0^2$ গৃহীত হবে যদি $\chi^2 < \chi^2_{1-\alpha,\, n}$ হয়।

(গ) H_0 বর্জিত হবে এবং $H_1 : \sigma^2 \neq \sigma_0^2$ গৃহীত হবে যদি $\chi^2 > \chi^2_{\alpha/2,\, n}$ অথবা $\chi^2 < \chi^2_{1-\alpha/2,\, n}$ হয়।

ক্ষেত্র—II

ধরা যাক পূর্ণক নিবেশনটি $N(\mu, \sigma^2)$, যেখানে μ অজানা। যদি $X_1, X_2, \ldots\ldots\ldots, X_n$ nটি নমুনা $N(\mu, \sigma^2)$ থেকে নেওয়া হয়। আমরা জানি $\frac{\Sigma(X_i - \bar{X})^2}{\sigma^2} = \frac{(\overline{n-1})s^2}{\sigma^2}$-এর নিবেশন (n – 1) স্বাতন্ত্র্যমাত্রা যুক্ত χ^2 নিবেশন। সুতরাং মুখ্য প্রকল্প অনুসারে $\frac{(n-1)s^2}{\sigma_0^2}$ এর নিবেশন (n – 1) স্বাতন্ত্র্যমাত্রা যুক্ত χ^2 নিবেশন।

এক্ষেত্রে χ^2-কে আমরা বিচার নমুনাঙ্ক হিসাবে ব্যবহার করব। তিনটি বৈকল্পিক প্রকল্পের জন্য বিচার পদ্ধতিকে নিম্নলিখিতভাবে বর্ণনা করা যেতে পারে।

1. n সংখ্যক নমুনা থেকে $\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$ বের করতে হবে।

2. সংশয় মাত্রা 100 α% হলে

(ক) H_0 বর্জিত হবে এবং $H_1 : \sigma^2 > \sigma_0^2$ গৃহীত হবে যদি $\chi^2 > \chi^2_{\alpha, n-1}$ হয়।

(খ) H_0 বর্জিত হবে এবং $H_1 : \sigma^2 < \sigma_0^2$ গৃহীত হবে যদি $\chi^2 > \chi^2_{1-\alpha, n-1}$ হয়।

(গ) H_0 বর্জিত হবে এবং $H_1 : \sigma^2 \neq \sigma_0^2$ গৃহীত হবে যদি $\chi^2 > \chi^2_{\alpha/2, n-1}$ অথবা $\chi^2 > \chi^2_{1-\alpha/2, n-1}$ হয়।

টীকা : χ^2 নিবেশনের এই সকল শতকরা বিন্দু পিয়ারসন ও হার্টলের বায়োমেট্রিক সারণিসমূহ থেকে পাওয়া যায়।

ক্ষেত্র—III

ধরা যাক, পূর্ণক নিবেশনটি অজানা এবং নমুনা আয়তন বৃহৎ ($n \geq 30$)। আমরা জানি যখন n বৃহৎ $\frac{s-\sigma}{\sigma.\sqrt{2n}}$-এর নিবেশনটি প্রায় সঠিক N(0, 1) হয়। সুতরাং মুখ্য প্রকল্প অনুসারে $Z = \frac{s-\sigma_0}{\sigma_0/\sqrt{2n}}$ এর নিবেশনটি প্রায় সঠিক N(0, 1)। এখানে আমরা Z-কে বিচার নমুনাঙ্ক হিসাবে ব্যবহার করব। বৈকল্পিক প্রকল্পের উপর নির্ভর করে বিচারগুলি একপুচ্ছ কিংবা উভপুচ্ছ নর্ম্যাল বিচার হবে। সুতরাং তিনটি বৈকল্পিক প্রকল্পের জন্য বিচার পদ্ধতিকে নিম্নলিখিতভাবে বর্ণনা করা যেতে পারে।

1. n সংখ্যক নমুনা থেকে $Z = \frac{s-\sigma_0}{\sigma_0/\sqrt{2n}}$ বের করতে হবে।

2. সংশয় মাত্রা 100 α% হলে

(ক) H_0 বর্জিত হবে এবং $H_1 : \sigma^2 > \sigma_0^2$ গৃহীত হবে যদি $Z > Z_\alpha$ হয়।

(খ) H_0 বর্জিত হবে এবং $H_1 : \sigma^2 < \sigma_0^2$ গৃহীত হবে যদি $Z < -Z_\alpha$ হয়।

(গ) H_0 বর্জিত হবে এবং $H_1 : \sigma^2 \neq \sigma_0^2$ গৃহীত হবে যদি $Z > Z_{\alpha/2}$ অথবা $Z < -Z_{\alpha/2}$ হয়।

বিভিন্ন ক্ষেত্রে σ^2-এর যে বিচারগুলি পেয়েছি সেগুলি সংক্ষিপ্ত আকারে পরের পাতার সারণিতে (৪৮.২) দেওয়া হল।

সারণী ৪৮.২ : পূর্ণক ভেদমান (σ^2)-এর বিচার

পূর্ণক নিবেশন সম্বন্ধে ধারণা	H_0	H_1	H_0 বর্জিত হবে যদি	বিচার নমুনাঙ্ক	মন্তব্য
পূর্ণক নিবেশন	$\sigma^2 = \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$\chi^2 > \chi^2_{\alpha,n}$	$\chi^2 = \frac{\sum_1 (X_i - \mu)^2}{\sigma_0^2}$	দক্ষিণপুচ্ছ বিচার
$N(\mu, \sigma^2)$	$\sigma^2 = \sigma_0^2$	$\sigma^2 < \sigma_0^2$	$\chi^2 < \chi^2_{1-\alpha,n}$		বামপুচ্ছ বিচার
μ জানা আছে	$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	$\chi^2 > \chi^2_{\alpha/2,\, n}$ অথবা $\chi^2 < \chi^2_{1-\alpha/2,\, n}$		উভপুচ্ছ বিচার
পূর্ণক নিবেশন	$\sigma^2 = \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$\chi^2 < \chi^2_{\alpha,\, n-1}$		দক্ষিণপুচ্ছ বিচার
$N(\mu, \sigma^2)$	$\sigma^2 = \sigma_0^2$	$\sigma^2 < \sigma_0^2$	$\chi^2 < \chi^2_{1-\alpha,\, n-1}$	$\chi^2 = \frac{\sum_1 (X_i - \bar{X})^2}{\sigma_0^2}$	বামপুচ্ছ বিচার
μ অজানা	$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	$\chi^2 > \chi^2_{\alpha/2,\, n-1}$ অথবা $\chi^2 < \chi^2_{1-\alpha/2,\, n-1}$	$= \frac{(n-1)s^2}{\sigma_0^2}$	উভপুচ্ছ বিচার
অজানা পূর্ণক	$\sigma^2 = \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$Z > Z_\alpha$		দক্ষিণপুচ্ছ বিচার
নমুনা আয়তন	$\sigma^2 = \sigma_0^2$	$\sigma^2 < \sigma_0^2$	$Z < -Z_\alpha$	$Z = \frac{s - \sigma_0}{\sigma_0 / \sqrt{2n}}$	বামপুচ্ছ বিচার
বৃহৎ	$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	$\lvert Z \rvert > Z_{\alpha/2}$		উভপুচ্ছ বিচার

৪৮.৩.৩ পূর্ণক অনুপাত (P) সম্বন্ধীয় প্রকল্প বিচার

অনেক বাস্তব পরিস্থিতিতে আমাদের পূর্ণকের কোনও নির্দিষ্ট বৈশিষ্ট্যমূলক বস্তুর অনুপাত P-এর প্রকল্প বিচারের প্রয়োজন হতে পারে। উদাহরণস্বরূপ P কোনও কোম্পানীর উৎপাদিত বস্তুর ত্রুটিপূর্ণ হবার সম্ভাবনা। এখানে আমরা তিন প্রকার প্রকল্পের বিষয়ে আলোচনা করব।

I	$H_0 : P = P_0$	$H_1 : P > P_0$
II	$H_0 : P = P_0$	$H_1 : P < P_0$
III	$H_0 : P = P_0$	$H_1 : P \neq P_0$

যেখানে P_o, P-এর পূর্ব ধারণালব্ধ একটি নির্দিষ্ট মান।

ধরা যাক nটি বস্তু আছে, যার মধ্যে Xটি বস্তু নির্দিষ্ট বৈশিষ্ট্যমূলক। তখন আমরা জানি $\frac{X}{n}$-এর নিবেশন প্রায় সঠিক $N\left(P, \frac{P(1-P)}{n}\right)$। সুতরাং $\frac{\frac{X}{n} - P}{\sqrt{\frac{P(1-P)}{n}}}$ এর নিবেশন N(0, 1) হবে। মুখ্য প্রকল্প অনুসারে,

$Z = \frac{\frac{X}{n} - P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}}$ এর নিবেশন প্রায় সঠিক N(0, 1) হবে।

এক্ষেত্রে আমরা Z-কে বিচার নমুনাঙ্ক হিসাবে ব্যবহার করব। তিনটি বৈকল্পিক প্রকল্পের জন্য বিচার পদ্ধতিকে নিম্নলিখিতভাবে বর্ণনা করা যেতে পারে।

1. n সংখ্যক নমুনা থেকে $Z = \frac{\frac{X}{n} - P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}}$ বের করতে হবে।

2. সংশয়মাত্রা 100 α% হলে

(ক) H_0 বর্জিত হবে এবং $H_1 : P > P_0$ গৃহীত হবে যদি $Z > Z_\alpha$ হয়।

(খ) H_0 বর্জিত হবে এবং $H_1 : P < P_0$ গৃহীত হবে যদি $Z < -Z_\alpha$ হয়।

(গ) H_0 বর্জিত হবে এবং $H_1 : P \neq P_0$ গৃহীত হবে যদি $|Z| > Z_{\alpha/2}$ হয়।

P-এর বিচারগুলিকে সংক্ষিপ্ত আকারে পরের পাতার সারণি (৪৮.৩) এ দেওয়া হল।

ক্ষেত্র—II

ধরা যাক দুটি পূর্ণকের নিবেশন $N(\mu_1, \sigma_1^2)$ এবং $N(\mu_2, \sigma_2^2)$, এবং $\sigma_1^2 = \sigma_2^2 = \sigma^2$ যেখানে σ^2 অজানা।

যেহেতু σ^2 অজানা, ক্ষেত্র–I এর নমুনাঙ্ক Z এখানে ব্যবহার করা যাবে না। এখানে σ^2-এর পরিবর্তে মিলিত নমুনা ভেদমান s^2 ব্যবহার করব যেখানে

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

আমরা জানি $\dfrac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{s\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$ -এর নিবেশন $\overline{n_1 + n_2 - 2}$ স্বাতন্ত্র্যমাত্রা যুক্ত t নিবেশন। মুখ্য প্রকল্প অনুসারে $t = \dfrac{\overline{X} - \overline{Y} - \delta}{s\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$ এর নিবেশন $\overline{n_1 + n_2 - 2}$ স্বাতন্ত্র্যমাত্রা যুক্ত t নিবেশন। এক্ষেত্রে t-কে আমরা বিচার নমুনাঙ্ক হিসাবে ব্যবহার করতে পারি। তিনটি বৈকল্পিক প্রকল্পের জন্য বিচার পদ্ধতিকে নিম্নলিখিতভাবে বর্ণনা করা যেতে পারে।

1. প্রথম এবং দ্বিতীয় পূর্ণক থেকে যথাক্রমে n_1 এবং n_2 সংখ্যক নমুনা নিয়ে $t = \dfrac{\overline{X} - \overline{Y} - \delta}{s\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$ বের করতে হবে।

2. সংশয়মাত্রা $100\alpha\%$ হলে

(ক) H_0 বর্জিত হবে এবং $H_1 : \mu_1 - \mu_2 > \delta$ গৃহীত হবে যদি $t > t_{\alpha,\, n_1 + n_2 - 2}$ হয়।

(খ) H_0 বর্জিত হবে এবং $H_1 : \mu_1 - \mu_2 < \delta$ গৃহীত হবে যদি $t < -t_{\alpha,\, n_1 + n_2 - 2}$ হয়।

(গ) H_0 বর্জিত হবে এবং $H_1 : \mu_1 - \mu_2 \neq \delta$ গৃহীত হবে যদি $|t| > t_{\alpha/2,\, n_1 + n_2 - 2}$ হয়।

ক্ষেত্র—III

ধরা যাক পূর্ণক নিবেশন দুটি অজানা দুটি নমুনার আয়তন n_1, n_2 বৃহৎ (≥ 20)। আমরা জানি বৃহৎ n_1 এবং n_2-এর জন্য $\overline{X}$, $\overline{Y}$ এর নিবেশন যথাক্রমে প্রায় সঠিক $N\left(\mu_1, \dfrac{\sigma_1^2}{n_1}\right)$ এবং $N\left(\mu_2, \dfrac{\sigma_2^2}{n_2}\right)$ হয়। সুতরাং $\overline{X} - \overline{Y}$ এর নিবেশন হবে $N\left(\mu_1 - \mu_2, \dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}\right)$।

I	$H_o : P = P_o$	$H_1 : P > P_o$
II	$H_o : P = P_o$	$H_1 : P < P_o$
III	$H_o : P = P_o$	$H_1 : P \neq P_o$

যেখানে P_o, P-এর পূর্ব ধারণালব্ধ একটি নির্দিষ্ট মান।

ধরা যাক nটি বস্তু আছে, যার মধ্যে Xটি বস্তু নির্দিষ্ট বৈশিষ্ট্যমূলক। তখন আমরা জানি $\frac{X}{n}$-এর নিবেশন প্রায় সঠিক $N\left(P, \frac{P(1-P)}{n}\right)$। সুতরাং $\dfrac{\frac{X}{n}-P}{\sqrt{\frac{P(1-P)}{n}}}$ এর নিবেশন N(0, 1) হবে। মুখ্য প্রকল্প অনুসারে,

$Z = \dfrac{\frac{X}{n}-P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}}$ এর নিবেশন প্রায় সঠিক N(0, 1) হবে।

এক্ষেত্রে আমরা Z-কে বিচার নমুনাঙ্ক হিসাবে ব্যবহার করব। তিনটি বৈকল্পিক প্রকল্পের জন্য বিচার পদ্ধতিকে নিম্নলিখিতভাবে বর্ণনা করা যেতে পারে।

1. n সংখ্যক নমুনা থেকে $Z = \dfrac{\frac{X}{n}-P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}}$ বের করতে হবে।

2. সংশয়মাত্রা 100 α% হলে

(ক) H_0 বর্জিত হবে এবং $H_1 : P > P_o$ গৃহীত হবে যদি $Z > Z_\alpha$ হয়।

(খ) H_0 বর্জিত হবে এবং $H_1 : P < P_o$ গৃহীত হবে যদি $Z < -Z_\alpha$ হয়।

(গ) H_0 বর্জিত হবে এবং $H_1 : P \neq P_o$ গৃহীত হবে যদি $|Z| > Z_{\alpha/2}$ হয়।

P-এর বিচারগুলিকে সংক্ষিপ্ত আকারে পরের পাতার সারণি (৪৮.৩) এ দেওয়া হল।

সারণি ৪৮.৩ : পূর্ণক অনুপাত P এর বিচার

পূর্ণক নিবেশন সম্বন্ধে ধারণা	H_0	H_1	H_0 বর্জিত হবে যদি	বিচার নমুনাঙ্ক	মন্তব্য
নমুনা আয়তন বৃহৎ ($\geq$ 100) P, O কিংবা 1-এর কাছাকাছি নয়	$P = P_o$ $P = P_o$ $P = P_o$	$P > P_o$ $P < P_o$ $P \neq P_o$	$Z > Z_\alpha$ $Z < -Z_\alpha$ $\lvert Z\rvert > Z_{\alpha/2}$	$Z = \dfrac{\frac{X}{n} - P_0}{\sqrt{\dfrac{P_0(1-P_0)}{n}}}$	দক্ষিণ পাক্ষিক নর্মাল বিচার বাম পাক্ষিক নর্মাল বিচার উভপাক্ষিক নর্মাল বিচার

৪৮.৪ পরিসংখ্যানগত বিচার—দুটি পূর্ণকের ক্ষেত্রে

শেষ অংশে একটি পূর্ণকের ক্ষেত্রে μ, σ^2 এবং P-এর প্রকল্প নিয়ে আলোচনা করা হয়েছে। কিন্তু অনেক পরিস্থিতির উদ্ভব হতে পারে যেখানে দুটি পূর্ণকের বিভিন্ন পূর্ণকাঙ্কের মধ্যে কোনও তফাৎ আছে কিনা সেই সমস্ত প্রকল্প বিচার করার প্রয়োজন হতে পারে। আমরা এই অংশে দুটি পূর্ণকের পূর্ণকাঙ্কের বিভিন্ন প্রকার প্রকল্পের বিচার নিয়ে আলোচনা করব।

৪৮.৪.১ দুটি পূর্ণক গড়ের তুলনা সম্বন্ধীয় প্রকল্প বিচার

ধরা যাক আমাদের দুটি পূর্ণক আছে। পূর্ণক–I এর গড় μ_1 এবং ভেদমান σ_1^2। পূর্ণক–II এর গড় μ_2 এবং ভেদমান σ_2^2। আমাদের দুটি পূর্ণক গড়ের তফাৎ $\mu_1 - \mu_2$ -এর প্রকল্পের বিচার করতে হবে। এখানে নমুনা রাশিতথ্যের মধ্যে দুটি সমসম্ভব নমুনার সমষ্টি থাকবে একটি নমুনা পূর্ণক–I থেকে এবং অন্যটি পূর্ণক–II থেকে। ধরা যাক, $X_1, X_2, \ldots\ldots, X_{n_1}$ n_1 আয়তনের সমসম্ভব নমুনা পূর্ণক–I থেকে এবং $Y_1, Y_2, \ldots\ldots, Y_{n_2}$ n_2 আয়তনের সমসম্ভব নমুনা পূর্ণক–II থেকে নেওয়া হয়েছে। সাধারণত দুটি নমুনা পরস্পর নিরপেক্ষ হয়। উদাহরণস্বরূপ, দুটি পূর্ণক দুটি বিভিন্ন পড়ানোর পদ্ধতি হতে পারে এবং সেখানে μ_1, μ_2 যথাক্রমে প্রথম পদ্ধতিতে এবং দ্বিতীয় পদ্ধতিতে পড়ানোর ফলে নম্বরের গড় হতে পারে। তখন প্রথম সমসম্ভব নমুনা n_1 ছাত্রের নম্বর যেখানে প্রথম পদ্ধতিতে পড়ানো হয়েছে এবং দ্বিতীয় সমসম্ভব নমুনা n_2 ছাত্রের নম্বর যেখানে দ্বিতীয় পদ্ধতিতে পড়ানো হয়েছে। যদি দাবী করা হয় যে দ্বিতীয় পদ্ধতি প্রথম পদ্ধতি থেকে ভালো তখন আমাদের বিচার করতে হবে $H_0 : \mu_1 = \mu_2$ বৈকল্পিক প্রকল্প $H_1 : \mu_1 < \mu_2$র সাপেক্ষে। এটি একটি বামপাক্ষিক বৈকল্পিক প্রকল্প। বর্তমানে আমরা তিন প্রকার বৈকল্পিক প্রকল্পের বিচার করব।

I $\quad H_0 : \mu_1 - \mu_2 = \delta$

$\quad H_1 : \mu_1 - \mu_2 > \delta$ দক্ষিণ পাক্ষিক বৈকল্পিক প্রকল্প

II $\qquad H_0 : \mu_1 - \mu_2 = \delta$

$\qquad H_1 : \mu_1 - \mu_2 < \delta$ বাম পাক্ষিক বৈকল্পিক প্রকল্প

III $\qquad H_0 : \mu_1 - \mu_2 = \delta$

$\qquad H_1 : \mu_1 - \mu_2 \neq \delta$

যেখানে δ পূর্বধারণা থেকে পাওয়া একটি নির্দিষ্ট বাস্তব সংখ্যা। উপরের উদাহরণে ধরা যাক $\overline{X}$ এবং $s_1^2 = \frac{1}{n_1 - 1} \sum_i (X_i - \overline{X})^2$ প্রথম পূর্ণকের নমুনা $X_1, X_2, \ldots\ldots\ldots, X_{n_1}$ থেকে পাওয়া নমুনা গড় এবং নমুনা ভেদমান। $\overline{Y}$ এবং $s_2^2 = \frac{1}{n_2 - 1} \sum_i (Y_i - \overline{Y})^2$ দ্বিতীয় পূর্ণকের নমুনা $Y_1, Y_2, \ldots\ldots\ldots, Y_{n_2}$ থেকে পাওয়া নমুনা গড় এবং নমুনা ভেদমান। একটি পূর্ণকের মতো এখানেও বর্জনাঞ্চল, পূর্ণক ভেদমান σ_1^2, σ_2^2 এবং n_1, n_2-এর উপর নির্ভর করবে। এখানে আমরা তিনটি বিভিন্ন ক্ষেত্রে প্রকল্প বিচার করব।

ক্ষেত্র—I

ধরা যাক দুটি পূর্ণকের নিবেশন যথাক্রমে $N(\mu_1, \sigma_1^2)$ এবং $N(\mu_2, \sigma_2^2)$, যেখানে μ_1^2, σ_2^2 জানা আছে। উপরোক্ত ধারণার পরিপ্রেক্ষিতে $\overline{X}$, $\overline{Y}$-এর নিবেশন যথাক্রমে $N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right)$ এবং $N\left(\mu_2, \frac{\sigma_2^2}{n_2}\right)$ যেহেতু $\overline{X}$ এবং $\overline{Y}$ নিরপেক্ষ, $\overline{X} - \overline{Y}$-এর নিবেশন $N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$ হয়।

সুতরাং $Z = \dfrac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$ এর নিবেশন $N(0,1)$ হবে।

মুখ্য প্রকল্প অনুসারে $Z = \dfrac{\overline{X} - \overline{Y} - \delta}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$ এর নিবেশন $N(0,1)$ হবে।

এক্ষেত্রে Z-কে আমরা বিচার নমুনাঙ্ক হিসাবে নিতে পারি। তিনটি বৈকল্পিক প্রকল্পের জন্য বিচার পদ্ধতিকে নিম্নলিখিতভাবে বর্ণনা করা যেতে পারে।

1. প্রথম এবং দ্বিতীয় পূর্ণক থেকে যথাক্রমে n_1 এবং n_2 নমুনা নিয়ে $Z = \dfrac{\overline{X} - \overline{Y} - \delta}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$ বের করতে হবে।

2. সংশয়মাত্রা $100\,\alpha\%$ হলে

(ক) H_0 বর্জিত হবে এবং $H_1 : \mu_1 - \mu_2 > \delta$ গৃহীত হবে যদি $Z > Z_\alpha$ হয়।

(খ) H_0 বর্জিত হবে এবং $H_1 : \mu_1 - \mu_2 < \delta$ গৃহীত হবে যদি $Z < -Z_\alpha$ হয়।

(গ) H_0 বর্জিত হবে এবং $H_1 : \mu_1 - \mu_2 \neq \delta$ গৃহীত হবে যদি $|Z| > Z_{\alpha/2}$ হয়।

ক্ষেত্র—II

ধরা যাক দুটি পূর্ণকের নিবেশন $N(\mu_1, \sigma_1^2)$ এবং $N(\mu_2, \sigma_2^2)$, এবং $\sigma_1^2 = \sigma_2^2 = \sigma^2$ যেখানে σ^2 অজানা।

যেহেতু σ^2 অজানা, ক্ষেত্র–I এর নমুনাঙ্ক Z এখানে ব্যবহার করা যাবে না। এখানে σ^2-এর পরিবর্তে মিলিত নমুনা ভেদমান s^2 ব্যবহার করব যেখানে

$$s^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$$

আমরা জানি $\dfrac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$ -এর নিবেশন $\overline{n_1+n_2-2}$ স্বাতন্ত্র্যমাত্রা যুক্ত t নিবেশন। মুখ্য প্রকল্প অনুসারে $t = \dfrac{\overline{X}-\overline{Y}-\delta}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$ এর নিবেশন $\overline{n_1+n_2-2}$ স্বাতন্ত্র্যমাত্রা যুক্ত t নিবেশন। এক্ষেত্রে t-কে আমরা বিচার নমুনাঙ্ক হিসাবে ব্যবহার করতে পারি। তিনটি বৈকল্পিক প্রকল্পের জন্য বিচার পদ্ধতিকে নিম্নলিখিতভাবে বর্ণনা করা যেতে পারে।

1. প্রথম এবং দ্বিতীয় পূর্ণক থেকে যথাক্রমে n_1 এবং n_2 সংখ্যক নমুনা নিয়ে $t = \dfrac{\overline{X}-\overline{Y}-\delta}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$ বের করতে হবে।

2. সংশয়মাত্রা $100\alpha\%$ হলে

(ক) H_0 বর্জিত হবে এবং $H_1 : \mu_1 - \mu_2 > \delta$ গৃহীত হবে যদি $t > t_{\alpha,\ n_1+n_2-2}$ হয়।

(খ) H_0 বর্জিত হবে এবং $H_1 : \mu_1 - \mu_2 < \delta$ গৃহীত হবে যদি $t < -t_{\alpha,\ n_1+n_2-2}$ হয়।

(গ) H_0 বর্জিত হবে এবং $H_1 : \mu_1 - \mu_2 \neq \delta$ গৃহীত হবে যদি $|t| > t_{\alpha/2,\ n_1+n_2-2}$ হয়।

ক্ষেত্র—III

ধরা যাক পূর্ণক নিবেশন দুটি অজানা দুটি নমুনার আয়তন n_1, n_2 বৃহৎ (≥ 20)। আমরা জানি বৃহৎ n_1 এবং n_2-এর জন্য $\overline{X}$, $\overline{Y}$ এর নিবেশন যথাক্রমে প্রায় সঠিক $N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right)$ এবং $N\left(\mu_2, \frac{\sigma_2^2}{n_2}\right)$ হয়। সুতরাং $\overline{X} - \overline{Y}$ এর নিবেশন হবে $N\left(\mu_1-\mu_2, \frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right)$।

সুতরাং $\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$ এর নিবেশন প্রায় সঠিক N(0, 1) হবে। মুখ্য প্রকল্প অনুসারে

$Z=\frac{\overline{X}-\overline{Y}-\delta}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$ এর নিবেশন প্রায় সঠিক N(0, 1) হবে। এখানে Z-কে আমরা বিচার নমুনাঙ্ক হিসাবে নিতে পারি। ক্ষেত্র–I এর সমস্ত বিচারগুলি এখানেও প্রযোজ্য যখন σ_1^2 এবং σ_2^2 অজানা, তখন σ_1^2 এবং σ_2^2-এর পরিবর্তে যথাক্রমে s_1^2 এবং s_2^2 নেওয়া হলেও z-এর নিবেশন প্রায় সঠিক N(0,1)-ই হবে। সুতরাং এক্ষেত্রেও ক্ষেত্র–I -এর সমস্ত বিচারগুলি প্রযোজ্য।

নীচের সারণি (৪৮.৪) তে $\mu_1-\mu_2$ এর বিচারগুলিকে সংক্ষিপ্ত আকারে দেওয়া হল।

সারণী ৪৮.৪ : দুটি পূর্ণক গড়ের তুলনা সম্বন্ধীয় বিচার

পূর্ণক নিবেশন সম্বন্ধে ধারণা	H_0	H_1	H_0 বর্জিত হবে যদি	বিচার নমুনাঙ্ক	মন্তব্য
পূর্ণক নিবেশন দুটি $N(\mu_1, \sigma_1^2)$ $N(\mu_2, \sigma_2^2)$ যেখানে σ_1^2, σ_2^2 জানা আছে। পরস্পর নিরপেক্ষ নিবেশন।	$\mu_1-\mu_2=\delta$ $\mu_1-\mu_2=\delta$ $\mu_1-\mu_2=\delta$	$\mu_1-\mu_2>\delta$ $\mu_1-\mu_2<\delta$ $\mu_1-\mu_2\neq\delta$	$Z>Z\alpha$ $Z<-Z\alpha$ $\lvert Z\rvert>Z_{\alpha/2}$	$Z=\frac{\overline{X}-\overline{Y}-\delta}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$	দক্ষিণপুচ্ছ নর্ম্যাল বিচার বামপুচ্ছ নর্ম্যাল বিচার উভপুচ্ছ নর্ম্যাল বিচার
পূর্ণক নিবেশন দুটি $N(\mu_1, \sigma_1^2)$ $N(\mu_2, \sigma_2^2)$ $\sigma_1^2=\sigma_2^2=\sigma^2$ σ^2 অজানা, পরস্পর নিরপেক্ষ নিবেশন	$\mu_1-\mu_2=\delta$ $\mu_1-\mu_2=\delta$ $\mu_1-\mu_2=\delta$	$\mu_1-\mu_2>\delta$ $\mu_1-\mu_2<\delta$ $\mu_1-\mu_2\neq\delta$	$t>t_{\alpha,\, n_1+n_2-2}$ $t<-t_{\alpha,\, n_1+n_2-2}$ $\lvert t\rvert>t_{\alpha/2,\, n_1+n_2-2}$	$t=\frac{\overline{X}-\overline{Y}-\delta}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$	দক্ষিণপুচ্ছ t-বিচার বামপুচ্ছ t-বিচার উভপুচ্ছ t-বিচার
পূর্ণক নিবেশন অজানা n_1, n_2 বৃহৎ পরস্পর নিরপেক্ষ নমুনা	$\mu_1-\mu_2=\delta$ $\mu_1-\mu_2=\delta$ $\mu_1-\mu_2=\delta$	$\mu_1-\mu_2>\delta$ $\mu_1-\mu_2<\delta$ $\mu_1-\mu_2\neq\delta$	$Z>Z\alpha$ $Z<-Z\alpha$ $\lvert Z\rvert>Z_{\alpha/2}$	$Z=\frac{\overline{X}-\overline{Y}-\delta}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$ যদি σ_1^2, σ_2^2 অজানা হয়, তাদের পরিবর্তে s_1^2 এবং s_2^2 নিতে হয়।	দক্ষিণপুচ্ছ নর্ম্যাল বিচার বামপুচ্ছ নর্ম্যাল বিচার উভপুচ্ছ নর্ম্যাল বিচার

৪৮.৪.২ দুটি পূর্ণক ভেদমানের তুলনা সম্বন্ধীয় প্রকল্প বিচার

ধরা যাক পুরুষ ও স্ত্রী মাকড়শা নিয়ে দুটি পূর্ণক আছে। σ_1^2, σ_2^2 যথাক্রমে পুরুষ ও স্ত্রী মাকড়শার দৈর্ঘ্যের ভেদমান। যদি আমরা দেখতে চাই যে পুরুষ মাকড়শার দৈর্ঘ্যের তারতম্য স্ত্রী মাকড়শার দৈর্ঘ্যের তারতম্য অপেক্ষা বেশী কিনা, তাহলে আমাদের বিচার করতে হবে $H_0 : \sigma_1^2 = \sigma_2^2$ বৈকল্পিক প্রকল্প $H_1 : \sigma_1^2 > \sigma_2^2$ এর সাপেক্ষে। এখানে আমরা তিন প্রকার বৈকল্পিক প্রকল্পের বিচার করব।

I $\quad H_0 : \dfrac{\sigma_1^2}{\sigma_2^2} = \xi_0^2$

$\quad H_1 : \dfrac{\sigma_1^2}{\sigma_2^2} > \xi_0^2$

II $\quad H_0 : \dfrac{\sigma_1^2}{\sigma_2^2} = \xi_0^2$

$\quad H_1 : \dfrac{\sigma_1^2}{\sigma_2^2} < \xi_0^2$

III $\quad H_0 : \dfrac{\sigma_1^2}{\sigma_2^2} = \xi_0^2$

$\quad H_1 : \dfrac{\sigma_1^2}{\sigma_2^2} \neq \xi_0^2$

যেখানে ξ_0 পূর্বধারণা থেকে পাওয়া একটি নির্দিষ্ট মান। উপরের উদাহরণে $\xi_0 = 1$। ধরা যাক $\overline{X}$ এবং s_1^2 প্রথম পূর্ণকের নমুনা $X_1 \ldots\ldots\ldots Xn_1$ থেকে পাওয়া নমুনা গড় এবং নমুনা ভেদমান এবং $\overline{Y}$ এবং s_2^2 দ্বিতীয় পূর্ণকের নমুনা $Y_1, Y_2 \ldots\ldots\ldots Yn_2$ থেকে পাওয়া নমুনা গড় এবং নমুনা ভেদমান। একটি পূর্ণকের মতো এখানেও বর্জনাঞ্চল পূর্ণক গড় μ_1 এবং μ_2 এর উপর নির্ভর করবে। এখানে আমরা দুটি বিভিন্ন ক্ষেত্রে বিচার পদ্ধতি আলোচনা করব।

ক্ষেত্র—I

ধরা যাক দুটি পূর্ণক যথাক্রমে $N(\mu_1, \sigma_1^2)$ এবং $N(\mu_2, \sigma_2^2)$, যেখানে μ_1, μ_2 জানা আছে এবং পূর্ণক দুটি পরস্পর নিরপেক্ষ। উপরোক্ত ধারণার পরিপ্রেক্ষিতে $\frac{\sum_i (X_i-\mu_1)^2/\sigma_1^2 n_1}{\sum_i (X_i-\mu_2)^2/\sigma_2^2 n_2}$ এর নিবেশন n_1, n_2 স্বাতন্ত্র্যমাত্রা যুক্ত F নিবেশন। মুখ্য প্রকল্প অনুসারে $F = \frac{s_1^2/s_2^2}{\xi_0^2}$ এর নিবেশন n_1, n_2 স্বাতন্ত্র্যমাত্রাযুক্ত F নিবেশন, যেখানে $s_1^{*2} = \frac{1}{n_1}\sum_i (X_i-\mu_1)^2$ এবং $s_2^{*2} = \frac{1}{n_2}\sum_i (Y_i-\mu_2)^2$ এক্ষেত্রে আমরা F-কে বিচার নমুনাঙ্ক হিসাবে নিতে পারি। তিনটি বৈকল্পিক প্রকল্পের জন্য বিচার পদ্ধতিকে নিম্নলিখিতভাবে বর্ণনা করা যেতে পারে।

1. প্রথম এবং দ্বিতীয় পূর্ণক থেকে n_1 এবং n_2 সংখ্যক নমুনা নিয়ে $F = \frac{s_1^{*2}/s_2^{*2}}{\xi_0^2}$ বের করতে হবে।

2. সংশয় মাত্রা $100\alpha\%$ হলে

(ক) H_0 বর্জিত হবে এবং $H_1: \frac{\sigma_1^2}{\sigma_2^2} > \xi_0^2$ গৃহীত হবে যদি $F > F_{\alpha, n_1, n_2}$ হয়।

(খ) H_0 বর্জিত হবে এবং $H_1: \frac{\sigma_1^2}{\sigma_2^2} < \xi_0^2$ গৃহীত হবে যদি $F < F_{1-\alpha, n_1, n_2}$ হয়।

(গ) H_0 বর্জিত হবে এবং $H_1: \frac{\sigma_1^2}{\sigma_2^2} \neq \xi_0^2$ গৃহীত হবে ,যদি $F > F_{\alpha/2, n_1, n_2}$ অথবা $F < F_{1-\alpha/2, n_1, n_2}$.

ক্ষেত্র—II

ধরা যাক, দুটি পূর্ণক যথাক্রমে $N(\mu_1, \sigma_1^2)$, $N(\mu_2, \sigma_2^2)$ যেখানে μ_1, μ_2 অজানা।

যেহেতু μ_1 এবং μ_2 এর মান অজানা সেজন্য ক্ষেত্র–I নমুনাঙ্ক এখানে ব্যবহার করা যাবে না। μ_1 μ_2 এর পরিবর্তে $\overline{X}$, $\overline{Y}$ ব্যবহার করা হবে। সুতরাং এ ক্ষেত্রে আমাদের বিচার নমুনাঙ্ক হবে $F = \frac{s_1^2/s_2^2}{\xi_0^2}$। যেক্ষেত্রে $s_1^2 = \frac{1}{n_1-1}\Sigma(X_i-\overline{X})^2$ ও $s_2^2 = \frac{1}{n_2-1}\Sigma(Y_i-\overline{Y})^2$ এবং এর নিবেশন হবে $\overline{n_1-1}, \overline{n_2-1}$ স্বাতন্ত্র্যমাত্রা যুক্ত F নিবেশন। তিনটি বৈকল্পিক প্রকল্পের জন্য বিচার পদ্ধতিকে নিম্নলিখিতভাবে বর্ণনা করা যেতে পারে।

1. প্রথম এবং দ্বিতীয় পূর্ণক থেকে যথাক্রমে n_1 ও n_2 নমুনা নিয়ে $F = \dfrac{s_1^2 / s_2^2}{\xi_0^2}$ বের করতে হবে।

2. সংশয়মাত্রা $100\alpha\%$ হলে

(ক) H_0 বর্জিত হবে এবং $H_1 : \dfrac{\sigma_1^2}{\sigma_2^2} > \xi_0^2$ গৃহীত হবে যদি $F > F_{\alpha, n_1 - 1, n_2 - 1}$ হয়।

(খ) H_0 বর্জিত হবে এবং $H_1 : \dfrac{\sigma_1^2}{\sigma_2^2} < \xi_0^2$ গৃহীত হবে যদি $F < F_{1 - \alpha, \overline{n_1 - 1}, \overline{n_2 - 1}}$ হয়।

(গ) H_0 বর্জিত হবে এবং $H_1 : \dfrac{\sigma_1^2}{\sigma_2^2} \neq \xi_0^2$ গৃহীত হবে যদি $F > F_{\alpha/2, n_1 - 1, n_2 - 1}$ অথবা $F < F_{1 - \alpha/2, n_1 - 1, n_2 - 1}$ হয়।

$\dfrac{\sigma_1^2}{\sigma_2^2}$ এর বিচারগুলিকে সংক্ষিপ্ত আকারে নীচের সারণি (৪৮.৫)-এ দেওয়া হল।

সারণির ৪৮.৫ : দুটি পূর্ণক ভেদমানের তুলনা সম্বন্ধীয় বিচার

পূর্ণক নিবেশন সম্বন্ধে ধারণা	H_0	H_1	H_0 বর্জিত হবে যদি	বিচার নমুনাঙ্ক	মন্তব্য
পূর্ণক নিবেশন দুটি $N(\mu_1, \sigma_1^2)$ $N(\mu_2, \sigma_2^2)$ যেখানে μ_1, μ_2 জানা আছে। পরস্পর নিরপেক্ষ নিবেশন	$\sigma_1^2 = \sigma_2^2\xi_0^2$	$\sigma_1^2 > \sigma_2^2\xi_0^2$	$F > F_{\alpha, n_1 - 1\, n_2 - 1}$	$F = \dfrac{s_1'^2}{s_2'^2 \xi_0^2}$	দক্ষিণপুচ্ছ F বিচার
	$\sigma_1^2 = \sigma_2^2\xi_0^2$	$\sigma_1^2 < \sigma_2^2\xi_0^2$	$F < F_{1-\alpha, \overline{n_1 - 1}, \overline{n_2 - 1}}$		বামপুচ্ছ F বিচার
	$\sigma_1^2 = \sigma_2^2\xi_0^2$	$\sigma_1^2 \neq \sigma_2^2\xi_0^2$	$F > F_{\alpha/2, n_1 - 1, n_2 - 1}$ অথবা $F < F_{1-\alpha/2, n_1 - 1, n_2 - 1}$		উভপুচ্ছ F বিচার
পূর্ণক নিবেশন দুটি $N(\mu_1, \sigma_1^2)$ $N(\mu_2, \sigma_2^2)$ যেখানে μ_1, μ_2 অজানা। পরস্পর নিরপেক্ষ নিবেশন।	$\sigma_1^2 = \sigma_2^2\xi_0^2$	$\sigma_1^2 > \sigma_2^2\xi_0^2$	$F > F_{\alpha, n_1 - 1, n_2 - 1}$	$F = \dfrac{s_1^2}{s_2^2 \xi_0^2}$	দক্ষিণপুচ্ছ F বিচার
	$\sigma_1^2 = \sigma_2^2\xi_0^2$	$\sigma_1^2 < \sigma_2^2\xi_0^2$	$F < F_{1-\alpha, n_1 - 1, n_2 - 1}$		বামপুচ্ছ F বিচার
	$\sigma_1^2 = \sigma_2^2\xi_0^2$	$\sigma_1^2 \neq \sigma_2^2\xi_0^2$	$F > F_{\alpha/2, n_1 - 1, n_2 - 1}$ অথবা $F < F_{1-\alpha/2, n_1 - 1, n_2 - 1}$		উভপুচ্ছ F বিচার

টীকা : F নিবেশনের এই সকল শতকরা বিন্দু পিয়ারসন ও হার্টলের বায়োমেট্রিকা সারণীসমূহ থেকে পাওয়া যাবে।

৪৮.৪.৩ দুটি অনুপাতের তুলনা সম্বন্ধীয় প্রকল্প বিচার

অনেক সময় আমাদের দুটি পূর্ণকে কোনও বস্তুর কোন নির্দিষ্ট বৈশিষ্ট্যের অনুপাত যথাক্রমে P_1 এবং P_2 সম্পর্কীত প্রকল্প বিচার করার প্রয়োজন হতে পারে। উদাহরণস্বরূপ, P_1 P_2 যথাক্রমে যন্ত্র I এবং II এর ত্রুটিপূর্ণ উৎপাদনের অনুপাত, যদি কোনও নিয়ন্ত্রক বিশেষজ্ঞ ধারণা করেন যে যন্ত্র I, যন্ত্র II এর থেকে 5% বেশী ত্রুটিপূর্ণ বস্তু উৎপাদন করে তখন যথার্থ মুখ্য প্রকল্প হবে।

$H_0 : P_1 - P_2 = 0.05$ এবং বৈকল্পিক প্রকল্প $H_1 : P_1 - P_2 > 0.05$

সাধারণত আমরা তিন প্রকার বৈকল্পিক প্রকল্পের বিচার করব।

I. $H_0 : P_1 - P_2 = \delta$

$H_1 : P_1 - P_2 > \delta$

II. $H_0 : P_1 - P_2 = \delta$

$H_1 : P_1 - P_2 < \delta$

III. $H_0 : P_1 - P_2 = \delta$

$H_1 : P_1 - P_2 \neq \delta$

যেখানে δ জানা আছে। উপরের উদাহরণে $\delta = 0.05$ ধরা যাক দুটি পূর্ণক থেকে যথাক্রমে n_1 এবং n_2 নমুনা নেওয়া হয়েছে। X_1 এবং X_2 যথাক্রমে বিশেষ বৈশিষ্ট্য সম্পন্ন বস্তুর সংখ্যা প্রথম এবং দ্বিতীয় পূর্ণকের ক্ষেত্রে তখন $P_1 - P_2$ প্রকল্প বিচারের জন্য যথার্থ বিচার নমুনা হবে $Z = \dfrac{\hat{P}_1 - \hat{P}_2 - \delta}{\sqrt{\dfrac{\hat{P}_1(1-\hat{P}_1)}{n_1} + \dfrac{\hat{P}_2(1-\hat{P}_2)}{n_2}}}$

যেখানে $\hat{P}_1 = \dfrac{X_1}{n_1}$ এবং $\hat{P}_2 = \dfrac{X_2}{n_2}$। Z এর নিবেশন প্রায় সঠিক N(0,1) হবে যখন n_1, n_2 বৃহৎ। সুতরাং Z কে আমরা $(P_1 - P_2)$ এর প্রকল্প বিচারের বর্জনাঞ্চল বের করার জন্য ব্যবহার করতে পারি। নীচের সারণি (৪৮.৬) তে $P_1 - P_2$ এর বিচারগুলিকে সংক্ষিপ্ত আকার দেওয়া হল।

বিশেষ ক্ষেত্রে মুখ্য প্রকল্প $H_0 : P_1 = P_2 = P$ বা $P_1 - P_2 = 0$ হলে

P-এর প্রাককলক দুটি নমুনার সমষ্টি থেকে হবে

$$\hat{P} = \frac{n_1 x_1 + n_2 x_2}{n_1 + n_2}$$

ও বিচার নমুনাঙ্ক হবে

$$Z = \frac{(\hat{P}_1 - \hat{P}_2) - 0}{\sqrt{\hat{P}(1-\hat{P})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

সারণী ৪৮.৬ : দুটি অনুপাতের তুলনার প্রকল্প বিচার

পূর্ণক নিবেশন সম্বন্ধে ধারণা	H_0	H_1	H_0 বর্জিত হবে যদি	বিচার নমুনাঙ্ক	মন্তব্য
পরস্পর নিরপেক্ষ নমুনা n_1 n_2 বৃহৎ	$P_1 - P_2 = \delta$	$P_1 - P_2 > \delta$	$Z > Z_\alpha$	$Z = \frac{\hat{P}_1 - \hat{P}_2 - \delta}{\sqrt{\frac{\hat{P}_1(1-\hat{P}_1)}{n_1} + \frac{\hat{P}_2(1-\hat{P}_2)}{n_2}}}$	দক্ষিণ পাক্ষিক নর্ম্যাল বিচার
	$P_1 - P_2 = \delta$	$P_1 - P_2 < \delta$	$Z < -Z_\alpha$		বাম পাক্ষিক নর্ম্যাল বিচার
	$P_1 - P_2 = \delta$	$P_1 - P_2 \neq \delta$	$\lvert Z \rvert > Z_{\alpha/2}$	$\hat{P}_1 = \frac{x_1}{n_1}, \hat{P}_2 = \frac{x_2}{n_2}$	উভ পাক্ষিক নর্ম্যাল বিচার

৪৮.৫ সারাংশ

এই এককে বর্ণিত বিষয়গুলিকে নিম্নে সারাংশ আকারে দেওয়া হল :

(ক) প্রকল্প বিচার সম্বন্ধে মৌলিক ধারণা।

(খ) একটি পূর্ণকের গড় (μ) প্রকল্প বিচারের জন্য নর্ম্যাল এবং t বিচার।

(গ) একটি পূর্ণক ভেদমানের বিচারের জন্য χ^2 এবং বৃহৎ নমুনা নর্ম্যাল বিচার।

(ঘ) একটি পূর্ণক অনুপাতের (P) প্রকল্প বিচারের জন্য বৃহৎ নমুনা নর্ম্যাল বিচার।

(ঙ) দুটি পূর্ণক গড়ের তফাৎ ($\mu_1 - \mu_2$)-এর প্রকল্প বিচারের জন্য নর্ম্যাল এবং t বিচার।

(চ) দুটি পূর্ণক ভেদমানের অনুপাতের প্রকল্প বিচারের জন্য F বিচার।

(ছ) দুটি অনুপাতের তফাৎ এর প্রকল্প বিচারের জন্য প্রায় সঠিক নর্ম্যাল বিচার।

৪৮.৬ অনুশীলনী ও উত্তরসংকেত

১। পোঁয়াস নিবেশনের গড় বড় জোর 2 হবে এই সম্বন্ধীয় প্রকল্প বর্জিত হবে যদি 4টি সমসম্ভব নমুনার ভিত্তিতে নমুনা গড় 2.5 এর বেশী হয়। এই বিচারের শক্তি অপেক্ষক কী হবে? এই বিচারের সংশয়মাত্রা বের করুন।

২। একটি মেশিনের উৎপাদন বন্ধ করা হবে যদি মেশিনটি 5% কিংবা তার বেশী ত্রুটিপূর্ণ বস্তু উৎপাদন করে। ত্রুটিপূর্ণ উৎপাদিত বস্তুর অনুপাতের উর্ধ্বসীমা 5% র মধ্যে আছে কিনা তা বিচার করার জন্য উৎপাদিত বস্তুর 20টি সমসম্ভব নমুনা নেওয়া হলো। যদি নমুনাতে ত্রুটিপূর্ণ বস্তুর সংখ্যা 1 এর বেশী হয় তাহলে আমরা মেশিনটি 5% কিংবা তার বেশী ত্রুটিপূর্ণ বস্তু উৎপাদন করে এই যুক্তিকে বর্জন করব।

I. এখানে যথাযথ প্রকল্প কী হবে?

II. বিচারের শক্তি অপেক্ষক বের করুন।

III. শক্তি অপেক্ষক থেকে বিচারের সংশয়মাত্রা বের করুন।

IV. শক্তি অপেক্ষক থেকে দ্বিতীয় প্রকার ভ্রান্তি বের করুন, যখন ত্রুটিপূর্ণ বস্তুর অনুপাত 10%।

৩। $N(\mu, 256)$ নিবেশনের nটি সমসম্ভব নমুনার ভিত্তি $H_0 : \mu = 100$ কে বিচার করতে হবে যখন $H_1 : \mu \neq 100$ ধরা যাক বিচারের বর্জনাঞ্চল $C : |\bar{x} - 100| > K$, K যেখানে একটি ধ্রুবক। যদি প্রথম প্রকার ভ্রান্তির সম্ভাবনা 0.01 এবং দ্বিতীয় প্রকার ভ্রান্তির সম্ভাবনা 0.2 হয় যখন $\mu = 92$। ধ্রুবক K এবং নমুনা আয়তন n বের করুন।

৪। ধরা যাক একটি পূর্ণকের নিবেশন $N(\mu, 4)$, $H_0 : \mu \leq 10$ কে বিচার করতে চাই যেখানে বৈকল্পিক প্রকল্প $H_1 : \mu > 10$ যেখানে $\bar{x}$, নর্মাল নিবেশন $N(\mu, 4)$ এর 9টি অবেক্ষনের গড়, সেক্ষেত্রে বর্জনাঞ্চল $C : \bar{x} \geq 11.5$ এই রূপে বর্ণিত হল।

I. বিচারের সংশয়মাত্রা কী হবে?

II. শক্তি অপেক্ষকটি চিত্রিত কর।

৫। নিম্নলিখিত পদগুলির সংজ্ঞা নির্দেশ করুন।

(i) প্রকল্প বিচার (ii) বিচারের সংশমাত্রা; (iii) প্রথম ও দ্বিতীয় প্রকার ভ্রান্তি; (iv) বিচারের শক্তি ও শক্তি রেখা।

উত্তরসংকেত :

১। এখানে যথার্থ প্রকল্প হবে

$H_0 : \lambda \leq 2$

এবং $H_1 : \lambda > 2$

বর্জনাঞ্চল $C : \bar{x} > 2.5$

এবং $n = 4$

শক্তি অপেক্ষক $\gamma(\lambda) = P_\lambda(\bar{x} > 2.5)$

$$= P_\lambda(\Sigma xi > 4 \times 2.5)$$

$$= P_\lambda(\Sigma xi > 10)$$

$$= 1-\sum_{\gamma=0}^{m} \frac{e^{-4\lambda}(4\lambda)^{\gamma}}{\gamma!}$$

যেহেতু $Y = \Sigma xi \sim P(4\lambda)$

সংশয় মাত্রা $\alpha = \upsilon(2)$

২। I. এখানে যথাযথ প্রকল্প হবে

$H_0 : P \leq 0.05$

$H_1 : P > 0.05$ যেখানে P ত্রুটিপূর্ণ বস্তুর অনুপাত

II. ধরা যাক X = ত্রুটিপূর্ণ বস্তুর সংখ্যা, n = 20 দেওয়া আছে

শক্তি অপেক্ষক $\gamma(p) = P_p(X \geq 2)$

$= 1 - P_p(X = 0) - P_p(X = 1)$

III. $1 - (1 - P)^{20} - \binom{20}{1}P(1 - P)^{19}$ α = সংশয় মাত্রা = $\gamma(0.05)$

$$= 1 - (1 - .05)^{20} - \binom{20}{1}.05\,(1 - .05)^{19}$$

$$= 1 - (1 - .05)^{20} - \binom{20}{1}(.05)\,(.95)^{19}$$

IV. $\beta(0.1)$ = দ্বিতীয় প্রকার ভ্রান্তির সম্ভাবনা যখন ত্রুটিপূর্ণ বস্তুর সংখ্যা 10

$$= 1 - \gamma(0.1)$$

$$= (0.9)^{20} + 20(0.1)\,(0.9)^{19}$$

৩। $\alpha l(100) = P_{200}(|\bar{X} - 100| > K) = 0.01$

$$\Rightarrow \frac{K}{16/\sqrt{n}} = Z_{0.005}\,\ldots\ldots\ldots\ldots(1)$$

$\beta(92) = P_{92}(|\bar{X} - 100| \leq K) = 0.02$

$$\Rightarrow \Phi\left(\frac{8+K}{16/\sqrt{n}}\right) - \Phi\left(\frac{8-K}{16/\sqrt{n}}\right) = 0.2\,\ldots\ldots\ldots\ldots(2)$$

(1) এবং (2) থেকে, নর্ম্যাল নিবেশনের সারণি ব্যবহার করে পাওয়া যাবে K = 46 এবং n = 89.

৪। I. $\alpha(10) = P_{10}(\bar{X} \geq 11.5)$

$$= 1 - \Phi\left(\frac{11.5-10}{2/\sqrt{9}}\right)$$

$$= 0.0122$$

II. $\gamma(\mu) = P_\mu(\bar{X} \geq 11.5) = 1 - \Phi\left(\frac{11.5-\mu}{2/\sqrt{9}}\right), \; -\infty < \mu < \infty$

নর্ম্যাল t, f বিচার সম্বন্ধীয় অনুশীলনী ও উত্তর সংকেত ঃ

১। কোন একটি শিক্ষা প্রতিষ্ঠানের ছাত্রদের পরিসংখ্যানে প্রাপ্ত গড় নম্বর 60 এবং সমক বিচ্যুতি 10। কমপিউটার দিয়ে পড়ানোর নতুন পদ্ধতি ব্যবহার করার ফলে 52 জন ছাত্রদের প্রাপ্ত গড় নম্বর দেখা গেল 62.75। এই রাশিতথ্য ব্যবহার করে পুরানো পদ্ধতি নতুনের থেকে ভাল কিনা তা বিচার করুন। (সংশয়মাত্রা যখন 5%)।

২। একটি নতুন স্বয়ংক্রিয় কাগজ ছাপার মেশিন ক্রয়ের জন্য সিদ্ধান্ত নিতে হবে। যদি গ্রহণযোগ্য কাগজের গড় সংখ্যা/ঘন্টা 1000 এর বেশী হয় তাহলে আমরা নতুন মেশিনটি সাশ্রয়ী বলতে পারি। এখন মুখ্য প্রকল্প $H_0 : \mu = 1000$ এবং বৈকল্পিক প্রকল্প $H_1 : \mu > 1000$ কে বিচার করার জন্য পরীক্ষামূলকভাবে মেশিনটিকে 1 ঘন্টা করে 64 বার চালানো হল। দেখা গেল গ্রহণযোগ্য কাগজের গড় সংখ্যা এবং সম্যক বিচ্যুতি যথাক্রমে 1028 এবং 126। এই রাশিতথ্য ব্যবহার করে প্রকল্প বিচার করুন। (সংশয়মাত্রা 5%)

৩। একটি টুথ পেষ্ট প্রস্তুতকারী সংস্থার দাবী যে তাদের তৈরী পেষ্ট 75% দন্তচিকিৎসক সুপারিশ করে। কিন্তু অন্য একটি সংস্থা মনে করে যে এই দাবী যথার্থ নয়। 390 জন দন্তচিকিৎসকের নমুনা নিয়ে দেখা গেল যে 270 জন ঐ পেষ্ট সুপারিশ করে। এই রাশিতথ্যের সাহায্যে দাবীর যথার্থতা বিচার করুন।

৪। ধরা যাক একটি ফুটকি চিহ্নিত ঘুঁটি (ডাই)-এর 6টি ফুটকি আসার সম্ভবনা P। যদি ঘুঁটিটি পক্ষপাতশূন্য হয়, তখন $P = \frac{1}{6}$ হয়। ঘুঁটিটিকে 150 বার ছোঁড়ার ফলে যদি 30 বার 6টা ফুটকি আসে তখন কি ঘুঁটিটিকে পক্ষপাতশূন্য বলা যাবে? (ধরে নিন $\alpha = 0.01$)

৫। পূর্ববর্তী রাশিতথ্য থেকে জানা আছে একজন অভিজ্ঞ ইনস্পেক্টরের পরিমাপগুলির ভেদমান 0.18 বর্গ সে. মি.। একজন নতুন ইনস্পেক্টর 100টি পরিমাপ নেওয়ার পর দেখা গেল পরিমাপগুলির ভেদমান 0.13 বর্গ সেমি। বিচার করে দেখুন যে নতুন ইনস্পেক্টরের পরিমাপগুলি কি সন্তোষজনক? (সংশয়মাত্রা যখন 5%)

৬। যেহেতু ফুটবল ম্যাচে মাঝে মাঝেই অতিরিক্ত সময়ের প্রয়োজন হয়, তাই একটি কোম্পানী পরীক্ষা করে দেখতে চায় যে একটি ফুটবল ম্যাচ টেলিভিশনে সরাসরি সম্প্রচার করতে হলে কতটা সময়ের দরকার। ফুটবল ম্যানেজাররা দাবী করেন অতিরিক্ত সময়ের সম্যক বিচ্যুতি 8 মিঃ। এই দাবীর যথার্থতা পরীক্ষা করার জন্য টেলিভিশন প্রোগ্রাম ডিরেক্টর 15টি পূর্ববর্তী খেলার একটি নমুনা নেন

145, 125, 120, 136, 137, 146, 123, 134, 129, 109, 140, 135, 145, 138, 134

দাবীর যথার্থতা বিচার করুন যখন সংশয়মাত্রা 10%

৭। খেলাধূলায় উৎসাহী 50টি ছাত্রের গড় দৈর্ঘ্য 68.2 ইঞ্চি এবং সম্যক বিচ্যুতি 2.5 ইঞ্চি, আবার খেলাধূলায় উৎসাহী নয় এমন 50টি ছাত্রের গড় দৈর্ঘ্য 67.5 ইঞ্চি এবং সম্যক বিচ্যুতি 2.8 ইঞ্চি। যেখানে সংশয়মাত্রা 5% এ বিচার করুন খেলাধূলায় উৎসাহ উচ্চতা বৃদ্ধির সহায়ক কি না।

৮। 10টি শূকরকে খাদ্য A এবং 12টি শূকরকে খাদ্য B একটি নির্দিষ্ট সময় খাওয়ানোর পর তাদের ওজনের বৃদ্ধি হয়েছে যথাক্রমে

খাদ্য A 10, 6, 16, 17, 13, 12, 8, 14, 15, 9

খাদ্য B 7, 13, 22, 15, 12, 14, 18, 8, 21, 23, 10, 17.

বিচার করুন I $H_0 : \sigma_1^2 = \sigma_2^2$

এবং $H_1 : \sigma_1^2 \neq \sigma_2^2$ (যেখানে সংশয়মাত্রা 1%)

II দুটি ক্ষেত্রে গড় ওজন বৃদ্ধির পরিমাণ এক কিনা তা বিচার করুন যখন সংশয়মাত্রা 5%

[ধরে নিন ওজনের নিবেশন দুটি ক্ষেত্রে যথাক্রমে $N(\mu_1, \sigma_1^2)$ এবং $N(\mu_2, \sigma_2^2)$]

৯। গুদামের ভিতরে এবং বাইরে মজুত করা কাঠ কতটা শক্ত, তার পরিমাপ নীচের সারণিতে দেওয়া আছে।

	ভিতর	বাহির
নমুনা আয়তন	13	11
নমুনা গড়	166	103
নমুনা ভেদমান	3500	2400

যখন সংশয়মাত্র 5%, গুদামের ভিতরে মজুত করা কাঠ, বাইরে মজুত করা কাঠের থেকে শক্ত হবে কিনা তা বিচার করুন।

১০। ওজন হ্রাসের একটি ওষুধের কার্যকারিতা দেখার জন্য ঐ ওষুধ 12 জন ব্যক্তির উপর প্রয়োগ করা হল। ওষুধ শুরু করার পর দ্বিতীয় ও চতুর্থ মাসে ঐ সকল ব্যক্তির গড় মাসিক ওজন হ্রাস নীচের সারণিতে দেওয়া হল।

ব্যক্তি নং	1	2	3	4	5	6	7	8	9	10	11	12
2য় মাস	3.71	4.58	1.26	7.25	5.43	3.49	2.66	1.32	3.93	12.50	7.95	2.66
4র্থ মাস	2.80	4.62	1.83	5.93	4.36	2.97	2.47	1.54	3.50	12.03	7.51	2.48

প্রতিমাসে ওষুধ ব্যবহার করার ফলে ওজন হ্রাসের পরিমাণ সময়ের সঙ্গে বাড়ছে কিনা তা বিচার করুন, যেখানে সংশয়মাত্রা 5%।

১১। কোনও একটি খেলায় 10টি গেমে 2 জন খেলোয়াড়ের অর্জিত পয়েন্ট নীচে দেওয়া আছে—

খেলোয়াড় 1 :	13	15	7	15	5	12	9	3	20	11
খেলোয়াড় 2 :	12	7	2	8	6	9	5	7	6	8

2 নং খেলোয়াড় কি 1 নং খেলোয়াড়ের থেকে বেশী ধারাবাহিক? [ধরে নিন $\alpha = 0.05$]

১২। একটি শহরে 600 জনের নমুনা নিয়ে দেখা গেল 400 জন ধূমপায়ী। অপরপক্ষে অন্য একটি শহরে 900 জনের নমুনা নিয়ে দেখা গেল 450 জন ধূমপায়ী। আমরা কি এই সিদ্ধান্তে উপনীত হতে পারি যে প্রথম শহরের ধূমপায়ীর অনুপাত দ্বিতীয় শহরের থেকে বেশী? যেখানে সংশয় মাত্রা 1%।

১৩। একটি মেশিনে উৎপাদিত বস্তুর 500টির নমুনা নিয়ে দেখা গেল 16টি বস্তু ত্রুটিপূর্ণ। উপযুক্ত রক্ষণাবেক্ষণের জন্য প্রতিশোধক নেওয়ার ফলে দেখা গেল 100টি নমুনার মধ্যে 3টি নমুনা ত্রুটিপূর্ণ হচ্ছে। মেসিনটি প্রতিশোধক নেওয়ার ফলে উন্নত হল কিনা বিচার করুন [$\alpha = 0.05$]

১৪। দুই ব্যক্তির পরিমাপের নির্ভুলতা বিচার করার জন্য প্রথম ব্যক্তি এবং দ্বিতীয় ব্যক্তিকে যথাক্রমে 25টি ও 30টি মাপ করতে দেওয়া হল। প্রথম ও দ্বিতীয় ব্যক্তির মাপগুলির সম্যক বিচ্যুতি যথাক্রমে 1.34 এবং 0.98, বিচার করুন দুই ব্যক্তির পরিমাপের নির্ভুলতা একই কি না। [ধরে নিন $\alpha = 0.1$]

উত্তর সংকেত :

১। ধরা যাক μ নতুন পদ্ধতিতে পড়ানো একটি ছাত্রের গড় নম্বর। তখন যথার্থ প্রকল্পগুলি হবে $H_0 : \mu = 60$ এবং $H_1 : \mu > 60$

দেওয়া আছে নমুনা আয়তন $n = 52$ (বৃহৎ)

এবং $\sigma = 10$

এখানে দক্ষিণ পুচ্ছ নর্ম্যাল বিচার প্রয়োগ করতে হবে।

$$Z = \frac{62.75 - 60}{10 / \sqrt{52}} = 1.983$$

$$Z_{0.05} = 1.64$$

যেহেতু $1.983 > 1.64$

সুতরাং $Z > Z_{0.05}$

অতএব H_0 বর্জন করব। অর্থাৎ নতুন শিক্ষার পদ্ধতি ছাত্রদের নম্বর বৃদ্ধিতে সাহায্য করবে।

২। দেওয়া আছে $H_0 : \mu = 1000$

এবং $H_1 : \mu > 1000$

নমুনা আয়তন $n = 64$ (বৃহৎ), $\overline{X} = 1028$ এবং $s = 126$ এখানে দক্ষিণ পুচ্ছ প্রায় সঠিক নর্ম্যাল বিচার ব্যবহার করতে হবে।

$$Z = \frac{1028 - 1000}{126 / \sqrt{64}} = 1.78$$

এবং $Z_{0.05} = 1.64$

যেহেতু $1.78 > 1.64$ অর্থাৎ $Z > Z_{.05}$

সুতরাং H_0 বর্জন করব।

অর্থাৎ নতুন মেশিনটি সাশ্রয়ী।

এখানে $H_0 : P = 0.75$

এখানে $H_1 : P < 0.75$

প্রায় সঠিক বামপুচ্ছ নর্ম্যাল বিচার এখানে ব্যবহার করা হবে,

$$Z = \frac{\frac{270}{390} - 0.75}{\sqrt{\frac{\frac{270}{390}\left(1 - \frac{120}{390}\right)}{390}}} = -2.468$$

$$Z_{0.05} = 1.64$$

যেহেতু, $-2.468 < 1.64$

অর্থাৎ $Z < Z_{0.05}$

সুতরাং H_0 বর্জিত হবে।

অতএব টুথপেষ্ট প্রস্তুতকারী সংস্থার দাবী যথার্থ নয়।

৪। এখানে $H_0 : P = \frac{1}{6}$

এবং $H_1 : P \neq \frac{1}{6}$

যেহেতু $n = 150$ (বৃহৎ), প্রায় সঠিক উভপুচ্ছ নর্ম্যাল বিচার এখানে ব্যবহার করা হবে।

$$|Z| = \left| \frac{\frac{30}{150} - \frac{1}{6}}{\frac{\sqrt{\frac{1}{6} \times \frac{5}{6}}}{150}} \right| = 1.095$$

$Z_{0.005} = 2.58$

যেহেতু $1.095 < 2.58$

অর্থাৎ $|Z| < Z_{0.005}$

সুতরাং H_0 গৃহীত হবে।

অতএব রাশিতথ্য ঘুঁটিটির পক্ষপাতশূন্যতা সম্পর্কীত প্রকল্প সমর্থন করে।

৫। এখানে যথার্থ মুখ্য প্রকল্প

$H_0 : \sigma^2 = 0.18$

এবং বৈকল্পিক প্রকল্প $H_1 : \sigma^2 \neq 0.18$

$n = 100$ (বৃহৎ)। প্রায় সঠিক উভয়পুচ্ছ নর্ম্যাল বিচার এখানে ব্যবহার করা হল।

দেওয়া আছে $s^2 = 0.13$

অতএব $|Z| = \left| \frac{s - \sigma_0}{\sigma_0 / \sqrt{200}} \right| = \left| \frac{0.36 - 0.42}{0.42 / \sqrt{200}} \right| = 2.02$

$Z_{0.025} = 1.96$

যেহেতু 2.02 > 1.96

সুতরাং $|Z| > Z_{0.025}$, H_0 বর্জিত হবে।

অতএব আমরা বলতে পারি যে নতুন ইনস্পেক্টরের পরিমাপগুলি সন্তোষজনক নয়।

৬। এখানে যথার্থ প্রকল্প হবে

$$H_0 : \sigma^2 = 64$$

এবং $H_1 : \sigma^2 > 64$

নমুনা আয়তন n = 15 (ছোট)

ধরা যাক খেলার সময়ের নিবেশন নর্মাল নিবেশন। এখানে দক্ষিণপুচ্ছ χ^2 বিচার প্রয়োগ করা হবে।

দেওয়া আছে $n = 15$, $s^2 = 106.29$

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

$$= \frac{14 \times 106.21}{64} = 23.233$$

χ^2 নিবেশনের সারণি পাই

$$\chi^2_{0.10,14} = 21.064$$

যেহেতু $\chi^2 > \chi^2_{0.10,14}$

সেহেতু H_0 বর্জিত হবে।

আমরা এই সিদ্ধান্তে উপনীত হতে পারি যে খেলার সময়ের সম্যক বিচ্যুতি' ৪ মিঃ এর বেশী।

৭। ধরা যাক, μ_1 এবং μ_2 যথাক্রমে যারা খেলাধুলায় উৎসাহী এবং উৎসাহী নয় এমন ছাত্রদের গড় উচ্চতা।

এখানে $H_0 : \mu_1 = \mu_2$

এবং $H_1 : \mu_1 > \mu_2$

দেওয়া আছে $n_1 = 50$, $n_2 = 50$, $\overline{X} = 68.2$, $s_1 = 2.5$, $\overline{Y} = 67.5$, $s_2 = 2.8$

যেহেতু নমুনা আয়তনগুলি বৃহৎ, আমরা উভপুচ্ছ প্রায় সঠিক নর্ম্যাল বিচার ব্যবহার করব।

$$Z = \frac{\overline{X} - \overline{Y} - \delta}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{68{:}2 - 67.5 - 0}{\sqrt{\frac{(2.5)^2}{50} + \frac{(2.8)^2}{50}}} = 1.32$$

নর্ম্যাল নিবেশনের সারণি থেকে পাই $Z_{0.05} = 1.64$

যেহেতু $Z < 1.64$

সুতরাং H_0 গৃহীত হবে।

অতএব, খেলাধূলায় উৎসাহী হলে উচ্চতা বেশী হবে— এই মত রাশিতথ্য সমর্থন করে না।

৮। ধরা যাক, μ_1 এবং μ_2 যথাক্রমে খাদ্য A এবং খাদ্য B প্রয়োগের ফলে শূকরের গড় ওজন বৃদ্ধি। দেওয়া আছে ওজনের নিবেশন দুটি ক্ষেত্রে $N(\mu_1, \sigma_1^2)$ এবং $N(\mu_2, \sigma_2^2)$

$n_1 = 10,\ n_2 = 12,\ \overline{X} = 12,\ \overline{Y} = 15,\ (n_1 - 1)\, s_1^2 = 120,\ (n_2 - 1)s_2^2 = 314$

I. এখানে $H_0 : \sigma_1^2 = \sigma_2^2$

$H_1 : \sigma_1^2 \neq \sigma_2^2$

এখানে যথাযথ বিচার হবে উভপুচ্ছ F বিচার, $F = \dfrac{s_1^2}{s_2^2} = \dfrac{120}{9} / \dfrac{314}{11} = 0.467$

F নিবেশনের সারণি থেকে পাই, $F_{0.01,\ 9,\ 11} = 4.62$

এখানে $F < 4.62$

সুতরাং H_0 গৃহীত হবে।

II. দুটি ক্ষেত্রে গড় ওজন বৃদ্ধির পরিমাণ সমান কিনা তা বিচার করার জন্য আমরা উভপুচ্ছ t বিচার ব্যবহার করব।

$$|t| = \left| \frac{\overline{X} - \overline{Y}}{s\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \right|,\ s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$= \frac{120 + 314}{20}$$

$$= 21.7$$

সুতরাং,

$$|t| = \left| \frac{12 - 15}{\sqrt{21.7\left(\dfrac{1}{10} + \dfrac{1}{12}\right)}} \right| = 0.323$$

t নিবেশনের সারণি থেকে পাই

$t_{0.025,\ 20,0} = 2.09$

যেহেতু $|t| < 2.09$, সুতরাং H_0 গৃহীত হবে।

অতএব বলা যেতে পারে যে দুটি ক্ষেত্রেই গড় ওজন বৃদ্ধির পরিমাপ একই।

৯। গুদামের বহিরে ও ভিতরে সংরক্ষিত কাঠ কতটা শক্ত তার গড় ধরা যাক যথাক্রমে μ_1 এবং μ_2 এবং ধরা যাক কঠিনতা নর্ম্যাল নিবেশন মেনে চলে।

প্রথমে আমরা বিচার করব

$$H_0 : \sigma_1^2 = \sigma_2^2$$

এবং $H_1 : \sigma_1^2 \neq \sigma_2^2$

এখানে যথাযথ বিচার হবে উভপুচ্ছ F বিচার।

$$F = \frac{2400}{3500} = 0.686$$

F নিবেশনের সারণি থেকে পাই—

$$F_{0.01,\ 10,\ 11} = 4.30$$

যেহেতু $F < 4.30$

সুতরাং H_0 গৃহীত হবে, অর্থাৎ $\sigma_1^2 = \sigma_2^2$

এবার আমরা বিচার করব $H_0 : \mu_1 - \mu_2 \geq 0$

এবং $H_1 : \mu_1 - \mu_2 < 0$

এখানে যথাযথ বিচার হবে বামপুচ্ছ t বিচার।

$$t = \frac{\overline{X} - \overline{Y} - 0}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

যেখানে $\overline{X} = 103$

$$\overline{Y} = 166$$

$$s^2 = \frac{(10 \times 2400) + (12 \times 3500)}{22}$$

$n_1 = 11,\ n_2 = 13$

অতএব $$t = \frac{103 - 166 - 0}{\sqrt{\frac{(10 \times 2400) + (12 \times 3500)}{22}\left[\frac{1}{11} + \frac{1}{13}\right]}} = -2.81$$

t নিবেশনের সারণি থেকে পাই—

$$t_{0.05,\ 22} = 1.717$$

যেহেতু $-2.81 < -1.717$

H_0 বর্জিত হবে।

সুতরাং আমরা বলতে পারি যে কাঠ গুদামের ভিতরে মজুত থাকলে তা বেশী শক্ত হবে।

১০। ধরা যাক

X = 2য় মাসের ওজন হ্রাসের পরিমাণ

Y = 4র্থ মাসের ওজন হ্রাসের পরিমাণ

$d = X - Y$

12 জন ব্যক্তির ক্ষেত্রে এই বিভেদ (d)-এর পরিমাপগুলি হল—

$d_1 = 0.91$, $d_2 = 0.132$, $d_3 = 0.19$, $d_4 = 0.47$, $d_5 = 0.04$, $d_6 = 0.107$, $d_7 = -0.22$, $d_8 = 0.44$, $d_9 = -0.57$, $d_{10} = 0.52$, $d_{11} = 0.43$, $d_{12} = 0.18$

এখানে যথার্থ মুখ্য প্রকল্প

$$H_0 : \mu_1 - \mu_2 \le 0$$

এবং $H_1 : \mu_1 - \mu_2 > 0$

$$t = \frac{\bar{d} - 0}{S_d / \sqrt{n}} = 2.52$$

আবার $t_{0.05,\ 11} = 1.796$

যেহেতু $t > 1.796$, আমরা H_0-কে বর্জন করব। অতএব, বলা যেতে পারে ওজন হ্রাসের পরিমাপ সময়ের সঙ্গে সঙ্গে বাড়ছে।

১১। এখানে যথার্থ মুখ্য প্রকল্প হবে

$$H_0 : \sigma_1^2 \le \sigma_2^2$$

এবং $H_1 : \sigma_1^2 > \sigma_2^2$

এখানে $n_1 = n_2 = 10$, রাশিতথ্য থেকে পাই $s_1^2 = 16.44$, $s_2^2 = 6.49$

এখানে যথাযথ বিচার হবে দক্ষিণপুচ্ছ F বিচার।

$$F = \frac{s_1^2}{s_2^2} = 2.39$$

আবার $F_{0.05, 9,9} = 3.19$

যেহেতু $F < 3.19$, আমরা 5% সংশয়মাত্রার ক্ষেত্রে H_0-কে বর্জন করব না। সুতরাং ২য় খেলোয়াড়কে 1ম খেলোয়াড়ের থেকে বেশী ধারাবাহিক বলা যাবে না।

১২। ধরা যাক প্রথম ও দ্বিতীয় শহরে ধূমপায়ীদের অনুপাত যথাক্রমে P_1 ও P_2

আমরা বিচার করতে চাই $H_0 : P_1 \leq P_2$

এবং $H_1 : P_1 > P_2$

এখানে দক্ষিণ পুচ্ছ নর্ম্যাল বিচার প্রয়োগ করা হবে। যখন $n_1 = 600$, $n_2 = 900$

$$\hat{P}_1 = \frac{800}{600},\ \hat{P}_2 = \frac{450}{900},\ \hat{P} = \frac{850}{1500}$$

$$Z = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{P}\left(1-\hat{P}\right)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = 6.55$$

আমরা নর্মাল নিবেশনের সারণি থেকে পাই—

$Z_{0.01} = 2.33$

যেহেতু $Z > 2.33$

অতএব H_0 বর্জিত হবে।

আমরা বলতে পারি যে প্রথম শহরের ধূমপায়ীর অনুপাত দ্বিতীয় শহরের থেকে বেশী।

১৩। মেশিনে উৎপাদিত বস্তুর প্রতিশোধক নেবার আগে ও পরে ত্রুটিপূর্ণ হবার অনুপাত যথাক্রমে P_1 এবং P_2

এখানে মুখ্য প্রকল্প

$H_0 : P_1 \leq P_2$

এবং $H_1 : P_1 > P_2$

দেওয়া আছে $n_1 = 500$, $n_2 = 100$

$$\hat{P}_1 = \frac{16}{500},\ \hat{P}_2 = \frac{3}{100}$$

$$\hat{P} = \frac{3+16}{100+500} = \frac{19}{600}$$

এখানে যথাযথ বিচার হবে দক্ষিণ পুচ্ছ নর্ম্যাল বিচার

$$Z = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{P}(1-\hat{P})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = 1.1$$

আবার $Z_{0.05} = 1.645$

যেহেতু $Z < 1.645$

সুতরাং H_0 গৃহীত হবে।

প্রতিশোধক নেবার ফলে মেশিনটি উন্নত হল না।

১৪। এখানে আমরা বিচার করতে চাই

$H_0 : \sigma_1^2 = \sigma_2^2$

এবং $H_1 : \sigma_1^2 \neq \sigma_2^2$

এখানে আমরা উভপুচ্ছ F বিচার প্রয়োগ করে পাই—

যখন $n_1 = 25,\ n_2 = 30,\ s_1 = 1.34,\ s_2 = 0.98$

$$F = \frac{s_1^2}{s_2^2} = 1.87$$

এখানে F নিবেশনের সারণি থেকে পাই—

$F_{0.05;\ 24,\ 29} = 1.89$

যেহেতু $F < 1.89$

সুতরাং H_0 গৃহীত হবে।

রাশিতথ্য বিচার করে বলা যেতে পারে দুই ব্যক্তির পরিমাপের নির্ভুলতা সমান।

১৫। t, F, ও χ^2 নমুনাঙ্কগুলির বিভিন্ন ব্যবহারগুলি আলোচনা করুন।

১৬। একটি পূর্ণকের গড় বা দুটি পূর্ণকের গড়ের পার্থক্য সম্পর্কীয় প্রকল্প কীভাবে বিচার করবেন? বিভিন্ন ক্ষেত্রে শর্তসমূহ লিখুন।

১৭। একটি পূর্ণকের সমক বিচ্যুতি বা দুটি পূর্ণকের সমকবিচ্যুতির অনুপাত সম্বন্ধে বিচারগুলি আলোচনা করুন। বিভিন্ন ক্ষেত্রে শর্তসমূহ লিখুন।

১৮। একটি পূর্ণকের অনুপাত বা দুটি পূর্ণকের অনুপাতের পার্থক্য সম্বন্ধে বিচারগুলি শর্তসহ লিখুন।

৪৮.৭ গ্রন্থপঞ্জী

১। রাশিবিজ্ঞানের মূলতত্ত্ব—প্রথম ও দ্বিতীয় খণ্ড (১৯৭৬)
ডঃ শৈলেশভূষণ চৌধুরী, ডঃ অরিজিৎ চৌধুরী,
ডঃ বিশ্বনাথ দাস (পশ্চিমবঙ্গ রাজ্য পুস্তক পর্ষৎ)

২। রাশিবিজ্ঞানের প্রয়োগ পদ্ধতি— do–

৩। Fundamentals of Statistics–vol I (1998) & II (2001)
Goon, A. M; Gupta M. K. & Dasgupta, B (World Press)

৪। An Outline of Statistical Theory–Vol I & II (1998)
Goon, A. M; Gupta, M. K & Dasgupta, B (World Press)

পরিশিষ্ট : সারণিসমূহ

সারণি ১: প্রামাণ্য নর্ম্যাল চলকের (গড় 0 ও সমকপার্থক্য 1) নিবেশনের
অক্ষরেখা (ordinate) ও ক্ষেত্রফল (area)

Z	$\phi(z)$	$\Phi(z)$	Z	$\phi(z)$	$\Phi(z)$	Z	$\phi(z)$	$\Phi(z)$
.00	.3989423	.5000000						
.01	.3989223	.5039894	.51	.3502919	.6949743	1.01	.2395511	.8437524
.02	.3988625	.5079783	.52	.3484925	.6984682	1.02	.2371320	.8461358
.03	.3987628	.5119665	.53	.3466677	.7019440	1.03	.2347138	.8484950
.04	.3986233	.5159534	.54	.3448180	.7054015	1.04	.2322970	.8508300
.05	.3984439	.5199388	.55	.3429439	.7088403	1.05	.2298821	.8531409
.06	3982248	.5239222	.56	.3410458	.7122603	1.06	.2274696	.8554277
.07	.3979661	.5279032	.57	.3391243	.7156612	1.07	.2250599	.8576903
.08	.3976677	.5318814	.58	.3371799	.7190427	1.08	.2226535	.8599289
.09	.3973298	.5358564	.59	.3352132	.7224047	1.09	.2202508	.8621434
.10	.3969525	.5398278	.60	.3332246	.7257469	1.10	.2178522	.8643339
.11	.3965360	.5437953	.61	.3312147	.7290691	1.11	.2154582	.8665005
.12	.3960802	.5477584	.62	.3291840	.7323711	1.12	.2130691	.8686431
.13	3955854	.5517168	.63	.3271330	.7356527	1.13	.2106856	.8707619
.14	3950517	.5556700	.64	.3250623	.7389137	1.14	.2083078	.8728568
.15	.3944793	.5596177	.65	.3229724	.7421539	1.15	.2059363	.8749281

Z	ϕ(z)	Φ(z)	Z	ϕ(z)	Φ(z)	Z	ϕ(z)	Φ(z)
.16	.3938684	.5635595	.66	.3208638	.7453731	1.16	.2035714	.8769756
.17	.3932190	.5674949	.67	.3187371	.7485711	1.17	.2012135	.8789995
.18	.3925315	.5714237	.68	.3165929	.7517478	1.18	.1988631	.8809999
.19	.3918060	.5753454	.69	.3144317	.7549029	1.19	.1965205	.8829768
.20	.3910427	.5792597	.70	.3122539	.7580363	1.20	.1941861	.8849303
.21	.3902419	.5831662	.71	.3100603	.7611479	1.21	.1918602	.8868606
.22	.3894038	.5870644	.72	.3078513	.7642375	1.22	.1895432	.8887676
.23	.3885286	.5909541	.73	.3056274	.7673049	1.23	.1872354	.8906514
.24	.3876166	.5948349	.74	.3033893	.7703500	1.24	.1849373	.8925123
.25	.3866681	.5987063	.75	.3011374	.7733726	1.25	.1826491	.8943502
.26	.3856834	.6025681	.76	.2988724	.7763727	1.26	.1803712	.8961653
.27	.3846627	.6064199	.77	.2965948	.7793501	1.27	.1781038	.8979577
.28	.3836063	.6102612	.78	.2943050	.7823046	1.28	.1758474	.8997274
.29	.3825146	.6140919	.79	.2920038	.7852361	1.29	.1736022	.9014747
.30	.3813878	.6179114	.80	.2896916	.7881446	1.30	.1713686	.9031995
.31	.3802264	.6217195	.81	.2873689	.7910299	1.31	.1691468	.9049021
.32	.3790305	.6255158	.82	.2850364	.7938919	1.32	.1669370	.9065825
.33	.3778007	.6293000	.83	.2826945	.7967306	1.33	.1647397	.9082409
.34	.3765372	.6330717	.84	.2803438	.7995458	1.34	.1625551	.9098773
.35	.3752403	.6368307	.85	.2779849	.8023375	1.35	.1603833	.9114920
.36	.3739106	.6405764	.86	.2756182	.8051055	1.36	.1582248	.9130850
.37	.3725483	.6443088	.87	.2732444	.8078498	1.37	.1560797	.9146565
.38	.3711539	.6480273	.88	.2708640	.8105703	1.38	.1539483	.9162067
.39	.3697277	.6517317	.89	.2684774	.8132671	1.39	.1518308	.9177356
.40	3682701	.6554217	.90	.2660852	.8159399	1.40	.1497275	.9192433
.41	.3667817	.6590970	.91	.2636880	.8185887	1.41	.1476385	.9207302
.42	.3652627	.6627573	.92	.2612863	.8212136	1.42	.1455641	.9221962
.43	.3637136	.6664022	.93	.2588805	.8238145	1.43	.1435046	.9236415
.44	.3621349	.6700314	.94	.2564713	.8263912	1.44	.1414600	.9250663
.45	.3605270	.6736448	.95	.2540591	.8289439	1.45	.1394306	.9264707
.46	.3588903	.6772419	.96	.2516443	.8314724	1.46	.1374165	.9278505
.47	.3572253	.6808225	.97	.2492277	.8339768	1.47	.1354181	.9292191
.48	.3555325	.6843863	.98	.2468095	.8364569	1.48	.1334353	.9305634
.49	.3538124	.6879331	.99	.2443904	.8389129	1.49	.1314684	.9318879
.50	.3520653	.6914625	1.00	.2419707	.8413447	1.50	.1295176	.9331928
1.51	.1275830	.9344783	2.01	0529192	.9777844	2.51	.0170947	.9939634
1.52	.1256646	.9357445	2.02	.0518636	.9783083	2.52	.0166701	.9941323

Z	φ(z)	Φ(z)	Z	φ(z)	Φ(z)	Z	φ(z)	Φ(z)
1.53	.1237628	.9369916	2.03	.0508239	.9788217	2.53	.0162545	.9942969
1.54	.1218775	.9382198	2.04	.0498001	.9793248	2.54	.0158476	.9944574
1.55	.1200090	.9394292	2.05	.0487920	.9798178	2.55	.0154493	.9946139
1.56	.1181573	.9406201	2.06	.0477996	.9803007	2.56	.0150596	.9947664
1.57	.1163225	.9417924	2.07	.0468226	.9807738	2.57	.0146782	.9949151
1.58	.1145048	.9429466	2.08	.0458611	.9812372	2.58	.0143051	.9950600
1.59	.1127042	.9440826	2.09	.0449148	.9816911	2.59	.0139401	.9952012
1.60	.1109208	.9452007	2.10	.0439836	.9821356	2.60	.0135830	.9953388
1.61	.1091548	.9463011	2.11	.0430674	.9825708	2.61	.0132337	.9954729
1.62	.1074061	.9473839	2.12	.0421661	.9829970	2.62	.0128921	.9956035
1.63	1056748	.9484493	2.13	.0412795	.9834142	2.63	.0125581	.9957308
1.64	.1039611	.9494974	2.14	.0404076	.9838226	2.64	.0122315	.9958547
1.65	.1022649	.9505285	2.15	.0395500	.9842224	2.65	.0119122	.9959754
1.66	.1005864	.9515428	2.16	.0387069	.9846137	2.66	.0116001	.9960930
1.67	.0989255	.9525403	2.17	.0378779	.9849966	2.67	.0112951	.9962074
1.68	0972823	.9535213	2.18	.0370629	.9853713	2.68	.0109969	.9963189
1.69	.0956568	.9544860	2.19	.0362619	.9857379	2.69	.0107056	.9964274
1.70	.0940491	.9554345	2.20	.0354746	.9860966	2.70	.0104209	.9965330
1.71	.0924591	.9563671	2.21	.0347009	.9864474	2.71	.0101428	.9966358
1.72	.0908870	.9572838	2.22	.0339408	.9867906	2.72	.0098712	.9967359
1.73	.0893326	.9581849	2.23	.0331939	.9871263	2.73	.0096058	.9968333
1.74	.0877961	.9590705	2.24	.0324603	.9874545	2.74	.0093466	.9969280
1.75	.0862773	.9599408	2.25	.0317397	.9877755	2.75	.0090936	.9970202
1.76	.0847764	.9607961	2.26	.0310319	.9880894	2.76	.0088465	.9971099
1.77	.0832932	.9616364	2.27	.0303370	.9883962	2.77	.0086052	.9971972
1.78	.0818278	.9624620	2.28	.0296546	.9886962	2.78	.0083697	.9972821
1.79	.0803801	.9632730	2.29	.0289847	.9889893	2.79	.0081398	.9973646
1.80	.0789502	.9640697	2.30	.0283270	.9892759	2.80	0079155	.9974449
1.81	.0775379	.9648521	2.31	.0276816	.9895559	2.81	.0076965	.9975229
1.82	.0761433	.9656205	2.32	.0270481	.9898296	2.82	.0074829	.9975988
1.83	.0747663	.9663750	2.33	.0264265	.9900969	2.83	.0072744	.9976726
1.84	.0734068	.9671159	2.34	.0258166	.9903581	2.84	.0070711	.9977443
1.85	.0720649	.9678432	2.35	0252182	.9906133	2.85	.0068728	.9978140
1.86	.0707404	.9685572	2.36	.0246313	.9908625	2.86	.0066793	.9978818
1.87	.0694333	.9692581	2.37	.0240556	.9911060	2.87	.0064907	.9979476
1.88	.0681436	.9699460	2.38	.0234910	.9913437	2.88	.0063067	.9980116
1.89	.0668711	.9706210	2.39	.0229374	.9915758	2.89	.0061274	.9980738

Z	$\phi(z)$	$\Phi(z)$	Z	$\phi(z)$	$\Phi(z)$	Z	$\phi(z)$	$\Phi(z)$
1.90	.0656158	.9712834	2.40	.0223945	.9918025	2.90	.0059525	.9981342
1.91	.0643777	.9719334	2.41	.0218624	.9920237	2.91	.0057821	.9981929
1.92	.0631566	.9725711	2.42	.0213407	.9922397	2.92	.0056160	.9982498
1.93	.0619524	.9731966	2.43	.0208294	.9924506	2.93	.0054541	.9983052
1.94	.0607652	.9738102	2.44	.0203284	.9926564	2.94	.0052963	.9983589
1.95	.0595947	.9744119	2.45	.0198374	.9928572	2.95	.0051426	.9984111
1.96	.0584409	.9750021	2.46	.0193563	.9930531	2.96	.0049929	.9984618
1.97	.0573038	.9755808	2.47	.0188850	.9932443	2.97	.0048470	.9985110
1.98	.0561831	.9761482	2.48	.0184233	.9934309	2.98	.0047050	.9985588
1.99	.0550789	.9767045	2.49	.0179711	.9936128	2.99	.0045666	.9986051
2.00	.0539910	.9772499	2.50	.0175283	.9937903	3.00	.0044318	.9986501
3.01	.0043007	.9986938	3.21	.0023089	.9993363	3.41	.0011910	.9996752
3.02	.0041729	.9987361	3.22	.0022358	.9993590	3.42	.0011510	.9996869
3.03	.0040486	.9987772	3.23	.0021649	.9993810	3.43	.0011122	.9996982
3.04	.0039276	.9988171	3.24	.0020960	.9994024	3.44	.0010747	.9997091
3.05	.0038098	.9988558	3.25	.0020290	.9994230	3.45	.0010383	.9997197
3.06	.0036951	.9988933	3.26	.0019641	.9994429	3.46	.0010030	.9997299
3.07	.0035836	.9989297	3.27	.0019010	.9994623	3.47	.0009689	.9997398
3.08	.0034751	.9989650	3.28	.0018397	.9994810	3.48	.0009358	.9997493
3.09	.0033695	.9989992	3.29	.0017803	.9994991	3.49	.0009037	.9997585
3.10	.0032668	.9990324	3.30	.0017226	.9995166	3.50	.0008727	.9997674
3.11	.0031669	.9990646	3.31	.0016666	.9995335	3.51	.0008426	.9997759
3.12	.0030698	.9990957	3.32	.0016122	.9995499	3.52	.0008135	.9997842
3.13	.0029754	.9991260	3.33	.0015595	.9995658	3.53	.0007883	.9997922
3.14	.0028835	.9991553	3.34	.0015084	.9995811	3.54	.0007581	.9997999
3.15	.0027943	.9991836	3.35	.0014587	.9995959	5.55	.0007317	.9998074
3.16	.0027075	.9992112	3.36	.0014106	.9996103	3.56	.0007061	.9998146
3.17	.0026231	.9992378	3.37	.0013639	.9996242	3.57	.0006814	.9998215
3.18	.0025412	.9992636	3.38	.0013187	.9996376	3.58	.0006575	.9998282
3.19	.0024615	.9992886	3.39	.0012748	.9996505	3.59	.0006343	.9998347
3.20	.0023841	.9993129	3.40	.0012322	.9996631	3.60	.0006119	.9998409

সারণি ২: প্রামাণ্য নর্ম্যাল চলকের নিবেশন : z_α -এর মানসমূহ

α	.05	.025	.01	.005
Z_α	1.645	1.960	2.326	2.576

সারণি ৩: χ^2-এর নিবেশন : $\chi^2_{\alpha,\,\nu}$ এর মানসমূহ

ν \ α	0.995	0.99	0.975	0.95	0.05	0.025	0.01	0.005
1	0.000	0.000	0.001	0.004	3.841	5.024	6.635	7.878
2	0.010	0.020	0.051	0.103	5.999	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	11.070	12.832	15.086	16.750
6	0.676	0.872	1.237	1.635	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	30.114	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	33.924	36.781	40.289	42.796
23	9.260	10.196	11.688	13.091	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	36.415	39.364	42.980	45.558
25	10.520	11.524	13.120	14.611	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	40.113	43.194	46.963	49.645
28	12.461	13.565	15.308	16.928	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	43.773	46.979	50.892	53.672
40	20.706	22.164	24.433	26.509	55.759	59.342	63.691	66.766
50	27.991	29.707	32.537	34.764	67.505	71.420	76.154	79.490

ν \ α	0.995	0.99	0.975	0.95	0.05	0.025	0.01	0.005
60	35.535	37.485	40.482	43.188	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	124.342	129.561	135.807	140.169

ν এর বৃহত্তর মানের জন্য $\sqrt{2x^2} - \sqrt{2\nu - 1}$ কে প্রমাণ নর্ম্যাল চলক হিসাবে ধরা যেতে পারে।

সারণী ৪: t-নিবেশন : $t_{\alpha,\nu}$ এর মানসমূহ

ν \ α	0.05	0.025	0.01	0.005
1	6.314	12.706	31.821	63.657
2	2.920	4.303	6.965	9.925
3	2.353	3.182	4.541	5.841
4	2.132	2.776	3.747	4.604
5	2.015	2.571	3.365	4.032
6	1.943	2.447	3.143	3.707
7	1.895	2.365	2.998	3.499
8	1.860	2.306	2.896	3.355
9	1.833	2.262	2.821	3.250
10	1.812	2.228	2.764	3.169
11	1.796	2.201	2.718	3.106
12	1.782	2.179	2.681	3.055
13	1.771	2.160	2.650	3.012
14	1.761	2.145	2.624	2.977
15	1.753	2.131	2.602	2.947
16	1.746	2.120	2.583	2.921
17	1.740	2.110	2.567	2.898
18	1.734	2.101	2.552	2.878
19	1.729	2.093	2.539	2.861
20	1.725	2.086	2.528	2.845
21	1.721	2.080	2.518	2.831
22	1.717	2.074	2.508	2.819
23	1.714	2.069	2.500	2.807
24	1.711	2.064	2.492	2.797
25	1.708	2.060	2.485	2.787
26	1.706	2.056	2.479	2.779

α / ν	0.05	0.025	0.01	0.005
27	1.703	2.052	2.473	2.771
28	1.701	2.048	2.467	2.763
29	1.699	2.045	2.462	2.756
30	1.697	2.042	2.457	2.750
40	1.684	2.021	2.423	2.704
60	1.671	2.000	2.390	2.660
120	1.658	1.980	2.358	2.617
∞	1.645	1.960	2.326	2.576

সারণি ৫: F নিবেশন : $F_{.05}$; ν_1, ν_2 এর মানসমূহ

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

সারণি ৫: F নিবেশন (পূর্বানুসরণ) : $F_{.01}$; ν_1, ν_2 এর মানসমূহ

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	4052	4999.5	5403	5625	5764	5859	5928	5982	6022	6056	6106	6157	6209	6235	6261	6287	6313	6339	6366
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.49	99.50
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.05	26.87	26.69	26.60	26.50	26.41	26.32	26.22	26.13
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	5.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	2.96	2.81	2.66	2.58	2.50	2.42	2.33	2.23	2.13
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38
	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00

ν_1, ν_2 এর অন্যান্য মানের জন্য $1/\nu_1$ ও $1/\nu_2$ -কে অনপেক্ষ চলক ধরে অন্তঃপ্রক্ষেপণ করা যেতে পারে।

4652	3819	8431	2150	2352	2472	0043	3488
9031	7617	1220	4129	7148	1943	4890	1749
2030	2327	7353	6007	9410	9179	2722	8445
0641	1489	0828	0385	8488	0422	7209	4950
8479	6062	5593	6322	9439	4996	1322	4918
9917	3490	5533	2577	4348	0971	2580	1943
6376	9899	9259	5117	1336	0146	0680	4052
7287	0983	3236	3252	0277	8001	6058	4501
0592	4912	3457	8773	5146	2519	3931	6794
6499	9118	3711	8838	0691	1425	7768	9544
0769	1109	7909	4528	8772	1876	2113	4781
8678	4873	2061	1835	5054	5026	2967	6560
0178	7794	6488	7364	4094	1649	2284	7753
3392	0963	6364	5762	0322	2592	3452	9002
0264	6009	1311	5873	5926	8597	9051	8995
4089	7732	8163	2798	1984	1292	0041	2500
9376	7365	7987	1937	2251	3411	6737	0367
3039	3780	2137	7641	4030	1604	2517	9211
8971	8653	1855	5285	5631	2649	6696	5475
0373	4153	5199	5765	2067	6627	3100	5716
9092	4773	0002	7000	7800	2292	2933	6125
2464	1038	3163	3569	7155	2029	2538	7080
3027	6215	3125	5856	9543	3660	0255	5544
5754	9247	1164	3283	1865	5274	5471	1346
4358	3716	6949	8502	1573	5763	5046	7135
7178	8324	8379	7365	4577	4864	0629	5100
5035	5939	3665	2160	6700	7249	1738	2721
3318	0220	3611	9887	4608	8664	2185	7290
9058	1735	7435	6822	6622	8286	8901	5534

7886	5182	7595	0305	4903	3306	8088	3899
3354	8454	7386	1333	5345	6565	3159	3991
3415	7671	0846	7100	1790	9449	6285	2525
3918	5872	7898	6125	2268	1898	0755	6034
6138	9045	6950	8843	6533	0917	6673	5721
3825	1704	2835	4677	4637	7329	3156	3291
1349	0417	9311	9787	1284	0769	8422	1077
4234	0248	7760	6504	2754	4044	0842	9080
6880	3201	7044	3657	5263	0374	7563	6599
0714	5008	5076	1134	5342	1608	5179	0967
3448	6421	3304	0583	1260	0662	7257	0766
5711	7343	7539	3684	9397	5335	4031	1486
2588	3301	0553	2427	3598	2580	7017	9176
8581	4253	7404	5264	5411	3431	3092	8573
8475	6322	3949	9675	6533	1133	8776	2216
0272	5624	8549	5552	7469	2799	2822	9620
7383	7795	7939	2652	4456	6993	2950	8573
5126	2089	7729	0945	3901	4445	7117	8186
2064	3760	0939	7319	5939	3432	2030	4752
9315	8185	7805	6294	7072	6491	4012	1016
6814	8752	3462	6001	3302	3895	7371	3432
4433	0247	9747	0412	3893	2503	2972	4154
9193	7314	1501	4702	7030	9601	0630	3727
4246	0693	6041	0931	2952	4968	8239	7729
6974	1051	8966	5157	2154	9558	7646	3043
5673	1602	8741	0513	8713	6108	7329	7698
7370	7319	4104	6025	4209	5042	4501	7824
6934	0165	3319	6222	4129	6524	4322	9422
1592	6953	7868	5874	0805	1138	9428	0189
4683	7249	1998	0956	8325	4001	2261	8844

4206	3295	1732	6780	8409	6957	5292	5041
5885	3316	1187	1217	3912	1107	7220	0035
2584	4222	9438	9652	0338	9712	8715	9587
1275	5976	4273	4895	5751	3112	5082	6050
6801	1709	0038	1231	5222	2473	8909	9970
6853	9282	1196	0347	3135	5902	2384	7929
3210	4345	4448	0229	0371	8269	4448	3348
1684	5742	1897	2503	1656	5702	4613	4108
2391	2897	3406	4844	8756	8011	0246	3663
2543	3913	1429	6379	3369	9040	5983	0436
6793	5986	8153	0769	3347	4014	7007	9018
8118	4646	9668	3408	8878	3534	5549	6929
4970	2717	9943	1136	9504	0519	5240	0991
4496	1109	8238	9173	6244	7230	0991	1463
9022	5050	5383	9582	1326	2516	5589	4051
4816	1007	1067	2866	7916	2674	5578	1675
8897	4869	3221	3266	3567	3365	3675	2195
4234	7491	8194	5072	6555	0799	1940	1232
6933	5786	6675	7853	8325	9408	3252	6799
0502	3633	7793	1529	4067	5459	8641	3247
6440	9450	8896	1441	7718	4849	3192	5958
1248	0405	4572	6861	3737	9558	1025	8707
3110	1168	6046	5837	6243	6745	2362	7710
8822	3604	7844	2085	7923	7979	0648	9003
8680	1201	2536	0308	8733	9722	4556	4684
5327	1250	9502	0340	9894	0438	2677	9200
3798	0805	8037	7474	0516	8715	8398	5552
2688	7601	3408	6525	2710	4547	9156	1623
8552	8348	7934	1530	3523	6882	4334	7237
8713	5638	7620	3148	4508	3123	4023	4560
2104	4716	7582	4576	8105	7527	9082	2426

6503	8499	3100	2209	3406	6314	6910	8051
0085	0711	9557	8428	4332	9685	6492	7422
3822	3407	5603	5431	0083	7074	6929	7054
2193	9184	4815	0566	1214	8433	2282	0916
5392	1390	7100	4578	5107	7946	4502	2765
4635	6166	3085	4297	8619	0912	6917	5364
0495	3715	6053	1723	0114	8257	4650	9901
3296	3067	3040	0852	2939	4015	6927	7710
1348	5573	7270	6840	7450	5933	6472	3750
3132	2603	5574	1528	8104	5520	7279	7940

মানুষের জ্ঞান ও ভাবকে বইয়ের মধ্যে সঞ্চিত করিবার যে একটা প্রচুর সুবিধা আছে, সে কথা কেহই অস্বীকার করিতে পারে না। কিন্তু সেই সুবিধার দ্বারা মনের স্বাভাবিক শক্তিকে একেবারে আচ্ছন্ন করিয়া ফেলিলে বুদ্ধিকে বাবু করিয়া তোলা হয়।

— রবীন্দ্রনাথ ঠাকুর

ভারতের একটা mission আছে, একটা গৌরবময় ভবিষ্যৎ আছে; সেই ভবিষ্যৎ ভারতের উত্তরাধিকারী আমরাই। নূতন ভারতের মুক্তির ইতিহাস আমরাই রচনা করছি এবং করব। এই বিশ্বাস আছে বলেই আমরা সব দুঃখ কষ্ট সহ্য করতে পারি, অন্ধকারময় বর্তমানকে অগ্রাহ্য করতে পারি, বাস্তবের নিষ্ঠুর সত্যগুলি আদর্শের কঠিন আঘাতে ধূলিসাৎ করতে পারি।

— সুভাষচন্দ্র বসু

Any system of education which ignores Indian conditions, requirements, history and sociology is too unscientific to commend itself to any rational support.

— Subhas Chandra Bose

Price : Rs. 225.00

Published by : Netaji Subhas Open University, 1 Woodburn Park, Kolkata-700 020 and
Printed at : Sailee, 4A, Maniktala Main Road, Kolkata 700 054