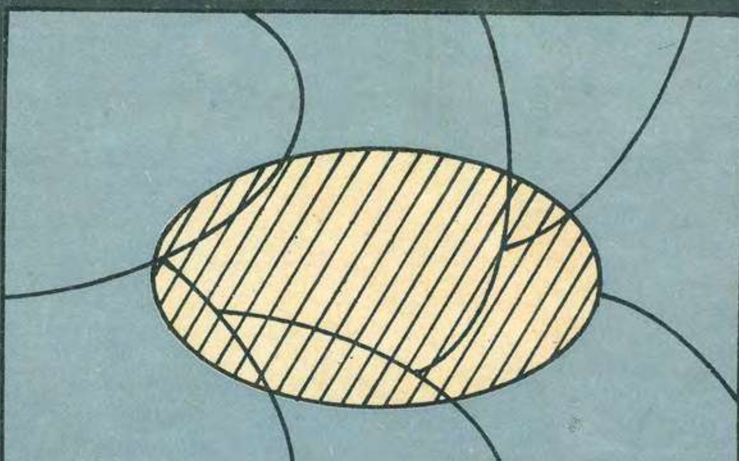
প্রাণবিজ্ঞানে পরিসংখ্যান

কেশব চন্দ্র ভূঞাঁ
স্বপন কুমার ধর



প্রাণবিজ্ঞানে পরিসংখ্যান

[দ্বিতীয় খণ্ড]

ড. কেশব চন্দ্র ভুঞা
পরিসংখ্যান বিভাগ
গ্যারিয়োনিস বিশ্ববিদ্যালয়
লিবিয়া

ড. স্বপন কুমার ধর
পরিসংখ্যান বিভাগ
গ্যারিয়োনিস বিশ্ববিদ্যালয়
লিবিয়া

বাংলা একাডেমী ঢাকা

প্রথম প্রকাশ
ফাল্গুন ১৪০৭
ফেব্রুয়ারি ২০০১
বাএ (২০০০-২০০১ পাঠ্যপুস্তক : ভৌ ও প্র ৬) ৪১১৩

মুদ্রণ সংখ্যা ১২৫০

পাণ্ডুলিপি প্রণয়ন ও মুদ্রণ তত্ত্বাবধান
ভৌতবিজ্ঞান ও প্রকৌশল উপবিভাগ
ভৌ ও প্র ২০৬

প্রকাশক
সুব্রত বিকাশ বড়ুয়া
পরিচালক
পাঠ্যপুস্তক বিভাগ
বাংলা একাডেমী, ঢাকা

মুদ্রক
মধুপুর প্রিন্টার্স
১৩/সি আজিমপুর রোড
ঢাকা।

প্রচ্ছদ
মামুন কায়সার

মূল্য : ১৪৫.০০

PRANBIJNANE PARISANGKHYAN (Statistics in Life Science Vol. II) by Dr. Keshab Chandra Bhuyan and Dr. Swapan Kumar Dhar. Published by Subrata Bikash Barua, Director, Textbook Division, Bangla Academy, Dhaka, Bangladesh. First edition, February 2001.
Price : Taka 145.00

ISBN 984-07-4112-8

ভূমিকা

গ্রন্থটির প্রথম খণ্ডে সাধারণ পরিসংখ্যান পদ্ধতি এবং সিদ্ধান্ত গ্রহণ পদ্ধতি জীববিজ্ঞান সম্বন্ধীয় উপাত্ত বিশ্লেষণের মাধ্যমে উপস্থাপন করা হয়েছে। সাধারণ পদ্ধতিগুলি প্রয়োগের পর উন্নত গবেষণার জন্য উন্নত পরিসংখ্যান পদ্ধতি প্রয়োগ যুক্তিসঙ্গত। উন্নত পরিসংখ্যান পদ্ধতির সাথে পরিচিতি লাভের সুবিধার্থে এই খণ্ডে নির্ভরণ ও সংশ্লেষণ, পরীক্ষণের নকশা ও ভেদাঙ্ক বিশ্লেষণ, নমুনায়ন পদ্ধতি, উৎপাদক বিশ্লেষণ এবং নির্ণায়ক বিশ্লেষণ এই পাঁচটি অধ্যায় আলোচনা করা হলো। এ সমস্ত পদ্ধতির প্রয়োগ দেখানোর জন্য প্রাণবিজ্ঞান সম্পর্কীয় উপাত্ত বিশ্লেষণ করা হয়েছে।

অনেক উন্নত এবং বড় আকারের নমুনা উপাত্ত বিশ্লেষণের জন্য কম্পিউটার প্রোগ্রাম ব্যবহার করা হয়ে থাকে। এখানে বিশেষ কোনো প্রোগ্রাম সম্পর্কে আলোচনা করা হয়নি। তবে বহুল প্রচলিত SPSS, STATISTICA, বা SAS STATGRAPHICS প্রোগ্রাম ব্যবহার করে বিভিন্ন অধ্যায়ে উপস্থাপিত উদাহরণগুলির উপাত্ত বিশ্লেষণ করা হয়েছে। পরিসংখ্যান সম্বন্ধীয় গবেষণায় নিয়োজিত ছাত্রছাত্রী এবং গবেষক ঐ সকল উদাহরণ পর্যালোচনা করে তাঁদের গবেষণা কাজে সহায়তা পেতে পারেন।

প্রাণবিজ্ঞানে পরিসংখ্যান (২য় খণ্ড) গ্রন্থটি রচনার জন্য যেসব বিদেশী লেখকের গ্রন্থের সহায়তা নেয়া হয়েছে আমরা সেসব লেখকের নিকট বিশেষভাবে ঋণী। যাদের উদ্দেশ্যে এ গ্রন্থটি লেখা হয়েছে তারা উপকৃত হলে আমাদের এ প্রচেষ্টা সার্থক হবে বলে মনে করি।

কেশব চন্দ্র ভূঞা

স্বপন কুমার ধর

সূচিপত্র

ষষ্ঠ অধ্যায়

নির্ভরণ ও সংশ্লেষণ

(Regression and Correlation)

৬.১ উপক্রমণিকা

জীববিজ্ঞান সম্পর্কীয় উপাত্ত আলোচনা করতে গিয়ে লক্ষ্য করা গেছে যে একই নমুনা বিন্দু হতে অনেক চলকের মান পরিমাপ করা হয়, যেমন—সারণি ১.১-এ (প্রথম খণ্ড) একই Slug- এর Body length, Body weight, Keel length, Shell length, Mantle length ইত্যাদি চলকের পরিমাপ করা হয়েছে। এই চলকগুলির মান যেহেতু একই নমুনা বিন্দু হতে সংগ্রহ করা, সে কারণে এগুলির মধ্যে একটি সম্পর্ক থাকা স্বাভাবিক। অনেকগুলি নমুনা বিন্দু হতে একই ধরনের চলকের মান পরিমাপ করে চলকগুলির সম্পর্ক পর্যালোচনা করার জন্য যে পরিসংখ্যান পদ্ধতি ব্যবহৃত হয় তাই নির্ভরণ (regression) এবং ঐ সম্পর্কের মাত্রা জানার জন্য যে পরিসংখ্যান পদ্ধতি ব্যবহৃত হয় তাই সংশ্লেষণ (correlation)।

চলকের সম্পর্ক পর্যালোচনা করতে গিয়ে লক্ষ্য করা যায় যে একটি চলকের মানের পরিবর্তনের সাথে সাথে অন্য চলকের মানেরও পরিবর্তন হয়ে থাকে। যেমন, এটি চিরন্তন সত্য যে স্বাভাবিক অবস্থায় মানুষের উচ্চতা বৃদ্ধির সাথে সাথে ওজন বৃদ্ধি পায়। এক্ষেত্রে অন্যান্য কারণের কথা উল্লেখ না করেও বা অন্যান্য কারণ স্বাভাবিক অবস্থায় বিবেচনা করেও বলা যায় যে ওজন বৃদ্ধির একটি কারণ হলো উচ্চতা বৃদ্ধি। আবার জমিতে সার প্রয়োগ করা হলে ফসলের উৎপাদন বৃদ্ধি পায়। এক্ষেত্রেও অন্যান্য কারণের কথা উল্লেখ না করেও বলা যায় যে সার প্রয়োগের কারণে ফসলের উৎপাদনে একটি প্রভাব পড়ে। চলকসমূহের ক্ষেত্রে এই কারণ (cause) এবং প্রভাব (effect) সম্পর্ক পর্যালোচনা করার যে পদ্ধতি তাই নির্ভরণ বিশ্লেষণ। এখানে যে চলকের প্রভাবে অন্য চলকের মানের পরিবর্তন হয় তাকে অনপেক্ষ চলক (independent variable) এবং যে চলকের মানের পরিবর্তন হয় তাকে নির্ভরশীল চলক (dependent variable)

বলা হয়। সুতরাং নির্ভরণ বিশ্লেষণ হলো এমন একটি পদ্ধতি যার মাধ্যমে নির্ভরশীল চলক ও অনপেক্ষ চলকের মধ্যে সম্পর্কের প্রকৃতি জানা যায় বা ঐ চলকগুলির মধ্যে সম্পর্কের সম্ভাব্য প্রকৃতি কি হবে সে সম্বন্ধে ধারণা লাভ করা যায়। ইংরেজ বিজ্ঞানী Sir Francis Galton (1822—1911) প্রথম নির্ভরণ-এর ধারণা ব্যক্ত করেন। তিনি মানুষের উচ্চতা পর্যালোচনা করতে গিয়ে ধারণা ব্যক্ত করেন যে, পূর্ণবয়স্ক ছেলে-মেয়েদের উচ্চতা, মা-বাবার উচ্চতা যাই হোক না কেন, গণসমষ্টির লোকদের গড় উচ্চতার মতই হয়। অর্থাৎ মা-বাবা লম্বাই হোক বা বেঁটেই হোক পূর্ণবয়স্ক ছেলে-মেয়ের উচ্চতা গণসমষ্টির লোকদের গড় উচ্চতায় ফিরে যায় (revert)। এই গড় উচ্চতায় ফিরে যাওয়ার ধারণা হতেই পরবর্তীতে তিনি নির্ভরণের ধারণা ব্যক্ত করেন।

সংশ্লেষণের ধারণাও এই ইংরেজ বিজ্ঞানী ১৮৮৮ সনে ব্যক্ত করেন। সংশ্লেষণ বিশ্লেষণ হলো অনপেক্ষ ও নির্ভরশীল চলকের মধ্যে সম্পর্কের মাত্রা নির্ধারণ করার একটি পদ্ধতি। যেহেতু সংশ্লেষণ চলকগুলির সম্পর্কের মাত্রা নির্ধারণ করে, সে কারণে এই বিশ্লেষণের জন্য চলকের শ্রেণীবিভাগ করার প্রয়োজন হয় না। শুধু চলকগুলি সম্পর্কিত এটি জানা থাকলেই চলে। অপরপক্ষে নির্ভরণ বিশ্লেষণের ক্ষেত্রে নির্ভরশীল চলক হলো অনপেক্ষ চলকের একটি ফাংশন। কাজেই এই বিশ্লেষণের মাধ্যমে অনপেক্ষ চলকের বিশেষ বিশেষ মানের জন্য নির্ভরশীল চলকের মান নির্ণয় করাও হয়ে থাকে।

উপরিউক্ত নির্ভরণ দুই প্রকার। যেমন: (১) সরল নির্ভরণ (simple regression) (২) বহুল নির্ভরণ (multiple regression)।

সরল নির্ভরণ: একটি নির্ভরশীল চলক ও একটি অনপেক্ষ চলকের মধ্যে যে সম্পর্ক তাই সরল নির্ভরণ। এই সম্পর্ক যদি সরলরৈখিক (linear) হয়, তবে নির্ভরণকে সরলরৈখিক নির্ভরণ (simple linear regression) বলা হয়।

বহুল নির্ভরণ: একটি নির্ভরশীল চলক একাধিক অনপেক্ষ চলকের ফাংশন হলে ঐ সম্পর্ক হলো বহুল নির্ভরণ। বহুল নির্ভরণের ক্ষেত্রেও চলকগুলির সম্পর্ক রৈখিক বা অরৈখিক হতে পারে।

বর্তমান অধ্যায়ে সরল নির্ভরণ বিশ্লেষণ এবং বহুল নির্ভরণ বিশ্লেষণ আলোচনা করা হবে। একই সাথে সংশ্লেষণ সম্পর্কেও আলোচনা করা হবে।

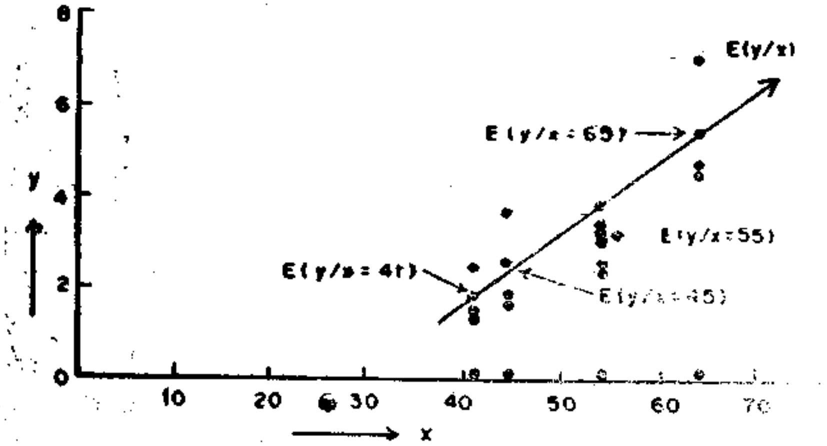
৬.২ নির্ভরণ প্রতিকৃতি (Regression Model)

আগেই উল্লেখ করা হয়েছে যে নির্ভরণ হলো গড় মানে ফিরে যাওয়া। এ ব্যাপারটি একটি উদাহরণের সাহায্যে ব্যাখ্যা করা যাক।

সারণি ১.১-এ (প্রথম খণ্ড) লক্ষ্য করা গেছে যে একই পরিমাপের Body length-এর জন্য বিভিন্ন পরিমাপের Body weight আছে যেমন,

Body length (in mm) x	Body weight (in gms) y	গড় Body weight E(y/x)
41	1.50, 2.30, 1.40	1.73
45	1.90, 1.88, 3.80	2.53
55	2.40, 3.40, 3.90, 3.00, 2.50, 3.90	3.18
65	4.40, 4.50, 7.20	5.37

দেখা যাচ্ছে যে, Body length (x)-এর যে কোনো মানের জন্য Body weight (y)-এর একাধিক মান আছে। কিন্তু y-এর একাধিক মান থাকা সত্ত্বেও x-এর যে কোনো মানের জন্য y-এর একটি গড় মান [E(y/x)] পাওয়া যায়। এই গড় মান নিরূপণ করার পদ্ধতি হলো নির্ভরণ বিশ্লেষণ। লক্ষ্য করা যাচ্ছে x-এর মান বৃদ্ধির সাথে সাথে E(y/x)-এর মান বৃদ্ধি পাচ্ছে। বিষয়টি একটি চিত্রের সাহায্যে লক্ষ্য করা যাক। চিত্র



চিত্র ৬.১ : Body length (x) ও Body weight (y)-এর বিক্ষেপ চিত্রের মধ্যে প্রদর্শিত E(y/x)।

হতেও লক্ষ্য করা যাচ্ছে যে x-এর মান বৃদ্ধির সাথে সাথে y-এর গড় মান বৃদ্ধি পাচ্ছে এবং এই গড় মান [E(y/x)] গুলি মোটামুটি একটি সরলরেখার উপর পড়ছে। এই সরলরেখাকে গাণিতিকভাবে প্রকাশ করা যায় নিম্নরূপভাবে :

$$E(y/x) = \alpha + \beta x$$

কিন্তু লক্ষ্য করা যাচ্ছে যে y-এর সকল মান $E(y/x)$-এর সমান নয়। y-এর মানসমূহ ও $E(y/x)$-এর বিভেদ (deviation) হলো

$$y - E(y/x) = e$$

অথবা $$y - \alpha - \beta x = e$$

অথবা $$y = \alpha + \beta x + e \qquad (৬.২.১)$$

এই শেষোক্ত গাণিতিক ফাংশন হলো y ও x এর সম্পর্ক পর্যালোচনার জন্য গাণিতিক প্রতিকৃতি। এই প্রতিকৃতিকে বলা হয় সরলরৈখিক নির্ভরণ প্রতিকৃতি (simple linear regression model)। এখানে α এবং β হলো গণসমষ্টি নির্ভরণ পরামান। β হলো x-এর এক একক পরিবর্তনের কারণে y-এর পরিবর্তনের $[dy/dx]$ হার, তাই β-কে বলা হয় নির্ভরণ সহগ (regression coefficient)। আবার জ্যামিতির ভাষায় β-কে বলা হয় ঢাল (slope) এবং α হলো ইন্টারসেপ্ট (intercept)।

প্রতিকৃতিতে e হলো y-এর মান এবং $E(y/x)$ এর পার্থক্য নির্দেশকারী একটি মান। এই e-কে বলা হয় দৈব বিচ্যুতি। $E(y/x)$-এর প্রতিটি মান একই সরলরেখায় অবস্থিত হলে $e = y - E(y/x) = 0$ হবে। বাস্তবে তা হয় না। কারণ x-এর একই মানের জন্য y-এর একটি মান পাওয়া যায় না। একই x-এর জন্য y-এর বিভিন্ন মান পাওয়ার কারণ হলো বিভিন্ন রকম এবং এই বিভিন্নতার পরিমাপও সবসময় সম্ভব হয় না। অর্থাৎ y-এর মানের বিভিন্নতার উৎস অন্য কিছুও হতে পারে যা পরিমাপ করা যায়নি। কিন্তু ঐ উৎস বা উৎসসমূহের প্রভাব y-এর উপর আছে বলেই ঐ y ও x-এর সম্পর্ক পর্যালোচনা করার জন্য বিচ্যুতি e প্রতিকৃতিতে সংযোগ করতে হয়।

এতক্ষণ দুটি চলক x ও y-এর সম্পর্ক পর্যালোচনা করার জন্য প্রতিকৃতির আলোচনা করা হয়েছে। এখানে y হলো নির্ভরশীল চলক। বাস্তবে y-এর মান আরো অনেক চলক দ্বারা প্রভাবিত হতে পারে যে চলকসমূহের মান পরিমাপ করা যায়। যেমন, কোনো রোগীর অক্সিজেন গ্রহণের পরিমাণ (y) নির্ভর করে তার Systolic blood pressure (x_1, mm Hg), total cholesterol (x_2, mg/DL), HDL cholesterol (x_3, mg/DL) এবং Triglycerides (x_4, mg/DL)-এর উপর। আবার কোনো রোগীর অসুখের জটিলতা (y) নির্ভর করে তার শিক্ষা (x_1), প্রোটিন খাদ্য গ্রহণের পরিমাণ (x_2), বাসস্থানের পরিবেশ (x_3), পরিবারের প্রতি সদস্যের জন্য প্রাপ্ত স্থান (x_4), ধূমপানের অভ্যাস (x_5), চিকিৎসার পর্যায় (x_6) এবং আরো অনেক উৎসের উপর। এক্ষেত্রে সরলরৈখিক প্রতিকৃতি দ্বারা y ও x-সমূহের সম্পর্ক পর্যালোচনা করা যায় না। y এবং একাধিক x-এর সম্পর্ক পর্যালোচনা করার পরিসংখ্যানিক পদ্ধতি হলো বহুল নির্ভরণ এবং এই সম্পর্ক নির্দেশ করার জন্য প্রয়োজন বহুল নির্ভরণ প্রতিকৃতি (multiple regression model)।

ল নির্ভরণ প্রতিকৃতি : বহুল নির্ভরণের ক্ষেত্রে সাধারণত y ও x-চলকসমূহের রৈখিক সম্পর্ক বিদ্যমানই বিবেচনা করা হয়। ধরা যাক একটি নির্ভরশীল [Dependent variable] $x_1, x_2, \ldots, x_k$ নামক k সংখ্যক অনপেক্ষ চলক ndent variable or Explanatory variable বা Predictor variable] খিকভাবে প্রভাবিত এবং চলকসমূহের রৈখিক সম্পর্ক হলো

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + e \qquad (৬.২.২)$$

প্রতিকৃতি (৬.২.২)-কে বলা হয় বহুল নির্ভরণ প্রতিকৃতি। এই প্রতিকৃতির β_0, β_k হলো নির্ভরণ পরামান এবং বিশেষভাবে β_i $(i=1, 2, \ldots k)$ হলো পর x_i-এর প্রভাব নির্ণয়কারী সহগ। এগুলিকে নির্ভরণ সহগ (regression ent) বলা হয়।

কৃতি (৬.২.২)-কে মেট্রিক্স চিহ্নেও লেখা যায়।

$$Y = XB + U \qquad (৬.২.৩)$$

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{n \times 1}, \quad X = \begin{bmatrix} 1 & x_{11} & x_{21} & \ldots & x_{k1} \\ 1 & x_{12} & x_{22} & & x_{k2} \\ 1 & \vdots & & & \vdots \\ \vdots & & & & \\ 1 & x_{1n} & x_{2n} & & x_{kn} \end{bmatrix}_{n \times (k+1)}$$

$$B = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}_{(k+1) \times 1} \text{ এবং } \quad U = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}_{n \times 1}$$

এখানে বিবেচনা করা হয়েছে যে, প্রতিটি চলকের মান n সংখ্যক নমুনা একক (sample unit sample point) হতে সংগৃহীত হয়েছে।

প্রতিকৃতি (৬.২.১)-কেও প্রতিকৃতি (৬.২.৩)-এর ন্যায় লেখা যায়, সেক্ষেত্রে $k = 1$ এবং

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}_{n \times 2} \quad \text{এবং} \quad B = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}_{2 \times 1}$$

৬.৩ নির্ভরণ বিশ্লেষণের জন্য অনুমান (Assumptions for Regression Analysis)

নির্ভরণ বিশ্লেষণ সরলই হোক আর বহুলই হোক এর ভিত্তি হলো কতকগুলি অনুমানের উপর। বিশেষ করে নির্ভরণ সহগ সম্পর্কে কোনো পরিসংখ্যানিক সিদ্ধান্ত নিতে হলে বা কোনো সহগের জন্য সংশয়মাত্রা (confidence interval) নিরূপণ করতে হলে ঐ অনুমানগুলি একান্তই প্রয়োজন। কারণ গণসমষ্টি নির্ভরণ সহগের মান আমাদের পক্ষে জানা সম্ভব নয় বা সম্ভব হলেও সহজসাধ্য নয়। সেক্ষেত্রে নমুনা তথ্যমানের ভিত্তিতে নির্ভরণ সহগ নিরূপণ করতে হয়। নিরূপিত মানসমূহ নমুনায়ন বিন্যাস অনুসরণ করে এবং জানা নমুনায়ন বিন্যাসের ক্ষেত্রে নমুনা উপাত্তের উৎস হলো পরিমিত বিন্যাস (normal distribution)। কাজেই বিশ্লেষণের নির্ভরশীল চলক পরিমিত বিন্যাস অনুসরণ করে এমন অনুমান অপরিহার্য। তাছাড়া নিরূপণ পদ্ধতি ব্যবহারের জন্যও কিছু অনুমান অপরিহার্য। নিচে অনুমানসমূহ উপস্থাপন করা হলো। এই অনুমানসমূহ প্রতিকৃতি (৬.২.৩)-এর জন্য প্রযোজ্য।

(i) সরল নির্ভরণ বিশ্লেষণের ক্ষেত্রে বা বহুল নির্ভরণ বিশ্লেষণের ক্ষেত্রে y-এর বিন্যাস হলো পরিমিত বিন্যাস। তাছাড়া অনপেক্ষ চলক x-এর যে কোনো একটি মানের জন্য বা x-সমূহের যে কোনো এক সেট মানের জন্য y চলকের উপ-গণসমষ্টি আছে এবং প্রতিটি উপ-গণসমষ্টির (sub-population) ভেদাঙ্ক সমান, σ^2।

(ii) y চলকের মানসমূহ অনপেক্ষ (independent) অর্থাৎ x চলকের যে কোনো একটি মানের জন্য নমুনায়িত y চলকের মান x চলকের অন্য একটি মানের জন্য নমুনায়িত y চলকের মান-এর উপর নির্ভরশীল নয়।

(iii) সরল নির্ভরণ বিশ্লেষণের ক্ষেত্রে x চলক অদৈব চলক (non-random variable) এবং বহুল নির্ভরণের ক্ষেত্রে x মেট্রিক্স-এর স্তম্ভসমূহ অদৈব চলকের মান নিয়ে গঠিত।

v) যেহেতু X মেট্রিক্স-এর অর্ডার হলো $n \times (k+1)$, rank $(X) = k+1$ বে বলা যায় x মেট্রিক্স-এর স্তম্ভসমূহ অনপেক্ষ। আরো সহজ করে বলা যায় ক চলকসমূহ অসংশ্লেষিত (uncorrelated) বা অনপেক্ষ (independent)

) $E[U] = 0$ এবং $E[UU'] = \sigma^2 I_n$

i) অনপেক্ষ চলকসমূহের মান পরিমাপে কোনো বিচ্যুতি থাকে না (no mea-ent error)।

মুনা নির্ভরণ সমীকরণ (Sample Regression Equation)

উল্লেখ করা হয়েছে যে, গণসমষ্টি নির্ভরণ সহগের মান জানা সম্ভব নয় বা সম্ভব সহজসাধ্য নয়, একারণেই নমুনাভিত্তিক নির্ভরণ সমীকরণ (regression on) নিরূপণ করতে হয়। এটি করার জন্য প্রাথমিক কাজ হলো (সরল ণের ক্ষেত্রে প্রাপ্ত উপাত্তের ভিত্তিতে বিক্ষেপ চিত্র (scatter diagram) আঁকা। কপ চিত্র হতেই বুঝা যাবে x ও y চলকের মধ্যে কি ধরনের সম্পর্ক বিদ্যমান। বিক্ষেপ চিত্র (৬.১) হতে বুঝা যাচ্ছে যে Body weight (y) ও Body l (x) সরলরৈখিকভাবে সম্পর্কিত। অনেক সময় বিক্ষেপ চিত্রের সহায়তা না বিশেষ করে বহুল নির্ভরণের ক্ষেত্রে, রৈখিক সম্পর্ক বিদ্যমান অনুমান করে ন বিশ্লেষণ করা হয় এবং বিশ্লেষণের পরে যে অনুমানের ভিত্তিতে বিশ্লেষণ হয়েছে সে অনুমানসমূহ বহাল আছে কিনা তা পর্যবেক্ষণ করা হয়। তাছাড়া যে ণ সমীকরণ মিল (fit) করা হয় তা x ও y-এর সম্পর্ক পর্যালোচনা করার জন্য ক্ষমতা রাখে তা নিরীক্ষা করা হয়। এ সম্পর্কে (৬.৬) ও (৬.৭) আলোচনা বে।

যাক x ও y-এর সম্পর্ক সরলরৈখিক এবং নমুনা হতে প্রাপ্ত α ও β এর নিরূপক ator) হলো যথাক্রমে a ও b। তাহলে নমুনাভিত্তিক নির্ভরণ রেখা হলো

$$y = a + bx \qquad (৬.৪.১)$$

নির্ভরণের ক্ষেত্রে নমুনাভিত্তিক নির্ভরণ সমীকরণ হলো

$$y = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_k x_k \qquad (৬.৪.২)$$

$b_0, b_1, b_2, \cdots, b_k$ হলো যথাক্রমে $\beta_0, \beta_1, \cdots, \beta_k$ এর নিরূপক।

নির্ভরণ পরামান নিরূপণ পদ্ধতি (Method of Estimation of Regress-arameters)

ক্ত অনুচ্ছেদে রৈখিক নির্ভরণের ক্ষেত্রে α ও β-এর নিরূপক বিবেচনা করা হয়েছে ম a ও b। নমুনা হতে এই a ও b-এর মান পাওয়ার সূত্র প্রয়োজন। এই র্ণয়ের পদ্ধতিই হলো পরামান নিরূপণ পদ্ধতি। সাধারণত রৈখিক নির্ভরণের পরামান নিরূপণ করা হয় ন্যূনতম বর্গপদ্ধতি (method of least squares)

প্রয়োগ করে। এই পদ্ধতি প্রয়োগের জন্য প্রয়োজনীয় অনুমান ৬.৩ অনুচ্ছেদে আলোচনা করা হয়েছে।

ন্যূনতম বর্গপদ্ধতি (Method of least squares) : নমুনা নির্ভরণ রেখার ভিত্তিতে বিচ্যুতির নিরূপক হলো

$$u = y - a - bx$$

এবং বিচ্যুতির বর্গসমষ্টি হলো

$$\varphi = \Sigma u^2 = \Sigma (y - a - bx)^2$$

ন্যূনতম বর্গপদ্ধতির মূল বক্তব্য হলো a ও b-এর মান এমনভাবে নির্ণয় করতে হবে যেন φ-এর মান ন্যূনতম হয়। φ-এর মান ন্যূনতম হবে যদি a ও b-এর মান যথাক্রমে

$$\frac{\partial \varphi}{\partial a} = 0 \quad \text{এবং} \quad \frac{\partial \varphi}{\partial b} = 0$$

তাহলে উক্ত সমীকরণ হতে পাওয়া যায়

$$\Sigma y = na + b\Sigma x \qquad (৬.৫.১)$$

$$\Sigma xy = a\Sigma x + b\Sigma x^2 \qquad (৬.৫.২)$$

উক্ত সমীকরণদ্বয়কে বলা হয় পরিমিত সমীকরণ (normal equations)। এই সমীকরণদ্বয় হতে পাওয়া যায়

$$b = \frac{\Sigma xy - \frac{\Sigma x \Sigma y}{n}}{\Sigma x^2 - \frac{(\Sigma x)^2}{n}} = \frac{SP(x, y)}{SS(x)}$$

এবং $$a = \bar{y} - b\bar{x}$$

বহুল নির্ভরণের ক্ষেত্রে বিচ্যুতির নিরূপককে লেখা যায়

$$\hat{U} = Y - \hat{B}X$$

এবং $$\varphi = \hat{U}'\hat{U} = (Y - \hat{B}X)'(Y - \hat{B}X)$$

$$= Y'Y - 2\hat{B}'X'Y + \hat{B}'X'X\hat{B}$$

ন্যূনতম বর্গপদ্ধতির মাধ্যমে পরিমিত সমীকরণ হলো

$$\frac{\partial \varphi}{\partial \hat{B}} = 0$$

অথবা $$(X'X)\hat{B} = X'y \qquad (৬.৫.৩)$$

অথবা $$\hat{B} = (X'X)^{-1}X'y$$

ন্যূনতম বর্গপদ্ধতির ধর্ম অনুসারে a, b বা $\hat{B}$ যথাক্রমে α, β বা B-এর নির্ঝুঁকি নিরূপক এবং এই নিরূপকগুলির ভেদাঙ্ক ন্যূনতম। সে কারণে ন্যূনতম বর্গপদ্ধতিতে প্রাপ্ত নিরূপককে বলা হয় সর্বোত্তম রৈখিক নির্ঝুঁকি নিরূপক (Best Linear Unbiased Estimator, BLUE)।

এখানে $$V(\hat{B}) = (X'X)^{-1}\sigma^2 \tag{৬.৫.৪}$$

সরলরৈখিক নির্ভরণের ক্ষেত্রে

$$V(a) = \frac{\Sigma x^2 \sigma^2}{n\Sigma x^2 - (\Sigma x)^2} = \sigma^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{SS(x)} \right]$$

এবং $$V(b) = \frac{\sigma^2}{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}$$

সকল ক্ষেত্রেই σ^2 হলো বিচ্যুতির ভেদাঙ্ক এবং বাস্তবে একে নমুনা হতে নিরূপণ করা হয়। এর নিরূপক হলো $\hat{\sigma}^2$ = গড় বর্গবিচ্যুতি (Mean square error, MS (error)। সরলরৈখিক নির্ভরণ রেখা অনেক সময় দ্বি-চলক ঘটনসংখ্যা বিন্যাস হতেও মিল (fit) করা হয়। ধরা যাক, একটি দ্বি-চলক ঘটনসংখ্যা বিন্যাসের আকৃতি নিম্নরূপ :

সারণি ৬.১ : দ্বি-চলক ঘটনসংখ্যা বিন্যাস।

X\Y	y_1	y_2	.....	y_j	.....	y_k	মোট f_i
x_1	f_{11}	f_{12}	. .	f_{1j}	. .	f_{1k}	f_1
x_2	$_1f_2$	f_{22}	. .	f_{2j}	. .	f_{2k}	f_2
$\vdots$	..		...	...	...	..	...
x_i	f_{i1}	f_{i2}	. .	f_{ij}	. .	f_{ik}	f_i
$\vdots$	...	...	...		...	...	...
x_m	f_{m1}	f_{m2}	. .	f_{mj}	. .	f_{mk}	f_m
মোট f_j	f_1	f_2	....	f_j	....	f_k	f_N

এখানে y_j এবং x_i $(i=1, 2, \ldots, m\,;\; j=1, 2, \ldots, k)$ হলো যথাক্রমে y চলক এবং x চলক-এর জন্য নির্ধারিত শ্রেণীর j-তম ও i-তম মধ্যমান, f_{ij} হলো x-এর i-তম শ্রেণীর প্রাসঙ্গিক y-এর j-তম শ্রেণীর ঘটনসংখ্যা, N হলো মোট ঘটনসংখ্যা।

তাহলে
$$b = \frac{\sum_{i}^{m}\sum_{j}^{k} f_{ij}x_i y_j - \dfrac{\sum^{m} f_i x_i \sum^{k} f_j y_j}{N}}{\sum^{m} f_i x_i^2 - \dfrac{\left(\sum^{m} f_i x_i\right)^2}{N}}$$

এবং
$$a = \bar{y} - b\,\bar{x}$$

এখানে
$$\bar{y} = \frac{1}{N}\sum_{j=1}^{k} f_j y_j\,, \qquad \bar{x} = \frac{1}{N}\sum_{i=1}^{m} f_i x_i$$

৬.৬ নির্ভরণ রেখার মূল্যায়ন (Evaluation of Regression Equation)

আগেই উল্লেখ করা হয়েছে যে, নমুনা হতে প্রাপ্ত উপাত্ত সরলরৈখিক নির্ভরণের জন্য যথাযথ কিনা তা বিক্ষেপ চিত্রের মাধ্যমে লক্ষ্য করতে হয়। বহুল নির্ভরণের ক্ষেত্রে অবশ্য এটি বাস্তবসম্মত নয়। সে কারণেই চলকসমূহের মধ্যে রৈখিক সম্পর্ক বিদ্যমান অনুমান করে নির্ভরণ রেখা মিল করতে হয়। তারপর ঐ মিল যথাযথ হয়েছে কিনা সেটি পর্যবেক্ষণ করতে হয়। এই পর্যবেক্ষণের অনেক ধাপ আছে। সেগুলি (১) নির্ভরণ রেখা তাৎপর্যপূর্ণ (significant) কিনা ? (২) নির্ভরণ সহগ তাৎপর্যপূর্ণ কিনা ? (৩) x চলকের ভিত্তিতে y-এর নির্ভরণ রেখা y-এর ভেদের কত অংশ ব্যাখ্যা করতে পারে ইত্যাদি।

অনেক সময় নির্ভরণ রেখা এবং নির্ভরণসহগ তাৎপর্যপূর্ণ হলেও মিল করা রেখা দ্বারা y চলকের ভেদের বৃহত্তর অংশ ব্যাখ্যা করা যায় না। সেক্ষেত্রে রৈখিক নির্ভরণ যুক্তিযুক্ত না হয়ে অন্য কোনো অরৈখিক রেখা যথাযথ হতে পারে। এটি লক্ষ্য করার জন্য এক ধরনের পরিমাপ আছে যা সিদ্ধান্ত সহগ (coefficient of determination, R^2) হিসেবে পরিচিত। এই সিদ্ধান্ত সহগের মান যত বেশি হবে মিল করা রেখা তত যথাযথ হবে। তাছাড়া দুই বা ততোধিক মিল করা রেখার ক্ষেত্রে যে রেখার জন্য সিদ্ধান্ত সহগ বড় হবে সে রেখাই ভাল মিল হয়েছে বলে সিদ্ধান্ত নিতে হয়। এখানে নির্ভরণের তাৎপর্য যাচাই এবং সিদ্ধান্ত সহগ নিয়ে আলোচনা করা হবে।

উদাহরণ ৬.১

একটি হাসপাতালের নথি থেকে লক্ষ্য করা গেছে যে, স্বাভাবিক সময়ের চেয়ে কম সময়ে জন্ম গ্রহণ করা বাচ্চাদের জন্মকালীন ওজন কম। নিচে ঐরূপ 10 বাচ্চার জন্মকালীন ওজন (y kg) এবং গর্ভকালীন বয়স (Gestation Age x সপ্তাহ) দেয়া হলো। জন্মকালীন ওজনের উপর গর্ভকালীন বয়সের প্রভাব কিরূপ নির্ণয় করা যাক এবং x-এর উপর ভিত্তি করে একটি সরলরৈখিক নির্ভরণ রেখা মিল করা যাক।

X :	28	32	33	30	35	29	30	32	27	28
Y :	1.1	1.5	1.4	1.2	1.6	1.2	1.3	1.3	0.9	1.0

ধরা যাক মিল করা রেখা হলো

$$\hat{y} = a + bx$$

এখানে

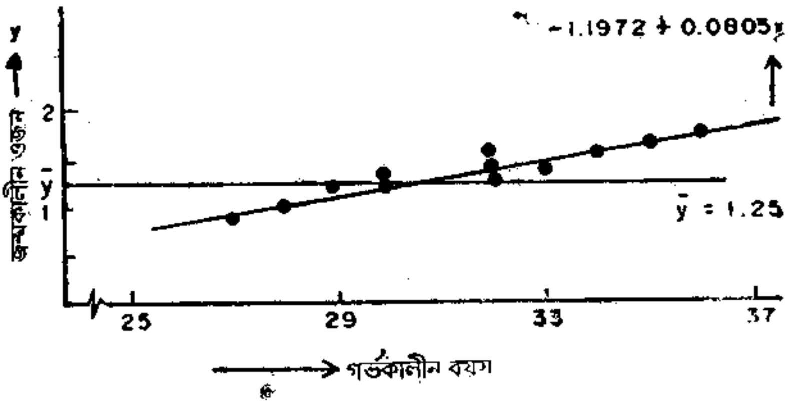
$$b = \frac{\Sigma xy - \frac{\Sigma x \Sigma y}{n}}{\Sigma x^2 - \frac{(\Sigma x)^2}{n}} = \frac{384.7 - \frac{304 \times 12.5}{10}}{9300 - \frac{(304)^2}{n}}$$

$$= \frac{4.7}{58.4} = 0.0805$$

$$a = \bar{y} - b\bar{x} = 1.25 - 0.0805 \times 30.4 = -1.1972$$

সুতরাং মিল করা নির্ভরণ রেখা হলো

$$\hat{y} = -1.1972 + 0.0805\,x \qquad (৬.৬.১)$$



চিত্র ৬.২ : বিক্ষেপ চিত্র, মিল করা নির্ভরণ রেখা ও ȳ-রেখা।

উপরিউক্ত উপাত্তকে একটি বিক্ষেপ চিত্রের মাধ্যমে প্রকাশ করে ঐ চিত্রে মিল করা রেখা এবং X-অক্ষের সমান্তরাল করে $\bar{y}$-রেখা আঁকা যাক। অঙ্কিত চিত্র হতে লক্ষ্য করা যাচ্ছে যে x-এর বিপরীতে প্রতিস্থাপিত y-এর মান নির্দেশকারী বিন্দুসমূহ $\hat{y}$ রেখা হতে কম বিক্ষিপ্ত। অপরদিকে ঐ বিন্দুগুলি $\bar{y}$ রেখা হতে বেশি বিক্ষিপ্ত। এখানে খাড়া রেখা (vertical line) দ্বারা y-এর i-তম মান ও $\bar{y}$-এর ব্যবধান, $(y_i - \bar{y})$-কে মোট ব্যবধান বা বিচ্যুতি (total deviation) বলা হয়। আবার $\hat{y}$ রেখা হতে $\bar{y}$ রেখা পর্যন্ত খাড়া রেখা দ্বারা $(\hat{y} - \bar{y})$ ব্যবধান নির্ণয় করা যায়। এই ব্যবধান নির্ণয়ের উদ্দেশ্য হলো নির্ভরণ রেখা মিল করার কারণে মোট ব্যবধান-এর কত অংশ ব্যবধান কমেছে। এই ব্যবধানকে বলা হয় ব্যাখ্যায়িত (explained) বা নির্ভরণ (regression) ব্যবধান। যেহেতু $(\hat{y} - \bar{y})$-হলো $(y_i - \bar{y})$-এর অংশ। সে কারণে লেখা যায়

$$(y_i - \bar{y}) = (\hat{y} - \bar{y}) + (y_i - \hat{y}) \qquad (৬.৬.২)$$

এখানে $(y_i - \hat{y})$ হলো y_i-এর এমন বিচ্যুতি যা নির্ভরণ রেখা দ্বারা ব্যাখ্যা করা যায় নি। একে বলা হয় অব্যাখ্যায়িত ব্যবধান (unexplained deviation)।

এখন (৬.৬.২) সমীকরণের উভয় পাশে বর্গ করে এবং i-এর সকল মানের জন্য যোগ করে পাওয়া যায়

$$\Sigma(y_i - \bar{y})^2 = \Sigma(\hat{y} - \bar{y})^2 + \Sigma(y_i - \hat{y})^2$$

এখানে $\Sigma(y_i - \bar{y})^2 =$ y-এর মোট বর্গসমষ্টি (Total sum of squares, SST)
$\Sigma(\hat{y} - \bar{y})^2 =$ রৈখিক নির্ভরণ রেখা দ্বারা ব্যাখ্যা করা বর্গসমষ্টি (Explained sum of squares) বা নির্ভরণ বর্গসমষ্টি (Regression sum of squares, SSR) এবং
$\Sigma(y_i - \hat{y})^2 =$ y-এর মোট বর্গসমষ্টির অব্যাখ্যা করা অংশ বা বিচ্যুতির বর্গসমষ্টি (Error sum of squares, SSE)

$$\therefore \quad SST = SSR + SSE \qquad (৬.৬.৩)$$

নির্ভরণ রেখা মিল করার উদ্দেশ্য হলো x ও y চলকের সম্পর্ক পর্যালোচনা করা। এখন x ও y-এর সম্পর্ক রৈখিক হলে ঐ সম্পর্ক পর্যালোচনা করার জন্য মিল করা রেখা যথেষ্ট অর্থবহ হওয়ার কথা। সেক্ষেত্রে y-এর মোট ভেদের একটি বৃহত্তর অংশ মিল করা রেখা দ্বারা ব্যাখ্যায়িত হওয়ার কথা। অর্থাৎ y-এর মোট বর্গসমষ্টির একটি বৃহত্তর অংশ হওয়া উচিত নির্ভরণ বর্গসমষ্টি। কাজেই নির্ভরণ রেখার মূল্যায়ন করার জন্য নির্ভরণ বর্গসমষ্টি y-এর মোট বর্গসমষ্টির কত অংশ তা দেখা যেতে পারে। এই অংশ হলো

$$r^2 = \frac{SSR}{SST} = \frac{\Sigma(\hat{y} - \bar{y})^2}{\Sigma(y_1 - \bar{y})^2}$$

এখানে r^2-কে বলা হয় সিদ্ধান্ত সহগ (coefficient of determination)।

এই r^2-এর মান সর্বোচ্চ 1 হতে পারে। সেক্ষেত্রে $SSR = SST$। এটি সম্ভব যখন y_1-এর প্রতিটি মান মিল করা নির্ভরণ রেখার উপর অবস্থিত হয়। বিষয়টি অন্য একটি উদাহরণের সাহায্যে লক্ষ্য করা যাক।

উদাহরণ ৬.২

নিচে x ও y চলকের জন্য কিছু কল্পিত মান দেয়া হলো। এই মানসমূহের ভিত্তিতে SSR ও SST নির্ণয় করা যাক।

x :	5	10	15	20	25
y :	2	4	6	8	10

ধরা যাক মিল করা রেখা হলো

$$y = a + bx$$

এখানে

$$b = \frac{\Sigma xy - \frac{\Sigma x \Sigma y}{n}}{\Sigma x^2 - \frac{(\Sigma x)^2}{n}} = \frac{550 - \frac{75 \times 30}{5}}{1375 - \frac{(75)^2}{5}} = \frac{100}{250} = 0.40$$

$$a = \bar{y} - b\bar{x} = 6 - 0.4 \times 15 = 0$$

$$\therefore \quad \hat{y} = 0.4x$$

এখন এই মিল করা রেখার ভিত্তিতে $\hat{y}_1$ সমূহ হলো

$$\hat{y}_1 = \quad 2 \quad 4 \quad 6 \quad 8 \quad 10$$

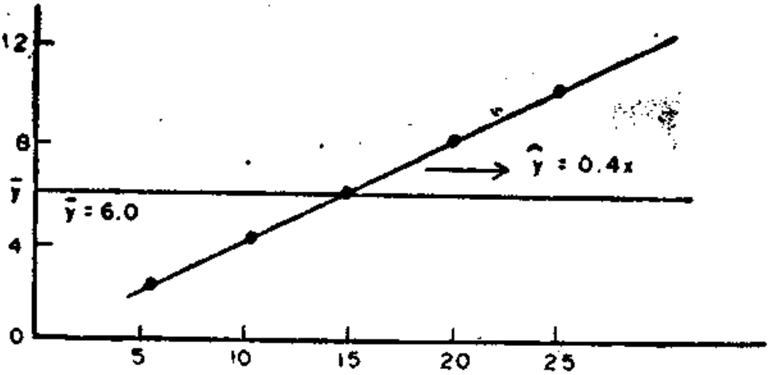
লক্ষ্য করা যাচ্ছে যে $\hat{y}_1$-এর প্রতিটি মান y_1-এর প্রতিটি প্রাসঙ্গিক মানের সমান। সুতরাং এখানে

$$SST = \Sigma(y_1 - \bar{y})^2 = \Sigma y_1^2 - \frac{(\Sigma y_1)^2}{n} = 220 - \frac{(30)^2}{5} = 40$$

এবং $SST = SSR$

কাজেই $\quad r^2 = \frac{SSR}{SST} = \frac{40}{40} = 1$

এখানে লক্ষণীয় বিষয় হলো x-এর মান একটি নির্দিষ্ট হারে বাড়ছে। অনুরূপভাবে y-এর মানও একটি নির্দিষ্ট হারে বাড়ছে। কোনো ক্ষেত্রেই চলকগুলির মান নির্দিষ্ট হারের পরিবর্তন থেকে বিচ্যুত নয়। সে কারণে x ও y-এর সরলরৈখিক সম্পর্ক যথাযথ এবং ঐ রেখা মিল করার করণে y-এর মোট ভেদ ব্যাখ্যা করা সম্ভব হয়েছে। বিষয়টি চিত্র অঙ্কন করেও লক্ষ্য করা যেতে পারে। দেখা যাচ্ছে যে, y-এর i-তম মান ও



চিত্র ৬.৩ : বিক্ষেপ চিত্র, মিল করা নির্ভরণ রেখা এবং $\bar{y}$-রেখা।

$\bar{y}$-এর ব্যবধান $[y_i - \bar{y}]$ এবং $(\hat{y} - \bar{y})$ একই। কারণ y_i-এর প্রতিটি মানই $\hat{y}$-রেখায় অবস্থিত। ফলে y_i-এর ভেদ নির্ভরণ রেখা দ্বারাই প্রকাশ করা যায়। এ কারণেই নির্ভরণ বর্গসমষ্টি [SSR] ও মোট বর্গসমষ্টি [SST]-এর অনুপাত মিল করা নির্ভরণ রেখার মূল্যায়ন করার জন্য ব্যবহৃত হয়।

এখন উদাহরণ ৬.১-এর ক্ষেত্রে এই r^2-এর মান নির্ণয় করা যাক। এখানে

$$SST = \Sigma y^2 - \frac{(\Sigma y)^2}{n} = 16.05 - \frac{(12.5)^2}{10} = 0.425$$

এবং

$$\begin{aligned} SSR &= \Sigma(\hat{y} - \bar{y})^2 \\ &= b^2 \Sigma(x - \bar{x})^2 \\ &= \frac{b\left[\Sigma xy - \frac{\Sigma x \Sigma y}{n}\right]\Sigma(x-\bar{x})^2}{\Sigma(x-\bar{x})^2} \\ &= b\left[\Sigma xy - \frac{\Sigma x \Sigma y}{n}\right] \\ &= b\,SP(xy) \\ &= 0.0805 \times 4.7 = 0.37835 \end{aligned}$$

কাজেই $$r^2 = \frac{SSR}{SST} = \frac{0.37835}{0.425} = 0.8902 \text{ বা } 89.02\%$$

লক্ষ্য করা যাচ্ছে যে, উপরিউক্ত উদাহরণের ক্ষেত্রে সরলরৈখিক নির্ভরণ রেখা মিল করে y-এর ভেদের 89.02% ব্যাখ্যা করা যায়। আবার উদাহরণ ৬.২-এর ক্ষেত্রে y-এর ভেদের 100% ব্যাখ্যা করা যায় এবং এটি সম্ভব হয়েছে যখন SST = SSR হয়েছে। কাজেই r^2-এর সর্বোচ্চ মান 1 হতে পারে এবং সর্বনিম্ন মান শূন্য হতে পারে [শূন্য তখনই হতে পারে, যখন নির্ভরণ রেখা মিল করে y-এর কোনো ভেদই ব্যাখ্যা করা যাবে না এবং এটি সম্ভব হয় যখন $\bar{y}$-রেখা এবং মিল করা নির্ভরণ রেখা একই অবস্থানে থাকে (coincide)]।

কাজেই লক্ষ্য করার বিষয় হলো r^2-এর মান বেশি হলে মিল করা নির্ভরণ রেখা যথাযথ হয়েছে বিবেচনা করা যেতে পারে। কিন্তু r^2-এর বেশি মান কত সেটির কোনো নির্দিষ্ট সীমা নেই। অনেক সময় $r^2 = 0.50$ হলেও মিল করা নির্ভরণ রেখা গ্রহণীয় সীমার মধ্যে পড়তে পারে এবং এটি নির্ধারণ করতে হবে যাচাই পদ্ধতির মাধ্যমে। যাচাই পদ্ধতি ৬.৭ অনুচ্ছেদে আলোচনা করা হবে, তার আগে বহুল নির্ভরণের ক্ষেত্রে এই r^2-এর মান কিভাবে নির্ণয় করা হয় তা দেখা যাক।

বহুল নির্ভরণের ক্ষেত্রে মিল করা রেখা হলো

$$Y = X\hat{B}$$

তাহলে, নিরূপণ করা বিচ্যুতি ভেক্টর হবে

$$\hat{U} = Y - X\hat{B}$$

সুতরাং বিচ্যুতির বর্গসমষ্টি, SSE হলো

$$SSE = \hat{U}'\hat{U} = (Y - X\hat{B})'(Y - X\hat{B})$$

সরল করে পাওয়া যায়

$$SSE = Y'Y - \hat{B}'X'Y$$
$$= SST_u - SSR_u$$

এখানে SST_u হলো অশোধিত (uncorrected)। y-এর মোট বর্গসমষ্টি হলো

$$SST = Y'Y - \frac{1}{n}(\Sigma y)^2 = SST_u - C.T \left[C.T = \frac{1}{n}(\Sigma y)^2 \right]$$

তেমনি SSR_u হলো অশোধিত নির্ভরণ বর্গসমষ্টি। কাজেই

$$SSR = SSR_u - C.T = \hat{B}'X'Y - \frac{1}{n}(\Sigma y)^2$$

যদি B-ভেক্টরের সব মানই নির্ভরণ সহগ হয় [এটি সম্ভব যদি নির্ভরণ প্রতিকৃতিতে ইন্টারসেপ্ট না থাকে বা নির্ভরণ গড় থেকে ব্যবধান নেয়া চলকের মানের ভিত্তিতে পর্যালোচনা করা হয়], তাহলে

$$SSR = SSR_{\alpha} = \hat{B}'X'Y$$

কাজেই সিদ্ধান্ত সহগ (coefficient of determination) হবে

$$R^2 = \frac{SSR}{SST}$$

এই R^2-কে অনেক সময় লেখা হয় $R^2_{y \cdot 1, 2, \ldots, k}$ এবং একে বলা হয় y ও x_1, x_2, ... x_k চলকসমূহের বহুল সংশ্লেষাঙ্ক (Multiple correlation coefficient)-এর বর্গ ।

এই R^2 দুই বা ততোধিক মিল করা রেখার মধ্যে তুলনামূলক বিশ্লেষণ করার জন্যও ব্যবহৃত হয়। যেহেতু R^2-এর বর্ধিত মান মিল করা নির্ভরণ রেখার একটি গ্রহণ-যোগ্য বৈশিষ্ট্য সে কারণে যে রেখার জন্য R^2 বড় তাকে প্রাপ্ত উপাত্তের জন্য গ্রহণ-যোগ্য মিল করা রেখা হিসেবে বিবেচনা করা হয়। কিন্তু একই গণসমষ্টি হতে ছোট আকারের ও বড় আকারের দুটি নমুনা চয়ন করা হলে বড় আকারের নমুনার ক্ষেত্রে মোট বর্গসমষ্টি (SST) এবং নির্ভরণ বর্গসমষ্টি (SSR) উভয়ই বড় হবে [কারণ, y_i ও $\bar{y}$ -এর ব্যবধান ধনাত্মক বা ঋণাত্মক যাই হোক না কেন $(y_i - \bar{y})^2$ ধনাত্মক হবে এবং এই ধনাত্মক রাশি যত বেশি যোগ করা হবে তাদের মোট তত বেশি হবে $\Sigma(y_i - \bar{y})^2$ । এ কারণে নমুনা আকারের পার্থক্যের কারণে দুই বা ততোধিক নমুনা হতে নির্ণয় করা R^2-এর মান সরাসরি তুলনীয় নয়। তুলনা করার জন্য R^2-কে সমন্বয় (adjust) করতে হয় । নিচে সমন্বিত $R^2_{(adj)}$-এর সূত্র দেয়া হলো ।

$$R^2_{(adj)} = 1 - \frac{n-1}{n-k-1}(1 - R^2)$$

$$= 1 - \frac{MSE\ (n-1)}{SST}$$

এই সমন্বিত সিদ্ধান্ত সহগ বিভিন্ন প্রতিকৃতিতে বিভিন্ন সংখ্যক অনপেক্ষ চলক ব্যবহৃত হলে ঐ প্রতিকৃতিসমূহের তুলনামূলক মিল (fit) পর্যবেক্ষণ করার জন্য ব্যবহৃত হয় ।

উদাহরণ ৬.১ এর ক্ষেত্রে $R^2_{(adj)} = 1 - \frac{10-1}{10-1-1}(1 - 0.8902)$

$$= 0.8765$$

৬.৭ নির্ভরণ সম্পর্কীয় যাচাই (Test Regarding Regression)

নির্ভরণ রেখা মূল্যায়ন করার জন্য r^2 বা R^2-এর ব্যবহার পূর্ববর্তী অনুচ্ছেদে আলোচনা করা হয়েছে। তবে এই মূল্যায়ন যথেষ্ট নয়। মূল্যায়নকে অধিক যুক্তিসঙ্গত ও ফলপ্রসূ করতে হলে পরিসংখ্যানিক যাচাই প্রয়োজন। বর্তমান অনুচ্ছেদে এই যাচাই পদ্ধতি নিয়ে আলোচনা করা হবে এবং যাচাই তথ্যজমানের ভিত্তিতে মিল করা নির্ভরণ রেখা হতে x-এর বিভিন্ন মানের জন্য y-এর মান নিরূপণ করে বা y-এর মান সম্পর্কে পূর্বাভাস (prediction) করে ঐ পূর্বাভাসের জন্য নিশ্চয়তা-পরিক্ষেপ (confidence interval) নিরূপণ করা হবে। তার আগে ৬.৬ অনুচ্ছেদে মোট বর্গসমষ্টিকে যে বিভিন্ন উৎসের বর্গসমষ্টিতে বিভক্ত করা হয়েছে [সমীকরণ (৫.৬.৩)] তা একটি সারণিতে

সারণি ৬.২ ঃ নির্ভরণের ক্ষেত্রে ভেদাঙ্ক বিশ্লেষণ সারণি।

ভেদের উৎস	স্বাধীনতার মাত্রা	বর্গসমষ্টি	গড় বর্গসমষ্টি	নির্ণেয় N
	d.f.	SS	MSS $= SS/d.f.$	
নির্ভরণ	k	SSR	MSR	MSR/MSE
বিচ্যুতি	$n-k-1$	SSE	MSE	
মোট	$n-1$	SST		

উপস্থাপন করা যাক। সরল নির্ভরণের ক্ষেত্রে সারণিতে দেয়া k-এর মান 1 এবং বহুল নির্ভরণের ক্ষেত্রে k হলো অনপেক্ষ চলকের সংখ্যা।

এই ভেদাঙ্ক বিশ্লেষণ সারণির মূল উদ্দেশ্য হলো সার্বিকভাবে মিল করা রেখা তাৎপর্যপূর্ণ (significant) কিনা যাচাই করা। এক্ষেত্রে নাস্তি কল্পনা (null hypothesis) হলো

H_0 : নির্ভরণ তাৎপর্যপূর্ণ নয়।

এবং বিকল্প কল্পনা (alternative hypothesis)

H_A : নির্ভরণ তাৎপর্যপূর্ণ।

এই নাস্তি কল্পনা যাচাই করার জন্য যাচাই তথ্যজমান (test statistic) হলো

$$F = \frac{MSR}{MSE}$$

এই F-এর বিন্যাস হলো k-এবং $(n-k-1)$ স্বাধীনতার মাত্রাবিশিষ্ট (d.f) ভেদাঙ্ক অনুপাত (variance ratio, F) বিন্যাস।

সুতরাং নির্ণেয় $F \geqslant F_{\alpha};\ k, (n-k-1)$

হলে নাস্তি কল্পনা বাতিল বলে গণ্য হয়। এখানে $F_{\alpha; 1,(n-k-1)}$ হলো k এবং $(n-k-1)$ স্বাধীনতার মাত্রাবিশিষ্ট F-বিন্যাস-এর $\alpha\%$ মান।

অনেক সময়, বিশেষ করে ফলিত গবেষণায়, নির্ণেয় F-কে F_{α}-এর সাথে তুলনা না করে F-এর P-মান-এর ভিত্তিতে সিদ্ধান্ত নেয়া হয়। এই P-মান হলো

$$\int_{F}^{\infty} f(F)\, dF = P$$

এই $P \leqslant 0.05 [\alpha = 0.05]$ হলে নাস্তি কল্পনা বাতিল বলে গণ্য হয়। $P > 0.05$ হলে নাস্তি কল্পনা বাতিলের পক্ষে নমুনা হতে যথেষ্ট প্রমাণ পাওয়া যায় না বলে সিদ্ধান্ত নিতে হয়। নিচে উদাহরণ ৬.১-এর জন্য নির্ভরণের যাচাই উপস্থাপন করা হলো।

সারণি ৬.৩ : উদাহরণ ৬.১-এর ক্ষেত্রে ভেদাঙ্ক বিশ্লেষণ সারণি।

ভেদের উৎস	স্বাধীনতার মাত্রা d.f	বর্গসমষ্টি SS	গড় বর্গসমষ্টি $MS = \frac{SS}{d.f.}$	F $= \frac{MSR}{MSE}$	P-মান
নির্ভরণ	1	0.37835	0.37835	64.90	0.00
বিচ্যুতি	8	0.04665	0.00583		
মোট	9				

এই বিশ্লেষণ সারণি হতে লক্ষ্য করা যাচ্ছে যে $P[F \geqslant 64.90] = 0.00 < 0.05$ সে কারণে মিল করা নির্ভরণ রেখা তাৎপর্যপূর্ণ বলে বিবেচনা করা যায়। অর্থাৎ গর্ভকালীন বয়স বৃদ্ধির সাথে সাথে জন্মকালীন ওজন তাৎপর্যপূর্ণভাবে বৃদ্ধি পায়।

নির্ভরণের তাৎপর্যতা (significance) যাচাই-এর পরবর্তী পদক্ষেপ হলো প্রতিটি নির্ভরণ সহগ বা নিভরাঙ্ক তাৎপর্যপূর্ণ কিনা তা পর্যবেক্ষণ করা। সরল নির্ভরণের ক্ষেত্রে এই তাৎপর্য যাচাই করার প্রয়োজন হয় না। কারণ সেক্ষেত্রে একটিই নির্ভরাঙ্ক এবং F যাচাই সে সম্পর্কে সিদ্ধান্ত গ্রহণ করতে সাহায্য করে [কারণ, $F = t^2$]। বহুল নির্ভরণের ক্ষেত্রে প্রতিটি নির্ভরণ সহগের তাৎপর্য যাচাই করতে হয়। কারণ নির্ভরণ তাৎপর্যপূর্ণ হলেও নির্ভরশীল চলকের উপর সকল অনপেক্ষ চলকের প্রভাব তাৎপর্যপূর্ণ

নাও হতে পারে। তবে যে অনুমানের ভিত্তিতে নির্ভরণ রেখা মিল করা হয় সেগুলি কোনো উপাত্তের জন্য বহাল থাকলে F-যাচাই-এর মাধ্যমে নির্ভরণ তাৎপর্যপূর্ণ হলে কমপক্ষে একটি নির্ভরণ পরামান তাৎপর্যপূর্ণ হবে।

নির্ভরণ পরামান বা নির্ভরণ সহগ-এর তাৎপর্য যাচাই করার জন্য নাস্তি কল্পনা হলো

$$H_0: \ \beta_i = 0 \qquad [i = 0, 1, 2, \ldots, k)$$

এবং $$H_A: \ \beta_i \neq 0$$

এই নাস্তি কল্পনা যাচাই-এর জন্য যাচাই তথ্যজমান হলো

$$t = \frac{b_i - \beta_i}{s.e(b_i - \beta_i)}$$

নাস্তি কল্পনার অধীনে

$$t = \frac{b_i}{s.e(b_i)}$$

এই 't'-এর বিন্যাস হলো $(n - k - 1)$ স্বাধীনতার মাত্রাবিশিষ্ট Student's 't' বিন্যাস।

এখানে b_i হলো β_i-এর নিরূপক এবং $s.e(b_i) = \sqrt{V(b_i)}$। জানা আছে $\hat{B}$ হলো B-এর নিরূপক এবং

$$V(\hat{B}_i) = (X'X)^{-1}\sigma^2$$

এর নিরূপক হলো

$$v(\hat{B}_i) = (X'X)^{-1}\hat{\sigma}^2$$

এখানে $$\hat{\sigma}^2 = MSE = \frac{SSE}{n-k-1}$$

উক্ত ভেদাঙ্ক-সহ-ভেদাঙ্ক মেট্রিক্স-এর i-তম কৌণিক মান হলো

$$V(b_i)$$

সরলরৈখিক নির্ভরণের ক্ষেত্রে

$$v(b) = \frac{\hat{\sigma}^2}{SS(x)}$$

এখন উদাহরণ ৬.১-এর উপাত্তের ভিত্তিতে নির্ভরণ সহগ β-এর তাৎপর্য যাচাই করা যাক। এক্ষেত্রে নাস্তি কল্পনা এবং বিকল্প কল্পনা হলো, যথাক্রমে

$$H_0: \ \beta = 0$$

$$H_A: \ \beta \neq 0$$

এবং যাচাই তথ্যজমান হলো

$$t = \frac{b - \beta}{\sqrt{\frac{SSE}{n-k-1}\frac{1}{SS(x)}}} = \frac{0.0805 - 0}{\sqrt{\frac{0.00583}{58.4}}} = 8.06$$

এই 't'-এর স্বাধীনতার মাত্রা হলো 8 এবং $P(t \geqslant 8.06) = 0.001$, কাজেই $P < 0.05$ হওয়াতে নাস্তি কল্পনা বাতিল। অর্থাৎ গর্ভকালীন বয়স এক সপ্তাহ বৃদ্ধি পেলে গর্ভকালীন ওজনের পরিবর্তনের হার [β] অতিশয় তাৎপর্যপূর্ণ।

অনুরূপভাবে সরলরৈখিক নির্ভরণ প্রতিকৃতির ক্ষেত্রে গণসমষ্টি ইন্টারসেপ্ট [α]-এর তাৎপর্যও যাচাই করা যায়। সেক্ষেত্রে

$$H_0 : \alpha = 0$$
$$H_A : \alpha \neq 0$$

এবং যাচাই তথ্যজমান হলো

$$t = \frac{a - \alpha}{\sqrt{\hat{\sigma}^2 \frac{\Sigma x^2}{n\Sigma x^2 - (\Sigma x)^2}}}$$

উদাহরণ ৬.১-এর ক্ষেত্রে পাওয়া যায়

$$t = \frac{-1.1972 - 0}{\sqrt{\frac{0.00583 \times 9300}{10 \times 9300 - (304)^2}}} = \frac{-1.1972}{0.3047} = -3.93$$

এই 't'-এরও স্বাধীনতার মাত্রা 8। তবে 't'-কে সারণিকৃত 't' (Tabulated 't')-এর সাথে তুলনা করতে হলে $|t|$ বিবেচনা করতে হবে। এখানে 5% সংশয় মাত্রায় সিদ্ধান্ত নিতে হলে সারণিকৃত 't'-এর মান $t_{.025,8} = 2.306$। নির্ণেয় $t > t_{.025}$ হওয়াতে আলোচিত নির্ভরণের ইন্টারসেপ্টও তাৎপর্যপূর্ণ। আবার লক্ষ্য করা যাচ্ছে যে, $P(t \leqslant -3.93) < 0.005$। এটি হতেও সিদ্ধান্ত নেয়া যায় যে ইন্টারসেপ্ট তাৎপর্যপূর্ণ (significant)।

নিশ্চয়তা পরিক্ষেপ (Confidence interval): এই পরিক্ষেপ সব কয়টি নির্ভরণ পরামানের জন্য নিরূপণ করা যায়। সাধারণত নির্ভরণ সহগ তাৎপর্যপূর্ণ হলে ঐ সহগের জন্য নিশ্চয়তা পরিক্ষেপ নিরূপণ করা হয়। ধরা যাক β_i হলো i-তম $[i = 1, 2, \ldots, k]$ অনপেক্ষ চলকের জন্য নির্ভরণ সহগ এবং b_i হলো নমুনা হতে β_i-এর নিরূপক। এই নিরূপকের নিরূপিত ভেদাঙ্ক পাওয়া যাবে

$$V(\hat{B}) = (X'X)^{-1}\hat{\sigma}^2$$

এর i-তম কৌণিক মান হতে। এখানে $\hat{\sigma}^2 = MSE$। তাহলে β_i-এর $(\alpha)\%$ সংশয় মাত্রায় (confidence level)-এ $100(1-\alpha)\%$ নিশ্চয়তা পরিক্ষেপ হলো

$$b_1 \pm t_{\alpha/2}\, s.e(b_1)$$

উদাহরণ ৬.১-এর ক্ষেত্রে β-এর 95% নিশ্চয়তা-পরিক্ষেপ হলো

$$b \pm t_{.025,8}\sqrt{\frac{\hat{\sigma}^2}{SS(x)}}$$

অথবা $$0.0805 \pm 2.306\sqrt{\frac{0.00583}{58.4}}$$

অথবা $$0.0805 \pm 0.0230$$

অর্থাৎ $$0.0575 \leqslant \beta \leqslant 0.1035$$

অনুরূপভাবে α-এর 95% নিশ্চয়তা-পরিক্ষেপ হলো

$$a \pm t_{.025,8}\sqrt{\frac{\hat{\sigma}^2 \Sigma x^2}{n\Sigma x^2-(\Sigma x)^2}}$$

$$-1.1972 \pm 2.306\sqrt{\frac{0.00583 \times 9300}{10 \times 9300-(304)^2}}$$

অথবা $$-1.1972 \pm 2.306 \times 0.3047$$

অথবা $$-1.1972 \pm 0.7026$$

অর্থাৎ $$-1.8998 \leqslant \alpha \leqslant -0.4946$$

৬.৮ নির্ভরশীল চলকের মান নিরূপণ এবং পূর্বাভাস (Estimation and Prediction of Value of Dependent Variable)

নির্ভরণ রেখা মিল করার পর তা তাৎপর্যপূর্ণ হলে সরল নির্ভরণের ক্ষেত্রে অনপেক্ষ চলক x-এর যে কোনো একটি মানের (x_0) জন্য y-এর একটি মান নিরূপণ করা যায়। ঐ নিরূপিত মানকে x-এর একটি মানের (x_0) জন্য y-এর গড় নিরূপিত মান $\hat{\mu}_0$ বিবেচনা করা যাক, তাহলে

$$\hat{\mu}_0 = a + bx_0$$

এবং $$V(\hat{\mu}_0) = \sigma^2\left[\frac{1}{n} + \frac{(x_0-\bar{x})^2}{\Sigma(x-\bar{x})^2}\right]$$

এক্ষেত্রে σ^2-কে তার নিরূপক $\hat{\sigma}^2 = MSE$ দ্বারা প্রতিস্থাপন করে $V(\hat{\mu}_0)$-এর নিরূপক পাওয়া যায় এবং এই নিরূপক হতে $\hat{\mu}_0$-এর পরিমিত বিচ্যুতি (standard error)-এর নিরূপক হলো

$$s.e(\hat{\mu}_0) = \sqrt{v(\hat{\mu}_0)} = \sqrt{\hat{\sigma}^2\left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\Sigma(x - \bar{x})^2}\right]}$$

উদাহরণ ৬.১-এর ক্ষেত্রে ধরা যাক $x_0 = 31$, তাহলে y-এর গড় মান হবে

$\hat{\mu}_0 = 1.1972 + 0.0805 \times 31 = 1.2983$ kg

এবং
$$s.e(\hat{\mu}_0) = \sqrt{\hat{\sigma}^2\left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\Sigma(x - \bar{x})^2}\right]}$$

$$= \sqrt{0.00583\left[\frac{1}{10} + \frac{(31 - 30.4)^2}{58.4}\right]}$$

$$= 0.0249$$

আবার x-এর যে কোনো মানের জন্য y-এর মান সম্পর্কে পূর্বাভাস করা যায়। ধরা যাক পূর্বাভাস করা y-এর মান হলো y_0। তাহলে

$$y_0 = a + bx_0$$

এখানে x_0 হলো x-এর যে কোনো একটি মান। এই y_0-এর ভেদাঙ্ক হলো

$$V(y_0) = \hat{\sigma}^2\left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\Sigma(x - \bar{x})^2}\right)$$

এবং এই ভেদাঙ্কের নিরূপক হলো

$$V(y_0) = \hat{\sigma}^2\left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\Sigma(x - \bar{x})^2}\right), \quad \text{এখানে } \hat{\sigma}^2 = MSE$$

এখন উদাহরণ ৬.১-এর ক্ষেত্রে $x_0 = 37$ সপ্তাহ হলে জন্মকালীন বাচ্চার ওজন কত হতে পারে (y_0) তার একটি পূর্বাভাস করা যাক। দেখা যাচ্ছে যে,

$$y_0 = -1.1972 + 0.0805 \times 37 = 1.7813 \text{ kg}$$

এবং
$$s.e(y_0) = \sqrt{\hat{\sigma}^2\left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\Sigma(x - \bar{x})^2}\right]}$$

$$= \sqrt{0.00583\left[1 + \frac{1}{10} + \frac{(37 - 30.4)^2}{58.4}\right]}$$

$$= 0.1037$$

বহুল নির্ভরণের ক্ষেত্রেও অনুরূপ পূর্বাভাস করা যায়। ধরা যাক $X_0 = [x_{10}\ x_{20} \ldots . x_{k0}]'$ হলো অনপেক্ষ চলকসমূহের মানের একটি ভেক্টর। এই মানসমূহের জন্য y-এর পূর্বাভাস করা মান হলো

$$y_0 = X_0\hat{B} \tag{৬.৮.১}$$

এবং
$$V(y_0) = [X_0'(X'X)^{-1}X_0 + 1]\sigma^2 \tag{৬.৮.২}$$

সমীকরণ (৬.৮.১) এবং (৬.৮.২)-কে লেখা যায়, যথাক্রমে

$$y_0 = b_0 + b_1x_{10} + b_2x_{20} + \ldots\ldots + b_kx_{k0}$$

এবং
$$V(y_0) = \sigma^2\left[1 + \frac{1}{n}\right] + \sum_{i=1}^{k}(x_{i0} - \bar{x}_i)^2V(b_i)$$

$$+ 2\sum_{i}\sum_{i<j}(x_{i0} - \bar{x}_i)(x_{j0} - \bar{x}_j)\,\mathrm{Cov}(b_i, b_j)$$

এখানে
$$\hat{B} = [b_0\ b_1\ b_2 \ldots \ldots b_k]'$$

এবং

$$V(\hat{B}) = \begin{bmatrix} V(b_0) & \mathrm{Cov}(b_0b_1) & \mathrm{Cov}(b_0b_2) & \ldots & \mathrm{Cov}(b_0b_k) \\ \mathrm{Cov}(b_0b_1) & V(b_1) & \mathrm{Cov}(b_1b_2) & \ldots & \mathrm{Cov}(b_1b_k) \\ \mathrm{Cov}(b_0b_2) & \mathrm{Cov}(b_1b_2) & V(b_2) & \cdots & \mathrm{Cov}(b_2b_k) \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \mathrm{Cov}(b_0b_k) & \mathrm{Cov}(b_1b_k)\ldots & \ldots & \cdots & V(b_k) \end{bmatrix}$$

$$= \sigma^2 \begin{bmatrix} n & \Sigma x_1 & \Sigma x_2 & \ldots & \ldots & \ldots & \Sigma x_k \\ \Sigma x_1 & \Sigma x_1^2 & \Sigma x_1x_2 & \cdots & \cdots & \cdots & \Sigma x_1x_k \\ \Sigma x_2 & \Sigma x_1x_2 & \Sigma x_2^2 & \cdots & \cdots & \cdots & \Sigma x_2x_k \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \Sigma x_k & \Sigma x_1x_k & \Sigma x_2x_k & \cdots & \cdots & \cdots & \Sigma x_k^2 \end{bmatrix}^{-1}$$

এখানে σ^2-এর পরিবর্তে তার নিরূপক

$$\hat{\sigma}^2 = \frac{SSE}{n-k-1} = \frac{1}{n-k-1}\left[\Sigma y_i^2 - \frac{(\Sigma by_i)^2}{n} - \sum_{i}^{k} b_i \sum_{j}^{n} x_{ij} y_j\right]$$

ব্যবহার করে $V(y_0)$-এর নিরূপক পাওয়া যায়। উক্ত ফলাফলের ভিত্তিতে y_0-এর $100(1-\alpha)\%$ নিশ্চয়তা পরিক্ষেপ হলো

$$y_0 \pm t_{\alpha/2,\, n-k-1}\ s.e\,(y_0)$$

উদাহরণ ৬.৩

সারণি ১.১-এর (১ম খণ্ড) উপাত্তের ক্ষেত্রে Slug-এর **Body weight** (y)-এর জন্য **Body length** (x_1), **Mantle length** (x_2), **Keel length** (x_3) এবং **Shell length** (x_4)-এর উপর নির্ভরণ রেখা মিল করা যাক।

বহুল নির্ভরণের বিশ্লেষণ পদ্ধতি অনুসারে উপরিউক্ত উপাত্তের ভিত্তিতে বিশ্লেষণ করা ফলাফল সারণি ৬.৪-এ দেয়া হলো। বিশ্লেষিত ফলাফল থেকে লক্ষ্য করা যাচ্ছে যে,

সারণি ৬.৪ : নির্ভরণ বিশ্লেষণের ফলাফল।

চলক	নির্ভরণ সহগের নিরূপক $[\hat{B}]$	পরিমিত বিচ্যুতি $s.e(\hat{B})$	t-এর মান	P-মান
ধ্রুবক	−5.843	0.625	−9.35	0.00
x_1	−0.015	0.019	−0.83	0.41
x_2	0.144	0.041	3.55	0.00
x_3	0.255	0.035	7.31	0.00
x_4	0.205	0.084	2.44	0.02

সারণি ৬.৪-এর ভেদাঙ্ক বিশ্লেষণ সারণি।

ভেদের উৎস	স্বাধীনতার মাত্রা d.f	বর্গসমষ্টি SS	গড় বর্গসমষ্টি MS	F	P-মান
নির্ভরণ	4	168.108	42.027	81.05	0.00
বিচ্যুতি	45	23.334	0.5185		
মোট	49				

Mantle length (x_2), Keel length (x_3) এবং Shell length (x_4)-এর পরিমাণ বৃদ্ধির কারণে Body weight (y) তাৎপর্যপূর্ণভাবে বৃদ্ধি পাচ্ছে।

উপরিউক্ত মিল করা নির্ভরণ রেখার মাধ্যমে Body length ($x_{10}=80.0$ mm), Mantle length ($x_{20}=38.0$ mm), Keel length ($x_{30}=25.0$ mm) এবং Shell length ($x_{40}=20$ mm) হলে Body weight (y) সম্পর্কে একটি পূর্বাভাস করা যেতে পারে। ধরা যাক পূর্বাভাস করা Body weight হলো y_0।

এখানে

$$
\begin{aligned}
y_0 &= b_0 + b_1x_{10} + b_2x_{20} + b_3x_{30} + b_4x_{40} \\
&= -5.843 - 0.015\times 80 + 0.144\times 38 + 0.255\times 25 + 0.205\times 20 \\
&= 8.904 \text{ gms}
\end{aligned}
$$

উপরিউক্ত বিশ্লেষণের ভিত্তিতে

$\bar{x}_1 = 52.78$, $\bar{x}_2 = 27.50$, $\bar{x}_3 = 15.5$, $\bar{x}_4 = 11.86$

$$
V(\hat{B}) = 0.5185 \begin{pmatrix} 50 & 2639 & 1375 & 775 & 593 \\ 2639 & 145377 & 75204 & 42421 & 32085 \\ 1375 & 75204 & 39377 & 22056 & 16744 \\ 775 & 42421 & 22056 & 12839 & 9392 \\ 593 & 32085 & 16744 & 9392 & 7231 \end{pmatrix}^{-1}
$$

$$
= \begin{pmatrix} 0.3904 & -0.0007 & 0.0012 & -0.0027 & -0.0282 \\ -0.0007 & 0.0004 & -0.0004 & -0.0002 & -0.0003 \\ 0.0012 & -0.0004 & 0.0016 & -0.0003 & -0.0016 \\ -0.0027 & -0.0002 & -0.0003 & 0.0012 & 0.0002 \\ -0.0282 & -0.0003 & -0.0016 & 0.0002 & 0.0070 \end{pmatrix}
$$

তাহলে $V(y_0)$-এর নিরূপক হবে

$$v(y_0) = \hat{\sigma}^2\left(1 + \frac{1}{n}\right) + \sum_{i=1}^{4} (x_{i0} - \bar{x}_i)^2\, V(b_i)$$

$$+ 2\sum_{i}^{4}\sum_{i<j}^{4} (x_{i0} - \bar{x}_i)(x_{j0} - \bar{x}_j)\ \mathrm{Cov}\ (b_i, b_j)$$

$$= 0.5185\left(1 + \frac{1}{50}\right) + (80 - 52.78)^2\, 0.0004$$
$$+ (38 - 27.50)^2\, 0.0016 + (25 - 15.5)^2\, 0.0012$$
$$+ (20 - 11.86)^2\, 0.007 + 2(80 - 52.78)\,(38 - 27.50) \times (-0.0004)$$
$$+ 2(80 - 52.78)\,(25 - 15.5)\,(-0.0002)$$
$$+ 2(80 - 52.78)\,(20 - 11.86)\,(-0.0003)$$
$$+ 2(38 - 27.50)\,(25 - 15.5)\,(-0.0003)$$
$$+ 2(38 - 27.50)\,(20 - 11.86)\,(-0.0016)$$
$$+ 2(25 - 15.5)\,(20 - 11.86)\,(0.0002)$$
$$= 0.8063$$

$$s.e(y_0) = \sqrt{0.8063} = 0.8979$$

সুতরাং পূর্বাভাস দেয়া y_0-এর জন্য 95% নিশ্চয়তা-পরিক্ষেপ হলো

$$y_0 \pm t_{.025,45}\ s.e(y_0)$$

অথবা $\quad 8.904 \pm 2.0141 \times 0.8979$

অথবা $\quad 6.289 \leq y_0 \leq 11.519$

সম্ভাবন-পরিক্ষেপ (Confidence band): অনপেক্ষ চলকের একগুচ্ছ মানের জন্য নির্ভরশীল চলকের একগুচ্ছ গড় মান নিরূপণ করা যায় এবং প্রতিটি নিরূপিত গড় মানের জন্য পরিমিত বিচ্যুতির নিরূপকও নির্ণয় করা যায়। এই অনুচ্ছেদে সে সম্পর্কে আলোচনা করা হয়েছে। অনপেক্ষ চলকের যে কোনো একটি মানের (x_0)-এর জন্য নির্ভরশীল চলকের গড় নিরূপিত মান $(\hat{\mu}_0)$-এর জন্য $100(1-\alpha)\%$ নিশ্চয়তা-পরিক্ষেপ হলো

$$\hat{\mu}_0 \pm t_{\alpha/2,\, n-k-1}\ \text{s.e}(\hat{\mu}_0)$$

ধরা যাক $\mu_{y/x}$ হলো x-এর যে কোনো মানের জন্য y-এর গড় মান। তাহলে $\mu_{x/y}$-এর জন্য নিশ্চয়তা-পরিক্ষেপের উচ্চসীমা (upper confidence limit) এবং নিশ্চয়তা-পরিক্ষেপের নিম্নসীমা (lower confidence limit) হবে যথাক্রমে

$$\hat{\mu}_0 + t_{\alpha/2,\, n-k-1}\ \text{s.e}(\hat{\mu}_0) \text{ এবং } \hat{\mu}_0 - t_{\alpha/2,\, n-k-1}\ \text{s.e}(\hat{\mu}_0)$$

অনপেক্ষ চলকের বিভিন্ন মানের জন্য এই উচ্চসীমা ও নিম্নসীমার মান বিভিন্ন হয়ে থাকে। এখন x-এর বিপরীতে $\mu_{y/x}$-এর উচ্চসীমা ও নিম্নসীমার নিরূপিত মানসমূহ বিক্ষেপ চিত্রে (scatter diagram) প্রতিস্থাপন করা হলে এবং উচ্চসীমার সকল মান একটি বক্ররেখা দ্বারা এবং নিম্নসীমার সকল মান অন্য একটি বক্ররেখা দ্বারা সংযোগ করা হলে যে চিত্র পাওয়া যায় তাকে সম্ভাবন-পরিক্ষেপ বলা হয়। উদাহরণ ৬.১-এর উপাত্তের ভিত্তিতে এরূপ একটি সম্ভাবন-পরিক্ষেপ আঁকা যেতে পারে।

উপরিউক্ত উপাত্তের ক্ষেত্রে নিরূপিত নির্ভরণ রেখা হলো

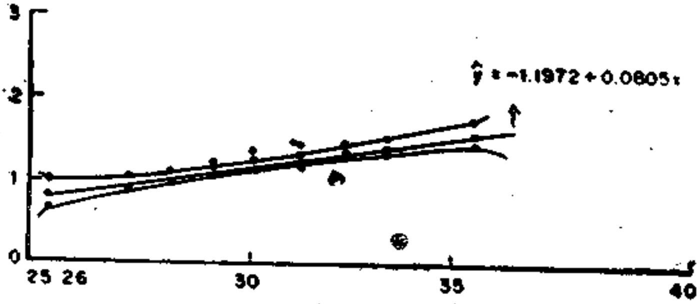
$$\hat{y} = -1.1972 + 0.0805\,x$$

ধরা যাক x-এর বিভিন্ন মান হলো 25, 27, 29, 31, 33 এবং 35। তাহলে y-এর গড় নিরূপিত মানসমূহ তাদের পরিমিত বিচ্যুতির নিরূপকসমূহ এবং তাদের নিশ্চয়তা-পরিক্ষেপের উচ্চসীমা ও নিম্নসীমা সারণি ৬.৫-এ দেখানো যেতে পারে এবং এই মানগুলি চিত্র ৬.৪-এ উপস্থাপন করা হলো।

সারণি ৬.৫ : $\mu_{y/x}$-এর 95% নিশ্চয়তা-পরিক্ষেপ।

x	$\hat{\mu}_0$	s.e$(\hat{\mu}_0)$	নিশ্চয়তা-পরিক্ষেপের	
			নিম্নসীমা	উচ্চসীমা
25	0.82	0.0591	0.68	0.96
27	0.98	0.0417	0.88	1.08
29	1.14	0.0279	1.08	1.20
31	1.30	0.0249	1.24	1.36
33	1.46	0.0355	1.38	1.54
35	1.62	0.0519	1.50	1.74
30.4	1.25	0.0241	1.19	1.30

চিত্র হতে লক্ষ্য করা যাচ্ছে যে, শেষ প্রান্তে অর্থাৎ x-এর সবচেয়ে ছোট এবং সবচেয়ে



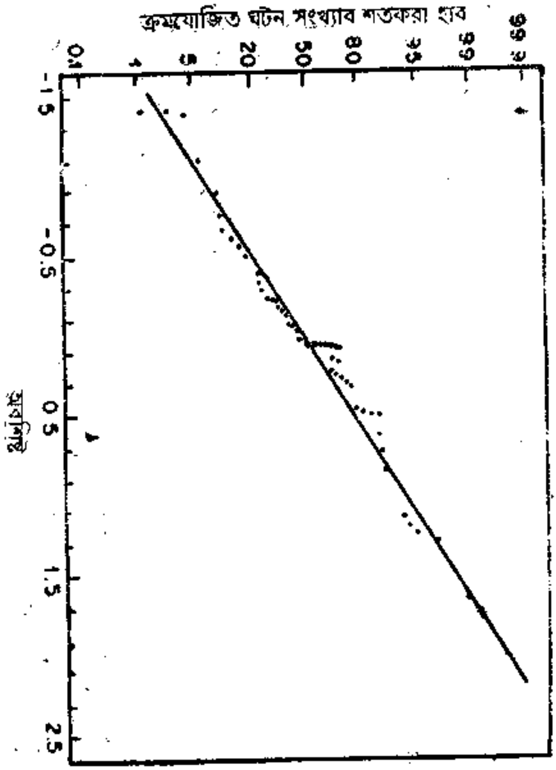
চিত্র ৬.৪ : 95% সম্ভাবন-পরিক্ষেপ।

বড় মানের জন্য সম্ভাবন-পরিক্ষেপ বেশি প্রশস্ত। এই প্রশস্ততা সবচেয়ে কম $x = \bar{x} = 30.4$-এর ক্ষেত্রে। $\bar{x}$-এর চেয়ে মান বাড়লে-কমলে সম্ভাবন-পরিক্ষেপের প্রশস্ততা বাড়তে থাকে।

উপরে যে পদ্ধতিতে নির্ভরশীল চলকের গড় মানের ($\mu_{y/x}$) জন্য সম্ভাবন-পরিক্ষেপ তৈরি করা হয়েছে অনুরূপ পদ্ধতিতে পূর্বাভাস করা মানের জন্যও এই পরিক্ষেপ তৈরি করা যায়।

৬.৯ নির্ভরণ রেখার মূল্যায়ন সম্পর্কীয় আরো তথ্য (More Information on Evaluation of regresion Equation)

এই অধ্যায়ের ৬.৬ অনুচ্ছেদে নির্ভরণ রেখার মূল্যায়ন পদ্ধতি আলোচনা করা হয়েছে। মূল্যায়নের মূল ভিত্তি হলো মিল করা নির্ভরণ রেখার বিশ্লেষিত ফলাফল, বিশেষ করে মিল করা রেখা দ্বারা নির্ভরশীল চলকের ব্যাখ্যায়িত ভেদের (explained sum of squares of y) পরিমাণ। নির্ভরণ রেখার মিল ভাল হওয়ার জন্য কতকগুলি শর্ত আবশ্যক। এই শর্তগুলিই নির্ভরণ বিশ্লেষণের জন্য অনুমান হিসেবে ৬.৩ অনুচ্ছেদে আলোচিত হয়েছে। এই অনুমানগুলির যে কোনো একটি বা একাধিক অনুমানের সত্যতার বিঘ্ন ঘটলেই নির্ভরণ রেখার মিল বিঘ্নিত হবে এবং মিল ভাল হয়েছে বলে দাবি করা যাবে না। কাজেই নির্ভরণ রেখার মূল্যায়নের জন্য যে অনুমানের ভিত্তিতে নির্ভরণ রেখা মিল করা হয়েছে সে অনুমানগুলি বহাল আছে কিনা তা পর্যবেক্ষণ করা উচিত। বর্তমান অনুচ্ছেদে ঐ অনুমানগুলি বহাল আছে কিনা তা পর্যবেক্ষণ পদ্ধতিতে আলোচনা করা হবে। এ সম্পর্কে বিস্তারিত জানার জন্য Bhuyan (১৯৯৫) পর্যালোচনা করা যেতে পারে।



চিত্র ৬.৫ : উদাহরণ ৬.৩-এর উপাত্তের ভিত্তিতে
পরিমিত সম্ভাবনা রেখা।

নির্ভরণ রেখা মিল করার জন্য একটি মূল অনুমান হলো প্রতিকৃতির বিচ্যুতি (error)-এর বিন্যাস পরিমিত বিন্যাস হবে। এই বিচ্যুতির বিন্যাস পরিমিত বিন্যাস না হলে নির্ভরণ বিশ্লেষণে অন্তর্ভুক্ত যাচাই পদ্ধতি অর্থবহ হবে না এবং যাচাই থেকে নেয়া সিদ্ধান্ত সঠিক হবে না। কাজেই বিচ্যুতির বিন্যাস পরিমিত বিন্যাস অনুসরণ করে কিনা তা পর্যবেক্ষণ করা প্রয়োজন। এই পর্যবেক্ষণ করা হয় অবশিষ্ট (residual) বিশ্লেষণের মাধ্যমে। এখানে অবশিষ্ট হলো

$$\hat{U} = Y - X\hat{\beta}$$

অথবা $\quad \hat{u}_1 = y_1 - b_0 - b_1x_{11} - b_2x_{21} - \ldots - b_kx_{k1}$

এখানে $\quad b_0, b_1, b_2, \ldots, b_k$ হলো যথাক্রমে $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$-এর নিরূপক।

উপরিউক্ত অবশিষ্ট ($\hat{u}_1$) ব্যবহার করে পরিমিত সম্ভাবনা রেখা (normal probability plot) অঙ্কন করে বিচ্যুতির পরিমিত বিন্যাস সম্পর্কে ধারণা লাভ করা যায়।

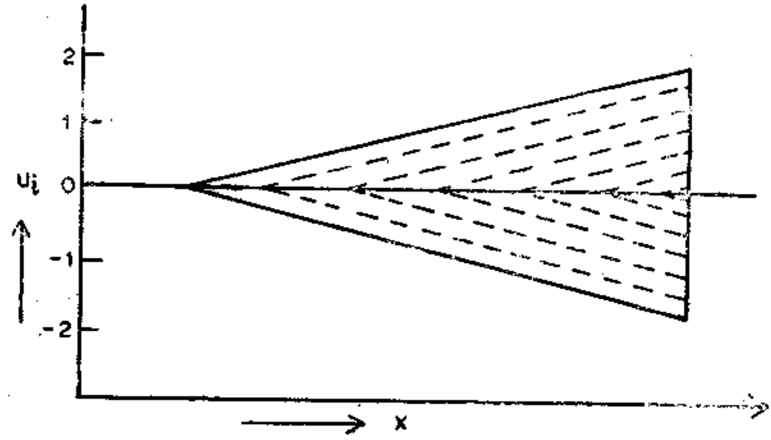
এক্ষেত্রে অবশিষ্টসমূহকে বা অবশিষ্টসমূহের শ্রেণী ব্যবধানের উচ্চ মানকে X-অক্ষে এবং বিভিন্ন শ্রেণীর ক্রমযোজিত ঘটনসংখ্যার শতকরা হারকে Y-অক্ষে প্রতিস্থাপন করে একটি চিত্র পাওয়া যায়। ঘটনসংখ্যার ক্রমযোজিত শতকরা হার-এর বিন্দুগুলি একটি উর্ধ্বমুখী সরলরেখায় অবস্থিত হলে বিচ্যুতি পরিমিত বিন্যাস অনুসরণ করে বলে সিদ্ধান্ত নেয়া যায়। বিন্দুগুলির অবস্থান সরলরেখা হতে বেশি দূরে হলে পরিমিত বিন্যাস-এর অনুমান মারাত্মকভাবে বিঘ্নিত হয়েছে বিবেচনা করা যায়। সেক্ষেত্রে বিশ্লেষিত ফলাফল যথাযথ হয় না এবং নির্ভরণ রেখার মিলও ভাল হয়নি বলে বিবেচনা করা যায়। এখানে উদাহরণ ৬.৩-এর উপাত্তের ক্ষেত্রে একটি পরিমিত সম্ভাবনা রেখা চিত্র ৬.৫-এ উপস্থাপন করা হলো।

এই চিত্র হতে লক্ষ্য করা যাচ্ছে যে, অবশিষ্টসমূহের ক্রমযোজিত শতকরা হার সঠিকভাবে একটি সরলরেখায় অবস্থিত নয়। কাজেই বিচ্যুতির বিন্যাস সম্পর্কীয় অনুমান বিঘ্নিত হয়েছে বলা যায়। সে কারণে ঐ মিল করা রেখা সম্পর্কীয় সিদ্ধান্ত যথাযথ নয়। এখানে বিচ্যুতির বিন্যাস পরিমিত বিন্যাস না হলেও তার কাছাকাছি। তাই নির্ভরণ বিশ্লেষণ একেবারে অর্থহীন নয়।

বিচ্যুতি সম্পর্কীয় আরেকটি গুরুত্বপূর্ণ অনুমান হলো ঐগুলির ভেদাঙ্ক ধ্রুবক বা স্থির। বিচ্যুতির ভেদাঙ্ক স্থির (constant) না হলে সাধারণ ন্যূনতম বর্গপদ্ধতি (ordinary least squares method) প্রয়োগ করে নির্ভরণ বিশ্লেষণ করা উচিত নয়। কারণ এই পদ্ধতি প্রয়োগের ফলে নির্ভরণ সহগসমূহের নিরূপকসমূহ নির্ঝুঁকি (unbiased) হলেও তাদের ভেদাঙ্ক ন্যূনতম হয় না। কাজেই ঐ ভেদাঙ্ক ব্যবহার করে কোনো যাচাই করা হলে বা কোনো নিশ্চয়তা-পরিক্ষেপ নির্ণয় করা হলে তা যথাযথ হয় না। সে কারণে বিচ্যুতির ভেদাঙ্ক-এর স্থিরতা সম্পর্কে একটি সিদ্ধান্তে আসা উচিত। বিচ্যুতির ভেদাঙ্কের স্থিরতা সম্পর্কেও সিদ্ধান্ত নেয়া যায় অবশিষ্ট বিশ্লেষণের মাধ্যমে। অবশ্য এ সিদ্ধান্ত নেয়ার জন্য যাচাই পদ্ধতির প্রয়োগও করা হয়।

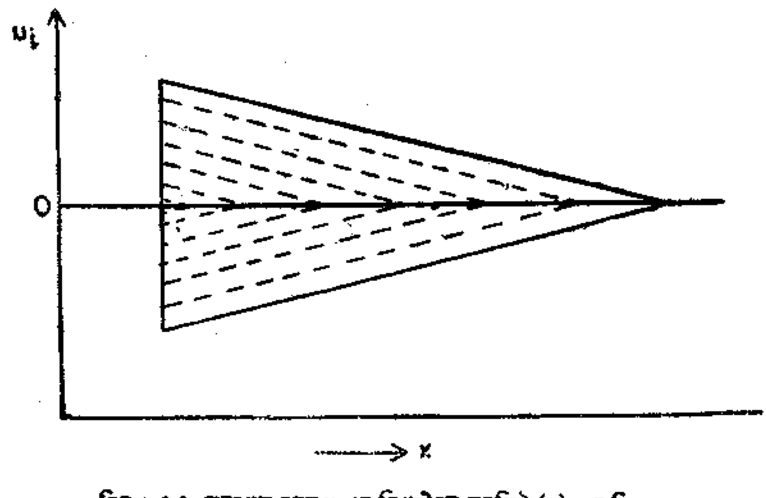
বিচ্যুতির ভেদাঙ্কের স্থিরতা নষ্ট হওয়ার কারণ অনেক। তবে বেশির ভাগ ক্ষেত্রেই ভেদাঙ্কের স্থিরতা নষ্ট হয় অনপেক্ষ চলকের মানের পরিবর্তনের কারণে। কাজেই অনপেক্ষ চলকের বিপরীতে অবশিষ্ট-এর চিত্র অঙ্কন করে বিচ্যুতির ভেদাঙ্কের স্থিরতা সম্পর্কে সিদ্ধান্তে উপনীত হওয়া যায়। এই চিত্রের জন্য X-অক্ষে একটি অনপেক্ষ চলকের মান এবং Y-অক্ষে অবশিষ্ট-এর মান প্রতিস্থাপন করা হয়। তারপর অবশিষ্ট-এর শূন্য মান থেকে Y-অক্ষ হতে X-অক্ষের সমান্তরাল করে একটি সরলরেখা টানা হয়। বিচ্যুতির ভেদাঙ্ক স্থির হলে অবশিষ্ট-এর জন্য প্রতিস্থাপিত বিন্দু ঐ রেখার উভয় পাশে মোটামুটি সমান দূরত্বে অবস্থিত হবে। অপরপক্ষে অনপেক্ষ চলকের মান

বৃদ্ধির সাথে বিচ্যুতির ভেদাঙ্ক বৃদ্ধি পেলে [$\sigma_i^2 \propto x_i^2$ বা $\sigma_i^2 \propto x_i$] অবশিষ্টসমূহ ক্রমান্বয়ে ঐ রেখার উভয় পাশে দূরে দূরে অবস্থান নিবে। অর্থাৎ x-এর বিপরীতে অবশিষ্ট প্রতিস্থাপন করে চিত্র আঁকা হলে চিত্রের আকার হবে নিম্নরূপ :



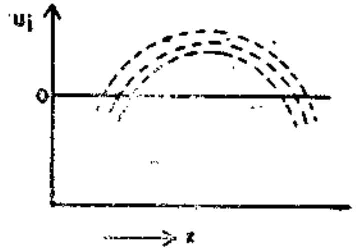
চিত্র ৬.৬ : অনপেক্ষ চলক x-এর বিপরীতে অবশিষ্ট (u)-এর চিত্র।

আবার অনপেক্ষ চলকের মান বৃদ্ধির সাথে সাথে বিচ্যুতির ভেদাঙ্কের মান কমতে থাকলে [$\sigma_i^2 \propto 1/x_i^2$ বা $\sigma_i^2 \propto 1/x_i$], এই চিত্রের আকার হবে নিম্নরূপ : এছাড়াও

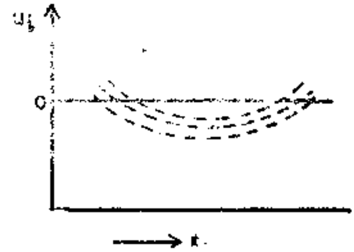


চিত্র ৬.৭ : অনপেক্ষ চলক x-এর বিপরীতে অবশিষ্ট (u)-এর চিত্র।

x-এর বিপরীতে আঁকা অবশিষ্ট-এর চিত্র আরো বিভিন্ন ধরনের আকৃতি ধারণ করে, যেমন,

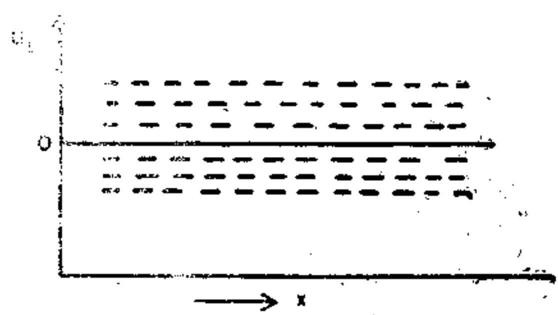


চিত্র ৬.৮ : অনপেক্ষ চলকের বিপরীতে অবশিষ্ট-এর চিত্র ।



চিত্র ৬.৯ : অনপেক্ষ চলকের বিপরীতে অবশিষ্ট-এর চিত্র ।

কিন্তু, বিচ্যুতির ভেদাঙ্ক স্থির হলে $[\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_n^2]$ এই চিত্রের আকার হবে নিম্নরূপ বা এর কাছাকাছি। অর্থাৎ x-এর বিপরীতে অবশিষ্ট প্রতিস্থাপন করা হলে-

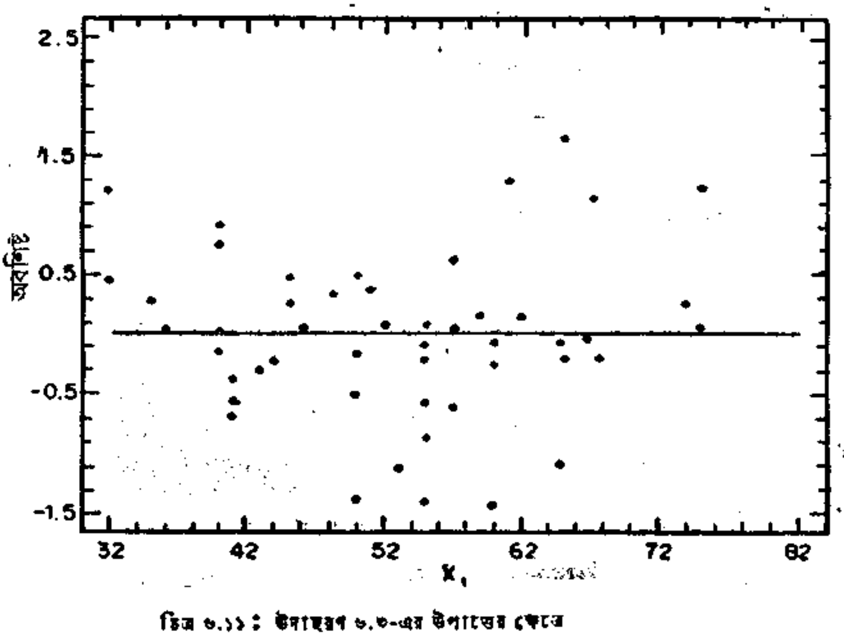


চিত্র ৬.১০ : অনপেক্ষ চলকের বিপরীতে অবশিষ্ট-এর চিত্র ।

অবশিষ্ট-এর চিত্র উপরে অঙ্কিত চিত্র (৬.৬), (৬.৭), (৬.৮) এবং (৬.৯)-এর ন্যায় কোনো বিশেষ আকার ধারণ করবে না। অর্থাৎ x-এর মানের পরিবর্তনের সাথে u-এর মানের পরিবর্তনের কোনো সম্পর্ক থাকবে না। বিষয়টি বাস্তব উদাহরণের ক্ষেত্রে লক্ষ্য করা যাক।

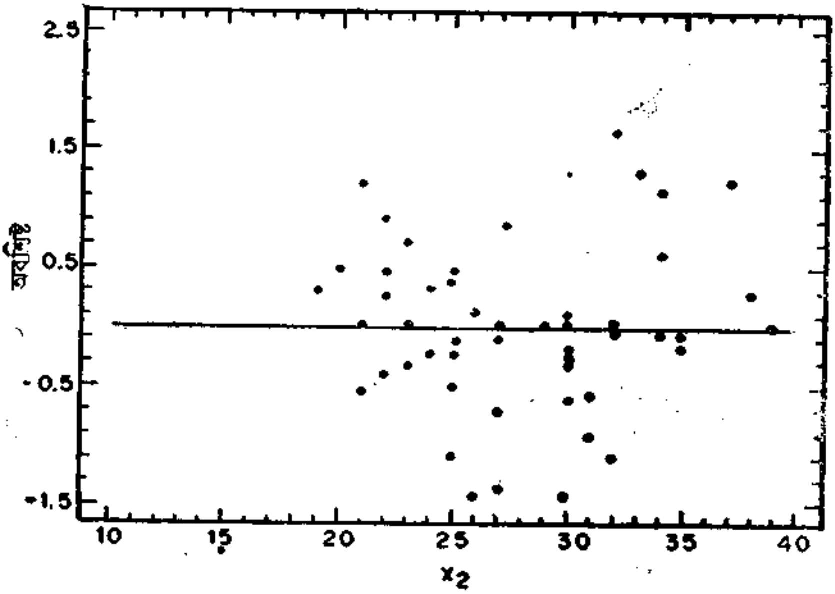
উদাহরণ ৬.৩-এর উপাত্তের ক্ষেত্রে অনপেক্ষ চলক x_1 ও x_2-এর বিপরীতে অঙ্কিত অবশিষ্ট-এর চিত্র [যথাক্রমে চিত্র ৬.১১ এবং চিত্র ৬.১২] লক্ষ্য করা যাক। উক্ত চিত্র হতে লক্ষণীয় বিষয় হলো অনপেক্ষ চলকের মানের পরিবর্তনের সাথে অবশিষ্ট-এর মান কোনো নির্দিষ্ট নিয়ম মেনে চলেনি। অর্থাৎ অবশিষ্ট-এর মানে কোনো গতি ধারা লক্ষণীয় নয়। কাজেই উপরিউক্ত উদাহরণের ক্ষেত্রে বিচ্যুতির ভেদাঙ্কের স্থিরতা বিদ্যমান আছে বিবেচনা করা যায়।

উপরে বর্ণিত চিত্রসমূহ প্রতিকৃতিতে অনপেক্ষ চলকের সংযুক্তি সম্পর্কে সিদ্ধান্ত নিতেও সাহায্য করে। প্রতিকৃতিতে সংযুক্ত কোনো অনপেক্ষ চলকের বিপরীতে অবশিষ্ট-এর চিত্র আঁকা হলে ঐ চিত্র যদি চলকের মানে এবং অবশিষ্ট-এর মানে কোনো সম্পর্ক নির্দেশ করে, তাহলে চলকটিকে প্রতিকৃতি হতে বাদ দেয়া যেতে পারে। আবার অবশিষ্টকে এমন কোনো অনপেক্ষ চলক, যা প্রতিকৃতিতে সংযুক্ত করা হয়নি, এর বিপরীতে প্রতিস্থাপন করে চিত্র আঁকা হলে ঐ চিত্র যদি অবশিষ্ট ও ঐ চলকের মধ্যে কোনো সম্পর্ক নির্দেশ করে, তাহলে চলকটিকে প্রতিকৃতিতে সংযুক্ত করে নতুন করে নির্ভরণ বিশ্লেষণ করতে হয়।



চিত্র ৬.১১ : উদাহরণ ৬.৩-এর উপাত্তের ক্ষেত্রে X_1-এর বিপরীতে অবশিষ্ট-এর চিত্র।

বিচ্যুতির ভেদাঙ্কের সমসত্বতা (homogeneity) পর্যবেক্ষণ করার জন্য অবশিষ্টকে পূর্বাভাস করা নির্ভরশীল চলকের $[\hat{y}]$ বিপরীতে প্রতিস্থাপন করেও চিত্র আঁকা যায়। ভেদাঙ্কের সমসত্বতা বজায় থাকলে চিত্রে $\hat{y}$ ও u-এর মধ্যে কোনো সম্পর্ক লক্ষণীয় হবে না। বিষয়টি উদাহরণ ৬.৩-এর উপাত্তের ক্ষেত্রে লক্ষ্য করা যাক। এই উপাত্তের

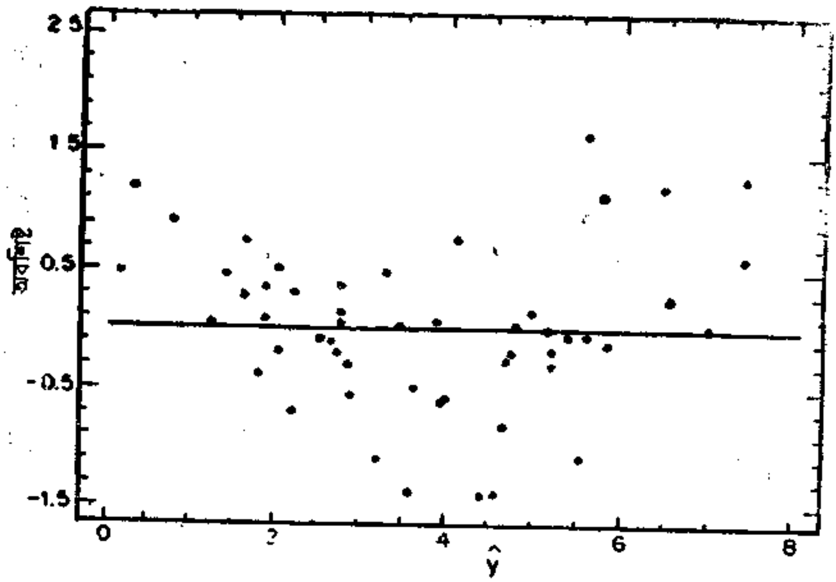


চিত্র ৬.১২ উদাহরণ ৬.৩-এর উপাত্তের ক্ষেত্রে x_2-এর বিপরীতে অবশিষ্ট-এর চিত্র।

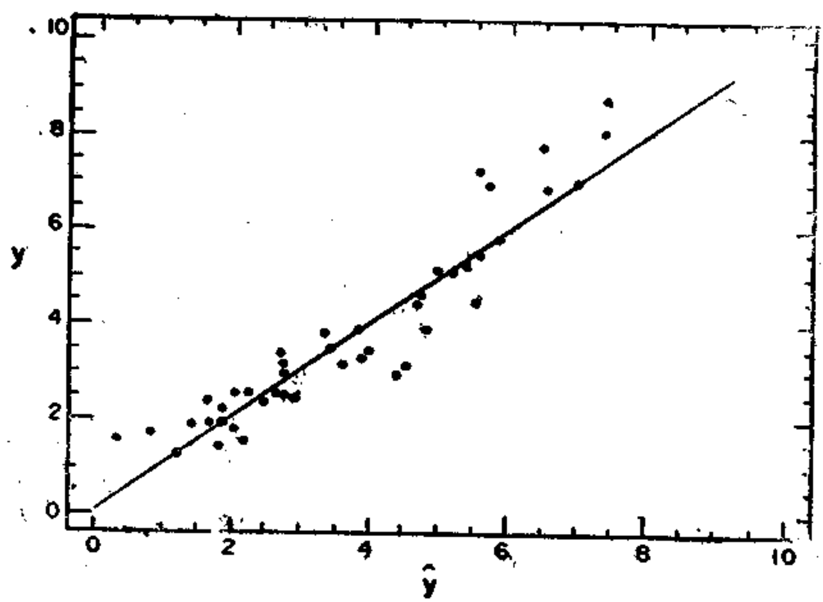
জন্য $\hat{y}$-এর বিপরীতে অবশিষ্ট u-কে প্রতিস্থাপন করে চিত্র আঁকা হয়েছে [চিত্র ৬.১২]। এই চিত্র হতে লক্ষ্য করা যাচ্ছে যে, $\hat{y}$-এর মানের পরিবর্তনের সাথে u-এর মানের পরিবর্তনের কোনো নির্দিষ্ট সম্পর্ক নেই। ফলে বিচ্যুতির ভেদাঙ্ক সমসত্ত্ব বিবেচনা করা যায়।

চিত্র ৬.১৩ প্রতিকৃতির রৈখিকতা (linearity) সম্পর্কে সিদ্ধান্ত নিতেও সাহায্য করে। কোনো উপাত্তের জন্য রৈখিক প্রতিকৃতি সঠিক হলে $\hat{y}$-এর বিপরীতে u প্রতিস্থাপন করে চিত্র আঁকা হলে ঐ চিত্র হতে $\hat{y}$ ও u-এর মধ্যে কোনো সম্পর্ক লক্ষ্য করা যাবে না। যেমনটি হয়েছে ৬.১৩ চিত্রে। এই চিত্র হতে বুঝা যাচ্ছে যে, উদাহরণ ৬.৩-এর জন্য রৈখিক প্রতিকৃতি সঠিক।

মিল করা নির্ভরণ রেখার মূল্যায়ন করার জন্য $\hat{y}$-এর বিপরীতে y-এর চিত্র আঁকা যায়। মিল ভাল হলে $\hat{y}$-এর বিপরীতে প্রতিস্থাপিত y-এর বিন্দুগুলি একটি সরলরেখার উভয় পাশে এমনভাবে অবস্থিত হবে যাতে সরলরেখা হতে বিন্দুগুলির দূরত্ব ন্যূনতম হয়। বিষয়টি ৬.১৩ চিত্রে লক্ষ্য করা যেতে পারে। চিত্র হতে বুঝা যায়, উদাহরণ ৬.৩-এর ক্ষেত্রে নির্ভরণ রেখার মিল ভাল হয়েছে।



চিত্র ৬.১৩ : উদাহরণ ৬.৩-এর উপাত্তের ক্ষেত্রে $\hat{y}$-এর বিপরীতে u-এর চিত্র।



চিত্র ৬.১৪ : উদাহরণ ৬.৩-এর উপাত্তের ক্ষেত্রে $\hat{y}$-এর বিপরীতে y-এর চিত্র।

নির্ভরণ রেখা মিল করার জন্য আরেকটি গুরুত্বপূর্ণ অনুমান হলো বিচ্যুতির অনপেক্ষতা (independence)। সাধারণত কালীন সারি (time series) উপাত্ত হলে নির্ভরশীল চলকের প্রতিটি মান তার পূর্ববর্তী সময়ের বা পূর্ববর্তী যে কোনো সময়ের মানের উপর নির্ভরশীল থাকে। যেমন, গোলাপ ফুলের উৎপাদনে নাইট্রোজেন সারের প্রভাব জানার জন্য কোনো পরীক্ষা পরপর কয়েক বার করা হলে চলতি বছরের উৎপাদন পূর্ববর্তী বছরের উৎপাদনের সাথে সম্পর্কিত হবে। কারণ বিভিন্ন বছরের উৎপাদন একই গাছ হতে একই সার প্রয়োগ করে পাওয়া যায়। এক্ষেত্রে বিচ্যুতির অনপেক্ষতা বিঘ্নিত হয়। তাছাড়া প্রতিকৃতিতে সম্ভাব্য সকল অনপেক্ষ চলক অন্তর্ভুক্ত না হলেও বিচ্যুতির অনপেক্ষতা বিঘ্নিত হয়। কাজেই নির্ভরণ রেখার মূল্যায়ন করার জন্য বিচ্যুতির অনপেক্ষতা যাচাই করে দেখা দরকার।

ধরা যাক ρ হলো বিচ্যুতির স্বয়ংসম্পর্ক সহগ (autocorrelation coefficient)। বিচ্যুতি অনপেক্ষ হলে $\rho = 0$ হতে হবে। কাজেই নাস্তি কল্পনা

$$H_0 : \rho = 0$$

এর বিপরীতে বিকল্প কল্পনা

$$H_A : \rho > 0 \text{ বা } H_A \, \rho < 0$$

যাচাই করে দেখা দরকার। এই যাচাই-এর জন্য যাচাই তথ্যজমান হলো

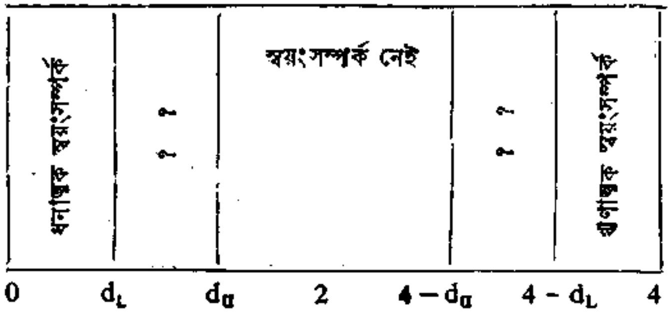
$$d = \frac{\sum_{t=2}^{n} (u_t - u_{t-1})^2}{\sum_{t=1}^{n} u_t^2}$$

এই 'd'-কে বলা হয় Durbin Watson 'd' statistic।

এখানে $u_t \, [t = 1, 2, \ldots, n]$ হলো t সময়ের জন্য অবশিষ্ট

$$[\text{Residual}; \; u_t = y_t - b_0 - b_1 x_{1t} - b_2 x_{2t} - \ldots - b_k x_{kt}]$$

Durbin and Watson k-এর এবং n-এর বিভিন্ন মানের জন্য 'd'-এর ন্যূনতম মান $[d_2]$ এবং উচ্চতম মান d_U একটি সারণিতে উপস্থাপন করেছেন। নির্ণেয় 'd' এবং সারণিকৃত d_L ও d_U ব্যবহার করে নিম্নরূপভাবে বিচ্যুতির অনপেক্ষতা সম্পর্কে সিদ্ধান্ত নিতে হয়। এই সিদ্ধান্ত নেয়ার জন্য ৬.১৫ চিত্র বিশেষ সহায়ক। চিত্র হতে লক্ষ্য করা যাচ্ছে যে



চিত্র ৬.১৫ : d-এর মানের অবস্থানের চিত্র।

(i) $d < d_L$ হলে $\rho = 0$ বাতিল এবং $\rho > 0$ গ্রহণযোগ্য।

(ii) $d > 4 - d_L$ হলে $\rho = 0$ বাতিল এবং $\rho < 0$ গ্রহণযোগ্য।

উভয় ক্ষেত্রেই বিচ্যুতির অনপেক্ষতা বিঘ্নিত হয় বলে সিদ্ধান্ত নেয়া যায়।

(iii) $d = 2$ হলে $\rho = 0$ গ্রহণযোগ্য, অর্থাৎ বিচ্যুতি অনপেক্ষভাবে বিন্যাসিত।

(iv) $d_L < d < d_U$ এবং $4 - d_U < d < 4 - d_L$ হলে কোনো সিদ্ধান্তে উপনীত হওয়া যায় না। এরূপ ক্ষেত্রে সম্ভব হলে আরো উপাত্ত সংগ্রহ করে পুনরায় যাচাই কাজ করা হয়।

উদাহরণ ৬.৩-এর ক্ষেত্রে $d = 1.28$। উক্ত উদাহরণে $n = 50$, $k = 4$ এবং 5% সংশয় মাত্রায় $d_L = 1.38$, $d_U = 1.72$। এখানে $d < d_L$ হওয়ায় বিচ্যুতির অনপেক্ষতা $[\rho = 0]$ বজায় নেই বিবেচনা করা যায়। বিচ্যুতির এই স্বয়ংসম্পর্ক-এর কারণ সম্ভবত প্রতিকৃতিতে যথাযথ অনপেক্ষ চলকের সংযোগ হয়নি। এখানে হয়ত আরো এক বা একাধিক চলক থাকতে পারে যেগুলি y-এর ভেদ ব্যাখ্যা করতে পারে বা এমন কোনো চলক আছে যা প্রতিকৃতি হতে বাদ দেওয়া প্রয়োজন।

নির্ভরণ বিশ্লেষণের জন্য একটি বিশেষ অনুমান হলো অনপেক্ষ চলকসমূহের মধ্যে কোনো সম্পর্ক থাকবে না। সাধারণ রৈখিক নির্ভরণ প্রতিকৃতি

$$Y = X\beta + U$$

এর ক্ষেত্রে এই অনুমান হলো Rank $(X) = k + 1$, যদি X ম্যাট্রিক্স-এর অর্ডার হয় $n \times (k+1)$। অর্থাৎ X ম্যাট্রিক্স-এর সকল স্তম্ভ অনপেক্ষ। এই অনুমান বিঘ্নিত হলে মাল্টিকোলিনিয়ারিটি (Multicollinearity) সমস্যার সম্মুখীন হতে হয়। তবে কি ধরনের সমস্যা সৃষ্টি হয় তা পরবর্তী অনুচ্ছেদে আলোচনা করা হবে। তার আগে এই অনুমান বিঘ্নিত হয় কিনা তা পর্যবেক্ষণের পদ্ধতি আলোচনা করা যাক।

ধরা যাক কোনো নির্ভরণের জন্য $X_1, X_2, \ldots, X_k$ অনপেক্ষ চলকসমূহ। এই চলকসমূহ সংশ্লেষিত হলে মাল্টিকোলিনিয়ারিটি সমস্যার সৃষ্টি হয়। ধরা যাক চলকসমূহের মধ্যে নিম্নলিখিত সম্পর্ক বিদ্যমান :

$$X_1 = \beta_{10} + \beta_{12} X_2 + \beta_{13} X_3 + \ldots + \beta_{1k} X_k + e_1$$

$$X_2 = \beta_{20} + \beta_{21} X_1 + \beta_{23} X_3 + \ldots + \beta_{2k} X_k + e_2$$

$$\ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots$$

$$X_k = X_{k0} + \beta_{k1} X_1 + \beta_{k2} X_2 + \ldots + \beta_{kk-1} X_{k-1} + e_k$$

এই রৈখিক সম্পর্কগুলি পরিসংখ্যানিকভাবে তাৎপর্যপূর্ণ হলে মাল্টিকোলিনিয়ারিটি সমস্যা বিদ্যমান বিবেচনা করা হয়। এখানে k সংখ্যক সমীকরণ বিদ্যমান। i-তম সমীকরণ $[i = 1, 2, \ldots, k]$ হতে সিদ্ধান্ত সহগ [coefficient of determination) হলো R_i^2। এই R_i^2 ব্যবহার করে i-তম সম্পর্কের তাৎপর্য যাচাই করার যাচাই তথ্যজমান হলো

$$F_i = \frac{R_i^2/(k-1)}{(1-R_i^2)/(n-k-2)}$$

এখানে k হলো অনপেক্ষ চলকসমূহের সংখ্যা এবং $(k-1)$ হলো প্রতি সমীকরণে সংযুক্ত অনপেক্ষ চলকের সংখ্যা। এই F_i-এর বিন্যাস হলো $(k-1)$ এবং $(n-k-2)$ স্বাধীনতার মাত্রাসহ ভেদাঙ্ক অনুপাত বিন্যাস। এই যাচাই Farrar and Glanber (1967) যাচাই হিসেবে পরিচিত। F_i-এর তাৎপর্যপূর্ণ মান নির্দেশ করে যে, X_i অন্যান্য অনপেক্ষ চলকের সাথে সংশ্লেশিত।

অনপেক্ষ চলকসমূহ সংশ্লেষিত হলেই যে নির্ভরণ বিশ্লেষণ অর্থহীন হয়ে যাবে তা নয়, সংশ্লেষিত চলক নিয়েও নির্ভরণ বিশ্লেষণ করা যায় যদি সংশ্লেষণের সমস্যা মারাত্মক না হয়। সংশ্লেষণের মারাত্মক অবস্থা চিহ্নিত করার জন্য ভেদাঙ্ক বর্ধিতকরণ উপাদান (Variance Inflation Factor, VIF) নির্ণয় করা হয়, এখানে

$$VIF_i = \frac{1}{1-R_i^2}$$

X_i অন্যান্য চলকের সাথে খুব বেশি সংশ্লেষিত হলে R_i^2-এর মান প্রায় 1 হবে এবং সেক্ষেত্রে VIF_i-এর মান বড় হবে। R_i^2-এর মান ছোট হওয়ার অর্থ হলো X_i অন্যান্য চলকের সাথে মারাত্মকভাবে সংশ্লেষিত নয়। সেক্ষেত্রে VIF_i ছোট হবে। Chatterjee and Price (1991) উল্লেখ করেছেন যে VIF_i-এর মান 10 অথবা তার বড় হলে কলিনিয়ারিটি সমস্যা প্রকট হয়।

মাল্টিকোলিনিয়ারিটি-এর একটি পরিমাপ করা হয় অনপেক্ষ চলকসমূহের সংশ্লেষণ ম্যাট্রিক্স-এর আইগেন মানের ভিত্তিতে। ধরা যাক $X_1, X_2, \cdots, X_K$ চলকের জন্য সংশ্লেষণ ম্যাট্রিক্স হলো

$$R = \begin{pmatrix} 1 & r_{12} & r_{13} & \cdots & r_{1k} \\ r_{21} & 1 & r_{23} & \cdots\ \cdots & r_{2k} \\ \cdots & \cdots & & \cdots & \cdots \\ r_{k1} & r_{k2} & & \cdots & 1 \end{pmatrix}_{k \times k}$$

মনে করা যাক এই R ম্যাট্রিক্স-এর আইগেন মানগুলি হলো $\lambda_1 \geq \lambda_2 \geq \lambda_3 .. \geq \lambda_k$।
এখন আইগেন মানগুলি ব্যবহার করে মাল্টিকোলিনিয়ারিটি-এর সার্বিক পরিমাপ করা হয় কন্ডিশন নম্বর (condition number) C_1 বের করে।

এখানে $$C_1 = \sqrt{\frac{\lambda_{max}}{\lambda_1}}$$

Chatterjee and Price (1991) উল্লেখ করেছেন যে C-এর মান 15-এর বেশি হলে মাল্টিকোলিনিয়ারিটি সমস্যা প্রকট হয়। সেক্ষেত্রে নির্ভরণ বিশ্লেষণের জন্য কিছু সংশোধিত পদ্ধতি প্রয়োগ করতে হয়। এ সম্পর্কে পরে আলোচনা করা হবে। এখন উদাহরণ ৬.৩-এর ক্ষেত্রে মাল্টিকোলিনিয়ারিটি-এর প্রবটতা যাচাই করে দেখা যাক। এই উপাত্তের জন্য R_1^2; F_1, λ_1 এবং VIF_1 এবং C-এর মান সারণি ৬.৬-এ উপস্থাপন করা হলো।

সারণি ৬.৬ : মাল্টিকোলিনিয়ারিটি-এর প্রকটতা যাচাই।

নির্ভরশীল চলক	অনপেক্ষ চলক	R_1^2	F_1	P-মান	VIF_1	λ_1	C_1
X_1	X_2, X_3, X_4	0.7581	48.05	0.000	4.13	3.1020	1.00
X_2	X_1, X_3, X_4	0.7986	60.82	0.000	4.96	0.5191	2.44
X_3	X_1, X_2, X_4	0.4827	14.31	0.000	1.93	0.2426	3.58
X_4	X_1, X_2, X_3	0.6216	25.67	0.000	2.64	0.1363	4.77

সারণি ৬.৬ হতে লক্ষ্য করা যাচ্ছে যে উদাহরণ ৬.৩-এ ব্যবহৃত অনপেক্ষ চলক-সমূহ সংশ্লেষিত [কারণ সকল F_1-এর P-মান < 0.05]। কাজেই মাল্টিকোলিনিয়ারিটি সমস্যা আছে। কিন্তু এই সমস্যা নির্ভরণ বিশ্লেষণকে ক্ষতিগ্রস্ত করেনি। কারণ কোনো

VIF_1-এর মান 10-এর অতিরিক্ত নয় এবং কোনো কন্ডিশন নম্বর 15-এর অতিরিক্ত নয়। সুতরাং উক্ত উদাহরণে মাল্টিকোলিনিয়ারিটি সমস্যা প্রকট নয়।

মাল্টিকোলিনিয়ারিটি সমস্যা চিহ্নিত করার এবং এই সমস্যার সমাধান করার একটি বিশেষ ব্যবস্থা হলো ভেদাঙ্ক সমানুপাত নির্ণয় করা [Bel'sley et al (1980) ; variance proportion]। এ সম্পর্কে কিছুটা আলোকপাত করা যাক।

ধরা যাক $X_1, X_2, \ldots, X_k$ হলো k সংখ্যক আদর্শায়িত অনপেক্ষ চলক। এগুলির সংশ্লেষণ ম্যাট্রিক্স R-এর আইগেন মান হলো $\lambda_1 \geqslant \lambda_2 \geqslant \lambda_3 \geqslant \cdots \geqslant \lambda_k$। এই আদর্শায়িত চলকসমূহের জন্য নির্ভরণ প্রতিকৃতি হলো

$$Y = \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + e$$

ধরা যাক j-তম $(j = 1, 2, \cdots, k)$ আইগেন মানের জন্য আইগেন ভেক্টর হলো

$$v_j = [\, a_{j1} \; a_{j2} \; \cdots \; a_{jk} \,]'$$

এই আইগেন ভেক্টর ব্যবহার করে X-সমূহকে নতুন চলক Z-এ পরিণত করা যায়, যেখানে Z_j $(j = 1, 2, \ldots, k)$ হলো $X_1, X_2, \cdots, X_k$ এর রৈখিক সংযোগ এবং এই সংযোগের জন্য সহগ হলো a_{ij} $(i \neq j = 1, 2, \cdots, k)$। যথা:

$$\begin{aligned}
Z_1 &= a_{11}X_1 + a_{12}X_2 + \ldots\ldots + a_{1k}X_k \\
Z_2 &= a_{21}X_1 + a_{22}X_2 + \ldots\ldots + a_{2k}X_k \\
&\cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \\
Z_j &= a_{j1}X_1 + a_{j2}X_2 + \ldots + a_{jk}X_k \\
&\cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \\
Z_k &= a_{k1}X_1 + a_{k2}X_2 + \cdots + a_{kk}X_k
\end{aligned}$$

এই নতুন চলক Z-কে বলা হয় প্রধান উপাদান (principal component)। এগুলি ব্যবহার করে Y-এর জন্য প্রতিকৃতি বিবেচনা করা যাক

$$Y = B_1 Z_1 + B_2 Z_2 + \cdots \cdots + B_k Z_k + \epsilon$$

ধরা যাক $\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_k$ হলো যথাক্রমে $\beta_1, \beta_2, \ldots, \beta_k$-এর নিরূপক এবং $b_1, b_2, \cdots, b_k$ হলো যথাক্রমে $B_1, B_2, \cdots \cdots, B_k$ এর-নিরূপক। উদ্দেশ্য হলো $\hat{\beta}_1, \hat{\beta}_2, \cdots, \hat{\beta}_k$ এর মান নির্ণয় করা। প্রধান উপাদান ব্যবহার করার কারণে [λ_j এর সকল মান ভিন্ন ভিন্ন হলে এবং যে কোনো $\lambda_j > 0.70$ হলে ; Jeffers (1967), Jolliffe (1972,1973)] Z গুলি অনপেক্ষ হয় এবং যথারীতি ন্যূনতম বর্গপদ্ধতি (Method of OLS) প্রয়োগ করে যথাযথ নিরূপক $b_1, b_2, \ldots, b_k$ পাওয়া যায়। এই নিরূপক হতে $\hat{\beta}_j$-এর মান পাওয়া যায় নিম্নরূপভাবে :

$$\hat{\beta}_1 = a_{11} b_1 + a_{21} b_2 + a_{31} b_3 + \cdots + a_{k1} b_k$$

$$\hat{\beta}_2 = a_{12} b_1 + a_{22} b_2 + a_{32} b_3 + \cdots + a_{k2} b_k$$

$$\cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots$$

$$\hat{\beta}_j = a_{1j} b_1 + a_{2j} b_2 + a_{3j} b_3 + \cdots + a_{kj} b_k$$

$$\cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots$$

$$\hat{\beta}_k = a_{1k} b_1 + a_{2k} b_2 + \cdots + a_{kk} b_k$$

এই নিরূপকগুলির নিরূপিত ভেদাঙ্ক (estimated variance) হলো

$$v(\hat{\beta}_j) = \left[\frac{a^2_{1j}}{\lambda_1} + \frac{a^2_{2j}}{\lambda_2} + \cdots + \frac{a^2_{kj}}{\lambda_k} \right] S^2$$

এখানে S^2 হলো প্রধান উপাদান (Z_j) ব্যবহার করে নির্ভরণ বিশ্লেষণ হতে প্রাপ্ত অবশিষ্টের গড় বর্গসমষ্টি (MSE) ।

উপরিউক্ত ভেদাঙ্ক সূত্র হতে লক্ষ্য করার বিষয় হলো যে নিরূপকের ভেদাঙ্ককে k অংশে বিভক্ত করা হয়েছে এবং প্রতিটি অংশ এক একটি আইগেন মানের সাথে জড়িত। আরো লক্ষ্য করার বিষয় হলো কোনো λ_j, $[\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_k]$ বিশেষ করে ক্ষুদ্র-তরগুলি খুব ছোট হলে তার সাথে জড়িত ভেদাঙ্ক উপাদানটি বড় হবে এবং তা $v(\hat{\beta}_j)$ কে বর্ধিত করবে। Belsley et al (1980) উল্লেখ করেছেন যে কোনো কন্ডিশন নম্বর বা কন্ডিশন সূচি (condition number or condition index) 15-এর বেশি হলে এবং তার প্রাসঙ্গিক কোনো দুই বা ততোধিক নিরূপকের ভেদাঙ্ক উপাদানের সমানুপাত (variance proportion) 0.5-এর বড় হলে মাল্টিকোলিনিয়ারিটি সমস্যা প্রকট এবং ঐ নিরূপকগুলি যে চলকগুলির সাথে সম্পর্কিত সেগুলি বেশি সংশ্লেষিত। এই সমস্যা হতে রেহাই পাওয়ার জন্য সংশ্লেষিত চলকগুলি নির্ভরণ বিশ্লেষণ হতে বাদ দিতে হয়।

এ জাতীয় বিশ্লেষণ আদর্শায়িত চলক ব্যবহার না করেও করা যায়। সেক্ষেত্রে কোনো নির্ভরণ পরামানের নিরূপকের ভেদাঙ্ক (k + 1) অংশে বিভক্ত হয়। এই বিশ্লে-ষণের জন্য Computer program ব্যবহার করা যায় বলে বিশ্লেষণে কোনো জটিলতাও

৬—

নেই। এ জাতীয় বিশ্লেষণ উদাহরণ ৬.৩-এর উপাত্তের ক্ষেত্রে করা যাক। এই বিশ্লেষিত ফলাফল সারণি ৬.৭-এ উপস্থাপন করা হলো। এখানে লক্ষণীয় বিষয় হলো যে

সারণি ৬.৭ : মাল্টিকোলিনিয়ারিটি চিহ্নিতকরণ।

ক্রমিক সংখ্যা	আইগেন মান	কন্ডিশন সূচি	ধ্রুবক	ভেদাঙ্ক সমানুপাত x_2	x_1	x_3	x_4
1	4.9303	1.00	0.0011	0.0003	0.0004	0.0013	0.0004
2	0.0354	11.80	0.2815	0.0004	0.0039	0.5016	0.0221
3	0.0208	15.38	0.4471	8.0714	0.0771	0.3879	0.0301
4	0.0082	24.57	0.2001	0.0014	0.5228	0.1026	0.6154
5	0.0053	30.61	0.0702	0.9264	0.3958	0.0066	0.3320

কন্ডিশন সূচি 24.57 এর প্রাসঙ্গিক x_1 ও x_4 চলকদ্বয়ের সাথে সম্পর্কিত ভেদাঙ্ক সমানুপাত 0.5 এর বেশি। সুতরাং এই চলকদ্বয় মাল্টিকোলিনিয়ারিটি সমস্যা সৃষ্টি করেছে প্রকটভাবে। কাজেই x_1 ও x_4 চলকদ্বয়কে নির্ভরণ বিশ্লেষণ হতে বাদ দেয়া উচিত।

উপরে আলোচিত মাল্টিকোলিনিয়ারিটি বিষয়ক সমস্যা চিহ্নিত করার পদ্ধতিগুলির গাণিতিক ভিত্তির চেয়ে গবেষণার ভিত্তিই বেশি প্রকাশিত এবং এখানে উল্লেখিত সকল পদ্ধতিই Computer program-এর মাধ্যমে করা যায়।

৬.১০ অনুমান বিঘ্নিত হলে প্রতিকার (Remedial Measures When Assumptions are Violated)

পূর্ববর্তী অনুচ্ছেদে নির্ভরণ বিশ্লেষণ করার জন্য যে সমস্ত অনুমান করা হয় সেগুলি বহাল থাকে কিনা তা পর্যবেক্ষণ পদ্ধতি আলোচনা করা হয়েছে। এই অনুমানগুলি বিঘ্নিত হওয়ার অর্থ হলো

(ক) e_i-এর বিন্যাস পরিমিত বিন্যাস নয়,

(খ) $V(e_i) \neq \sigma^2$,

(গ) $Cov(e_i , e_j) \neq 0$,

এবং (ঘ) X_i-গুলি সংশ্লেষিত,

এর মধ্যে X_i গুলি সংশ্লেষিত হলে প্রতিকার পদ্ধতি পূর্ববর্তী অনুচ্ছেদের শেষে করা হয়েছে। আরো কিছু প্রতিকার পদ্ধতি আছে। সব কিছু বিস্তারিত আলোচনা করা

এখানে সম্ভব নয়। তবে আগ্রহী পাঠকগণ Johnston (1984), Neter and Wasserman (1974) Chatterjee and Price (1991), Draper and Smith (1981) ইত্যাদি গ্রন্থকারগণের বই পড়তে পারেন। এখানে অন্য তিনটি অনুমান বিঘ্নিত হলে কি প্রতিকার করা যায় তা সংক্ষেপে আলোচনা করা হলো।

(ক) e_i-এর পরিমিত বিন্যাস না হলে প্রতিকার : ধরা যাক প্রতিকৃতি হলো

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + e_i$$

এবং e_i-এর বিন্যাস পরিমিত বিন্যাস নয়। ধরা যাক e_i-এর বিন্যাস হলো n এবং P পরামানসহ দ্বিপদী বিন্যাস। তাহলে e_i-কে

$$\sin^{-1}\sqrt{\frac{e_i}{n}} \quad \text{বা} \quad \sin^{-1}\sqrt{\frac{e_i + 3/8}{n + 3/4}}$$

দ্বারা পরিবর্তন করা যায়। এর অর্থ হলো নির্ভরণ বিশ্লেষণ করার আগে y এবং x চলকগুলির প্রতিটিকে উপরিউক্ত সূত্রের সাহায্যে পরিবর্তন করতে হবে।

অনুরূপভাবে e_i-এর বিন্যাস পঁয়সো বিন্যাস হলে পরিবর্তন হবে $\sqrt{e_i}$ বা $\sqrt{e_i + 3/8}$। এছাড়া আরো কিছু পরিবর্তন সূত্র হলো।

(i) $\log_e e_i$ (ii) $\log_{10} e_i$, (iii) $\log_e (e_i + 1)$, (iv) $\log_{10} (e_i + 1)$ ইত্যাদি।

(খ) $V(e_i) \neq \sigma^2$ না হলে প্রতিকার: ধরা যাক $V(e_i) = \sigma_i^2$। আরো ধরা যাক $W_i = 1/\sigma_i$। এখন প্রতিকৃতি

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + e_i$$

এর উভয় পাশে W_i দ্বারা গুণ করে পাওয়া যায়

$$W_i y_i = \beta_0 W_i + \beta_1 x_{1i} W_i + \beta_2 x_{2i} W_i + \cdots + \beta_k x_{ki} W_i + W_i e_i$$

অথবা $$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \epsilon_i$$

এখানে $Y_i = W_i y_i$ এবং $X_{ji} = x_{ji} W_i$ $(j = 1, 2, \cdots, k)$, $W_i e_i = \epsilon_i$

$$V(\epsilon_i) = V(W_i e_i) = W_i^2 \sigma_i^2 = 1$$

অর্থাৎ বিচ্যুতি ϵ_i-এর ভেদাঙ্ক স্থির। সুতরাং পরিবর্তিত প্রতিকৃতির ক্ষেত্রে সাধারণ ন্যূনতম বর্গপদ্ধতি (OLS method) প্রয়োগ করা যায়। এই আদি চলকসমূহকে বিচ্যুতির পরিমিত ব্যবধান (σ_i)-এর ব্যতিহারী (reciprocal) দ্বারা গুণ করা হয়। এই পদ্ধতিকে বলা হয় ভার আরোপিত ন্যূনতম বর্গপদ্ধতি (Weighted Least Squares Method, WLS method)।

WLS পদ্ধতি প্রয়োগ করার ক্ষেত্রে একটি অসুবিধা হলো যে σ_i^2-এর মান জানা থাকে না। সেক্ষেত্রে σ_i^2-এর নিরূপক পূর্ববর্তী কোনো বিশ্লেষণ হতে বা একটি প্রাক বিশ্লেষণ হতে পেতে হয়। এখানে ভার (W_i) নিরূপিত হওয়ার কারণে β_j $[j=0, 1, 2, \cdots k]$ গুলির নিরূপক নির্ঝুঁকি হলেও তাদের দক্ষতা (efficiency) কম হয় এবং এ কারণে β_j সম্পর্কিত কোনো সিদ্ধান্ত ক্ষতিগ্রস্ত হয়। তবে σ_i^2-এর নিরূপক বড় আকারের স্বাধীনতার মাত্রাসহ (n large) পেলে সমস্যার সমাধান হয়।

অনেক সময়

$$V(e_i) \propto x_{ji} \text{ বা } V(e_i) \propto x_{ji}^2$$

হয়। এখানে $V(e_i) = \sigma^2 x_{ji}$ এবং $V(e_i) = \sigma^2 x_{ji}^2$। এখন প্রথম ক্ষেত্রে $W_i = 1/\sqrt{x_{ji}}$ এবং দ্বিতীয় ক্ষেত্রে $W_i = 1/x_{ji}$ $(j=1, 2, \cdots, k)$। এরূপ ক্ষেত্রে WLS প্রয়োগ করা হলে কোনো সমস্যার সৃষ্টি হয় না।

(গ) $\text{Cov}(e_i, e_j) \neq 0$ হলে প্রতিকার : ধরা যাক

$$\text{Cov}[e_t, e_{t+s}] = \rho^s \sigma^2 \,;\; t \neq S = 1, 2, \cdots, n$$

অথবা $$\text{Cov}[e_t \,.\, e_{t+1}] = \rho\sigma^2$$

এক্ষেত্রে $H_0: \rho = 0$ নাস্তি কল্পনা বাতিল হলেই সমস্যার সৃষ্টি হয়। ধরা যাক $\rho \neq 0$। এই ρ-এর নিরূপক হলো

$$r = \frac{\sum_{t=2}^{n} u_t u_{t-1}}{\sum_{t=2}^{n} u^2_{t-1}}$$

তাহলে, এই সমস্যার সমাধান করার জন্য চলকসমূহের নিম্নরূপ পরিবর্তন করতে হয় এবং পরিবর্তিত চলক নিয়ে সাধারণ ন্যূনতম বর্গপদ্ধতি (OLS method) প্রয়োগ করে নির্ভরণ বিশ্লেষণ করতে হয়,

$$y_t - ry_{t-1} \,;\; x_{1t} - rx_{1t-1} \,;\; \cdots \,;\; x_{kt} - r\,x_{kt-1}$$

এই পরিবর্তনের ফলে n সংখ্যক তথ্যমানের পরিবর্তে $(n-1)$ সংখ্যক তথ্যমান পাওয়া যায়। প্রথম তথ্যমানটি পাওয়া যায় না। সে কারণে প্রথম তথ্যমানটি পাওয়ার জন্য নিম্নরূপ পরিবর্তন করতে হয়

$$\sqrt{1-r^2}\,y_1 \,;\; \sqrt{1-r^2}\,x_{j1} \;\; (j=1, 2, \cdots, k)$$

উপরিউক্ত পরিবর্তনের ফলে বিভিন্ন চলকের নির্ভরণ সহগের নিরূপক ক্ষতিগ্রস্ত হয় না। কিন্তু প্রতিকৃতিতে ধ্রুবকের নিরূপক প্রভাবিত হয়। বিষয়টি দুই চলক-বিশিষ্ট সরলরৈখিক নির্ভরণ প্রতিকৃতির ক্ষেত্রে লক্ষ্য করা যাক।

মনে করি x ও y চলকদ্বয় প্রতিকৃতি

$$y_t = \beta_0 + \beta_1 x_t + e_t \qquad (৬.১০.১)$$

এর মাধ্যমে সম্পর্কিত এবং

$$\text{Cov}[e_t, e_{t+1}] = \rho\sigma^2 \text{ এবং } e_t = \rho e_{t-1} + \epsilon_t ; \; |\rho| < 1$$

পরিবর্তিত চলক $y_t^* = y_t - \rho y_{t-1}$ এবং $x_t^* = x_t - \rho x_{t-1}$। ধরা যাক

$$E[\epsilon_t, \epsilon_{t+s}] = 0 \text{ , যদি } S \neq 0$$

$$= \sigma^2_\epsilon \text{ , যদি } S = 0$$

রৈখিক প্রতিকৃতি হতে লেখা যায়

$$\rho y_{t-1} = \rho\beta_0 + \beta_1 \rho x_{t-1} + \rho e_{t-1} \qquad (৬.১০.২)$$

(৬.১০.১)−(৬.১০.২) হতে পাওয়া যায়

$$y_t - \rho y_{t-1} = \beta_0(1-\rho) + \beta_1(x_t - \rho x_{t-1}) + e_t - \rho e_{t-1}$$

$$y_t^* = \beta_0^* + \beta_1 x_t^* + \epsilon_t$$

এখানে $\beta_0^* = \beta_0(1-\rho)$

সুতরাং $\beta_0 = \beta_0^*/(1-\rho)$। কাজেই পরিবর্তিত চলক ব্যবহার করে OLS প্রয়োগ করে β_0^*-এর নিরূপক পাওয়া যাবে এবং β_0^*-এর নিরূপককে $(1-r)$ দ্বারা ভাগ করে β_0-এর নিরূপক পাওয়া যাবে।

৬.১১ নির্ভরণ বিশ্লেষণের জন্য চলক নির্বাচন (Selection of Variable for Regression Analysis)

নির্ভরণ বিশ্লেষণের বিভিন্ন উদ্দেশ্য আছে। যেমন: (১) নির্ভরণ রেখার দ্বারা নির্ভরশীল চলকের পূর্বাভাস করা, (২) নির্ভরণ রেখার মাধ্যমে নির্ভরশীল চলক ও অনপেক্ষ চলকসমূহের মধ্যে সংশ্লেষণের ধরন ব্যাখ্যা করা এবং (৩) নির্ভরশীল চলকের একটি নির্দিষ্ট মানের জন্য অনপেক্ষ চলকের মান নিয়ন্ত্রণ করা। প্রথম উদ্দেশ্য সফল করার জন্য প্রতিকৃতিতে অনপেক্ষ চলক এমনভাবে সংযুক্ত করা হয় যেন বিচ্যুতির গড় বর্গ-সমষ্টি (MSE) ন্যূনতম হয়। উদ্দেশ্য যাই হোক না কেন নির্ভরণ বিশ্লেষণের জন্য একগুচ্ছ অনপেক্ষ চলক নির্বাচন করতেই হয় এবং নির্বাচিত চলকগুচ্ছ সর্বোত্তম গুচ্ছ

হওয়া উচিত। অবশ্য এক উদ্দেশ্যের জন্য নির্বাচিত সর্বোত্তম গুচ্ছ অন্য উদ্দেশ্যের ক্ষেত্রে সর্বোত্তম নাও হতে পারে।

সাধারণত নির্ভরণ সমীকরণে অনপেক্ষ চলকের সংযুক্তির জন্য নিম্নলিখিত পদ্ধতিগুলি ব্যবহৃত হয়।

(ক) ফরওয়ার্ড নির্বাচন পদ্ধতি (Forward selection procedure)

(খ) ব্যাকওয়ার্ড দূরকরণ পদ্ধতি (Backward elimination procedure)

এবং (গ) ধাপে ধাপে নির্বাচন পদ্ধতি (Stepwise selection procedure)

উপরিউক্ত পদ্ধতিগুলির মধ্যে সবচেয়ে বেশি ব্যবহৃত পদ্ধতি হলো ধাপে ধাপে নির্বাচন পদ্ধতি। সবগুলি পদ্ধতিরই কিছু সুবিধা অসুবিধা আছে। এ সম্পর্কে বিস্তারিত জানার জন্য Drapper and Smith (1981)-এর বই আলোচনা করা যেতে পারে। এ পদ্ধতিগুলির প্রয়োগ করার জন্য বিভিন্ন Computer program আছে বিধায় বাস্তবে কোনো অসুবিধা নেই।

উপরিউক্ত তিনটি পদ্ধতি ছাড়াও অনপেক্ষ চলকের গুচ্ছ নির্বাচন করার জন্য সম্ভাব্য সকল নির্ভরণ সমীকরণের মিল করে কিছু বৈশিষ্ট্য চিহ্নিত করা হয়। এই বৈশিষ্ট্যগুলি হলো (ক) R_p^2, (খ) MSE_p, (গ) C_p। এখানে

$$R_p^2 = \frac{SSR_p}{SST}$$

এখানে $SSR_p = P$ পরামানবিশিষ্ট নির্ভরণ সমীকরণ মিল হতে প্রাপ্ত নির্ভরণ বর্গসমষ্টি।

SST = ঐ নির্ভরণ বিশ্লেষণের ক্ষেত্রে নির্ভরশীল চলকের মোট বর্গসমষ্টি $[\Sigma(y_i - \bar{y})^2]$। অবশ্য R_p^2 ব্যবহার করতে গিয়ে সমন্বিত (adjusted) R_{ap}^2 ব্যবহার করা হয়। এখানে

$$R_{ap}^2 = 1 - \frac{MSE_p\,(n-1)}{SST} = 1 - \frac{(n-1)(1-R_a{}^2)}{n-P}$$

সম্ভাব্য সকল নির্ভরণ রেখা মিল করতে গিয়ে P-এর কোনো মানের জন্য R_{ap}^2 সর্বোচ্চ হলে k অনপেক্ষ চলকই সর্বোত্তম অনপেক্ষ চলকের গুচ্ছ বিবেচিত হয়। এখানে $k = P - 1$

সম্ভাব্য সকল নির্ভরণ রেখা মিল করে বিশ্লেষণ করার পর বিচ্যুতির গড় বর্গসমষ্টি (MSE)-কে নির্ভরণের মূল্যায়নের জন্য ব্যবহার করা হয়। ধরা যাক P পরামানবিশিষ্ট নির্ভরণ রেখার ক্ষেত্রে বিচ্যুতির গড় বর্গসমষ্টি হলো MSE_p। এই MSE_p-এর মান যে সকল k চলকের জন্য ন্যূনতম হবে সেগুলিই সর্বোত্তম অনপেক্ষ চলকগুচ্ছ হিসেবে বিবেচিত হবে।

উপরে আলোচিত অনপেক্ষ চলকগুচ্ছ নির্বাচনের পদ্ধতির ক্ষেত্রে সম্ভাব্য সকল নির্ভরণের কথা উল্লেখ করা হয়েছে। কোনো নির্ভরণের জন্য k অনপেক্ষ চলক থাকলে মোট 2^k নির্ভরণ রেখা মিল করা যায়। এগুলির মধ্যে একটি রেখা হলো কোনো অনপেক্ষ চলকবিহীন। কমপক্ষে একটি অনপেক্ষ চলকসহ সম্ভাব্য নির্ভরণ রেখার সংখ্যা হলো $(2^k - 1)$। প্রতিটি ক্ষেত্রেই বিচ্যুতির বর্গসমষ্টি [SSE], বিচ্যুতির গড় বর্গসমষ্টি [MSE] এবং সমন্বিত সিদ্ধান্ত সহগ $[R_a^2]$ নির্ণয় করা হয়। এছাড়া সকল k চলক প্রতিকৃতিতে সংযুক্ত করেও নির্ভরণ রেখা মিল করা হয়। ধরা যাক সকল k চলক $[P = k + 1$ পরামানবিশিষ্ট] নিয়ে নির্ভরণ রেখা মিল করতে গিয়ে বিচ্যুতির ভেদাঙ্ক $[\sigma^2]$-এর নিরূপক হলো $\hat{\sigma}^2 [\hat{\sigma}^2 = SSE/n - P]$ এবং P-এর যে কোনো মান $[P = 1, 2, \cdots, k+1]$-এর জন্য নির্ভরণ বিশ্লেষণ হতে প্রাপ্ত বিচ্যুতির বর্গসমষ্টি [SSE] হলো SSE_p। উপরিউক্ত দুটি বিচ্যুতির বর্গসমষ্টি ব্যবহার করে চলকগুচ্ছ নির্বাচনের তথা নির্ভরণের মূল্যায়নের জন্য Mallows (1973) একটি তথ্যজমানের (statistic) প্রস্তাব করেছেন। এই তথ্যজমানটি হলো C_p, যেখানে

$$C_p = \frac{SSE_p}{\hat{\sigma}^2} + (2P - n)$$

এখানে n হলো নমুনা সংখ্যা। Mallows-এর মতে P পরামানবিশিষ্ট নির্ভরণ রেখায় কোনো ঝোঁক না থাকলে $C_p = P$ হয়। সুতরাং Cp ও p-এর ব্যবধান লক্ষ্য করে নির্ভরণ রেখার মূল্যায়ন করা যায় বা $Cp = P$ হলে $k = P - 1$ চলকগুচ্ছই নির্ভরণ বিশ্লেষণের জন্য নির্বাচিত চলকগুচ্ছ হিসেবে বিবেচিত হয়।

আগেই উল্লেখ করা হয়েছে যে MSE-এর মানের ভিত্তিতে চলকগুচ্ছ-এর নির্বাচন হতে পারে। তবে চলকগুচ্ছ চলকের সংখ্যা এক হলেও MSE সবগুচ্ছের জন্য সমান হয় না। ধরা যাক কোনো নির্ভরণের জন্য $x_1, x_2, \cdots, x_k$ হলো k সংখ্যক অনপেক্ষ চলক। তাহলে $P = 2$ হবে k সংখ্যক বার, $P = 3$ হবে $\binom{k}{2}$ বার, $P = 4$ হবে $\binom{k}{3}$ বার $\cdots$ $P = (k+1)$ হবে $\binom{k}{k}$ বা এক বার। ধরা যাক $P = 2$। তাহলে $P = 2$ বিশিষ্ট নির্ভরণ রেখার সংখ্যা হবে k। এখানে প্রতিটি নির্ভরণ রেখা মিল করে MSE পাওয়া যাবে। এ ব্যাপারটি P এর যে কোনো মানের জন্য সত্য। নির্ভরণ রেখার পরিবর্তনের সাথে সাথে MSE-এর মানও পরিবর্তিত হয়। কাজেই চলকগুচ্ছ নির্বাচন এমনভাবে করতে হয় যেন MSE-এর মান ন্যূনতম হয়, R_a^2-এর মান বেশি হয়

এবং $Cp = P$ হয়। তাছাড়া MSE যেহেতু বিচ্যুতির ভেদাঙ্ক (σ^2)-এর নিরূপক, সে কারণে এ নিরূপকটি যাতে স্থির (stable) হয় সেদিকেও লক্ষ্য রাখতে হয়। চিত্রের সাহায্যে $\hat{\sigma}^2$-এর স্থিরতা লক্ষ্য করা হয়। এ ক্ষেত্রে P-এর বিভিন্ন মানের জন্য MSE-এর গড় মান নির্ণয় করা হয় এবং X-অক্ষে P বসিয়ে ও Y-অক্ষে MSE-এর গড় মান বসিয়ে P-এর বিপরীতে বিন্দু স্থাপন করা হয়। বিন্দুগুলি মুক্ত হস্তে যোগ করা হলে একটি চিত্র পাওয়া যায়। লক্ষণীয় বিষয় হলো, P-এর মান বাড়ার সাথে সাথে গড় MSE কমতে থাকে এবং এক পর্যায়ে P-এর মান যতই বাড়ুক না কেন গড় MSE-এর মান আর ছোট হয় না। অর্থাৎ মুক্ত হস্তে যোগ করা রেখার গতি নিম্নমুখী হয়েও এক পর্যায়ে X-অক্ষের প্রায় সমান্তরালে চলতে থাকে। যে বিন্দু থেকে ঐ রেখা X-অক্ষের সমান্তরালে চলা শুরু করে সে বিন্দুর P-এর মানই স্থির MSE সরবরাহ করে থাকে।

চলক নির্বাচনের শেষোক্ত তিনটি পদ্ধতি উদাহরণ ৬.৩-এর উপাত্তের ক্ষেত্রে প্রয়োগ করা যাক। উক্ত উপাত্তের জন্য সম্ভাব্য সকল নির্ভরণ হতে প্রাপ্ত R_a^2, MSE C_p এবং গড় MSE সারণি ৬.৮-এ উপস্থাপন করা হলো। লক্ষ্য করা যাচ্ছে যে,

সারণি ৬.৮ : উদাহরণ ৬.২-এর উপাত্তের ক্ষেত্রে সম্ভাব্য সকল নির্ভরণের জন্য R_a^2, MSE, C_p এবং গড় MSE-এর মান।

চলকসমূহের সাবসক্রিপ্ট		পরামানের সংখ্যা P	R_a^2		MSE		Cp		MSE
1	2	2	0.584	0.710	1.624	1.134	104.4	58.9	
3	4	,,	0.703	0.524	1.161	1.861	61.5	126.3	1.445
12	13	3	0.713	0.774	1.122	0.883	57.7	36.1	
14	23	,,	0.644	0.855	1.391	0.564	82.1	7.1	0.954
24	34	,,	0.716	0.832	1.110	0.656	56.7	15.5	
123	124	4	0.853	0.716	0.574	1.109	9.0	56.4	0.712
134	234	,,	0.834	0.868	0.649	0.515	15.6	3.7	
1234	—	5	0.867	—	0.518	—	5.0	—	0.518

নির্ভরণে চলক x_2, x_3 এবং x_4 ব্যবহার করা হলে সমন্বিত সিদ্ধান্ত সহগ $[R_a^2]$ সবচেয়ে বড়, বিচ্যুতির গড় বর্গসমষ্টি [MSE] সবচেয়ে ছোট এবং C_p-এর মান পরামান P-এর সমান। সুতরাং উদাহরণ ৬.৩-এর উপাত্তের ক্ষেত্রে Slug-এর Body weight-এর ভেদ ব্যাখ্যা করার জন্য Mantle length (x_2), Keel length (x_3) এবং Shell length (x_4) সবচেয়ে ভাল চলকগুচ্ছ। ধাপে ধাপে চলক নির্বাচনের ফরওয়ার্ড নির্বাচন পদ্ধতিও এই চলকগুচ্ছকে সবচেয়ে ভাল চলকগুচ্ছ হিসেবে নির্বাচন করেছে।

ধাপে ধাপে চলকগুচ্ছ নির্বাচনের ফরওয়ার্ড নির্বাচন পদ্ধতির জন্য Computer program আছে। এখানে উদাহরণ ৬.৩-এর উপাত্তের ভিত্তিতে এই নির্বাচন পদ্ধতি ব্যাখ্যা করা যাক। আলোচিত নির্বাচন পদ্ধতি প্রয়োগের জন্য প্রথমেই নির্ভরশীল চলক (y) এবং অনপেক্ষ চলক x_1, x_2, x_3 ও x_4-এর সংশ্লেষাঙ্ক (r_i) নির্ণয় করতে হয়, এখানে

$$r_i = \frac{\sum_{j=1}^{n} x_{ij}y_j - \frac{1}{n}\sum_j x_{ij}\Sigma y_j}{\sqrt{\left[\sum_j x_{ij}^2 - \frac{(\Sigma x_{ij})^2}{n}\right]\left[\Sigma y_j^2 - \frac{(\Sigma y_j)^2}{n}\right]}}$$

$i = 1,2,3,4$

এই r এর তাৎপর্য যাচাই করার জন্য F যাচাই তথ্যজমান নির্ণয় করা হয়, যেখানে

$$F_i = \frac{(n-2)\,r_i^2}{(1-r_i^2)}$$

সাধারণত পূর্ব থেকে F-এর একটি মান নির্দিষ্ট করে দেয়া হয়। নির্ণেয় F_i-এর মান F-এর নির্দিষ্ট মানের চেয়ে বড় হলে i-তম চলক নির্ভরণ রেখায় সংযুক্ত হবে। F-এর নির্দিষ্ট মান বিবেচনা করার ক্ষেত্রে 5% সংশয়মাত্রা এবং k ও $(n-k-1)$ স্বাধীনতার মাত্রা বিবেচনা করা যেতে পারে। আলোচিত উপাত্তের ক্ষেত্রে চলক নির্বাচনের বিভিন্ন ধাপের মান সারণি ৬.৯-এ উপস্থাপন করা হলো। এই পদ্ধতি অনুযায়ী যে

চলকের জন্য r_i $(i=1,2,3,4)$-এর মান সব চেয়ে বড় হবে প্রথমে সে চলক নির্ভরণ রেখায় অন্তর্ভুক্ত হবে।

সারণি ৬.৯ : উদাহরণ ৬.৩-এর উপাত্তের ক্ষেত্রে অনপেক্ষ চলক নির্বাচনের জন্য ফরওয়ার্ড নির্বাচন পদ্ধতির বিভিন্ন ধাপের মান।

অনপেক্ষ চলক	r_i	F_i
x_1	0.7699	69.866
x_2	0.8460	120.863
x_3	0.8420	116.896
x_4	0.7303	54.865

প্রথম অন্তর্ভুক্ত চলক	নির্ভরাঙ্ক	F-এর মান	অন্তর্ভুক্ত নয় এমন চলক	সংশ্লেষাঙ্ক	F-এর মান
x_2	0.2959	120.863	x_1	0.1747	1.48
			x_3	0.7159	49.42
			x_4	0.2021	2.00
দ্বিতীয় অন্তর্ভুক্ত চলক	**নির্ভরাঙ্ক**	**F-এর মান**	**অন্তর্ভুক্ত নয় এমন চলক**	**সংশ্লেষাঙ্ক**	**F-এর মান**
x_3	0.2428	49.42	x_1	0.0598	0.16
			x_4	0.3271	5.51
তৃতীয় এবং শেষ ধাপে অন্তর্ভুক্ত চলকসমূহ	**নির্ভরাঙ্ক**	**F-এর মান**	**অন্তর্ভুক্ত নয় এমন চলক**	**সংশ্লেষাঙ্ক**	**F-এর মান**
x_2	0.1253	13.73	x_1	0.1222	0.68
x_3	0.2456	55.36			
x_4	0.1931	5.51			

সারণি ৬.৯-এ দেখা যাচ্ছে যে $r_2=0.846$ সবচেয়ে বড়। তাই x_2 প্রথমে নির্ভরণ প্রতিকৃতিতে সংযুক্ত। অর্থাৎ প্রথম ধাপে মিল করা নির্ভরণ রেখা হলো

$$\hat{y}_1 = a + b_2 x_2$$

এখানে $\hat{y}_1$ হলো প্রথম ধাপে নিরূপিত y-এর মান। এই রেখা হতে অবশিষ্ট-এর নিরূপক হলো

$$u_1 = y - \hat{y}_1$$

প্রথম ধাপে নির্ভরাঙ্ক (b_2) তাৎপর্যপূর্ণ হলে x_2 প্রতিকৃতিতে থাকবে এবং দ্বিতীয় ধাপে আরো একটি চলক সংযুক্ত করার জন্য প্রচেষ্টা চালাতে হয়। দ্বিতীয় ধাপে চলক সংযুক্ত করার জন্য u_1 এবং বাকি x সমূহের সংশ্লেষাঙ্ক নির্ণয় করতে হয় এবং যে x-এর জন্য এই সংশ্লেষাঙ্ক সবচেয়ে বড় তা প্রতিকৃতিতে সংযুক্ত হয়। এই সংশ্লেষাঙ্ক নির্ণয়ের সূত্র হলো

$$r_{ju_1} = \frac{\sum_{j}^{n} u_{ij} x_{ij} - \dfrac{\hat{j}\sum^{n} u_{ij} \sum^{n} x_{ij}}{n}}{\sqrt{\left[\Sigma u_{ij}^2 - \dfrac{(\Sigma u_{ij})^2}{n}\right]\left[\sum_{j} x_{ij}^2 - \dfrac{(\Sigma x_{ij})^2}{n}\right]}}$$

লক্ষ্য করা যাচ্ছে যে $r_{3u_1} = 0.7159$ সবচেয়ে বড়। তাই দ্বিতীয় ধাপে চলক x_3 প্রতিকৃতিতে অন্তর্ভুক্ত। এই পর্যায়ে মিল করা নির্ভরণ রেখা হলো

$$\hat{y}_2 = a + b_2x_2 + b_3x_3$$

এই রেখা হতে অবশিষ্টের নিরূপক হলো

$$u_2 = y - \hat{y}_2$$

এই পর্যায়ের পর অন্তর্ভুক্ত চলকের নির্ভরাঙ্ক তাৎপর্যপূর্ণ হলে তৃতীয় চলকের অন্তর্ভুক্তির প্রচেষ্টা নেয়া হয় একই পদ্ধতিতে। অর্থাৎ u_2 ও বাকি x সমূহের সংশ্লেষাঙ্ক নির্ণয় করা হয় এবং যে x-এর জন্য সংশ্লেষাঙ্ক $[r_{ju_2}]$ বড় হবে তা প্রতিকৃতিতে অন্তর্ভুক্ত হবে এবং ঐ x-এর নির্ভরাঙ্ক তাৎপর্যপূর্ণ কিনা তা পর্যবেক্ষণ করতে হবে। এভাবে চলক অন্তর্ভুক্ত হতে হতে কোনো x-এর জন্য নির্ভরাঙ্ক তাৎপর্যহীন হলে চলক নির্বাচনের কাজ শেষ হবে। লক্ষ্য করা যাচ্ছে যে দ্বিতীয় ধাপে অন্তর্ভুক্ত চলক x_3-এর নির্ভরাঙ্ক $b_3 = 0.2428$ এবং তা তাৎপর্যপূর্ণ $[F = 49.42]$। এ ধাপের পর $r_{4u_2} = 0.3271$ সবচেয়ে বড়। তাই x_4 প্রতিকৃতিতে অন্তর্ভুক্ত এবং এর নির্ভরাঙ্ক $(b_4 = 0.1931)$ তাৎপর্যপূর্ণ $[F = 5.51]$। x_1 ও u_3 এর সংশ্লেষাঙ্ক হলো 0.1222 এবং এই সংশ্লেষাঙ্ক যাচাই-এর তথ্যজমান $F = 0.68$ নির্দিষ্ট $F\ [F = 2]$ এর চেয়ে ছোট হওয়ায় চলক x_1 প্রতিকৃতিতে অন্তর্ভুক্ত হবে না। এখানে

$$u_3 = y - \hat{y}_3 \text{ এবং } \hat{y}_3 = a + b_2x_2 + b_3x_3 + b_4x_4$$

কাজেই x_2, x_3 এবং x_4 চলকগুচ্ছই নির্বাচিত চলকগুচ্ছ। এই গুচ্ছের ভিত্তিতে নির্ভরণ বিশ্লেষণের ফলাফল সারণি ৬.১০-এ উপস্থাপন করা হলো।

সারণি ৬.১০ : নির্ভরণ বিশ্লেষণের ফলাফল।

চলক	নির্ভরাঙ্ক-এর নিরূপক $[\hat{b}_i]$	পরিমিত বিচ্যুতি $s.e(\hat{b}_i)$	t-মান	P-মান
ধ্রুবক	− 5.875	0.621	9.45	0.000
x_2	0.125	0.033	3.73	0.000
x_3	0.246	0.033	7.44	0.000
x_4	0.193	0.082	2.35	0.023

ভেদাঙ্ক বিশ্লেষণ সারণি

ভেদের উৎস	স্বাধীনতার মাত্রা d.f	বর্গসমষ্টি SS	গড় বর্গসমষ্টি MS	নির্ণেয় F	P-মান
নির্ভরণ	3	167.754	55.918	108.59	0.000
বিচ্যুতি	46	23.687	0.5149		
মোট	49				

উপরিউক্ত নিয়মে চলকগুচ্ছ নির্বাচন করার পর নির্ভরণ বিশ্লেষণের জন্য প্রয়োজনীয় অনুমানগুলি বহাল আছে কি না তা পর্যবেক্ষণ করতে হয়। বিশেষ করে নির্বাচিত চলকগুচ্ছ মাল্টিকোলিনিয়ারিটি দ্বারা প্রভাবিত হয় কিনা তা পর্যবেক্ষণ করতে হয়। মাল্টিকোলিনিয়ারিটি সমস্যা চিহ্নিতকরণ এবং অন্যান্য অনুমানসমূহ বিঘ্নিত না হওয়ার শর্তসমূহ এই অধ্যায়ে ৬.১ অনুচ্ছেদে আলোচনা করা হয়েছে।

৬.১২ নির্ভরণ বিশ্লেষণের জন্য প্রতিকৃতি নির্বাচন (Selection of Model for Regression Analysis)

নির্ভরণ বিশ্লেষণের জন্য রৈখিক প্রতিকৃতি (linear model) মিল করার পদ্ধতি এতক্ষণ আলোচনা করা হয়েছে। কোনো উপাত্তের জন্য রৈখিক প্রতিকৃতি যথাযথ হবে কিনা সে সম্পর্কে সিদ্ধান্ত নেয়ার জন্য বিক্ষেপ চিত্রের ব্যবহারও আলোচনা করা হয়েছে। কিন্তু বাস্তবে সকল উপাত্তের জন্য রৈখিক প্রতিকৃতি যথাযথ হবে এমন কোনো কথা নেই। নির্ভরশীল চলকের ভেদ ব্যাখ্যা করার জন্য প্রতিকৃতি প্রধানত দু' ধরনের। যথা : (১) রৈখিক প্রতিকৃতি, এবং (২) অরৈখিক প্রতিকৃতি (non-linear model)। বর্তমান অনুচ্ছেদে রৈখিক প্রতিকৃতি সম্পর্কে সিদ্ধান্ত নেয়ার জন্য যাচাই পদ্ধতি আলোচনা করা হবে এবং অরৈখিক প্রতিকৃতি মিল করার পদ্ধতি আলোচনা করা হবে।

প্রতিকৃতির রৈখিকতা সম্পর্কে যাচাই (Test for linearity of model) : ধরা যাক x একটি অনপেক্ষ চলক এবং y একটি নির্ভরশীল চলক। x-এর j-তম স্তরের জন্য y-এর i-তম মান হলো y_{ij} $[i=1, 2, \cdots, n_j\ ;\ j=1, 2, \cdots q]$ উদাহরণ ৫.৩-এর উপাত্তের ক্ষেত্রে Slug-এর Body weight (y) এবং Body length (x)-এর উপাত্ত সারণি ৬.১১-এ উপস্থাপন করা হলো।

সারণি ৬.১১ : Slug-এর Body length (x) এবং Body weight (y)-এর উপাত্ত।

তথ্যমানের সংখ্যা, n_j \ x_j	32	35	36	40	41	43	44	45	46
1	0.60	2.50	1.30	2.39	1.50	2.50	1.80	1.90	1.90
2	1.50			1.70	2.30			1.88	
3				2.50	1.40			3.80	
4				4.80					
গড়, $\bar{y}_j$	1.05	2.50	1.30	2.85	1.73	2.50	1.80	2.53	1.90

n_j \ x_j	48	50	51	52	53	55	57	59	60	61	62
1	2.20	2.50	3.10	2.80	2.10	2.40	3.50	2.90	4.40	8.70	5.10
2		3.10				3.40	8.00		5.30		
3		5.70				3.90	3.30		3.10		
4		2.20				3.90					
5						3.00					
6						2.50					
গড়, $\bar{y}_j$	2.20	3.37	3.10	2.80	2.10	3.18	4.93	2.90	4.27	8.70	5.10

n_j \ x_j	65	67	68	74	75
1	5.50	6.90	5.00	6.80	7.70
2	4.40	5.10			7.00
3	4.50				
4	7.20				
গড়, $\bar{y}_j$	5.40	6.00	5.00	6.80	7.35

ধরা যাক নির্ভরশীল চলক (y) এবং অনপেক্ষ চলক (x)-এর মধ্যে একটি সম্পর্ক আছে। এখানে নাস্তি কল্পনা হলো

$$H_0 : y = \alpha + \beta x + e \text{ বা } E(y) = \alpha + \beta x$$

এবং বিকল্প কল্পনা হলো

$$H_A : E(y) \neq \alpha + \beta x$$

নাস্তি কল্পনা দ্বারা বুঝা যাচ্ছে যে x ও y-এর সম্পর্ক রৈখিক। এই রৈখিক সম্পর্কের ভিত্তিতে নির্ভরণ রেখা মিল করার জন্য সকল অনুমান বহাল আছে বিবেচনা করা যাক। তাহলে, বিশ্লেষিত ফলাফল হতে লেখা যায়

$$SST = \Sigma\Sigma(y_{ij} - \bar{y})^2 = \Sigma\Sigma y_{ij}^2 - \frac{(\Sigma\Sigma y_{ij})^2}{n}$$

$$SSR = b\, SP(xy)$$

$$SSE = SST - SSR$$

এখানে b হলো নির্ভরাঙ্ক β-এর নিরূপক।

$$b = \frac{SP(xy)}{SS(x)}, \quad SP(xy) = \sum_{i}^{n_j}\sum_{j}^{q} x_j y_{ij} - \frac{\Sigma n_j x_j \Sigma\Sigma y_{ij}}{n}$$

$$n = \Sigma n_j, \quad SS(x) = \Sigma n_j x_j^2 - \frac{(\Sigma n_j x_j)^2}{n}$$

আবার x-এর যে কোনো মানের জন্য y-এর গড় হতে ব্যবধানের মোট বর্গসমষ্টি হলো

$$\sum_{i=1}^{n_j}(y_{ij} - \bar{y}_j)^2$$

এবং x-এর সকল মানের জন্য এই বর্গসমষ্টি হলো

$$SPE = \Sigma\Sigma(y_{ij} - \bar{y}_j)^2$$

$$= \sum_{i}\sum_{j} y_{ij}^2 - \sum_{j}\frac{\left(\sum_{i} y_{ij}\right)^2}{n_j}$$

এই বর্গসমষ্টির স্বাধীনতার মাত্রা $(n-q)$। এই বর্গসমষ্টিই হলো নির্ভরণ রেখা মিল করার ক্ষেত্রে খাঁটি বিচ্যুতির বর্গসমষ্টি (pure error sum of squares)। কিন্তু বিচ্যুতির বর্গসমষ্টি হলো SSE। তাহলে নির্ভরণ রেখা মিল ভাল না হওয়ার কারণে বর্গসমষ্টি হলো (lack of fit sum of squares)

$$SLF = SSE - SPE$$

এই SLF-এর স্বাধীনতার মাত্রা হলো $(q-2)$। তাহলে নাস্তি কল্পনার জন্য যাচাই তথ্যজমান হলো

$$F = \frac{SLF/(q-2)}{SPE/(n-q)}$$

এই F-এর বিন্যাস হলো $(q-2)$ এবং $(n-q)$ স্বাধীনতার মাত্রাবিশিষ্ট ভেদাঙ্ক অনুপাত বিন্যাস।

$F \geqslant F_{0\cdot05\,;\,q-2,\,n-q}$ হলে 5% সংশয়মাত্রায় রৈখিক প্রতিকৃতির নাস্তি কল্পনা বাতিল বলে গণ্য হবে। অন্যথায় y ও x-এর সম্পর্ক রৈখিক হবে বলে সিদ্ধান্ত নেয়া যায়।

সারণি ৬.১১-এর উপাত্তের ক্ষেত্রে $q = 25$,

$n_1 = 2,\ n_2 = 1,\ n_3 = 1,\ n_4 = 4,\ n_5 = 3,\ n_6 = 1,\ n_7 = 1,\ n_8 = 3,\ n_9 = 1,$
$n_{10} = 1,\ n_{11} = 4,\ n_{12} = 1,\ n_{13} = 1,\ n_{14} = 1,\ n_{15} = 6,\ n_{16} = 3,\ n_{17} = 1$
$n_{18} = 3,\ n_{19} = 1,\ n_{20} = 1,\ n_{21} = 4,\ n_{22} = 2,\ n_{23} = 1,\ n_{24} = 1,\ n_{25} = 2;$
$n = 50$;

$$SP(xy) = 10541.9 - \frac{2639 \times 183.47}{50} = 831.3534$$

$$SS(x) = 145377 - \frac{(2639)^2}{50} = 6090.58$$

$$b = \frac{831.3534}{6090.58} = 0.1365$$

$$SSR = 0.1365 \times 831.3534 = 113.4797$$

$$SST = 864.6656 - \frac{(183.47)^2}{50} = 191.4408$$

$$SSE = SST - SSR = 77.9611$$

$$SPE = 864.6656 - 822.5693 = 42.0963$$

$$SLF = SSE - SPE = 35.8648$$

$$\therefore \qquad F = \frac{SLF/(q-2)}{SPE/(n-q)} = \frac{35.8648/(25-2)}{42.0963/(50-25)} = 0.93$$

$F_{0.05;\, 23, 25} = 2.225 > F$ । কাজেই নাস্তি কল্পনা

$H_0 :\ E(y) = \alpha + \beta x$-এর বিপক্ষে কোনো যুক্তি নেই। অর্থাৎ আলোচিত উদাহরণের ক্ষেত্রে রৈখিক প্রতিকৃতি যথাযথ।

উদাহরণ ৬.৪

একজন গবেষণাকারী একপ্রকার ক্ষুদ্র প্রাণীর ডিম থেকে বাচ্চা হওয়ার পর প্রতি দুই সপ্তাহ অন্তর অন্তর কতটি বাচ্চা জীবিত থাকে তা নথিভুক্ত করেন। পরীক্ষাটি 12 সপ্তাহ যাবৎ তিন গ্রুপ বাচ্চার ক্ষেত্রে পরিচালিত হয়। নিচে সারণি ৬.১২-এ 2 সপ্তাহ পর পর জীবিত বাচ্চাদের সংখ্যা দেয়া হলো।

সারণি ৬.১২ : দুই সপ্তাহ পরপর জীবিত বাচ্চার সংখ্যা।

জীবিত	সপ্তাহ সংখ্যা (x_j)						তথ্যমানের সংখ্যা
বাচ্চার সংখ্যা	2	4	6	8	10	12	n_j
	500	150	75	20	4	2	1
	541	116	58	27	6	3	2
	550	135	65	25	5	4	3
গড়, $\bar{y}_j$	530.33	133.67	66.00	24.00	5.00	3.00	

সময় (x)-এর সাথে জীবিত প্রাণীর সংখ্যা (y) সম্পর্কে পর্যালোচনা করা গবেষণাকারীর উদ্দেশ্য।

ধরা যাক x ও y-এর সম্পর্ক রৈখিক। অর্থাৎ

$$y = \alpha + \beta x + e$$

তাহলে নাস্তি কল্পনা যাচাই করতে হবে

$$H_0 :\ E(y) = \alpha + \beta x$$

এবং বিকল্প কল্পনা হলো

$$H_A :\ E(y) \neq \alpha + \beta x$$

ধরা যাক উপরিউক্ত নির্ভরণ রেখা মিল করার জন্য প্রতিকৃতির রৈখিকতা ভিন্ন অন্যান্য অনুমান বহাল আছে। তাহলে

$$SST = \Sigma\Sigma y_{ij}^2 - \frac{(\Sigma\Sigma y_{ij})^2}{n}$$

$$= 914436 - \frac{(2286)^2}{18} = 624114$$

এখানে $q = 6$; $j = 1, 2, 3, 4, 5, 6$

j-এর যে কোনো মানের জন্য $n_j = 3$; $n = 18$ ।

$$SS(x) = \Sigma n_j x_j^2 - \frac{(\Sigma n_j x_j)^2}{n} = 1092 - \frac{(126)^2}{18} = 210$$

$$Sp(xy) = \sum_i \sum_j x_j y_{ij} - \frac{\Sigma n_j x_j \; \Sigma\Sigma y_{ij}}{n} = 6808 - \frac{126 \times 2286}{18}$$

$$= -9144$$

$$b = \frac{Sp(xy)}{SS(x)} = -43.7805, \; SSR = bSP(xy) = 402522.0762$$

$$SSE = SST - SSR = 221591.9238$$

$$SPE = \Sigma\Sigma y_{ij}^2 - \sum_j \frac{(\Sigma y_{ij})^2}{n_j} = 914436 - 864733.3333$$

$$= 49702.6667$$

$$SLF = SSE - SPE = 171889.2571$$

$$\therefore \quad F = \frac{SLF/(q-2)}{SPE/(n-q)} = 10.38 \; ; \; F_{.05;4,12} = 3.26$$

সুতরাং নাস্তিকল্পনা বাতিল। অর্থাৎ উপরিউক্ত উদাহরণের ক্ষেত্রে রৈখিক প্রতিকৃতি সঠিক নয়।

বহুল নির্ভরণের ক্ষেত্রে একাধিক অনপেক্ষ চলক থাকে বিধায় প্রতিকৃতির রৈখিকতা সম্পর্কে ধারণা ব্যক্ত করার জন্য প্রতিটি অনপেক্ষ চলকের সাথে নির্ভরশীল চলকের রৈখিকতা সম্পর্কে অনুরূপভাবে F-যাচাই তথ্যজমানের মাধ্যমে সিদ্ধান্ত নিতে হয়। অধিকাংশ অনপেক্ষ চলক নির্ভরশীল চলকের সাথে রৈখিকভাবে সম্পর্কিত হলে ঐ সমস্ত অনপেক্ষ চলক ব্যবহার করে রৈখিক প্রতিকৃতি মিল করা যেতে পারে।

অরৈখিক প্রতিকৃতি (non-linear model) : কোনো উপাত্তের জন্য রৈখিক প্রতিকৃতি সঠিক না হলে অরৈখিক প্রতিকৃতি মিল করার বিষয় চিন্তা করতে হয়। এক বা একাধিক অনপেক্ষ চলকের ক্ষেত্রে নিম্নলিখিত অরৈখিক প্রতিকৃতিসমূহ মিল করা যেতে পারে।

(i) $y = \alpha x^{\beta} \in$

(ii) $y = \alpha \beta^{x} \in$

(iii) $y = \alpha e^{\beta x} \in$

(iv) $y = e^{\alpha + \beta x + \in}$

(v) $y = \dfrac{1}{\alpha + \beta x + \in}$

(vi) $y = \alpha + \beta x + \gamma x^2 + \in$

(vii) $y = \beta_0 x_1^{\beta_1} x_2^{\beta_2} \cdots x_k^{\beta_k} \in$

(viii) $y = e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \in}$

(ix) $y = \dfrac{1}{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \in}$

(x) $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_p x^p + \in$

ইত্যাদি।

উপরিউক্ত প্রতিকৃতিসমূহের মধ্যে প্রতিকৃতি (vi) এবং (x) ভিন্ন অন্য সব প্রতিকৃতিকে কোনো পরিবর্তনের মাধ্যমে রৈখিক প্রতিকৃতিতে পরিবর্তন করা যায়। যেমন

$$y = \alpha x^{\beta} \in$$

প্রতিকৃতিকে লগারিদমিক পরিবর্তনের মাধ্যমে লেখা যায়

$$\log y = \log \alpha + \beta \log x + \log \in$$

অথবা $$Y = A + \beta X + e$$

এখানে $$Y = \log y, \quad X = \log x, \quad e = \log \in$$

এই শেষোক্ত প্রতিকৃতিটি রৈখিক। এখন রৈখিক প্রতিকৃতি মিল করার জন্য কোনো অনুমান বিঘ্নিত না হলে ন্যূনতম বর্গপদ্ধতি প্রয়োগ করে A ও B এর নিরূপক পাওয়া যায়। এই নিরূপকগুলি হলো

$$\hat{A} = \bar{Y} - \hat{\beta}\bar{X}, \quad \hat{\beta} = \frac{SP(XY)}{SS(X)}$$

এবং $\hat{\alpha} = \text{Anti log } \hat{A}$;

অনুরূপভাবে প্রতিকৃতি (ii), (iii), (iv) এবং (v)-কে যথাক্রমে নিম্নরূপভাবে রৈখিক প্রতিকৃতিতে রূপান্তর করা যায়।

$$\log y = \log \alpha + x \log \beta + \log \in$$

$$ln\ y = ln\ \alpha + \beta x + ln \in$$

$$ln\ y = \alpha + \beta x + \in$$

এবং $1/y = \alpha + \beta x + \in$

এখন রৈখিক প্রতিকৃতি মিল করার নিয়ম অনুযায়ী উপরিউক্ত চারটি প্রতিকৃতি মিল করা যায়। উদাহরণ ৬.৪-এর উপাত্তের ভিত্তিতে এই প্রতিকৃতিগুলি মিল করা যাক।

প্রথমে বিবেচনা করা যাক প্রতিকৃতি (i)

$$\log y = \log\alpha + \beta \log x + \log \in$$

অথবা $Y = A + \beta X + e$

এখানে $Y = \log y, \quad A = \log\alpha, \quad X = \log x, \quad e = \log\in$

তাহলে $$\hat{\beta} = \frac{SP(XY)}{SS(X)} = \frac{17.83659 - \dfrac{13.99054 \times 27.58454}{18}}{12.11645 - \dfrac{(13.99054)^2}{18}}$$

$$= -2.90076$$

$$SST = 53.54528 - \frac{(27.58454)^2}{18} 11.27267$$

$$SSR = \hat{\beta} SP(XY) = 10.45305, \quad SSE = SST - SSR = 0.81962$$

$$R^2 = \frac{SSR}{SST} = 0.9273, \quad F = \frac{SSR/1}{SSE/(n-2)} = 204.06$$

$P = 0.00$। অর্থাৎ মিল করা নির্ভরণ রেখা তাৎপর্যপূর্ণ।

এখানে $\hat{A} = \bar{Y} - \hat{\beta}\bar{X} = 1.53247 + 2.90076 \times 0.77725$

$= 3.78708$

এবং $\hat{\alpha} = \text{Anti log } \hat{A} = 6124.71254$

সুতরাং মিল করা রেখা হলো

$$y = 6124.71254\ x^{-2.90076}$$

প্রতিকৃতি (ii) হতে লেখা যায়

$$\log\ y = \log\alpha + x\log\beta + \log\epsilon$$

$$Y = A + Bx + e$$

$$Y = \log y.\ B = \log\beta,\ A = \log\alpha,\quad e = \log\epsilon$$

$$\hat{B} = \frac{SP(XY)}{SS(x)} = \frac{144.93402 - \frac{126\times27.58454}{18}}{1092 - \frac{(126)^2}{18}} = \frac{-48.15775}{210}$$

$$= -0.22932$$

$SST = 11.27267 \quad SSR = bSP(xY) = 11.04366$

$SSE = 0.22901, \quad F = 771.57, \quad R^2 = 0.9797$

$P = 0.00$, অর্থাৎ মিল করা নির্ভরণ রেখা তাৎপর্যপূর্ণ।

এখানে $\hat{A} = \bar{Y} - \hat{B}\bar{X} = 1.53247 + 0.22901 \times 7 = 3.1355$

$\hat{\alpha} = \text{Anti log } 3.1355 = 1366.2809$

$\hat{\beta} = \text{Anti log } -0.22932 = 0.5898$

সুতরাং মিল করা নির্ভরণ রেখা হলো

$$y = 1366.2809 \times (0.5898)^x$$

প্রতিকৃতি (iii) হতে লেখা যায়

$$ln\ y = ln\alpha + \beta x + ln\epsilon$$

$$Y = A + \beta x + e$$

$$Y = ln\ y,\ A = ln\alpha,\quad e = ln\epsilon$$

$$\hat{\beta} = \frac{sp(xy)}{ss(x)} = \frac{333.72292 - \frac{126\times63.51575}{18}}{1092 - \frac{(126)^2}{18}}$$

$$= -0.52803$$

$$SST = 283.8916 - \frac{(63.51575)^2}{18} = 59.7665$$

$SSR = 58.55183$, $SSE = 1.21467$, $R^2 = 0.9797$, $F = 771.26$, $P = 0.00$

অর্থাৎ মিল করা নির্ভরণ রেখা তাৎপর্যপূর্ণ। এখানে

$$\hat{A} = \bar{Y} - \hat{\beta}\bar{x} = 3.52867 + 0.52803 \times 7 = 7.22488$$

$$\hat{\alpha} = \text{Anti } ln\hat{A} = 1373.17382$$

সুতরাং মিল করা নির্ভরণ রেখা হলো

$$y = 1373.17382\ e^{-0.52803x}$$

প্রতিকৃতি (iv) হতে লেখা যায়

$$ln y = \alpha + \beta x + \in$$

$$Y = \alpha + \beta x + \in, \ Y = ln\ y$$

$$\therefore \quad \hat{\beta} = \frac{sp(xy)}{ss(x)} = -0.52803, \ \hat{\alpha} = \bar{Y} - \hat{\beta}\bar{x} = 7.22488$$

$$SST = 59.7665, \ SSR = 58.55183, \ SSE = 1.21467$$

$$R^2 = 0.9797, \ F = 771.26, \ P = 0.00$$

সুতরাং, মিল করা নির্ভরণ রেখা হলো

$$y = e^{7.22488 - 0.52803x}$$

এবং এই রেখা তাৎপর্যপূর্ণ।

প্রতিকৃতি (v) হতে লেখা যায়

$$1/y = \alpha + \beta x + \in$$

অথবা $$Y = \alpha + \beta x + \in, \ Y = 1/y$$

$$\therefore \quad \hat{\beta} = \frac{sp(xY)}{ss(x)} = \frac{20.56083 - \dfrac{126 \times 1.90136}{18}}{210} = \frac{7.25131}{210}$$

$$= 0.03453$$

$$SST = 0.56026 - \frac{(1.90136)^2}{18} = 0.35941$$

$$SSR = 0.25038,\quad SSE = 0.10903,\quad R^2 = 0.6966$$
$$F = 36.74,\quad P = 0.00$$
$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{x} = 0.10563 - 0.03453 \times 7 = -0.13608$$

সুতরাং মিল করা নির্ভরণ রেখা হলো

$$y = \frac{1}{-0.13608 + 0.03453x}$$

এই নির্ভরণ রেখাও তাৎপর্যপূর্ণ।

প্রতিকৃতি (vi) হতে লেখা যায়

$$\begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \\ \hat{\gamma} \end{bmatrix} = \begin{bmatrix} n & \Sigma x & \Sigma x^2 \\ \Sigma x & \Sigma x^2 & \Sigma x^3 \\ \Sigma x^2 & \Sigma x^3 & \Sigma x^4 \end{bmatrix}^{-1} \begin{bmatrix} \Sigma y \\ \Sigma xy \\ \Sigma x^2 y \end{bmatrix}$$

$$= \begin{bmatrix} 18 & 126 & 1092 \\ 126 & 1092 & 10584 \\ 1092 & 10584 & 109200 \end{bmatrix}^{-1} \begin{bmatrix} 2286 \\ 6808 \\ 27312 \end{bmatrix}$$

$$= \begin{bmatrix} 1.066667 & -0.325000 & 0.020833 \\ -0.325000 & 0.114137 & -0.007812 \\ 0.020833 & -0.007812 & 0.000558 \end{bmatrix} \begin{bmatrix} 2286 \\ 6808 \\ 27312 \end{bmatrix}$$

$$= \begin{bmatrix} 794.7916 \\ -179.2666 \\ 9.6802 \end{bmatrix}$$

$$SSR = [\hat{\alpha}\ \hat{\beta}\ \hat{\gamma}] \begin{bmatrix} \Sigma y \\ \Sigma xy \\ \Sigma x^2 y \end{bmatrix} - \frac{1}{n}(\Sigma y)^2$$

$$=[794.7916 \;\; -179.2666 \;\; 9.6802]\begin{bmatrix}2286\\6808\\27312\end{bmatrix}-\frac{(2286)^2}{18}$$

$$=570510.2066$$

$$SST=914436-\frac{(2286)^2}{18}=624114$$

$SSE=53603.7934, \quad R^2=0.9141$

$F=79.82,\ P=0.00$। এখানে F-এর স্বাধীনতার মাত্রা হলো 2 এবং 15 সুতরাং মিল করা নির্ভরণ রেখা হলো

$$y=794.7916-179.2666\,x+9.6802x^2$$

এবং এই রেখা তাৎপর্যপূর্ণ।

আবার, উপরিউক্ত উপাত্তের জন্য সরলরৈখিক প্রতিকৃতি মিল করা হলে [উদাহরণ ৬.৪] $R^2=0.6449$। আগেই উল্লেখ করা হয়েছে যে উক্ত উপাত্তের জন্য রৈখিক প্রতিকৃতি সঠিক নয়। এটি R^2-এর মান হতেও বুঝা যায়। প্রতিকৃতি (ii), (iii) এবং (iv) মিল করার ফলে $R^2=0.9797$ পাওয়া গেছে। কাজেই আলোচিত উপাত্তের এই তিনটি প্রতিকৃতির যে কোনো একটি মিল করা যুক্তিসঙ্গত।

এখানে নির্ভরণ প্রতিকৃতি নির্বাচন করার জন্য রৈখিকতার যাচাই এবং সিদ্ধান্ত সহগ (R^2)-এর ব্যবহার উল্লেখ করা হয়েছে। সরল নির্ভরণের ক্ষেত্রে বিক্ষেপ চিত্রই প্রতিকৃতির ধরন সম্পর্কে ধারণা ব্যক্ত করতে সহায়ক। বিক্ষেপ চিত্র কোনো সরলরেখার আকার ধারণ করে কিনা তা লক্ষ্য করার বিষয়। সরলরৈখিক নির্ভরণের ক্ষেত্রে x চলকের মান বাড়ার সাথে সাথে y-চলকের মান হয় বাড়বে বা কমবে। অর্থাৎ y-এর পরিবর্তনের একটি নির্দিষ্ট গতি থাকবে। বিক্ষেপ চিত্রের বিক্ষিপ্ত বিন্দুগুলির মধ্য দিয়ে কোনো সরলরেখা আঁকা হলে ঐ রেখার উভয় পাশে মোটামুটি সমান সংখ্যক বিন্দু থাকবে। কোনো উপাত্তের জন্য নির্ভরণ রেখা বক্ররেখা হলে x-এর পরিবর্তনের সাথে y-এর মানের পরিবর্তনের কোনো নির্দিষ্ট গতিধারা থাকবে না। ঐরূপ ক্ষেত্রে নির্ভরণ বিশ্লেষণ করার জন্য x চলক বা y চলক বা উভয় চলকের জন্য কতকগুলো রূপান্তর (transformation) করা হয়ে থাকে। এখানে log রূপান্তর এবং বিপরীত (reciprocal) রূপান্তর প্রয়োগ করে নির্ভরণ রেখা মিল করা হয়েছে। এদুটো রূপান্তর ছাড়া বর্গমূল রূপান্তর (square root transformation) এবং উপাত্ত অনুপাত নির্দেশ করলে arcsine রূপান্তর করা যেতে পারে।

৬.১৩ দৈব অনপেক্ষ চলকের ক্ষেত্রে সরলরৈখিক নির্ভরণ (Simple Linear Regression in Case of Random Independent Variable)

রৈখিক নির্ভরণের ক্ষেত্রে অনুমান করা হয়েছে যে অনপেক্ষ চলক (x) অদৈব চলক বা স্থির বা চলকের মান গবেষণাকারী কর্তৃক পূর্ব নির্ধারিত। যথা, সারণি ৬.১১-এর উপাত্ত বিবেচনা করা যাক। উক্ত উপাত্তে Body length (x) গবেষণাকারী কর্তৃক পূর্ব নির্ধারিত হলে তা অদৈব চলক হবে। কিন্তু x এর প্রতিটি মানের জন্য প্রাপ্ত Body weight (y) অদৈব চলক। এক্ষেত্রে Slug-এর Body length (x)-এর বিপরীতে Mantle length বা shell length বা keel length পর্যালোচনা করা হলে উক্ত চলকগুলি দৈব চলক হবে এবং Body weight (y)-কে নির্ভরশীল চলক বিবেচনা করে Shell length-এর উপর নির্ভরণ রেখা মিল করা হলে Shell length দৈব চলক হবে। এরূপ ক্ষেত্রে নির্ভরাঙ্ক নিরূপণ করার জন্য ন্যূনতম বর্গপদ্ধতি যথাযথ নয়।

ধরা যাক দৈব অনপেক্ষ চলকের ক্ষেত্রে নির্ভরাঙ্ক-এর নিরূপক হলো b´ এবং ইন্টারসেপ্ট-এর নিরূপক হলো a´। তাহলে

$$b' = \pm \frac{s_y}{s_x} \quad \text{এবং} \quad a' = \bar{y} - b'\bar{x}$$

এখানে $$s_y^2 = \frac{1}{n-1} \Sigma(y - \bar{y})^2 \quad \text{এবং} \quad s_x^2 = \frac{1}{n-1} \Sigma(x - \bar{x})^2$$

b´-এর চিহ্ন +ve বা −ve হতে পারে। চিহ্নের ব্যাপারে সিদ্ধান্ত নিতে হবে

$$sp(xy) = \Sigma(x - \bar{x})(y - \bar{y})$$

এর চিহ্নের ভিত্তিতে।

উপরিউক্ত নির্ভরণের ক্ষেত্রে x ও y চলক দৈব চলক হওয়ায় নির্ভরশীল চলকের কোনো সঠিক সংজ্ঞায়ন করা যায় না। উপরন্তু এই নির্ভরণ বিশ্লেষণের ক্ষেত্রে নির্ভরাঙ্ক এর তাৎপর্য যাচাই করার জন্য 't'-যাচাই তথ্যজমান প্রয়োগ করা হয় না। কেননা উক্ত যাচাই-এর জন্য নাস্তি কল্পনা হলো $H_0 : \beta = 0$, যেখানে β হলো গণসমষ্টি নির্ভরাঙ্ক এবং $\beta = 0$ হওয়ার অর্থ হলো y চলকের পরিমিত ব্যবধান [σ_y] শূন্য যা সম্ভব নয়। এক্ষেত্রে ভেদাঙ্ক বিশ্লেষণের মাধ্যমে F যাচাই তথ্যজমান ব্যবহার করে নির্ভরণের তাৎপর্য যাচাই করা হয়, এখানে

$$F = \frac{SSR/1}{SSE/(n-2)}$$

$$SSR = b' \; SP(xy), \; SST = \Sigma(y-\bar{y})^2, SSE = SST - SSR$$

এ সম্পর্কে আরো বিস্তারিত জানার জন্য Sokal and Rohlf (1981) আলোচনা করা যেতে পারে।

উদাহরণ ৬.৫

একটি মাছ ধরা নৌকা হতে বিশেষ ওজনের দশটি মাছ চয়ন করে ঐগুলির Length (y mm) এবং Otolith length (x mm) নথিভুক্ত করা হয়েছে। উক্ত উপাত্ত হতে x ও y এর সম্পর্ক পর্যালোচনা করা যাক।

x :	5.5	7.6	6.2	10.2	8.5	7.7	6.5	9.7	9.5	8.8
y :	115	119	118	128	123	121	120	123	124	124

ধরা যাক x ও y রৈখিকভাবে সম্পর্কিত এবং এই সম্পর্কের জন্য প্রতিকৃতি হলো

$$y = \alpha + \beta x + e$$

তাহলে β এবং α-এর নিরূপক হলো যথাক্রমে

$$b' = \pm \frac{s_y}{s_x} \text{ এবং } a' = \bar{y} - b'\bar{x}$$

এখানে $$SP(xy) = \Sigma xy - \frac{\Sigma x \; \Sigma y}{n}, \; n = 10$$

$$= 9793.6 - \frac{80.2 \times 1215}{10} = 49.3$$

$$SST = 147745 - \frac{(1215)^2}{10} = 122.5$$

$$s_y = \sqrt{\frac{1}{n-1}\left[\Sigma y^2 - \frac{(\Sigma y)^2}{n}\right]} = \sqrt{\frac{122.5}{9}} = 3.6893$$

$$s_x = \sqrt{\frac{1}{n-1}\left[\Sigma x^2 - \frac{(\Sigma x)^2}{n}\right]} = \sqrt{\frac{1}{9}\left[660.06 - \frac{(80.2)^2}{10}\right]} = 1.5936$$

এখানে $SP(xy) + Ve$ হওয়ায়

$$b' = \frac{s_y}{s_x} = \frac{3.6893}{1.5936} = 2.3151$$

এবং $$a' = 121.5 - 2.3151 \times 8.02 = 102.9329$$

সুতরাং x এর উপর y এর নির্ভরণ রেখা হলো

$$y = 102.9329 + 2.3151\,x$$

এই বিশ্লেষণ হতে আরো পাওয়া যায়

$$SSR = b'SP(xy) = 114.1344,\ \ SSE = SST - SSR = 8.3656$$

$$F = \frac{SSR/1}{SSE/(n-2)} = 109.15;\ F_{.05,1,8} = 5.32$$

এখানে $F > F_{.05}$ হওয়াতে নির্ভরণ তাৎপর্যপূর্ণ। $R^2 = 0.9317$

৬.১৪ গুণগত চলকের ক্ষেত্রে নির্ভরণ বিশ্লেষণ (Regression Analysis in Case of Qualitative Variables)

এতক্ষণ নির্ভরণ বিশ্লেষণ আলোচনা করতে যে চলকসমূহের উল্লেখ করা হয়েছে সেগুলি সবই পরিমাণগত (quantitative) চলক। বাস্তবে নির্ভরশীল ও অনপেক্ষ চলক উভয়ই গুণগত (qualitative) চলক হতে পারে। যেমনঃ জন্মকালীন বাচ্চার ওজন গর্ভকালীন সময়, মায়ের স্বাস্থ্য, মায়ের গর্ভকালীন শুশ্রূষা ইত্যাদির উপর নির্ভরশীল বিবেচনা করা হলে বাচ্চার ওজন এবং গর্ভকালীন সময় পরিমাপ করে সংখ্যার দ্বারা প্রকাশ করা যায়। কিন্তু মায়ের স্বাস্থ্য বা শুশ্রূষা পরিমাপ করে সংখ্যার দ্বারা প্রকাশ করা যায় না। এক্ষেত্রে মায়ের স্বাস্থ্য খারাপ, মোটামুটি এবং ভাল প্রকাশ করার জন্য যথাক্রমে সংখ্যা 0, 1 এবং 2 দ্বারা পরিমাপ করা হলেও এই পরিমাপের কোনো পরিমাণগত অর্থ হয় না। এখানে 0, 1 এবং 2 দ্বারা স্বাস্থ্যের তিনটি অবস্থাকে বুঝানো হয়েছে। অনুরূপভাবে মায়ের শুশ্রূষা খারাপ, মধ্যম এবং উত্তম হয়েছে বিবেচনা করে এই তিনটি অবস্থাকে যথাক্রমে 0,1 এবং 2 দ্বারা প্রকাশ করা যায়। আলোচিত চলকদ্বয়কে নির্ভরণ বিশ্লেষণে সংযুক্ত করার জন্য এরূপ সংখ্যার মাধ্যমে প্রকাশ করা হয়। এ জাতীয় চলককে নির্দেশকারী (indicator) চলক বা নকল (dummy) চলক বলা হয়।

উপরে আলোচিত নকল চলকদ্বয় অনপেক্ষ চলক। নির্ভরশীল চলকের ক্ষেত্রেও নকল চলকের ব্যবহার লক্ষ্য করা যায়। ধরা যাক কৃষি জমির পোকামাকড় মারার জন্য পেস্টিসাইড ব্যবহার করা হয়েছে বিভিন্ন জমিতে। কোনো জমির পোকামাকড় মরেছে, কোনো জমির পোকামাকড় সম্পূর্ণ ধ্বংস হয়নি বিবেচনা করা হলে এবং পোকামাকড় ধ্বংস হওয়া পেস্টিসাইডের পরিমাণের উপর নির্ভরশীল বিবেচনা করা হলে এই নির্ভরশীল চলকের ক্ষেত্রে নকল চলকের ব্যবহার করতে হবে। বর্তমান অনুচ্ছেদে অনপেক্ষ চলক এবং নির্ভরশীল চলক নকল হলে কিভাবে নির্ভরণ বিশ্লেষণ করা হয় তা আলোচনা করা হবে।

নকল অনপেক্ষ চলক (Dummy independent variable) : নকল অনপেক্ষ চলকের উদাহরণ আগেই উল্লেখ করা হয়েছে। এই চলক নির্ভরণ বিশ্লেষণে সহজেই ব্যবহার করা যায়। এই চলক ব্যবহারের ফলে মাল্টিকোলিনিয়ারিটি সমস্যা ছাড়া নির্ভরণ বিশ্লেষণের আর কোন সমস্যা সৃষ্টি হয় না। কাজেই সাধারণ ন্যূনতম বর্গপদ্ধতি প্রয়োগ করে ৬.৫ অনুচ্ছেদে আলোচিত পদ্ধতিতে নির্ভারাঙ্কসমূহ নিরূপণ করা যায়। মাল্টিকোলিনিয়ারিটি সমস্যা সৃষ্টি হলে নির্ভরণ বিশ্লেষণ করার জন্য ৬.৯ অনুচ্ছেদে আলোচিত পদ্ধতি প্রয়োগ করা যায়। এখানে মাল্টিকোলিনিয়ারিটি সমস্যা কিভাবে সৃষ্টি হতে পারে তা একটি উদাহরণের সাহায্যে ব্যাখ্যা করা যাক।

ধরা যাক এক ধরনের ফসল উৎপাদন করার জন্য কিছু জমিতে ফসফরাস সার এবং কিছু জমিতে পটাশ সার প্রয়োগ করে জমি চাষ করা হয়েছে। উৎপাদিত ফসলের উপর এই সার দুটির কি প্রভাব আছে তা পর্যালোচনা করার জন্য নির্ভরণ বিশ্লেষণ করা যেতে পারে। এক্ষেত্রে উৎপাদিত ফসলের পরিমাণ (y kg), ফসফরাস সারের ব্যবহার (x_1 ; $x_1 = 0$ ফসফরাস ব্যবহার করা হয়নি, $x_1 = 1$ ফসফরাস ব্যবহার করা হয়েছে), পটাশ সারের ব্যবহার (x_2 ; $x_2 = 0$ পটাশ সার ব্যবহার করা হয়নি, $x_2 = 1$ পটাশ সার ব্যবহার করা হয়েছে) এবং জমির পরিমাণ (x_3 হেক্টর) চলক বিবেচনা করে ধরা যাক ছয়টি জমি হতে এই চলকগুলির মান নথিভুক্ত করা হয়েছে। তাহলে চলকগুলির মান নিম্নরূপ হতে পারে। এক্ষেত্রে

y	x_1	x_2	x_3
y_1	0	1	x_{31}
y_2	0	1	x_{32}
y_3	0	1	x_{33}
y_4	1	0	x_{34}
y_5	1	0	x_{35}
y_6	1	0	x_{36}

নির্ভরণ বিশ্লেষণের জন্য রৈখিক প্রতিকৃতি

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + e$$

বিবেচনা করা মাল্টিকোলিনিয়ারিটি সমস্যার সম্মুখীন হতে হবে। কারণ উপরিউক্ত প্রতিকৃতিকে

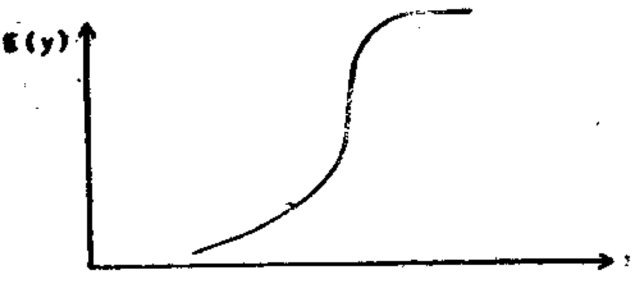
$$Y = X\beta + U$$

আকারের প্রকাশ করা হলে

$$X = \begin{bmatrix} 1 & 0 & 1 & x_{31} \\ 1 & 0 & 1 & x_{32} \\ 1 & 0 & 1 & x_{33} \\ 1 & 1 & 0 & x_{34} \\ 1 & 1 & 0 & x_{35} \\ 1 & 1 & 0 & x_{36} \end{bmatrix}_{6 \times 4}$$

এখানে X ম্যাট্রিক্সের দ্বিতীয় ও তৃতীয় স্তম্ভ যোগ করা হলে প্রথম স্তম্ভের সমান হয় বিধায় X ম্যাট্রিক্সের স্তম্ভগুলি অনপেক্ষ (independent) নয়। X ম্যাট্রিক্সের স্তম্ভগুলি অনপেক্ষ না হওয়ার কারণে বলা যায় মাল্টিকোলিনিয়ারিটি সমস্যার সৃষ্টি হয়েছে।

নকল নির্ভরশীল চলক (Dummy dependent variable) : নকল নির্ভরশীল চলকের উদাহরণ আগেই উল্লেখ করা হয়েছে। নির্ভরশীল চলক নকল হলে অধিকাংশ ক্ষেত্রেই নির্ভরণ রেখা বক্ররেখা হয়ে থাকে এবং বক্ররেখার আকার নিম্নরূপ ধারণ



চিত্র ৬.১৬ : লজিস্টিক নির্ভরণ রেখা।

করে। এই রেখাকে বলা হয় লজিস্টিক নির্ভরণ রেখা। এই রেখার গাণিতিক ফাংশন হলো

$$E(y) = \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)} \qquad (৬.১৪.১)$$

এখানে একটি অনপেক্ষ চলক (x) বিবেচনা করা হয়েছে। উক্ত ফাংশনকে রৈখিকভাবে প্রকাশ করে লজিস্টিক নির্ভরণ রেখা মিল করা যায়। নিচে তা আলোচনা করা হলো।

ধরা যাক $E(y) = P$, এখানে $E(y)$ হলো নির্ভরশীল চলকের গণসমষ্টি গড় এবং নির্ভরশীল চলক নকল চলক হওয়াতে এই গড় হলো একটি সম্ভাবনা (probability)। এই P ব্যবহার করে একটি লজিট পরিবর্তন

$$P' = ln\frac{P}{1-P} \qquad (৬.১৪.২)$$

এর মাধ্যমে (৬.১৪.১) হতে পাওয়া যায়

$$P' = \beta_0 + \beta_1 x \qquad (৬.১৪.৩)$$

এখানে শেষোক্ত ফাংশন (৬.১৪.৩) একটি রৈখিক ফাংশন। উপরিউক্ত ফাংশন মিল করার পদ্ধতি নিচে আলোচনা করা হলো।

ধরা যাক চলক x-এর q স্তর $x_1, x_2, \cdots x_j, \cdots, x_q$ আছে এবং j-তম স্তরের $[j = 1, 2, \cdots, q]$ n_j তথ্যমান আছে। এই n_j তথ্যমানের জন্য নির্ভরশীল চলকের n_j মান আছে। ধরা যাক ঐ n_j মানের মধ্যে R_j হলো 1 এবং $(n_j - R_j)$ হলো 0। তাহলে, x_j-এর ক্ষেত্রে 1 এর সমানুপাত হলো

$$\bar{P}_j = \frac{R_j}{n_j} \qquad (৬.১৪.৪)$$

এই নমুনা সমানুপাতের উপর লজিট পরিবর্তন করে পাওয়া যায়

$$\bar{P}_j' = ln\frac{P_j}{1-P_j} \qquad (৬.১৪.৫)$$

এখন $\bar{P}_j'$-কে নির্ভরশীল চলকের মান হিসেবে বিবেচনা করে (৬.১৪.৩) রৈখিক ফাংশনটি মিল করা যায়।

উপরিউক্ত $\bar{P}_j'$-এর নমুনা ভেদাঙ্ক হলো

$$v(\bar{P}_j') = \frac{1}{n_j \bar{P}_j(1-\bar{P}_j)} \qquad (৬.১৪.৬)$$

সুতরাং ফাংশন (৬.১৪.৩) মিল করার ক্ষেত্রে সাধারণ ন্যূনতম বর্গপদ্ধতি (OLS method) এর পরিবর্তে ৬.১০ অনুচ্ছেদে আলোচিত ভার আরোপিত ন্যূনতম বর্গপদ্ধতি (WLS method) প্রয়োগ করতে হয়। এখানে ভার হলো

$$W_j = n_j \bar{P}_j (1 - \bar{P}_j) \qquad (৬.১৪.৭)$$

এই ভার আরোপের ফলে নতুন নির্ভরশীল চলক এবং অনপেক্ষ চলক হবে, যথাক্রমে

$$p_j = \bar{P}_j' W_j \quad \text{এবং} \quad X_j = x_j W_j$$

ধরা যাক b_0 এবং b_1 হলো যথাক্রমে β_0 এবং β_1-এর নিরূপক। তাহলে

$$b_1 = \frac{SP(X_0 p_0)}{SS(X_0)} \quad \text{এবং} \quad b_0 = \bar{p}_0 - b_1 \bar{X}$$

যেখানে
$$\bar{p} = \frac{\Sigma n_j p_j}{\Sigma n_j} \quad \text{এবং} \quad \bar{X} = \frac{\Sigma n_j X_j}{\Sigma n_j}$$

সুতরাং মিল করা ফাংশন হলো

$$\hat{P} = b_0 + b_1 x$$

এবং
$$\hat{P} = \frac{\exp(b_0 + b_1 x)}{1 + \exp(b_0 + b_1 x)}$$

এ ধরনের বিশ্লেষণে বর্তমানে কোনো অসুবিধা নেই। একাধিক অনপেক্ষ চলকের ক্ষেত্রেও এ ধরনের বিশ্লেষণের জন্য কম্পিউটার প্রোগ্রাম আছে বিধায় যে কোনো গবেষকই এই লজিস্টিক ফাংশন মিল করতে পারেন। এখানে এই ফাংশন মিল করার একটি উদাহরণ দেওয়া হলো।

উদাহরণ ৬.৬

একটি পরীক্ষাগারে Sumicidin প্রয়োগ করে Wood louse মারা যায় কিনা তা পর্যবেক্ষণ করার জন্য Sumicidin (x)-এর চারটি স্তর [$x = 0.5$, $x = 1$, $x = 2$ এবং $x = 3$] ত্রিশটি Wood lice-এর উপর প্রয়োগ করে ১৫ দিন যাবৎ পর্যবেক্ষণ করে প্রতিদিনের ফলাফল নথিভুক্ত করা হয়েছে (সারণি ৬.১৩)।

সারণি ৬.১৩ : Sumicidin প্রয়োগে wood louse মারা গেছে কিনা তার ফলাফল।

দিন	Sumicidin (x)-এর বিভিন্ন স্তর প্রয়োগে Wood louse মারা যাওয়া ($y = 1$) এবং মারা না যাওয়া ($y = 0$)			
	$x = 0.5$	$x = 1$	$x = 2$	$x = 3$
1	0	1	1	1
2	1	1	0	1
3	0	1	0	1
4	1	0	0	0
5	0	0	0	1
6	0	0	1	1
7	0	0	1	0

দিন	$x=0.5$	$x=1$	$x=2$	$x=3$
8	0	0	0	0
9	0	0	0	0
10	0	0	0	0
11	1	0	1	0
12	0	0	1	0
13	0	1	1	1
14	0	1	1	1
15	1	1	0	1

লজিস্টিক নির্ভরণ-এর মাধ্যমে Wood louse মারা যাওয়ার উপর sumicidin-এর প্রভাব পর্যালোচনা করা যাক।

সারণি ৬.১৪ : লজিস্টিক ফাংশন মিল করার জন্য বিভিন্ন ধাপের হিসাব-নিকাশ।

Sumicidin এর স্তর x_j	প্রতি x_j-এর জন্য প্রাপ্ত তথ্য মান, n_j	মোট কয়-দিন Wood louse মারা গেছে R_j	মারা যাওয়া দিনের সমানুপাত $\bar{P}_j = \frac{R_j}{n_j}$	রূপান্তরিত সমানুপাত $\bar{P}'_j = \ln \frac{\bar{P}_j}{1-\bar{P}_j}$	ভার $W_j = n_j \bar{P}_j \times (1-\bar{P}_j)$
0.5	15	4	0.266667	−1.01160	2.93333
1	15	6	0.40	−0.40546	3.60
2	15	7	0.466667	−0.13353	3.73333
3	15	8	0.533333	0.13353	3.73333

$p_j = \bar{P}'_j W_j$	$X_j = x_j W_j$	$n_j p_j X_j$	$n_j X_j$	$n_j p_j$
−2.96736	1.466665	−65.28184	21.99997	−44.5104
−1.45966	3.60	−6.08192	54.00	−21.8949
−0.49851	7.466666	−55.83311	111.99999	−7.47765
0.49851	11.199999	83.74967	167.99998	7.47765
মোট		−43.4472	355.99994	−66.4053

$$SP(Xp) = \Sigma n_j X_j p_j - \frac{\Sigma n_j X_j \Sigma n_j p_j}{\Sigma n_j}$$

$$= -43.4472 + \frac{355.99994 \times 66.4053}{60}$$

$$= 350.55751$$

$$SS(X) = \Sigma n_j X_j^2 - \frac{(\Sigma n_j X_j)^2}{\Sigma n_j} = 2944.532775 - \frac{(355.99994)^2}{60}$$

$$= 832.26682$$

$$b_1 = \frac{SP(Xp)}{SS(X)} = 0.42121$$

$$b_0 = \bar{p} - b_1 \bar{X}, \qquad \bar{p} = \frac{\Sigma n_j p_j}{\Sigma n_j} = \frac{-66.4053}{60}$$

$$= 1.39243 \qquad\qquad = -1.10675$$

$$\bar{X} = \frac{\Sigma n_j X_j}{\Sigma n_j} = \frac{355.99994}{60}$$

$$= 5.93333$$

সুতরাং $\hat{P} = 1.39243 + 0.42121x$ (৬.১৪.৮)

এখানে $SSR = b_1\, SP(Xp) = 147.65833$

$$SST = \Sigma n_j p_j^2 - \frac{(\Sigma n_j p_j)^2}{\Sigma n_j}$$

$$= 171.49286 - \frac{(-66.4053)^2}{60} = 97.99846$$

$$SSE = SST - SSR = 49.65987$$

$\therefore$ $$F = \frac{SSR/1}{SSE/(n-2)} = 178.40, \qquad n = \Sigma n_j = 60$$

এখানে $F_{0.05,1,58} = 4.01$ বা $(P = 0.00)$ হওয়ায় মিল করা ফাংশন (৬.১৪.৮) তাৎপর্যপূর্ণ।

এই ফাংশন মিল করার মাধ্যমে 15 দিনের মধ্যে কি সমানুপাতে Wood louse মারা যায় তা নিরূপণ করা যেতে পারে। এই নিরূপণ Sumicidin-এর যে কোনো স্তরের জন্য করা যেতে পারে। ধরা যাক $x = 2$, তাহলে

$$\hat{P}' = 2.23485$$

$$\frac{\hat{P}}{1-\hat{P}} = \text{Anti ln } 2.23485 = 9.3451$$

$$\therefore \quad \hat{P} = 0.90$$

উপরে একটি অনপেক্ষ চলক নিয়ে লজিস্টিক ফাংশন মিল করার পদ্ধতি আলোচনা করা হয়েছে। বাস্তবে একাধিক অনপেক্ষ চলক নিয়েও এই ফাংশন মিল করার প্রয়োজন হতে পারে। সেক্ষেত্রে লজিস্টিক ফাংশন হলো

$$E(y) = \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k)}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k)} \qquad (৬.১৪.৯)$$

এই ফাংশন মিল করার জন্য সর্বোত্তম সম্ভাবনা নিয়ম (maximum likelihood method) প্রয়োগ করা যেতে পারে এবং ফাংশনের পরামানসমূহ নিরূপণ করার জন্য তৈরি কম্পিউটার প্রোগ্রাম ব্যবহার করা যেতে পারে।

উপরিউক্ত পদ্ধতিতে ফাংশন মিল করার পর পরামানসমূহের তাৎপর্য যাচাই করা যেতে পারে। ধরা যাক, $b_0, b_1, b_2, \cdots, b_k$ যথাক্রমে $\beta_0, \beta_1, \beta_2, \cdots, \beta_k$-এর নিরূপক। তাহলে যে কোনো পরামান β_i $(i = 1, 2, \cdots, k)$ এর তাৎপর্য যাচাই করার জন্য Wald Statistic (WS) ব্যবহার করা যায়, যেখানে

$$WS = \left[\frac{b_i}{s.e(b_i)}\right]^2$$

এখানে $s.e(b_i)$ কম্পিউটার বিশ্লেষণের মাধ্যমেই পাওয়া যায়। এই WS-এর বিন্যাস হলো কাইবর্গ (Chi-square) বিন্যাস এবং শ্রেণী বিভক্ত [categorical] চলকের ক্ষেত্রে Chi-square-এর স্বাধীনতার মাত্রা হলো (শ্রেণী সংখ্যা-1)। নির্ণেয় WS-এর মান $\chi^2_{0.05}$ অপেক্ষা বড় হলে পরামান β_i তাৎপর্যপূর্ণ বলে বিবেচিত হয়।

প্রতিকৃতি (৬.১৪.৯) মিল করা ভাল হয়েছে কিনা তাও χ^2 কাইবর্গ তথ্যজমান দ্বারা যাচাই করা যায়, যেখানে

$$\chi^2 = \sum_{i=1}^{n} \frac{(y_i - \hat{P}_i)^2}{\hat{P}_i(1-\hat{P}_i)}$$

এখানে y_i হলো i-তম নির্ভরশীল চলকের মান $(i = 1, 2, \cdots, n)$, $\hat{P}_i$ হলো x সমূহের মানের ভিত্তিতে y_i-এর নিরূপিত মান। এই χ^2-এর স্বাধীনতার মাত্রা হলো $(n - P)$। এখানে P হলো প্রতিকৃতিতে পরামানের সংখ্যা। নির্ণেয় $\chi^2 \geqslant \chi^2_{0.05}$ হলে প্রতিকৃতির মিল ভাল হওয়ার নাস্তি কল্পনা বাতিল বলে গণ্য হবে।

৬.১৫ নির্ভরণ রেখার সমতা (Equality of Regression Lines)

উপরিউক্ত বিভিন্ন অনুচ্ছেদে নির্ভরণ বিশ্লেষণের বিভিন্ন দিক আলোচনা করা হয়েছে। সকল ক্ষেত্রেই কোনো গণসমষ্টি হতে একটি নমুনার ভিত্তিতে নির্ভরণ রেখা, রৈখিক বা অরৈখিক মিল করার পদ্ধতি আলোচনা করা হয়েছে। কিন্তু যে চলকসমূহের ভিত্তিতে নির্ভরণ বিশ্লেষণ করা হবে ঐ চলকসমূহের উপাত্ত যদি ভিন্ন ভিন্ন নমুনার মাধ্যমে সংগৃহীত হয় বা একই চলকগুচ্ছ ভিন্ন ভিন্ন সময়ে সংগৃহীত হয় বা কীটপতঙ্গের ক্ষেত্রে ভিন্ন ভিন্ন Species বা Family হতে সংগৃহীত হয়, তাহলে ভিন্ন ভিন্ন উপাত্তের জন্য একই নির্ভরণ রেখা হবে না ভিন্ন নির্ভরণ রেখা হবে তা যাচাই করে দেখার প্রয়োজন হয়। উদাহরণ হিসেবে সারণি ১.১ (প্রথম খণ্ড)-এর উপাত্তের কথা বিবেচনা করা যাক। ঐ উপাত্ত একটি কীট Slug হতে সংগৃহীত হয়েছে। কিন্তু Slug-এর চারটি Species, যথা, M. Rusticus, M. Sowerbyi, M. Gagates এবং Limux Tenellus আছে। এখন Slug-এর Body weight-এর জন্য অন্যান্য শারীরিক বৈশিষ্ট্যের উপর নির্ভরণ রেখা মিল করা হলে চার প্রকার Slug-এর জন্য চারটি নির্ভরণ রেখা হবে না সকল Slug-এর জন্য একটি নির্ভরণ রেখা হবে তা পর্যালোচনা করা প্রয়োজন। চার প্রকার Slug-এর জন্য চারটি নির্ভরণ রেখা মিল করা হলে এবং যাচাই-এর মাধ্যমে যদি চারটি রেখার মধ্যে কোনো তাৎপর্যপূর্ণ পার্থক্য (significant difference) পরিলক্ষিত না হয়, তাহলে নির্ভরণ রেখাগুলির সমতা প্রমাণিত হবে এবং সেক্ষেত্রে চারটি নির্ভরণ রেখার পরিবর্তে সকল Slug-এর উপাত্ত ব্যবহার করে একটি নির্ভরণ রেখার মাধ্যমে তাদের শারীরিক বৈশিষ্ট্যের সম্পর্ক পর্যালোচনা করা যাবে। বর্তমান অনুচ্ছেদে এরূপ নির্ভরণ রেখাসমূহের সমতা যাচাই পদ্ধতি আলোচনা করা হবে।

ধরা যাক $(k-1)$ অনপেক্ষ চলক $x_r (r = 2, 3, \ldots, k)$ এর ভিত্তিতে একটি নির্ভরশীল চলক y-এর রৈখিক সম্পর্ক পর্যালোচনা করার জন্য n সংখ্যক তথ্যমান আছে। ধরা যাক এই n সংখ্যক উপাত্ত একটি গণসমষ্টি হতে P নমুনা চয়ন করে সংগ্রহ করা হয়েছে, যেখানে i-তম নমুনার $[i = 1, 2, \ldots, P]$ তথ্যমান সংখ্যা n_i এবং n হলো n_i সমূহের যোগফল $[\, n = \Sigma n_i \,]$। ধরা যাক i-তম নমুনার চলকসমূহের উপাত্তকে

$$y_{ij}\ ;\ x_{rij}\ \ (\ j=1,2,\cdots,n_i\ )$$

দ্বারা চিহ্নিত করা যায়। এই চলকসমূহের মধ্যে রৈখিক সম্পর্ক বিবেচনা করা যাক।

$$y_{ij}=\beta_1+\beta_2x_{2ij}+\beta_3x_{3ij}+\cdots+\beta_kx_{kij}+e_{ij} \qquad (৬.১৫.১)$$

এই প্রতিকৃতির ক্ষেত্রে ন্যূনতম বর্গপদ্ধতির সকল অনুমান বহাল আছে বিবেচনা করে নির্ভরণ বিশ্লেষণ করে পাওয়া যায়

$$SST=\sum_i^p\sum_j^{n_i}y_{ij}^2-\frac{(\Sigma\Sigma y_{ij})^2}{n}$$

বিচ্যুতির বর্গসমষ্টি $=S_1^2=SST-SSR$

$$=SST-b_2SP(x_2y)-b_3SP(x_3y)\cdots-b_k\,SP(x_ky)$$

এখানে $b_2,\ b_3,\ \cdots\ b_k$ হলো যথাক্রমে $\beta_2,\ \beta_3,\cdots,\ \beta_k$ এর নিরূপক এবং

$$SP(x_ry)=\Sigma\Sigma\ y_{ij}\,x_{rij}-\frac{\Sigma\Sigma y_{ij}\ \Sigma\Sigma x_{rij}}{n}$$

$b_2,\ b_3,\ \cdots,\ b_k$-এর মান নির্ণয় পদ্ধতি ৬.৫ অনুচ্ছেদে আলোচনা করা হয়েছে। এই মানসমূহ কম্পিউটার প্রোগ্রাম-এর মাধ্যমে নির্ণয় করা হয়। এখানে S_1^2-এর স্বাধীনতার মাত্রা হলো $(n-k)$।

উপরিউক্ত নির্ভরণ বিশ্লেষণের ভিত্তি হলো যে সকল নমুনার জন্য একটি নির্ভরণ রেখা হবে। বাস্তবক্ষেত্রে সকল নমুনার জন্য একটি নির্ভরণ রেখা নাও হতে পারে। P-নমুনার জন্য P নির্ভরণ রেখা সঠিক হতে পারে। সেক্ষেত্রে নির্ভরণ রেখাসমূহের মধ্যে পার্থক্য থাকার সম্ভাব্য কারণসমূহ হতে পারে :

(i) ভিন্ন ভিন্ন নমুনার নির্ভরণ রেখার ক্ষেত্রে ইন্টারসেপ্টসমূহ ভিন্ন ভিন্ন।

(ii) ভিন্ন ভিন্ন নমুনার নির্ভরণ রেখার ক্ষেত্রে ঢালসমূহ (slopes) বা নির্ভরাঙ্কসমূহ ভিন্ন ভিন্ন।

(iii) ভিন্ন ভিন্ন নমুনার নির্ভরণ রেখার ক্ষেত্রে ইন্টারসেপ্টসমূহ এবং নির্ভরাঙ্কসমূহ ভিন্ন ভিন্ন।

প্রথমোক্ত কারণে নির্ভরণ রেখাসমূহকে লেখা যায়

$$\begin{aligned}
y_{1j}&=\beta_1+\beta_2\,x_{21j}+\beta_3\,x_{31j}+\cdots+\beta_kx_{k1j}+e_{1j}\\
y_{2j}&=(\alpha_2+\beta_1)+\beta_2\,x_{22j}+\beta_3x_{32j}+\cdots+\beta_kx_{k2j}+e_{2j} \qquad (৬.১৫.২)\\
\cdots&\quad\cdots\quad\cdots\quad\cdots\quad\cdots\quad\cdots\\
y_{pj}&=(\alpha_p+\beta_1)+\beta_2x_{2pj}+\beta_3x_{3pj}+\cdots+\beta_kx_{kpj}+e_{pj}
\end{aligned}$$

(৬.১৫.২) দ্বারা চিহ্নিত প্রতিকৃতিসমূহের ক্ষেত্রে ন্যূনতম বর্গপদ্ধতির অনুমানসমূহ বহাল আছে বিবেচনা করে নির্ভরণ বিশ্লেষণ করা হলে ঐ বিশ্লেষণ থেকে বিচ্যুতির বর্গসমষ্টি (SSE) পাওয়া যায়

$$S_2^2 = \sum_i^p \sum_j^{n_i} y_{ij}^2 - a_2 \sum_j y_{2j} - a_3 \sum_j y_{3j} - \cdots - a_p \sum_j y_{pj}$$
$$- b_2' \Sigma\Sigma x_{2ij}\, y_{ij} - b_3' \Sigma\Sigma x_{3ij}\, y_{ij} - \cdots\cdots$$
$$- b_k' \Sigma\Sigma x_{kij}\, y_{ij} - b_1' \Sigma\Sigma y_{ij}$$

এই S_2^2-এর স্বাধীনতার মাত্রা হলো $(n - P - k + 1)$। এখানে $a_2, a_3, \ldots, a_p$ হলো যথাক্রমে $\alpha_2, \alpha_3, \ldots, \alpha_p$-এর নিরূপক এবং $b_1', b_2', b_3', \ldots, b_k'$ হলো এই বিশ্লেষণের ক্ষেত্রে যথাক্রমে $\beta_1, \beta_2, \beta_3, \cdots, \beta_k$-এর নিরূপক।

এই বিশ্লেষণও ৬.৫ অনুচ্ছেদে আলোচিত পদ্ধতিতে করা হয় এবং কম্পিউটার প্রোগ্রাম ব্যবহার করে বিশ্লেষণ করা হলে বিচ্যুতির বর্গসমষ্টি S_2^2 সরাসরি পাওয়া যায়। তবে বিশ্লেষণের জন্য $x_2, x_3, \ldots, x_k$ চলকসমূহের মান ভিন্ন কতকগুলি নতুন চলক $z_2, z_3, \cdots, z_p$ বিবেচনা করতে হয়, যেখানে z_2-এর প্রথম n_1 মান 0, দ্বিতীয় n_2 মান 1 এবং বাকি মানসমূহ 0, z_3-এর প্রথম n_1 ও n_2 মান 0, তৃতীয় n_3 মান 1 এবং পরবর্তী মানসমূহ 0, অনুরূপভাবে z_p-এর শেষ n_p মান 1 এবং বাকি মানসমূহ 0 বিবেচনা করতে হয়।

এখন নাস্তি কল্পনা

H_0 : ইন্টারসেপ্টসমূহ ভিন্ন নয় (৬.১৫.৩)

বিবেচনা করা হলে যাচাই তথ্যজমান হবে

$$F_1 = \frac{S_3^2/(P-1)}{S_2^2/(n-P-k+1)}$$

এখানে $S_3^2 = S_1^2 - S_2^2$। এই F_1-এর বিন্যাস $(P-1)$ এবং $(n-P-k+1)$ স্বাধীনতার মাত্রাবিশিষ্ট ভেদাঙ্ক অনুপাত বিন্যাস। এই F_1-এর মান 5% সংশয় মাত্রায় সারণিকৃত F-এর মান অপেক্ষা বড় বা সমান হলে নাস্তি কল্পনা বাতিল বলে গণ্য হবে।

নির্ভরণ রেখাসমূহের মধ্যে পার্থক্য থাকার তৃতীয় সম্ভাব্য কারণের [উপরিউক্ত (iii)] ক্ষেত্রে প্রতিকৃতি বিবেচিত হতে পারে

$$y_{ij} = \beta_{1i} + \beta_{2i}x_{2ij} + \beta_{3i}x_{3ij} + \dots + \beta_{ki}x_{kij} + e_{ij} \qquad (৬.১৫.৪)$$

$$i = 1, 2, \dots, P$$

এখানে P নমুনার জন্য P প্রতিকৃতি বিবেচনা করা হয়েছে। প্রতিটি প্রতিকৃতির ক্ষেত্রে ন্যূনতম বর্গপদ্ধতি প্রয়োগ করার জন্য অনুমানসমূহ বহাল আছে বিবেচনা করে বিশ্লেষণ করা হলে i-তম $(i = 1, 2, \cdots, P)$ প্রতিকৃতি হতে বিচ্যুতির বর্গসমষ্টি (SSE) পাওয়া যায়

$$S_{1i}^2 = \sum_j y_{ij}^2 - b_{1i}\Sigma y_{ij} - b_{2i}\sum_j x_{2ij}y_{ij} - \dots - b_{ki}\sum_j x_{kij}y_{ij}$$

এই S_{1i}^2-এর স্বাধীনতার মাত্রা হলো $(n_i - k)$। এখানে $b_{1i}, b_{2i}, \dots, b_{ki}$ হলো যথাক্রমে i-তম নমুনার ক্ষেত্রে $\beta_{1i}, \beta_{2i}, \dots, \beta_{ki}$-এর নিরূপক। এই নিরূপকসমূহ ৬.৫ অনুচ্ছেদে আলোচিত পদ্ধতিতে পাওয়া যায়। এখন নাস্তি কল্পনা

$$H_0 : \text{সকল নমুনার নির্ভরাঙ্কসমূহ সমসত্ত্ব} \qquad (৬.১৫.৫)$$

যাচাই করার জন্য যাচাই তথ্যজমান হবে.

$$F_2 = \frac{(S_2^2 - S_4^2)/(Pk - P - k + 1)}{S_4^2/(n - Pk)}$$

এখানে
$$S_4^2 = \sum_{i=1}^{p} S_{1i}^2$$

এই S_4^2-এর স্বাধীনতার মাত্রা হলো $(n - Pk)$। এই F_2-এর বিন্যাস হলো $(Pk - P - k + 1)$ এবং $(n - Pk)$ স্বাধীনতার মাত্রাবিশিষ্ট ভেদাঙ্ক অনুপাত বিন্যাস।

আবার নাস্তি কল্পনা

H_0 : সকল নমুনার জন্য নির্ভরণ রেখাসমূহ সমসত্ত্ব (৬.১৫.৬) যাচাই করার জন্য যাচাই তথ্যজমান হবে

$$F_3 = \frac{(S_3^2 + S_2^2 - S_4^2)/k(P - 1)}{S_4^2/(n - Pk)}$$

এই F_3-এর বিন্যাস হলো $k(P - 1)$ এবং $(n - Pk)$ স্বাধীনতার মাত্রাবিশিষ্ট ভেদাঙ্ক অনুপাত বিন্যাস। F_1 নির্ণয় করে যেভাবে সিদ্ধান্ত নিতে হয় F_2 ও F_3 নির্ণয় করে সেই একই পদ্ধতিতে সিদ্ধান্ত নিতে হয়।

এতক্ষণ ন্যূনতম বর্গপদ্ধতি প্রয়োগ করে নির্ভরণ বিশ্লেষণ করার বিষয় আলোচনা করা হয়েছে। ধরা যাক প্রতিকৃতি (৬.১৫.৪)-এর ক্ষেত্রে $V(e_{ij}) = \sigma_i^2 \; (i = 1, 2, \cdots, P)$। তাহলে সাধারণ ন্যূনতম বর্গপদ্ধতির [OLS method] পরিবর্তে ভরারোপিত ন্যূনতম বর্গপদ্ধতি [WLS method] প্রয়োগ করে নির্ভরণ বিশ্লেষণ করতে হবে। এখন ভরারোপিত ন্যূনতম বর্গপদ্ধতি প্রয়োগ করে নাস্তি কল্পনা (৬.১৫.৩), (৬.১৫.৫) এবং (৬.১৫.৬) যাচাই পদ্ধতি আলোচনা করা হবে।

প্রতিকৃতি (৬.১৫.৪)-এর ক্ষেত্রে i-তম নমুনার ভিত্তিতে প্রাপ্ত বিচ্যুতির বর্গসমষ্টি হলো $S^2_{1i} (i = 1, 2, \cdots, P)$। এই S^2_{1i} ব্যবহার করে Bartlett's (1937) χ^2-যাচাই প্রয়োগ করে নাস্তি কল্পনা

$$H_0 : \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_p^2 \qquad (৬.১৫.৭)$$

যাচাই করতে হয়। এই নাস্তি কল্পনা বাতিল হলে ভরারোপিত ন্যূনতম বর্গপদ্ধতি প্রয়োগ করতে হয়। সেক্ষেত্রে, ধরা যাক

$$W_i = 1/s^2_{1i}, \qquad s^2_{1i} = S^2_{1i}/(n_i - k)$$

তাহলে বিশ্লেষণের পূর্বে y_{ij}, x_{rij} এবং z_{ij}-এর সকল মানকে W_i দ্বারা গুণ করে নিতে হয় এবং গুণ করে প্রাপ্ত পরিবর্তিত চলকসমূহের মান ব্যবহার করে সাধারণ ন্যূনতম বর্গপদ্ধতি প্রয়োগ করে বিশ্লেষণ করতে হয়। এই পরিবর্তিত চলকের মান হতেই S_1^2 ও S_2^2 নির্ণয় করতে হয়। এক্ষেত্রে $S_4^2 = \sum\limits_{i}^{p} W_i \, S^2_{1i}$। অন্যান্য তথ্যজমানের নির্ণয় পদ্ধতি এই অনুচ্ছেদে প্রথমাংশের অনুরূপ।

উদাহরণ ৬.৭

সারণি ১.১ (১মখণ্ড)-এ দেয়া উপাত্তের ভিত্তিতে চার প্রকার Slug যথা, M. Sowerbyi, M. Rusticus, M. Gagates এবং L. Tenellus-এর ক্ষেত্রে Body length (x_1 in mm), Mantle length (x_2 in mm), Keel length (x_3 in mm) এবং Shell length (x_4 in mm) এর উপর Body weight (y in gms)-এর চারটি নির্ভরণ রেখা মিল করে ঐ রেখাগুলি সমসত্ত্ব (homogeneous) কিনা তা পর্যালোচনা করা যাক।

উপরিউক্ত উপাত্তের ক্ষেত্রে $n_1 = 15$, $n_2 = 15$, $n_3 = n_4 = 10$, $P = 4$, $k = 5$ এবং অনপেক্ষ চলকের সংখ্যা $= k - 1 = 4$ । প্রতিকৃতি (৬.১৫.৪) অনুসরণ করে সাধারণ ন্যূনতম বর্গপদ্ধতিতে প্রাপ্ত নির্ভরণ বিশ্লেষণের ফলাফল M. Sowerbyi, M. Rusticus, M. Gagates এবং L. Tenellus-এর ক্ষেত্রে যথাক্রমে সারণি ৬.১৫, সারণি ৬.১৬, সারণি ৬.১৭ এবং সারণি ৬.১৮-এ উপস্থাপন করা হলো।

সারণি ৬.১৫ : M. Sowerbyi-এর Body weight-এর নির্ভরণ বিশ্লেষিত ফলাফল।

অনপেক্ষ চলক	নির্ভরাঙ্কের নিরূপক			
	b_i	$s.e(b_{ii})$	t-এর মান	P-মান
ধ্রুবক	− 5.114	0.768	6.66	0.00
x_1	0.042	0.032	1.31	0.22
x_2	0.067	0.050	1.34	0.20
x_3	0.191	0.050	3.81	0.00
x_4	0.139	0.088	1.58	0.14

ভেদাঙ্ক বিশ্লেষণ সারণি

ভেদের উৎস	স্বাধীনতার মাত্রা	গড় বর্গসমষ্টি	F	P-মান	R^2
নির্ভরণ	4	9.326	58.62	0.00	0.959
বিচ্যুতি	10	0.1591			
মোট	14				

সারণি ৬.১৬ : M. Rusticus-এর Body weight-এর নির্ভরণ বিশ্লেষিত ফলাফল।

অনপেক্ষ চলক	নির্ভরাঙ্কের নিরূপক			
	b_i	$s.e(b_i)$	t-এর মান	P-মান
ধ্রুবক	− 4.444	1.773	2.51	0.03
x_1	0.050	0.061	0.82	0.43
x_2	0.223	0.120	1.85	0.09
x_3	0.102	0.126	0.81	0.43
x_4	−0.176	0.288	0.61	0.55

ভেদাঙ্ক বিশ্লেষণ সারণি

ভেদের উৎস	স্বাধীনতার মাত্রা	গড় বর্গসমষ্টি	F	P-মান	R^2
নির্ভরণ	4	14.224	20.92	0.00	0.893
বিচ্যুতি	10	0.6798			
মোট	14				

সারণি ৬.১৭ : M. Gagates-এর ক্ষেত্রে Body weight-এর নির্ভরণ বিশ্লেষিত ফলাফল।

অনপেক্ষ চলক নির্ভরাঙ্কের নিরূপক

	b_1	$s.e(b_1)$	t-এর মান	P-মান
ধ্রুবক	−6.926	1.583	4.37	0.01
x_1	0.027	0.047	0.57	0.59
x_2	0.137	0.074	1.83	0.13
x_3	0.216	0.086	2.50	0.05
x_4	0.229	0.127	1.79	0.13

ভেদাঙ্ক বিশ্লেষণ সারণি

ভেদের উৎস	স্বাধীনতার মাত্রা	গড় বর্গসমষ্টি	F	P-মান	R^2
নির্ভরণ	4	12.807	29.24	0.00	0.959
বিচ্যুতি	5	0.4380			
মোট	9				

সারণি ৬.১৮ : L. Tenellus-এর ক্ষেত্রে Body weight-এর নির্ভরণ বিশ্লেষিত ফলাফল।

অনপেক্ষ চলক নির্ভরাঙ্কের নিরূপক

	b_1	$s.e(b_1)$	t-এর মান	P-মান
ধ্রুবক	−5.247	0.926	5.66	0.00
x_1	0.086	0.041	2.09	0.09
x_2	0.111	0.058	1.91	0.11
x_3	−0.090	0.071	1.27	0.26
x_4	0.157	0.159	0.99	0.37

ভেদাঙ্ক বিশ্লেষণ সারণি

ভেদের উৎস	স্বাধীনতার মাত্রা	গড় বর্গসমষ্টি	F	P-মান	R^2
নির্ভরণ	4	2.399	22.81	0.00	0.948
বিচ্যুতি	5	0.1052			
মোট	9				

এই বিশ্লেষণ হতে পাওয়া যায় $s^2_{11} = 0.1591, s^2_{12} = 0.6798, s^2_{13} = 0.4380$ এবং $s^2_{14} = 0.1052$; এগুলির স্বাধীনতার মাত্রা হলো যথাক্রমে $\gamma_1 = 10$, $\gamma_2 = 10$, $\gamma_3 = 5$ এবং $\gamma_4 = 5$। উপরিউক্ত চারটি নির্ভরণের ক্ষেত্রে বিচ্যুতির ভেদাঙ্ক সমসত্ব কিনা যাচাই করার জন্য নাস্তি কল্পনা হলো

$$H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 \qquad (৬.১৫.৮)$$

এবং যাচাই তথ্যাজমান হলো

$$\chi^2 = \frac{1}{C}\left[\gamma \ln s^2 - \sum_{i=1}^{p} \gamma_i \ln s^2_{1i}\right]$$

এখানে $\gamma = \Sigma\gamma_i = 30$, $C = 1 + \dfrac{\Sigma\dfrac{1}{\gamma_i} - \dfrac{1}{\gamma}}{3(P-1)}$

$$= 1.063$$

$$s^2 = \frac{\Sigma\gamma_i s^2_{1i}}{\gamma} = 0.3702 \text{ এবং}$$

$\chi^2 = 7.35$। এই χ^2-এর বিন্যাস হলো $(P-1) = 3$ স্বাধীনতার মাত্রাবিশিষ্ট কাইবর্গ বিন্যাস। এই $\chi^2 < \chi^2_{0.05, 3}$ হওয়ায় নাস্তি কল্পনার বিপক্ষে কোনো যুক্তি নেই সিদ্ধান্ত নেয়া যায়।

যেহেতু চরাটি নির্ভরণের ক্ষেত্রে বিচ্যুতির ভেদাঙ্ক সমসত্ব, সে কারণে নির্ভরণ রেখার সমসত্বতা যাচাই করার জন্য সাধারণ ন্যূনতম বর্গপদ্ধতি প্রয়োগ করা যায়। এই পদ্ধতি অনুসারে প্রতিকৃতি (৬.১৫.১) ও প্রতিকৃতি (৬.১৫.২)-এর নির্ভরণ বিশ্লেষণের ফলাফল যথাক্রমে সারণি ৬.১৯ এবং সারণি ৬.২০-এ উপস্থাপন করা হলো।

সারণি ৬.১৯ : সকল Slug-এর Body weight-এর নির্ভরণ বিশ্লেষিত ফলাফল।

অনপেক্ষ চলক	নির্ভরাঙ্কের নিরূপক b_1	$s.e(b_1)$	t-এর মান	P-মান
ধ্রুবক	− 5.843	0.625	9.35	0.00
x_1	− 0.015	0.019	0.83	0.41
x_2	0.144	0.041	3.55	0.00
x_3	0.255	0.035	7.31	0.00
x_4	0.205	0.084	2.44	0.02

ভেদাঙ্ক বিশ্লেষণ সারণি

ভেদের উৎস	স্বাধীনতার মাত্রা	গড় বর্গসমষ্টি	F	P-মান	R^2
নির্ভরণ	4	42.027	81.05	0.00	0.878
বিচ্যুতি	45	0.5185			
মোট	49				

সারণি ৬.২০ : সকল প্রকার Slug-এর Body weight-এর ক্ষেত্রে নির্ভরণ রেখার সমসত্বতা যাচাই-এর সাথে সম্পর্কিত নির্ভরণ বিশ্লেষণের ফলাফল।

অনপেক্ষ চলক	নির্ভরাঙ্কের নিরূপক	পরিমিত বিচ্যুতি	t-এর মান	P-মান
ধ্রুবক	− 5.690	0.559	10.17	0.00
x_1	0.011	0.019	0.58	0.56
x_2	0.122	0.035	3.48	0.00
x_3	0.214	0.036	5.95	0.00
x_4	0.169	0.072	2.35	0.02
z_2	0.277	0.230	1.20	0.23
z_3	0.297	0.157	1.89	0.06
z_4	− 0.220	0.085	2.58	0.01

ভেদাঙ্ক বিশ্লেষণ সারণি

ভেদের উৎস	স্বাধীনতার মাত্রা	গড় বর্গসমষ্টি	F	P-মান	R^2
নির্ভরণ	7	25.075	66.18	0.00	0.917
বিচ্যুতি	42	0.3789			
মোট	49				

উক্ত বিশ্লেষণ হতে পাওয়া যাচ্ছে

$$S_1^2 = 23.334, \quad S_2^2 = 15.914, \quad S_3^2 = S_1^2 - S_2^2 = 7.42$$

$$S_4^2 = \sum_{i=1}^{4} S_{1i}^2 = 11.105, \quad n = 50. \quad k = 5; \quad P = 4$$

তাহলে $F_1 = \dfrac{S_3^2/P-1}{S_2^2/(n-P-k+1)} = 6.53$, $F_{.05,\,3,42} = 2.83$

$F_1 > 2.83$ হওয়ায় চার প্রকার Slug-এর জন্য চারটি নির্ভরণ রেখার ইন্টারসেপ্টসমূহ তাৎপর্যপূর্ণভাবে ভিন্ন। আবার

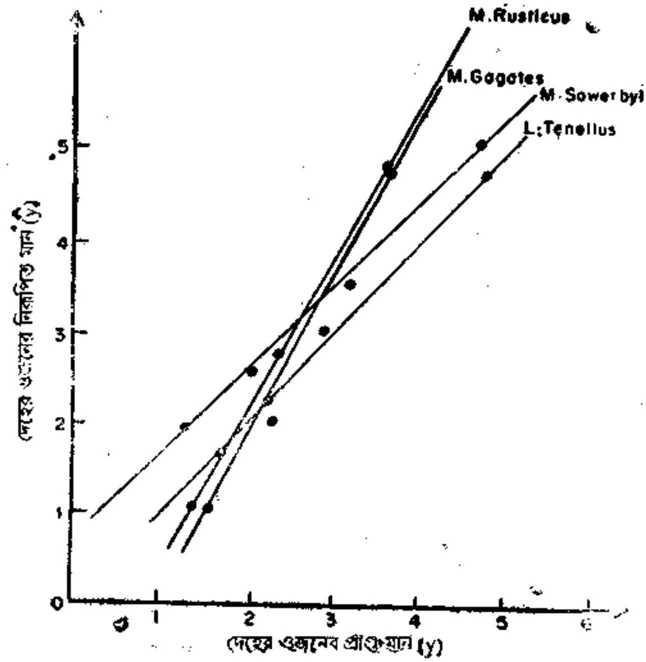
$$F_2 = \frac{(S_2^2 - S_4^2)/(P-k-P-k+1)}{S_4^2/(n-Pk)} = 1.08$$

$F_{.05\,,\,12,30} = 2.09$ । $F_2 < 2.09$ হওয়ায় চারটি নির্ভরণ রেখার ঢালসমূহ [নির্ভরাঙ্কসমূহ] সমসত্ব। এছাড়া

$$F_3 = \frac{(S_3^2 + S_2^2 - S_4^2)/k(P-1)}{S_4^2/(n-Pk)} = 2.20$$

$F_{.05\,,\,15,30} = 2.01$ । $F_3 > 2.01$ হওয়ায় বলা যায় যে, চার প্রকার Slug-এর জন্য চারটি নির্ভরণ রেখা সাবিকভাবে ভিন্ন ভিন্ন। তবে F_2 যাচাই তথ্যজমান হতে বলা যায় যে, সকল প্রকার Slug-এর Body weight অন্যান্য Body বৈশিষ্ট্য দ্বারা সমরূপভাবে প্রভাবিত। উদাহরণ হিসেবে Body length x_1-এর প্রভাবের কথা বিবেচনা করা যাক। F_2 যাচাই তথ্যজমানের উপর ভিত্তি করে বলা যায় y-এর উপর x_1-এর প্রভাব সকল প্রকার Slug-এর ক্ষেত্রে সমসত্ব। অনুরূপ সমসত্ব প্রভাব x_2, x_3 এবং x_4-এর ক্ষেত্রেও লক্ষণীয়।

চার প্রকার Slug-এর শারীরিক বৈশিষ্ট্য পর্যালোচনার জন্য যে চারটি নির্ভরণ রেখা মিল করা হয়েছে ঐগুলি যে সমসত্ব নয় তা Slug-এর Body weight-এর প্রাপ্ত তথ্যমানের বিপরীতে নিরূপিত মানের চিত্র [চিত্র ৬.১৭] হতেও বুঝা যায়। চিত্র হতে লক্ষ্য করা যাচ্ছে যে, চার প্রকার Slug-এর জন্য অঙ্কিত চারটি রেখার গতি ধারা ভিন্ন ভিন্ন।



চিত্র ৬.১৭ : চার প্রকার Slug-এর Body weight (y)-এর প্রাপ্ত মানের বিপরীতে নিরূপিত মানের চিত্র।

বর্তমান অনুচ্ছেদে নির্ভরণ রেখার সমতা যাচাই-এর একটি সংক্ষিপ্ত আলোচনা করা হয়েছে। এ সম্পর্কে আরো বিস্তারিত এবং কোনো বিশেষ চলকের নির্ভরাঙ্কসমূহের সমতা যাচাই পদ্ধতি জানার জন্য Bhuyan (1996) এবং Bhuyan and Majumder (1996) পর্যালোচনা করা যেতে পারে।

৬.১৬ অপরামাত্রিক নির্ভরণ বিশ্লেষণ (Non-Parametric Regression Analysis)

নির্ভরণ বিশ্লেষণের একটি গুরুত্বপূর্ণ অংশ হলো নির্ভরণ সহগসমূহের যাচাই এবং পূর্বাভাস করা মানের নিশ্চয়তা-পরিক্ষেপ নির্ণয়। উভয় ক্ষেত্রেই একটি প্রয়োজনীয় শর্ত হলো বিচ্যুতির বিন্যাস পরিমিত বিন্যাস হতে হবে এবং বিচ্যুতির প্রত্যাশিত

মান শূন্য হতে হবে $[E(e_i)=0]$। মোট কথা গতানুগতিক নির্ভরণ বিশ্লেষণের জন্য কতকগুলি অনুমান বহাল থাকতে হবে। বাস্তবক্ষেত্রে ঐ অনুমানগুলির বিঘ্ন ঘটতে পারে। বিশেষ করে বিচ্যুতি পরিমিত বিন্যাস অনুসরণ না করে অন্য কোনো বিন্যাস অনুসরণ করতে পারে যা অজানা। সেরূপক্ষেত্রে নির্ভরণ সহগসমূহ নিরূপণের জন্য সাধারণ ন্যূনতম বর্গপদ্ধতি প্রয়োগ না করে অপরামাত্রিক নির্ভরণ বিশ্লেষণ করা হয়। বর্তমান অনুচ্ছেদে সরলরৈখিক নির্ভরণের ক্ষেত্রে অপরামাত্রিক বিশ্লেষণ পদ্ধতি প্রয়োগ করে ইন্টারসেপ্ট ও নির্ভরাঙ্ক নিরূপণ পদ্ধতি আলোচনা করা হবে। এ সম্পর্কে বিস্তারিত জানার জন্য Daniel (1989) পর্যালোচনা করা যেতে পারে।

ধরা যাক সরলরৈখিক নির্ভরণ প্রতিকৃতি হলো

$$y_i=\alpha+\beta x_i+e_i\,; \quad i=1,\ 2,\ \ldots,\ n$$

এই প্রতিকৃতি মিল করার জন্য $(x_1,\ y_1),\ (x_2,\ y_2),\ \ldots\ (x_n,\ y_n)$ n যুগ্ম মান আছে এবং বিবেচনা করা যাক যে $x_1<x_2<\cdots<x_n$। এখানে অনুমান করতে হবে যে $e_i(i=1,\ 2,\ \ldots,\ n)$ সমূহ পরস্পর অনপেক্ষ। এছাড়া e_i সম্পর্কে আরো দুটি অনুমান করা যেতে পারে। যেমন: (i) e_i প্রতিসম বিন্যাস অনুসরণ করে এবং (ii) e_i প্রতিসম বিন্যাস অনুসরণ করে না।

বিচ্যুতি e_i সম্পর্কে অনুমান (i) বা (ii) যাই হোক না কেন নির্ভরাঙ্ক β-এর নিরূপকে কোনো পরিবর্তন হয় না। এই নিরূপক নির্দেশ করেছেন Theil (1950)। তাঁর নির্দেশিত পদ্ধতিতে প্রথমে

$$S_{ij}=(y_j-y_i)/(x_j-x_i)\ ,\ i<j=1,\ 2,\ \ldots\ ;\ n$$

নির্ণয় করতে হয়। নমুনা আকার n হলে $\binom{n}{2}$ সংখ্যক S_{ij} পাওয়া যায়। সেক্ষেত্রে β-এর নিরূপক b হলো

$$b=S_{ij}\text{-এর মধ্যমা}$$

এই b-এর মান প্রয়োগ করে (y_i-bx_i) $(i=1,\ 2,\ \cdots,\ n)$ পাওয়া যায়। বিচ্যুতির জন্য অনুমান (ii)-এর ক্ষেত্রে α-এর নিরূপক a হলো

$$a=(y_i-bx_i)\text{-এর মধ্যমা}$$

অনুমান (i)-এর ক্ষেত্রে (y_i-bx_i)-এর সকল মানকে প্রথমে মানের ক্রম অনুসারে সাজাতে হয় এবং ঐ মানগুলির জোড়াসমূহের গড় নির্ণয় করতে হয়। (y_i-bx_i) এর মান এবং জোড়াসমূহের গড়সহ মোট $n(n+1)/2$ মান পাওয়া যায়। সেক্ষেত্রে

$$a=n(n+1)/2\text{ মানের মধ্যমা।}$$

Dietz (1989) a-এর মান নির্ণয় পদ্ধতি আলোচনা করেছেন। একটি উদাহরণের সাহায্যে a ও b-এর মান নির্ণয় পদ্ধতি ব্যাখ্যা করা যাক।

উদাহরণ ৬.৮

নিচে এক ধরনের কীটের Body weight (y in gms) এবং heart beat (x per minute) দেয়া হলো:

y :	5.5	8.3	8.8	11.3	14.1	15.7
x :	10	15	12	11	12	9

উক্ত উপাত্ত ব্যবহার করে অপরামাত্রিক নির্ভরণ বিশ্লেষণের মাধ্যমে x-এর উপর y-এর একটি সরলরৈখিক নির্ভরণ মিল করা যাক।

উক্ত উপাত্তের ক্ষেত্রে $S_{ij} = (y_j - y_i)/(x_j - x_i)$, $i < j = 1, 2, \cdots, n$-এর মান গুলিকে মানের ক্রম অনুসারে সাজিয়ে নিচে লেখা হলো :

$$-10.20, 2.90, -2.50, -2.30, -2.20, -1.23, -0.75, -0.40,$$
$$-0.17, 0.56, 1.40, 1.65, 2.87, 5.3, 5.8$$

$\therefore \quad b = S_{ij}$-এর মধ্যমা $= -0.40$

এখন $(y_i - bx_i)$-এর মানসমূহ হলো নিম্নরূপ :

$$9.5, 13.6, 14.3, 15.7, 19.3, 19.3$$

এই মানগুলিসহ এগুলির জোড়ায় জোড়ায় গড় মানকে মানের ক্রম অনুসারে সাজিয়ে পাওয়া যায়

9.5, 11.55, 11.90, 12.6, 13.60, 13.95, 14.30, 14.40, 14.40, 14.65
15.0, 15.7, 16.45, 16.45, 16.80, 16.80, 17.5, 17.5, 19.3, 19.3, 19.3

বিচ্যুতি প্রতিসম বিন্যাস অনুসরণ করলে

$$a = \text{উপরিউক্ত } n(n+1)/2\text{-এর মধ্যমা} = 15.00$$

বিচ্যুতি প্রতিসম বিন্যাস অনুসরণ না করলে

$$a = (y_i - bx_i)\text{-এর মধ্যমা} = 15.00$$

সুতরাং মিল করা নির্ভরণ রেখা হলো

$$\hat{y} = 15.0 - 0.40x$$

৬.১৭ কালক্রমিক নির্ভরণ (Periodic Regression)

জীববিজ্ঞানে অনেক ঘটনা একটি নির্দিষ্ট সময়ের মধ্যে ঘটে থাকে এবং ঘটে যাওয়া ফলাফল একটি নির্দিষ্ট সময় অন্তর অন্তর নথিভুক্ত করা গবেষকের জন্য জরুরি হয়ে পড়ে। যেমন, গবেষণাগারে পোকামাকড় মারার জন্য কোনো ওষুধ প্রয়োগ করে পরীক্ষা

পরিচালনা করা হলে ঐ পরীক্ষায় পোকামাকড়ের মৃত্যু দিনের একটি নির্দিষ্ট সময়ের মধ্যে ঘটে যেতে পারে। সেক্ষেত্রে কতক্ষণ পরপর কি পরিমাণ পোকা মারা যাচ্ছে তা লক্ষ্য করার বিষয়, কারণ ঐ তথ্য হতে ওষুধের কার্যকারিতা সম্পর্কে সিদ্ধান্ত নেয়া যাবে। আবার কোনো কীটপতঙ্গের ডিম হতে বাচ্চা হওয়ার একটি নির্দিষ্ট সময়ের মধ্যে ঐগুলি বয়:প্রাপ্ত হয়ে যায়। ঐ ক্ষেত্রে ঐ কীট-পতঙ্গের শারীরিক পরিবর্তন ঐ নির্দিষ্ট সময়ের মধ্যে কিভাবে হচ্ছে তা লক্ষ্য করার বিষয়। চিকিৎসা শাস্ত্রে জীবন সঙ্কটাপন্ন রোগীকে একটি নির্দিষ্ট সময় অন্তর অন্তর নিরীক্ষা করা হয় এবং সেভাবে চিকিৎসার ব্যবস্থা করা হয়। এখানে প্রতিক্ষেত্রেই সময়ের বিপরীতে পরীক্ষালব্ধ ফলাফল গবেষণার বিষয় এবং সময় কিভাবে ঐ ফলাফলকে প্রভাবিত করে তা পর্যালোচনা করা গুরুত্বপূর্ণ বিষয়।

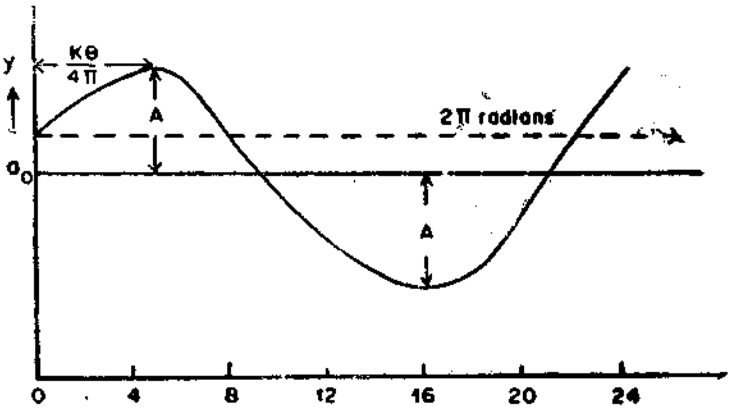
উপরিউক্ত ঘটনাসমূহ একটি নির্দিষ্ট কালচক্র (cycle or period) অনুসরণ করে এবং এই ঘটনাসমূহের ফলাফল বিশ্লেষণ করার একটি পদ্ধতি হলো ফলাফলকে সময়ের বিপরীতে প্রতিস্থাপন করে চিত্র আঁকা। এরূপ উপাত্তের গাণিতিক বিশ্লেষণের একটি পদ্ধতি হলো কালক্রমিক নির্ভরণ। এই নির্ভরণের উদ্দেশ্য হলো সময়ের পরিবর্তনের সাথে জীবন সংক্রান্ত নিরীক্ষার ফলাফল পরিবর্তনের সংশ্লেষণ পর্যালোচনা করা।

জীববিজ্ঞানে কালক্রমিক উপাত্তের বিশ্লেষণ করার অনেক পদ্ধতিই আছে। এই বিশ্লেষণের সাথে তিনটি পরামান জড়িত: (১) কালচক্রের দৈর্ঘ্য, (২) কালচক্রের দৈর্ঘ্য -এর Amplitude বা প্রাপ্ত ফলাফলের পরিসর, এবং (৩) Phase angle বা কালচক্রের মধ্যে সর্বোচ্চ ফলাফল প্রাপ্তির সময়ের কৌণিক মান। কালচক্র যাই হোক না কেন, এরূপ উপাত্ত বিশ্লেষণের জন্য সাধারণত Harmonic বিশ্লেষণ এবং Fourier series বিশ্লেষণ করা হয়। বর্তমান অনুচ্ছেদে এরূপ বিশ্লেষণের জন্য Sine curve fit করার পদ্ধতি ব্যাখ্যা করা হবে। অন্যান্য বিশ্লেষণ পদ্ধতি বা এসম্পর্কে বিস্তারিত জানার জন্য Bliss (১৯৭০) পর্যালোচনা করা যেতে পারে।

ধরা যাক t-তম পিরিয়ড়ে $[t=0,1,2,\cdot\cdot,k-1]$ কোনো পরীক্ষালব্ধ ফলাফলের মান চলক y দ্বারা পরিমাপ করা হয়। মনে করা যাক k পিরিয়ডের প্রতি পিরিয়ড়ে y-এর f সংখ্যক মান পাওয়া যায়। তাহলে y-এর মোট মানের সংখ্যা হবে $N=fk$। এই fk মানের গড় ধরা যাক a_0। সেক্ষেত্রে t-এর প্রতিমানের জন্য y-এর প্রত্যাশিত মানকে নিম্ন সমীকরণ দ্বারা চিহ্নিত করা যায়

$$y=a_0+A\cos(Ct-\theta) \qquad (৬.১৭.১)$$

এখানে $C=2\pi/k$ এবং $2\pi=360^\circ=6.283185$ এবং $t=0,1,2,\cdots,(k-1)$ পিরিয়ডিক মানকে C দ্বারা কৌণিক মানে পরিবর্তন করা হয়, θ হলো রেডিয়ানে পরিমাপ করা Phase angle বা $t_0=0$ সময়ে y-এর মান পরিমাপ করা শুরু করে যে সময়ে y-এর মান সর্বোচ্চ হয় তা। নিচে Sine curve-এর একটি চিত্র দেয়া হলো।



চিত্র ৬.১৮ : সাইন কার্ভ।

সমীকরণ (৬.১৭.১)-কে লেখা যায়

$$y = a_0 + a_1u_1 + b_1v_1 \tag{৬.১৭.২}$$

এখানে $\Sigma u_1 = \Sigma v_1 = \Sigma u_1v_1 = 0$

এবং $u_1 = \cos\dfrac{360\,t}{k}$, $v_1 = \sin\dfrac{360\,t}{k}$, $t = 0,1,2,\ldots,(k-1)$

এই সমীকরণ হতে পাওয়া যায়

$$a_1 = \frac{\Sigma u_1 y}{\Sigma u_1^2} = \frac{\Sigma u_1 y}{\frac{1}{2}fk} = \frac{\Sigma u_1 T_t}{\frac{1}{2}fk}$$

$$b_1 = \frac{\Sigma v_1 y}{\Sigma v_1^2} = \frac{\Sigma v_1 y}{\frac{1}{2}fk} = \frac{\Sigma v_1 T_t}{\frac{1}{2}fk}$$

এখানে $T_t = \sum_{1}^{f} y$; $t = 0,1,2,\ldots,k-1$

= যে কোনো পিরিয়ডে y-সমূহের যোগফল।

যখন $f = 1$; $T_t = y$ এবং

$$a_1 = \frac{\Sigma u_1 y}{\frac{1}{2}k},\ b = \frac{\Sigma v_1 y}{\frac{1}{2}k}$$

$$A = \sqrt{a_1^2 + b_1^2}$$

এই সমীকরণ মিল করে y-এর গড় পরিসর নির্ণয় করা হয় 2A দ্বারা এবং y-এর সর্বোচ্চ মান পাওয়ার পিরিয়ড হলো

$$t_{max} = \frac{k\,\theta}{2\pi}$$

এখানে $$\tan \theta' = \left|\frac{b_1}{a_1}\right| \text{ radians, এবং}$$

$$\begin{aligned}
\theta &= \pi - \theta', \text{ যদি } a_1 - Ve \text{ এবং } b_1 + Ve \text{ হয়} \\
&= \pi + \theta', \text{ যদি } a_1 - Ve \text{ এবং } b_1 - Ve \text{ হয়} \\
&= 2\pi - \theta', \text{ যদি } a_1 + Ve \text{ এবং } b_1 - Ve \text{ হয়} \\
&= \theta' \quad, \text{ যদি } a_1 + Ve \text{ এবং } b_1 + Ve \text{ হয়}
\end{aligned}$$

ধরা যাক k পিরিয়ডের $[t = 0, 1, 2, \cdots, k-1]$ সময়ের মান হলো x_0, x_1, $x_2, \cdots, x_{k-1}$। তাহলে, y-এর সর্বোচ্চ মান পাওয়ার সময় হবে

$$x_{max} = x_0 + i\, t_{max}$$

এখানে $$i = (x\text{-এর কালচক্রের দৈর্ঘ্য})/k$$

Sine curve প্রতিসম হওয়ার কারণে y-এর ন্যূনতম মান হবে কালচক্রের দৈর্ঘ্যের অর্ধেকের আগে বা পরে, অর্থাৎ কালচক্রের দৈর্ঘ্য 24 ঘণ্টা হলে y-এর ন্যূনতম মান হবে $(x_{max} - 12)$ ঘন্টা পরে বা আগে।

উপরিউক্ত Sine curve মিল করার ক্ষেত্রে সাধারণত বিবেচনা করা হয় যে, কালচক্র হবে k সংখ্যক সমান পিরিয়ডে বিভক্ত। এই Sine curve-কে আরো বর্ধিত আকারে নিম্নরূপভাবে প্রকাশ করা যায় :

$$y = a_0 + a_1u_1 + b_1v_1 + a_2u_2 + b_2v_2 + \cdots \qquad (৬.১৭.৩)$$

যেখানে $$u_1 = \frac{\cos(i\ 360\ t)}{k}, \quad v_1 = \frac{\sin(i\ 360\ t)}{k}$$

$$a_1 = \frac{\Sigma u_1\ T_t}{\frac{1}{2} fk}, \quad b_1 = \frac{\Sigma v_1\ T_t}{\frac{1}{2} fk}$$

পরপর দুই কালচক্রের ক্ষেত্রে

$$y = a_0 + a_1u_1 + b_1v_1 + a_2u_2 + b_2v_2 \qquad (৬.১৭.৪)$$

এই কার্ভ-এর বিশ্লেষণ করার জন্য ভেদাঙ্ক বিশ্লেষণ পদ্ধতি প্রয়োগ করা যায়। সেক্ষেত্রে

$$S_1 = SS\,(a_1 \text{ and } b_1) = \frac{(\Sigma u_1\ T_t)^2 + (\Sigma v_1\ T_t)^2}{\frac{1}{2} f k}$$

$$S_2 = SS\,(a_2 \text{ and } b_2) = \frac{(\Sigma u_2 T_t)^2 + (\Sigma v_2\ T_t)^2}{\frac{1}{2} fk}$$

$$S_3 = SS\ (\text{বিক্ষিপ্ততা}) = \frac{\Sigma T_i^2}{f} - C.T - S_1 - S_2$$

$$\text{মোট বর্গসমষ্টি},\ TSS = \Sigma y^2 - C.T, \quad C.T = \frac{(\Sigma y)^2}{fk}$$

$$SS\ (\text{বিচ্যুতি}) = S_4 = TSS - S_1 - S_2 - S_3$$

ভেদাঙ্ক বিশ্লেষণ সারণি

ভেদের উৎস	স্বাধীনতার মাত্রা, d.f.	বর্গসমষ্টি SS	গড় বর্গসমষ্টি MS	নির্ণেয় F
a_1 and b_1	2	S_1	s_1	s_1/s_4
a_2 and b_2	2	S_2	s_2	s_2/s_4
বিক্ষিপ্ততা	$k-5$	S_3	s_3	s_3/s_4
বিচ্যুতি	$k(f-1)$	S_4	s_4	
মোট	$fk-1$			

উদাহরণ ৬.৯

একটি পরীক্ষাগারে Gramaxon প্রয়োগ করে Gravid females of P. Scaber-কে মারার জন্য পরীক্ষা পরিচালনা করা হয়। উক্ত পরীক্ষায় 500 ppm এর herbicide প্রয়োগ করা হয় প্রতি পাঁচটি Gravid female-এর উপর। এই পাঁচটি female-কে রাখা হয় একটি Petri dish-এ। পরীক্ষায় পাঁচটি Petri dish ব্যবহার করা হয়। পরীক্ষা শুরু হওয়ার পর থেকে 24 ঘন্টা পরপর Wood lice মারা যাওয়ার সংখ্যা নথিভুক্ত করা হয় এবং এই নথিভুক্ত করার কাজ 14 দিন চলতে থাকে। নিচে প্রতি দিনের পর মারা যাওয়া Wood lice-এর সংখ্যা সারণি ৬.২১-এ দেখানো হলো

সারণি ৬.২১ : Gramaxon প্রয়োগে প্রতি 24 ঘন্টা পরপর মৃত Gravid females of P. Scaber-এর সংখ্যা।

পুনরায়ন	প্রতিদিন পর মৃত Wood lice-এর সংখ্যা													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	0	1	1	0	0	0	0	1	0	0	0	1	0	0
2	0	0	1	0	0	0	0	1	0	0	0	0	1	1
3	1	0	1	0	0	0	1	0	0	0	0	0	0	1
4	0	0	1	0	1	0	0	0	0	0	0	0	0	0
5	0	0	1	0	1	0	0	0	0	0	0	1	0	0
মোট T_i	1	1	5	0	2	1	1	2	0	0	0	2	1	2

উপরিউক্ত উপাত্তের ভিত্তিতে সর্বোচ্চ সংখ্যক Wood lice মারা যাওয়ার ন্যূনতম দিন নির্ণয় করা যাক।

এই উপাত্তের ক্ষেত্রে $t=0,1,2,\ldots 13$; $f=5$, $k=14$ $N=fk=70$।
এখানে u_1 ও v_1-এর মান নিচে দেখানো হলো।

t	$u_1=\cos\left[\frac{i\,360\,t}{k}\right]$	$v_1=\sin\left[\frac{i\,360\,t}{k}\right]$	t	u_1	v_1
0	1.000	0	7	−1.000	0
1	0.901	0.434	8	−0.901	−0.434
2	0.623	0.782	9	−0.623	−0.782
3	0.222	0.975	10	−0.222	−0.975
4	−0.222	0.975	11	0.222	−0.975
5	−0.623	0.782	12	0.623	−0.782
6	−0.901	0.434	13	0.901	−0.434

উপরিউক্ত u_1 ও v_1-এর ভিত্তিতে মৃত Wood lice-এর সংখ্যার জন্য Sine curve হলো

$$y=a_0+a_1u_1+b_1v_1$$

যেখানে
$$a_1=\frac{\Sigma u_1T_t}{\frac{1}{2}fk}=\frac{2\times3.917}{5\times14}=0.1119$$

$$b_1=\frac{\Sigma v_1T_t}{\frac{1}{2}fk}=\frac{2\times3.91}{5\times14}=0.1117$$

$$a_0=\frac{\Sigma T_t}{fk}=\frac{18}{5\times14}=0.2571$$

সুতরাং মিল করা Sine curve হলো

$$y=0.2571+0.111\,u_1+0.1117\,v_1$$

$$\tan\theta'=\left|\frac{b_1}{a_1}\right|=\left|\frac{0.1117}{0.1119}\right|=0.9982\ \text{radians}$$

এখানে $\theta'=\theta$, কারণ a_1 ও b_1-এর মান ধনাত্মক $(+Ve)$

$\therefore$ $\theta=\tan^{-1}0.9982\ \text{radians}=0.7845$

এখন সর্বোচ্চ সংখ্যক Wood lice মারার জন্য ন্যূনতম দিন হলো

$$t_{max} = \frac{k\,\theta}{2\pi} = \frac{14 \times 0.7845}{6.283185} = 1.75$$

এই সময়কে দিনে রূপান্তরিত করে পাওয়া যায়

$$\begin{aligned} x_{max} &= x_0 + i\, t_{max} \\ &= 1 + 1 \times 1.75 \quad \because i = 1 \\ &= 2.75 \end{aligned}$$

অর্থাৎ, গড়ে তৃতীয় দিনের মধ্যেই বেশিরভাগ Wood lice মারা যায়।

এখন $$SS(a_1 \text{ and } b_1) = \frac{(\Sigma u_1 T_t)^2 + (\Sigma v_1 T_t)^2}{\frac{1}{2} fk}$$

$$= \frac{2[(3.917)^2 + (3.91)^2]}{5 \times 14} = 0.8752$$

$$C.T = \frac{(\Sigma y)^2}{fk} = \frac{(18)^2}{5 \times 14} = 4.6286$$

মোট বর্গসমষ্টি, $TSS = \Sigma y^2 - C.T. = 18 - 4.6286 = 13.3714$

$$S_2 = SS\ (\text{বিক্ষিপ্ততা}) = \frac{\Sigma T_t^2}{f} - C.T - S_1 = \frac{46}{5} - 4.6286 - 0.8752$$

$$= 3.6962$$

$$SS\ (\text{বিচ্যুতি}) = TSS - S_1 - S_2 = 13.3714 - 0.8752 - 3.6962 = 8.8$$

ভেদাঙ্ক বিশ্লেষণ সারণি

ভেদের উৎস	স্বাধীনতার মাত্রা, d.f.	বর্গসমষ্টি SS	গড় বর্গসমষ্টি $MS = \frac{SS}{d.f}$	নির্ণেয় F	$F_{0.05}$
a_1 and b_1	2	0.8752	0.4376	2.88	3.16
বিক্ষিপ্ততা	9	3.6962	0.4107	2.71	2.05
বিচ্যুতি	58	8.8	0.1517		
মোট	69				

এই বিশ্লেষণ হতে লক্ষণীয় যে, উপরিউক্ত উপাত্তের ক্ষেত্রে Simple sine curve উপযুক্ত নয় [কারণ MS(a_1 and b_1) তাৎপর্যহীন]।

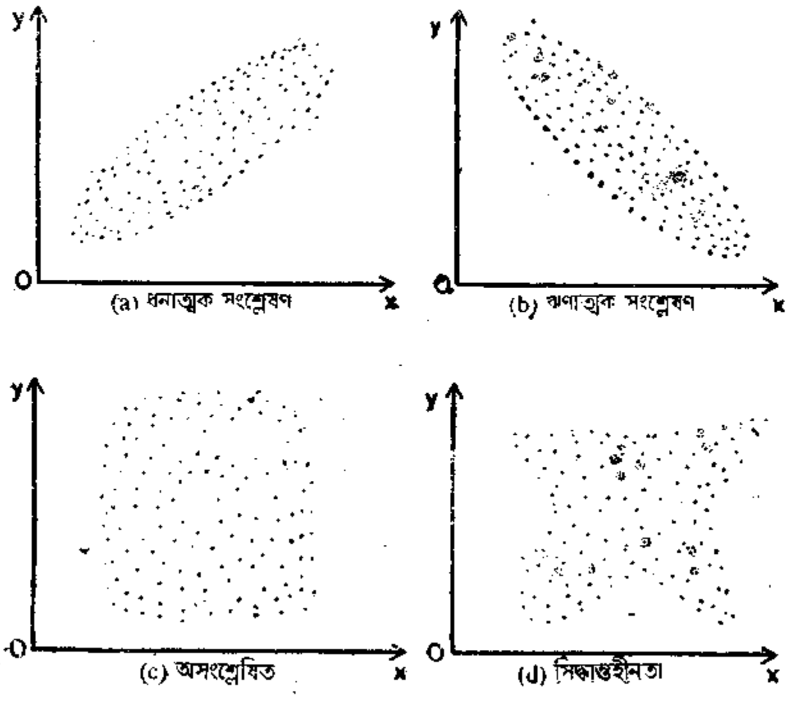
৬.১৮ সংশ্লেষণ (Correlation)

দুই বা ততোধিক চলকের সম্পর্ক পর্যালোচনা করার জন্য পরিসংখ্যানিক পদ্ধতি হলো নির্ভরণ। এ সম্বন্ধে পূর্ববর্তী অনুচ্ছেদসমূহে আলোচনা করা হয়েছে। এই সম্পর্কের মাত্রা নির্ণয় পদ্ধতি হলো সংশ্লেষণ বিশ্লেষণ। নির্ভরণ বিশ্লেষণ আলোচনা করার সময় চলকসমূহের সম্পর্ক কি পরিমাণ শক্তিশালী তা আলোচনা করতে গিয়ে বা নির্ভরণ রেখার মূল্যায়ন করতে গিয়ে সিদ্ধান্ত সহগ (coefficient of determination, R^2)-এর উল্লেখ করা হয়েছে। এখানে R মূলত চলকসমূহের রৈখিক সম্পর্কের মাত্রা বা চলকসমূহের সংশ্লেষণ পরিমাপ করে। এই R-কে বলা হয় বহুল সংশ্লেষক (multiple correlation coefficient) । এখানে সরল সংশ্লেষণ এবং সংশ্লেষাঙ্ক নিয়ে আলোচনা করা হবে।

দুই চলকের মধ্যে সম্পর্ক সরল সংশ্লেষণ হিসেবে পরিচিত এবং এই সম্পর্কের মাত্রা নির্ণয় করার জন্য ব্যবহৃত তথ্যজমান মান হলো সংশ্লেষাঙ্ক (correlation coefficient) এবং একে r দ্বারা চিহ্নিত করা হয়। সংশ্লেষণ ধনাত্মক বা ঋণাত্মক উভয়ই হতে পারে। ধনাত্মক সংশ্লেষণ দ্বারা বুঝা যায় একটি চলকের মান বৃদ্ধি পেলে অপর চলকের মানও বৃদ্ধি পাবে। আবার কোনো চলকের মান বৃদ্ধি পেলে তার সাথে সম্পর্কিত চলকের মান যদি কমতে থাকে, তাহলে সংশ্লেষণ ঋণাত্মক হয়। যেমন, কৃষি জমিতে একটি নির্দিষ্ট সীমার মধ্যে Pesticide-এর পরিমাণ বৃদ্ধি করা হলে পোকা-মাকড়ের আক্রমণ কমতে থাকে। এখানে Pesticide-এর পরিমাণ এবং পোকামাকড়ের পরিমাণ ঋণাত্মকভাবে সংশ্লেষিত বলা যায়। আবার জমিতে একটি নির্দিষ্ট সীমার মধ্যে সারের পরিমাণ বৃদ্ধি করা হলে কৃষি উৎপাদনের পরিমাণ বৃদ্ধি পাবে। এখানে সারের পরিমাণ এবং কৃষি উৎপাদনের পরিমাণ ধনাত্মকভাবে সংশ্লেষিত। এখানে সারের নির্দিষ্ট সীমার কথা উল্লেখ করা হয়েছে এ কারণে যে, সারের পরিমাণ সীমাহীনভাবে বৃদ্ধি করলে উৎপাদনের পরিমাণ সীমাহীনভাবে বৃদ্ধি পাবে না বরং কৃষি উৎপাদন ব্যাহত হতে পারে।

উপরিউক্ত আলোচনা হতে বুঝা যায় যে, সংশ্লেষণ দ্বি-চলক (bivariate) তথ্য মানের ক্ষেত্রে প্রযোজ্য এবং ঐ দ্বি-চলক একই নমুনা বিন্দু হতে পর্যালোচিত। এছাড়া ঐ দ্বি-চলকের বিন্যাস হতে হবে দ্বি-চলক পরিমিত বিন্যাস (bivariate normal distribution) । ধরা যাক দ্বি-চলকের একটি হলো x এবং অপরটি y । এখানে x ও y উভয়েই দৈব চলক হওয়াতে নির্ভরণের ন্যায় একটিকে অনপেক্ষ চলক এবং অপরটিকে নির্ভরশীল চলক হিসেবে চিহ্নিত করা হয় না। সে কারণেই একটি চলক অপরটি দ্বারা প্রভাবিত এমন বিবেচনা করা হয় না বরং চলকদ্বয় সংশ্লেষিত (correlated) বিবেচনা করা হয়।

দ্বি-চলক পরিমিত বিন্যাসের পাঁচটি পরামানের একটি হলো P, যেখানে P হলো x ও y-এর গণসমষ্টি সংশ্লেষাঙ্ক। চলক x ও y-এর গণসমষ্টি সিদ্ধান্ত সহগকে P^2 দ্বারা চিহ্নিত করা হয়। সে কারণে P হলো P^2-এর বর্গমূল। আগেই আলোচনা করা হয়েছে যে সিদ্ধান্ত সহগের মান 0 থেকে 1 পর্যন্ত হয়। কাজেই P-এর মান -1 থেকে $+1$ পর্যন্ত হয়। যদি $P = +1$ হয়, তাহলে বুঝা যায় চলক x ও y সঠিক-ভাবে ধনাত্মকভাবে এবং রৈখিকভাবে সংশ্লেষিত। অর্থাৎ x-এর মান যে হারে বাড়ে বা কমে ঠিক সে হারেই y-এর মান বাড়ে বা কমে। আবার $P = -1$ হওয়ার অর্থ হলো x-এর মান যে হারে বাড়ে (কমে) ঠিক সে হারেই y-এর মান কমে (বাড়ে)। এক্ষেত্রে x ও y-এর সম্পর্ক সঠিকভাবে এবং উল্টাভাবে (inversely) সংশ্লেষিত। যদি $P = 0$ হয়, তাহলে x ও y চলকদ্বয় অসংশ্লেষিত এবং রৈখিক সম্পর্কের ক্ষেত্রে $P = 0$ হলে x ও y অনপেক্ষ (independent) হিসেবে বিবেচিত হয়। সংশ্লেষাঙ্ক-এর চিহ্ন হবে x ও y চলকের সহ-ভেদাঙ্কের (covariance) চিহ্ন। এখানে সহ-ভেদাঙ্ক হলো



চিত্র ৬.১৮ : দ্বি-চলক বিন্যাসের বিক্ষেপ চিত্র।

$$Cov(x,y) = \frac{1}{N}\left[\Sigma xy - \frac{\Sigma x \Sigma y}{N}\right]$$

এবং N = গণসমষ্টি তথ্যমানের সংখ্যা।

গণসমষ্টি সংশ্লেষাঙ্ক P-এর নিরূপক হলো r। এই r-এর মান নমুনা তথ্যমান হতে নির্ণয় করা হয় এবং এর মানও -1 থেকে $+1$ পর্যন্ত হয়। সংশ্লেষাঙ্ক r-এর সঠিক মান গাণিতিক সূত্রের সাহায্যে নির্ণয় করা হয়। তবে r-এর মান সম্পর্কে একটি ধারণা লাভ করা যায় বিক্ষেপ চিত্র হতে। পূর্ব পৃষ্ঠায় চারটি বিক্ষেপ চিত্র অঙ্কন করে দেখানো হলো।

চিত্র (৬.১৮a) হতে লক্ষ্য করা যাচ্ছে যে, x-এর মান বাড়ার সাথে y-এর মান বাড়ছে। এক্ষেত্রে x ও y ধনাত্মকভাবে সংশ্লেষিত। চিত্র (৬.১৮b) নির্দেশ করছে যে, x-এর মান বাড়লে y-এর মান কমে। কাজেই x ও y ঋণাত্মকভাবে সংশ্লেষিত। চিত্র (৬.১৮c) x-এর মানের পরিবর্তনের সাথে y-এর মানের পরিবর্তনের কোনো স্বাভাবিক গতিধারা বুঝা যাচ্ছে না। এক্ষেত্রে x ও y অসংশ্লেষিত। কিন্তু চিত্র (৬.১৮d) হতে x ও y-এর সম্পর্ক সম্বন্ধে সিদ্ধান্ত নেয়া সহজ নয়।

বিক্ষেপ চিত্র হতে সংশ্লেষণ সম্বন্ধে ধারণা লাভ করা গেলেও সংশ্লেষাঙ্ক-এর মান সম্বন্ধে ধারণা করা যায় না। এই মান গাণিতিক সূত্রের সাহায্যে নির্ণয় করতে হয়। বিক্ষেপ চিত্র হতে একটি ব্যাপার লক্ষণীয়, তাহলো সংশ্লেষণ x ও y-এর যৌথ ভেদের একটি বহিঃপ্রকাশ। এই যৌথ ভেদ পরিমাপ করা হয় সহ-ভেদাঙ্ক (Covariance) দ্বারা। নমুনা উপাত্তের ক্ষেত্রে

$$Cov(x,y) = \frac{1}{n-1}\Sigma(x-\bar{x})(y-\bar{y}) = \frac{1}{n-1}\left[\Sigma xy - \frac{\Sigma x \Sigma y}{n}\right]$$

$$= \frac{1}{n-1} SP(xy)$$

সুতরাং সংশ্লেষাঙ্ক-এর একটি পরিমাপ হিসেবে x ও y-এর সহ-ভেদাঙ্ক নির্ণয় করা যায়। কিন্তু এই পরিমাপের একটি অসুবিধা হলো যে, এর মান x ও y চলকদ্বয়ের এককের (unit) উপর নির্ভরশীল। ফলে সংশ্লেষাঙ্ক নির্ণয় করে ভিন্ন এককের দুই বা ততোধিক গণসমষ্টির সংশ্লেষাঙ্ক তুলনা করা যায় না। এছাড়া পূর্বে উল্লিখিত P^2-এর মান 0 হতে 1 পর্যন্ত হয় না। এ কারণে সংশ্লেষাঙ্ক r-এর মান নির্ণয় করার জন্য Cov(x, y)-এর মানকে x ও y-এর পরিমিত ব্যবধান (standard deviation) দ্বারা ভাগ করা হয়। সুতরাং

$$r = \frac{Cov(xy)}{s_x \; s_y}$$

এখানে

$$s_x = \sqrt{\left[\frac{1}{n-1}\Sigma(x-\bar{x})^2\right]} \quad \text{এবং} \quad s_y = \sqrt{\left[\frac{1}{n-1}\Sigma(y-\bar{y})^2\right]}$$

বাস্তব ক্ষেত্রে r নির্ণয় করার সহজ সূত্র হলো

$$r = \frac{\Sigma xy - \frac{\Sigma x \Sigma y}{n}}{\sqrt{\left[\Sigma x^2 - \frac{(\Sigma x)^2}{n}\right]\left[\Sigma y^2 - \frac{(\Sigma y)^2}{n}\right]}}$$

সরল নির্ভরণের ক্ষেত্রে লক্ষ্য করা হয়েছে যে

$$r^2 = \frac{SSR}{SST} = \frac{b\,SP(xy)}{\Sigma y^2 - \frac{(\Sigma y)^2}{n}}$$

এখানে $SP(xy) = \Sigma xy - \frac{\Sigma x \Sigma y}{n}$, $b = \frac{SP(xy)}{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}$

কাজেই $$r^2 = \frac{b^2[\Sigma x^2 - (\Sigma x)^2/n]}{\Sigma y^2 - (\Sigma y)^2/n}$$

এক্ষেত্রে $r = \sqrt{r^2}$ এবং r-এর চিহ্ন হবে SP(xy)-এর চিহ্ন। আবার,

$$b_{yx} = \frac{SP(xy)}{\Sigma x^2 - \frac{(\Sigma x)^2}{n}} \quad \text{এবং} \quad b_{xy} = \frac{SP(xy)}{\Sigma y^2 - \frac{(\Sigma y)^2}{n}}$$

এখানে b_{yx} এবং b_{xy} হলো যথাক্রমে x-এর উপর y-এর নির্ভরাঙ্ক এবং y-এর উপর x-এর নির্ভরাঙ্ক। সুতরাং r^2-কে লেখা যায়

$$r^2 = b_{yx} \cdot b_{xy} = \frac{[SP(xy)]^2}{\left[\Sigma x^2 - \frac{(\Sigma x)^2}{n}\right]\left[\Sigma y^2 - \frac{\Sigma y^2}{n}\right]}$$

এই r-কে বলা হয় প্রোডাক্ট মোমেন্ট সংশ্লেষাঙ্ক।

অনেক সময় r-এর মান নির্ণয় করার ক্ষেত্রে চলক x ও y-এর মানসমূহকে সংক্ষিপ্ত করার জন্য এগুলির উপর একটি পরিবর্তন করা হয়। ধরা যাক পরিবর্তিত চলকদ্বয় হলো u এবং v, যেখানে

$$u = \frac{x-A}{h} \quad \text{এবং} \quad v = \frac{y-B}{k}$$

এখানে A, B, h এবং k হলো যথেচ্ছা (arbitrary) নির্বাচিত মান। দ্বি-চলক ঘটনসংখ্যা বিন্যাসের ক্ষেত্রে A এবং B হতে পারে যথাক্রমে x ও y চলকদ্বয়ের মধ্য শ্রেণী বা প্রায় মধ্য শ্রেণীর মধ্য বিন্দু (mid point) এবং h ও k হতে পারে যথাক্রমে x ও y চলকদ্বয়ের শ্রেণী ব্যবধানের পরিমাণ। এরূপ পরিবর্তন করে u ও v চলকের সংশ্লেষাঙ্ক r_{uv} নির্ণয় করা হলেই r_{xy} পাওয়া যায়। কারণ সংশ্লেষাঙ্ক আদি বিন্দু বা মাপনি দ্বারা প্রভাবিত হয় না। কাজেই

$$r_{xy}=r_{uv}=\frac{\Sigma uv-\dfrac{\Sigma u\Sigma v}{n}}{\left[\Sigma u^2-\dfrac{(\Sigma u)^2}{n}\right]\left[\Sigma v^2-\dfrac{(\Sigma v)^2}{n}\right]}$$

উদাহরণ ৬.৯

সারণি ১.৪-এ (প্রথম খণ্ড) দেয়া রোগীদের চিকিৎসার পূর্বে (B) এবং চিকিৎসার পরে (A) F.B.S. এর মধ্যে যে সংশ্লেষণ আছে তা পর্যালোচনা করা যাক।

এই পর্যালোচনার জন্য B এবং A-এর মানসমূহের সংশ্লেষাঙ্ক (r_{AB}) নির্ণয় করা যাক। এখানে

$$r_{AB}=\frac{\Sigma AB-\dfrac{\Sigma A\Sigma B}{n}}{\sqrt{\left[\left\{\Sigma A^2-\dfrac{(\Sigma A)^2}{n}\right\}\left\{\Sigma B^2-\dfrac{(\Sigma B)^2}{n}\right\}\right]}},\quad n=25$$

$$=\frac{1442823-\dfrac{5926\times5442}{25}}{\sqrt{\left[\left\{1580784-\dfrac{(5926)^2}{25}\right\}\left\{1379078-\dfrac{(5442)^2}{25}\right\}\right]}}$$

$$=\frac{152851.32}{\sqrt{176084.96\times194463.44}}$$

$$=0.83$$

লক্ষ্য করা যাচ্ছে যে, চিকিৎসার পূর্বের এবং পরের F. B. S-এর মধ্যে একটি উচ্চ সংশ্লেষণ আছে।

সংশ্লেষাঙ্ক-এর তাৎপর্য যাচাই (Test of significance of correlation cofficient) : উদাহরণ ৬.৯-এর ক্ষেত্রে উল্লেখ করা হয়েছে উচ্চ সংশ্লেষণের কথা। আগেই উল্লেখ করা হয়েছে $-1\leq r\leq 1$। স্বাভাবিকভাবেই $r=\pm 1$ বা এর কাছাকাছি হলেই উচ্চ সংশ্লেষণ বলা হয়। কিন্তু r-এর মান $|1|$-এর চেয়ে অনেক ছোট হলেও

চলকদ্বয়ের মধ্যে সংশ্লেষণ তাৎপর্যপূর্ণ হতে পারে। আবার r-এর |1|-কাছাকাছি হলে সংশ্লেষণ তাৎপর্যপূর্ণ নাও হতে পারে। তাই সংশ্লেষণ উচ্চ কি কম সিদ্ধান্ত অপেক্ষা তা তাৎপর্যপূর্ণ কিনা এ সিদ্ধান্তই প্রয়োজন। এরূপ সিদ্ধান্তের জন্যই সংশ্লেষাঙ্কের তাৎপর্য যাচাই করতে হয়।

সংশ্লেষাঙ্ক-এর তাৎপর্য যাচাই করার জন্য নাস্তি কল্পনা হলো

$$H_0: \ \rho = 0$$

এবং বিকল্প কল্পনা হলো

$$H_A: \ \rho \neq 0$$

এক্ষেত্রে অনুমান করতে হয় যে, নমুনা তথ্যমান এমন একটি দ্বি-চলক পরিমিত বিন্যাস হতে চয়ন করা হয়েছে যার গণসমষ্টি সংশ্লেষাঙ্ক ρ। এই নাস্তি কল্পনা যাচাই-এর জন্য যাচাই তথ্যজমান হলো

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

এই 't'-এর বিন্যাস হলো $(n-2)$ স্বাধীনতার মাত্রাবিশিষ্ট Student's 't' বিন্যাস। এখন 5% সংশয় মাত্রায় সিদ্ধান্ত নিতে হলে $t > t_{.025,\,n-2}$ এর ক্ষেত্রে নাস্তি কল্পনা বাতিল বলে গণ্য হবে। অন্যথায় নাস্তি কল্পনার বিপক্ষে কোনো যুক্তি নেই বলে সিদ্ধান্ত নিতে হবে।

এখানে উদাহরণ ৬.৯-এর ক্ষেত্রে সংশ্লেষাঙ্ক ρ-এর তাৎপর্য যাচাই করে দেখা যেতে পারে। উক্ত উদাহরণের ক্ষেত্রে

$$t = \frac{0.83\sqrt{25-2}}{\sqrt{1-(0.83)^2}} = 7.14$$

$t_{0.025,23} = 2.069$। $t > 2.069$ হওয়ায় নাস্তি কল্পনা বাতিল। সুতরাং চিকিৎসার পূর্বের এবং পরের F.B.S. এর সংশ্লেষণ তাৎপর্যপূর্ণ।

উপরিউক্ত t-এর সূত্র হতে লক্ষণীয় যে t-এর মান নমুনা আকার n-এর উপর অধিক নির্ভরশীল। r-এর মান বড় বা ছোট যাই হোক না কেন n খুব বড় হলে t-এর মান বড় হবে এবং 't' বর্জন ক্ষেত্রে (critical region) পড়বে। উদাহরণ হিসেবে $r = 0.9$-এর কথা বিবেচনা করা যাক। ধরা যাক এই $r = 0.9$ নির্ণয় করা হয়েছে $n = 3$ তথ্যমানের ভিত্তিতে। তাহলে $t = 2.06$ এবং এটি $t_{.025,1} = 12.706$ অপেক্ষা ছোট হওয়ায় সংশ্লেষণ তাৎপর্যহীন। আবার ধরা যাক $n = 300$ এবং এই নমুনা হতে প্রাপ্ত $r = 0.2$। এক্ষেত্রে $t = 3.52 > t_{.025,298}$ হওয়ায় সংশ্লেষণ তাৎপর্যপূর্ণ। এ আলোচনা হতে বুঝা যাচ্ছে যে r নির্ণয় করে সিদ্ধান্ত নেয়ার জন্য নমুনা আকার গুরুত্বপূর্ণ।

এতক্ষণ নাস্তি কল্পনা $H_0: \rho = 0$-এর যাচাই পদ্ধতি আলোচনা করা হয়েছে। এই নাস্তি কল্পনার অর্থ হলো চলক x ও y -এর মধ্যে কোনো সংশ্লেষণ নেই। বাস্তবে

x ও y সংশ্লেষিত এবং তাদের সংশ্লেষাঙ্কের একটি নির্দিষ্ট মান আছে বিবেচিত হতে পারে। এক্ষেত্রে নাস্তি কল্পনা এবং বিকল্প কল্পনা হলো যথাক্রমে

$$H_0: \quad \rho = \rho_0$$

এবং $$H_A: \quad \rho \neq \rho_0 \quad \text{বা} \quad [\rho < \rho_0 \quad \text{বা} \quad \rho > \rho_0]$$

এই নাস্তি কল্পনা যাচাই করার জন্য Fisher (১৯২১) পরিমিত যাচাই (normal test) পদ্ধতি নির্দেশ করেছেন। তাঁর মতে

$$z = \frac{z_r - z_\rho}{1/\sqrt{n-3}} \sim N(0,1)$$

এখানে

$$z_r = \frac{1}{2}\, ln \frac{1+r}{1-r} \quad \text{এবং} \quad z_\rho = \frac{1}{2}\, ln \frac{1+\rho_0}{1-\rho_0}$$

এই $|z| \geq 1.96$ হলে 5% সংশয় মাত্রায় নাস্তি কল্পনা বাতিল বলে গণ হবে।

উদাহরণ ৬.১০

দূষিত মাছ খাওয়ার মাধ্যমে কিছু লোক মিথাইল মারকারি গ্রহণ করেছে। এতে ঐ লোকদের রক্তে মারকারির পরিমাণ পরিবর্তিত হয়ে গেছে। নিচে একটি দ্বি-চলক ঘটন-সংখ্যা বিন্যাসের মাধ্যমে ঐ লোকদের মারকারি গ্রহণের পরিমাণ (x μg Hg/day) এবং রক্তে মারকারির পরিমাণ (y ng/g) দেখানো হলো। এই x ও y-এর সংশ্লেষাঙ্ক, $\rho = 0.5$ কিনা যাচাই করে দেখা যাক।

সারণি ৬.২১ : মিথাইল মারকারি গ্রহণ করেছে এমন লোকদের রক্তে মারকারির পরিমাণ নির্দেশকারী ঘটনসংখ্যা বিন্যাস।

মিথাইল মারকারি গ্রহণের পরিমাণ x(μg Hg/day)	রক্তে মারকারির পরিমাণ (y ng/g) 60 – 110	110 – 160	160 – 210	210 – 260	260 – 310	মোট f_x
100 – 150	9	2	1			12
150 – 200	4	5	2			11
200 – 250	2	8	6			16
250 – 300	4	6	3	2	3	18
300 – 350		3	5	4	10	22
350 – 400		7	2	8	4	21
মোট f_y	19	31	19	14	17	100

সারণি ৬.২২ : x ও y এর সংশ্লেষাঙ্ক নির্ণয়ের কিছু হিসাব। x এর $u = \dfrac{X-275}{50}$ y এর মধ্যবিন্দু।

মধ্যবিন্দু X		Y 85	135	185	235	285		
	$v=\dfrac{Y-185}{50}$	−2	−1	0	1	2	fu	$\sum_v$fuv.uv
125	−3	9	2	1			12	60
175	−2	4	5	2			11	26
225	−1	2	8	6			16	12
275	0	4	6	3	2	3	18	0
325	1		3	5	4	10	22	21
375	2		7	2	8	4	21	18
	fv	19	31	19	14	17	100	
$\sum_u$fuv.uv		74	7	0	20	36		

$$r_{xy} = r_{uv} = \frac{\Sigma\Sigma f_{uv}.uv - \dfrac{\Sigma f_u.u \; \Sigma f_v.v}{N}}{\sqrt{\left[\left\{\Sigma f_u.u^2 - \dfrac{(\Sigma f_u.u)^2}{N}\right\}\left\{\Sigma f_v.v^2 - \dfrac{(\Sigma f_v.v)^2}{N}\right\}\right]}}$$

এখানে $N = \Sigma\Sigma f_{uv} = \Sigma f_u = \Sigma f_v = 100$

সুতরাং $$r_{uv} = \frac{137 - \dfrac{(-10)(-21)}{100}}{\sqrt{\left[\left\{274 - \dfrac{(-10)^2}{100}\right\}\left\{189 - \dfrac{(-21)^2}{100}\right\}\right]}}$$

$$= \frac{134.9}{\sqrt{273 \times 184.59}}$$

$$= 0.60$$

যাচাই করতে হবে

$$H_0 : \rho = \rho_0 = 0.50$$

এবং $$H_A : \rho \neq 0.50$$

এখন $$z_r = \frac{1}{2}\ln\frac{1+r}{1-r} = \frac{1}{2}\ln\frac{1+0.60}{1-0.60} = 0.6931$$

$$z_\rho = \frac{1}{2}\ln\frac{1+\rho_0}{1-\rho_0} = \frac{1}{2}\ln\frac{1+0.50}{1-0.50} = 0.5493$$

তাহলে $$z = \frac{z_r - z_\rho}{1/\sqrt{n-3}} = \frac{0.6931-0.5493}{1/\sqrt{100-3}} = 1.42$$

$z < 1.96$ হওয়ায় 5% সংশয় মাত্রায় নাস্তি কল্পনার বিপক্ষে কোনো প্রমাণ নেই।

উপরিউক্ত যাচাই তথ্যজমানের ভিত্তিতে ρ এর জন্য নিশ্চয়তা-পরিক্ষেপও (confidence interval) নির্ণয় করা যায়। $100(1-\alpha)\%$ নিশ্চয়তা-পরিক্ষেপ হলো

$$z_r \pm z_{\alpha/2}(1/\sqrt{n-3})$$

এখানে $z_{\alpha/2}$ হলো পরিমিত বিন্যাসের $\alpha\%$ সারণিকৃত মান। সাধারণভাবে 95% নিশ্চয়তা-পরিক্ষেপের জন্য $z_{\alpha/2} = 1.96$। তাহলে উদাহরণ ৬.১০ এর উপাত্তের ক্ষেত্রে ρ এর জন্য 95% নিশ্চয়তা-পরিক্ষেপ হলো

$$0.6931 \pm 1.96(1/\sqrt{100-3})$$

বা $$0.6931 \pm 0.1990$$

বা $$0.4941 \leqslant \rho \leqslant 0.8921$$

অনুক্রমিক সংশ্লেষণ (Rank correlation) : সংশ্লেষণ আলোচনা করতে উল্লেখ করা হয়েছে যে, চলক x ও y দ্বি-চলক পরিমিত বিন্যাস অনুসরণ করে। এই চলকদ্বয় পরিমিত বিন্যাস অনুসরণ করে বিধায় তারা পরিমাণগত (quantitative) চলক। বাস্তব ক্ষেত্রে যে কোনো নমুনা বিন্দু হতে এমন চলকের পরিমাপ করা হতে পারে যেগুলির সব কয়টি বা কিছু গুণগত (qualitative) চলক বা ঐগুলির পরিমাপ করা হয় Ordinal scale-এ। যেমন, Blood sugar level এবং Kidney সমস্যা সংশ্লেষিত কিনা তা পর্যালোচনা করার সময় Blood sugar level সংখ্যার মাধ্যমে পরিমাপ করা যায়। কিন্তু Kidney সমস্যা আছে কি নেই তা পরিমাপ করার জন্য নকল চলক [যথা সমস্যা আছে = 1 এবং সরলতা নেই = 0 ধরা যায়] ব্যবহার করতে হয়। এই নকল চলক পরিমিত বিন্যাস অনুসরণ করে না। আবার Heart সমস্যা ও ধূমপান সংশ্লেষিত কিনা তা পর্যালোচনা করার জন্য Heart সমস্যা পরি-

মাপ করার জন্য Ordinal scale বা নকল চলক ব্যবহার করতে হবে। ধূমপান পরিমাপ করার ক্ষেত্রেও নকল চলক ব্যবহার করতে হবে [যথা ধূমপান করে না = 0, মাঝে মাঝে করে = 1, নিয়মিত করে = 2 এবং খুব বেশি ধূমপান করে = 3]। এছাড়া রক্তে কোলেস্টরাল এবং আর্থিক সঙ্গতির মধ্যে সংশ্লেষণ আছে কিনা তা পর্যালোচনা করার ক্ষেত্রেও রক্তে কোলেস্টরালের পরিমাণ Interval scale-এ বা পরিমাণগতভাবে পরিমাপ করা যায়। আবার আর্থিক সঙ্গতি Interval scale এবং Ordinal scale বা পরিমাণগত ও গুণগত উভয়ভাবেই পরিমাপ করা যায়।

সংশ্লেষণ পর্যালোচনা করার জন্য যে কোনো একটি চলক Ordinal scale-এ বা গুণগত চলক হিসেবে পরিমাপ করা হলে প্রোডাক্ট মোমেন্ট সংশ্লেষাঙ্ক (r) নির্ণয় করা যায় না। সেক্ষেত্রে সংশ্লেষণ হলো অনুক্রমিক সংশ্লেষণ এবং এটি পরিমাপ করা হয় অনুক্রমিক সংশ্লেষাঙ্ক r_s দ্বারা। এই সংশ্লেষাঙ্ককে Spearman rank correlation coefficient বলা হয়। এখানে

$$r_s = 1 - \frac{6\,\Sigma d_i^2}{n(n^2-1)}$$

যেখানে d_i = i-তম নমুনা বিন্দুর অনুক্রমের পার্থক্য এবং অনুক্রম হলো প্রাপ্ত উপাত্তের মানের ক্রম সংখ্যা (rank)। নিচে অনুক্রম নির্ণয় পদ্ধতি আলোচনা করা হলো :

ধরা যাক কোনো উপাত্তে n জোড়া তথ্যমান আছে। তাহলে,

(i) যে কোনো চলকের ক্ষেত্রে সর্বনিম্নমানের অনুক্রম 1, পরবর্তী মানের অনুক্রম 2 এবং এভাবে সর্বোচ্চমানের অনুক্রম হবে n।

যেমন, x :	15	18	3	5	12	7	(n = 6)
অনুক্রম (x) :	5	6	1	2	4	3	

(ii) যে কোনো চলকের ক্ষেত্রে একই মান একাধিক হলে ঐগুলির প্রত্যেকের অনুক্রম হবে তাদের গড় অনুক্রম। গড় নির্ণয় করার আগে বিবেচনা করতে হবে যে চলকের মান ভিন্ন ভিন্ন এবং সবগুলির অনুক্রম হবে যথা নিয়মে।

যেমন, x :	8	12	2	7	9	9	12	12	(n = 8)
অনুক্রম (x) :	3	7	1	2	4.5	4.5	7	7	

এখানে 9 দুটি এবং 12 তিনটি। যথা নিয়মে একটি 9-এর অনুক্রম 4 এবং অপর 9-এর অনুক্রম 5। এই 4 এবং 5-এর গড় মান 4.5। সুতরাং 9-এর অনুক্রম 4.5। যেহেতু 9-এর অনুক্রম 4 এবং 5, সে কারণে পরবর্তী রাশি 12-এর অনুক্রম হবে 6 এবং অপর দুটি 12-এর অনুক্রম হবে 7 এবং 8। এই 6, 7 ও 8-এর গড় মান 7। সুতরাং 12-এর অনুক্রম 7।

উপরিউক্ত r_s নির্ণয় করার ক্ষেত্রে দুটি বিষয় লক্ষ্য রাখা উচিত। এগুলি হলো
(i) n-এর মান 7-এর ছোট হওয়া উচিত নয়
(ii) কোনো চলকের একই মান [tied] n/2-এর বেশি হওয়া উচিত নয়।

উদাহরণ ৬.১১

একটি পরীক্ষাগারে কিছু Wood lice এনে স্বাভাবিক অবস্থায় রাখা হয়েছে এবং প্রতি দুই দিন অন্তর অন্তর কয়টি Wood lice মারা যায় তা লক্ষ্য করা হয়েছে। অনুমান করা হয়েছে যে, স্বাভাবিক অবস্থায় রাখা হলেও পরীক্ষাগারে বেশি দিন থাকার কারণে Wood lice মারা যাবে। নিচে দুই দিন অন্তর অন্তর মৃত Wood lice-এর সংখ্যা দেয়া হলো।

তথ্যমান নথিভুক্ত করার পিরিয়ড, x	মৃত Wood lice-এর সংখ্যা y	x-এর অনুক্রম	y-এর অনুক্রম	দুই অনুক্রমের অন্তর d
1st	5	1	10	−9
2nd	2	2	7	−5
3rd	0	3	1	2
4th	1	4	4	0
5th	3	5	8	−3
6th	1	6	4	2
7th	1	7	4	3
8th	1	8	4	4
9th	4	9	9	0
10th	1	10	4	6

এখানে $$r_s = 1 - \frac{6\,\Sigma d^2}{n(n^2-1)} = 1 - \frac{6 \times 184}{10(100-1)} = -0.115$$

লক্ষ্য করা যাচ্ছে যে, স্বাভাবিক অবস্থায়ও পরীক্ষাগারে Wood louse প্রথম দুই দিনেই বেশি মারা গিয়েছে। এরপর যতই দিন গিয়েছে মৃত্যু হার ততই কমেছে। তাই অতিক্রান্ত পিরিয়ডের সঙ্গে Wood louse-এর সংশ্লেষণ ঋণাত্মক। অবশ্য, প্রাপ্ত সংশ্লেষাঙ্ক তাৎপর্যহীন।

অনুক্রমিক সংশ্লেষাঙ্কের তাৎপর্য যাচাই করার জন্য প্রাপ্ত সংশ্লেষাঙ্কের মানকে Spearman's rank correlation coefficient-এর (Appendix 3) বর্জনীয় মানের সাথে তুলনা করতে হয়। এক্ষেত্রেও নাস্তি কল্পনা Product moment correlation coefficient-এর জন্য প্রস্তাবিত নাস্তি কল্পনার অনুরূপ। প্রাপ্ত $|r_s| <$ সারণিকৃত r_s হলে নাস্তি কল্পনার বিপক্ষে কোনো যুক্তি নেই বিবেচিত হয়। উপরিউক্ত উদাহরণের ক্ষেত্রে $n = 10$ হলে সারণিকৃত $r_s = 0.648$ (5% সংশয় মাত্রায়) এবং এটি প্রাপ্ত r_s অপেক্ষা বড় বিধায় Wood louse-এর মৃত্যুর সাথে পরীক্ষাগারে অধিক দিন Wood louse-কে রাখার কোনো সংশ্লেষণ নেই বিবেচনা করা যায়।

৬.১৯ আংশিক সংশ্লেষণ (Partial Correlation)

ধরা যাক x_1, x_2 এবং x_3 চলক তিনটি পরস্পর সংশ্লেষিত এবং $r_{12} = x_1$ ও x_2-এর সরল সংশ্লেষাঙ্ক, $r_{13} = x_1$ ও x_3-এর সরল সংশ্লেষাঙ্ক, $r_{23} = x_2$ ও x_3 -এর সরল সংশ্লেষাঙ্ক। এরূপ ক্ষেত্রে x_3 -এর প্রভাব দূর করে x_1 ও x_2 -এর সংশ্লেষণ বা x_2-এর প্রভাব দূর করে x_1 ও x_3 -এর সংশ্লেষণ বা x_2-এর প্রভাব দূর করে x_1 ও x_3 এর সংশ্লেষণ হলো আংশিক সংশ্লেষণ। এই সংশ্লেষণের পরিমাপ করা হয় আংশিক সংশ্লেষাঙ্ক (Parial correlation coefficient) দ্বারা। এখানে x_3-এর প্রভাব দূর করে x_1 ও x_2 এর সংশ্লেষাঙ্ক হলো $r_{12\cdot3}$। এটি নির্ণয় করার সূত্র হলো

$$r_{12\cdot3} = \frac{r_{12} - r_{13}\, r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}}$$

অনুরূপভাবে x_2-এর প্রভাব দূর করে x_1 ও x_3 এর সংশ্লেষাঙ্ক হলো

$$r_{13.2} = \frac{r_{13} - r_{12}\, r_{23}}{\sqrt{(1 - r_{12})^2(1 - r_{23}^2)}}$$

আবার x_1 এর প্রভাব দূর করে x_2 ও x_3 এর সংশ্লেষাঙ্ক হলো

$$r_{23\cdot1} = \frac{r_{23} - r_{12}\, r_{13}}{\sqrt{(1 - r_{12}^2)(1 - r_{13}^2)}}$$

উদাহরণ ৬.১৩

সারণি ১.১ (১ম খণ্ড) এ দেয়া উপাত্তের ক্ষেত্রে Slug এর Body weight (x_2) Body length (x_1) এবং Mantle length (x_3) এর সাথে সংশ্লেষিত। আবার x_1 ও x_3 সংশেষিত। এখন x_3 এর প্রভাব দূর করে x_1 ও x_2 এর সংশ্লেষাঙ্ক নির্ণয় করা যাক।

এখানে $r_{12} = x_1$ ও x_2-এর সংশ্লেষাঙ্ক $= 0.77$

$r_{13} = x_1$ ও x_3-এর সংশ্লেষাঙ্ক $= 0.85$

$r_{23} = x_2$ ও x_3-এর সংশ্লেষাঙ্ক $= 0.85$

তাহলে, x_3-এর প্রভাব দূর করে x_1 ও x_2-এর আংশিক সংশ্লেষাঙ্ক হলো

$$r_{12.3} = \frac{r_{12} - r_{13}\, r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}} = \frac{0.77 - 0.85 \times 0.85}{\sqrt{[1 - (0.85)^2][1 - (0.85)^2]}}$$
$$= 0.17$$

লক্ষ্য করা যাচ্ছে যে, Body length এবং Mantle length অধিক সংশ্লেষিত [$r_{13} = 0.85$]। আবার Body weight (x_2) এবং Body length (x_1) তাৎপর্যপূর্ণভাবে সংশ্লেষিত [$t = 8.36$, $p = 0.00$]। কিন্তু Mantle length (x_3) এর প্রভাব দূর করা হলে x_1 ও x_2-এর সংশ্লেষণের মাত্রা [$r_{12.3}$] তাৎপর্যপূর্ণভাবে সংশ্লেষিত হয় না [$t = 1.19$, $p > 0.05$]।

উপরে তিনটি চলক পরস্পর সংশ্লেষিত হলে একটির প্রভাব দূর করে অন্য দুটির আংশিক সংশ্লেষাঙ্ক নির্ণয় করার সূত্র আলোচনা করা হয়েছে। চারটি চলক x_1, x_2, x_3 এবং x_4 পরস্পর সংশ্লেষিত হলে যে কোনো দুটির প্রভাব দূর করে অন্য দুটির সংশ্লেষাঙ্ক ও আংশিক সংশ্লেষাঙ্ক এবং ধরা যাক $r_{12.34}$ হলো x_3 ও x_4 এর প্রভাব দূর করে x_1 ও x_2 এর সংশ্লেষাঙ্ক। এই সংশ্লেষাঙ্ক নির্ণয়ের সূত্র হলো

$$r_{12\cdot 34} = \frac{r_{12\cdot 3} - r_{14\cdot 3}\, r_{24\cdot 3}}{\sqrt{(1 - r_{14\cdot 3}^2)(1 - r_{24\cdot 3}^2)}}$$

উদাহরণ ৬.১৪

সারণি ১.১ (১ম খণ্ড)-এর উপাত্তের ক্ষেত্রে Slug-এর Body length (x_1), Mantle length (x_3) এবং Keel length (x_4) সংশ্লেষিত। আবার এই তিনটি চলক Body weight (x_2)-এর সাথে সংশ্লেষিত। এক্ষেত্রে x_3 ও x_4-এর প্রভাব দূর করে x_1 ও x_2-এর আংশিক সংশ্লেষাঙ্ক নির্ণয় করা যাক। এখানে

$r_{12} = x_1$ ও x_2-এর সরল সংশ্লেষাঙ্ক $= 0.77$ [$t = 8.36$, $p = 0.00$]

$r_{13} = x_1$ ও x_3-এর সরল সংশ্লেষাঙ্ক $= 0.85$ [$t = 11.18$, $p = 0.00$]

$r_{14} = x_1$ ও x_4-এর সরল সংশ্লেষাঙ্ক $= 0.67$ [$t = 6.25$, $p = 0.00$]

$r_{23} = x_1$ ও x_3-এর সরল সংশ্লেষাঙ্ক $= 0.85$ [$t = 11.18$, $p = 0.00$]

$r_{24} = x_2$ ও x_3-এর সরল সংশ্লেষাঙ্ক $= 0.84$ [$t = 10.72$, $p = 0.00$]

$r_{34} = x_3$ ও x_4-এর সরল সংশ্লেষাঙ্ক $= 0.65$ [$t = 5.92$, $p = 0.00$]

লক্ষ্য করা যাচ্ছে যে, প্রতিটি চলকই পরস্পর তাৎপর্যপূর্ণভাবে সংশ্লেষিত। এখন Mantle length (x_3) এবং Keel length (x_4)-এর প্রভাব দূর করে Body weight (x_2) ও Body length (x_1)-এর সংশ্লেষণের মাত্রা লক্ষ্য করা যাক। এই মাত্রা নির্ণয় করার জন্য প্রথমেই $r_{12\cdot3}$, $r_{14\cdot3}$ এবং $r_{24\cdot3}$ নির্ণয় করতে হয়। এখানে

$r_{12\cdot3} = 0.17$ [উদাহরণ ৫.১২]

$$r_{14\cdot3} = \frac{r_{14} - r_{13}\, r_{34}}{\sqrt{(1 - r_{13}^2)(1 - r_{34}^2)}}$$

$$= \frac{0.67 - 0.85 \times 0.65}{\sqrt{[1 - (0.85)^2][1 - (0.65)^2]}} = 0.29$$

$$r_{24\cdot3} = \frac{r_{24} - r_{23}\, r_{34}}{\sqrt{(1 - r_{23}^2)(1 - r_{34}^2)}}$$

$$r_{24\cdot3} = \frac{0.84 - 0.85 \times 0.65}{\sqrt{[1 - (0.85)^2][1 - (0.65)^2]}} = 0.72$$

এখন
$$r_{12\cdot34} = \frac{r_{12\cdot3} - r_{14\cdot3}\, r_{24\cdot3}}{\sqrt{(1 - r_{14\cdot3}^2)(1 - r_{24\cdot3}^2)}}$$

$$= \frac{0.17 - 0.29 \times 0.72}{\sqrt{[1 - (0.29)^2][1 - (0.72)^2]}} = -0.06$$

লক্ষ্য করা যাচ্ছে যে, Slug-এর Body weight এবং Body length তাৎপর্যপূর্ণভাবে সংশ্লেষিত এবং Body length বৃদ্ধির সাথে সাথে Body weight বৃদ্ধি পাচ্ছে। আবার Body length, Mantle length এবং Keel length-এর সাথেও তাৎপর্যপূর্ণভাবে সংশ্লেষিত। এখানে $r_{12\cdot34}$ নির্দেশ করছে যে, এই শেষোক্ত সংশ্লেষণের কারণেই Body weight ও Body length তাৎপর্যপূর্ণভাবে সংশ্লেষিত। কারণ, Mantle length এবং Keel length-এর প্রভাব দূর করার পরে Body weight এবং Body length-এর সংশ্লেষণ তাৎপর্যহীন [$t = -0.42$, $p > 0.05$]।

আবার Body length (x_1) এবং Mantle length (x_3) এর সংশ্লেষণও অন্য দুটি চলক (x_2, x_4)-এর প্রভাব দূর করে পর্যালোচনা করা যায়। এই সংশ্লেষণের মাত্রা হলো

$$r_{13\cdot24}=\frac{r_{13\cdot4}-r_{12\cdot4}\,r_{23\cdot4}}{\sqrt{[1-r_{12\cdot4}^2][1-r_{23\cdot4}^2]}}$$

এখানে
$$r_{13.4}=\frac{r_{13}-r_{14}\,r_{34}}{\sqrt{[1-r_{14}^2][1-r_{34}^2]}}$$

$$=\frac{0.85-0.67\times0.65}{\sqrt{[1-(0.67)^2][1-(0.65)^2]}}=0.73$$

$$r_{12\cdot4}=\frac{r_{12}-r_{14}\,r_{24}}{\sqrt{[1-r_{14}^2][1-r_{24}^2]}}$$

$$=\frac{0.77-0.67\times0.84}{\sqrt{[1-(0.67)^2][1-(0.84)^2]}}=0{\cdot}51$$

$$r_{23\cdot4}=\frac{r_{23}-r_{24}\,r_{34}}{\sqrt{[1-r_{24}^2][1-r_{34}^2]}}$$

$$=\frac{0.85-0.84\times0.65}{\sqrt{[1-(0.84)^2][1-(0.65)^2]}}=0.74$$

তাহলে
$$r_{13\cdot24}=\frac{r_{13\cdot4}-r_{12\cdot4}\,r_{23.4}}{\sqrt{[1-r_{12\cdot4}^2][1-r_{23\cdot4}^2]}}$$

$$=\frac{0.73-0.51\times0.74}{\sqrt{[1-(0.51)^2][1-(0.74)^2]}}$$
$$=0.61$$

অনুরূপভাবে Body length (x_1) এবং Keel length (x_4)-এর সংশ্লেষাঙ্ক [$r_{14\cdot23}$] অন্য দুটি চলকের প্রভাব দূর করে নির্ণয় করা যায়। এই সংশ্লেষাঙ্ক হলো

$$r_{14\cdot23}=\frac{r_{14\cdot3}-r_{12\cdot3}\,r_{24\cdot3}}{\sqrt{[1-r_{12\cdot3}^2][1-r_{24\cdot3}^2]}}$$

$$=\frac{0.29-0.17\times0.72}{\sqrt{[1-(0.17)^2][1-(0.72)^2]}}$$
$$=0.24$$

৬.২০ কানুনী সংশ্লেষণ বিশ্লেষণ (Canonical Correlation Analysis)

আমরা ৬.১৮ অনুচ্ছেদে সরল সংশ্লেষণ সম্বন্ধে পড়েছি। একটি নির্ভরশীল চলক এবং একাধিক অনপেক্ষ চলকের ক্ষেত্রে বহুল সংশ্লেষাঙ্ক (multiple correlation coefficient) সম্বন্ধে ৬.৬ অনুচ্ছেদে কিছুটা আলোকপাত করা হয়েছে। ঐ অনুচ্ছেদে সিদ্ধান্ত সহগ R^2-এর ধনাত্মক বর্গমূলই হলো বহুল সংশ্লেষাঙ্ক যার মান 0 থেকে 1 পর্যন্ত হতে পারে।

বাস্তব ক্ষেত্রে নির্ভরশীল চলক একাধিক হতে পারে। যেমন, জন উর্বরতা (fertility) এবং শিশু-মৃত্যু (child mortality) চলক দুটি পিতা-মাতার শিক্ষা, তাদের পেশা, আর্থিক অবস্থা, বয়স, বিবাহিত সময়কাল ইত্যাদির উপর নির্ভরশীল। এখানে জন উর্বরতা এবং শিশু-মৃত্যু চলকদ্বয় নির্ভরশীল চলক এবং বাকি চলকগুলি অনপেক্ষ চলক। এ চলকগুলির আন্তঃসম্পর্ক যুগপৎ পর্যালোচনা করতে হলে বহুল সংশ্লেষণ বা বহুল নির্ভরণ যথেষ্ট নয়। এক্ষেত্রে চলকসমূহকে দুটি গুচ্ছে ভাগ করা হয়। একটি নির্ভরশীল চলকসমূহের গুচ্ছ এবং অপরটি অনপেক্ষ চলকসমূহের গুচ্ছ। ধরা যাক নির্ভরশীল চলক গুচ্ছে p চলক আছে এবং এটি Y-গুচ্ছ হিসেবে চিহ্নিত এবং অনপেক্ষ চলকগুচ্ছে q চলক আছে যা X-গুচ্ছ হিসেবে চিহ্নিত। ধরা যাক $q \geqslant p$। এখন X-গুচ্ছ ও Y-গুচ্ছের ক্ষেত্রে দুটি রৈখিক সমাবেশ যথাক্রমে

$$X^* = a_1x_1 + a_2x_2 + \cdots + a_qx_q$$
$$Y^* = b_1y_1 + b_2y_2 + \cdots + b_py_p$$

বিবেচনা করা যাক। এখানে

$$V(X) = \Sigma_{XX} = \begin{pmatrix} V(x_1) & Cov(x_1\, x_2) & \ldots & Cov(x_1\, x_q) \\ Cov(x_1x_2) & V(x_2) & \cdots & Cov(x_2\, x_q) \\ \cdots & \cdots & \cdots & \cdots \\ Cov(x_1\, x_q) & Cov(x_2x_q) & \ldots & V(x_q) \end{pmatrix}_{q \times q}$$

$$V(Y) = \Sigma_{YY} = \begin{pmatrix} V(y_1) & Cov(y_1y_2) & \ldots & Cov(y_1\, y_p) \\ Cov(y_1y_2) & V(y_2) & \cdots & Cov(y_2y_p) \\ \cdots & \cdots & \cdots & \cdots \\ Cov(y_1y_p) & Cov(y_2y_p) & \cdots & V(y_p) \end{pmatrix}_{p \times p}$$

$$\mathrm{Cov}(XY)=\Sigma_{XY}=\begin{bmatrix}\mathrm{Cov}(x_1y_1) & \mathrm{Cov}(x_2y_1) & \cdots & \mathrm{Cov}(x_qy_1)\\ \mathrm{Cov}(x_1y_2) & \mathrm{Cov}(x_2y_2) & \cdots & \mathrm{Cov}(x_qy_2)\\ \cdots & \cdots & \cdots & \cdots\\ \mathrm{Cov}(x_1y_p) & \mathrm{Cov}(x_2y_p) & \cdots & \mathrm{Cov}(x_qy_p)\end{bmatrix}_{p\times q}$$

এখন X^* ও Y^*-এর সংশ্লেষাঙ্ক হলো

$$\rho_{X^*Y^*}=\frac{a'\,\Sigma_{XY}\,b}{\left[(a'\Sigma_{XX}a)(b'\Sigma_{YY}b)\right]^{1/2}}$$

এখানে

$$a'=[a_1\ a_2\ \ldots\ a_q],\ X=\begin{bmatrix}x_1\\x_2\\\vdots\\x_q\end{bmatrix}\ b'=[b_1\ b_2\ \ldots\ b_p],\ Y=\begin{bmatrix}y_1\\y_2\\\vdots\\y_p\end{bmatrix}$$

কানুনী সংশ্লেষণের উদ্দেশ্য হলো X^* ও Y^*-এর মান এমনভাবে নির্ণয় করা যেন $\rho_{X^*Y^*}$ সর্বোচ্চ হয়।

বহুল নির্ভরণের ক্ষেত্রে Y-গুচ্ছে একটি চলক থাকে $[p=1]$ এবং X^*-এর মান এমনভাবে নির্ণয় করা হয় যেন X^* Y-এর সর্বোচ্চ ভেদ ব্যাখ্যা করতে পারে। সেক্ষেত্রে কানুনী সংশ্লেষণ হলো বহুল নির্ভরণের অনুরূপ যেখানে একাধিক নির্ভরশীল চলক থাকে। এই কানুনী সংশ্লেষণ বিশ্লেষণ এবং বহুল নির্ভরণ বিশ্লেষণ একই বিশ্লেষণে পরিণত হয় যদি নির্ভরশীল চলক একটি হয় $(p=1)$।

উপরিউক্ত আলোচনা হতে বুঝা যাচ্ছে যে, কানুনী সংশ্লেষণ হলো নির্ভরশীল চলকগুচ্ছ ও অনপেক্ষ চলকগুচ্ছের যুগপৎ সম্পর্ক পর্যালোচনা করার একটি পদ্ধতি। বাস্তবক্ষেত্রে এই বিশ্লেষণ করার জন্য কম্পিউটার প্রোগ্রাম থাকায় এই বিশ্লেষণ পদ্ধতির প্রয়োগ সহজ হয়েছে। এখানে বিশ্লেষণের একটি সংক্ষিপ্ত পদ্ধতি এবং বিশ্লেষিত ফলাফলের ব্যাখ্যা দেয়ার চেষ্টা করা হয়েছে।

এই বিশ্লেষণ X-গুচ্ছের এবং Y-গুচ্ছের সংশ্লেষাঙ্ক ম্যাট্রিক্স হতে করা যায়। ধরা যাক

R_{XX} = X-গুচ্ছের চলকসমূহের সংশ্লেষাঙ্ক ম্যাট্রিক্স,

R_{YY} = Y-গুচ্ছের চলকসমূহের সংশ্লেষাঙ্ক ম্যাট্রিক্স।

$R_{yx} = R'_{xy}$ = X-গুচ্ছ ও Y-গুচ্ছের চলকসমূহের সংশ্লেষাঙ্ক ম্যাট্রিক্স। এই ম্যাট্রিক্সগুলি হতে অন্য দুটি ম্যাট্রিক্স

$$R_{xx}^{-1} R_{xy} R_{yy}^{-1} R_{yx} \text{ অথবা } R_{yy}^{-1} R_{yx} R_{xx}^{-1} R_{xy}$$

পাওয়া যায়। ধরা যাক এ ম্যাট্রিক্স দুটির যে কোনো একটির আইগেন মান (eigen value) হলো

$$\lambda_1 > \lambda_2 > \cdots > \lambda_p, \text{ যদি } p \leqslant q \text{ হয়}$$

এখন i-তম আইগেন মান λ_i-এর জন্য আইগেন ভেক্টর হলো a_i, যেখানে

$$a_i = \begin{bmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{qi} \end{bmatrix}$$

এই a_i হলো প্রথম ম্যাট্রিক্সের প্রাসঙ্গিক i-তম আইগেন ভেক্টর। অনুরূপভাবে দ্বিতীয় ম্যাট্রিক্স হতে i-তম আইগেন ভেক্টর হলো

$$b_i = \begin{bmatrix} b_{1i} \\ b_{2i} \\ \vdots \\ b_{pi} \end{bmatrix}$$

তাহলে i-তম কানুনী চলক জোড়া হলো

$$a_i'X = a_{1i}x_1 + a_{2i}\, x_2 + \cdots + a_{qi}\, x_q$$
$$b_i'X = b_{1i}y_1 + b_{2i}y_2 + \cdots + b_{pi}y_p$$

এখানে $a_i' = [a_{1i}\ a_{2i}^2 \cdots a_{qi}]$ এবং $b_i' = [b_{1i}\ b_{2i} \cdots b_{pi}]$

হলো i-তম কানুনী চলক জোড়া (canonical variate pair)-এর কানুনী ভার (canonical weights) এবং i-তম কানুনী চলক জোড়ার কানুনী সংশ্লেষাঙ্ক হলো $\sqrt{\lambda_i}$ ।

উপরিউক্ত কানুনী ভার দ্বারা বুঝা যায় কোন কোন চলক দ্বারা কানুনী চলক জোড়া বেশি প্রভাবিত হয়। X-চলক গুচ্ছে যদি মাল্টিকোলিনিয়ারিটি সমস্যা না থাকে, তাহলে কানুনী ভারের বড় মান যে চলকের সাথে সম্পর্কিত ঐ চলক কানুনী

চলক জোড়াকে বেশি প্রভাবিত করে। অন্যভাবে বলা যায় কানুনী ভারের বড় মান যে চলকগুলির সাথে সম্পর্কিত ঐ চলক গুচ্ছই X-set ও Y-set-এর সম্পর্ক পর্যালোচনা করার জন্য অধিক প্রভাবশালী। এই কানুনী ভার বহুল নির্ভরণের ক্ষেত্রে নির্ভরাঙ্কের অনুরূপ।

X-গুচ্ছের চলকসমূহের মধ্যে মাল্টিকোলিনিয়ারিটি সমস্যা বিদ্যমান থাকলে কানুনী ভার চলকের প্রভাব চিহ্নিত করার জন্য যথাযথ নয়। সেক্ষেত্রে কানুনী লোডিং (canonical loading) নির্ণয় করতে হয়। এই কানুনী লোডিং নির্ণয় করার সূত্র হলো

$$R_{X^*X(i)} = R_{XX}a_i = \begin{pmatrix} 1 & r_{12} & r_{13} \cdots & r_{1q} \\ r_{21} & 1 & r_{23} \cdots & r_{2q} \\ \cdots & \cdots & \cdots & \cdots \\ r_{q1} & r_{q2} & \cdots & 1 \end{pmatrix} \begin{pmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{qi} \end{pmatrix} = \begin{pmatrix} a'_{1i} \\ a'_{2i} \\ \vdots \\ a'_{qi} \end{pmatrix}$$

এখানে r_{ij} ($i \neq j = 1, 2, \dots, q$) হলো X-গুচ্ছের চলকসমূহের সংশ্লেষাঙ্ক।

$$R_{Y^*Y(i)} = R_{YY}b_i = \begin{pmatrix} 1 & r_{12} & r_{13} \dots & r_{1p} \\ r_{21} & 1 & r_{22} \; . & r_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ r_{p1} & r_{p2} & \cdots \quad \cdots & 1 \end{pmatrix} \begin{pmatrix} b_{1i} \\ b_{2i} \\ \vdots \\ b_{pi} \end{pmatrix} = \begin{pmatrix} b'_{1i} \\ b'_{2i} \\ \vdots \\ b'_{pi} \end{pmatrix}$$

এই r_{ij} ($i \neq j = 1, 2, \dots, p$) হলো Y-গুচ্ছের চলকসমূহের সংশ্লেষাঙ্ক।

বিশ্লেষণের এই পর্যায়ে কানুনী চলক দ্বারা X-গুচ্ছের ভেদের কত সমানুপাত (Proportion of variation of Y-set) ব্যাখ্যা করা যায় তা নির্ণয় করা যায়। এই সমানুপাত হলো

$$R^2_{(i)y} = \frac{1}{p}\left[b'^2_{1i} + b'^2_{2i} + \cdots + b'^2_{pi}\right]$$

অনুরূপভাবে X-গুচ্ছের ভেদের প্রকাশিত সমানুপাত হলো

$$R^2_{(i)x} = \frac{1}{q}\left[a'^2_{1i} + a'^2_{2i} + \cdots + a'^2_{qi}\right]$$

আবার X-গুচ্ছ দ্বারা Y-গুচ্ছের ভেদের প্রকাশিত সমানুপাত হলো

$$R^2_{(i)y/x} = \lambda_i \, Y^2_{(i)y}$$

সবশেষে Y-গুচ্ছের যে কোনো চলকের সাথে X-গুচ্ছের যে কোনো চলকের সম্পর্ক পরিমাপ করার জন্য ক্রস ভার (cross-weights) নির্ণয় করা হয়। i-তম কানুনী চলক জোড়ার ক্ষেত্রে Y-গুচ্ছের l-তম ($l = 1, 2, \cdots ; p$) চলকের জন্য এই ক্রস ভার হলো

$$\sqrt{\lambda_1}\ b'_{l1} (l = 1, 2, \cdots, p)$$

অনুরূপভাবে X-গুচ্ছের j-তম ($j = 1, 2, \cdots, q$) চলকের জন্য এই ক্রস ভার হলো

$$\sqrt{\lambda_1}\ a'_{j1}$$

এতক্ষণ কানুনী সংশ্লেষণের বিভিন্ন ধাপ এবং এই বিশ্লেষণ হতে চলকগুলির সম্পর্কের ব্যাখ্যা সম্বন্ধে একটি সংক্ষিপ্ত ধারণা ব্যক্ত করা হয়েছে। উল্লেখ করা হয়েছে যে, কানুনী সংশ্লেষণ বিশ্লেষণ করতে p আইগেন মান পাওয়া যায়, যেখানে $p \leqslant q$। সঠিকভাবে আইগেন মানের সংখ্যা হলো (p ও q)-এর মধ্যে যেটি ন্যূনতম এবং যতটি আইগেন মান পাওয়া যায় ঠিক ততটি কানুনী চলক জোড়া পাওয়া যায়। এর মধ্যে সব কানুনী চলক জোড়া যে তাৎপর্যপূর্ণ হবে এমন কোনো কথা নেই। তাৎপর্যপূর্ণ কানুনী চলক জোড়াই X-গুচ্ছ Y-গুচ্ছ চলকের সংশ্লেষণ পর্যালোচনার জন্য যথেষ্ট। তাই বিস্তারিত বিশ্লেষণের পূর্বে তাৎপর্যপূর্ণ কানুনী চলক জোড়া চিহ্নিত করতে হয়। যেহেতু $\sqrt{\lambda_i}$ ($i = 1, 2, \ldots, p$) i-তম কানুনী চলক জোড়ার কানুনী সংশ্লেষাঙ্ক, সে কারণে λ_i ব্যবহার করে তাৎপর্যপূর্ণ কানুনী চলক জোড়া চিহ্নিত করা যায়। এই পর্যায়ে নাস্তি কল্পনা হলো

$$H_0 : \quad \Sigma_{XY} = 0$$

এই নাস্তি কল্পনার অর্থ হলো X-গুচ্ছ ও Y-গুচ্ছ এর মধ্যে কোনো সম্পর্ক নেই। এই নাস্তি কল্পনার জন্য যাচাই তথ্যজমান হলো

$$\chi^2_{pq} = -\{n - \tfrac{1}{2}(p + q + 3)\} \sum_{i=1}^{k} ln\,(1 - \lambda_i)$$

এখানে k = ন্যূনতম (p,q)। এই χ^2_{pq}-এর বিন্যাস হলো pq স্বাধীনতার মাত্রাবিশিষ্ট কাই বর্গ বিন্যাস। $\chi^2_{pq} > \chi^2_{0.05}$ হলে নাস্তি কল্পনা বাতিল বলে গণ্য হবে।

এই নাস্তি কল্পনা বাতিল হলেই কানুনী সংশ্লেষণ বিশ্লেষণ অর্থবহ হবে। সে ক্ষেত্রে প্রথম কানুনী চলক জোড়া তাৎপর্যপূর্ণ বলে বিবেচিত হবে। Bartlett (১৯৪৭) প্রস্তাব করেছেন যে, k = ন্যূনতম (p,q) এর মধ্যে প্রথম $s (s = 1, 2, \ldots, k)$ সংখ্যক কানুনী

চলক জোড়া [অন্যভাবে বলা যায় $\lambda_1 > \lambda_2 > \ldots > \lambda_s > 0$] তাৎপর্যপূর্ণ যাচাই করার জন্য যাচাই তথ্যজমান হলো

$$\chi^2_{(p-s)(q-s)} = -\{n - \tfrac{1}{2}(p+q+3)\} \sum_{i=s+1}^{k} ln(1-\lambda_i)$$

এই $\chi^2_{(p-s)(q-s)}$-এর স্বাধীনতার মাত্রা হলো $(p-s)(q-s)$। এক্ষেত্রে $\chi^2_{(p-s)(q-s)} < \chi^2_{0.05}$ হলে প্রথম s কানুনী চলক জোড়া তাৎপর্যপূর্ণ বলে বিবেচিত হবে।

উদাহরণ ৬.১২

Morphological বৈশিষ্ট্য পর্যালোচনা করার জন্য ভূমধ্যসাগরীয় উপকূলবর্তী অঞ্চলের একটি বাজার হতে বিভিন্ন Order এর বিভিন্ন Family ভুক্ত বিভিন্ন প্রজাতির (species) কিছু মাছ সংগ্রহ করে ঐ গুলির Caudal vertebrae এর বিভিন্ন অংশের পরিমাপ করা হয়েছে [Osama (১৯৯৭)] নিচে Osteich thyes শ্রেণীর Perciformes order এর Serranidae family ভুক্ত ছয় প্রজাতির মাছের Vertebrae এর বিভিন্ন অংশের পরিমাপ দেয়া হলো। এই উপাত্তের ভিত্তিতে কানুনী সংশ্লেষণ বিশ্লেষণ করা যাক।

সারণি ৬.২৩ : মাছের vertebra এর বিভিন্ন অংশের পরিমাপ।

Vertebrae-এর বিভিন্ন অংশ

ক্রমিক সংখ্যা	মাছের প্রজাতি	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
1	Epine-	4.42	3.75	3.70	3.93	0.60	2.20	2.20	0.6	67	67
2	phelus	4.36	3.80	3.78	3.85	0.60	2.20	2.20	0.6	65	53
3	aeneus	4.59	3.86	3.72	3.77	0.90	2.78	2.20	0.6	60	47
4		4.63	3.75	3.62	3.70	0.90	2.78	1.67	0.6	60	43
5		4.74	3.64	3.59	3.58	0.60	3.30	1.10	0.6	50	47
6		4.92	3.63	3.60	3.55	0.60	3.30	1.10	0.6	45	33
7		5.06	3.62	3.54	3.43	0.60	2.78	1.10	0.6	45	47
8		5.23	3.59	3.58	3.38	0.70	2.78	1.10	0.6	45	30

9		5.42	3.55	3.61	3.43	1.10	2.78	1.10	0.6	47	52
10		5.67	3.68	3.57	3.46	1.10	2.78	1.10	0.6	45	44
11		5.74	3.66	3.54	3.36	1.30	2.20	1.10	1.1	37	45
12	Epinephelus	10.11	8.57	8.38	8.60	2.20	4.40	4.40	0.6	65	65
13	Caninus	10.10	8.42	8.12	8.59	2.20	5.20	3.30	0.6	75	45
14		10.04	8.12	8.11	8.41	3.30	6.67	3.30	0.8	70	38
15		10.32	8.15	7.71	8.06	2.20	6.67	2.20	0.7	62	40
16		10.64	7.78	7.62	7.77	2.78	7.20	2.20	0.7	54	47
17		11.10	7.60	7.48	7.63	2.78	7.78	2.20	0.6	53	45
18		11.04	7.53	7.46	7.67	2.78	7.20	2.78	1.3	52	45
19		1.64	7.58	7.461	7.48	2.78	7.20	2.20	1.3	47	45
20		11.82	7.59	7.63	7.47	2.78	6.67	2.20	1.1	42	50
21		12.27	7.66	7.72	7.42	2.78	6.67	2.20	1.4	43	45
22		12.03	7.64	7.50	7.18	2.78	6.10	1.67	1.3	35	46
23		10.88	7.55	6.89	7.31	1.10	3.30	1.10	0.6	37	36
24	Epinephelus	8.69	7.83	7.79	8.30	1.67	3.89	4.40	0.6	63	60
25	Costae	8.72	7.73	7.49	7.90	1.67	3.30	3.30	0.6	65	50
26		8.59	7.41	7.22	7.41	2.20	4.40	2.78	0.6	75	35
27		8.49	7.23	6.93	7.16	1.67	4.40	2.78	0.7	65	47
28		8.62	7.02	6.82	7.06	2.20	5.56	2.20	1.1	65	43
29		8.83	6.87	6.60	6.81	2.20	5.56	2.20	1.1	58	47
30		9.05	6.69	6.39	6.74	2.20	5.56	2.20	1.1	55	50
31		9.15	6.48	6.15	6.64	1.67	4.40	2.20	1.1	55	45
32		9.39	6.34	6.15	6.55	2.20	5.00	1.67	1.1	54	45
33		9.15	6.25	6.06	6.25	2.20	5.00	1.10	0.8	50	45
34		8.76	6.05	5.99	6.03	2.20	4.40	1.10	1.1	47	38
35		7.36	6.00	5.55	5.90	1.10	3.30	1.10	1.1	53	40

36	Epinephelus	4.65	4.01	3.96	4.02	0.90	1.67	2.20	0.6	75	70
37	haifensis	4.51	3.90	3.81	4.10	1.00	2.20	2.20	0.6	72	60
38		4.40	3.75	3.88	4.02	1.10	2.78	2.20	0.6	65	65
39		4.46	3.83	3.71	3.80	1.10	2.78	1.67	0.6	60	66
40		4.58	3.69	3.64	3.72	1.00	3.10	1.10	0.6	55	60
41	Epinephelus	4.67	3.69	3.53	3.50	1.00	3.10	1.10	0.6	55	63
42	marginatus	4.71	3.57	3.40	3.46	1.00	2.89	1.10	0.6	70	45
43		4.79	3.42	3.30	3.41	1.10	2.56	1.10	0.6	55	65
44		4.80	3.40	3.25	3.37	1.10	2.56	1.10	0.7	53	55
45		5.07	3.39	3.27	3.35	1.10	2.40	1.10	0.9	50	54
46		5.01	3.38	3.28	3.20	1.10	2.20	1.10	0.9	42	45
47		4.87	3.37	3.14	3.13	0.90	1.89	1.10	0.6	38	48
48	Myeteroperca	13.52	11.30	11.21	11.81	1.10	5.56	3.89	0.6	51	47
49	rubra	13.39	11.19	11.15	11.22	1.67	3.30	3.89	0.6	60	75
50		12.90	10.99	10.92	11.21	1.67	4.40	3.89	0.6	68	65
51		13.20	10.95	10.80	11.41	2.20	3.30	4.40	0.6	69	57
52		13.06	10.82	10.63	11.07	2.20	6.10	4.40	0.6	64	62
53		13.02	10.68	10.50	10.87	2.20	6.67	3.89	0.6	60	65
54		13.18	10.57	10.43	10.56	2.20	6.67	3.30	0.6	55	53
55		13.34	10.59	10.20	10.27	2.20	7.20	3.30	0.6	51	53
56		13.85	10.41	10.00	10.27	2.78	6.67	3.89	1.1	47	53
57		14.26	10.18	9.85	10.09	2.20	6.67	3.30	1.1	44	50
58		14.69	10.00	9.75	9.95	2.78	7.20	3.30	1.1	45	55
59		14.84	9.91	9.85	9.90	3.30	7.20	3.30	1.1	45	43
60		15.07	10.15	10.05	9.78	2.78	6.67	3.30	1.1	40	42
61		14.90	10.40	10.55	9.65	2.20	4.40	2.78	1.1	37	45
62		13.02	10.61	9.95	9.50	0.60	2.20	2.20	0.6	37	35

এখানে x_1 = Centrum-এর দৈর্ঘ্য [mm],

x_2 = Anterior হতে নেয়া Centrum-এর প্রশস্ততা

x_3 = Posterior হতে নেয়া Centrum-এর প্রশস্ততা

x_4 = Centrum-এর উচ্চতা

x_5 = Neural spine-এর ভিত্তি হতে Posterior face পর্যন্ত দূরত্ব

x_6 = Haemal spine-এর ভিত্তি হতে Posterior face পর্যন্ত দূরত্ব

x_7 = Anterior Zygapophysis-এর দৈর্ঘ্য

x_8 = Posterior Zygapophysis-এর দৈর্ঘ্য

x_9 = Neural spine এবং Longitudinal axis-এর মধ্যে কোণ

x_{10} = Haemal spine এবং Logittudinal axis-এর মধ্যে কোণ

উপরিউক্ত চলকসমূহের মধ্যে x_9 এবং x_{10} অন্যান্য চলকের উপর নির্ভরশীল। সুতরাং x_9 ও x_{10} হলো নির্ভরশীল চলকগুচ্ছ (Y-set) এবং x_1, x_2, x_3, x_4, x_5, x_6, x_7 ও x_8 হলো অনপেক্ষ চলকগুচ্ছ (X-set)। উক্ত চলকগুচ্ছের ভিত্তিতে সংশ্লেষণ ম্যাট্রিক্সগুলি নিম্নরূপ :

$$R_{yy} = \begin{array}{c} \\ x_9 \\ x_{10} \end{array} \begin{array}{c} \begin{array}{cc} x_9 & x_{10} \end{array} \\ \begin{bmatrix} 1.00 & 0.41 \\ 0.41 & 1.00 \end{bmatrix} \end{array}, \quad R_{XX} = \begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \end{array} \begin{array}{c} \begin{array}{cccccccc} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 \end{array} \\ \begin{bmatrix} 1.00 & & & & & & & \\ 0.96 & 1.00 & & & & & & \\ 0.96 & 0.99 & 1.00 & & & & & \\ 0.95 & 0.99 & 0.99 & 1.00 & & & & \\ 0.75 & 0.66 & 0.67 & 0.67 & 1.00 & & & \\ 0.77 & 0.69 & 0.69 & 0.69 & 0.89 & 1.00 & & \\ 0.70 & 0.82 & 0.83 & 0.85 & 0.51 & 0.49 & 1.00 & \\ 0.39 & 0.21 & 0.21 & 0.18 & 0.59 & 0.49 & -0.03 & 1.00 \end{bmatrix} \end{array}$$

$$R_{xy} = \begin{array}{c|cc} & x_9 & x_{10} \\ \hline x_1 & -0.22 & -0.08 \\ x_2 & -0.02 & 0.02 \\ x_3 & -0.01 & 0.04 \\ x_4 & 0.04 & 0.06 \\ x_5 & -0.06 & -0.13 \\ x_6 & -0.12 & -0.18 \\ x_7 & 0.36 & 0.33 \\ x_8 & -0.45 & -0.31 \end{array}$$

এখন $[R_{xx}^{-1} R_{xy} R_{yy}^{-1} R_{yx}]$ অথবা $[R_{yy}^{-1} R_{yx} R_{xx}^{-1} R_{xy}]$ ম্যাট্রিক্স-এর আইগেন মান হলো $\lambda_1 = 0.7978$ এবং $\lambda_2 = 0.2142$ । এখানে $p=2$ হওয়ায় এবং $q=8 > p$ হওয়ায় p ও q-এর ন্যূনতম মান $[\min(p, q)=2]$ 2 এবং দুটি আইগেনই পাওয়া যায় এবং সে কারণেই $p=2$ জোড়া কানুনী চলক পাওয়া যায়।

উপরিউক্ত আইগেন মানের ভিত্তিতে X-গুচ্ছ হতে প্রাপ্ত আইগেন ভেক্টর এবং Y-গুচ্ছ হতে প্রাপ্ত আইগেন ভেক্টর সারণি ৬.২৪-এ দেয়া হলো।

সারণি ৬.২৪ : X-গুচ্ছ ও Y-গুচ্ছ হতে প্রাপ্ত আইগেন ভেক্টর।

চলকসমূহের প্রাসঙ্গিক আইগেন ভেক্টরের মান	প্রাসঙ্গিক আইগেন মান X-গুচ্ছ এর জন্য a_i		Y-গুচ্ছ-এর জন্য b_i	
	λ_1	λ_2	λ_1	λ_2
x_1	3.0391	3.9867	—	—
x_2	0.2472	−9.3503	—	—
x_3	0.0008	5.3844	—	—
x_4	−2.6084	−0.0239	—	—
x_5	−0.6782	−0.8561	—	—
x_6	0.1400	−0.5128	—	—
x_7	−0.2689	1.4573	—	—
x_8	0.0799	−0.1357	—	—
x_9	—	—	−0.9503	−0.5489
x_{10}	—	—	−0.1088	1.0920

এখানে প্রথম কানুনী চলক জোড়া হলো

$$a_1' X = 3.0391\,x_1 + 0.2472\,x_2 + 0.0008\,x_3 - 2.6084\,x_4 - 0.6782\,x_5 + 0.14\,x_6 - 0.2619\,x_7 + 0.0799\,x_8$$

$$b_1' Y = -0.9503\,x_9 - 0.1088\,x_{10}$$

এই বিশ্লেষণে $\lambda_1 = 0.7978$ বড় হওয়ায় তার প্রাসঙ্গিক আইগেন ভেক্টরই প্রথম কানুনী চলক জোড়ার সহগসমূহ। দ্বিতীয় কানুনী চলক জোড়া হলো

$$a_2' X = 3.9867\,x_1 - 9.3503\,x_2 + 5.3844\,x_3 - 0.0239\,x_4 - 0.8561\,x_5 - 0.5128\,x_6 + 1.4573\,x_7 - 0.1357\,x_8$$

$$b_2' Y = -0.5489\,x_9 + 1.0920\,x_{10}$$

লক্ষ্য করার বিষয় হলো প্রাপ্ত কানুনী চলক জোড়াদ্বয় তাৎপর্যপূর্ণ কিনা। এই জন্য নাস্তি কল্পনা হলো

$$H_0 : \Sigma_{xy} = 0 \text{ এবং}$$

বিকল্প কল্পনা $\qquad H_A : \Sigma_{xy} \neq 0$

এই নাস্তি কল্পনা যাচাই-এর সাথে জড়িত ফলাফল সারণি ৬.২৫-এ উপস্থাপন করা হলো। দেখা যাচ্ছে যে, প্রথম কানুনী চলক জোড়াই তাৎপর্যপূর্ণ। এই কানুনী

সারণি ৬.২৫ : কানুনী চলক জোড়াগুলি তাৎপর্যপূর্ণ কিনা তা যাচাই-এর সাথে জড়িত ফলাফল।

কানুনী চলক জোড়া	আইগেন মান, λ_1	χ^2	স্বাধীনতার মাত্রা, d.f.	p-মান	কানুনী সংশ্লেষাঙ্ক $\sqrt{\lambda_1}$
1	0.7978	102.10	16	0.00	0.8932
2	0.2142	13.38	7	0.06	0.4628

চলক জোড়া Centrum-এর দৈর্ঘ্য (x_1) এবং Neural spine ও Longitudinal axis-এর মধ্যে কোণ দ্বারা (x_4) বেশি প্রভাবিত। এই কানুনী চলক জোড়ার জন্য পরবর্তী গুরুত্বপূর্ণ চলক হলো Centrum-এর উচ্চতা (x_4) এবং Haemal spine ও Longitudinal axis-এর মধ্যে কোণ (x_{10})। এই প্রথম কানুনী চলক জোড়া উপাত্তের ভেদের 89.32% ব্যাখ্যা করতে পারে।

দ্বিতীয় কানুনী চলক জোড়া তাৎপর্যপূর্ণ না হলেও তা উপাত্তের ভেদের 46.28% ব্যাখ্যা করতে পারে এবং এক্ষেত্রে গুরুত্বপূর্ণ চলক হলো Anterior হতে নেয়া Centrum-এর প্রশস্ততা (x_2) এবং x_{10}। এখানে দ্বিতীয় কানুনী চলক জোড়া তাৎপর্যপূর্ণ

না হয়েও উপাত্তের ভেদের একটি বড় অংশ ব্যাখ্যা করছে। এই সন্দেহজনক ফলাফল X-গুচ্ছের চলকসমূহের মধ্যে মাল্টিকোলিনিয়ারিটি সমস্যা বিদ্যমান থাকার কারণে হতে পারে। সে কারণে কানুনী সংশ্লেষণের ক্ষেত্রে চলকের গুরুত্বপূর্ণ ভূমিকা লক্ষ্য করার জন্য কানুনী লোডিং নির্ণয় করতে হয়। X-গুচ্ছের চলকসমূহের জন্য এই লোডিং হলো

$$R_{X^*X(i)} = R_{XX}\ a_i = a_i'$$

এবং Y-গুচ্ছের চলকসমূহের জন্য এটি

$$R_{Y^*Y(i)} = R_{YY}\ b_i = b_i'$$

এখানে সারণি ৬.২৬-এ এই X-গুচ্ছ ও Y-গুচ্ছের জন্য কানুনী লোডিং দেয়া হলো। এই বিশ্লেষণ হতে লক্ষ্য করা যাচ্ছে যে, প্রথম কানুনী চলক জোড়ার ক্ষেত্রে চলক x_8

সারণি ৬.২৬ : X-গুচ্ছের জন্য কানুনী লোডিং $[a_i']$ এবং Y-গুচ্ছের জন্য কানুনী লোডিং $[b_i']$।

চলকসমূহ	আইগেন মানের প্রাসঙ্গিক আইগেন মান λ_1	λ_2	a_i' ও b_i'-এর মান আইগেন মান λ_1	λ_2
	X-গুচ্ছের জন্য a_i'		Y-গুচ্ছের জন্য b_i'	
x_1	0.2413	0.0870	—	—
x_2	0.0285	0.0314	—	—
x_3	0.0165	0.1848	—	—
x_4	−0.0477	0.1241	—	—
x_5	0.0518	−0.2390	—	—
x_6	0.1552	−0.3103	—	—
x_7	−0.4350	0.3457	—	—
x_8	0.5242	−0.2181	—	—
x_9	—	—	−0.9949	−0.1012
x_{10}	—	—	−0.4984	0.8669

X-গুচ্ছের চলকসমূহের সাথে বেশি সংশ্লেষিত এবং x_9 Y-গুচ্ছের চলকসমূহের সাথে বেশি সংশ্লেষিত। পরবর্তী গুরুত্বপূর্ণ চলক হলো x_7 এবং x_{10}।

প্রতিটি কানুনী চলক জোড়া X-গুচ্ছ ও Y-গুচ্ছের চলকগুলির মোট ভেদের কিছু অংশ ব্যাখ্যা করতে পারে। i-তম কানুনী চলক জোড়া দ্বারা Y-গুচ্ছের ও X-গুচ্ছের চলকসমূহের ভেদের ব্যাখ্যা করা সমানুপাত হলো যথাক্রমে $R^2_{(i)y}$ এবং $R^2_{(i)x}$। সারণি ৬.২৭-এ কানুনী চলক জোড়া দ্বারা দুই গুচ্ছের চলকসমূহের ভেদের ব্যাখ্যা করা সমানুপাত দেয়া হলো। দেখা যাচ্ছে যে, প্রথম কানুনী চলক জোড়া X-গুচ্ছের চলকসমূহের ভেদের 69% এবং Y-গুচ্ছের চলকসমূহের ভেদের 61.91% ব্যাখ্যা করেছে। এছাড়া X-গুচ্ছ দ্বারা Y-গুচ্ছের ভেদের প্রকাশিত সমানুপাত হলো

সারণি ৬.২৭ : কানুনী চলক জোড়া দ্বারা X-গুচ্ছ ও Y-গুচ্ছের চলকসমূহের ভেদের ব্যাখ্যা সমানুপাত।

কানুনী চলক জোড়া	ভেদের ব্যাখ্যা করা সমানুপাত	
	X-গুচ্ছ, $R^2_{(i)x}$	Y-গুচ্ছ, $R^2_{(i)y}$
1	0.0690	0.6191
2	0.0473	0.3809

$$\lambda_1 R^2_{(1)y} = 0.4939 \text{ এবং } \lambda_2 R^2_{(2)y} = 0.0816$$

এই বিশ্লেষণের মূল উদ্দেশ্য হলো Y-গুচ্ছের যে কোনো চলকের সাথে X-গুচ্ছের যে কোনো চলকের সংশ্লেষণ পর্যালোচনা করা। এই সংশ্লেষণের পরিমাণ নির্ণয় করা হয় ক্রস ভার নির্ণয় করে। i-তম কানুনী চলক জোড়ার ক্ষেত্রে Y-গুচ্ছের l-তম $[l = 1, 2, \cdots, p]$ চলকের জন্য এই ক্রস ভার হলো $\sqrt{\lambda_i}\, b_{li}'$। অর্থাৎ প্রথম কানুনী চলক জোড়ার ক্ষেত্রে Y-গুচ্ছের x_9 এবং X-গুচ্ছের অন্যান্য চলকের মধ্যে সংশ্লেষণের পরিমাণ হলো $\sqrt{0.7978}\ (-0.9949) = -0.8886$। আবার Y-গুচ্ছের x_{10} এবং X-গুচ্ছের অন্যান্য চলকের মধ্যে সংশ্লেষণের পরিমাণ হলো $\sqrt{0.7978}\ (-0.4984) = -0.4452$। অনুরূপভাবে i-তম কানুনী চলক জোড়ার ক্ষেত্রে X-গুচ্ছের j-তম চলকের জন্য ক্রস ভার হলো $\sqrt{\lambda_i}\, a_{ji}'$। উভয় চলক জোড়ার ক্ষেত্রে এই ক্রস ভার সারণি ৬.২৮-এ উপস্থাপন করা হলো।

সারণি ৬.২৮ : X-গুচ্ছের এবং Y-গুচ্ছের চলকসমূহের জন্য ক্রস ভার।

চলকসমূহ	i-তম কানুনী চলকের ক্ষেত্রে			
	X-গুচ্ছের চলকের ক্রস ভার		Y-গুচ্ছের চলকের ক্রস ভার	
	λ_1	λ_2	λ_1	λ_2
x_1	0.2155	0.0403	—	—
x_2	0.0254	0.0145	—	—

x_3	0.0147	0.0855	—	—
x_4	−0.0426	0.0574	—	—
x_5	0.0463	−0.1106	—	—
x_6	0.1386	0.1436	—	—
x_7	−0.3885	0.1600	—	—
x_8	0.4686	−0.1009	—	—
x_9	—	—	−0.8886	−0.0468
x_{10}	—	—	−0.0.4452	0.4012

এই বিশ্লেষণ হতে সিদ্ধান্ত নেয়া যায় যে, Vertebrae-এর বৈশিষ্ট্য পর্যালোচনা করার ক্ষেত্রে চলক x_9 ও x_{10} চলক x_7 ও x_8 দ্বারা বেশি প্রভাবিত। Centrum-এর দৈর্ঘ্য (x_1) ও কোণদ্বয়ের $[x_9, x_{10}]$ পরিমাণে গুরুত্বপূর্ণ ভূমিকা পালন করে।

সপ্তম অধ্যায়

পরীক্ষণের নকশা ও ভেদাঙ্ক বিশ্লেষণ
(Design of Experiments and Analysis of Variance)

৭.১ সূচনা (Introduction)

কৃষি, শিল্প, জীববিজ্ঞান এবং সামাজিক বিজ্ঞানে এমন অনেক ধরনের পরীক্ষা পরিচালনা করা হয়ে থাকে যেগুলির মূল উদ্দেশ্য হলো অনুমান সিদ্ধ কল্পনা যাচাই করা। ধরা যাক এক প্রকার কীটকে মারার জন্য পাঁচ প্রকার Pesticide প্রয়োগ করে কোনো পরীক্ষাগারে একটি পরীক্ষা পরিচালনা করা হয়েছে। এই পরীক্ষার মূল উদ্দেশ্য হলো Pesticide গুলি কীট মারার জন্য অনুমোদিত হওয়ায় কোনো বিশেষ Pesticide কীটকে বেশি পরিমাণে বা তাড়াতাড়ি মারতে পারে তা চিহ্নিত করা। এখানে বিশেষ প্রকার কীট বা অনেক ধরনের কীট সংগ্রহ করে একটি বিশেষ অবস্থায় বা বিভিন্ন অবস্থায় বিভিন্ন প্রকার Pesticide প্রয়োগ করে কীটের মৃত্যুর পরিমাণ বা কয় দিনে কীটগুলি মারা যায় তা পর্যবেক্ষণ করা হয়। এরূপ পর্যবেক্ষণকে বলা হয় পরীক্ষা (experiment)। এ জাতীয় পরীক্ষার প্রাথমিক কাজ হলো উপাত্ত সংগ্রহ করা (data collection)। এই উপাত্ত সংগ্রহ করার বিভিন্ন পদ্ধতি আছে। এখানে পূর্ব নির্ধারিত উদ্দেশ্যে উপাত্ত সংগ্রহ করার পদ্ধতিকে বলা হয় পরীক্ষণের নকশা (design of experiment)।

উদ্দেশ্য অনুসারে পরীক্ষণের নকশার প্রকারভেদ হতে পারে। আবার পরীক্ষারও প্রকারভেদ থাকতে পারে। যেমন, কতকগুলো Slug সংগ্রহ করে সেগুলোর উপর বিভিন্ন Pesticide প্রয়োগ করে ঐগুলোর Heart beat/minute লক্ষ্য করা যেতে পারে (সারণি ১.৩)। এখানে উদ্দেশ্য হতে পারে Heart beat-এর উপর Pesticide-এর প্রভাব কি রকম এবং কোন Pesticide Heart beat-কে বেশি প্রভাবিত করে—তা পর্যবেক্ষণ করা। এই উদ্দেশ্যে Slug-এর প্রকারভেদ বিবেচনা করে উপাত্ত সংগ্রহ করা যেতে পারে। আবার কোন প্রকার Slug-এর ক্ষেত্রে কোন Pesticide Heart beat-কে বেশি প্রভাবিত করে তাও পর্যবেক্ষণ করা যেতে পারে। এই শেষোক্ত উদ্দেশ্যে উপাত্ত সংগ্রহ করার ক্ষেত্রে Slug-এর প্রকারভেদ এবং একই সময়ে Pesticide-এর প্রকারভেদের বিষয় বিবেচনা করতে হবে। কাজেই বুঝা যাচ্ছে যে, উদ্দেশ্য

অনুসারে এবং নিরীক্ষার জন্য ব্যবহৃত সামগ্রীর ভেদ অনুসারে পরীক্ষণের নকশা বিভিন্ন হয়।

উদ্দেশ্য যাই হোক না কেন এবং পরীক্ষণের নকশা যাই প্রয়োগ করা হোক না কেন পরীক্ষালব্ধ ফলাফলে ভেদ থাকবেই। সকল উপাত্তের ভেদ পরিমাপ করার জন্য ভেদাঙ্ক (variance) বা মোট বর্গসমষ্টি (Total Sum of Squares, TSS) নির্ণয় করা হয়। এই ভেদাঙ্ক অনেক কারণে হতে পারে। যেমন, সারণি ১.৩-এর (১ম খণ্ড) ক্ষেত্রে ভেদাঙ্কের উৎস দুটি, (১) Slug-এর প্রকারভেদের জন্য ভেদাঙ্ক এবং (২) Pesticide এর প্রকারভেদের জন্য ভেদাঙ্ক। সুতরাং পরীক্ষা লব্ধ ফলাফল বিশ্লেষণের একটি অঙ্গ হলো ভেদাঙ্ক বিশ্লেষণ এবং ভেদাঙ্ক বিশ্লেষণ হলো উপাত্তের মোট ভেদাঙ্ককে ভেদাঙ্কের চিহ্নিত উৎস অনুসারে কয়েকটি অংশে বিভক্ত করা এবং ঐগুলির তুলনা করা। R.A. Fisher সর্ব প্রথম ভেদাঙ্ক বিশ্লেষণের পদ্ধতি উদ্ভাবন করেন। তাঁর ভাষায়, "ভেদাঙ্ক বিশ্লেষণ হলো একগুচ্ছ কারণের জন্য উদ্ভূত ভেদাঙ্ক হতে বিভিন্ন চিহ্নিতকরণের জন্য উদ্ভূত ভেদাঙ্ককে পৃথকীকরণ। বর্তমান অধ্যায়ে সংক্ষিপ্তভাবে পরীক্ষণের নকশার প্রয়োগ এবং ঐ নকশা হতে প্রাপ্ত ফলাফলের ভেদাঙ্ক বিশ্লেষণ পদ্ধতি আলোচনা করা হবে। এ সম্পর্কে বিস্তারিত জানার জন্য Cochran and Cox (1957), Kempthorne (1952), Federer (1955), John (1971), Winer (1971) ইত্যাদি লেখকের বই পর্যালোচনা করা যেতে পারে।

৭.২ পরীক্ষণের সাথে জড়িত কিছু শব্দাবলি (Terms Related with Experiment)

(১) চর্যা (treatment): পরীক্ষায় যে বিষয়বস্তুর প্রভাব পরিমাপ করা হয়, তুলনা করা হয় এবং যার প্রভাব সম্বন্ধে অনুমান সিদ্ধ কল্পনা যাচাই করা হয় তাকে চর্যা বলে। যেমন, কৃষি উৎপাদনে সারের বিভিন্ন স্তর, পোকামাকড় মারার জন্য বিভিন্ন কীট-নাশক, দুধ উৎপাদন বৃদ্ধিতে বিভিন্ন গো-খাদ্য, ডায়াবেটিস নিয়ন্ত্রণে রাখার জন্য বিভিন্ন স্তরের ইনসুলিন ইত্যাদি বিষয়ের প্রভাব পরিমাপ করা যায় এবং এগুলির প্রভাব সম্বন্ধে নাস্তি কল্পনা যাচাই করা যায়। এ কারণে এগুলিকে চর্যা বলে।

(২) পরীক্ষণের খণ্ড বা একক (Experimental plot or unit): যে কোনো পরীক্ষায় যে কোনো একটি চর্যাকে যে বস্তুতে বা যে জমিতে বা যে পশু বা পশুসমূহে বা যে গাছ বা গাছসমূহে প্রয়োগ করা হয় তাকে পরীক্ষণের খণ্ড বা একক বলে। যেমন, Pesticide প্রয়োগে Slug-এর Heart beat-এ কি পরিবর্তন হয় তা জানার জন্য যে কোনো একটি Slug বা কতগুলি Slug-এর উপর Pesticide-এর একটি স্তর প্রয়োগ করা হলে ঐ Slug বা Slug-গুলিকে পরীক্ষণের একক বলা হবে।

(৩) উৎপাদন (Yield) : পরীক্ষা কাজ পরিচালনা করার পর উক্ত কাজের ফলাফলকে উৎপাদন বলা হয়। যেমন, Pesticide প্রয়োগে Slug-এর Heart beat-এর পরিবর্তন লক্ষ্য করার জন্য পরীক্ষা কাজ পরিচালনা করা হলে পরীক্ষা পরবর্তী Slug-এর Heart beat হলো উৎপাদন। Blood sugar স্তর কমাবার জন্য চিকিৎসা করা হলে চিকিৎসা পরবর্তী Blood sugar স্তর হলো উৎপাদন। ফুলের উৎপাদন বৃদ্ধির জন্য সার প্রয়োগ করা হলে সার প্রয়োগ করা গাছের ফুলের পরিমাণ হলো উৎপাদন।

(৪) ব্লক (Block) : একগুচ্ছ সমমাত্রিক পরীক্ষণ একককে ব্লক হয়। যেমন, সারণি ১.৩-এ (১ম খণ্ড) Milux Rusticus নামক Slug বা Milux Sowerbyi নামক কতকগুলি Slug-এর উপর বিভিন্ন Pesticide প্রয়োগ করা হয়েছে। এখানে Milux Rusticus Slug বা Milux Sowerbyi Slug-গুলি একটি ব্লক হিসেবে বিবেচিত হতে পারে।

(৫) পরীক্ষণের বিচ্যুতি (Experiment error) : যে কোনো নিরীক্ষায় পরীক্ষণের বিষয়বস্তু (চর্যা) ছাড়া অন্য কোনো উৎস দ্বারাও পরীক্ষার ফলাফল প্রকাশিত হয়ে থাকে। যেমন, ফুলের উৎপাদনে সারের প্রভাব নিরূপণ করার জন্য নিরীক্ষা কাজ পরিচালিত হলে ফুলের উৎপাদন পোকামাকড় দ্বারা বা অতিবৃষ্টি দ্বারা ক্ষতিগ্রস্ত হতে পারে। এরূপ ক্ষেত্রে পোকামাকড়ের উপদ্রব বা অতিবৃষ্টির প্রভাব সব সময় পরীক্ষণের মাধ্যমে নিয়ন্ত্রণ করা সম্ভব নয়। এ জাতীয় অনিয়ন্ত্রিত উৎসগুলির কারণে সঠিক উৎপাদনের পরিবর্তে বিচ্যুতি (error) সহ উৎপাদন নথিভুক্ত হয়। এখানে অনিয়ন্ত্রিত উৎসের কারণে যে বিচ্যুতির উদ্ভব হয় তাকে পরীক্ষণের বিচ্যুতি বলা হয়। পরীক্ষণের বিচ্যুতি যতোটি কমানো যায় সেদিকে নিরীক্ষকের লক্ষ্য রাখা দরকার। কারণ, পরীক্ষণের বিচ্যুতি কম হলেই পরীক্ষণটি যথাযথ হবে।

(৬) দৈবায়িতকরণ (Randomisation) : পরীক্ষা কাজ পরিচালনা করার জন্য দৈবায়িতকরণ একটি গুরুত্বপূর্ণ ধাপ। দৈবায়িতকরণের অর্থ হলো চর্যাকে পরীক্ষা এককে দৈব পদ্ধতিতে বণ্টন করা। গবেষক নিজের সুবিধামত বা ইচ্ছামত কোনো চর্যাকে কোনো এককে বণ্টন করতে পারবে না। চর্যার বণ্টন ব্যবস্থা দৈব পদ্ধতিতে করা হলে পরীক্ষণ বিচ্যুতি ন্যূনতম হয় এবং বিচ্যুতিগুলি অনপেক্ষ (independent) হয়।

দৈবায়িতকরণের জন্য মোট পরীক্ষণ একক হতে কতকগুলি একক দৈব পদ্ধতিতে চয়ন করে (selecting units on the basis of random number table) ঐগুলিতে যে কোনো একটি চর্যা বণ্টন করা যেতে পারে।

(৭) পুনরায়ন (Replication) : পরীক্ষার কাজ বারবার করা বা কোনো চর্যাকে একাধিক পরীক্ষণ এককে বণ্টন করা হলো পুনরায়ন। পরীক্ষণ বিচ্যুতির ভেদাঙ্কের

নিরূপক পাওয়ার জন্য পুনরায়ন করা হয় এবং এই নিরূপক পরীক্ষার যথার্থতা পরিমাপ করার জন্য ব্যবহৃত হয়।

ধরা যাক কোনো চর্যাকে r এককে দৈব পদ্ধতি বণ্টন করা হয়েছে। তাহলে r হবে ঐ চর্যার পুনরায়নের সংখ্যা। আবার প্রতিটি চর্যাকেও r এককে বণ্টন করা যায়। সেক্ষেত্রে পরীক্ষাটির পুনরায়নের সংখ্যা হবে r। এখানে

$$r = \frac{2s^2 t_0^2}{d^2}$$

যেখানে $d =$ দুটি চর্যার প্রভাবের পার্থক্য

$s^2 =$ বিচ্যুতির ভেদাঙ্কের নিরূপক

$t_0 =$ ঈপ্সিত স্তরের জন্য (সাধারণত 5% level) t-বিন্যাসের বর্জনীয় (critical value) মান।

কোনো পরীক্ষার ক্ষেত্রে s^2 জানা থাকলে বা যদি এটি বড় আকারের কোনো পরীক্ষা থেকে নিরূপিত হয়, তাহলে t_0-এর পরিবর্তে পরিমিত বিন্যাসের মান (সাধারণত 5% স্তরে $Z = 1.96$) ব্যবহার করতে হয়।

স্থানীয় নিয়ন্ত্রণ (Local control) : স্থানীয় নিয়ন্ত্রণ হলো পরীক্ষণের একক-গুলিকে সমমাত্রিক বা সমগোত্রীয় দলে বা গুচ্ছে ভাগ করা। যেমন, Slug-এর Heart beat-এ Pesticide-এর প্রভাব নিরূপণ করার পরীক্ষণ কাজে একই জাতের বা একই ওজনের বা একই দৈর্ঘ্যের কতকগুলি Slug চয়ন করা হলো স্থানীয় নিয়ন্ত্রণ।

স্থানীয় নিয়ন্ত্রণ পরীক্ষার কার্যকারিতা এবং যথার্থতা বৃদ্ধি করে। এটি পরীক্ষণ বিচ্যুতি কমাতে সাহায্য করে। এ কারণে একে বিচ্যুতির নিয়ন্ত্রণও বলা হয়।

উপরে আলোচিত দৈবায়িতকরণ, পুনরায়ন এবং স্থানীয় নিয়ন্ত্রণ হলো পরীক্ষণের নকশার মূলনীতি (basic principle)।

৭.৩ ভেদাঙ্ক বিশ্লেষণের জন্য প্রতিকৃতি (Model for Analysis of Variance)

আগেই উল্লেখ করা হয়েছে, ভেদাঙ্ক বিশ্লেষণ হলো পরীক্ষালব্ধ উপাত্তের মোট ভেদাঙ্ককে তার চিহ্নিত উৎসসমূহের ভেদাঙ্কে পৃথককরণ। এখানে এক একটি চিহ্নিত উৎসকে এক একটি উপাদান (factor) বলা যায়। যেমন, সারণি ১.৩-এর (১ম খণ্ড) উপাত্তের ক্ষেত্রে Slug-এর Heart beat-এর ভেদাঙ্কের দুটি চিহ্নিত উৎস আছে— একটি হলো Pesticide এবং অপরটি হলো Slug-এর প্রকার ভেদ। এখানে Pesticide হলো একটি উপাদান এবং Slug-এর প্রকারভেদ হলো অন্য একটি উপাদান। বিভিন্ন Slug-এর বিভিন্ন Hart beat হওয়ার কারণ হলো বিভিন্ন Pesticide-এর বিভিন্ন প্রভাব এবং বিভিন্ন প্রকার Slug-এর

বিভিন্ন প্রকার প্রভাব। আবার একই প্রকার Slug-এর উপর একটি Pesticide প্রয়োগ করেও দুই রকম Heart beat লক্ষ্য করা যাচ্ছে। এখানে শেষোক্ত ক্ষেত্রে Heart beat-এর বিভিন্নতার উৎস হলো পরীক্ষণ বিচ্যুতি। কাজেই বুঝা যাচ্ছে যেপরীক্ষালব্ধ ফলাফল কতকগুলি উপাদানের যৌথ প্রভাব। এ কারণেই পরীক্ষালব্ধ ফলাফলকে কতকগুলি প্রভাবের ফাংশন হিসেবে প্রকাশ করা হয়। এই প্রভাবগুলি দৈব চলক (random variable) হতে পারে। আবার গাণিতিক চলকও (mathematical variable) হতে পারে।

দৈব প্রভাব (Random effect) : কোনো পরীক্ষণ কাজের জন্য চর্যার স্তর যদি গণসমষ্টি স্তর হতে দৈব পদ্ধতিতে চয়ন করা হয়, তাহলে ঐ পরীক্ষার ক্ষেত্রে চর্যার প্রভাব দৈব প্রভাব হবে বা চর্যার প্রভাব দৈব চলক হিসেবে বিবেচিত হবে। যেমন, Wood louse মারার জন্য Pesticide প্রয়োগ করা একটি পরীক্ষণের উদ্দেশ্য। এখন Pesticide-এর বিভিন্ন শ্রেণী আছে। যথা : (১) Supracide, (২) Pomex, (৩) Sumicidin (৪) Glyphosate (৫) Milkrup ইত্যাদি। এখন কোনো পরীক্ষায় এই পাঁচটি Pesticide প্রয়োগ না করে যদি দৈব পদ্ধতিতে তিনটি চয়ন করে {যেমনটি করা হয়েছে সারণি ১.৫-এর (১ম খণ্ড) উপাত্তের ক্ষেত্রে} প্রয়োগ করা হয়, তাহলে Pesticide-এর প্রভাব দৈব প্রভাব হবে।

স্থির প্রভাব (Fixed effect) : কোনো পরীক্ষায় চর্যার সম্ভাব্য সকল স্তর ব্যবহৃত হলে, চর্যার প্রভাব স্থির প্রভাব হবে বা চর্যার প্রভাব গাণিতিক চলক হিসেবে বিবেচিত হবে। যেমন, কোনো পরীক্ষাগারে পাঁচটি নতুন প্রোটিন আবিষ্কার করা হয়েছে এবং এই পাঁচটির মধ্যে কোনটি অধিক ফলপ্রসূ তা নিরীক্ষা করার জন্য একটি পরীক্ষা পরিচালিত হবে। এক্ষেত্রে আবিষ্কৃত পাঁচটি প্রোটিন কোনো জীব-জন্তুর উপর প্রয়োগ করে পরীক্ষা পরিচালনা করা হলে, প্রোটিনের প্রভাব স্থির প্রভাব হবে।

চর্যার বা উপাদানের প্রভাব দৈব হবে কি স্থির হবে তা ভেদাঙ্ক বিশ্লেষণের ক্ষেত্রে একটি গুরুত্বপূর্ণ বিবেচ্য বিষয়। এই বিবেচনার উপর ভিত্তি করে বিশ্লেষণের বিভিন্ন ধাপ পরিচালিত হয়। বিশ্লেষণের প্রাথমিক ধাপ হিসেবে পরীক্ষালব্ধ ফলাফলকে প্রতিকৃতি (model)-এর মাধ্যমে প্রকাশ করা হয় এবং প্রতিকৃতিতে ব্যবহৃত বিভিন্ন চলক (স্থির বা দৈব) সম্বন্ধে অনুমান করতে হয়। অনুমান সম্পর্কে পরবর্তী অনুচ্ছেদে আলোচনা করা হবে। এখানে প্রতিকৃতি সম্পর্কে একটি সংক্ষিপ্ত ধারণা ব্যক্ত করা হলো।

প্রতিকৃতি (Model) : প্রতিকৃতি বা মডেল হলো একটি গাণিতিক সমীকরণ যা দৈব চলক, গাণিতিক চলক এবং পরামান-এর সমন্বয়ে গঠিত। নির্ভরণ বিশ্লেষণের ক্ষেত্রে এ সম্পর্কে আলোচনা করা হয়েছে। এই প্রতিকৃতিকে প্রধানত দুইভাগে ভাগ

করা যায়, যথা (১) রৈখিক প্রতিকৃতি এবং (২) অরৈখিক প্রতিকৃতি। ভেদাঙ্ক বিশ্লেষণের জন্য রৈখিক প্রতিকৃতিই সাধারণত ব্যবহৃত হয়।

রৈখিক প্রতিকৃতি (Linear model) : যে সমীকরণে দৈব চলক, গাণিতিক চলক এবং পরামান থাকে এবং সমীকরণটি দৈবচলক ও পরামানের ভিত্তিতে রৈখিক (linear) হয় তাকে রৈখিক প্রতিকৃতি বা রৈখিক মডেল বলা হয়। যেমন :

$$y_{ij} = \mu + \alpha_i + e_{ij} \; ; \; i = 1, 2, \ldots, p \; ; \; j = 1, 2, \cdots, q \qquad (৭.৩.১)$$

এখানে y_{ij}-গুলি হলো একটি চর্যার p স্তর ব্যবহার করে প্রতি স্তরকে q বার পুনরায়ন করে পরিচালিত পরীক্ষালব্ধ ফলাফল।

$\mu = y_{ij}$-গুলির গণসমষ্টি (population) গড় (mean)।

α_i = i-তম চর্যার প্রভাব,

e_{ij} = j-তম পুনরায়নে i-তম চর্যার ফলাফলে বিচ্যুতি।

এখানে e_{ij} দৈব চলক এবং সে কারণে y_{ij}-ও দৈব চলক। μ পরামান, α_i পরামানও (স্থির প্রভাব) হতে পারে, আবার দৈব চলকও (দৈব প্রভাব) হতে পারে।

চর্যার প্রভাবের স্থিরতা বা দৈবতা-এর কারণে প্রতিকৃতির নামকরণও বিভিন্ন হয়। যেমন : (১) স্থির প্রভাব প্রতিকৃতি, (২) দৈব প্রভাব প্রতিকৃতি এবং (৩) মিশ্র প্রভাব প্রতিকৃতি।

স্থির প্রভাব প্রতিকৃতি (Fixed effect model) : স্থির প্রভাব সমন্বয়ে গঠিত কোনো প্রতিকৃতিকে বলা হয় স্থির প্রভাব প্রতিকৃতি। উপরে (৭.৩.১) প্রতিকৃতির ক্ষেত্রে α_i হলো চর্যার প্রভাব। এই প্রভাব স্থির হলে (৭.৩.১) প্রতিকৃতিকে বলা হবে স্থির প্রভাব প্রতিকৃতি এবং α_i হলো পরামান।

দৈব প্রভাব প্রতিকৃতি (Random effect model) : কোনো প্রতিকৃতির সবগুলি প্রভাবই দৈব চলক হলে ঐ প্রতিকৃতিকে দৈব প্রভাব প্রতিকৃতি বলা হয়। ধরা যাক (৭.৩.১) প্রতিকৃতির i-তম চর্যার প্রভাব দৈব চলক। তাহলে ঐ প্রতিকৃতি দৈব প্রভাব প্রতিকৃতি হবে। প্রতিকৃতি স্থির প্রভাব হোক বা দৈব প্রভাব হোক না কেন গণসমষ্টি গড় μ সব সময়ই পরামান।

মিশ্র প্রভাব প্রতিকৃতি (Mixed effect model) : কোনো প্রতিকৃতির কিছু প্রভাব পরামান (স্থির প্রভাব) এবং কিছু প্রভাব দৈব চলক (দৈব প্রভাব) হলে ঐ প্রতিকৃতিকে মিশ্র প্রভাব প্রতিকৃতি বলা হয়। ধরা যাক সারণি ১.৩-এর (১ম খণ্ড) ক্ষেত্রে Slug-এর চারটি প্রকার আছে এবং ঐ চারটি প্রকার হতে দুটি দৈব পদ্ধতিতে চয়ন করে ঐগুলির উপর সম্ভাব্য সকল প্রকার Pesticide প্রয়োগ করে Slug-এর Heart beat/minute লক্ষ্য করা হয়েছে। প্রতিটি Pesticide যে কোনো প্রকার Slug-এর

দুটির উপর প্রয়োগ করে দুটিরই Heart beat নথিভুক্ত করা হয়েছে। এক্ষেত্রে Heart beat-এর উপর Slug-এর প্রকারের যে প্রভাব তা দৈব প্রভাব এবং Pesticide-এর প্রভাব স্থির প্রভাব। ধরা যাক $\alpha_i (i = 1, 2)$ হলো i-তম Slug-এর প্রকারের প্রভাব এবং β_j হলো $(j = 1, 2, \cdots, 6)$ j-তম Pesticide-এর প্রভাব। তাহলে উক্ত নিরীক্ষার ক্ষেত্রে পরীক্ষালব্ধ ফলাফলের জন্য প্রতিকৃতি হবে

$$y_{ijl} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ijl} \qquad (৭.৩.২)$$

এখানে y_{ijl} = i-তম Slug-এর প্রকারের উপর j-তম Pesticide প্রয়োগ করার পর l-তম $(l = 1, 2)$ Slug-এর Heart beat,

$\mu = y_{ijl}$ তথ্যমানসমূহের গণসমষ্টি গড়। একে সাধারণত বলা হয় সাধারণ গড় (general mean),

$(\alpha\beta)_{ij}$ = j-তম Pesticide ও i-তম Slug-এর প্রকারের যৌথ প্রভাব ỉ (interaction effect),

e_{ijl} = দৈব বিচ্যুতি।

এখানে α_i দৈব প্রভাব হওয়াতে $(\alpha\beta)_{ij}$-ও দৈব প্রভাব। সুতরাং প্রতিকৃতি (৭.৩.২) হলো মিশ্র প্রভাব প্রতিকৃতি।

৭.৪ ভেদাঙ্ক বিশ্লেষণের জন্য অনুমান (Assumptions for Analysis of Variance)

আগেই উল্লেখ করা হয়েছে যে, ভেদাঙ্ক বিশ্লেষণ হলো এক গুচ্ছ উপাত্তের মোট ভেদাঙ্ককে কতকগুলি চিহ্নিত উৎস-এর ভেদাঙ্কে পৃথককরণ। প্রতিটি উৎস-এর প্রভাবকে পরামান বা দৈব চলক দ্বারা চিহ্নিত করা হয় এবং ভেদাঙ্ক বিশ্লেষণের একটি অঙ্গ হিসেবে ঐ পরামান সম্বন্ধে (স্থির প্রভাব প্রতিকৃতির ক্ষেত্রে) বা দৈব চলকের (দৈব প্রভাব প্রতিকৃতির ক্ষেত্রে) ভেদাঙ্ক-এর তাৎপর্য সম্বন্ধে নাস্তি কল্পনা যাচাই করা হয়। সে কারণেই বিশ্লেষণের জন্য কিছু অনুমান করতে হয়। এখানে ঐ অনুমান-সমূহ আলোচনা করা হলো।

ধরা যাক কোনো পরীক্ষায় উপাদান B-এর j-তম স্তরের $[j = 1, 2, \ldots, q]$ প্রাসঙ্গিক উপাদান A-এর i-তম স্তরের $[i = 1, 2, \cdots, p]$ তথ্য মানের $[y_{ij}]$ বিচ্যুতি e_{ij}, A-এর i-তম স্তরের প্রভাব α_i এবং B-এর j-তম স্তরের প্রভাব β_j। তাহলে এই পরীক্ষণ হতে প্রাপ্ত তথ্যমান $[y_{ij}]$-এর জন্য প্রতিকৃতি হবে

$$y_{ij} = \mu + \alpha_i + \beta_j + e_{ij} \qquad (৭.৪.১)$$

এই প্রতিকৃতি অনুসারে ভেদাঙ্ক বিশ্লেষণ করার জন্য প্রাথমিক অনুমান হলো

(i) প্রতিকৃতি স্থির প্রভাববিশিষ্ট। অন্যান্য অনুমান হলো

(ii) e_{ij} পরিমিত বিন্যাস (normal distribution) অনুসরণ করে যার

$$E(e_{ij}) = 0$$

$$E(e_{ij}, e_{ik}) = 0\,, \quad \text{যদি } j \neq k$$

$$= \sigma^2, \quad \text{যদি } j = k$$

অর্থাৎ বিচ্যুতি শূন্য গড় এবং সাধারণ ভেদাঙ্ক (Common variance বা Homogeneous variance) σ^2 সহ বিন্যাসিত।

(iii) e_{ij}-গুলি অনপেক্ষভাবে বিন্যাস। অনুমান (ii) এবং (iii)-কে সংক্ষিপ্তভাবে লেখা যায়

$$e_{ij} \sim \text{NID}\,(0, \sigma^2)$$

প্রতিকৃতির কোনো প্রভাব দৈব হলে (দৈব প্রভাব প্রতিকৃতির ক্ষেত্রে বা মিশ্র প্রভাব প্রতিকৃতির ক্ষেত্রে) সে সম্পর্কেও অনুমান করতে হয়। সাধারণত অনুমান করা হয় যে, যে কোনো দৈব প্রভাব শূন্য গড় এবং একটি সমমাত্রিক ভেদাঙ্কসহ পরিমিত বিন্যাস অনুসরণ করে এবং যে কোনো প্রভাবের স্তরসমূহ অনপেক্ষভাবে বিন্যাসিত। যেমন, প্রতিকৃতি (৭.৪.১)-এর ক্ষেত্রে ধরা যাক α_i দৈব প্রভাব। তাহলে, অনুমান হবে $\alpha_i \sim \text{NID}\,(0, \sigma^2)$।

৭.৫ সম্পূর্ণ দৈবায়িত নকশা (Completely Randomized Design)

ধরা যাক কোনো পরীক্ষার জন্য P চর্যা আছে এবং i-তম $[\, i = 1, 2, \cdots, P \,]$ চর্যাকে n_i বার পুনরায়ন করে পরীক্ষণ কাজটি পরিচালিত করতে হবে। তাহলে P চর্যার জন্য মোট $\sum_{i=1}^{p} n_i = n$ পরীক্ষণ এককের [খণ্ডের] প্রয়োজন হবে। ধরা যাক এই n একক সমমাত্রিক (homogeneous)। এখন n একক হতে দৈব পদ্ধতিতে n_1 একক চয়ন করে ঐগুলিতে প্রথম চর্যাকে বণ্টন (allocate) করা যাক। পরবর্তীতে $(n - n_1)$ একক হতে n_2 একক দৈব পদ্ধতিতে চয়ন করে ঐগুলিতে

দ্বিতীয় চর্যাকে বণ্টন করা যেতে পারে। এভাবে সকল চর্যা পরীক্ষণ এককে বণ্টিত না হওয়া পর্যন্ত দৈব পদ্ধতিতে একক চয়ন করা এবং চর্যার বণ্টন ব্যবস্থা চলতে থাকবে। চর্যার এরূপ বণ্টন পদ্ধতিতে যে নকশার উদ্ভব হয় তাকে সম্পূর্ণ দৈবায়িত নকশা বলা হয়।

এই নকশার ক্ষেত্রে পরীক্ষণ একক সমমাত্রিক হওয়ায় পরীক্ষণ বিচ্যুতি ভিন্ন পরীক্ষালব্ধ ফলাফলে ভেদের চিহ্নিত উৎস হলো একটি। এই উৎস হলো চর্যা। সুতরাং পরীক্ষালব্ধ উপাত্তের মোট ভেদাঙ্ককে একটি চিহ্নিত উৎসের ভেদাঙ্ক এবং বিচ্যুতির ভেদাঙ্কে পৃথকীকরণ করা যায় বলে এই নকশা হতে প্রাপ্ত উপাত্তের বিশ্লেষণকে এক-মুখী শ্রেণী বিন্যাস বলা হয়। এই বিশ্লেষণের জন্য প্রতিকৃতি হলো

$$y_{ij} = \mu + \alpha_i + e_{ij} \qquad (৭.৫.১)$$

$$i = 1, 2, \cdots, P; \quad j = 1, 2, \cdots, n_i$$

এখানে y_{ij} = i-তম চর্যার j-তম পুনরায়নের ফলাফল,

μ = সাধারণ গড়,

α_i = i-তম চর্যার প্রভাব,

e_{ij} = দৈব বিচ্যুতি।

অনুমান : (i) $e_{ij} \sim NID(0, \sigma^2)$, (ii) প্রতিকৃতি স্থির প্রভাববিশিষ্ট। সুতরাং সাধারণ ন্যূনতম বর্গপদ্ধতি (OLS method) প্রয়োগ করে বিশ্লেষণ করা যায়।

এই বিশ্লেষণের ক্ষেত্রে উপাত্তের

মোট বর্গসমষ্টি, $TSS = \Sigma\Sigma(y_{ij} - \bar{y}_{..})^2$

চর্যার বর্গসমষ্টি, $S_1 = \Sigma n_i(\bar{y}_{i.} - \bar{y}_{..})^2$

এবং বিচ্যুতির বর্গসমষ্টি, $S_2 = \Sigma\Sigma(y_{ij} - \bar{y}_{i.})^2$

এখানে, $TSS = S_1 + S_2 = \Sigma n_i(\bar{y}_{i.} - \bar{y}_{..})^2 + \Sigma\Sigma(y_{ij} - \bar{y}_{i.})^2$

$$\bar{y}_{..} = \frac{1}{n}\Sigma\Sigma y_{ij}, \quad \bar{y}_{i.} = \frac{1}{n_i}\sum_j y_{ij}$$

পরীক্ষালব্ধ ফলাফলকে সারণিকৃত করে সাজিয়ে লেখা যায় নিম্নরূপ :

সারণি ৭.১ : সম্পূর্ণ দৈবায়িত নকশা হতে প্রাপ্ত ফলাফল।

পুনরায়ন	চর্যা					
	1	2	⋯	i	⋯	P
	y_{11}	y_{21}		y_{i1}	⋯	y_{p1}
	y_{12}	y_{22}	⋯	y_{i2}	⋯	y_{p2}
	⋮	⋮	⋯	⋮	⋯	⋮
	y_{1n_1}	y_{2n_2}	⋯	y_{in_i}	⋯	y_{pn_p}
মোট	$y_{1.}$	$y_{2.}$	⋯	$y_{i.}$	⋯	$y_{p.}$
গড়	$\bar{y}_{1.}$	$\bar{y}_{2.}$	⋯	$\bar{y}_{i.}$	⋯	$\bar{y}_{p.}$

উপরিউক্ত নকশার ক্ষেত্রে $n_1 = n_2 = \cdots\cdots = n_p = q$ হলে

$$\Sigma\Sigma(y_{ij} - \bar{y}_{..})^2 = q\Sigma(\bar{y}_{i.} - \bar{y}_{..})^2 + \Sigma\Sigma(y_{ij} - \bar{y}_{i.})^2$$

এবং $$\bar{y}_{..} = \frac{1}{pq}\Sigma y_{ij}\ ,\quad \bar{y}_{i.} = \frac{1}{q}\sum_{j} y_{ij}$$

এই বিশ্লেষিত ফলাফলকে ভেদাঙ্ক বিশ্লেষণ সারণিতে সাজিয়ে লেখা যায়।

সারণি ৭.২ : সম্পূর্ণ দৈবায়িত নকশার ক্ষেত্রে ভেদাঙ্ক বিশ্লেষণ সারণি।

ভেদের উৎস	স্বাধীনতার মাত্রা d.f	বর্গসমষ্টি SS	গড় বর্গসমষ্টি MS = SS/d.f	F	$F_{.05}$
চর্যা	$P-1$	S_1	s_1	s_1/s_2	
বিচ্যুতি	$n-P$	S_2	s_2		
মোট	$n-1$				

এই বিশ্লেষণের ক্ষেত্রে অনুমান করা হয় যে, P চর্যার উপাত্ত P গণসমষ্টি হতে চয়ন করা হয়েছে। যেখানে i-তম $[i = 1, 2, \cdots, P]$ গণসমষ্টির গড় μ_i। সে কারণে বিশ্লেষণের মূল উদ্দেশ্য হলো নাস্তি কল্পনা

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_p = \mu$$

অথবা $$H_0 : \mu_1 - \mu = \mu_2 - \mu = \cdots\cdots = \mu_p - \mu = 0$$

অথবা $$H_0 : \alpha_1 = \alpha_2 = \cdots\cdots = \alpha_p = 0$$

চর্যা। এক্ষেত্রে পেস্টিসাইড ভিন্ন মৃত পোকামাকড় হবে নিয়ন্ত্রিত চর্যার ফলাফল। এই নিয়ন্ত্রিত চর্যার সাথে পরীক্ষণ চর্যার তুলনা করা হয় Dunnett যাচাই তথ্যজমান প্রয়োগ করে। এখানে Dunnett যাচাই তথ্যজমান হলো

$$D = d_{0.05;p-1,\, n-p}\sqrt{\frac{2s_2}{H.M(n_i)}}$$

অথবা

$$D = d_{.05\,;\,p-1,\, n-p}\sqrt{\frac{2s_2}{q}}$$

এখানে, $d_{0\cdot05;p-1,\,n-p}$ হলো 5% সংশয় মাত্রায় $(n-p)$ স্বাধীনতার মাত্রাবিশিষ্ট $(p-1)$ গড়ের জন্য Dunnett প্রস্তাবিত সারণিকৃত মান।

ধরা যাক আলোচিত p চর্যার মধ্যে প্রথম চর্যাটি নিয়ন্ত্রিত চর্যা এবং এর গড় হলো $\bar{y}_{1\cdot}$। তাহলে

$$|\bar{y}_{i\cdot} - \bar{y}_{1\cdot}| \geqslant D\ [i = 2, 3, \cdots, p]$$

হলে নিয়ন্ত্রিত চর্যার সাথে i-তম চর্যার পার্থক্য তাৎপর্যপূর্ণ বলে বিবেচিত হবে।

উদাহরণ ৭.১

Serranidae পরিবারভুক্ত ৬ প্রকার মৎস্য প্রজাতির (species) Centrum-এর দৈর্ঘ্যের মধ্যে কোনো তাৎপর্যপূর্ণ পার্থক্য আছে কিনা যাচাই করার জন্য ভূমধ্যসাগরীয় উপকূলবর্তী অঞ্চলের একটি বাজার হতে প্রতি প্রজাতির পাঁচটি মাছ সংগ্রহ করে ঐগুলির Centrum এর দৈর্ঘ্য নথিভুক্ত করা হয়েছে। নিচে প্রাপ্ত উপাত্ত দেওয়া হলো। এই উপাত্তের ভিত্তিতে বিভিন্ন প্রজাতির মধ্যে পার্থক্য আছে কিনা যাচাই করা যাক এবং শেষোক্ত

সারণি ৭.৩ : Serranidae পরিবারভুক্ত ছয় প্রজাতির মাছের Centrum এর দৈর্ঘ্য [mm] $[y_{ij}]$ মাছের প্রজাতি

Epine phelus aeneus (1)	Epine phelus caninus (2)	Epine phelus costae (3)	Epine phelus haifensis (4)	Epine phelus marginatus (5)	Myetero perea rubra (6)
4.42	10.11	8.69	4.65	4.67	13.52
4.36	10.10	8.72	4.51	4.66	13.39
4.59	10.04	8.59	4.40	4.55	12.90
4.63	10.32	8.49	4.46	4.50	13.20
4.74	10.64	8.62	4.58	4.50	13.06
মোট $y_{i\cdot}$ 22.74	51.21	43.11	22.60	22.88	66.07

প্রজাতির মাছের Centrum-এর দৈর্ঘ্য অন্যান্য প্রজাতির মাছের ঐ দৈর্ঘ্যের থেকে পার্থক্য কিনা তাও যাচাই করা যাক।

এখানে চর্যা সংখ্যা, $p = 6$; পুনরায়ন সংখ্যা, $q = 5$ মোট বর্গসমষ্টি, $TSS = \Sigma\Sigma y^2_{ij}$

$$- C.T, \quad C.T = \frac{G^2}{pq}, \; G = 228.61 \quad C.T = \frac{(228.61)^2}{6 \times 5} = 1742.0844$$

$$TSS = 2080.1951 - 1742.0844 = 338.1107$$ চর্যার বর্গসমষ্টি

[বর্গসমষ্টি (প্রজাতি)], $S_1 = \dfrac{\Sigma y_{i.}^2}{q} - C.T = \dfrac{10397.5431}{5} - 1742.0844$

$$= 337.4242$$

বিচ্যুতির বর্গসমষ্টি, $S_2 = TSS - S_1 = 338.1107 - 337.4242 = 0.6865$

ভেদাঙ্ক বিশ্লেষণ সারণি

ভেদের উৎস	স্বাধীনতার মাত্রা d.f	বর্গসমষ্টি SS	গড় বর্গসমষ্টি MS = SS/d.f	F	$F_{0.05}$
প্রজাতি	5	337.4242	67.4848	2359.61	2.62
বিচ্যুতি	24	0.6865	0.0286		
মোট	29				

এখানে নাস্তি কল্পনা হলো

H_0 : বিভিন্ন প্রজাতির মধ্যে কোনো পার্থক্য নেই।

নির্ণেয় $F = 2359.61 > F_{.05} = 2.62$ হওয়ায় [$p = 0.00$] বিভিন্ন প্রজাতির মাছের Centrum-এর দৈর্ঘ্য খুবই তাৎপর্যপূর্ণ পার্থক্য আছে। এই পার্থক্য পরিলক্ষিত হচ্ছে গড় দৈর্ঘ্যের মাধ্যমে। এখানে গড় দৈর্ঘ্যসমূহ হলো

মাছের প্রজাতি :	E_p. aeneus	E_p. caninus	E_p. costae
Centrum এর গড় দৈর্ঘ্য :	4.548	10.242	8.622
মাছের প্রজাতি :	E_p. haifensis	E_p. marginatus	Myet. rubra
Centrum এর গড় দৈর্ঘ্য :	4.52	4.576	13.214

যেহেতু গড় দৈর্ঘ্যের মধ্যে তাৎপর্যপূর্ণ পার্থক্য আছে, সে কারণে কোন কোন প্রজাতি দৈর্ঘ্যের ভিত্তিতে একইরূপ এবং কোনগুলি ভিন্ন তা DMRT-এর মাধ্যমে যাচাই করে দেখা যেতে পারে। এখানে যাচাই তথ্যজমান হলো

$$R_2 = r_{0.05;2,24}\sqrt{\frac{0.0286}{5}} = 2.93 \times 0.076 = 0.221$$

$$R_3 = r_{0.05;3,24}\sqrt{\frac{0.0286}{5}} = 3.076 \times 0.076 = 0.234$$

$$R_4 = r_{0.05;4,24}\sqrt{\frac{0.0286}{5}} = 3.156 \times 0.076 = 0.240$$

$$R_5 = r_{0.05;5,24}\sqrt{\frac{0.0286}{5}} = 3.23 \times 0.076 = 0.245$$

$$R_6 = r_{0.05;6,24}\sqrt{\frac{0.0286}{5}} = 3.28 \times 0.076 = 0.249$$

গড়গুলিকে মানের ক্রম অনুসারে সাজিয়ে লেখা যায়

$\bar{Y}_{1.} = 4.52$, $\bar{Y}_{2.} = 4.548$. $\bar{Y}_{3.} = 4.576$, $\bar{Y}_{4.} = 8.622$, $\bar{Y}_{5.} = 10.242$, $\bar{Y}_{6.} = 13.214$

$\bar{Y}_{6.} - \bar{Y}_{1.} = 8.694 > R_6$; গড়গুলির মধ্যে তাৎপর্যপূর্ণ পার্থক্য আছে।

$\bar{Y}_{5.} - \bar{Y}_{1.} = 5.722 > R_5$; $\bar{Y}_{5.}$ এবং $\bar{Y}_{1.}$-এর মধ্যে তাৎপর্যপূর্ণ পার্থক্য আছে।

$\bar{Y}_{6.} - \bar{Y}_{2.} = 8.666 > R_5$; $\bar{Y}_{6.}$ এবং $\bar{Y}_{2.}$-এর মধ্যে তাৎপর্যপূর্ণ পার্থক্য আছে।

$\bar{Y}_{4.} - \bar{Y}_{1.} = 4.102 > R_4$; $\bar{Y}_{4.}$ এবং $\bar{Y}_{1.}$-এর মধ্যে তাৎপর্যপূর্ণ পার্থক্য আছে

$\bar{Y}_{5.} - \bar{Y}_{2.} = 5.694 > R_4$; $\bar{Y}_{5.}$ এবং $\bar{Y}_{1.}$-এর মধ্যে তাৎপর্যপূর্ণ পার্থক্য আছে

$\bar{Y}_{6.} - \bar{Y}_{3.} = 8.638 > R_4$; $\bar{Y}_{6.}$ এবং $\bar{Y}_{3.}$-এর মধ্যে তাৎপর্যপূর্ণ পার্থক্য আছে

$\bar{Y}_{3.} - \bar{Y}_{1.} = 0.056 < R_3$; $\bar{Y}_{1.}$ থেকে $\bar{Y}_{3.}$ গড়গুলির মধ্যে কোনো তাৎপর্যপূর্ণ পার্থক্য নেই। এই তিনটি একই গুচ্ছভুক্ত।

$\bar{Y}_{4.} - \bar{Y}_{2.} = 4.074 > R_3$; $\bar{Y}_{4.}$ এবং $\bar{Y}_{2.}$ এর মধ্যে তাৎপর্যপূর্ণ পার্থক্য আছে

$\bar{Y}_{5.} - \bar{Y}_{3.} = 5.666 > R_3$; $\bar{Y}_{5.}$ এবং $\bar{Y}_{3.}$-এর মধ্যে তাৎপর্যপুর্ণ পার্থক্য আছে

$\bar{Y}_{6.} - \bar{Y}_{4.} = 4.592 > R_3$; $\bar{Y}_{6.}$ এবং $\bar{Y}_{4.}$-এর মধ্যে তাৎপর্যপূর্ণ পার্থক্য আছে

$$SS(\text{সময়}) = \frac{\Sigma y_{i.}^2}{q} - C.T = \frac{2.179697}{4} - 0.4485 = 0.0964$$

$$SS(\text{পেস্টিসাইড}) = \frac{\Sigma y_{.j}^2}{P} - C.T = \frac{2.319679}{5} - 0.4485 = 0.0154$$

$$SS(\text{বিচ্যুতি}) = TSS - SS\ (\text{সময়}) - SS\ (\text{পেস্টিসাইড})$$

$$= 0.008601$$

ভেদাঙ্ক বিশ্লেষণ সারণি

ভেদের উৎস	স্বাধীনতার মাত্রা d.f.	বর্গসমষ্টি SS	গড় বর্গসমষ্টি MS = SS/d.f.	F	$F_{.05}$	P-মান
সময়	4	0.0964	0.0241	33.61	3.26	0.00
পেস্টিসাইড	3	0.0154	0.00513	7.15	3.49	0.00
বিচ্যুতি	12	0.008601	0.000717			
মোট	19					

উপরিউক্ত বিশ্লেষণ হতে লক্ষণীয় যে, সময় অতিক্রান্ত হওয়ার সাথে সাথে Slug-এর অক্সিজেন গ্রহণের পরিমাণ তাৎপর্যপূর্ণভাবে বৃদ্ধি পায়। আবার বিভিন্ন পেস্টিসাইডে রাখার কারণে অক্সিজেন গ্রহণের পরিমাণে তাৎপর্যপূর্ণ পার্থক্য পরিলক্ষিত হচ্ছে। এই পার্থক্য Control এবং পেস্টিসাইডের ক্ষেত্রে কিরূপ তা পর্যালোচনা করা যাক।

Control চর্যা অন্যান্য পেস্টিসাইড হতে ভিন্ন কিনা তা Dunnett's যাচাই-এর মাধ্যমে লক্ষ্য করা যাক। এখানে

$$D = d_{0.05,3,12}\sqrt{\frac{2s_3}{P}}$$

$$= 2.68\sqrt{\frac{2\times 0.000717}{5}} = 0.0454$$

বিভিন্ন চর্যার গড়গুলি হলো

$\bar{C} = 0.1966,\ \bar{P} = 0.1238,\ \bar{G} = 0.1396\ ,\ \bar{M} = 0.1390$

$|\bar{C} - \bar{P}| = 0.0728 > D$; Control-এর পরিবর্তে Pomex-এ রাখার কারণে গড় অক্সিজেন গ্রহণের পরিমাণ তাৎপর্যপূর্ণভাবে কম।

$|\bar{C}-\bar{G}| = 0.057 > D$; Glyphosate-এ রাখার কারণে অক্সিজেন গ্রহণের পরিমাণ তাৎপর্যপূর্ণভাবে কম

$|\bar{C}-\bar{M}| = 0.0576 > D$; Milkrup-এ রাখার কারণে অক্সিজেন গ্রহণের পরিমাণ তাৎপর্যপূর্ণভাবে কম।

লক্ষণীয় বিষয় হলো কন্ট্রোল ভিন্ন পেস্টিসাইডে রাখার কারণে M. Rusticus Slug-এর অক্সিজেন গ্রহণের পরিমাণ তাৎপর্যপূর্ণভাবে কম। এখন তিন প্রকার পেস্টিসাইডের মধ্যে কোনো পার্থক্য আছে কিনা তা DMRT-এর মাধ্যমে লক্ষ্য করা যাক। এখানে

$$R_4 = r_{0.05,4,12}\sqrt{\frac{s_3}{P}} = 3.33\sqrt{\frac{0.000717}{5}} = 0.0399$$

$$R_3 = r_{0.05,3,12}\sqrt{\frac{s_3}{P}} = 3.23\sqrt{\frac{0.000717}{5}} = 0.0387$$

$$R_2 = r_{0.05,2,12}\sqrt{\frac{s_3}{P}} = 3.08\sqrt{\frac{0.000717}{5}} = 0.0369$$

গড়গুলিকে মানের ক্রম অনুসারে সাজিয়ে লেখা যায় নিম্নরূপ :

$$\bar{P} = 0.1238, \quad \bar{M} = 0.1390, \quad \bar{G} = 0.1396 \quad \bar{C} = 0.1966$$

$\bar{C} - \bar{P} = 0.0728 > R_4$, $\bar{C}$ এবং $\bar{P}$ তাৎপর্যপূর্ণভাবে পার্থক্য

$\bar{G} - \bar{P} = 0.0158 < R_3$, $\bar{P}$ থেকে $\bar{G}$ পর্যন্ত গড়গুলির মধ্যে কোনো তাৎপর্যপূর্ণ পার্থক্য নেই।

$\bar{C} - \bar{M} = 0.0576 > R_3$, $\bar{C}$ এবং $\bar{M}$ তাৎপর্যপূর্ণভাবে পার্থক্য

$\bar{C} - \bar{G} = 0.057 > R_2$, $\bar{C}$ এবং $\bar{G}$ তাৎপর্যপূর্ণভাবে পার্থক্য

উপরিউক্ত ফলাফলের ভিত্তিতে বলা যায় যে, অক্সিজেন গ্রহণের পরিমাণের ভিত্তিতে তিন প্রকার পেস্টিসাইড-এর প্রভাব একইরূপ এবং এগুলি Control-এর প্রভাব থেকে ভিন্নতর। যে গড়গুলির মধ্যে পার্থক্য নেই ঐগুলিকে নিচে দাগ দিয়ে দেখানো হলো।

$$\underline{\bar{P},\ \bar{M},\ \bar{G}},\ \bar{C}$$

নাস্তি কল্পনাসমূহ যাচাই করা। নাস্তি কল্পনা (৭.৬.২) এবং (৭.৬.৩)-এর জন্য যাচাই তথ্যজমান হলো, যথাক্রমে $F_1 = s_1/s_3$ এবং $F_2 = s_2/s_3$। নির্ণেয় F-এর মান সারণিকৃত F-এর মান অপেক্ষা বড় হলে নাস্তি কল্পনা বাতিল হবে।

নাস্তি কল্পনা (৭.৬.৩) বাতিল হলে

$$H_0 : \beta_j = \beta_{j'} \qquad (৭.৬.৪)$$
$$H_A : \beta_j \neq \beta_{j'} \quad (j \neq j' = 1, 2, \cdots, q)$$

যাচাই করতে হয়। এই যাচাই-এর উদ্দেশ্য হলো j-তম এবং j′-তম চর্যার মধ্যে কোনো তাৎপর্যপূর্ণ পার্থক্য আছে কিনা তা লক্ষ্য করা। j-তম এবং j′-তম নির্দিষ্ট হলে যাচাই তথ্যজমান হবে

$$t = \frac{\bar{y}_{\cdot j} - \bar{y}_{\cdot j'}}{\sqrt{\frac{2s_3}{p}}}$$

এবং $|t| \geq t_{0\cdot 025\,;\,(p-1)(q-1)}$ হলে 5% সংশয় মাত্রায় নাস্তি কল্পনা বাতিল হবে।

j-তম এবং j′-তম চর্যার দ্বারা যে কোনো জোড়া চর্যা বুঝালে এবং চর্যাগুলিকে সমমাত্রিক (homogeneous) গুচ্ছে বিভক্ত করতে হলে যাচাই তথ্যজমান হবে **DMRT**, যেখানে

$$R_k = r_{0\cdot 05\,,\,k,\,(p-1)(q-1)}\sqrt{\frac{s_3}{p}}\ ;\ k = 2, 3, \cdots, q$$

এই যাচাই তথ্যজমান হতে সিদ্ধান্ত গ্রহণ পদ্ধতি ৭.৫ অনুচ্ছেদে আলোচিত পদ্ধতির অনুরূপ হবে। এক্ষেত্রে R_k-কে k চর্যার গড়ের পরিসরের সাথে তুলনা করতে হয়। অর্থাৎ

$$\bar{Y}_{k\cdot} - \bar{Y}_{l\cdot} \geq R_k$$

হলে চর্যাদ্বয়ের মধ্যে পার্থক্য তাৎপর্যপূর্ণ হবে। এখানে $\bar{Y}_{k\cdot}$ এবং $\bar{Y}_{l\cdot}$ হলো যথাক্রমে চর্যার সর্বোচ্চ ও সর্বনিম্ন গড়।

j-তম চর্যা নিয়ন্ত্রিত চর্যা হলে এবং j′-তম ≠ j-তম [2, 3, ⋯, q] চর্যা অন্য যে কোনো চর্যা হলে যাচাই তথ্যজমান হবে **Dunnett's 'D' test**, যেখানে

$$D = d_{0\cdot 05\,;\,q-1,\,(p-1)(q-1)}\sqrt{\frac{2s_3}{p}}$$

এক্ষেত্রেও সিদ্ধান্ত নেয়ার জন্য পূর্ববর্তী অনুচ্ছেদে আলোচিত পন্থা অবলম্বন করতে হবে।

উদাহরণ ৭.২

একটি পরীক্ষাগারে M. Rusticus Slug-এর অক্সিজেন গ্রহণের পরিমাণে কি পরিবর্তন হয় তা লক্ষ্য করার জন্য কিছু Slug-কে চার প্রকার রাসায়নিক পেস্টিসাইডের মধ্যে রাখা হয় এবং সময় (ঘণ্টা) অতিক্রান্ত হওয়ার সাথে সাথে অক্সিজেন গ্রহণের পরিমাণের পরিবর্তন লক্ষ্য করা হয়। নিচে সময়ভিত্তিক এবং পেস্টিসাইডভিত্তিক অক্সিজেন গ্রহণের পরিমাণ দেয়া হলো। এই উপাত্তের ভিত্তিতে অক্সিজেন গ্রহণের

সারণি ৭.৪ : বিভিন্ন রাসায়নিক পেস্টিসাইডে রাখার পর M. Rusticus Slug কর্তৃক বিভিন্ন সময়ে গ্রহণ করা অক্সিজেনের পরিমাণ [y_{ij}]

সময় (ঘণ্টা)	পেস্টিসাইড (চর্যা)					
	Control	Pomex	Glyphosate	Milkrup	মোট $y_{i\cdot}$	$\bar{y}_{i\cdot}$
0	0.080	0.040	0.040	0.090	0.25	0.062
2	0.096	0.050	0.051	0.099	0.296	0.074
48	0.227	0.143	0.161	0.155	0.686	0.171
96	0.274	0.180	0.207	0.171	0.832	0.208
144	0.306	0.206	0.239	0.180	0.931	0.233
মোট $y_{\cdot j}$	0.983	0.619	0.698	0.695	2.995	
$\bar{y}_{\cdot j}$	0.1966	0.1238	0.1396	0.1390	0.14975	

পরিমাণের Control অন্যান্য পেস্টিসাইডের তুলনায় ভিন্ন প্রভাব বিস্তার করে কিনা তা পর্যালোচনা করা যাক। এছাড়া পেস্টিসাইডগুলিকে গড় অক্সিজেন গ্রহণের ভিত্তিতে গুচ্ছায়ন করা যাক।

এখানে $p[\text{ব্লক}] = 5,\ q[\text{চর্যা}] = 4,\ G = 2.995$

$$C.T = \frac{G^2}{pq} = \frac{(2.995)^2}{5 \times 4} = 0.4485$$

মোট বর্গসমষ্টি $TSS = \Sigma\Sigma y_{ij}^2 - C.T = 0.568901 - 0.4485 = 0.12040$

এই নকশা হতে প্রাপ্ত তথ্যমানে ভেদের চিহ্নিত উৎস হলো দুটি, যথা (১) ব্লক এবং (২) চর্যা। এক ব্লকের এককসমূহ অন্য ব্লকের এককসমূহ হতে ভিন্নতর বলে ব্লকের কারণেও উপাত্তে ভেদের সৃষ্টি হয়। সে কারণে উপাত্তের মোট ভেদাঙ্ককে বিচ্যুতির ভেদাঙ্ক ছাড়াও ব্লক ভেদাঙ্ক এবং চর্যার ভেদাঙ্কে বিভক্ত করা হয়। বিচ্যুতির ভেদাঙ্ক ছাড়া উপাত্তের মোট ভেদাঙ্ককে আরো দুটি চিহ্নিত উৎসের ভেদাঙ্কে বিভক্ত করা হয় বলে দৈবায়িত ব্লক নকশার উপাত্তের বিশ্লেষণকে দ্বি-মুখী শ্রেণী বিন্যাস (two way classification) বলা হয়। এই বিশ্লেষণের জন্য প্রতিকৃতি হলো

$$y_{ij} = \mu + \alpha_i + \beta_j + e_{ij} \tag{৭.৬.১}$$

$$i = 1, 2, \cdots, p\,; \;\; j = 1, 2, \cdots, q$$

এখানে, y_{ij} = i-তম ব্লকে j-তম চর্যার উৎপাদন। মনে করা হচ্ছে যে, q চর্যা প্রতিটিকে p ব্লকে পুনরায়ন করা হচ্ছে।

μ = সাধারণ গড়,

α_i = i-তম ব্লকের প্রভাব,

β_j = j-তম চর্যার প্রভাব,

e_{ij} = দৈব বিচ্যুতি।

অনুমান : (i) প্রতিকৃতি স্থির প্রভাববিশিষ্ট,

(ii) $e_{ij} \sim NID\,(0, \sigma^2)$

প্রতিকৃতি হতে নিরূপিত বিচ্যুতির বর্গসমষ্টি হলো

$$\varphi = \Sigma\Sigma\, e_{ij}^2 = \Sigma\Sigma(y_{ij} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j)^2$$

সুতরাং ন্যূনতম বর্গপদ্ধতি প্রয়োগ করে পরিমিত সমীকরণ হলো

$$\frac{\partial\varphi}{\partial\hat{\mu}} = 0\,, \;\; \frac{\partial\varphi}{\partial\hat{\alpha}_i} = 0 \;\; \text{এবং} \;\; \frac{\partial\varphi}{\partial\hat{\beta}_j} = 0$$

এই সমীকরণগুলিকে সমাধান করে $\Sigma\alpha_i = \Sigma\beta_j = 0$ শর্তাধীনে

$$\hat{\mu} = \bar{y}_{..}\,,\;\; \hat{\alpha}_i = \bar{y}_{i.} - \bar{y}_{..} \;\; \text{এবং} \;\; \hat{\beta}_j = \bar{y}_{.j} - \bar{y}_{..}$$

উপাত্তের মোট বর্গসমষ্টি হলো

$$TSS = \sum_i \sum_j (y_{ij} - \bar{y}_{..})^2$$

একে পৃথকীকরণ করা যায় নিম্নরূপে

$$\Sigma\Sigma(y_{ij} - \bar{y}_{..})^2 = \Sigma\Sigma[(\bar{y}_{i.} - \bar{y}_{..}) + (\bar{y}_{.j} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})]^2$$

সরল করে পাওয়া যায়

$$\Sigma\Sigma(y_{ij} - \bar{y}_{i.})^2 = q\sum_i(\bar{y}_{i.} - \bar{y}_{..})^2 + p\sum_j(\bar{y}_{.j} - \bar{y}_{..})^2 + \Sigma\Sigma(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2$$

$$= S_1 + S_2 + S_3$$

এখানে $S_1 = SS\ (\text{ব্লক}) = q\,\Sigma(\bar{y}_{i.} - \bar{y}_{..})^2 = \dfrac{\Sigma y_{i.}^2}{q} - C.T$

$S_2 = SS\ (\text{চর্যা}) = p\,\Sigma(y_{.j} - \bar{y}_{..})^2 = \dfrac{\Sigma y_{.j}^2}{p} - C.T$

$S_3 = SS\ (\text{বিচ্যুতি}) = \Sigma\Sigma(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} - \bar{y}_{..})^2$

$= TSS - SS\ (\text{ব্লক}) - SS\ (\text{চর্যা})$

এখানে $C.T = \dfrac{G^2}{pq}$; G = উপাত্তের সর্বমোট।

সারণি ৭.৩ ঃ দৈবায়িত ব্লক নকশার ক্ষেত্রে ভেদাঙ্ক বিশ্লেষণ সারণি।

ভেদের উৎস	স্বাধীনতার মাত্রা d.f	বর্গসমষ্টি SS	গড় বর্গসমষ্টি MS = SS/d.f.	F	$F_{.05}$
ব্লক	$p-1$	S_1	s_1	s_1/s_3	
চর্যা	$q-1$	S_2	s_2	s_2/s_3	
বিচ্যুতি	$(p-1)(q-1)$	S_3	s_3		
মোট	$pq-1$				

এই বিশ্লেষণের মূল উদ্দেশ্য হলো

$$H_0 : \alpha_i = 0 \qquad (৭.৬.২)$$
$$H_A : \alpha_i \neq 0$$

এবং

$$H_0 : \beta_j = 0 \qquad (৭.৬.৩)$$
$$H_A : \beta_j \neq 0$$

$\bar{Y}_{4.}-\bar{Y}_{3.}=4.046>R_2$; $\bar{Y}_{4.}$ এবং $\bar{Y}_{3.}$ এর মধ্যে তাৎপর্যপূর্ণ পার্থক্য আছে।

$\bar{Y}_{5.}-\bar{Y}_{4.}=1.62>R_2$; $\bar{Y}_{5.}$ এবং $\bar{Y}_{4.}$ এর মধ্যে তাৎপর্যপূর্ণ পার্থক্য আছে।

$\bar{Y}_{6.}-\bar{Y}_{5.}=2.972>R_2$; $\bar{Y}_{6.}$ এবং $\bar{Y}_{4.}$ এর মধ্যে তাৎপর্যপূর্ণ পার্থক্য আছে।

উপরিউক্ত ফলাফলের ভিত্তিতে নিচে দাগ দেয়া প্রজাতিগুলি একই শ্রেণীভুক্ত বিবেচিত হতে পারে (centrum এর দৈর্ঘ্যের ভিত্তিতে)।

E_p. haif. ; E_p. aen. ; E_p. marg. E_p. cost. ; E_p. Can. ; Myet. rub.

DMRT-এর মাধ্যমে লক্ষণীয় বিষয় হলো সারণি ৭.৩-এ দেয়া মাছের প্রথম, চতুর্থ এবং পঞ্চম প্রজাতি Centrum-এর দৈর্ঘ্যের ভিত্তিতে অনুরূপ। এখন এই তিনটি প্রজাতির প্রথম দুটির মধ্যে কোনো পার্থক্য আছে কিনা তা 't'-যাচাই তথ্যজমানের মাধ্যমে যাচাই করা যাক। এজন্য নাস্তি কল্পনা হলো

$$H_0: \mu_1=\mu_4$$

এবং

$$H_A: \mu_1\neq\mu_4$$

$$t=\frac{\left|\bar{y}_{1.}-\bar{y}_{4.}\right|}{\sqrt{\frac{2s_2}{q}}}=\frac{|4.548-4.52|}{\sqrt{\frac{2\times0.0286}{5}}}=0.26$$

এখানে $t<t_{0.025,24}=2.0639$ হওয়ায় নাস্তি কল্পনার বিপক্ষে কোনো যুক্তি এই নমুনা হতে পাওয়া যাচ্ছে না।

Myeteroperca rubra (6) প্রজাতির মাছকে অন্যান্য প্রজাতির মাছের সাথে দৈর্ঘ্যের ভিত্তিতে তুলনা করার জন্য Myet. rub.-কে নিয়ন্ত্রিত চর্যা বিবেচনা করে Dunnett's যাচাই পদ্ধতি প্রয়োগ করা যায়। এক্ষেত্রে যাচাই তথ্যজমান হলো

$$D=d_{.05,5,24}\sqrt{\frac{2s_2}{q}}=2.70\times\sqrt{\frac{2\times0.0286}{5}}=0.289$$

এখন

$\left|\bar{y}_{6.}-\bar{y}_{1.}\right|=8.666>D$; চর্যা-6 চর্যা-1 হতে তাৎপর্যপূর্ণভাবে ভিন্ন

$\left|\bar{y}_{6.}-\bar{y}_{2.}\right|=2.972>D$; চর্যা-6 চর্যা-2 হতে তাৎপর্যপূর্ণভাবে ভিন্ন

$\left|\bar{y}_{6.} - \bar{y}_{3.}\right| = 4.592 > D$; চর্যা-6 চর্যা-3 হতে তাৎপর্যপূর্ণভাবে ভিন্ন

$\left|\bar{y}_{6.} - \bar{y}_{4.}\right| = 8.694 > D$; চর্যা-6 চর্যা-4 হতে তাৎপর্যপূর্ণভাবে ভিন্ন

$\left|\bar{y}_{6.} - \bar{y}_{5.}\right| = 8.638 > D$; চর্যা-6 চর্যা-5 হতে তাৎপর্যপূর্ণভাবে ভিন্ন

অর্থাৎ Myeteroperca rubra প্রজাতির মাছের Centrum-এর দৈর্ঘ্য অন্যান্য প্রজাতির মাছের অনুরূপ দৈর্ঘ্য হতে তাৎপর্যপূর্ণভাবে ভিন্ন।

৭.৬ দৈবায়িত ব্লক নকশা (Randomized Block Design)

সম্পূর্ণ দৈবায়িত নকশার ক্ষেত্রে শর্ত হলো পরীক্ষণ এককসমূহ সমমাত্রিক (homogeneous) হতে হবে। বাস্তবে অনেক পরীক্ষার ক্ষেত্রেই ব্যবহৃত এককসমূহে একমুখী ভেদ (one-way variation) থাকতে পারে। যেমন, কোনো ডেইরী ফার্ম-এ বর্ধিত দুধ উৎপাদনের জন্য রাসায়নিক প্রক্রিয়াজাত বিভিন্ন গো-খাদ্য ব্যবহার করার পরীক্ষা পরিচালিত করা হলে লক্ষ্য করা যাবে যে বিভিন্ন গো-খাদ্য একই জাতীয় গরুকে খাওয়ানো সম্ভব হচ্ছে না। এমন হতে পারে যে গরুগুলির মধ্যে বয়সের পার্থক্য থাকতে পারে বা জাতের পার্থক্য থাকতে পারে। এদিকে জাতের পার্থক্যের কারণে বা বয়সের পার্থক্যের কারণে দুধ উৎপাদনে পার্থক্য হবে। সেক্ষেত্রে বিভিন্ন গো-খাদ্যের প্রভাব সঠিকভাবে জানতে হলে সম্পূর্ণ দৈবায়িত নকশা প্রয়োগ করে পরীক্ষা পরিচালনা করা যাবে না। এক্ষেত্রে গরুগুলিকে জাত অনুযায়ী বা বয়স অনুযায়ী (একই জাতের হলে) গুচ্ছে (group) ভাগ করে নিয়ে প্রতি গুচ্ছের গরুগুলিকে পরীক্ষায় ব্যবহৃত সকল প্রকার গো-খাদ্য খাওয়াতে হবে। এতে করে কোনো বিশেষ জাতের গরু বা বিশেষ বয়সের গরু বেশি বা কম দুধ দিতে পারে এমন প্রশ্ন উঠবে না এবং গো-খাদ্যের সঠিক প্রভাব জানা যাবে। এখানে গরুকে গুচ্ছে বিভক্ত করার কারণ হলো গরুগুলি এক জাতীয় নয়। গরুগুলির মধ্যে একমুখী ভেদ (জাত বা বয়স) আছে। দৈবায়িত ব্লক নকশার উদ্দেশ্য হলো এই একমুখী ভেদকে নিয়ন্ত্রণ করে চর্যার সঠিক প্রভাব নিরূপণ করা।

এই নকশা প্রয়োগের ক্ষেত্রে পরীক্ষণ এককসমূহকে ঐগুলির সমমাত্রিকতার ভিত্তিতে গুচ্ছে বিভক্ত করা হয় যেন প্রতি গুচ্ছের এককসমূহ সমমাত্রিক হয় এবং বিভিন্ন গুচ্ছ অসমমাত্রিক হয়। এখানে একটি সমমাত্রিক এককের গুচ্ছকে ব্লক বলা হয়। ব্লকের এককসমূহকে চর্যার সংখ্যার গুণিতকে (multiple) বিভক্ত করে প্রতি ভাগে এক একটি চর্যা দৈব পদ্ধতিতে (randomly) বণ্টন করা হলে যে নকশার উদ্ভব হয় তাকে দৈবায়িত ব্লক নকশা বলা হয়।

পরীক্ষণের শুরুতে এবং পরীক্ষণের শেষে Slug কর্তৃক অক্সিজেন গ্রহণের পরিমাণে কোনো পার্থক্য আছে কিনা তা যাচাই করে দেখা যেতে পারে। এই যাচাই-এর জন্য নাস্তি কল্পনা হলো

$$H_0: \alpha_1 = \alpha_5 \qquad (৭.৬.৫)$$

$$H_A: \alpha_1 \neq \alpha_5$$

এবং যাচাই তথ্যজমান হলো

$$t = \frac{\bar{y}_{1.} - \bar{y}_{5.}}{\sqrt{\frac{2s_3}{q}}} = \frac{0.062 - 0.233}{\sqrt{\frac{2 \times 0.000717}{4}}} = -9.03$$

$|t| > t_{0.025, 12} = 2.179$ হওয়ায় 5% সংশয় মাত্রায় নাস্তি কল্পনা বাতিল বলে গণ্য হলো। অর্থাৎ দীর্ঘ সময় অতিক্রান্ত হওয়াতে Slug কর্তৃক অক্সিজেন গ্রহণের পরিমাণ তাৎপর্যপূর্ণভাবে বৃদ্ধি পেয়েছে।

দৈবায়িত ব্লক নকশার ক্ষেত্রে একটি তথ্যমান লুপ্ত হলে বিশ্লেষণ পদ্ধতি : ধরা যাক দৈবায়িত ব্লক নকশার ক্ষেত্রে i-তম ব্লকের j-তম চর্যার উৎপাদন নষ্ট হয়ে গেছে বা তা লুপ্ত হয়েছে। এই উৎপাদনটির মান মনে করা যাক x। তাহলে মোট উপাত্তকে সাজিয়ে লেখা যায় নিম্নরূপ :

ব্লক	চর্যা						
	1	2	…	j	…	q	মোট
1	y_{11}	y_{12}	…	y_{1j}	…	y_{1q}	$y_{1.} = B_1$
2	y_{21}	y_{22}	…	y_{2j}	…	y_{2q}	$y_{2.} = B_2$
⋮	…	…	…	…	…	…	…
i	y_{i1}	y_{i2}	…	x	…	y_{iq}	$y_{i.} + x = B_i + x$
⋮	…	…	…	…	…	…	…
P	y_{p1}	y_{p2}	…	y_{pj}	…	y_{pq}	$y_{p.} = B_p$
মোট	$y_{.1}$	$y_{.2}$	…	$y_{.j} + x$		$y_{.q}$	$G + x$
	T_1	T_2	…	$T_j + x$		T_q	

এখানে
$$x = \frac{pB_i + qT_j - G}{(p-1)(q-1)}$$

এই উপাত্তের বিশ্লেষণ করার সময় লুপ্ত মানের স্থলে x-এর নিরূপিত মান বসিয়ে স্বাভাবিক বিশ্লেষণ করতে হয়। তবে, একটি মান লুপ্ত হওয়ার কারণে মোট স্বাধীনতার মাত্রা হতে এবং সে কারণে বিচ্যুতির স্বাধীনতার মাত্রা হতে 1 বিয়োগ করতে হয়। এখানে

$$C.T = \frac{(G+x)^2}{pq}, \text{ মোট বর্গসমষ্টি, } TSS = \Sigma\Sigma y_{ij}^2 + x^2 - C.T$$

$$SS\ (\text{ব্লক}) = \frac{\Sigma B_i'^2}{q} + \frac{(B_i + x)^2}{q} - C.T. = S_1$$

$$SS\ (\text{চর্যা}) = \frac{\Sigma T_j'^2}{p} + \frac{(T_j + x)^2}{p} - C.T. = S_2$$

$$SS\ (\text{বিচ্যুতি}) = TSS - SS\ (\text{ব্লক}) - SS\ (\text{চর্যা}) = S_3$$

সারণি ৭.৫ : ভেদাঙ্ক বিশ্লেষণ সারণি

ভেদের উৎস	স্বাধীনতার মাত্রা, d.f.	বর্গসমষ্টি SS	গড় বর্গসমষ্টি MS = SS/d.f.	F	$F_{.05}$
ব্লক	$p-1$	S_1	s_1	s_1/s_3	
চর্যা	$q-1$	S_2	s_2	s_2/s_3	
বিচ্যুতি	$(p-1)(q-1)-1$	S_3	s_3		
মোট	$Pq-2$				

এই বিশ্লেষণের ক্ষেত্রে চর্যার প্রভাব F-যাচাই এর মাধ্যমে তাৎপর্যপূর্ণ হলে নাস্তি কল্পনা

$$H_0: \beta_j = \beta_{j'} \qquad (j \neq j' = 1, 2, \cdots, q)$$
$$H_A: \beta_j \neq \beta_{j'}$$

যাচাই করার জন্য যাচাই তথ্যজমান হবে

$$t = \frac{\bar{y}_{.j} - \bar{y}_{.j'}}{\sqrt{s_3\left[\frac{2}{p} + \frac{q}{p(p-1)(q-1)}\right]}}$$

এখানে j-তম চর্যা হলো লুপ্তমানবিশিষ্ট চর্যা এবং j´-তম চর্যা হলো অন্য যে কোনো একটি চর্যা। এই যাচাই-এর ক্ষেত্রে $|t| \geqslant t_{.025,\,(p-1)(q-1)-1}$ হলে নাস্তি কল্পনা বাতিল হবে।

উদাহরণ ৭.৩

একটি কৃষি খামারে রজনীগন্ধা ফুলের চাষ করতে গিয়ে জমিতে পটাশ (K) এবং ইউরিয়া (N) সার প্রয়োগ করা হয়েছে। ইউরিয়া সারের স্তরসমূহ হলো

$N_1 = 30$ kg/ha, $N_2 = 60$ kg/ha, $N_3 = 90$ kg/ha, $N_4 = 120$ kg/ha এবং পটাশ সারের স্তরসমূহ হলো $K_1 = 30$ kg/ha, $K_2 = 45$ kg/ha, $K_3 = 60$kg/ha। পটাশ সারের যে কোনো স্তর প্রয়োগ করা চারটি খণ্ডে দৈবায়িত পদ্ধতিতে ইউরিয়া সারের স্তরগুলি প্রয়োগ করে নিরীক্ষা কাজটি পরিচালনা করা হয়েছে। নিচে সারণি ৭.৬-এ প্রতি খণ্ডে উৎপাদিত রজনীগন্ধা ফুলের সংখ্যা দেয়া হলো। কিন্তু পটাশের দ্বিতীয় স্তরে N_3 দেয়া খণ্ডের উৎপাদন নষ্ট হয়ে গেছে। উক্ত উপাত্তের বিশ্লেষণ করা যাক।

সারণি ৭.৬ : প্রতি স্তরের পটাশ সার প্রয়োগ করা জমিতে দৈবায়িত পদ্ধতিতে বণ্টন করা ইউরিয়া সারের বিভিন্ন স্তর প্রয়োগে জমি খণ্ডে উৎপাদিত রজনীগন্ধা ফুলের সংখ্যা $[y_{ij}]$।

পটাশ সার (K)		ইউরিয়া সার (N)				
		N_1	N_2	N_3	N_4	মোট $y_{i.}$
	K_1	50	55	54	65	$224 = B_1$
	K_2	54	60	x	72	$186 + x = B_2 + x$
	K_3	57	64	68	75	$264 = B_3$
মোট	$y_{.j}$	161	179	$122 + x$	212	$674 + x = G + x$
		T_1	T_2	$T_3 + x$	T_4	

এখানে
$$x = \frac{p\,B_i + q\,T_j - G}{(p-1)(q-1)}$$

$$B_i = B_2 = 186,\ T_j = T_3 = 122,\ G = 674,\ p = 3,\ q = 4$$

$$\therefore \quad x = \frac{3 \times 186 + 4 \times 122 - 674}{(3-1)(4-1)} = 62$$

তাহলে $$C.T = \frac{(G+x)^2}{pq} = \frac{(674+60)^2}{3\times 4} = 45141.33$$

মোট বর্গসমষ্টি $$TSS = \Sigma\Sigma y_{ij}^2 + x^2 - C.T$$

$$= 41960 + 3844 - 45141.33$$

$$= 662.67$$

$$SS\,(\text{পটাশ}) = \frac{B_1^2 + B_3^2}{4} + \frac{(B_2+x)^2}{4} - C.T$$

$$= \frac{119872}{4} + \frac{61504}{4} - 45141.33$$

$$= 202.67$$

$$SS\,(\text{ইউরিয়া}) = \frac{T_1^2 + T_2^2 + T_4^2}{3} + \frac{(T_3+x)^2}{3} - C.T$$

$$= \frac{102906}{3} + \frac{33856}{3} - 45141.33$$

$$= 446.00$$

$$SS\,(\text{বিচ্যুতি}) = TSS - SS\,(\text{পটাশ}) - SS\,(\text{ইউরিয়া})$$

$$= 662.67 - 202.67 - 446.00$$

$$= 14.00$$

ভেদাঙ্ক বিশ্লেষণ সারণি

ভেদের উৎস	স্বাধীনতার মাত্রা d.f.	বর্গসমষ্টি SS	গড় বর্গসমষ্টি $MS = \frac{SS}{d.f}$	F	$F_{.05}$	p-মান
পটাশ	2	202.67	101.335	36.19	5.79	0.00
ইউরিয়া	3	446.00	148.667	53.09	5.41	0.00
বিচ্যুতি	5	14.00	2.8			
মোট	10					

লক্ষ্য করা যাচ্ছে যে, ইউরিয়া সারের বিভিন্ন স্তর প্রয়োগ করার ফলে রজনীগন্ধা ফুলের উৎপাদনে তাৎপর্যপূর্ণ পার্থক্য হয়েছে। একই সিদ্ধান্ত পটাশ সার প্রয়োগের ক্ষেত্রে নেয়া যায়। এখন লক্ষ্য করা যাক N_2 ও N_3-এর উৎপাদনে পার্থক্য আছে কিনা। এর জন্য যাচাই তথ্যজমান হলো

$$t = \frac{\bar{N}_2 - \bar{N}_3}{\sqrt{s_3\left[\frac{2}{p} + \frac{q}{p(p-1)(q-1)}\right]}}$$

এখানে $\bar{N}_2 = 59.67$, $\bar{N}_3 = 61.33$, s_3 = গড় বর্গসমষ্টি (বিচ্যুতি) $= 2.8$

$$\therefore \qquad t = \frac{59.67 - 61.33}{\sqrt{2.8\left[\frac{2}{3} + \frac{4}{3(3-1)(4-1)}\right]}} = -1.05$$

এখানে $|t| < t_{0\cdot025, 5} = 2.571$ হওয়ায় N_2 এবং N_3-এর মধ্যে কোনো তাৎপর্যপূর্ণ পার্থক্য আছে বলা যায় না।

৭.৭ প্রতি কোষে একাধিক (সমান) তথ্যমানবিশিষ্ট দ্বি-মুখী শ্রেণীবিন্যাস (Two way Classification with Several (Equal) Observations Per Cell)

দৈবায়িত ব্লক নকশার ক্ষেত্রে লক্ষ্য করা গেছে যে, উপাত্তের মোট ভেদের চিহ্নিত উৎস দুটি—(১) ব্লক এবং (২) চর্যা। সে কারণে দৈবায়িত ব্লক নকশা হতে প্রাপ্ত উপাত্তের বিশ্লেষণকে দ্বি-মুখী শ্রেণীবিন্যাস (two-way classification) বলা হয়। এক্ষেত্রে ব্লকের প্রতি খণ্ডকে কয়েকটি উপ-খণ্ডে (sub-plot) বিভক্ত করে, ধরা যাক r উপ-খণ্ড, যে কোনো চর্যাকে একই ব্লকে r বার পুনরায়ন করা যায়। এখানে প্রতি খণ্ডের r উপ-খণ্ডকে r কোষ বিবেচনা করা যেতে পারে। এরূপ প্রতি খণ্ডে r কোষ-বিশিষ্ট কয়েকটি খণ্ড নিয়ে গঠিত হয় একটি ব্লক এবং এ জাতীয় নকশা হতে প্রাপ্ত উপাত্তের বিশ্লেষণকে বলা হয় প্রতি কোষে একাধিক তথ্যমানবিশিষ্ট দ্বি-মুখী শ্রেণী বিন্যাস। প্রতি খণ্ডে কোষের সংখ্যা সমান হতে পারে, আবার অসমানও হতে পারে। কোষের সংখ্যা অসমান হলে বিশ্লেষণে জটিলতা সৃষ্টি হয়। এ বই-এ সেই জটিল বিশ্লেষণ নিয়ে আলোচনা করা হবে না।

দৈবায়িত ব্লক নকশায় নিরীক্ষা একটি দুই উপাদানী নিরীক্ষা। মনে করা যাক A এবং B হলো দুটি উপাদান, যেখানে A ও B-এর স্তর হলো যথাক্রমে p এবং q। ধরা যাক A ও B-এর যে কোনো যুগ্মস্তরের জন্য r তথ্যমান আছে এবং B-এর j-তম স্তরের $[j=1, 2, \ldots, q]$ প্রাসঙ্গিক A-এর i-তম স্তরের $[i=1, 2, \cdots p,]$ l-তম তথ্যমান $[l=1, 2, \ldots, r]$ হলো y_{ijl}। এই y_{ijl} তথ্যমানের জন্য প্রতিকৃতি হলো

$$y_{ijl}=\mu+\alpha_i+\beta_j+(\alpha\beta)_{ij}+e_{ijl} \qquad (৭.৭.১)$$

এখানে μ = সাধারণ গড়, α_i = A-এর i-তম স্তরের প্রভাব, β_j = B-এর j-তম স্তরের প্রভাব, $(\alpha\beta)_{ij}$ = B এর j-তম স্তরের সাথে A-এর i-তম স্তরের মিশ্র প্রভাব, e_{ijl} = দৈব বিচ্যুতি।

এই বিশ্লেষণের জন্য অনুমান হলো

(১) প্রতিকৃতি ৭.৭.১ স্থির প্রভাব বিশিষ্ট,

(২) $e_{ijl} \sim \text{NID}\ (0, \sigma^2)$

উপরিউক্ত অনুমানের ভিত্তিতে প্রতিকৃতির পরামানসমূহ ন্যূনতম বর্গপদ্ধতি প্রয়োগ করে নিরূপণ করা যায়। এখানে নিরূপিত বিচ্যুতির বর্গসমষ্টিকে লেখা যায়

$$\varphi=\sum_i \sum_j \sum_l [\, y_{ijl}-\hat{\mu}-\hat{\alpha}_i-\hat{\beta}_j-(\hat{\alpha\beta})_{ij}]^2 \qquad (৭.৭.২)$$

সুতরাং $\hat{\mu}$, $\hat{\alpha}_i$, $\hat{\beta}_j$ এবং $(\hat{\alpha\beta})_{ij}$ এর মান পাওয়ার জন্য পরিমিত সমীকরণগুলি হলো যথাক্রমে

$$\frac{\partial\varphi}{\partial\hat{\mu}}=0\,,\ \ \frac{\partial\varphi}{\partial\hat{\alpha}_i}=0\,,\ \ \frac{\partial\varphi}{\partial\hat{\beta}_j}=0 \ \ \text{এবং}\ \ \frac{\partial\varphi}{\partial(\hat{\alpha\beta})_{ij}}=0$$

এই সমীকরণগুলি সমাধান করে

$$\Sigma\,\hat{\alpha}_i=\Sigma\,\hat{\beta}_j=\sum_i(\alpha\hat{\beta})_{ij}=\sum_j(\alpha\hat{\beta})_{ij}=0$$

শর্তাধীনে পাওয়া যায়

$$\hat{\mu}=\bar{y}_{\ldots},\ \ \hat{\alpha}_i=\bar{y}_{i\ldots}-\bar{y}_{\ldots},\ \ \hat{\beta}_j=\bar{y}_{.j.}-\bar{y}_{\ldots}$$

এবং $$(\hat{\alpha\beta})_{ij}=\bar{y}_{ij.}-\bar{y}_{i..}-\bar{y}_{.j.}+\bar{y}_{\ldots}$$

উপাত্তের মোট বর্গসমষ্টিকে নিম্নরূপভাবে পৃথকীকরণ করা যায়

$$\sum_i \sum_j \sum_l (y_{ijl} - \bar{y}_{...})^2 = \Sigma\Sigma\Sigma[(\bar{y}_{i..} - \bar{y}_{...}) + (\bar{y}_{.j.} - \bar{y}_{...})$$

$$+ (\bar{y}_{ij.} - \bar{y}_{.j.} - \bar{y}_{i..} + \bar{y}_{...}) + (y_{ijl} - \bar{y}_{ij.}]^2$$

সরল করে পাওয়া যায়

$$\Sigma\Sigma\Sigma(y_{ijl} - \bar{y}_{...})^2 = qr \sum_i (\bar{y}_{i..} - \bar{y}_{...})^2 + Pr \sum_j (\bar{y}_{.j.} - \bar{y}_{...})^2$$

$$+ \sum_i \sum_j (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} - \bar{y}_{...})^2 + \Sigma\Sigma\Sigma (y_{ijl} - \bar{y}_{ij.})^2$$

অথবা $\Sigma\Sigma\Sigma(y_{ijl} - \bar{y}_{...})^2 = SS(A) + SS(B) + SS(AB) + SS(error)$

$$= S_1 + S_2 + S_3 + S_4 \qquad (৭.৭.৩)$$

সারণি ৭.৭ : ভেদাঙ্ক বিশ্লেষণ সারণি।

ভেদের উৎস	স্বাধীনতার মাত্রা d.f.	বর্গসমষ্টি SS	গড় বর্গসমষ্টি MS = SS/d.f.	F	$F_{0.05}$
A	$p-1$	S_1	s_1	s_1/s_4	
B	$q-1$	S_2	s_2	s_2/s_4	
AB	$(p-1)(q-1)$	S_3	s_3	s_3/s_4	
বিচ্যুতি	$pq(r-1)$	S_4	s_4		
মোট	$Pqr-1$				

এই বিশ্লেষণের মুখ্য উদ্দেশ্য হলো নাস্তি কল্পনা

$$H_0 : \alpha_i = 0 \qquad (৭.৭.৪)$$

বিকল্প কল্পনা $H_A : \alpha_i \neq 0$

$$H_0 : \beta_j = 0 \qquad (৭.৭.৫)$$

বিকল্প কল্পনা $H_A : \beta_j \neq 0$

$$H_0 : (\alpha\beta)_{ij} = 0 \qquad (৭.৭.৬)$$

বিকল্প কল্পনা $H_A : (\alpha\beta)_{ij} \neq 0$

যাচাই করা। এখানে নাস্তি কল্পনা (৭.৭.৪), (৭.৭.৫) এবং (৭.৭.৬)-এর জন্য যাচাই তথ্যজমান হলো। যথাক্রমে

$$F = s_1/s_4,\ F = s_2/s_4 \text{ এবং } F = s_3/s_4.$$

নির্ণেয় F সমূহ ঐগুলির স্বস্ব স্বাধীনতার মাত্রাবিশিষ্ট $F_{0.05}$-এর সমান বা বড় হলে প্রাসঙ্গিক নাস্তি কল্পনা বাতিল বলে গণ্য হবে।

এই বিশ্লেষণের ক্ষেত্রে A এর যে কোনো দুটি স্তরের সমতা যাচাই করার জন্য নাস্তি কল্পনা হলো

$$H_0 : \alpha_i = \alpha_{i'} \ (i \neq i' = 1, 2, \cdots, p) \qquad (৭.৭.৭)$$
$$H_A : \alpha_i \neq \alpha_{i'}$$

এবং যাচাই তথ্যজমান হলো

$$t = \frac{\bar{y}_{i\cdot} - \bar{y}_{i'\cdot}}{\sqrt{\frac{2s_4}{q}}}$$

A-এর স্তরগুলিকে জোড়ায় জোড়ায় তুলনা করার জন্য DMRT যাচাই তথ্যজমান হলো।

$$R_k = r_{0.05, k, pq(r-1)} \sqrt{\frac{s_4}{q}},\ k = 2, 3, \ldots, P$$

B এর যে কোনো দুটি স্তরের সমতা যাচাই করার জন্য নাস্তি কল্পনা হলো

$$H_0 : \beta_j = \beta_{j'} ;\ i \neq j' = 1, 2, \cdots, q \qquad (৭.৭.৮)$$

বিকল্প কল্পনা $\quad H_A : \beta_j \neq \beta_{j'}$

এবং যাচাই তথ্যজমান হলো

$$t = \frac{\bar{y}_{\cdot j} - \bar{y}_{\cdot j'}}{\sqrt{\frac{2s_4}{p}}}$$

B এর স্তরগুলোকে জোড়ায় জোড়ায় তুলনা করার জন্য DMRT যাচাই তথ্যজমান হলো।

$$R_h = r_{0.05, h, pq(r-1)} \sqrt{\frac{s_4}{P}};\ h = 2, 3, \cdots, q$$

উপরিউক্ত যাচাই তথ্যজমানগুলো থেকে সিদ্ধান্ত ৭.৬ অনুচ্ছেদে আলোচিত উপায়ে নিতে হবে।

উদাহরণ ৭.৪

সারণি ১.৩ এ দেয়া উপাত্তের ক্ষেত্রে Pesticide প্রয়োগে Slug-এর Heart beat কিভাবে প্রভাবিত হয়েছে তা পর্যবেক্ষণ করার জন্য দ্বিমুখী শ্রেণীবিন্যাসের পদ্ধতি অনুসরণ করে বিশ্লেষণ করা যাক। নিয়ন্ত্রিত চর্যার (control) তুলনায় Pesticide গুলি Heart beat এর উপরে কিরূপ প্রভাব বিস্তার করেছে তাও লক্ষ্য করা যাক।

উপরিউক্ত সারণির ক্ষেত্রে Slug একটি উপাদান এবং Pesticide একটি উপাদান বিবেচিত হতে পারে। সেক্ষেত্রে প্রতিটি উপাদান স্তরের জন্য চারটি Slug এর Heart beat [y_{ijl} ; $i=1, 2,$, $j=1, 2, 3, 4, 5, 6$, $l=1, 2, 3, 4$] দেয়া আছে। নিচে চারটি Slug এর মোট Heart beat সারণি ৭.৮-এ উপস্থাপন করা হলো।

সারণি ৭.৮ : বিভিন্ন Pesticide এর বিপরীতে দুই প্রকার Slug-এর মোট Heart beat [$y_{ij\cdot}$]

Slug	Pesticides						
	Control	Supra-cide	Pomex	Sumi-cidin	Glypho-sate	Milkrup	মোট $y_{i\cdot}$
M. Rusticus	51	34	37	36	30	36	224
M. Sowerbyi	52	33	43	39	47	51	265
মোট $y_{\cdot j\cdot}$	103	67	80	75	77	87	489
	12.875	8.375	10.00	9.375	9.625	10.875	

এখানে $p=2,\quad q=6,\quad r=4,\quad G=489$

$$C.T=\frac{G^2}{pqr}=\frac{(489)^2}{2\times 6\times 4}=4981.6875$$

মোট বর্গসমষ্টি, $TSS=\Sigma\Sigma\Sigma\, y^2_{ijl}-C.T$

$$=5165-4981.6875=183.3125$$

$$SS\ (Slug)=\frac{\Sigma y_{i\cdot\cdot}^2}{q}-C.T=\frac{120401}{6\times 4}-4981.6875=35.0208$$

$$SS\ (Pesticide)=\frac{\Sigma y^2_{\cdot j\cdot}}{pr}-C.T=\frac{40621}{2\times 4}-4981.6875=95.9375.$$

$$SS(\text{Slug} \times \text{Pesticide}) = \frac{\Sigma\Sigma y_{ij\cdot}^2}{r} - C.T - SS(\text{Slug}) - SS(\text{Pesticide})$$

$$= \frac{20591}{4} - 4981.6875 - 35.0208 - 95.9395$$

$$= 35.1042$$

$$SS\ (\text{Error}) = TSS - SS\ (\text{Slug}) - SS\ (\text{Pesticide}) - SS\ (\text{Slug} \times \text{Pesticide})$$
$$= 183.3125 - 35.0208 - 95.9375 - 35.1042 = 17.25$$

ভেদাঙ্ক বিশ্লেষণ সারণি

ভেদের উৎস	স্বাধীনতার মাত্রা, d.f.	বর্গসমষ্টি SS	গড় বর্গসমষ্টি MS = SS/d.f.	F	$F_{.05}$	P-মান
Slug	1	35.0208	35.0208	73.08	4.12	0.00
Pesticide	5	95.9375	19.1875	40.04	2.46	0.00
Slug × Pesticide	5	35.1042	7.0208	14.65	2.46	0.00
মোট বিচ্যুতি	36	17.25	0.4792			
মোট	47					

ভেদাঙ্ক বিশ্লেষণ সারণি হতে লক্ষ্য করা যাচ্ছে যে, Slug-এর প্রাসঙ্গিক $F = 73.08 > F_{.05} = 4.2$ বা $P < 0.05$। সে কারণে Heart beat-এর প্রসঙ্গে M. Rusticus এবং M. Sowerbyi তাৎপর্যপূর্ণভাবে ভিন্ন। এখানে দুই প্রকার Slug হওয়ায় জোড়ায় জোড়ায় Slug-এর গড় Heart beat-এর পার্থক্য পর্যবেক্ষণ করার জন্য পুনরায় t যাচাই অথবা DMRT যাচাই তথ্যজমান নির্ণয় করার প্রয়োজন হবে না।

আবার Pesticide এর ভিন্নতা যাচাই করার জন্য নির্ণেয় $F = 40.04 > F_{.05} = 2.46$ অথবা $P < 0.05$ হওয়ায় বলা যায় যে, ভিন্ন ভিন্ন Pesticide Slug-এর উপর ভিন্ন ভিন্ন প্রভাব ফেলেছে। এখানে Pesticide গুলির জোড়ায় জোড়ায় পার্থক্য পর্যবেক্ষণ করার জন্য DMRT যাচাই তথ্যজমান নির্ণয় করা যেতে পারে।

Pesticide গুলির গড়কে মানের ক্রম অনুসারে সাজিয়ে পাওয়া যায়

Supracide	Sumicidin	Glyphosate	Pomex	Milkrup
$\bar{y}_{\cdot 2\cdot} = 8.375$	$\bar{y}_{\cdot 4\cdot} = 9.375$	$\bar{y}_{\cdot 5\cdot} = 9.625$	$\bar{y}_{\cdot 3\cdot} = 10.00$	$\bar{y}_{\cdot 6\cdot} = 10.875$

Control

$\bar{y}_{\cdot 1\cdot} = 12.875$

এখন DMRT $(R_k,\ k = 2,3,4,5,6)$ হলো

$$R_2 = r_{0\cdot05,2,36}\sqrt{\frac{s_4}{pr}} = 2.872\sqrt{\frac{0.4792}{2\times4}} = 0.703$$

$$R_3 = r_{0\cdot05,3,36}\sqrt{\frac{s_4}{pr}} = 3.022\sqrt{\frac{0.4792}{2\times4}} = 0.740$$

$$R_4 = r_{0\cdot05,4,36}\sqrt{\frac{s_4}{pr}} = 3.102\sqrt{\frac{0.4792}{2\times4}} = 0.759$$

$$R_5 = r_{0\cdot05,5,36}\sqrt{\frac{s_4}{pr}} = 3.182\sqrt{\frac{0.4792}{2\times4}} = 0.779$$

$$R_6 = r_{0\cdot05,6,36}\sqrt{\frac{s_4}{pr}} = 3.232\sqrt{\frac{0.4792}{2\times4}} = 0.791$$

$\bar{y}_{\cdot1\cdot} - \bar{y}_{\cdot2\cdot} = 4.5 > R_6$, $\bar{y}_{\cdot1\cdot} - \bar{y}_{\cdot4\cdot} = 3.50 > R_5$, $\bar{y}_{\cdot5\cdot} - \bar{y}_{\cdot1\cdot} = 2.50 > R_5$

$\bar{y}_{\cdot1\cdot} - \bar{y}_{\cdot5\cdot} = 3.25 > R_4$, $\bar{y}_{\cdot6\cdot} - \bar{y}_{\cdot4\cdot} = 1.50 > R_4$, $\bar{y}_{\cdot3\cdot} - \bar{y}_{\cdot1\cdot} = 1.625 > R$

$\bar{y}_{\cdot1\cdot} - \bar{y}_{\cdot3\cdot} = 2.875 > R_3$, $\bar{y}_{\cdot6\cdot} - \bar{y}_{\cdot5\cdot} = 1.25 > R_3$, $\bar{y}_{\cdot3\cdot} - \bar{y}_{\cdot4\cdot} = 0.625 < R_3$

সুতরাং $\bar{y}_{\cdot4\cdot}$, $\bar{y}_{\cdot5\cdot}$ এবং $\bar{y}_{\cdot3\cdot}$-এর মধ্যে কোনো পার্থক্য নেই।

$\bar{y}_{\cdot5\cdot} - \bar{y}_{\cdot2\cdot} = 1.25 > R_3$, $\bar{y}_{\cdot1\cdot} - \bar{y}_{\cdot6\cdot} = 2.00 > R_2$, $\bar{y}_{\cdot6\cdot} - \bar{y}_{\cdot3\cdot} = 0.875 > R_2$

$\bar{y}_{\cdot4\cdot} - \bar{y}_{\cdot2\cdot} = 1.00 > R_2$

উপরিউক্ত ফলাফল থেকে বলা যায় যে, **Sumicidin, Glyphosate** এবং **Pomex Slug**-এর **Heart beat**-এর উপর একইরূপ প্রভাব ফেলেছে।

Control চর্যার সাথে অন্যান্য **Pesticide**-এর পার্থক্য পর্যবেক্ষণ করার জন্য **Dunnett's** যাচাই প্রযোজ্য হলো

$$D = d_{0\cdot05,5,36}\sqrt{\frac{2s_4}{pr}} = 2.636\sqrt{\frac{2\times0.4792}{2\times4}} = 0.912$$

$$|\bar{y}_{\cdot1\cdot} - \bar{y}_{\cdot2\cdot}| = 4.5 > D, \quad |\bar{y}_{\cdot1\cdot} - \bar{y}_{\cdot3\cdot}| = 2.875 > D$$

$$|\bar{y}_{\cdot1\cdot} - \bar{y}_{\cdot4\cdot}| = 3.5 > D, \quad |\bar{y}_{\cdot1\cdot} - \bar{y}_{\cdot5\cdot}| = 3.25 > D$$

$$|\bar{y}_{\cdot1\cdot} - \bar{y}_{\cdot6\cdot}| = 2.0 > D$$

লক্ষ্য করা যাচ্ছে যে, Control চর্যার গড় [$\bar{y}_{.1.}$] অন্যান্য Pesticide-এর গড় হতে তাৎপর্যপূর্ণভাবে পার্থক্য। এখানে Control-এর তুলনায় যে কোনো Pesticide Slug-এর Heart beat-কে তাৎপর্যপূর্ণভাবে কমিয়েছে।

৭.৮ লাতিন বর্গ নকশা (Latin Square Design)

পরীক্ষণ এককে একমুখী ভেদ থাকলে তাকে নিয়ন্ত্রণ করার জন্য দৈবায়িত ব্লক নকশা প্রয়োগ করা হয়। বাস্তবক্ষেত্রে পরীক্ষণ এককে দুই মুখী ভেদ থাকতে পারে। যেমন, কোনো নিরীক্ষার ক্ষেত্রে এক একটি বা কয়েকটি Slug হলো পরীক্ষণ একক। নিরীক্ষাটির উদ্দেশ্য হলো Pesticide প্রয়োগ করে Slug-কে কিভাবে ধ্বংস করা যায় তা পর্যবেক্ষণ করা। এক্ষেত্রে Pesticide-এর বিভিন্ন স্তর (বিভিন্ন প্রকার Pesticide বা একই Pesticide-এর বিভিন্ন ঘনত্ব) হলো চর্যা। এই চর্যা Slug-এর উপর প্রয়োগ করতে লক্ষ্য করা যাবে যে Slug-গুলি এক জাতীয় নয় এবং একই জাতীয় Slug-এর বয়স বা Body weight বা Body length একই রকম নয়। অর্থাৎ Slug-সমূহের মধ্যে দুই রকমের ভেদ বিদ্যমান—(১) Slug-এর প্রকার, (২) Slug-এর Body length। Pesticide প্রয়োগ করে Slug-এর মৃত্যুর উপর প্রভাব নিরূপণ করতে হলে Slug-সমূহ একই জাতীয় হওয়া উচিত। তা না হলে Slug-এর মৃত্যু Pesticide-এর ঘনত্বের জন্য হলো নাকি Slug-এর প্রকার ভেদের জন্য হলো নাকি Slug-এর Body length ভেদের জন্য হলো তা নির্দিষ্ট করে বলা যাবে না। এক্ষেত্রে Slug-এর মধ্যে যে দুই প্রকারের ভেদ আছে তা নিয়ন্ত্রণ করে Pesticide প্রয়োগ করতে হবে। পরীক্ষণ এককে বিদ্যমান দুই প্রকার ভেদকে নিয়ন্ত্রণ করে পরীক্ষা পরিচালনা করার পদ্ধতি হলো লাতিন বর্গ নকশা।

লাতিন বর্গ নকশা প্রয়োগের ক্ষেত্রে পরীক্ষণ এককসমূহকে ঐগুলির ভেদের উৎস অনুসারে সারি (row) এবং স্তম্ভে (column) বিভক্ত করতে হয়। যেমন, চার প্রকার Slug হলে এক একপ্রকারের Slug-কে এক একটি সারি হিসেবে বিভক্ত করা যেতে পারে। আবার Slug-এর Body length-এর পরিমাণ অনুযায়ী ঐগুলিকে চারটি ভাগে বিভক্ত করা যেতে পারে। শেষোক্ত এক একভাগকে স্তম্ভ (Column) বলা হয়। এক্ষেত্রে Slug-এর উপর চার প্রকার Pesticide প্রয়োগ করা যেতে পারে। এই পরীক্ষার জন্য Slug-এর উপর Pesticide (চর্যা) এমনভাবে প্রয়োগ করতে হবে যেন প্রতি সারির এবং প্রতি স্তম্ভের Slug-এর উপর যে কোনো Pesticide কেবল একবারই প্রয়োগ করা হয়। চর্যার বণ্টন পরীক্ষণ এককে উপরিউক্তভাবে করা হলে প্রাপ্ত নকশাকে লাতিন বর্গ নকশা বলা হয়।

লাতিন বর্গ নকশার ক্ষেত্রে সারি, স্তম্ভ এবং চর্যার সংখ্যা সমান হতে হয়। ধরা যাক, কোনো পরীক্ষার জন্য চারটি চর্যা হলো A, B, C এবং D। তাহলে সারি এবং স্তম্ভে চর্যার বণ্টন ব্যবস্থা হবে নিম্নরূপ। আবার তিনটি চর্যা A, B এবং C

সারি	স্তম্ভ			
	C_1	C_2	C_3	C_4
R_1	A	B	C	D
R_2	B	C	D	A
R_3	C	D	A	B
R_4	D	A	B	C

চিত্র ৭.১ : 4x4 লাতিন বর্গ নকশা।

হলে লাতিন বর্গ নকশা হবে নিম্নরূপ :

সারি	স্তম্ভ		
	C_1	C_2	C_3
R_1	A	B	C
R_2	B	C	A
R_3	C	A	B

চিত্র ৭.২ : 3x3 লাতিন বর্গ নকশা।

এখানে লক্ষণীয় যে, উপরিউক্ত পরীক্ষার ক্ষেত্রে তিনটি উপাদান আছে (সারি, স্তম্ভ এবং চর্যা)। প্রতিটি উপাদানের স্তর সমান। প্রতিটি উপাদানের চারটি করে স্তর থাকলে মোট স্তরের সংখ্যা হওয়া উচিত $4 \times 4 \times 4$। কিন্তু বাস্তবে লক্ষ্য করা যাচ্ছে যে চারটি চর্যাবিশিষ্ট লাতিন বর্গ নকশা হতে প্রাপ্ত তথ্যমানের সংখ্যা হলো 4×4। এ কারণেই চার চর্যাবিশিষ্ট লাতিন বর্গ নকশাকে 4×4 লাতিন বর্গ নকশা বলা হয়।

চর্যা, সারি এবং স্তম্ভের সংখ্যা সমান হওয়ার কারণে বাস্তবে 8×8 বা 10×10 লাতিন বর্গ নকশার চেয়ে বড় নকশা ব্যবহৃত হয় না। এই নকশার ব্যবহার পরীক্ষাগারেই এবং জীব-জন্তু নিয়ে পরীক্ষার ক্ষেত্রেই বেশি ব্যবহৃত হয়।

মনে করা যাক কোনো পরীক্ষার জন্য k চর্যা আছে। তাহলে চর্যা এবং সারির সংখ্যাও k হবে। ধরা যাক i-তম সারির j-তম স্তম্ভে l-তম চর্যার উৎপাদন হলো y_{ijl} $[i = j = l = 1, 2, \ldots, k]$। এই y_{ijl} তথ্যমানের জন্য প্রতিকৃতি হলো

$$y_{ijl} = \mu + \alpha_i + \beta_j + \gamma_l + e_{ijl} \qquad (৭.৮.১)$$

এখানে μ = সাধারণ গড়, α_i = i-তম সারির প্রভাব, β_j = j-তম স্তম্ভের প্রভাব, γ_l = l-তম চর্যার প্রভাব এবং e_{ijl} = দৈব বিচ্যুতি।

অনুমান : (i) প্রতিকৃতি (৭.৮.১) স্থির প্রভাববিশিষ্ট,

(ii) $e_{ijl} \sim \text{NID}(0, \sigma^2)$

উপরিউক্ত অনুমানের ভিত্তিতে প্রতিকৃতি (৭.৮.১) হতে নিরূপিত বিচ্যুতির বর্গ-সমষ্টি হলো

$$\varphi = \sum_i^k \sum_j^k \sum_l^k \left[y_{ijl} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j - \hat{\gamma}_l \right]^2 \qquad (৭.৮.২)$$

এবং $\hat{\mu}$, $\hat{\alpha}_i$, $\hat{\beta}_j$ ও $\hat{\gamma}_l$-এর মান পাওয়ার জন্য পরিমিত সমীকরণসমূহ হলো যথাক্রমে

$$\frac{\partial\varphi}{\partial\hat{\mu}} = 0, \quad \frac{\partial\varphi}{\partial\hat{\alpha}_i} = 0, \quad \frac{\partial\varphi}{\partial\hat{\beta}_j} = 0, \quad \frac{\partial\varphi}{\partial\hat{\gamma}_l} = 0$$

এখন $\Sigma\hat{\alpha}_i = \Sigma\hat{\beta}_j = \Sigma\hat{\gamma}_l = 0$ শর্তাধীনে উপরিউক্ত সমীকরণগুলি সমাধান করে পাওয়া যায়

$$\hat{\mu} = \bar{y}_{\ldots}, \quad \hat{\alpha}_i = \bar{y}_{i..} - \bar{y}_{\ldots}, \quad \hat{\beta}_j = \bar{y}_{.j.} - \bar{y}_{\ldots}, \quad \hat{\gamma}_l = \bar{y}_{..l} - \bar{y}_{\ldots}$$

উপাত্তের মোট বর্গসমষ্টিকে পৃথককরণ করে পাওয়া যায় নিম্নরূপ :

$$\sum_i^k \sum_j^k \sum_l^k (y_{ijl} - \bar{y}_{\ldots})^2 = \sum_i^k \sum_j^k \sum_l^k [(\bar{y}_{i..} - \bar{y}_{\ldots}) + (\bar{y}_{.j.} - \bar{y}_{\ldots})$$

$$+ (\bar{y}_{..l} - \bar{y}_{\ldots}) + (y_{ijl} - \bar{y}_{i..} - \bar{y}_{.j.} - \bar{y}_{..l} + 2\bar{y}_{\ldots})]^2$$

$$= k\Sigma(\bar{y}_{i..} - \bar{y}_{...})^2 + k\,\Sigma(\bar{y}_{.j.} - \bar{y}_{...})^2$$
$$+ k\,\Sigma(\bar{y}_{..l} - \bar{y}_{...})^2 + \Sigma\Sigma\Sigma(y_{ijl} - \bar{y}_{i..} - \bar{y}_{.j.} - \bar{y}_{..l} + \bar{y}_{...})^2$$
$$= \text{SS Row}) + \text{SS(Column)} + \text{SS(Treatment)}$$
$$+ \text{SS(error)}$$
$$= S_1 + S_2 + S_3 + S_4$$

সারণি ৭.৯ : লাতিন বর্গ নকশার ক্ষেত্রে ভেদাঙ্ক বিশ্লেষণ সারণি।

ভেদের উৎস	স্বাধীনতার মাত্রা, df,	বর্গসমষ্টি SS	গড় বর্গসমষ্টি MS = SS/d,f	F	$F_{.05}$
সারি	$k-1$	S_1	s_1	s_1/s_4	
স্তম্ভ	$k-1$	S_2	s_2	s_2/s_4	
চর্যা	$k-1$	S_3	s_3	s_3/s_4	
বিচ্যুতি	$(k-1)(k-2)$	S_4	s_4		
মোট	k^2-1				

এই বিশ্লেষণের মূল উদ্দেশ্য হলো নাস্তি কল্পনা

$$H_0 : \quad \gamma_l = 0 \qquad (৭.৮.৩)$$

বিকল্প কল্পনা $\quad H_A : \quad \gamma_l \neq 0$

যাচাই করা। যাচাই তথ্যজমান $F = s_3/s_4$ $(k-1)$ এবং $(k-1)(k-2)$ স্বাধীনতার মাত্রাবিশিষ্ট $F_{0.05}$ হতে বড় অথবা তার সমান হলে নাস্তি কল্পনা বাতিল বলে গণ্য হবে। অনেক সময় নাস্তি কল্পনা $H_0 : \alpha_i = 0$ এবং $H_0 : \beta_j = 0$ যাচাই করে দেখতে হয়। এই শেষোক্ত নাস্তি কল্পনাদ্বয়ের জন্য যাচাই তথ্যজমানদ্বয় হলো যথাক্রমে $F = s_1/s_4$ এবং $F = s_2/s_4$ এই দুটি যাচাই তথ্যজমান সম্পর্কেও সিদ্ধান্ত পূর্বোক্ত সিদ্ধান্ত নেয়ার পদ্ধতিতে নিতে হয়।

নাস্তি কল্পনা (৭.৮.৩) বাতিল হলে যে কোনো দুটি চর্যার পার্থক্য যাচাই করার জন্য নাস্তি কল্পনা

$$H_0 : \quad \gamma_l = \gamma_{l'} \quad (l \neq l' = 1, 2, \cdots, k) \qquad (৭.৮.৪)$$

বিকল্প কল্পনা $\quad H_A : \quad \gamma_l \neq \gamma_{l'}$

যাচাই করতে হয়। এক্ষেত্রে যাচাই তথ্যজমান হলো

$$t = \frac{\bar{y}_{..l} - \bar{y}_{..l'}}{\sqrt{\dfrac{2s_4}{k}}} \qquad (৭.৮.৫)$$

এখন $|t| \geqslant t_{0\cdot025, (k-1)(k-2)}$ হলে নাস্তি কল্পনা বাতিল বলে গণ্য হবে।

নাস্তি কল্পনা (৭.৮.৪)-এর উদ্দেশ্য যদি প্রতি জোড়া চর্যার তুলনা করা হয়, তাহলে যাচাই তথ্যজমান হবে DMRT। এখানে DMRT হলো

$$R_p = r_{0\cdot05, p, (k-1)(k-2)} \sqrt{\frac{s_4}{k}} \qquad (৭.৮.৬)$$

$$p = 2, 3, \ldots, k$$

আবার কোনো একটি চর্যা [ধরা যাক l-তম চর্যা] নিয়ন্ত্রিত চর্যা হলে তাকে অন্য চর্যাসমূহের সাথে তুলনা করার জন্য যাচাই তথ্যজমান হবে Dunnett's test, যেখানে

$$D = d_{0\cdot05, k-1, (k-1)(k-2)} \sqrt{\frac{2 s_4}{k}} \qquad (৭.৮.৭)$$

এখানে (৭.৮.৬) ও (৭.৮.৭) তথ্যজমান সম্পর্কে সিদ্ধান্ত ৭.৬ অনুচ্ছেদে আলোচিত পদ্ধতি নিতে হয়।

উদাহরণ ৭.৫

একটি পরীক্ষাগারে Slug কে মারার জন্য 4 প্রকার Pesticide প্রয়োগ করে একটি পরীক্ষা পরিচালনা করা হয়েছে। পরীক্ষায় ব্যবহৃত Pesticide গুলি হলো Supracide, Pomex, Sumicidin এবং Glyphosate। প্রতিটি Pesticide এর 4 প্রকার ঘনত্ব (concentration) যেমন, (i) 100 mg/kg(A), (ii) 125 mg/kg(B), (iii) 150 mg/kg(C) এবং (iv) 200 mg/kg(D)সম্পন্ন তরল 4 প্রকার Slug-এর উপর প্রয়োগ করা হয়েছে। 21 দিন পর Slug গুলির মৃত্যু হার নথিভুক্ত করা হয়েছে। প্রতিটি Pesticide-এর প্রতিটি ঘনত্ব প্রতি প্রকার Slug-এর 10 (দশটির) উপর প্রয়োগ করা হয়েছে। নিচে সারণি ৭.১০-এ Slug-এর মৃত্যু হার দেখানো হলো।

সারণি ৭.১০ : বিভিন্ন ঘনত্বসহ Pesticide-এর বিভিন্ন স্তর বিভিন্ন Slug-এর উপর প্রয়োগ করার পর Slug-এর মৃত্যু হার $[y_{ijl}]$

Slug	Pesticide				
	Supracide	Pomex	Sumicidin	Glyphosate	মোট $y_{i..}$
M.Rusticus	A – 43.3	B – 56.7	C – 43.3	D – 13.3	156.6
M. Sowerbyi	B – 50.0	C – 10.0	D – 23.3	A – 30.0	113.3
M. Gagates	C – 90.0	D – 60.0	A – 20.0	B – 33.3	203.3
L. Tenellus	D – 43.3	A – 36.7	B – 23.3	C – 96.7	200.0
মোট $y_{.j.}$	226.6	163.4	109.9	173.3	673.2
গড় $\bar{y}_{.j.}$	56.65	40.85	27.475	43.325	

উপাত্ত হতে Slug-এর মৃত্যু হারের উপর Pesticide এবং ঐগুলির ঘনত্বের প্রভাব পর্যালোচনা করা যাক।

এখানে $k=4$, $G=673.2$, ঘনত্ব অনুযায়ী Slug-এর মোট মৃত্যু হার এবং গড় মৃত্যু হার হলো

$y_{..l}$: $A=130 \quad B=163.3 \quad C=240 \quad D=139.9$

$\bar{y}_{..l}$: $\bar{A}=32.5 \quad \bar{B}=40.825 \quad \bar{C}=60.00 \quad \bar{D}=34.975$

$$C.T=\frac{G^2}{k^2}=\frac{673.2}{4\times 4}=28324.89$$

মোট বর্গসমষ্টি, $TSS=\Sigma\Sigma\Sigma y_{ijl}^2-C.T=37508.9-28324.89$

$$=9184.01$$

$$SS\ (\text{ঘনত্ব})=\frac{\Sigma y_{..l}^2}{k}-C.T=\frac{120738.9}{4}-28324.89$$

$$=1859.835$$

$$SS(\text{Slug})=\frac{\Sigma y_{i..}^2}{k}-C.T=\frac{118691.34}{4}-28324.89=1347.945$$

$$SS(\text{Pesticide})=\frac{\Sigma_{.j.}^2}{k}-C.T=\frac{120158.02}{4}-28324.89=1714.615$$

$$SS(\text{বিচ্যুতি})=TSS-SS(\text{Slug})-SS(\text{Pesticide})-SS(\text{ঘনত্ব})$$

$$=9184.01-1347.945-1714.615-1859.835$$

$$=4261.615$$

ভেদাঙ্ক বিশ্লেষণ সারণি

ভেদের উৎস	স্বাধীনতার মাত্রা, d.f.	বর্গসমষ্টি SS	গড় বর্গসমষ্টি MS = SS/d.f	F	$F_{.05}$
Slug	3	1347.945	449.315	0.63	4.76
Pesticide	3	1714.615	571.538	0.80	4.76
ঘনত্ব	3	1859.835	619.945	0.87	4.76
বিচ্যুতি	6	4261.615	710.269		
মোট	15				

লক্ষ্য করা যাচ্ছে যে, নির্ণেয় সব F-ই $F_{.05}$ অপেক্ষা ছোট [$p > 0.05$]। এতে বুঝা যায় (i) Pesticide প্রয়োগে সব ধরনের Slug-এর মৃত্যু হার একই রকম, (ii) কোনো Pesticide-ই Slug-এর মৃত্যুর জন্য বেশি ক্রিয়াশীল নয়, (iii) ব্যবহৃত Pesticide-এর কোনো ঘনত্বই Slug-এর মৃত্যুর উপর বেশি প্রভাব ফেলতে পারেনি।

উপরিউক্ত বিশ্লেষণে যেহেতু কোনো উপাদানেরই তাৎপর্যপূর্ণ পার্থক্য পরিলক্ষিত হয়নি, সে কারণে যাচাই তথ্যজমান (৭.৮.৬) বা (৭.৮.৭) প্রয়োগ করার কোনো প্রয়োজন নেই।

লাতিন বর্গনকশার ক্ষেত্রে একটি তথ্যমান লুপ্ত হলে বিশ্লেষণ পদ্ধতি (Analysis with one missing observation in case of Latin square design) : ধরা যাক l-তম চর্যার তথ্যমান i-তম সারির j-তম স্তম্ভে লুপ্ত হয়েছে। মনে করি এই তথ্যমানটি x_1 তাহলে, সারি, স্তম্ভ এবং চর্যার মোট উৎপাদনকে লেখা যাবে নিম্নরূপভাবে :

k × k লাতিন বর্গনকশা

সারি	স্তম্ভ						মোট
	1	2	…	j	…	k	
1							R_1
2							R_2
⋮							
i				x			$R_i + x$
⋮							
k							R_k
মোট	C_1	C_2		$C_j + x$		C_k	$G + x$

চর্যা

	1	2	…	l	…	k
মোট	T_1	T_2	…	$T_l + x$	…	T_k

ন্যূনতম বর্গপদ্ধতি প্রয়োগ করে x-এর মান পাওয়া যায়, যেখানে

$$x = \frac{k(R_1 + C_j + T_l) - 2G}{(k-1)(k-2)} \qquad (৭.৮.৮)$$

এই বিশ্লেষণের শুরুতে প্রাপ্ত উপাত্তের ভিত্তিতে x-এর মান (৭.৮.৮) সূত্র প্রয়োগ করে নিরূপণ করতে হয় এবং x-এর পরিবর্তে নিরূপিত মান বসিয়ে স্বাভাবিক বিশ্লেষণ করতে হয়। তবে একটি তথ্যমান নষ্ট হওয়ার কারণে মোট স্বাধীনতার মাত্রা $[k^2 - 1]$ হতে 1 বিয়োগ করতে হয় এবং সে কারণে বিচ্যুতির স্বাধীনতার মাত্রা $[(k-1)(k-2)]$ হতেও 1 বিয়োগ করতে হয়।

x-এর মান বসিয়ে বিশ্লেষণের ক্ষেত্রে

$C.T = \frac{(G+x)^2}{k^2}$, মোট বর্গসমষ্টি, $TSS = \Sigma\Sigma\Sigma y_{ijl}^2 + x^2 - C.T$

$$SS(\text{সারি}) = \frac{\Sigma R_{i'}^2}{k} + \frac{(R_1 + x)^2}{k} - C.T = S_1$$

$$SS(\text{স্তম্ভ}) = \frac{\Sigma C_{j'}^2}{k} + \frac{(C_j + x)^2}{k} - C.T = S_2$$

$$SS(\text{চর্যা}) = \frac{\Sigma T_{l'}^2}{k} + \frac{(T_l + x)^2}{k} - C.T = S_3$$

$S_4 = SS(\text{বিচ্যুতি}) = TSS - SS(\text{সারি}) - SS(\text{স্তম্ভ}) - SS(\text{চর্যা})$

এই বিশ্লেষণের ক্ষেত্রে ভেদাঙ্ক বিশ্লেষণ সারণি হবে নিম্নরূপ :

সারণি ৭.১১ : একটি লুপ্তমানের ক্ষেত্রে লাতিন বর্গনকশার জন্য ভেদাঙ্ক বিশ্লেষণ সারণি।

ভেদের উৎস	স্বাধীনতার মাত্রা, d.f.	বর্গসমষ্টি SS	গড় বর্গসমষ্টি MSS/d.f	F
সারি	$k-1$	S_1	s_1	s_1/s_4
স্তম্ভ	$k-1$	S_2	s_2	s_2/s_4
চর্যা	$k-1$	S_3	s_3	s_3/s_4
বিচ্যুতি	$(k-1)(k-2)-1$	S_4	s_4	
মোট	k^2-2			

এই বিশ্লেষণের ক্ষেত্রে লুপ্তমানবিশিষ্ট চর্যার সাথে অন্য চর্যার তুলনা করার জন্য যাচাই তথ্যজমান হলো

$$t = \frac{\bar{y}_{\cdot\cdot l} - \bar{y}_{\cdot\cdot l'}}{\sqrt{s_4\left[\frac{2}{k} + \frac{1}{(k-1)(k-2)}\right]}}$$

এখন $|t| \geqslant t_{0.025,(k-1)(k-2)-1}$ হলে l-তম চর্যা ও l'-তম [$l' \neq l = 1, 2, \ldots, k$] চর্যার পার্থক্য তাৎপর্যপূর্ণ বলে বিবেচিত হবে।

উদাহরণ ৭.৬

একটি দুগ্ধ উৎপাদন খামারে চার প্রকার গো-খাদ্য [F_1, F_2, F_3, F_4] হতে সবচেয়ে ভাল একটি গো-খাদ্য চিহ্নিত করার জন্য নিরীক্ষার পরিকল্পনা করা হয়েছে। এই গো-খাদ্যগুলির মধ্যে F_1-এর ফলাফল পূর্ব থেকে জানা। উক্ত নিরীক্ষার জন্য চার জাতের গরু নির্বাচিত করে ঐগুলিকে দুধ দেয়ার স্তর (lactation period) অনুযায়ী চারভাগে ভাগ করে প্রতি ভাগের এক একটি গরুকে লাতিন বর্গনকশার নিয়ম অনুযায়ী গো-খাদ্য খাওয়ানো হয়েছে এবং নিরীক্ষাকালীন সময়ে প্রতি গরুর দৈনিক গড় দুধ উৎপাদন (kg) নথিভুক্ত করা হয়েছে। নিচে প্রতি জাতের, প্রতি দুধ দেয়ার স্তরের গরুর গড় উৎপাদন দেয়া হলো। কিন্তু নথি অনুযায়ী দেখা যাচ্ছে যে, প্রথম দুধ দেয়ার স্তরের একটি গরুর দুধ উৎপাদনের পরিমাণ নেই। প্রাপ্ত উপাত্তের বিশ্লেষণ পূর্বক সবচেয়ে ভাল গো-খাদ্য চিহ্নিত করা যাক এবং F_2, F_3 ও F_4 F_1 অপেক্ষা ভাল কিনা সে সম্পর্কে সিদ্ধান্ত নেয়া যাক।

সারণি ৭.১২ : লাতিন বর্গনকশার মাধ্যমে প্রাপ্ত চার প্রকার গো-খাদ্যের বিপরীতে দুধের উৎপাদন (y_{ijl} kg)।

গরুর জাত	দুধ দেয়ার স্তর				
	L_1	L_2	L_3	L_4	মোট $y_{i\cdot\cdot} = R_i$
Holland, H	$F_1 - 30.5$	$F_2 - 38.4$	$F_3 - 40.6$	$F_4 - 35.2$	144.7
Australia, A	$F_2 - x$	$F_3 - 42.5$	$F_4 - 38.2$	$F_1 - 28.6$	$109.3 + x$
Newzealand N	$F_3 - 25.6$	$F_4 - 29.2$	$F_1 - 30.1$	$F_2 - 25.0$	109.9
Local, L	$F_4 - 18.2$	$F_1 - 20.5$	$F_2 - 22.3$	$F_3 - 19.0$	80.0
মোট $y_{\cdot j\cdot} = C_j$	$74.3 + x$	130.6	131.2	107.8	$443.9 + x$

এখানে $R_i = 109.3$, $C_j = 74.3$, $T_l = 85.7$, $G = 443.9$, $k = 4$

সুতরাং $$x = \frac{k(R_i + C_j + T_l) - 2G}{(k-1)(k-2)}$$

$$= \frac{4(109.3 + 74.3 + 85.7) - 2 \times 443.9}{(4-1)(4-2)} = 31.6$$

x-এর মান বসিয়ে বিশ্লেষিত ফলাফল নিম্নরূপ :

$R_1 = 144.7,\ R_2 = 109.3 + 31.6 = 140.9,\ R_3 = 109.9,\ R_4 = 80.0$

$C_1 = 74.3 + 31.6 = 105.9,\ C_2 = 130.6,\ C_3 = 131.2,\ C_4 = 107.8$

$T_1 = 109.7,\ T_2 = 85.7 + 31.6 = 117.3,\ T_3 = 127.7, T_4 = 120.8$

$$C.T = \frac{G + x)^2}{k^2} = \frac{(443.9 + 31.6)^2}{4 \times 4} = 14131.266$$

$$TSS = \Sigma\Sigma\Sigma y_{ijl}^2 - C.T = 15023.01 - 14131.266$$
$$= 891.744$$

$$SS(\text{সারি}) = \frac{R_1^2 + R_2^2 + R_3^2 + R_4^2}{4} - C.T = \frac{59268.91}{4} - 14131.266$$
$$= 685.961$$

$$SS(\text{স্তম্ভ}) = \frac{C_1^2 + C_2^2 + C_3^2 + C_4^2}{4} - C.T = \frac{57105.45}{5} - 14131.266$$
$$= 145.096$$

$$SS(\text{চর্যা}) = \frac{T_1^2 + T_2^2 + T_3^2 + T_4^2}{4} - C.T = \frac{56693.31}{4} - 14131.266$$
$$= 42.061$$

$$SS(\text{বিচ্যুতি}) = TSS - SS(\text{সারি}) - SS(\text{স্তম্ভ}) - SS(\text{চর্যা})$$
$$= 891.744 - 685.961 - 145.096 - 42.061$$
$$= 18.626$$

ভেদাঙ্ক বিশ্লেষণ সারণি

ভেদের উৎস	স্বাধীনতার মাত্রা d.f	বর্গসমষ্টি SS	গড় বর্গসমষ্টি $MS = \frac{SS}{d.f}$	F	$F_{.05}$
সারি	3	685.961	228.654	61.38	5.41
স্তম্ভ	3	145.096	48.365	12.98	5.41
চর্যা	3	42.061	14.020	3.76	5.41
বিচ্যুতি	5	18.626	3.7252		
মোট	14				

এখানে চর্যা হিসেবে ব্যবহৃত হয়েছে চার প্রকার গো-খাদ্য। চর্যার প্রাসঙ্গিক $F = 3.76 < F_{.05}$ হওয়ায় গো-খাদ্যগুলির মধ্যে কোনো পার্থক্য থাকার প্রমাণ পাওয়া যায়নি। সারিতে আছে গরুর জাত। সারির প্রাসঙ্গিক $F = 61.38 > F_{.05}$ হওয়ায় দুধ উৎপাদনে গরুর জাতের মধ্যে তাৎপর্যপূর্ণ পার্থক্য আছে সিদ্ধান্ত নেয়া যায়। আবার স্তম্ভে ব্যবহৃত হয়েছে গরুর দুধ দেয়ার স্তর (lactation period)। স্তম্ভের প্রাসঙ্গিক $F = 12.98 > F_{.05}$ হওয়ায় দুধ দেয়ার স্তরের পার্থক্যের কারণে দুধ উৎপাদনে তাৎপর্যপূর্ণ পার্থক্য পরিলক্ষিত হচ্ছে।

এই বিশ্লেষণ হতে কোনো গো-খাদ্যকেই সবচেয়ে ভাল বলা যাবে না। সবগুলির প্রভাবই অনুরূপ। সে কারণে Dunnett's যাচাই প্রয়োগ করে F_2, F_3 এবং F_4 F_1 অপেক্ষা ভাল বা মন্দ বলার প্রয়োজন নেই। তবে দুধ দেয়ার স্তরের পার্থক্যের কারণে গড় দুধ উৎপাদনে কিরূপ পার্থক্য আছে তা দেখার জন্য DMRT প্রয়োগ করা যায়। এখানে

$$R_p = r_{0.05, p, 5}\sqrt{\frac{s_4}{k}}\ , \ p = 2, 3, 4$$

$$R_2 = 3.64\sqrt{\frac{3.7252}{4}} = 3.513$$

$$R_3 = 3.74\sqrt{\frac{3.7252}{4}} = 3.609,\ R_4 = 3.79\sqrt{\frac{3.7252}{4}} = 3.657$$

দুধ দেয়ার স্তরের পার্থক্যের কারণে গড় দুধ উৎপাদনকে মানের ক্রম অনুসারে সাজিয়ে পাওয়া যায়

$$\bar{C}_1 = \bar{L}_1 = 26.475, \bar{C}_4 = \bar{L}_4 = 26.95,\ \bar{C}_2 = \bar{L}_2 = 32.65, \bar{C}_3 = \bar{L}_3 = 32.80$$

$$\bar{L}_3 - \bar{L}_1 = 6.325 > R_4\ ;\ \bar{L}_2 - \bar{L}_1 = 6.175 > R_3$$

$$\bar{L}_3 - \bar{L}_4 = 5.85 > R_3\ ;\ \bar{L}_4 - \bar{L}_1 = 0.475 < R_2$$

$$\bar{L}_2 - \bar{L}_4 = 5.70 > R_2\ ;\ \bar{L}_3 - \bar{L}_2 = 0.15 < R_2$$

লক্ষ্য করা যাচ্ছে যে, প্রথম দুধ দেয়ার স্তরে (L_1) এবং চতুর্থ দুধ দেয়ার স্তরের (L_4) গরু একইরূপ দুধ দেয়। আবার দ্বিতীয় এবং তৃতীয় দুধ দেয়ার স্তরের গরুও একইরূপ দুধ দেয় এবং এই দুই স্তরের গরুর দুধ দেয়ার পরিমাণ অন্য দুই স্তরের গরুর দুধ দেয়ার পরিমাণ অপেক্ষা বেশি। নিচে দুধ দেয়ার স্তর অনুযায়ী গরুগুলিকে গুচ্ছভুক্ত করে দেখানো হলো

$$(L_1\ L_4),\ (L_2\ L_3)$$

অনুরূপভাবে গরুগুলিকে ঐগুলির জাত অনুসারেও গুচ্ছভুক্ত করে দেখানো যায়। সেক্ষেত্রে

$$R_p = r_{0.05, p, 6}\sqrt{\frac{s_4}{k}}, \quad p = 2, 3, 4$$

এবং জাত অনুসারে গরুর গড় দুধের পরিমাণ মানের ক্রম অনুসারে সাজিয়ে পাওয়া যায় :

$$\bar{L} = 20.00, \ \bar{N} = 27.475, \ \bar{A} = 35.225, \ \bar{H} = 36.175$$

এখানে

$$\bar{H} - \bar{L} = 16.175 > R_4, \ \bar{A} - \bar{L} = 15.225 > R_3, \ \bar{H} - \bar{N} = 8.7 > R_3$$

$$\bar{N} - \bar{L} = 7.475 > R_2, \ \bar{A} - \bar{N} = 7.475 > R_2, \ \bar{H} - \bar{A} = 0.95 < R_2$$

এখানে একই গুচ্ছভুক্ত গরু হলো

$$\{H, A\}, \ \{N\}, \ \{L\}$$

৭.৯ উপাদানী পরীক্ষা (Factorial Experiment)

পূর্ববর্তী অনুচ্ছেদসমূহে লক্ষ্য করা গেছে যে খণ্ডে বা ব্লকে বা সারি এবং স্তম্ভে চর্যা বণ্টন করা হয় বিভিন্ন পদ্ধতিতে। সেক্ষেত্রে চর্যা হলো যে কোনো উপাদানের কয়েকটি স্তর। যেমন, উদাহরণ ৭.৩-এ ইউরিয়া সারের চারটি স্তর হলো চারটি চর্যা। ঐ উদাহরণেই ব্লক হিসেবে বিবেচনা করা হয়েছে পটাশ সারের তিনটি স্তরকে। এখানে পটাশ সার আরেকটি উপাদান। এই দুটি উপাদানের বিভিন্ন স্তরের বণ্টন ব্যবস্থা আলাদা। বাস্তবে উপাদানদ্বয়ের বিভিন্ন স্তরের বণ্টন ব্যবস্থা আলাদা না করে ঐ স্তরগুলিকে মিশ্রিত (combined) করা হয় এবং মিশ্রিত স্তরগুলিকে চর্যা হিসেবে বিবেচনা করা হয়। এরূপ চর্যা নিয়ে কোনো পরীক্ষা পরিচালনা করা হলে তাকে উপাদানী পরীক্ষা বলা হয়। যেমন, উদাহরণ ৭.৩-এর ক্ষেত্রে ইউরিয়া সারের এবং পটাশ সারের মিশ্রিত স্তর হলো

$$N_1K_1, N_2K_1, N_3K_1, N_4K_1, N_1K_2, N_2K_2, N_3K_2$$
$$N_4K_2, N_1K_3, N_2K_3, N_3K_3, N_4K_3$$

এই বারটি [4 × 3] মিশ্রিত স্তরকে কোনো পরীক্ষায় চর্যা হিসেবে ব্যবহার করা হলে ঐ পরীক্ষাকে উপাদানী পরীক্ষা বলা হয়।

উপরিউক্ত উপাদানী পরীক্ষার ক্ষেত্রে দুটি উপাদানের অসমান স্তরের মিশ্রণ বিবেচনা করা হয়েছে। বাস্তবে উপাদানের সংখ্যা 2-এর (n, যে কোনো সংখ্যা) অধিক হতে পারে এবং প্রতি উপাদানের সমান এবং অসমান স্তর হতে পারে। প্রতি উপাদানের স্তর সমান হলে তাকে প্রতিসম (symmetric) উপাদানী পরীক্ষা বলা হয়। এই প্রতিসম উপাদানী পরীক্ষাকে লেখা হয় p^n-উপাদানী পরীক্ষা। এখানে n উপাদানের প্রতিটির p স্তর আছে। আবার বিভিন্ন উপাদানের বিভিন্ন স্তর হলে তাকে অপ্রতিসম (asymmetric) উপাদানী পরীক্ষা বলা হয়। যেমন, উপরিউক্ত উদাহরণের ক্ষেত্রে ইউরিয়া সারের স্তর চারটি এবং পটাশ সারের স্তর তিনটি। একে লেখা হয় (4×3) উপাদানী পরীক্ষা। সাধারণত m উপাদান $F_1, F_2, \ldots, F_m$ হলে এবং তাদের স্তর যথাক্রমে $l_1, l_2, \cdots, l_m$ হলে উপাদানী পরীক্ষাকে লেখা হয় $(l_1 \times l_2 \times \cdots \times l_m)$-উপাদানী পরীক্ষা। এই পরীক্ষার ক্ষেত্রে উপাদানের স্তরসমূহের মিশ্রণ হলো $(l_1 \times l_2 \times \ldots \times l_m)$ এবং এই মিশ্রিত স্তরই চর্যা হিসেবে ব্যবহৃত হয়।

উপাদানী পরীক্ষার ক্ষেত্রে চর্যাকে কোনো পরীক্ষণ নকশার মাধ্যমে খণ্ডে বণ্টন করা হয়। ধরা যাক ইউরিয়া সারের 4 স্তর এবং পটাশ সারের 3 স্তরবিশিষ্ট অপ্রতিসম উপাদানী পরীক্ষাটি একটি সম্পূর্ণ দৈবায়িত নকশার মাধ্যমে পরিচালনা করা হবে। তাহলে, ঐ নকশায় মোট চর্যা সংখ্যা $4 \times 3 = 12$ এর জন্য 12 এর যে কোনো গুণিতক খণ্ড প্রয়োজন এবং খণ্ডে 12 চর্যা সম্পূর্ণ দৈবায়িত পদ্ধতিতে বণ্টন করতে হবে। আবার এই 12 চর্যা নিয়ে দৈবায়িত ব্লক নকশার মাধ্যমে পরীক্ষা পরিচালনা করতে হলে ব্লকে খণ্ডের সংখ্যা হতে হবে 12 (12 এর গুণিতকও হতে পারে) এবং 12 খণ্ডে দৈবায়িত পদ্ধতিতে চর্যাগুলিকে বণ্টন করতে হবে।

এই 4×3 উপাদানী পরীক্ষা এবং উদাহরণ ৭.৩-এ উল্লেখিত পরীক্ষাটির মধ্যে পার্থক্য হলো প্রথমোক্ত ক্ষেত্রে 12টি চর্যা সম্পূর্ণ দৈবায়িত পদ্ধতিতে হয় খণ্ডে বা ব্লকে বণ্টন করা হয় এবং শেষোক্ত ক্ষেত্রে দ্বিতীয় উপাদানের প্রতিটি স্তরে প্রথম উপাদানের 4টি স্তরকে দৈবায়িত পদ্ধতিতে বণ্টন করা হয়। এছাড়া উদাহরণ ৭.৩ এর ক্ষেত্রে শুধু ইউরিয়া সার ও পটাশ সারের প্রভাব নিরূপণ করা হয় অবশ্য পরীক্ষাটি r বার পুনরায়ন করা হলে ইউরিয়া সার ও পটাশ সারের মিশ্র প্রভাবও (interaction effect) নিরূপণ করা যায়। কিন্তু পরীক্ষাটি উপাদানী পরীক্ষা হলে প্রতি উপাদানের স্বস্ব প্রভাব এবং তাদের যে কোনো দুটি বা ততোধিক উপাদানের

মিশ্রপ্রভাব নিরূপণ করা যায়। এটিই উপাদানী পরীক্ষার সুবিধা। প্রাণবিজ্ঞান, জীব-বিজ্ঞান এবং চিকিৎসা বিজ্ঞানে দুই বা ততোধিক উপাদানের মিশ্রণে পরীক্ষা পরিচালনা করা এবং ঐ উপাদানসমূহের মিশ্রপ্রভাব নিরূপণ করা একটি স্বাভাবিক কাজ। সে কারণে উপরিউক্ত বিজ্ঞানের অনুচ্ছেদসমূহে উপাদানী পরীক্ষার ব্যাপক প্রয়োগ লক্ষ্য করা যায়।

উপাদানী পরীক্ষার কিছু অসুবিধাও আছে। উপাদান বা স্তর বেশি হলে যুগ্ম স্তরের সংখ্যা বা চর্যার সংখ্যা বেশি হয়। সেক্ষেত্রে এক জাতীয় খণ্ডে পরীক্ষা পরিচালনা করতে সমস্যার সম্মুখীন হওয়া স্বাভাবিক। কারণ চর্যা সংখ্যা বেশি হলে [যেমন, $s_1=3$, $s_2=3$, $s_3=3$ এবং $s_4=3$ হলেই যুগ্ম স্তরের সংখ্যা $3^4=81$ হয়] একই জাতীয় বেশি খণ্ড পাওয়ার বাস্তব অসুবিধা থাকে। যেমন, কৃষিক্ষেত্রে 3^4-উপাদানী পরীক্ষা পরিচালনা করতে হলে একই উর্বরা শক্তিসম্পন্ন $3^4=81$ খণ্ড জমি পাওয়া সহজসাধ্য নয়। অবশ্য এ জাতীয় সমস্যা দূর করার পদ্ধতি আছে। বর্তমান অনুচ্ছেদে সে সম্পর্কে আলোচনা করা হবে না। ইচ্ছুক পাঠকগণ Bhuyan 1993 পর্যালোচনা করতে পারেন।

উপাদানী পরীক্ষা এবং অ-উপাদানী পরীক্ষার (সম্পূর্ণ দৈবায়িত নকশা, দৈবায়িত ব্লক নকশা বা লাতিন বর্গ নকশা) মধ্যে পার্থক্য সম্পর্কে উপরে কিছুটা আলোকপাত করা হয়েছে। বিশ্লেষণ পদ্ধতিতে এই দুই পরীক্ষার মধ্যে পার্থক্য লক্ষ্য করা যায়। উপাদানী পরীক্ষার ক্ষেত্রে উপাদানসমূহের স্ব-স্ব মূল প্রভাব এবং ঐগুলির মিশ্রপ্রভাব সহজেই নিরূপণ করা যায় এবং মিশ্রপ্রভাবসমূহকে পৃথকীকরণ করা যায়। অ-উপাদানী পরীক্ষার ক্ষেত্রে মিশ্রপ্রভাবসমূহকে পৃথকীকরণ করা যায় না। বর্তমান অনুচ্ছেদ 2^n-উপাদানী পরীক্ষার সহজ বিশ্লেষণ পদ্ধতি আলোচনা করা হবে।

2^2-উপাদানী পরীক্ষা (2^2-Factorial Experiment) : ধরা যাক A এবং B দুটি উপাদান। তাদের প্রতিটির দুটি করে স্তর আছে। মোট যুগ্ম স্তরের সংখ্যা হলো $2\times2=4$। এই স্তরগুলিকে সাধারণত লেখা হয় (1) a, b এবং ab। এখানে a এবং b দ্বারা যথাক্রমে উপাদান A ও B-এর উপস্থিতি বুঝানো হয়েছে এবং (1) দ্বারা উভয় উপাদানের অনুপস্থিতি বুঝানো হয়েছে। অনেক সময় অনু-পস্থিতিকে উপাদানের প্রথম স্তর এবং উপস্থিতিকে তার দ্বিতীয় স্তর বিবেচনা করা হয়। তাই

(1) = উভয় উপাদানের প্রথম স্তর,

a = A-এর দ্বিতীয় স্তর এবং B-এর প্রথম স্তর ,

b = B-এর দ্বিতীয় স্তর এবং A-এর প্রথম স্তর,

ab = A ও B উভয়ের দ্বিতীয় স্তর।

এই পরীক্ষার ক্ষেত্রে কোনো উপাদানের প্রভাব নিরূপণ করার জন্য উপাদানের দ্বিতীয় স্তরের উৎপাদন হতে প্রথম স্তরের উৎপাদন বিয়োগ করা হয়। অর্থাৎ

$a-(1)=$ B-এর প্রথম স্তরের ক্ষেত্রে A-এর পভাব,

$ab-b=$ B-এর দ্বিতীয় স্তরের ক্ষেত্রে A-এর প্রভাব।

এই দুটি প্রভাবকে বলা হয় সরল প্রভাব এবং এগুলির যোগ করে পাওয়া যায় A-এর মোট প্রভাব (total effect)। সুতরাং

$$[A]=ab-b+a-(1)=(a-1)(b+1)$$

অনুরূপভাবে

$b-(1)=$ A-এর অনুপস্থিতিতে B-এর প্রভাব,

$ab-a=$ A-এর উপস্থিতিতে B-এর প্রভাব।

সুতরাং B-এর মোট প্রভাব হলো

$$[B]=ab-a+b-(1)=(a+1)(b-1)$$

B-এর উপস্থিতিতে A-এর প্রভাব হতে B-এর অনুপস্থিতিতে A-এর প্রভাব বিয়োগ করে পাওয়া যায় AB মিশ্রপ্রভাব। সুতরাং

$$[AB]=(ab-b)-(a-(1))=(a-1)(b-1)$$

এই বিশ্লেষণ হতে গড় প্রভাবসমূহ হলো

$$A=\frac{1}{2r}(a-1)(b+1)$$

$$B=\frac{1}{2r}(a+1)(b-1)$$

$$AB=\frac{1}{2r}(a-1)(b-1)$$

এখানে r হলো পরীক্ষাটির পুনরায়নের সংখ্যা।

লক্ষ্য করা যাচ্ছে যে, A, B ও AB-এর মোট প্রভাব নিরূপণ করার জন্য প্রতিটি চর্যার উৎপাদনের রৈখিক সংযোগ নির্ণয় করা হয়েছে। এই রৈখিক সংযোগের ক্ষেত্রে লক্ষণীয় বিষয় হলো যে কোনো প্রভাবের ক্ষেত্রেই অর্ধেক $[2^{n-1}]$ চর্যার উৎপাদন যোগ চিহ্নবিশিষ্ট এবং অর্ধেক $[2^{n-1}]$ চর্যার উৎপাদন বিয়োগ চিহ্নবিশিষ্ট। চিহ্ন সারণি (sign table)-এর মাধ্যমে এই চিহ্নগুলি নির্ণয় করা হয়। চিহ্ন সারণি তৈরি করার সময় চর্যায় কোনো উপাদান উপস্থিত থাকলে তার জন্য যোগ চিহ্ন এবং অনুপস্থিত থাকলে বিয়োগ চিহ্ন বিবেচনা করা হয়। চর্যায় যতোগুলি উপাদান থাকে তাদের চিহ্ন বিবেচনা করে ঐগুলির গুণফল করা হয় এবং গুণফল হতে প্রাপ্ত চিহ্নই ঐ চর্যার চিহ্ন বিবেচিত হয়। নিচে 2^2-উপাদানী পরীক্ষার ক্ষেত্রে চিহ্ন সারণি তৈরি করে দেখানো হলো :

সারণি ৭.১৩ : 2^2-উপাদানী পরীক্ষার ক্ষেত্রে চিহ্ন সারণি।

প্রভাবসমূহ	যুগ্ম স্তরসমূহ			
	(1)	a	b	ab
A	−	+	−	+
B	−	−	+	+
AB	+	−	−	+

এই পরীক্ষার ক্ষেত্রে প্রভাবসমূহের বর্গসমষ্টি নির্ণয় করার মোট প্রভাবকে বর্গ করে ঐ বর্গফলকে $2^n r$ দ্বারা ভাগ করা হয়। মোট বর্গসমষ্টি [TSS] এবং SS (ব্লক) বা SS(পুনরায়ন) স্বাভাবিক নিয়মেই নির্ণয় করা হয়। সুতরাং

$$SS(A) = \frac{[A]^2}{2^2 r}, \quad SS(B) = \frac{[B]^2}{2^2 r}, \quad SS(AB) = \frac{[AB]^2}{2^2 r}$$
$$= S_2 \qquad\qquad = S_3 \qquad\qquad = S_4$$

ধরা যাক i-তম ব্লকের মোট উৎপাদন হলো $B_i (i = 1, 2, \cdots, r)$ তাহলে,

$$SS(\text{ব্লক}) = \frac{\Sigma B_i^2}{2^2} - C.T, \quad C.T = \frac{G^2}{2^2 r}$$
$$= S_4$$

মোট বর্গসমষ্টি $TSS = \Sigma\Sigma\Sigma y^2_{ijl} - C.T$, এখানে y_{ijl} = i-তম ব্লকে A-এর j-তম স্তরের এবং B-এর l-তম স্তরের উৎপাদন।

এই বিশ্লেষণের ক্ষেত্রে ভেদাঙ্ক বিশ্লেষণ সারণি হলো নিম্নরূপ :

সারণি ৭.১৪ : 2^2-উপাদানী পরীক্ষার ক্ষেত্রে ভেদাঙ্ক বিশ্লেষণ সারণি।

ভেদের উৎস	স্বাধীনতার মাত্রা d.f.	বর্গসমষ্টি SS	গড় বর্গসমষ্টি $MS = \frac{SS}{d.f}$	F	$F_{.05}$
ব্লক	$r-1$	S_1	s_1	s_1/s_5	
A	1	S_2	s_2	s_2/s_5	
B	1	S_3	s_3	s_3/s_5	
AB	1	S_4	s_4	s_4/s_5	
বিচ্যুতি	$(2^2-1)(r-1)$	S_5	s_5		
মোট	$2^2 r - 1$				

এখানে $S_5 = TSS - S_1 - S_2 - S_3 - S_4$

এই বিশ্লেষণ হতে সিদ্ধান্ত পূর্ববর্তী অনুচ্ছেদসমূহের ক্ষেত্রে গৃহীত সিদ্ধান্তের পদ্ধতি অনুসরণ করেই নিতে হয়।

2^3-উপাদানী পরীক্ষা (2^3-Factorial experiment) : ধরা যাক A, B এবং C তিনটি উপাদানের প্রতিটির দুটি করে স্তর আছে। উপাদানত্রয়ের মোট যুগ্ম স্তর হলো আটটি এবং স্তরগুলি হলো

$$(1),\ a,\ b,\ ab,\ c,\ ac,\ bc,\ abc$$

এই পরীক্ষা হতে 7 স্বাধীনতার মাত্রাবিশিষ্ট মূল প্রভাব ও মিশ্রপ্রভাব নিরূপণ করা যায়। এগুলি হলো, A, B, C, AB, AC, BC এবং ABC। এগুলির মোট প্রভাব নির্ণয়ের সূত্র হলো যথাক্রমে

$$[A] = abc - bc + ab - b + ac - c + a - (1) = (a-1)(b+1)(c+1)$$
$$[B] = abc - ac + bc - c + ab - a + b - (1) = (a+1)(b-1)(c+1)$$
$$[C] = abc - ab + ac - a + bc - b + c - (1) = (a+1)(b+1)(c-1)$$
$$[AB] = abc + ab + c + (1) - a - b - ac - bc = (a-1)(b-1)(c+1)$$
$$[AC] = abc + ac + b + (1) - a - c - ab - bc = (a-1)(b+1)(c-1)$$
$$[BC] = abc + bc + a + (1) - b - c - ac - ab = (a+1)(b-1)(c-1)$$
$$[ABC] = abc + a + b + c - ab - ac - bc - (1) = (a-1)(b-1)(c-1)$$

পরীক্ষাটি r বার পুনরায়ন করা হলে গড় প্রভাবসমূহ নির্ণয়ের সূত্র হলো

$$A = \frac{[A]}{2^2 r},\quad B = \frac{[B]}{2^2 r},\quad C = \frac{[C]}{2^2 r},\ AB = \frac{[AB]}{2^2 r},\ AC = \frac{[AC]}{2^2 r},$$

$$BC = \frac{[BC]}{2^2 r},\ ABC = \frac{[ABC]}{2^2 r}$$

এই বিশ্লেষণের ক্ষেত্রে প্রভাবসমূহের বর্গসমষ্টি পাওয়ার সূত্র নিচে দেয়া হলো। ধরা যাক X হলো কোনো একটি প্রভাব। তাহলে X-এর বর্গসমষ্টি হলো

$$SS\ (X) = \frac{[X]^2}{2^3 r}$$

এই বর্গসমষ্টি সহজেই নির্ণয় করা হয় Yates Table তৈরি করে। নিচে Yate's Table তৈরির নিয়ম আলোচনা করা হলো।

(i) Yate's Table তৈরি করতে প্রথম স্তম্ভ ধারাবাহিকভাবে চর্যাসমূহ লেখা হয়

(ii) দ্বিতীয় স্তম্ভে বিভিন্ন পুনরায়ন হতে প্রাপ্ত চর্যাসমূহের মোট উৎপাদন লেখা হয় এবং ঐগুলিকে জোড়ায় জোড়ায় গুচ্ছ করা হয়।

(iii) তৃতীয় স্তম্ভে মোট উৎপাদনের জোড়ায় জোড়ায় যোগ ও বিয়োগ ফল নির্ণয় করা হয়। বিয়োগফল নির্ণয়ের সময় জোড়ার দ্বিতীয় রাশি হতে প্রথম রাশি বিয়োগ করা হয়।

(iv) পরবর্তী স্তম্ভসমূহের ফলাফল অনুরূপভাবে জোড়ায় জোড়ায় যোগ ও বিয়োগ ফল হতে পাওয়া যায়। চতুর্থ স্তম্ভের ফলাফল নির্ণয় করা হয় তৃতীয় স্তম্ভ থেকে, পঞ্চম স্তম্ভের ফলাফল পাওয়া যায় চতুর্থ স্তম্ভ থেকে এবং এ পদ্ধতিতে n-th স্তম্ভের

ফলাফল নির্ণয় করা হয় $(n-1)$ th স্তম্ভ হতে। 2^n-উপাদানী পরীক্ষার ক্ষেত্রে nth স্তম্ভের ফলাফল হলো মূল প্রভাব ও মিশ্রপ্রভাবের মোট। এই স্তম্ভের প্রথম মান হলো পরীক্ষার সর্বমোট উৎপাদন।

(v) প্রভাবের মোট নির্দেশকারী স্তম্ভের পরবর্তী স্তম্ভে মূল প্রভাব ও মিশ্রপ্রভাবগুলি লেখা হয়। এক্ষেত্রে যে ধারাবাহিকভাবে চর্যা লেখা হয় সে ধারাবাহিকভাবে প্রভাবগুলিকেও লেখা হয়।

(vi) মোট প্রভাবগুলিকে বর্গ করে ঐ বর্গফলকে $2^n r$ দ্বারা ভাগ করে বর্গসমষ্টি [SS] পাওয়া যায়। নিচে ৭.১৫ সারণিতে 2^3-উপাদানী পরীক্ষার ক্ষেত্রে Yate's table তৈরি করে দেখানো হলো।

সারণি ৭.১৫ : 2^3-উপাদানী পরীক্ষার ক্ষেত্রে বর্গসমষ্টি নির্ণয় করার Yate's table।

চর্যা	মোট উৎপাদান	যোগ-বিয়োগ করার কাজ প্রথম	দ্বিতীয়	তৃতীয় $=[\ ]$	প্রভাব-সমূহ	বর্গসমষ্টি $SS=\frac{[\ ]^2}{2^3r}$
(1)	x_1	$x_2+x_1=X_1$	$X_2+X_1=z_1$	$z_2+z_7=y_1$	G	$G^2/2^3r=C.T$
a	x_2	$x_4+x_3=X_2$	$X_4+X_3=z_2$	$z_4+z_3=y_2$	A	$y_2^2/2^3r=S_1$
b	x_3	$x_5+x_6=X_3$	$X_6+X_5=z_3$	$z_6+z_5=y_3$	B	$y_3^2/2^3r=S_2$
ab	x_4	$x_8+x_7=X_4$	$X_8+X_7=z_4$	$z_8+z_7=y_4$	AB	$y_4^2/2^3r=S_3$
c	x_5	$x_2-x_1=X_5$	$X_2-X_1=z_5$	$z_2-z_1=y_5$	B	$y_5^2/2^3r=S_4$
ac	x_6	$x_4-x_3=X_6$	$X_4-X_3=z_6$	$z_4-z_3=y_6$	AC	$y_6^2/2^3r=S_5$
bc	x_7	$x_6-x_5=X_7$	$X_6-X_5=z_7$	$z_6-z_5=y_7$	BC	$y_7^2/2^3r=S_6$
abc	x_8	$x_8-x_7=X_8$	$X_8-X_7=z_8$	$z_8-z_7=y_8$	ABC	$y_8^2/2^3r=S_7$

পরীক্ষাটি r ব্লকে পরিচালিত হলে ধরা যাক i-তম ব্লকের মোট উৎপাদন হলো $B_i[i = 1, 2, \ldots, r]$। তাহলে

$$SS(\text{ব্লক}) = \frac{\Sigma B_i^2}{2^3} - C.T = S$$

$$\text{মোট বর্গসমষ্টি, } TSS = \sum_i^r \sum_j^2 \sum_l^2 \sum_k^2 y_{ijlk}^2 - C.T$$

$$SS(\text{বিচ্যুতি}) = TSS - SS(\text{ব্লক}) - S_1 - S_2 - S_3 - S_4 - S_5 - S_6 - S_7$$

সারণি ৭.১৬ : 2^3-উপাদানী পরীক্ষার ক্ষেত্রে ভেদাঙ্ক বিশ্লেষণ সারণি।

ভেদের উৎস	স্বাধীনতার মাত্রা, d.f.	বর্গসমষ্টি SS	গড় বর্গসমষ্টি $MS = \frac{SS}{d.f.}$	F
ব্লক	$r-1$	S	s	s/s_8
A	1	S_1	s_1	s_1/s_8
B	1	S_2	s_2	s_2/s_8
C	1	S_3	s_3	s_3/s_8
AB	1	S_4	s_4	s_4/s_8
AC	1	S_5	s_5	s_5/s_8
BC	1	S_6	s_6	s_6/s_8
ABC	1	S_7	s_7	s_7/s_8
বিচ্যুতি	$(2^3-1)(r-1)$	S_8	s_8	
মোট	2^3r-1			

উদাহরণ ৭.৭

একটি পরীক্ষাগারে Wood louse-এর মৃত্যুতে Abate(A), Benlate(B) এবং Gramaxon(G)-এর কিরূপ প্রভাব আছে জানার জন্য পরীক্ষা পরিচালনা করা হয়। উক্ত পরীক্ষায় প্রতিটি পেস্টিসাইডের দুটি করে স্তর প্রয়োগ করা হয় (Abate-এর স্তর হলো 500 ppm এবং 1000 ppm ; Benlate-এর স্তর হলো 2000 ppm এবং 2500 ppm ; Gramaxon-এর স্তর হলো 500 ppm এবং 1000 ppm) পরীক্ষার

শুরুতে পেস্টিসাইডসমূহের যুগ্ম স্তরের যে কোনো একটিকে 30 Wood lice-এর উপর প্রয়োগ করা হয় এবং 7 দিন পর কয়টি Wood lice মারা গেছে তা লক্ষ্য করা হয়। পরীক্ষাটি 3 বার পুনরায়ন করা হয়। নিচে সারণি ৭.১৭-এ পেস্টিসাইডের যুগ্মস্তর অনুযায়ী মৃত Wood lice-এর সংখ্যা দেয়া হলো। উপাত্তটি বিশ্লেষণ করে

সারণি ৭.১৭ : তিন জাতীয় পেস্টিসাইডের যুগ্মস্তর প্রয়োগে মৃত Wood lice-এর সংখ্যা [y_{ijk}]।

পেস্টিসাইডের যুগ্ম স্তর	পুনরায়নের সংখ্যা মাধ্যমে মৃত Wood lice-এর			মোট
	1	2	3	
(1)	17	16	15	48
a	18	17	19	54
b	18	19	17	54
ab	20	19	22	61
g	28	26	26	80
ag	23	20	22	65
bg	25	24	23	72
agb	27	26	24	77
মোট B_1	176	167	168	

Wood lice মারার জন্য পেস্টিসাইডগুলির তাৎপর্যপূর্ণ প্রভাব আছে কিনা লক্ষ্য করা যাক।

এখানে $r=3$, মোট উপাত্ত সংখ্যা $= 2^3 \times 3 = 24$

$$G = 511, \quad C.T = \frac{G^2}{2^3 r} = \frac{(511)^2}{8 \times 3} = 10880.04$$

মোট বর্গসমষ্টি $TSS = \Sigma\Sigma\Sigma\Sigma y^2_{ijlk} - C.T$

$$= 11223 - 10880.04 = 342.96$$

$$SS(\text{ব্লক}) = \frac{\Sigma B_1^2}{2^3} - C.T = \frac{87089}{8} - 10880.04 = 6.08$$

$$SS(\text{বিচ্যুতি}) = TSS - SS(\text{ব্লক}) - SS(A) - SS(B) - SS(G) - SS(AB)$$
$$- SS(AG) - SS(BG) - SS(ABG)$$

$$= 342.96 - 6.08 - 22.04 - 12.04 - 247.04 - 18.37 - 0.37 - 3.37 - 15.04$$
$$= 18.61$$

সারণি ৭.১৮ : বর্গসমষ্টি নির্ণয়ের জন্য Yate's Table.

যুগ্ম স্তর	মোট উৎপাদন	যোগ-বিয়োগ করার কাজ 1	2	3 = [ ]	প্রভাবসমূহ G	বর্গসমষ্টি $SS = \frac{[\ ]^2}{2^3 r}$
(1)	48	102	217	511	G	10880.04
a	54	115	294	23	A	22.04
b	54	145	13	17	B	12.04
ab	61	149	10	21	AB	18.37
g	80	6	13	77	G	247.04
ga	65	7	4	−3	AG	0.37
bg	72	−15	1	−9	BG	3.37
abg	77	5	20	19	ABG	15.04

ভেদাঙ্ক বিশ্লেষণ সারণি

ভেদের উৎস	স্বাধীনতার মাত্রা d.f.	বর্গসমষ্টি SS	গড় বর্গসমষ্টি $MS = \frac{SS}{d.f.}$	F	$F_{.05}$
ব্লক	2	6.08	3.04	2.28	3.74
A	1	22.04	22.04	16.57	4.60
B	1	12.04	12.04	9.05	4.60
G	1	247.04	247.04	185.74	4.60
AB	1	18.37	18.37	13.81	4.60
AG	1	0.37	0.37	0.28	4.60
BG	1	3.37	3.37	2.53	4.60
ABG	1	15.04	15.04	11.31	4.60
বিচ্যুতি	14	18.61	1.33		
মোট	23				

লক্ষ্য করা যাচ্ছে যে, Abate, Benlate এবং Gramaxon-এর প্রতিটির Wood louse মারার তাৎপর্যপূর্ণ ক্ষমতা আছে। এই তিনটি পেস্টিসাইডের মিশ্রণও Wood louse-কে ধ্বংস করে। তবে Abate ও Gramaxon এবং Benlate ও Gramaxon মিশিয়ে কার্যকরী প্রভাব লক্ষ্য করা যায়নি।

উপরে 2^3-উপাদানী পরীক্ষার বিশ্লেষণ পদ্ধতি আলোচনা করা হয়েছে। লক্ষ্য করা যাচ্ছে যে, এই পরীক্ষার ক্ষেত্রে মোট $(2^3-1)=7$ প্রভাব ও মিশ্রপ্রভাব (effects and interactions) নিরূপণ করা হয়েছে। প্রভাবগুলির মোট (total effect) নির্ণয় করা হয়েছে Yate's পদ্ধতি অনুযায়ী। আগেই আলোচনা করা হয়েছে যে, Yate's পদ্ধতি n-এর যে কোনো মানের জন্য 2^n-উপাদানী পরীক্ষার প্রভাবগুলির মোট নির্ণয়ের জন্য প্রয়োগ করা যায়। সেক্ষেত্রে যোগ-বিয়োগের n-তম ধাপ হলো প্রভাবগুলির মোট। ধরা যাক এই পরীক্ষার ক্ষেত্রে X হলো কোনো একটি প্রভাব। তাহলে X-এর বর্গসমষ্টি (এখানে X হতে পারে A বা B বা C বা AB বা AC বা ABC বা ABCD বা ABCDE ··· ইত্যাদি)

$$SS(X)=\frac{[X]^2}{2^n r}$$

এখানে r হলো 2^n-উপাদানী পরীক্ষার পুনরায়নের সংখ্যা।

এই পরীক্ষার ক্ষেত্রে মোট (2^n-1) প্রভাব ও মিশ্রপ্রভাব নিরূপণ করা যায়। যে কোনো প্রভাবের মোট হলো চর্যাসমূহ বা যুগ্ম স্তরসমূহের রৈখিক সংযোগ (linear combination)। এই রৈখিক সংযোগের ক্ষেত্রে 2^{n-1} যুগ্মস্তরে যোগ চিহ্ন [+Ve] এবং 2^{n-1} যুগ্ম স্তরে বিয়োগ চিহ্ন [−Ve] বসে। যে কোনো প্রভাব X-এর ক্ষেত্রে গড় প্রভাব (mean effect) নির্ণয়ের সূত্র হলো

$$X=\frac{[X]}{2^{n-1}r}$$

যেমন, উদাহরণ ৭.৭-এর ক্ষেত্রে Abate-এর গড় প্রভাব হলো

$$A=\frac{[A]}{2^{3-1}r}=\frac{23}{4\times 3}=1.92$$

আবার Abate, Benlate এবং Gramaxon-এর তিন উপাদানী মিশ্রপ্রভাবের ক্ষেত্রে গড় প্রভাব হলো

$$ABG=\frac{[ABG]}{2^{3-1}r}=\frac{19}{4\times 3}=1.58$$

সংমিশ্রিত উপাদানী পরীক্ষা (Confounded factorial experiment) : উপরে সংক্ষিপ্তভাবে 2^n-উপাদানী পরীক্ষার বিশ্লেষণ পদ্ধতি আলোচনা করা হয়েছে। এই পরীক্ষার একটি অসুবিধা হলো n-এর মান বড় হলে চর্যা সংখ্যা বড় হয়। যেমন, $n=6$ হলে $2^6=64$ । এই 64 চর্যা নিয়ে দৈবায়িত ব্লকে পরীক্ষা পরিচালনা করার ক্ষেত্রে বাস্তব অসুবিধার সম্মুখীন হওয়ার সম্ভাবনা থাকে। কারণ 64 চর্যার জন্য এক জাতীয় 64 খণ্ড পাওয়া সহজসাধ্য নয়। যেমন, Abate, Benlate, Gramaxon, Sumicidin, Pesguard এবং Glyphosate-এর প্রতিটির দুটি ঘনত্ব (concentration) বিশিষ্ট তরল কোনো পোকামাকড় মারার জন্য ব্যবহার করা হলে এবং এর যে কোনো একটি পেস্টিসাইডের প্রভাব জানতে চাওয়া হলে একই জাতীয়, একই ওজনের এবং একই দৈর্ঘ্যের 64 পোকামাকড় প্রয়োজন হবে। বাস্তবে অনুরূপ বৈশিষ্ট্যবিশিষ্ট 64 পোকামাকড় পাওয়া সহজ ব্যাপার নয়। এরূপক্ষেত্রে 64-এর অর্ধেক 32 বা এক চতুর্থাংশ 16টি অনুরূপ পোকামাকড় পাওয়া সহজসাধ্য হতে পারে। তাই 64 চর্যা প্রয়োগ করার জন্য একই জাতীয় 32টি করে বা 16টি করে পোকামাকড় সংগ্রহ করে ঐগুলির উপর চর্যা প্রয়োগ করা যেতে পারে। এ জাতীয় উপাদানী পরীক্ষাকে সংমিশ্রিত উপাদানী পরীক্ষা বলা হয়।

সংমিশ্রিত উপাদানী পরীক্ষার উদ্দেশ্য হলো ব্লকে খণ্ডের সংখ্যা কমানো যেন যে কোনো ব্লকের খণ্ডগুলি সমজাতীয় হয়। আগেই উল্লেখ করা হয়েছে যে, যে কোনো প্রভাবের মোট নির্ণয়ের জন্য রৈখিক সমাবেশে (linear combination) অর্ধেক চর্যা যোগ চিহ্ন [+Ve] বিশিষ্ট এবং অর্ধেক চর্যা বিয়োগ চিহ্ন [−Ve] বিশিষ্ট। তাই সংমিশ্রিত উপাদানী পরীক্ষার কোনো প্রভাবের জন্য যোগ চিহ্নবিশিষ্ট চর্যাগুলিকে এক ব্লকে এবং বিয়োগ চিহ্নবিশিষ্ট চর্যাগুলিকে অন্য ব্লকে বণ্টন করা হলে ব্লকের আকার ছোট হবে। ধরা যাক 2^3-উপাদানী পরীক্ষার ক্ষেত্রে ব্লকের আকার 8-এর পরিবর্তে 4 করা প্রয়োজন। এই পরীক্ষার ABC প্রভাবের জন্য যোগ চিহ্ন ও বিয়োগ চিহ্নবিশিষ্ট চর্যা নিম্নরূপ :

	(1)	a	b	ab	c	ac	bc	abc
ABC	−	+	+	−	+	−	−	+

এখন কোনো পুনরায়নে এক ব্লকের পরিবর্তে নিম্নরূপভাবে দুই ব্লকে 8 চর্যা বণ্টন করা যায়।

						মোট
পুনরায়ন-1	ব্লক-1	a	b	c	abc	B_1
	ব্লক-2	(1)	ab	ac	bc	B_2

উপরিউক্ত বণ্টনের ফলে ব্লকের আকার ছোট হলো। এতে করে পরীক্ষার দক্ষতা (efficiency) বাড়ে।

উপরিউক্ত বণ্টন পদ্ধতির ফলে লক্ষ্য করা যাচ্ছে যে

$[ABC] = (abc + a + b + c) - ((1) + ab + ac + bc)$ [চিহ্ন সারণি অনুযায়ী]

$= B_1 - B_2 =$ ব্লকের পার্থক্য,

অর্থাৎ ABC মিশ্র প্রভাবটি ব্লকের পার্থক্যের অনুরূপ। এটি গাণিতিক সত্য। তাই ABC প্রভাব সঠিকভাবে জানা যায় না। কারণ যেটি ABC প্রভাব সেটিই ব্লকের প্রভাব। সে কারণে ABC প্রভাব নিরূপণ করা হয় না। তাই কোনো উপাদানী পরীক্ষার ক্ষেত্রে কোনো উচ্চ অর্ডারের মিশ্রপ্রভাব জানা নিষ্প্রয়োজন হলে বা উচ্চ অর্ডারের প্রভাব তাৎপর্যহীন হলে ঐ মিশ্রপ্রভাবের জন্য যোগ চিহ্নবিশিষ্ট চর্যাসমূহ এক ব্লকে এবং বিয়োগ চিহ্নবিশিষ্ট চর্যাসমূহ অন্য ব্লকে বণ্টন করা হলে মিশ্র প্রভাবটি ব্লকের সাথে সংমিশ্রিত হবে এবং তার প্রভাব নিরূপণ করা যাবে না। উপরিউক্ত উদাহরণে ABC প্রভাব ব্লকের সাথে সংমিশ্রিত।

সংমিশ্রিত উপাদানী পরীক্ষার ক্ষেত্রে সংমিশ্রণ এমনভাবে করা হয় যেন কোনো মূল প্রভাব সংমিশ্রিত না হয়। এই উপাদানী পরীক্ষার বিশ্লেষণ পদ্ধতিও স্বাভাবিক উপাদানী পরীক্ষার বিশ্লেষণ পদ্ধতির অনুরূপ। শুধু পার্থক্য হলো সংমিশ্রিত মিশ্র প্রভাবের বর্গসমষ্টি নির্ণয় করা হয় না। এখানে বিশ্লেষণ পদ্ধতি তাই ভিন্নভাবে আলোচনা করা হলো না। 2^3-উপাদানী পরীক্ষার ক্ষেত্রে ABC সংমিশ্রিত হলে ভেদাঙ্ক বিশ্লেষণ সারণি হবে নিম্নরূপ :

সারণি ৭.১৯ : সংমিশ্রিত উপাদানী পরীক্ষার ক্ষেত্রে ভেদাঙ্ক বিশ্লেষণ সারণি।

ভেদের উৎস	স্বাধীনতার মাত্রা, d.f.
পুনরায়নে ব্লক	r
A	1
B	1
C	1
AB	1
AC	1
BC	1
বিচ্যুতি	$(2^3 - 1)(r - 1)$
মোট	$2^3 r - 1$

এখানে r হলো পুনরায়নের সংখ্যা।

$$SS\,[\text{পুনরায়নে ব্লক}] = \sum_{i}^{r}\left[\frac{B_{1i}^2 + B_{2i}^2}{2^2} - \frac{(B_{1i} + B_{2i})^2}{2^3}\right]$$

B_{1i} = i-তম পুনরায়নে ব্লক-1 এর মোট

B_{2i} = i-তম পুনরায়নে ব্লক-2 এর মোট।

৭.১০ খণ্ডিত-প্লট নকশা (Split-Plot Design)

মনে করি **A** একটি উপাদান যার p স্তর আছে এবং **B** একটি উপাদান যার q স্তর আছে। **A** উপাদানটি এমন যেখানে তার এক একটি স্তরকে বণ্টন (allocate) করার জন্য বড় আকারের পরীক্ষণ এককের প্রয়োজন। এই বড় আকারের পরীক্ষণ একককে q সংখ্যক উপ-খণ্ডে বিভক্ত করা যেতে পারে। তাহলে একবার পুনরায়ন করার জন্য ব্লকে যতোগুলি পরীক্ষণ একক থাকবে ঐগুলিকে p ভাগে বিভক্ত করে A-এর p স্তর দৈব পদ্ধতিতে বণ্টন করতে হবে। তারপর এক একভাগের পরীক্ষণ এককসমূহকে q উপ-খণ্ডে বিভক্ত করে B-এর q স্তরকে দৈব পদ্ধতিতে বণ্টন করতে হবে। এভাবে A ও B-এর স্তরসমূহকে দৈব পদ্ধতিতে যথাক্রমে খণ্ডে এবং উপ-খণ্ডে বণ্টন করা r বার পুনরায়ন করা হলে প্রাপ্ত নকশাকে খণ্ডিত-প্লট নকশা বলা হয়। এই নকশার ক্ষেত্রে প্রাপ্ত উপাত্তকে নিম্নরূপভাবে সাজিয়ে লেখা যায় :

পুনরায়ন	A-এর স্তরসমূহ	B-এর স্তরসমূহ						মোট
		B_1	B_2	…	B_j	…	B_q	$y_{ij\cdot}$
	A_1	y_{111}	y_{112}	…	y_{11j}	…	y_{11q}	$y_{11\cdot}$
	A_2	y_{121}	y_{122}	…	y_{12j}	…	y_{12q}	$y_{12\cdot}$
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1	A_j	y_{1j1}	y_{1j1}	…	y_{1jj}	…	y_{1jq}	$y_{1j\cdot}$
	…	…	…	…	…	…	…	…
	A_p	y_{1p1}	y_{1p2}	…	y_{1pj}	…	y_{1pq}.	$y_{1p\cdot}$
	A_1	y_{211}	y_{212}	…	y_{21j}	…	y_{21q}	$y_{21\cdot}$
	A_2	y_{221}	y_{222}	…	y_{22j}	…	y_{22q}	$y_{22\cdot}$
	…	…	…	…	…	…	…	…
2	A_j	y_{2j1}	y_{2j2}	…	y_{2jj}	…	y_{2jq}	$y_{2j\cdot}$

পুনরায়ন	A-এর স্তরসমূহ	B-এর স্তরসমূহ						মোট
	…	…	…	…	…	…	…	…
	A_p	y_{2p1}	y_{2p2}	…	y_{2pl}	…	y_{2pq}	$y_{2p\cdot}$
	…	…	…	…	…	…	…	…
	A_1	y_{r11}	y_{r12}	…	y_{r1l}	…	y_{r1q}	$y_{r1\cdot}$
	A_2	y_{r21}	y_{r22}	…	y_{r2l}	…	y_{r2q}	$y_{r2\cdot}$
r	⋮	…	…	…	…	…	…	…
	A_j	y_{rj1}	y_{rj2}	…	y_{rjl}	…	y_{rjq}	$y_{rj\cdot}$
	…	…	…	…	…	…	…	…
	A_p	y_{rp1}	y_{rp2}	…	y_{rpl}	…	y_{rpq}	$y_{rp\cdot}$
মোট	$y_{\cdot\cdot l}$	$y_{\cdot\cdot 1}$	$y_{\cdot\cdot 2}$	…	$y_{\cdot\cdot l}$	…	$y_{\cdot\cdot q}$	$y_{\cdot\cdot\cdot}$

খণ্ডিত-প্লট নকশার ব্যবহার কৃষি ক্ষেত্রেই বেশি হয়ে থাকে। ধরা যাক ইউরিয়া সারের কয়েকটি স্তর, যেমন : $N_1 = 30$ kg/ha, $N_2 = 60$ kg/ha, $N_3 = 90$ kg/ha এবং $N_4 = 120$ kg/ha প্রয়োগ করে বিভিন্ন স্তরের পানি সেচের মাধ্যমে, যেমন— I_1 = জমি ভিজা থাকবে, I_2 = গাছের গোড়ায় এক ইঞ্চি পরিমাণ পানি থাকবে এবং I_3 = গাছের গোড়ায় দুই ইঞ্চি পরিমাণ পানি রেখে ইরি ধানের চাষ করা হবে। এক্ষেত্রে পানি সেচের জন্য বড় আকারের জমি খণ্ড প্রয়োজন। পরীক্ষা এক-বার পুনরায়ন করার জন্য প্রাপ্ত মোট জমিকে তিন বড় জমি খণ্ডে বিভক্ত করে এক এক খণ্ডে এক স্তরের পানি সেচ দৈবায়িতভাবে করা যায়। আবার প্রতিটি বড় জমি খণ্ডকে চারটি উপ-খণ্ডে বিভক্ত করে ইউরিয়া সারের চারটি স্তরকে দৈব পদ্ধতিতে চার উপ-খণ্ডে বণ্টন করা যায়। এই পরীক্ষা কয়েকবার পুনরায়ন করা হলে প্রাপ্ত নকশাকে খণ্ডিত-প্লট নকশা বলা হয়।

খণ্ডিত-প্লট নকশা হতে প্রাপ্ত উপাত্তের জন্য প্রতিকৃতি হলো

$$y_{ijl} = \mu + \alpha_i + \beta_j + \gamma_l + (\beta\gamma)_{jl} + e_{ijl} \qquad (৭.১০.১)$$

$$i = 1, 2, \cdots, r\,,\; j = 1, 2, \cdots, p\,,\; l = 1, 2, \cdots, q$$

এখানে y_{ijl} = i-তম পুনরায়নে A-এর j-তম স্তরের প্রাসঙ্গিক B-এর l-তম স্তরের উৎপাদন,

μ = সাধারণ গড়

α_i = i-তম পুনরায়নের প্রভাব,

β_j = A-এর j-তম স্তরের প্রভাব,

γ_l = B-এর l-তম স্তরের প্রভাব,

$(\beta\gamma)_{jl}$ = A-এর j-তম স্তরের সাথে B-এর l-তম স্তরের মিশ্রপ্রভাব,

e_{ijl} = দৈব উপাদান।

এখানে $E(e_{ijl}) = 0$

$$E[e_{ijl}, e_{i'j'l'}] = \sigma^2, \text{ যখন } i = i', j = j', l = l'$$
$$= \rho\sigma^2, \text{ যখন } i = i', j = j', l \neq l'$$
$$= 0, \text{ অন্যথায়}$$

খণ্ডিত-প্লট নকশায় চর্যার বণ্টন ব্যবস্থা দুই ধাপে দৈবায়িত পদ্ধতিতে করা হয় বলে এই নকশা হতে প্রাপ্ত উপাত্তের বিশ্লেষণও দুই ধাপে হয়ে থাকে। একটি বিশ্লেষণকে বলা হয় প্রধান খণ্ড বিশ্লেষণ এবং অপরটি হলো উপ-খণ্ড বিশ্লেষণ। প্রতি বিশ্লেষণ হতে বিচ্যুতির বর্গসমষ্টি পাওয়া যায়। প্রধান খণ্ড বিশ্লেষণ হতে প্রাপ্ত বিচ্যুতির বর্গসমষ্টিকে বলা হয় বিচ্যুতির বর্গসমষ্টি-1 এবং উপ-খণ্ড বিশ্লেষণের ক্ষেত্রে বিচ্যুতির বর্গসমষ্টি হলো বিচ্যুতির বর্গসমষ্টি-2। নিচে বিভিন্ন বর্গসমষ্টি নির্ণয়ের সূত্র দেয়া হলো :

$$SS(\text{পুনরায়ন}) = S_1 = \frac{\Sigma y_{i..}^2}{pq} - C.T.\ C.T = \frac{G^2}{pqr}$$

$$SS(A) = S_2 = \frac{\Sigma y_{.j.}^2}{rq} - C.T,\ SS(B) = S_4 = \frac{\Sigma y_{..l}^2}{pr} - C.T$$

$$SS(\text{বিচ্যুতি-1}) = S_3 = \frac{\Sigma\Sigma y_{ij.}^2}{q} - C.T - SS(\text{পুনরায়ন}) - SS(A)$$

$$SS(AB) = S_5 = \frac{\Sigma\Sigma y_{.jl}^2}{r} - C.T - SS(A) - SS(B)$$

$$SS(\text{বিচ্যুতি-2}) = S_6 = TSS - SS(\text{পুনরায়ন}) - SS(A)$$
$$- SS(\text{বিচ্যুতি-1}) - SS(B) - SS(AB)$$

এখানে $TSS = \Sigma\Sigma\Sigma y_{ijl}^2 - C.T$

সারণি ৭.২০ : খণ্ডিত-প্লট নকশার ক্ষেত্রে ভেদাঙ্ক বিশ্লেষণ সারণি।

ভেদের উৎস	স্বাধীনতার মাত্রা, d.f	বর্গসমষ্টি SS	গড় বর্গসমষ্টি $MS = \frac{SS}{d.f}$	F	$F_{.05}$
পুনরায়ন	$r-1$	S_1	s_1	s_1/s_3	
A	$p-1$	S_2	s_2	s_2/s_3	
বিচ্যুতি-1	$(r-1)(p-1)$	S_3	s_3	—	
B	$q-1$	S_4	s_4	s_4/s_6	
AB	$(p-1)(q-1)$	S_5	s_5	s_5/s_6	
বিচ্যুতি-2	$p(r-1)(q-1)$	S_6	s_6		
মোট	$pqr-1$				

উদাহরণ ৭.৮

একটি পরীক্ষাগারে Milux Rusticus-এর মৃত্যুর উপর Pomex, Supracide, Sumicidin এবং Glyphosate-এর কি প্রভাব আছে জানার জন্য প্রতি দশটি Rusticus-এর উপর দৈব পদ্ধতিতে উক্ত পেস্টিসাইডগুলি প্রয়োগ করা হয়েছে। পেস্টিসাইড প্রয়োগ করার আগেই Slug গুলিকে 18^0C, 20^0C এবং 22^0C তাপমাত্রায় রাখা হয়েছিল। পরীক্ষাটি পরিচালনা করার জন্য খণ্ডিত প্লট নকশা প্রয়োগ করা হয়েছে, যেখানে প্রধান খণ্ডের চর্যা হলো তাপমাত্রার স্তর এবং উপ-খণ্ডের চর্যা হলো পেস্টিসাইডের স্তর। প্রতি পেস্টিসাইড প্রয়োগে মৃত Rusticus-এর সংখ্যা সারণি ৭.২১-এ দেয়া হলো। উপাত্ত বিশ্লেষণ করে পেস্টিসাইড এবং তাপমাত্রার প্রভাব। তাৎপর্যপূর্ণ কিনা যাচাই করা যাক।

সারণি ৭.২১ : বিভিন্ন তাপমাত্রায় পেস্টিসাইড প্রয়োগে মৃত Rusticus-এর সংখ্যা $[y_{ijl}]$

পুনরায়ন	তাপমাত্রা	পেস্টিসাইড Pomex	Supracide	Sumicidin	glyphosate	মোট $y_{ij.}$
	18^0	4	7	2	5	18
1	20^0	6	9	2	5	22
	22^0	5	8	3	8	24
	18^0	5	5	3	6	19

পুনরায়ন	তাপমাত্রা	পেস্টিসাইড				
		Pomex	Supracide	Sumicidin	glyphosate	মোট $y_{ij\cdot}$
2	20^0	5	7	2	5	19
	22^0	6	8	3	8	25
	18^0	4	6	2	6	18
3	20^0	5	8	3	6	22
	22^0	5	8	3	9	25
মোট	$y_{\cdot\cdot l}$	45	66	23	58	192

এখানে $r = 3, \quad p = 3, \quad q = 4, \quad G = 192$

$$C.T = \frac{G^2}{pqr} = \frac{(192)^2}{3 \times 4 \times 3} = 1024$$

$$TSS = \Sigma\Sigma\Sigma y^2_{ijl} - C.T = 1178 - 1024 = 154$$

বিভিন্ন পুনরায়নে এবং বিভিন্ন তাপমাত্রায় মৃত Rusticus এর সংখ্যা $[y_{ij\cdot}]$

পুনরায়ন	তাপমাত্রা			মোট $y_{i\cdot\cdot}$
	18°	20°	22°	
1	18	22	24	64
2	19	19	25	63
3	18	22	25	65
মোট $y_{\cdot j\cdot}$	55	63	74	

$$SS(\text{পুনরায়ন}) = \frac{\Sigma y^2_{i\cdot\cdot}}{pq} - C.T = \frac{12290}{3 \times 4} - 1024 = 0.17$$

$$SS(\text{তাপমাত্রা}) = \frac{\Sigma y^2_{\cdot j\cdot}}{rq} C.T = \frac{12470}{3 \times 4} - 1024 = 15.17$$

$$SS(\text{বিচ্যুতি-1}) = \frac{\Sigma\Sigma y^2_{ij\cdot}}{q} - C.T - SS(\text{পুনরায়ন}) - SS(\text{তাপমাত্রা})$$

$$= \frac{4164}{4} - 1024 - 0.17 - 15.17 = 1.66$$

তাপমাত্রা এবং পেস্টিসাইড অনুযায়ী মৃত্যু Rusticus এর সংখ্যা $[y_{.ji})$

তাপমাত্রা	পেস্টিসাইড				
	Pomex	Supracide	Sumicidin	Glyphosate	গড় $\bar{y}_{.j}$
18^0	13	18	7	17	4.58
20^0	16	24	7	16	5.25
22^0	16	24	9	25	6.17
মোট $\bar{y}_{..i}$	45	66	23	58	
গড় $y_{..i}$	5.00	7.33	2.55	6.44	

$$SS(\text{পেস্টিসাইড}) = \frac{\Sigma y^2_{..i}}{rp} - C.T = \frac{10274}{3 \times 3} - 1024 = 117.55$$

$$SS(\text{তাপমাত্রা} \times \text{পেস্টিসাইড}) = \frac{\Sigma\Sigma y^2_{.ji}}{r} - C.T - SS\,(\text{তাপমাত্রা}) - SS(\text{পেস্টিসাইড})$$

$$= \frac{3506}{3} - 1024 - 15.17 - 117.55$$

$$= 11.95$$

$$SS(\text{বিচ্যুতি-2}) = TSS - SS\,(\text{পুনরায়ন}) - SS\,(\text{তাপমাত্রা}) - SS\,(\text{বিচ্যুতি} - 1) - SS\,(\text{পেস্টিসাইড}) - SS\,(\text{তাপমাত্রা} \times \text{পেস্টিসাইড})$$

$$= 154 - 0.17 - 15.17 - 1.66 - 117.55 - 11.95$$

$$= 7.5$$

ভেদাঙ্ক বিশ্লেষণ সারণি

ভেদের উৎস	স্বাধীনতার মাত্রা d.f.	বর্গসমষ্টি SS	গড় বর্গসমষ্টি $MS = \frac{SS}{d.f.}$	F	$F_{.05}$
পুনরায়ন	2	0.17	0.085	0.20	6.94
তাপমাত্রা	2	15.17	7.585	18.28	6.94
বিচ্যুতি-1	4	1.66	0.415	—	
পেস্টিসাইড	3	117.55	39.183	93.96	3.16
তাপমাত্রা × পেস্টিসাইড	6	11.95	1.992	4.78	2.66
বিচ্যুতি-2	18	7.50	0.417		
মোট	35				

বিশ্লেষিত ফলাফল হতে লক্ষ্য করা যাচ্ছে যে, তাপমাত্রার পরিবর্তনের কারণে Rusticus-এর মৃত্যুর মধ্যে তাৎপর্যপূর্ণ পার্থক্য পরিলক্ষিত হচ্ছে। অনুরূপ তাৎপর্যপূর্ণ পার্থক্য পেস্টিসাইড-১ এর পরিবর্তনের কারণেও লক্ষণীয়। এখন DMRT-এর মাধ্যমে লক্ষ্য করা যাক Rusticus-এর গড় মৃত্যুর পরিপ্রেক্ষিতে কোনো কোনো তাপমাত্রা একই জাতীয়। এখানে

$$R_k = r_{0.05,k,4}\sqrt{\frac{\text{MS(বিচ্যুতি-1)}}{rq}}\ , \ k=2,\ 3,\ 4$$

$$R_2 = r_{0.05,2,4}\sqrt{\frac{\text{MS (বিচ্যুতি-1)}}{rq}} = 3.93\sqrt{\frac{0.415}{3\times 4}} = 0.73$$

$$R_3 = r_{0.05,3,4}\sqrt{\frac{\text{MS (বিচ্যুতি-1)}}{rq}} = 4.01\sqrt{\frac{0.415}{3\times 4}} = 0.74$$

তাপমাত্রা অনুযায়ী মৃত Rusticus-এর গড়কে মানের ক্রম অনুসারে সাজিয়ে পাওয়া যায় :

$\bar{y}_{.1.} = 4.58,\quad \bar{y}_{.2.} = 5.25,\quad \bar{y}_{.3.} = 6.17$

$\bar{y}_{.3.} - \bar{y}_{.1.} = 15.9 > R_3,\quad \bar{y}_{.2.} - \bar{y}_{.1.} = 0.67 < R_2$

$\bar{y}_{.3.} - \bar{y}_{.2.} = 92 > R_2$

লক্ষ্য করা যাচ্ছে যে, 18°C এবং 20°C-এ মৃত্য Rusticus-এর গড় অনুরূপ।

পেস্টিসাইড-এর কারণে মৃত Rusticus-এর গড়ের পার্থক্য লক্ষ্য করার জন্য DMRT হলো

$$R_k = r_{.05,k,18}\sqrt{\frac{\text{MS (বিচ্যুতি-2)}}{rp}}\ , \ k=2,3,4$$

$$R_4 = r_{0.05,4,18}\sqrt{\frac{\text{MS (বিচ্যুতি-2)}}{rp}} = 3.21\sqrt{\frac{0.417}{3\times 3}} = 0.69$$

$$R_3 = r_{0.05,3,18}\sqrt{\frac{\text{MS (বিচ্যুতি-2)}}{rp}} = 3.12\sqrt{\frac{0.417}{3\times 3}} = 0.67$$

$$R_2 = r_{0.05,2,18}\sqrt{\frac{\text{MS (বিচ্যুতি-2)}}{rp}} = 2.97\sqrt{\frac{0.417}{3\times 3}} = 0.64$$

পেস্টিসাইড অনুযায়ী মৃত্যু Rusticus-এর গড়কে মানের ক্রম অনুসারে সাজিয়ে পাওয়া যায় :

Sumicidin	Pomex	Glyphosate	Supracide
$\bar{y}_{..3} = 2.55$	$\bar{y}_{..1} = 5.00$	$\bar{y}_{..4} = 6.44$	$\bar{y}_{..2} = 7.33$

$\bar{y}_{..2} - \bar{y}_{..3} = 4.78 > R_4$, $\quad \bar{y}_{..2} - \bar{y}_{..1} = 2.33 > R_3$

$\bar{y}_{..4} - \bar{y}_{..3} = 3.89 > R_3$, $\quad \bar{y}_{..1} - \bar{y}_{..3} = 2.45 > R_2$

$\bar{y}_{..4} - \bar{y}_{..1} = 1.44 > R_2$, $\quad \bar{y}_{..2} - \bar{y}_{..4} = 0.89 > R_2$

লক্ষণীয় বিষয় হলো প্রতিটি পেস্টিসাইডেরই Rusticus মারার ক্ষমতা তাৎপর্যপূর্ণভাবে ভিন্ন। তবে Supracide-এর ক্ষমতা সবচেয়ে বেশি।

অষ্টম অধ্যায়

নমুনায়ন পদ্ধতি
(Sampling Technique)

৮.১ সূচনা

প্রথম অধ্যায়ে (প্রথম খণ্ড) উল্লেখ করা হয়েছে যে, পরিসংখ্যানিক ব্যাখ্যাদানের জন্য গণসমষ্টির উপাত্তের পরিবর্তে তার প্রতিনিধিত্বমূলক অংশবিশেষের ভিত্তিতে সংখ্যাভিত্তিক বৈশিষ্ট্য পর্যালোচনা করা হয়। এখানে গণসমষ্টি (population) হলো যে একক-সমূহের বৈশিষ্ট্য পর্যালোচনা করা হবে ঐগুলির একটি সমাহার (aggregation or collection)। উদাহরণ হিসেবে বলা যায় সুন্দরবনে কতো সুন্দরী বৃক্ষ আছে তা নিরূপণ করতে হবে। সেক্ষেত্রে সুন্দরবনের সকল সুন্দরী বৃক্ষ গণসমষ্টি একক হিসেবে চিহ্নিত হবে। পদ্মা নদীতে কোনো এক বছরে কতো ইলিশ মাছ উৎপাদন হবে তা নিরূপণ করতে হলে পদ্মা নদীর সব ইলিশ মাছ গণসমষ্টি একক হবে। সাভার ডেইরী ফার্মে গাভী প্রতি কতো কেজি দুধ উৎপাদিত হয় তা নিরূপণ করার জন্য ঐ ফার্মের সকল গাভী গণসমষ্টি একক হিসেবে চিহ্নিত হবে।

উপরিউক্ত গণসমষ্টিসমূহের বৈশিষ্ট্য পর্যালোচনা করার জন্য উপাত্ত সংগ্রহ করা হয়। উপাত্ত সংগ্রহের দুটি পদ্ধতি আছে। যেমন: (১) শুমারী (census), (২) নমুনা জরিপ (sample survey)। গণসমষ্টির সকল একক হতে উপাত্ত সংগ্রহ করা হলে তাকে বলা হয় শুমারী। যেমন, আদম শুমারী (population census)। এই শুমারীতে কোনো একটি দেশের প্রতিটি নাগরিকের বয়স, শিক্ষা, পেশা ইত্যাদির তথ্য সংগ্রহ করা হয়। এ জাতীয় শুমারী কৃষি, গৃহপালিত পশু, গৃহায়ণ ইত্যাদি ক্ষেত্রেও করা হয়। নমুনা জরিপের ক্ষেত্রে গণসমষ্টির সকল একক হতে উপাত্ত সংগ্রহ করা হয় না। বরং চিহ্নিত গণসমষ্টির প্রতিনিধিত্বমূলক কিছু একক চয়ন (select) করা হয় এবং ঐ চয়ন করা এককগুলি হতে উপাত্ত সংগ্রহ করা হয়। এখানে চয়ন করা এককগুলিকে বলা হয় নমুনা তথ্যমান (sample observations) বা নমুনা একক (sample units) এবং যে পদ্ধতিতে ঐ এককগুলিকে চয়ন করা হয় তাকে বলা হয় নমুনায়ন পদ্ধতি (sampling technique)। এই নমুনা জরিপের বা নমুনায়ন

পদ্ধতির উদ্দেশ্য হলো অল্প সময়ে অল্প খরচে, দক্ষ জনশক্তি ব্যবহার করে গণসমষ্টি পরামানের (parameter) সঠিক এবং যথাযথ নিরূপক নির্ণয় করা। বর্তমান অধ্যায়ে নমুনায়ন পদ্ধতি সংক্ষিপ্তভাবে আলোচনা করা হবে।

৮.২ নমুনায়নের সাথে জড়িত কিছু শব্দাবলি (Terms Related to Sampling)

নমুনায়ন সম্পর্কিত আলোচনার প্রথমেই উল্লেখ করতে হয় গণসমষ্টি সম্পর্কে। পূর্বোক্ত অনুচ্ছেদে গণসমষ্টির সংজ্ঞা ব্যাখ্যা করা হয়েছে। গণসমষ্টি দুই প্রকার। যেমন: (১) সসীম গণসমষ্টি (finite population), (২) অসীম গণসমষ্টি (infinite population)।

সসীম গণসমষ্টি (Finite population): যে গণসমষ্টির এককের একটি নির্দিষ্ট সংখ্যা আছে এবং ঐগুলির একটি নির্দিষ্ট অবস্থান আছে তাকে সসীম গণসমষ্টি বলা হয়। যেমন, কোনো এক বছরে একটি হাসপাতালের বহির্বিভাগে চিকিৎসার জন্য আসা মোট রোগী গণসমষ্টিভুক্ত হবে যদি ঐ রোগীদের সম্পর্কে কোনো তথ্য জরিপের মাধ্যমে জানতে চাওয়া হয়।

অসীম গণসমষ্টি (Infinite population): যে গণসমষ্টির এককের কোনো নির্দিষ্ট অবস্থান চিহ্নিত করা যায় না এবং এককের কোনো নির্দিষ্ট সংখ্যা জানা যায় না তাকে অসীম গণসমষ্টি বলা হয়। যেমন, কোনো সমুদ্রের মাছের বৈশিষ্ট্য জানতে চাওয়া হলে ঐ মাছের গণসমষ্টি হবে অসীম।

নমুনায়ন একক (Sampling unit): গণসমষ্টির প্রতিটি একককে নমুনায়ন একক বলা হয়। কারণ নমুনায়নের মাধ্যমে গণসমষ্টি যে কোনো একক নমুনায় অন্তর্ভুক্ত হতে পারে।

কাঠামো (Frame): গণসমষ্টির সকল নমুনায়ন এককের তালিকাকে কাঠামো বা ফ্রেম বলা হয়। যেমন, ডায়াবেটিক সেন্টারে আগত রোগীদের সুগার-এর পর্যালোচনা করা হবে। এক্ষেত্রে কোনো একটি নির্দিষ্ট সময়ে আগত একজন ডায়াবেটিক রোগী হলো নমুনায়ন একক এবং ঐ নির্দিষ্ট সময়ে আগত সকল রোগীর তালিকা হলো ফ্রেম। অসীম গণসমষ্টির ক্ষেত্রে ফ্রেম পাওয়া যায় না। অনেক সময় সসীম গণসমষ্টির ক্ষেত্রে ফ্রেম পাওয়া অসম্ভব না হলেও কষ্টকর। যেমন, কোনো দেশের সন্তান দেয়ার ক্ষমতাশীল দম্পতি গণসমষ্টি হলে ঐ দম্পতিদের ফ্রেম পাওয়া অসম্ভব না হলেও সহজসাধ্য নয়।

প্রশ্নমালা (Questionnaire) : নমুনা একক হতে উদ্দেশ্য অনুযায়ী উপাত্ত সংগ্রহ করার জন্য সহজবোধ্য কতকগুলি প্রশ্নের সমাহারকে বলা হয় প্রশ্নমালা। এই প্রশ্নমালাই হলো উপাত্তের মূল ভিত্তি। কারণ প্রতিটি নমুনায়ন একক প্রশ্নমালা অনুযায়ী উপাত্ত সরবরাহ করে থাকে। উপাত্তের সরবরাহ পদ্ধতি অবশ্য দুই ধরনের। যেমন, (১) প্রশ্নমালা নিয়ে নমুনায়ন একককে পরিদর্শন করা এবং সরাসরি কথোপকথনের মাধ্যমে প্রশ্নমালা পূরণ করা, (২) ডাকযোগে প্রশ্নমালা নমুনায়ন এককের নিকট পাঠানো এবং তা পূরণ করে ফেরৎ পাঠাবার ব্যবস্থা করা।

প্রাক জরিপ (Pilot survey) : যে কোনো জরিপ কাজ পরিচালনা করার পূর্বেই গণসমষ্টি সম্পর্কে ধারণা লাভ করার জন্য বা গণসমষ্টির নমুনায়ন এককসমূহকে সহজ পদ্ধতিতে। অল্প সময়ে এবং অল্প খরচে পরিদর্শনের জন্য ব্যবস্থা করা সম্পর্কে ধারণা লাভ করার জন্য একটি জরিপ কাজ পরিচালনা করা হয়। এই জরিপকে বলা হয় প্রাক-জরিপ।

নমুনায়ন নকশা (Sampling design) : গণসমষ্টি একক হতে ঐগুলির একটি প্রতিনিধিত্বমূলক অংশ চয়ন করা হলো নমুনায়ন। এই নমুনায়নের জন্য যে পদ্ধতি প্রয়োগ করা হয় তাকে নমুনায়ন নকশা বলা হয়। নমুনায়ন নকশা সম্পর্কে সিদ্ধান্ত নেয়ার আগে নমুনায়ন এককের যে বৈশিষ্ট্য পর্যালোচনা করা হবে তা নমুনায়ন এককের ভেদের সাথে সম্পর্কিত কিনা তা লক্ষ্য রাখতে হয়। কারণ বৈশিষ্ট্যের ভেদ এককের ভেদের সাথে সম্পর্কিত হলে এমনভাবে নমুনা চয়ন করতে হবে যেন সকল প্রকার নমুনায়ন একক নমুনায় অন্তর্ভুক্ত হয়। সাধারণত সরল দৈব নমুনায়ন (simple random sampling), স্তরিত দৈব নমুনায়ন (stratified random sampling), ধারাবাহিক নমুনায়ন (systematic sampling), গুচ্ছ নমুনায়ন (cluster sampling) ইত্যাদি নমুনা নকশা হিসেবে বিবেচিত হয়। জীববিজ্ঞানে বিশেষ করে জীবজন্তু সম্পর্কীয় জরিপে আরো একটি নমুনা নকশার ব্যবহার হয়ে থাকে যা কেপচার-রিকেপচার প্রতিকৃতি (capture re-capture model) হিসেবে চিহ্নিত। এই প্রতিকৃতির মূল কথা হলো কতকগুলি জীবজন্তুকে প্রথমে ধরা হয় এবং ঐগুলির উপর কোনো পরীক্ষণ কাজ পরিচালনা করে ঐগুলিকে কোনোভাবে চিহ্নিত করে তাদের স্বাভাবিক জীবন প্রণালীতে ছেড়ে দেয়া হয়। তারপর একটি নির্দিষ্ট সময় পরে আবার ঐ জাতীয় কিছু জীবজন্তু ধরা হয়। এই দ্বিতীয় পর্যায়ের ধরাতে পূর্ব চিহ্নিত জীবজন্তু কিছু থাকা স্বাভাবিক। এভাবে দ্বিতীয় পর্যায়ে জীবজন্তু চয়ন করার পদ্ধতি হলো কেপচার-রিকেপচার প্রতিকৃতি। অন্যান্য নমুনায়ন নকশা সম্পর্কে পরে আলোচনা করা হবে।

দৈব নমুনায়ন (Random samplings) : কোনো সম্ভাবনা তত্ত্বের ভিত্তিতে নমুনায়ন একককে গণসমষ্টি হতে চয়ন করা হলে ঐ নমুনায়নকে দৈব নমুনায়ন বলা হয়। দৈব নমুনায়নের ক্ষেত্রে

(ক) প্রতিটি নমুনায়ন একক নমুনায় অন্তর্ভুক্ত হওয়ার সম্ভাবনা সমান হতে পারে।

(খ) ভিন্ন ভিন্ন নমুনায়ন এককের নমুনায় অন্তর্ভুক্ত হওয়ার সম্ভাবনা ভিন্ন ভিন্ন হতে পারে।

দৈব নমুনায়নের জন্য দুটি পদ্ধতি প্রয়োগ করা হয়। এগুলি হলো—(১) দৈব সংখ্যার সারণি (random number table)-এর প্রয়োগ এবং (২) লটারি (lottery)।

দৈব সংখ্যার সারণি-এর প্রয়োগ (Use of random number table) : দৈব সংখ্যার সারণি হলো কতকগুলি রাশির সারি এবং স্তম্ভে সাজানো রূপ। Fisher and Yates তাঁদের বই "Statistical Tables for Biological, Agricultural and Medical Research"-এ এই সারণি সরবরাহ করেছেন। এই সারণির সারি বা স্তম্ভে দেয়া রাশি দ্বারা তৈরি সংখ্যা অনুযায়ী গণসমষ্টি হতে একক চয়ন করা হলে নমুনায়ন দৈব বিবেচনা করা হয়।

এই সারণি প্রয়োগ করার জন্য প্রথমে গণসমষ্টি এককের একটি তালিকা তৈরি করতে হয়। তালিকায় প্রতিটি এককের বিপরীতে একটি ক্রমিক সংখ্যা দেয়া হয়। ধরা যাক কোনো গণসমষ্টি 100 একক আছে। তাহলে ঐগুলিকে ক্রমিক সংখ্যা দ্বারা চিহ্নিত করা হলে সংখ্যাগুলি হবে 00, 01, 02, 03, ··· , 99। মনে করা যাক এই গণসমষ্টি হতে 10 এককের একটি নমুনা চয়ন করা হবে। সেক্ষেত্রে পরিসংখ্যানের কোনো বই-এ দেয়া Random number table-এর কোনো একটি হতে সারি অনুযায়ী বা স্তম্ভ অনুযায়ী দুই রাশিবিশিষ্ট (যেহেতু গণসমষ্টির এককসমূহের ক্রমিক সংখ্যা সর্বোচ্চ দুই রাশিবিশিষ্ট) পরপর দশটি দৈব সংখ্যা চয়ন করতে হয়। এই দৈব সংখ্যাবিশিষ্ট গণসমষ্টি একক নমুনায় অন্তর্ভুক্ত করতে হয়। মনে রাখতে হবে দৈব সংখ্যা চয়ন করার সময় কোনো সংখ্যা যেন দুইবার চয়ন করা না হয়। অনেক সময় দৈব সংখ্যা চয়ন করতে লক্ষ্য করা যায় যে সারণি কোনো সংখ্যা গণসমষ্টি এককের বৃহত্তম ক্রমিক সংখ্যা চেয়ে বড়। সেক্ষেত্রে ঐ দৈব সংখ্যাটি বাদ দিয়ে পরবর্তী দৈব সংখ্যা চয়ন করতে হয়।

দৈব নমুনায়নের একটি সহজ পদ্ধতি হলো লটারি পদ্ধতি : এই পদ্ধতির জন্য গণসমষ্টি এককের মোট সংখ্যার সমান কতগুলি সমান আকারের কাগজে গণসমষ্টি এককের ক্রমিক সংখ্যাগুলি লেখা হয়। সংখ্যা লেখা কাগজগুলি পরে এমনভাবে ভাঁজ করতে হয় যেন ভিতরের লেখা সংখ্যা দেখা না যায় এবং প্রতিটি ভাঁজ করা কাগজের আকার সমান হয়। এরপর সংখ্যা লেখা কাগজগুলিকে ভাল করে মিশিয়ে একটি থলিতে বা ঝুড়িতে রাখা হয়। গণসমষ্টি হতে কয়টি একক নমুনায়ন অন্তর্ভুক্ত করতে হবে সে কয়টি কাগজ থলি হতে চয়ন করতে হয়। কাগজে লেখা ক্রমিক সংখ্যাবিশিষ্ট গণসমষ্টি এককগুলি নমুনায়ন অন্তর্ভুক্ত হবে।

৮.৩ সরল দৈব নমনায়ন (Simple Random Sampling)

কোনো দৈব নমুনায়নের ক্ষেত্রে যদি প্রতিটি নমুনায়ন একক অনপেক্ষভাবে সমান সম্ভাবনাসহ চয়ন করা হয়, তাহলে নমুনায়নকে সরল দৈব নমুনায়ন বলা হয়। ধরা যাক কোনো গণসমষ্টিতে N একক আছে এবং n আকারের একটি নমুনা চয়ন করতে হবে। এই গণসমষ্টি হতে মোট $\binom{N}{n}$ নমুনায়ন করা যায়। প্রতিটি নমুনা চয়ন করার সম্ভাবনা সমান হলে নমুনায়ন পদ্ধতিতে সরল দৈব নমুনায়ন বলা হয়।

ধরা যাক গণসমষ্টি এককসমূহের যে বৈশিষ্ট্য পর্যালোচনা করা হবে প্রতি এককের ক্ষেত্রে ঐ বৈশিষ্ট্যের মান হলো $y_1, y_2, \ldots, y_N$। এই গণসমষ্টি হতে n আকারের নমুনা চয়ন করা হলে নমুনায়ন অন্তর্ভুক্ত এককসমূহের মান হলো $y_1, y_2, \ldots, y_n$। তাহলে

গণসমষ্টির ক্ষেত্রে	নমুনার ক্ষেত্রে
মোট, $Y = y_1 + y_2 + \ldots + y_N = \sum^{N} y_i$	মোট, $y = y_1 + y_2 + \ldots + y_n = \sum^{n} y_i$
গড়, $\bar{Y} = \frac{1}{N}(y_1 + y_2 + \ldots + y_N)$	গড়, $\bar{y} = \frac{1}{n}(y_1 + y_2 + \ldots + y_n)$

সরল দৈব নমুনায়নের ক্ষেত্রে নমুনা গড় $\bar{y}$ গণসমষ্টি গড় $\bar{Y}$-এর নির্ঝুঁকি নিরূপক (unbiased estimator)। এই নিরূপকের ভেদাঙ্ক হলো

$$V(\bar{y}) = \frac{N-n}{Nn} S^2 = \frac{1-f}{n} S^2, \text{ এখানে } f = \frac{n}{N}$$

এবং $S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{Y})^2 = \frac{1}{N-1}\left[\sum^{N} y_i^2 - \frac{\left(\sum^{N} y_i\right)^2}{N}\right]$

আবার $V(\bar{y})$-এর নিৰ্ঝুঁকি নিরূপক হলো

$$v(\bar{y}) = \frac{N-n}{Nn}\ s^2,\ \text{এখানে}$$

$$s^2 = \frac{1}{n-1}\left[\sum^{n} y_i^2 - \frac{\left(\sum^{n} y_i\right)^2}{n}\right]$$

এই নমুনায়ন হতে গণসমষ্টি মোট Y-এর নিৰ্ঝুঁকি নিরূপক হলো

$$\hat{Y} = N\bar{y}$$

এবং $$V(\hat{Y}) = N^2 V(\bar{y}) = \frac{N(N-n)}{n}\ S^2 = \frac{N^2(1-f)S^2}{n}$$

এই ভেদাঙ্কের নিৰ্ঝুঁকি নিরূপক হলো

$$v(\hat{Y}) = \frac{N(N-n)}{n}s^2 = \frac{N^2(1-f)}{n}s^2$$

উপরিউক্ত n/N-কে বলা হয় নমুনায়ন ভগ্নাংশ (sampling fraction)। এই ভগ্নাংশ খুব ছোট হলে, অর্থাৎ গণসমষ্টি এককের তুলনায় নমুনা আকার ছোট হলে f-কে বাদ দেয়া যায়। সেক্ষেত্রে

$$v(\bar{y}) = \frac{s^2}{n}$$

এবং $$v(\hat{Y}) = \frac{N^2 s^2}{n}$$

উপরিউক্ত নমুনায়ন পদ্ধতির মাধ্যমে গণসমষ্টি গড়ের 95% নিশ্চয়তা-পরিক্ষেপ (confidence interval) হলো

$$\hat{\bar{Y}}_L = \bar{y} - t_{.025,\, n-1}\sqrt{\frac{(1-f)}{n}s^2}$$

$$\hat{\bar{Y}}_U = \bar{y} + t_{.025,\, n-1}\sqrt{\frac{(1-f)}{n}s^2}$$

এখানে $\hat{\bar{Y}}_L = \bar{Y}$-এর ন্যূনতম মানের নিরূপক

$\hat{\bar{Y}}_u = \bar{Y}$-এর সর্বোচ্চ মানের নিরূপক

$t_{0\cdot025,n-1} = (n-1)$ স্বাধীনতার মাত্রবিশিষ্ট 't'-বিন্যাসের 5% সারণিকৃত মান।

অনুরূপভাবে গণসমষ্টি মোট-এর 95% নিশ্চয়তা-পরিক্ষেপও নির্ণয় করা যায়। $\hat{Y}_L$ ও $\hat{Y}_u$ যথাক্রমে গণসমষ্টি মোট-এর ন্যূনতম ও সর্বোচ্চ মান ধরা হলে এগুলি নির্ণয়ের সূত্র হলো

$$\hat{Y}_L = N\bar{y} - t_{0\cdot025,\, n-1}\sqrt{\frac{N^2(1-f)s^2}{n}}$$

এবং

$$\hat{Y}_u = N\bar{y} - t_{0\cdot025,\, n-1}\sqrt{\frac{N^2(1-f)s^2}{n}}$$

উদাহরণ ৮.১

নিচে একটি উপজেলার বিভিন্ন গ্রামের কাঁঠাল গাছের সংখ্যা দেয়া হলো :

গ্রামের ক্রমিক সংখ্যা	কাঁঠাল গাছের সংখ্যা, y_i	গ্রামের ক্রমিক সংখ্যা	কাঁঠাল গাছের সংখ্যা, y_i	গ্রামের ক্রমিক সংখ্যা	কাঁঠাল গাছের সংখ্যা, y_i
00	150	03	51	06	168
01	180	04	62	07	55
02	10	05	72	08	12
09	228	31	36	53	168
10	76	32	48	54	52

গ্রামের ক্রমিক সংখ্যা	কাঁঠাল গাছের সংখ্যা, y_i	গ্রামের ক্রমিক সংখ্যা	কাঁঠাল গাছের সংখ্যা, y_i	গ্রামের ক্রমিক সংখ্যা	কাঁঠাল গাছের সংখ্যা, y_i
11	22	33	62	55	32
12	18	34	50	56	28
13	70	35	52	57	27
14	55	36	68	58	42
15	52	37	31	59	60
16	24	38	17	60	75
17	28	39	16	61	46
18	37	40	29	62	30
19	45	41	42	63	42
20	51	42	20	64	48
21	12	43	39	65	28
22	17	44	67	66	27
23	18	45	88	67	56
24	20	46	92	68	31
25	50	47	36	69	19
26	68	48	18	70	117
27	117	49	19	71	112
28	250	50	59	72	82
29	108	51	26	73	96
30	95	52	22	74	78

দশটি গ্রাম দৈব পদ্ধতিতে চয়ন করে গ্রাম প্রতি কয়টি কাঁঠাল গাছ আছে তা নিরূপণ করা যাক। এই নিরূপকের 95% নিশ্চয়তা-পরিক্ষেপ নির্ণয় করা যাক। এছাড়া উক্ত উপজেলায় মোটকাঁঠাল গাছের 95% নিশ্চয়তা-পরিক্ষেপও নিরূপণ করা যাক।

এখানে $N = 75$, $n = 10$ । এখন Probability and Statistics in Engineering and Management Science, 2nd, edition, by W.W. Hines and D.C. Montgomery, Wiley, New York, 1980 বই হতে Random Number Table ব্যবহার করে নিচের দশটি দৈব সংখ্যা পাওয়া গেছে :

10, 22, 24, 42, 37, 28, 63, 09, 07, 51

সুতরাং নমুনায় অন্তর্ভুক্ত গ্রাম ও ঐগুলির কাঁঠাল গাছের সংখ্যা নিচে দেয়া হলো :

গ্রামের ক্রমিক সংখ্যা :	10	22	24	42	37	28	63	09	07	51
কাঁঠাল গাছের সংখ্যা, y_i :	76	17	20	20	31	250	42	228	55	42

এখানে গ্রাম প্রতি কাঁঠাল গাছের সংখ্যার নিরূপক হলো

$$\bar{y} = \frac{1}{n}\Sigma y_i = \frac{781}{10} = 78.1 = 78$$

এখানে
$$s^2 = \frac{1}{n-1}\left[\Sigma y_i^2 - \frac{(\Sigma y_i)^2}{n}\right] = \frac{1}{9}\left[128863 - \frac{(781)^2}{10}\right]$$
$$= 7540.77$$

গ্রাম প্রতি কাঁঠাল গাছের 95% নিশ্চয়তা-পরিক্ষেপ হলো

$$\hat{Y}_L = \bar{y} - t_{.025,9}\sqrt{\frac{(1-f)}{n}s^2}\ ,\ f = \frac{n}{N} = \frac{10}{75} = 0.13$$

$$= 78 - 2.262\sqrt{\frac{(1-0.13)7540.77}{10}}$$

$$= 78 - 2.262 \times 25.61 = 20$$

$$\hat{Y}_U = \bar{y} + t_{0.025}\sqrt{\frac{(1-f)s^2}{n}}$$

$$= 78 + 2.262 \times 25.61 = 135.6 = 136$$

উপজেলায় মোট কাঁঠাল গাছের নিরূপক হলো

$$\hat{Y} = N\bar{y} = 75 \times 78.1 = 5857.5 = 5858$$

মোট কাঁঠাল গাছের 95% নিশ্চয়তা-পরিক্ষেপ হলো

$$\hat{Y}_L = N\bar{y} - t_{0.025,9}\sqrt{\frac{N^2(1-f)s^2}{n}}$$

$$= 5858 - 2.262\sqrt{\frac{(75)^2(1-0.13)7540.77}{10}}$$

$$= 5858 - 2.262 \times 1921.01 = 1513$$

$$\hat{Y}_U = N\bar{y} + t_{0.025}\sqrt{\frac{N^2(1-f)s^2}{n}}$$

$$= 5858 + 2.262 \times 1921.01$$

$$= 10203$$

৮.৪ সরল দৈব নমনায়নের ক্ষেত্রে গণসমষ্টি সমানুপাত নিরূপণ (Estimation of Population Proportion in Case of Simple Random Sampling)

গণসমষ্টির বৈশিষ্ট্য পর্যালোচনা করার ক্ষেত্রে ঐ বৈশিষ্ট্যের প্রাসঙ্গিক চলক পরিমাণগত (quantitative) বা গুণগত (qualitative) হতে পারে। পরিমাণগত চলকের ক্ষেত্রে গণসমষ্টি গড় ও গণসমষ্টি মোট নিরূপণ পদ্ধতি পূর্ববর্তী অনুচ্ছেদে করা হয়েছে। বর্তমান অনুচ্ছেদে গুণগত চলকের ক্ষেত্রে গণসমষ্টি গড় বা সমানুপাত নিরূপণ পদ্ধতি আলোচনা করা হবে।

উপরিউক্ত নিরূপণের ক্ষেত্রে যে বৈশিষ্ট পর্যালোচনা করা হবে তার প্রাসঙ্গিক চলকের মান কোনো নমুনায়ন এককে হয় বিদ্যমান থাকতে পারে নাও থাকতে পারে। পরিসংখ্যানিক বিশ্লেষণের জন্য ঐ বৈশিষ্ট্য বিদ্যমান থাকলে চলকের মান ধরা হয় 1 এবং বিদ্যমান না থাকলে চলকের মান ধরা হয় 0 [$y_i = 0$, যদি i-তম এককে ঐ বৈশিষ্ট্য বিদ্যমান থাকে, $y_i = 0$, যদি i-তম এককে ঐ বৈশিষ্ট্য বিদ্যমান না থাকে]। উদাহরণ হিসেবে ধরা যাক সারণি ৬.২৩-এ দেয়া মাছের প্রজাতির মধ্যে Epinephelus aeneus প্রজাতির সমানুপাত (proportion) কত তা জানতে হবে। এক্ষেত্রে মোট মাছের মধ্যে যেগুলি Epinephelus aeneus প্রজাতির সেগুলির জন্য $y_i = 1$ এবং যেগুলি এই প্রজাতির নয় সেগুলির ক্ষেত্রে $y_i = 0$ ধরা যেতে পারে। ধরা যাক কোনো গণসমষ্টিতে একটি বিশেষ বৈশিষ্ট্যের অধিকারী এককের সংখ্যা হলো A। নমুনায় অন্তর্ভুক্ত এ জাতীয় এককের সংখ্যা ধরা যাক a। যেমন, সারণি ৬.২৩-এ দেয়া উপাত্তের ক্ষেত্রে Epinephelus aeneus প্রজাতির মাছের সংখ্যা 11, অর্থাৎ A = 11। তাহলে

গণসমষ্টির ক্ষেত্রে	:	নমুনার ক্ষেত্রে
গণসমষ্টি গড়, $\bar{Y}$		নমুনা গড় বা সমানুপাত

অথবা সমানুপাত

$$P = \frac{1}{N}[y_1 + y_2 + \ \dots \ + y_N] \qquad \bar{y} = p = \frac{1}{n}(y_1 + y_2 + \cdots + y_n)$$

$$= \frac{A}{N} \qquad\qquad = \frac{a}{n}$$

গণসমষ্টি মোট — নমুনা মোট

$$Y = A = y_1 + y_2 + \cdots + y_N \qquad a = y_1 + y_2 + \ \cdots \ + y_n$$

$$= \sum^{N} y_i \qquad\qquad = \sum^{n} y_i$$

এখানে নমুনা সমানুপাত p গণসমষ্টি সমানুপাত P-এর নিখুঁত নিরূপক। এই নিরূপকের ভেদাঙ্ক হলো

$$V(p) = \frac{N-n}{Nn}\,\frac{NPQ}{N-1} = \frac{N-n}{N-1}\,\frac{PQ}{n} = \frac{N(1-f)}{N-1}\,\frac{PQ}{n}$$

$$= \frac{NPQ}{n(N-1)}, \quad \text{যদি } f = \frac{n}{N} \text{ ছোট হয়।}$$

এখানে $Q = 1 - P$ । এই ভেদাঙ্কের নিরূপক হলো

$$v(p) = \frac{N-n}{Nn}\,\frac{npq}{n-1} = \frac{(1-f)}{n-1}pq\,, \quad q = 1 - p$$

$$= \frac{pq}{n-1}, \quad \text{যদি } f = \frac{n}{N} \text{ ছোট হয়।}$$

উপরিউক্ত নিরূপকগুলির ভিত্তিতে গণসমষ্টি সমানুপাত P-এর 95% নিশ্চয়তা-পরিক্ষেপ হলো

$$\hat{P}_L = \hat{p} - t_{0.025,\,n-1}\sqrt{\frac{(1-f)pq}{n-1}}$$

এবং
$$\hat{P}_U = \hat{p} + t_{0.025,\,n-1}\sqrt{\frac{(1-f)pq}{n-1}}$$

এই নিরূপণের ক্ষেত্রে গণসমষ্টি মোট A-এর 95% নিশ্চয়তা-পরিক্ষেপ হলো

$$\hat{A}_L = \hat{A} - t_{.025,\,n-1}\sqrt{\frac{N^2(1-f)pq}{n-1}}\;,\quad \hat{A} = Np$$

এবং $$\hat{A}_U = \hat{A} + t_{.025,\,n-1}\sqrt{\frac{N^2(1-f)pq}{n-1}}$$

এখানে $\hat{A}$ হলো A-এর নির্ঝ়ঁকি নিরূপক, $\hat{A}_L$ এবং $\hat{A}_U$ হলো যথাক্রমে A-এর ন্যূনতম মান এবং সর্বোচ্চ মানের নিরূপক।

উদাহরণ ৮.২

সারণি ৬.২৩-এ দেয়া উপাত্ত হতে দশটি মাছের একটি নমুনা চয়ন করে Epinephelus aeneus প্রজাতির মাছের সমানুপাত নিরূপণ করা যাক এবং সমানুপাতের 95% নিশ্চয়তা-পরিক্ষেপ নিরূপণ করা যাক। এই প্রজাতির মোট মাছের সংখ্যার 95% নিশ্চয়তা-পরিক্ষেপও নিরূপণ করা যাক।

উদাহরণ ৮.১-এ ব্যবহৃত দৈব সংখ্যা এই উদাহরণের ক্ষেত্রেও ব্যবহার করা যায়। এই দৈব সংখ্যাগুলি হলো

10, 22, 24, 42, 37, 28, 63, 09, 07, 51

উক্ত দৈব সংখ্যা অনুযায়ী সারণি ৬.২৩ হতে চয়ন করা মাছের ক্রমিক সংখ্যা এবং Epinephelus aeneus প্রজাতির মাছের ক্ষেত্রে $y_j = 1$ এবং এই প্রজাতি না হলে $y_j = 0$ নিচে দেখানো হলো

মাছের ক্রমিক সংখ্যা :	10	22	24	42	37	28	63	09	07	51
Epinephelus aeneus y_j :	1	0	0	0	0	0	0	1	1	0

এখানে $a = \Sigma y_j = 3,\; n = 10$

Epinephelus aeneus প্রজাতির মাছের সমানুপাত হলো

$$p = \frac{a}{n} = \frac{3}{10} = 0.3\;,\quad f = \frac{n}{N} = \frac{10}{62} = 0.16$$

সুতরাং $q = 1 - p = 1 - 0.3 = 0.7$, এখানে $N = 62$

$$v(p) = \frac{(1-f)pq}{n-1} = \frac{(1-0.16)0.3\times 0.7}{10-1} = 0.0196$$

সুতরাং গণসমষ্টিতে Epinephelus aeneus মাছের সমানুপাতের 95% নিশ্চয়তা-পরিক্ষেপ হলো

$$\hat{P}_L = \hat{p} - t_{.025,9}\sqrt{\frac{(1-f)pq}{n-1}} = 0.3 - 2.262\sqrt{0.0196} = -0.02$$

$$\hat{P}_U = \hat{p} + t_{.025,9}\sqrt{\frac{(1-f)pq}{n-1}} = 0.3 + 2.262\sqrt{0.0196} = 0.62$$

উক্ত গণসমষ্টিতে **Epinephelus aeneus** প্রজাতির মোট মাছের সংখ্যার নিরূপক হলো

$$\hat{A} = N\hat{p} = 62 \times 0.3 = 18.6 \doteq 19$$

এবং $$v(\hat{A}) = N^2 v(\hat{p}) = \frac{N^2(1-f)pq}{n-1} = \frac{(62)^2(1-0.16)0.3 \times 0.7}{10-1}$$

$$= 75.3424$$

এখানে উক্ত প্রজাতির মোট মাছের সংখ্যার 99% নিশ্চয়তা-পরিক্ষেপ হলো

$$\hat{A}_L = \hat{A} - t_{0 \cdot 005,9}\sqrt{\frac{N^2(1-f)pq}{n-1}} = 19 - 3.25\sqrt{75.3424}$$

$$= -9.21 \doteq -9$$

$$\hat{A}_U = \hat{A} - t_{.005,9}\sqrt{\frac{N^2(1-f)pq}{n-1}} = 19 + 3.25\sqrt{75.3424}$$

$$= 47.21$$

$$\doteq 47$$

উপরিউক্ত নিরূপণ পদ্ধতি সরল দৈব নমুনায়নের মাধ্যমে প্রাপ্ত নমুনার ভিত্তিতে করা হয়েছে। এই নমুনায়নের একটি সুবিধা হলো যে নমুনা একক চয়ন করা সহজ এবং অল্প খরচে এই নমুনায়ন করা যায়। এই নমুনায়নের মাধ্যমে প্রাপ্ত নিরূপক অধিক নির্ভরযোগ্য। এই নমুনায়নের একটি অসুবিধা হলো যে গণসমষ্টির ফ্রেম না থাকলে এই নমুনায়ন পদ্ধতি প্রয়োগ করা যায় না। অধিকন্তু গণসমষ্টি এককের সংখ্যা বেশি বড় হলে এই নমুনায়ন পদ্ধতি সুবিধাজনক নয়।

৮.৫ স্তরিত নমুনায়ন (Stratified Sampling)

কোনো কোনো গণসমষ্টির ক্ষেত্রে নমুনায়ন এককসমূহের বৈশিষ্ট্য সমজাতীয় (homogeneous) নয়। যেমন, উদাহরণ ৭.৪-এ উল্লেখিত তথ্য অনুযায়ী লক্ষ্য করা যাচ্ছে যে পেপ্টিসাইড প্রয়োগে দুই প্রকার Slug এর Heart beat দুই রকম। অর্থাৎ পেপ্টিসাইড প্রয়োগের পর Heart beat এর প্রসঙ্গে Slug-সমূহ সমজাতীয় নয়। এক্ষেত্রে বৈশিষ্ট্যের সমমাত্রিকতা (homogeneity) অনুযায়ী নমুনায়ন একক-সমূহকে বিভিন্ন স্তরে (strata) ভাগ করতে হয়। যেমন, উদাহরণ ৭.৪-এর ক্ষেত্রে Slug-সমূহকে তাদের Heart beat-এর প্রসঙ্গে দুই স্তরে ভাগ করা যায়। এক স্তরে হলো M. Rusticus এবং অন্য স্তরে হলো M. Sowerbyi। এরপর প্রতি স্তর থেকে ভিন্ন ভিন্নভাবে সরল দৈব নমুনায়নের মাধ্যমে নমুনা একক চয়ন করা হলে ঐ নমুনায়ন পদ্ধতিকে স্তরিত নমুনায়ন বলা হয়। যেমন, Slug-এর গড় Heart beat পরিমাপ করার জন্য M. Rusticus হতে কিছু Slug এবং M. Sowerbyi হতে কিছু Slug ভিন্ন ভিন্নভাবে সরল দৈব নমুনায়নের মাধ্যমে চয়ন করা হলে প্রাপ্ত নমুনা হবে স্তরিত নমুনা।

স্তরিত নমুনায়নের জন্য গণসমষ্টি এককের সমমাত্রিকতা অনুযায়ী ঘটনসংখ্যা বিন্যাস তৈরি করে স্তর তৈরি করা হয়। এই স্তর তৈরি করার ক্ষেত্রে পূর্ব অভিজ্ঞতা কাজে লাগানো হয়। যেমন, সারণি ১.৩ (প্রথম খণ্ড)-এর উপাত্ত বিশেষ করে জানা যায় যে, Heart beat-এর প্রসঙ্গে M. Rusticus এবং M. Sowerbyi সমমাত্রিক নয়। সুতরাং উক্ত সারণিতে দেয়া 48 Slug-কে দুটি শ্রেণীতে ভাগ করে অর্থাৎ M. Rusticus নিয়ে একটি স্তর এবং M. Sowerbyi নিয়ে অন্য একটি স্তর গঠন করা যায়।

ধরা যাক কোনো গণসমষ্টিতে N একক আছে। এই N একককে L স্তরে ভাগ করা যায় এমনভাবে যেন প্রথম স্তরের একক সংখ্যা হয় N_1, দ্বিতীয় স্তরের একক সংখ্যা N_2, ... এভাবে L-তম স্তরের একক সংখ্যা N_L হয় এবং $N = N_1 + N_2 + \cdots + N_L$ হয়। আরো ধরা যাক i-তম স্তরের $[i = 1, 2, \cdots, L]$ j-তম এককের যে বৈশিষ্ট্য পর্যালোচনা করা হবে তা $[j = 1, 2, \cdots, N_i]$

$$y_{ij} : [y_{i1}, n_{i2}, \cdots, y_i N_i]$$

তাহলে i-তম স্তরের গণসমষ্টি গড় হলো

$$\bar{Y}_i = \frac{1}{N_i}\left(y_{i1} + y_{i2} + \cdots + y_{iN_i}\right) = \frac{\sum_{j=1}^{N_i} y_{ij}}{N_i}$$

এবং গণসমষ্টি ভেদাঙ্ক হলো

$$S_i^2 = \frac{1}{N_i - 1}\sum_{j=1}^{N_i}(y_{ij} - \bar{Y}_i)^2 = \frac{1}{N_i - 1}\left(\sum_j y_{ij}^2 - \frac{(\sum_j y_{ij})^2}{N}\right)$$

এক্ষেত্রে গণসমষ্টি গড় হলো

$$\bar{Y} = \frac{1}{N}\sum_{i=1}^{L}\sum_{j=1}^{N_i} y_{ij} = \sum_{i=1}^{L}\frac{N_i\bar{Y}_i}{N} = \sum_{i=1}^{L} W_i\bar{Y}_i,\ W_i = \frac{N_i}{N}$$

এবং গণসমষ্টির সকল এককের ভেদাঙ্ক হলো

$$S^2 = \frac{1}{N-1}\sum_{i}^{L}\sum_{j}^{N_i}(y_{ij} - \bar{Y})^2$$

$$(N-1)S^2 = \sum_i\sum_j\left[(y_{ij} - \bar{Y}_i) + (\bar{Y}_i - \bar{Y})\right]^2$$

সরল করে পাওয়া যায়

$$(N-1)S^2 = \sum_i (N_i - 1)S_i^2 + \Sigma N_i(\bar{Y}_i - \bar{Y})^2$$

এখন $1/N_i$ এবং $1/N$ খুব ছোট হলে সরল করে পাওয়া যায়

$$S^2 = \frac{1}{N}\Sigma N_i S_i^2 + \frac{1}{N}\Sigma N_i(\bar{Y}_i - \bar{Y})^2$$

অথবা
$$\frac{S^2}{n} = \frac{1}{Nn}\Sigma N_i S_j^2 + \frac{1}{Nn}\Sigma N_i(\bar{Y}_i - \bar{Y})^2$$

এই নমুনায়নের ক্ষেত্রে ধরা যাক n আকারের একটি নমুনা চয়ন করতে হবে এমনভাবে যেন প্রথম স্তর থেকে n_1 একক, দ্বিতীয় স্তর থেকে n_2 একক এবং এভাবে

i-তম স্তর থেকে n_2 একক নমুনায় অন্তর্ভুক্ত হয় এবং $n = n_1 + n_2 + \dots + n_2$ হয়। ধরা যাক i-তম স্তরের j-তম এককের $(i = 1, 2, \dots, L; j = 1, 2, \cdots, n_i)$ পর্যালোচিত বৈশিষ্ট্যের মান y_{ij}। তাহলে i-তম স্তরের নমুনা গড় হলো

$$\bar{y}_i = \frac{1}{n_i}(y_{i1} + y_{i2} + \cdots + y_{in_i}) = \frac{1}{n_i}\sum_{j}^{n_i} y_{ij}$$

এবং এই স্তরের নমুনা ভেদাঙ্ক হলো

$$s_i^2 = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2 = \frac{1}{n_i - 1}\left[\sum_{j} y_{ij}^2 - \frac{(\Sigma y_{ij})^2}{n_i}\right]$$

i-তম স্তর থেকে সরল দৈব নমুনায়নের মাধ্যমে n_i একক চয়ন করা হয় বলে $\bar{y}_i$ এবং s_i^2 যথাক্রমে $\bar{Y}_i$ ও S_i^2-এর নিরঝুঁকি নিরূপক। তাই এই নমুনায়নের ক্ষেত্রে গণসমষ্টি গড়ের নিরঝুঁকি নিরূপক হলো

$$\bar{y}_{st} = \frac{1}{N}\sum_{i=1}^{L} N_i \bar{y}_i = \sum_{i=1}^{L} W_i \bar{y}_i \, . \quad W_i = \frac{N_i}{N}$$

এই নিরূপকের ভেদাঙ্ক হলো

$$V(\bar{y}_{st}) = \frac{1}{N^2}\sum_{i=1}^{L} N_i^2 \frac{(N_i - n_i)}{N_i n_i} S_i^2$$

$$= \Sigma W_i^2 (1 - f_i)\frac{S_i^2}{n_i}\,, \quad f_i = \frac{n_i}{N_i}$$

i-তম স্তরের নমুনায়ন ভগ্নাংশ f_i খুব ছোট হলে

$$V(\bar{y}_{st}) = \sum_{i=1}^{L} W_i^2 \frac{S_i^2}{n_i}$$

এই নমুনায়নের ক্ষেত্রে i-তম স্তর থেকে চয়ন করা n_i একক ঐ স্তরের গণসমষ্টি একক N_i-এর সমানুপাতিক হলে, অর্থাৎ $n_i \propto N_i$

অথবা $\quad n_i = \frac{n}{N} N_i$ হলে

$$v(\overline{y}_{st}) = \sum \frac{N_i^2}{N^2}\left(1 - \frac{n_i}{N_i}\right)\frac{S_i^2}{n_i}$$

অথবা
$$V(\overline{y}_{st}) = \sum \frac{N_i^2}{N^2}\left(1 - \frac{\frac{n}{N}N_i}{N_i}\right)\frac{S_i^2}{\frac{n}{N}N_i}$$

$$= \sum^{L} \frac{N_i}{N}\left(1 - f\right)\frac{S_i^2}{n}, \quad f = \frac{n}{N}$$

$$= \sum W_i \frac{S_i^2}{n}, \quad \text{যদি } f = \frac{n}{N} \text{ খুব ছোট হয়।}$$

এই ভেদাঙ্কের নির্ঝুঁকি নিরূপক হলো

$$v(\overline{y}_{st})_{porp} = \sum_{i=1}^{L} \frac{N_i}{N}\left(1 - f\right)\frac{S_i^2}{n}$$

$$= \sum_{i=1}^{L} \frac{N_i}{N}\frac{s_i^2}{n}, \quad \text{যদি } f = \frac{n}{N} \text{ খুব ছোট হয়।}$$

অনেক সময় স্তরিত দৈব নমুনায়নের মাধ্যমে গণসমষ্টি মোট Y-কে নিরূপণ করতে হয়। এই নিরূপক হলো

$$\hat{Y}_{st} = N\overline{y}_{st}$$

সমানুপাতিক চয়নের ক্ষেত্রে $\left[n_i \propto N_i \text{ অথবা } n = \frac{n}{N}N_i\right]$ $\overline{Y}_{st}$-এর ভেদাঙ্কের নির্ঝুঁকি নিরূপক হলো

$$v(\hat{y}_{st}) = N^2 v(\overline{y}_{st})$$

$$= N\sum_{i=1}^{L} N_i(1 - f)\frac{s_i^2}{n}$$

$$= \frac{N}{n}\sum_{i=1}^{L} N_i s_i^2, \quad \text{যদি } f = \frac{n}{N} \text{ খুব ছোট হয়।}$$

অনেক সময় i-তম স্তর থেকে n_i-এর মান নিরূপণ করা হয় Neyman বণ্টনের মাধ্যমে। এই বণ্টন নিয়ম অনুযায়ী

$$n_i = \frac{nN_iS_i}{\Sigma N_iS_i} \quad [\, n_i \propto N_iS_i \,]$$

সেক্ষেত্রে $$v(\bar{y}_{st})_{opt} = \frac{1}{n}(\Sigma W_is_i)^2 - \frac{1}{N}\Sigma W_is_i^2\,,\ \ W_i = \frac{N_i}{N}$$

$$= \frac{1}{n}(\Sigma W_is_i)^2\,,\ \ \text{যদি}\ \ f_i = \frac{n_i}{N_i}\ \ \text{খুব ছোট হয়}$$

এবং $$v(\hat{Y}_{st})_{opt} = N^2\left[\frac{1}{n}(\Sigma W_is_i)^2 - \frac{1}{N}\,\Sigma W_is_i^2\right]$$

$$= N^2\frac{1}{n}(\Sigma W_is_i)^2\,,\ \ \text{যদি}\ \ f_i = \frac{n_i}{N_i}\ \ \text{খুব ছোট হয়।}$$

উদাহরণ ৮.৩

সারণি ১.৩ (১ম খণ্ড)-এর উপাত্তের ক্ষেত্রে M. Rusticus-কে একটি স্তরভুক্ত এবং M. Sowerbyi-কে অন্য স্তর বিবেচনা করে (১) সমানুপাতিক বণ্টন এবং (২) Neyman বণ্টনের মাধ্যমে বারটি Slug চয়ন করে ঐ Slug গণসমষ্টির গড় Heart beat নিরূপণ করা যাক এবং গণসমষ্টি গড়ের 95% নিশ্চয়তা-পরিক্ষেপ নির্ণয় করা যেতে পারে।

আলোচিত উপাত্তের ক্ষেত্রে $N_1 = 24$, $N_2 = 24$, $N = 48$, $i = 1,2$ $(L = 2)$, $n = 12$।

স্তর-1		স্তর-2		স্তর-1		স্তর-2	
ক্রমিক নং	y_{1j}	ক্রমিক নং	y_{2j}	ক্রমিক নং	y_{1j}	ক্রমিক নং	y_{2j}
01	14	01	12	13	9	13	10
02	11	02	14	14	9	14	10
03	13	03	12	15	9	15	10
04	13	04	14	16	9	16	9
05	9	05	8	17	7	17	11
06	8	06	8	18	8	18	12

07	8	07	9	19	7	19	12
08	9	08	8	20	8	20	12
09	9	09	11	21	9	21	13
10	10	10	10	22	10	22	13
11	9	11	11	23	9	23	12
12	9	12	11	24	8	24	13

$$S_1^2 = \frac{1}{N_1 - 1}\left[\Sigma y_{1j}^2 - \frac{(\Sigma y_{1j})^2}{N_1}\right], \quad S_2^2 = \frac{1}{N_2 - 1}\left[\Sigma y_{2j}^2 - \frac{(\Sigma y_{2j})^2}{N_2}\right]$$

$$= \frac{1}{23}\left[2704 - \frac{(250)^2}{24}\right] \qquad = \frac{1}{23}\left[2461 - \frac{(239)^2}{24}\right]$$

$$= 4.34 \qquad = 3.52$$

সমানুপাতিক বণ্টনের ক্ষেত্রে

$$n_1 = \frac{n}{N} \cdot N_1, \quad \therefore\ n_1 = \frac{12}{48} \times 24 = 6\ ;\ n_2 = \frac{12}{48} \times 24 = 6$$

এখন Appendix 7-এ দেয়া Random numvers-এর প্রথম দুই স্তম্ভ হতে নির্বাচিত 6 দৈব সংখ্যা হলো 10, 22. 24, 09, 07, 02। এই সংখ্যাগুলির পরবর্তী 6 দৈব সংখ্যা হলো 01, 07, 02, 05, 09, 17। সুতরাং প্রথম স্তর ও দ্বিতীয় স্তর হতে চয়ন করা Slug-এর Heart beat হলো যথাক্রমে

স্তর-1	দৈব সংখ্যা	:	10	22	24	09	07	02
	y_{1j}	:	10	10	8	9	8	11
স্তর-2	দৈব সংখ্যা	:	01	07	02	05	09	17
	y_{2j}	:	12	9	14	8	11	11

এখানে $$\bar{y}_1 = \frac{1}{n_1}\Sigma y_{1j} = \frac{56}{6} = 9.33$$

$$s_1^2 = \frac{1}{n_1 - 1}\left[\Sigma y_{1j}^2 - \frac{(\Sigma y_{1j})^2}{n_1}\right] = \frac{1}{5}\left[530 - \frac{(56)^2}{6}\right] = 1.47$$

$$\bar{y}_2 = \frac{1}{n_2}\Sigma y_{2j} = \frac{65}{6} = 10.83$$

$$s_2^2=\frac{1}{n_2-1}\left[\Sigma y_{2j}^2-\frac{(\Sigma y_{2j})^2}{n_2}\right]=\frac{1}{5}\left[727-\frac{(65)^2}{6}\right]=4.57$$

সুতরাং Heart beat-এর গণসমষ্টি গড়ের নিরূপক হলো

$$\bar{y}_{st}=\frac{1}{N}\Sigma N_i\bar{y}_i=\frac{1}{48}\left[24\times 9.33+24\times 10.83\right]=10.08$$

-এর ভেদাঙ্কের নিরূপক হলো

$$v(\bar{y}_{st})_{prop}=\Sigma\frac{N_i}{N}(1-f)\frac{s_i^2}{n}$$

$$=\frac{1-f}{Nn}\left[\Sigma N_i s_i^2\right],\quad f=\frac{n}{N}=\frac{12}{48}=0.25$$

$$=\frac{1-0.25}{48\times 12}[24\times 1.47+24\times 4.57]=0.19$$

এই নিরূপণ হতে গণসমষ্টি গড়ের 95% নিশ্চয়তা-পরিক্ষেপ হলো

$$\hat{\bar{Y}}_L=\bar{y}_{st}-t_{.025,\,n_e}\sqrt{v(\bar{y}_{st})_{prop}}\,,\quad n_e=\frac{(\Sigma g_i s_i^2)^2}{\Sigma\frac{g_i^2\,s_i^4}{n_i-1}}$$

$$=10.08-t_{.025,\,8}\sqrt{0.19}$$

$$=10.08-2.306\sqrt{0.19}=9.07\qquad g_i=\frac{N_i(N_i-1)}{n_i}$$

$$\hat{\bar{Y}}_U=\bar{y}_{st}+t_{.025,\,n_e}\sqrt{v(\bar{y}_{st})_{Prop.}}\qquad \therefore\ g_1=g_2=72$$

$$=\bar{y}_{st}+t_{.025,8}\sqrt{0.19}\qquad \therefore\ n_e=8$$

$$=10.08+2.306\sqrt{0.19}=11.08$$

প্রতিস্তর হতে নমুনা আকার বড় হলে $t_{.025}$-এর পরিবর্তে $z_{.025}$-এর মান বসাতে হয়।

Neyman বণ্টন : এখানে

$$n_1=\frac{nN_1S_1}{\Sigma N_iS_i}=\frac{12\times 24\times 1.47}{24\times 1.47+24\times 4.57}=\frac{423.36}{144.96}=3$$

$$n_2=\frac{nN_2S_2}{\Sigma N_iS_i}=\frac{12\times 24\times 4.57}{144.96}=9\quad [\text{অথবা, } n_2=n-3=9]$$

প্রথমোক্ত ক্ষেত্রে যেভাবে নমুনা চয়ন করা হয়েছে অনুরূপভাবে চয়ন করা নমুনা হলো:

স্তর-1	দৈব সংখ্যা :	10	22	24						
	y_{1j} :	10	10	8						
স্তর-2	দৈব সংখ্যা :	01	07	02	05	09	17	14	13	16
	y_{2j} :	12	9	14	8	11	11	10	10	9

তাহলে, $\bar{y}_1 = \frac{1}{n_1}\,\Sigma y_{1j} = \frac{28}{3} = 9.33$

$$s_1^2 = \frac{1}{n_1-1}\left[\Sigma y_{1j}^2 - \frac{(\Sigma y_{1j})^2}{n_1}\right] = \frac{1}{2}\left[264 - \frac{(28)^2}{3}\right] = 1.33$$

$$\bar{y}_2 = \frac{1}{n_2}\,\Sigma\, y_{2j} = \frac{94}{9} = 10.44$$

$$s_2^2 = = \frac{1}{n_2-1}\left[\Sigma\, y_{2j}^2 - \frac{(\Sigma y_{2j})^2}{n_2}\right] = \frac{1}{8}\left[1008 - \frac{(94)^2}{9}\right] = 3.28$$

সুতরাং $\bar{y}_{st} = \frac{1}{N}\,\Sigma N_i \bar{y}_i = \frac{1}{48}\left[24 \times 9.33 + 24 \times 10.44\right] = 9.88$

এই নিরূপকের ভেদাঙ্কের নিরূপক হলো

$$v(\bar{y}_{st})_{opt} = \frac{1}{n}\,(\Sigma W_i s_i)^2 - \frac{1}{N}\,\Sigma W_i s_i^2$$

এখানে $W_1 = \frac{N_1}{N} = \frac{24}{48} = 0.5$, $W_2 = \frac{N_2}{N} = \frac{24}{48} = 0.5$

$$s_1 = 1.13, \qquad s_2 = 1.81$$

$$v(\bar{y}_{st})_{opt} = \frac{1}{12}\left[0.5 \times 1.13 + 0.5 \times 1.81\right]^2 - \frac{1}{48}\left[0.5 \times 1.33 + 0.5 \times 3.28\right]$$

$$= \frac{2.1609}{12} - \frac{2.305}{48} = 0.132$$

স্তরিত নমুনায়নের ক্ষেত্রে প্রতি স্তরের নমুনায়ন এককগুলির মধ্যে ভেদাঙ্ক কম হওয়ার কারণে নমুনা নিরূপক অধিক যথাযথ হয়। এই নমুনায়নে একটি অসুবিধা হলো যে স্তর গঠন করা সহজ নয়।

৮.৬ Quota নমুনায়ন

স্তরিত নমুনায়নের ক্ষেত্রে পূর্ব অভিজ্ঞতা না থাকলে স্তর গঠন সহজ নয়। তাই স্তর থেকে দৈব নমুনায়নও সহজ নয়। এরূপ ক্ষেত্রে বিভিন্ন স্তর থেকে নমুনায়নের মাধ্যমে তথ্য সংগ্রহ করতে হলে i-তম স্তর থেকে n_i এককের $[i = 1, 2, \ldots, L]$ তথ্য সংগ্রহ করার জন্য উপাত্ত সংগ্রহকারীকে (enumerator) নির্দেশ দেয়া হয়। উপাত্ত সংগ্রহকারী জরিপ চলাকালীন সময়ে বিভিন্ন একক হতে তথ্য সংগ্রহ করে এককগুলিকে তাদের স্তরভুক্তির জন্য পূর্ব নির্ধারিত কোনো নির্দেশকের মাধ্যমে স্তরে বিভক্ত করে থাকে। i-তম স্তরে n_i একক হতে উপাত্ত সংগ্রহ না হওয়া পর্যন্ত জরিপ কাজ চলতে থাকে। এই পদ্ধতিতে নমুনায়ন করাকে বলে Quota নমুনায়ন।

এই নমুনায়ন করা হয় যখন i-তম স্তরের ফ্রেম না থাকে। তবে এই নমুনায়নের ক্ষেত্রে নমুনায়নের কোনো তত্ত্ব প্রয়োগ করা যায় না। কিন্তু বাস্তবে অনেক ক্ষেত্রেই এই নমুনায়ন সহজ। ধরা যাক, একই পরিবারভুক্ত কোনো মাছের ছয়টি প্রজাতি আছে এবং বিভিন্ন প্রজাতির বৈশিষ্ট্য বিভিন্ন। এক্ষেত্রে প্রজাতি অনুযায়ী মাছকে স্তরে ভাগ করা বা প্রজাতি অনুযায়ী মাছের ফ্রেম পাওয়া সহজ নয়। তাই কোনো বিশেষ বৈশিষ্ট্য পর্যালোচনা করার জন্য কোন প্রজাতির কয়টি মাছ হতে তথ্য সংগ্রহ করা হবে তা আগেই ঠিক করে নিতে হয় এবং জরিপ চলাকালীন সময়ে তথ্য সংগ্রহ করে মাছকে প্রজাতি অনুযায়ী ভাগ করে নিতে হয়। এভাবে বিশেষ প্রজাতির বিশেষ সংখ্যক মাছ হতে তথ্য সংগৃহীত না হওয়া পর্যন্ত জরিপ কাজ চলতে থাকে।

৮.৭ ধারাবাহিক নমুনায়ন (Systematic Sampling)

সরল দৈব নমুনায়নের ক্ষেত্রে নমুনার প্রতিটি একককে দৈব পদ্ধতিতে চয়ন করা হয়। অনেক সময় প্রথম একটি একক দৈব পদ্ধতিতে চয়ন করে পরবর্তী এককগুলিকে একটি নির্দিষ্ট ব্যবধানে চয়ন করা হয়। এই চয়ন পদ্ধতি একটি নির্দিষ্ট ধারায় হয় বলে একে ধারাবাহিক নমুনায়ন বলা হয়। এই নমুনায়নের ক্ষেত্রে n আকারের নমুনার জন্য গণসমষ্টি একককে প্রথমে n গুচ্ছে ভাগ করা হয় এবং প্রতি গুচ্ছ থেকে একটি একক (প্রথমটি ভিন্ন) ধারাবাহিকভাবে চয়ন করা হয়।

ধরা যাক কোনো গণসমষ্টিতে N একক আছে এবং এই গণসমষ্টি হতে n আকারের নমুনা চয়ন করতে হবে। তাহলে পুরো গণসমষ্টির এককগুলিকে n গুচ্ছে এমনভাবে ভাগ করতে হবে যেন প্রতিগুচ্ছে k সংখ্যক একক থাকে। অর্থাৎ $N = nk$

হয় এবং k একটি গোটা সংখ্যা হয়। অবশ্য গুচ্ছায়ন বিবেচনা করার সময় শেষ গুচ্ছের মান k থেকে ছোট বা বড় হতে পারে। যেমন, $N = 499$ এবং $n = 50$, তাহলে $k = 10$। এক্ষেত্রে প্রথম 49 গুচ্ছেরই $k = 10$ কিন্তু শেষ গুচ্ছের $k = 9$। আবার $N = 502$ হলে শেষ গুচ্ছের $k = 12$। এরূপ গুচ্ছায়নের বিবেচনা করা হলেও বাস্তব ক্ষেত্রে গণসমষ্টির এককগুলিকে 1 হতে N পর্যন্ত সংখ্যা দ্বারা চিহ্নিত করা হয় এবং প্রথম 1 হতে k পর্যন্ত একক হতে একটি একক দৈব পদ্ধতিতে চয়ন করা হয়। পরবর্তী এককগুলিকে ধারাবাহিকভাবে পূর্বে চয়ন করা যে কোনো এককের k সংখ্যক দূরত্বে চয়ন করা হয়। অর্থাৎ প্রথম এককটি i-তম একক হলে $[i = 1, 2, \ldots, k]$ দ্বিতীয়, তৃতীয়, ... n-তম একক হবে গণসমষ্টিতে $(i + k)$-তম, $(i + 2k)$-তম, ... $\{i + (n - 1)k\}$-তম একক।

এই নমুনায়নের ক্ষেত্রে i-এর মান শেষ গুচ্ছের k-এর মান অপেক্ষা বড় হলে নমুনায় অন্তর্ভুক্ত এককের সংখ্যা হবে $(n - 1)$ এবং ছোট হলে নমুনা আকার হবে n বা $(n + 1)$। অবশ্য N যদি k-এর গোটা সংখ্যা গুণিতক হয়, তাহলে নমুনা এককের এই সমস্যা থাকে না। উদাহরণ হিসেবে উদাহরণ ৮.১-এ দেয়া উপাত্ত হতে $n = 10$ আকারের একটি নমুনা চয়ন করা যাক। ঐ উপাত্তে $N = 75$, তাহলে $nk = N$ হওয়ার কারণে $k = 8$ হতে হয়। এক্ষেত্রে শেষ গুচ্ছে এককের সংখ্যা হবে 3। এখন নমুনা চয়ন করার জন্য 1 হতে 8 রাশির মধ্য হতে একটি রাশি দৈব পদ্ধতিতে চয়ন করতে হবে। এখানে পরিশিষ্টে দেয়া দৈব সংখ্যা সারণি অনুযায়ী প্রথম দৈব সংখ্যাটি 1। তাহলে নমুনায় অন্তর্ভুক্ত এককসমূহ হবে

1st, 9th, 17th, 25th. 33rd, 41st, 49th, 57th, 65th. 73rd

অর্থাৎ চয়ন করা নমুনাটি হবে

y_j : 150, 12, 24, 20, 48, 29, 18, 28, 48, 96

কিন্তু প্রথম চয়ন করা দৈব সংখ্যাটি 4 হলে ($i = 4$, শেষ গুচ্ছের k অপেক্ষা বড়) নমুনায়ন অন্তর্ভুক্ত এককগুলি হবে

4th, 12th, 20th, 28th. 36th, 44th, 52nd, 60th, 68th।

এখানে $n = 9$ এবং চয়ন করা নমুনা এককের মান (y_j)-গুলি হবে নিম্নরূপ :

y_j : 51, 22, 45, 117, 52. 39, 26, 60, 56

ধারাবাহিক নমুনায়নের ক্ষেত্রে গণসমষ্টি গড় $\overline{Y}$-এর নিরূপক হলো

$$\bar{y}_{sys} = \frac{1}{n} \sum_{j=1}^{n} y_j$$

এবং এই নিরূপকের ভেদাঙ্কের নিরূপক হলো

$$\text{এখানে} \quad v(\bar{y}_{sys}) = \frac{N-n}{Nn}\, s^2 \text{ , এখানে } s^2 = \frac{1}{n-1} \sum_{j=1}^{n} (y_j - \bar{y}_{sys})^2$$

ভেদাঙ্কের এই নিরূপক নির্ণয়ের শর্ত হলো যে গণসমষ্টিতে নমুনায়ন এককগুলি দৈবক্রমে থাকবে। কোনো বিশেষ একক বিশেষ ক্রম অনুসারে থাকবে না। তাহলে প্রথম এককটি দৈব পদ্ধতিতে চয়ন করার কারণে সমগ্র নমুনাটি দৈব নমুনা হিসেবে চিহ্নিত হতে পারে এবং সেক্ষেত্রে ভেদাঙ্কের নিরূপক যথাযথ হবে।

এখানে প্রথম নমুনার ভিত্তিতে উপজেলায় মোট কাঁঠাল গাছের সংখ্যার নিরূপক হলো

$$Y_{sys} = N\bar{y}_{sys} \quad , \qquad \bar{y}_{sys} = \frac{1}{n}\Sigma y_j = \frac{473}{10}$$

$$= 75 \times 47.3 \qquad\qquad = 47.3$$

$$= 3548$$

আবার, $$s^2 = \frac{1}{n-1}\left\{\Sigma y_j^2 - \frac{(\Sigma y_j)^2}{n}\right\} = \frac{1}{9}\left\{39393 - \frac{(473)^2}{10}\right\}$$

$$= 1891.12$$

তাহলে, $$v(\hat{Y}_{sys}) = N^2\ v(\bar{y}_{sys}) = \frac{N(N-n)}{n}\, s^2$$

$$= \frac{75(75-10)}{10} \times 1891.12 = 921921$$

এই নমুনায়নের একটি সুবিধা হলো যে ফ্রেম আধুনিক হলে অল্প সময়ে বড় আকারের [n বড়] নমুনা চয়ন করা যায়। তবে গণসমষ্টিতে নমুনায়ন এককগুলি দৈবক্রমে না থাকলে এই নমুনায়ন পদ্ধতি প্রয়োগ করে নমুনা নিরূপকের ভেদাঙ্কের নিরূপক সরাসরি পাওয়া যায় না।

সুবিধা এবং অসুবিধা যাই থাকুক না কেন বাস্তব ক্ষেত্রে এই নমুনায়নের প্রয়োগ বেশি হয়ে থাকে। যেমন :

(i) জীববিজ্ঞানে মাটির তাপমাত্রা পরপর কয়েকদিন নথিভুক্ত হলে তা ধারাবাহিক নমুনায়নের পর্যায়ভুক্ত।

(ii) ভূমিতে কোনো পোকামাকড়-এর সংখ্যা নিরূপণের জন্য একটি নির্দিষ্ট দূরত্বে নির্দিষ্ট পরিমাণ জায়গায় কত পোকামাকড় আছে তা নথিভুক্ত করা হলে ঐ নমুনায়নকে ধারাবাহিক নমুনায়ন বলা হয়।

(iii) জমিতে কোনো বিশেষ শস্যের চারা গজাবার ক্ষমতা পর্যালোচনা করার জন্য একটি নির্দিষ্ট দূরত্বে নির্দিষ্ট পরিমাণ জায়গায় কত বীজ লাগানো হয়েছে এবং ঐ বীজ হতে কত চারা গজালো তা নথিভুক্ত করা হলে তা ধারাবাহিক নমুনায়নের পর্যায়ভুক্ত।

(iv) কোনো এলাকায় উৎপন্ন ফুলের বা ফলের পরিমাণ নিরূপণ করার জন্য একটি নির্দিষ্ট দূরত্বে নির্দিষ্ট পরিমাণ জায়গায় [আবাদকৃত জমিতে] কত ফুল বা ফল উৎপন্ন হয়েছে তা নথিভুক্ত করা হলে ঐ নমুনায়নকে ধারাবাহিক নমুনায়ন বলা হয়।

৮.৮ গুচ্ছ নমুনায়ন (Cluster Sampling)

ধারাবাহিক নমুনায়নের ক্ষেত্রে গণসমষ্টি একককে n গুচ্ছে বিভক্ত করার কথা উল্লেখ করা হয়েছে। বাস্তবে গুচ্ছে বিভক্ত করতে হয় না। শুধু k একক বিশিষ্ট n গুচ্ছ বিবেচনা করতে হয়। অনেক ক্ষেত্রে গণসমষ্টি একককে কতকগুলি চিহ্নিত গুচ্ছে বিভক্ত করা হয়। গুচ্ছের এককের সংখ্যা সমান হতেও পারে আবার অসমানও হতে পারে। যেমন, একটি সামুদ্রিক মাছের কিছু বৈশিষ্ট্য পর্যালোচনা করা হবে সিদ্ধান্ত গৃহীত হয়েছে। ঐ সামুদ্রিক মাছ কোনো এক মাছ বাজারের বিভিন্ন দোকানদার ক্রয় করে থাকে এবং প্রতি দোকানদারের কাছে অনেকগুলি মাছ থাকে। সেক্ষেত্রে একজন দোকানদারের বিশেষ একটি জাতের সকল মাছ নিয়ে একটি গুচ্ছ হতে পারে। কোনো বনে কি পরিমাণ শাল গাছ আছে তা নিরূপণ করার জন্য সমগ্র বনকে কতকগুলি চিহ্নিত ছোট ছোট এলাকায় বিভক্ত করে নেয়া হলে এক একটি ছোট এলাকাকে গুচ্ছ বিবেচনা করা যেতে পারে। উপরিউক্ত পদ্ধতিতে গুচ্ছে বিভক্ত করে ঐ গুচ্ছসমূহ হতে কিছু গুচ্ছ দৈব পদ্ধতিতে চয়ন করা হলে ঐ নমুনায়নকে গুচ্ছ নমুনায়ন বলা হয়। এই নমুনায়নের ক্ষেত্রে গুচ্ছের প্রতিটি একক হতে উপাত্ত সংগ্রহ করা হয়।

ধরা যাক কোনো গণসমষ্টিতে M_0 একক আছে এবং এককগুলি N গুচ্ছে বিভক্ত, যেখানে i-তম গুচ্ছে এককের সংখ্যা $[\, i = 1, 2, \ldots, N \,]$ M_i। তাহলে

$$M_0 = \sum_{i=1}^{N} M_i$$

। এই N গুচ্ছ হতে n গুচ্ছ দৈব পদ্ধতিতে চয়ন করতে হবে।

ধরা যাক i-তম চয়ন করা গুচ্ছের j-তম এককের উপাত্ত হলো y_{ij} $(i = 1, 2, \ldots, n;$ $j = 1, 2, \ldots, M_i)$। তাহলে গণসমষ্টি গড়,

$$\bar{Y} = \frac{1}{M_0} \sum_{i}^{N} \sum_{j}^{M_i} y_{ij}$$

$$= \frac{1}{M_0} \sum_{i=1}^{N} M_i \bar{Y}_i$$

এখানে $\bar{Y}_i = \frac{1}{M_i} \sum_{j=1}^{M_i} y_{ij}$ = i-তম গুচ্ছের গণসমষ্টি গড়। তাহলে

$$\bar{Y} = \frac{1}{\bar{M}} \sum \frac{M_i \bar{Y}_i}{N}, \quad \text{এখানে} \quad \bar{M} = \frac{M_0}{N} = \frac{\Sigma M_i}{N}$$

$$= \frac{1}{N} \Sigma \frac{M_i}{\bar{M}} \bar{Y}_i = \frac{1}{N} \sum_{i=1}^{N} u_i \bar{Y}_i, \quad u_i = \frac{M_i}{\bar{M}}$$

কিন্তু $M_1 = M_2 = \ldots = M_N = M$ হলে গণসমষ্টিতে মোট এককের সংখ্যা হবে NM। সেক্ষেত্রে

$$\bar{Y} = \frac{1}{NM} \sum_{i}^{N} \sum_{j}^{M} y_{ij}$$

$$= \frac{1}{N} \Sigma \bar{Y}_i, \quad \bar{Y}_i = \frac{1}{M} \sum_{j=1}^{M} y_{ij}$$

এই $\bar{Y}$-এর নির্ঝুঁকি নিরূপক হলো

$$\bar{y}_c = \frac{N}{nM_0} \sum_{i=1}^{n} M_i \bar{Y}_i = \frac{1}{n\bar{M}} \sum_{i=1}^{n} M_i \bar{Y}_i = \frac{1}{n} \sum_{i=1}^{n} u_i \bar{Y}_i$$

কিন্তু $M_1 = M_2 = \cdots = M_N = M$ হলে $M_i = \bar{M}$ এবং $u_i = 1$

$$\text{সেক্ষেত্রে } \bar{y}_c = \frac{1}{n}\sum_{i=1}^{n}\bar{Y}_i$$

এই নিরূপকের ভেদাঙ্ক হলো

$$V(\bar{y}_c) = \frac{N-n}{Nn}S_b^2 \quad , \qquad S_b^2 = \frac{1}{N-1}\sum_{i}^{N}\left(\frac{M_i\bar{Y}_i}{\bar{M}} - \bar{Y}\right)^2$$

এই ভেদাঙ্কের নিরূপক হলো

$$v(\bar{y}_c) = \frac{N-n}{Nn}s_b^2 \ ; \ s_b^2 = \frac{1}{n-1}\sum_{i}^{n}\left(\frac{M_i\bar{Y}_i}{\bar{M}} - y_c\right)^2$$

$$= \frac{1-f}{n}s_b^2$$

$$= \frac{s_b^2}{n} \text{ , যদি } f = \frac{n}{N} \text{ খুব ছোট হয় ।}$$

গুচ্ছের আকার সমান হলে $[M_1 = M_2 = \cdots\cdots = M_N]$

$$s_b^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\bar{Y}_i - \bar{y}_c)^2$$

উদাহরণ ৮.৪

সারণি ৬.২৩-এ দেয়া Serranidae family ভুক্ত মাছের Centrum-এর গড় দৈর্ঘ্য পরিমাপ করার জন্য গুচ্ছ নমুনায়ন করা যাক।

উপরিউক্ত সারণিতে যেহেতু এক পরিবারভুক্ত মাছের উপাত্ত দেয়া আছে [Centrum-এর দৈর্ঘ্য x_i], সে কারণে ঐ পরিবারভুক্ত মাছগুলিকে তাদের প্রজাতি অনুযায়ী গুচ্ছে বিভক্ত বিবেচনা করা যাক। তাহলে, মোট গুচ্ছ সংখ্যা, $N=6$ এবং $M_1 = 11$, $M_2 = 12$, $M_3 = 12$, $M_4 = 5$, $M_5 = 7$, $M_6 = 15$ এবং $M_0 = 62$। এখন 6 গুচ্ছ হতে ধরা যাক $n = 3$ গুচ্ছ দৈব পদ্ধতিতে চয়ন করতে হবে। পরিশিষ্ট ১-এ দেয়া দৈব সংখ্যা সারণি অনুযায়ী চয়ন করা দৈব সংখ্যা হলো 1,2, এবং 4। অর্থাৎ প্রথম গুচ্ছ, দ্বিতীয় গুচ্ছ এবং চতুর্থ গুচ্ছ নমুনায় অন্তর্ভুক্ত। এখানে নমুনায় অন্তর্ভুক্ত এককসমূহের তথ্য নিচে দেয়া হলো :

গুচ্ছ	এককের মান y_{ij}	গড় $\overline{Y}_i$
1	4.42, 4.36, 4.59, 4.63, 4.74, 4.92, 5.06, 5.23, 5.42, 5.67, 5.74	4.98
2	10.11, 10.10, 10.04, 10.32, 10.64, 11.10, 11.04, 11.64 11.82, 12.27, 12.03, 10.88	11.00
4	4.65, 4.51, 4.40, 4.46, 4.58	4.52

সুতরাং মাছ প্রতি Centrum-এর গড় দৈর্ঘ্যের নিরূপক হলো

$$\overline{y}_c = \frac{N}{nM_0}\sum_{i=1}^{n} M_i\overline{Y}_i = \frac{6}{3\times 62}[11\times 4.98 + 12\times 11.00 + 5\times 4.52]$$

$$= 6.75$$

আবার,

$$s_b^2 = \frac{1}{n-1}\sum\left(\frac{M_i\overline{Y}_i}{\overline{M}} - \overline{y}_c\right)^2 \quad . \quad \overline{M} = \frac{M_0}{N} = \frac{62}{6} = 10.33$$

$$= \frac{1}{3-1}\left[\left(\frac{11\times 4.98}{10.33} - 6.75\right)^2 + \left(\frac{12\times 11.00}{10.33} - 6.75\right)^2 + \left(\frac{5\times 4.52}{10.33} - 6.75\right)^2\right] = 29.624$$

$$\therefore \qquad v(\overline{y}_c) = \frac{N-n}{Nn}s_b^2 = \frac{6-3}{6\times 3}\times 29.624 = 4.94$$

নবম অধ্যায়

উৎপাদক বিশ্লেষণ
(Factor Analysis)

৯.১ সূচনা

অনেকগুলি আন্তঃসম্পর্কীয় চলকের সম্পর্ক উপস্থাপন করার জন্য অল্প সংখ্যক উৎপাদক (factor) চিহ্নিত করার পরিসংখ্যানিক পদ্ধতি হলো উৎপাদক বিশ্লেষণ। এই অল্প সংখ্যক উৎপাদকে চিহ্নিত করার মূল ধারণা হলো যে আদি চলকসমূহের (original variables) মধ্যে এমন কতকগুলি সাধারণ উৎপাদক থাকে যেগুলির মান সহজে গবেষক সংগ্রহ করতে পারেন না অথচ ঐ উৎপাদকগুলির মান থাকে। যেমন, কোনো শস্য বীজের অঙ্কুরিত হওয়ার ক্ষমতা পর্যালোচনা করার জন্য সার, পানি সেচ, তাপমাত্রা, জমির পরিমাণ ইত্যাদি নিয়ন্ত্রণ করে কোনো পরীক্ষা পরিচালনা করা হলে লক্ষ্য করা যাবে যে, অঙ্কুরিত বীজের ভেদ অন্যান্য চলকের সাথে সম্পর্কিত। এখানে অঙ্কুরিত বীজের সংখ্যা সার, পানি সেচ, তাপমাত্রা এবং জমির পরিমাণের সাথে সম্পর্কিত। এই চলকসমূহের মধ্যে সার এবং পানি সেচ জমির উর্বরা শক্তির উপর নির্ভরশীল। এখানে গবেষক উর্বরা শক্তির পরিমাণ পরিমাপ হয়ত করেন না বা করতে পারেন না। এখানে সার এবং পানি সেচ চলক দুটির মধ্যে উর্বরা শক্তি একটি সাধারণ উৎপাদক (common factor) এবং এই সাধারণ উৎপাদক ছাড়াও চলকদ্বয়ের মানের পরিবর্তন অন্য কোনো উৎপাদকের কারণে হতে পারে যে উৎপাদককে বলা হয় লেটেন্ট উৎপাদক (latent factor)। উৎপাদক বিশ্লেষণের উদ্দেশ্য হলো পরিসংখ্যানিক পদ্ধতিতে সাধারণ উৎপাদকগুলি চিহ্নিত করা এবং অল্পসংখ্যক উৎপাদকের মাধ্যমে চলকগুলির আন্তঃসম্পর্ক উপস্থাপন করা।

উৎপাদক বিশ্লেষণের ক্ষেত্রে বিবেচনা করা হয় যে, সাধারণ উৎপাদকের তথ্য সরাসরি সংগ্রহ করা যায় না। যে চলকের তথ্য সংগ্রহ করা হয় তার মান সাধারণ উৎপাদকের মানের উপর এবং একটি দৈব বিচ্যুতির উপর নির্ভরশীল। অর্থাৎ প্রতিটি চলকই কতকগুলি সাধারণ উৎপাদক এবং দৈব বিচ্যুতির ফাংশন। এখানে চলক, সাধারণ উৎপাদক এবং দৈব বিচ্যুতির মধ্যে একটি রৈখিক ফাংশন বিবেচনা করে উৎপাদক বিশ্লেষণ আলোচনা করা হবে। এই রৈখিক ফাংশনের সহগগুলিই নির্দেশ

করে কোন চলকের ক্ষেত্রে কোন সাধারণ উৎপাদকের গুরুত্ব কতটুকু। এই সহগ নিরূপণ এমনভাবে করা হয় যেন চিহ্নিত সাধারণ উৎপাদকগুলি আদি চলকসমূহের ভেদের একটি বৃহত্তর অংশ ব্যাখ্যা করতে পারে। এটি সম্ভব হলেই যে কোনো পরিসংখ্যানিক পর্যালোচনায় নমুনায়ন এককের যে কোনো বৈশিষ্ট্যের ভেদের জন্য কোন কোন চলক গুরুত্বপূর্ণ সে সম্পর্কে সিদ্ধান্ত নেয়া যায়।

এই বিশ্লেষণ করতে গিয়ে লক্ষ্য করা যায় যে, যে কোনো সাধারণ উৎপাদকই আদি চলকসমূহের রৈখিক ফাংশন। ফলে রৈখিক ফাংশনের ক্ষেত্রে চলকের বৃহত্তম সহগই (চিহ্নবর্জিত) উপাত্তের ভেদ ব্যাখ্যা করার জন্য বেশি গুরুত্বপূর্ণ।

৯.২ উৎপাদক গঠনের গাণিতিক প্রতিকৃতি (Mathematical Model for Factor Structure)

ধরা যাক n আকারের নমুনার প্রতি নমুনা একক হতে $x_1, x_2, \ldots, x_p$ চলকগুলির মান নথিভুক্ত হয়েছে এবং এই চলকগুলি পরস্পর সংশ্লেষিত। বিবেচনা করা যাক যে এই চলকগুলির মধ্যে m সাধারণ উৎপাদক (common factor) w_j আছে $(j=1, 2, \ldots m)$ এবং i-তম $(i=1, 2, \ldots, p)$ চলকের ক্ষেত্রে e_i হলো একক উৎপাদক (unique factor) যার মান গবেষক সংগ্রহ করতে পারেন না বা পারা যায় না। এখানে চলক হলো w_j-সমূহের এবং e_i-এর রৈখিক সংযোগ, সুতরাং লেখা যায়

$$
\begin{aligned}
x_1 &= \lambda_{11} w_1 + \lambda_{12} w_2 + \cdots + \lambda_{1m} w_m + e_1 \\
x_2 &= \lambda_{21} w_1 + \lambda_{22} w_2 + \cdots + \lambda_{2m} w_m + e_2 \\
&\cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \qquad (৯.২.১) \\
x_p &= \lambda_{p1} w_1 + \lambda_{p2} w_2 + \cdots + \lambda_{pm} w_m + e_m
\end{aligned}
$$

এখানে λ_{ij} = i-তম চলকের ক্ষেত্রে j-তম সাধারণ উৎপাদকের গুরুত্ব নির্দেশকারী পরিমাণ। এই λ_{ij}-কে $[i=1,2, \cdots, p;\ j=1,2, \cdots, m]$ বলা হয় উৎপাদক ভার (factor loading)

উৎপাদক বিশ্লেষণের মূল উদ্দেশ্য হলো λ_{ij}-এর মান এমনভাবে নিরূপণ করা যেন m সাধারণ উৎপাদক $w_1, w_2, \ldots, w_m$ চলক $x_1, x_2, \ldots, x_p$-এর ভেদের একটি বৃহত্তম অংশ ব্যাখ্যা করতে পারে। এখানে λ_{ij} হলো i-তম চলক ও j-তম সাধারণ উৎপাদকের সংশ্লেষাঙ্ক (correlation coefficient)। সুতরাং λ_{ij}-এর মান থেকেই বলা যায় কোন চলক কোন সাধারণ উৎপাদকের জন্য কত বেশি গুরুত্বপূর্ণ। অন্যভাবে বলা যায় উপাত্তের বৈশিষ্ট্য পর্যালোচনা করার জন্য কোন চলক বেশি গুরুত্বপূর্ণ।

বাস্তব ক্ষেত্রে λ_{ij} হলো $x_1, x_2, \cdots, x_p$ চলকসমূহের সংশ্লেষাঙ্ক ম্যাট্রিক্স হতে প্রাপ্ত i-তম আইগেন মানের প্রাসঙ্গিক আইগেন ভেক্টরের j-তম মান। এই মান পাওয়ার জন্য প্রাথমিকভাবে প্রধান উপাদান বিশ্লেষণ (principal component analysis) পদ্ধতি প্রয়োগ করা হয়। এই পদ্ধতির জন্য অনুমান হলো ঃ

$$E(x_i) = 0, \quad E(w_j) = 0, \quad V(w_j) = 1, \quad Cov.(w_j, w_j') = 0,$$
$$Cov(x_i, w_j) = E(x_i, w_j) = \lambda_{ij}, \quad E(e_i) = 0, \quad V(e_i) = \psi_i,$$
$$Cov(w_j, e_i) = 0. \quad V(x_i) = \sigma_i^2 \text{ ।}$$

এছাড়া অনুমান করা হয় যে, x_i-গুলি এবং w_j-গুলি বহুচলক পরিমিত বিন্যাস (Multivariate normal distribution) অনুসরণ করে। অবশ্য এই শেষোক্ত অনুমান বহাল না থাকলে উৎপাদক বিশ্লেষণ যে অর্থহীন হয়ে যাবে তা নয়।

৯.৩ উৎপাদক বিশ্লেষণের ধাপসমূহ (Steps of Factor Analysis)

উৎপাদক বিশ্লেষণের একটি উদ্দেশ্য হলো এমন কতকগুলি উৎপাদক (factors) চিহ্নিত করা যেগুলি চলকসমূহের সংশ্লেষণ ব্যাখ্যা করতে পারে। সে কারণে উৎপাদক বিশ্লেষণের শুরুতেই চলকসমূহের সংশ্লেষণ তাৎপর্যপূর্ণ কিনা যাচাই করা দরকার। কারণ সংশ্লেষণ তাৎপর্যপূর্ণ হলে চলকগুলির মধ্যে সাধারণ উৎপাদক থাকার সম্ভাবনা কম।

ধরা যাক চলক $x_1, x_2, \ldots, x_p$-এর ক্ষেত্রে নমুনা সংশ্লেষণ ম্যাট্রিক্স হলো

$$R = \begin{pmatrix} 1 & r_{12} & r_{13} & \cdots & r_{1p} \\ r_{21} & 1 & r_{23} & \cdots & r_{2p} \\ \cdots & \cdots & \cdots & & \cdots \\ r_{p1} & r_{p2} & r_{p3} & \cdots & 1 \end{pmatrix}_{P \times P}$$

যেখানে

$$r_{ij} = \frac{\Sigma x_i x_j - \frac{\Sigma x_i \, \Sigma x_j}{n}}{\sqrt{\left[\Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n}\right]\left[\Sigma x_j^2 - \frac{(\Sigma x_j)^2}{n}\right]}}, \quad i \neq j = 1, 2, \cdots, P$$

এই চলকগুলির গণসমষ্টি সংশ্লেষণ ম্যাট্রিক্স ধরা যাক P। তাহলে যাচাই করতে হবে নাস্তি কল্পনা

$$H_0: \quad P = I \qquad (৯.৩.১)$$

এবং বিকল্প কল্পনা $H_0: \quad P \neq I$

এখানে I হলো আইডেনটিটি ম্যাট্রিক্স এবং নাস্তি কল্পনা দ্বারা বুঝা যায় চলকগুলির মধ্যে কোনো সংশ্লেষণ নেই। এই নাস্তি কল্পনার জন্য যাচাই তথ্যজমান হলো

$$-2\ ln\ \lambda = n\ ln\ |R| \qquad (৯.৩.২)$$

n বড় হলে $-2ln\lambda$-এর বিন্যাস হলো কাইবর্গ বিন্যাস যার স্বাধীনতার মাত্রা $P(P-1)/2$। Box (1949) দেখিয়েছেন যে χ^2-এর বিন্যাস খুব ভাল হয় যদি n-এর পরিবর্তে $n-(2P+11)/6$ ব্যবহার করা হয়। উপরিউক্ত নাস্তি কল্পনা (৯.৩.১) বাতিল হলে উপাত্ত উৎপাদক বিশ্লেষণের জন্য উপযুক্ত বলে বিবেচিত হয়।

উপাত্ত উৎপাদক বিশ্লেষণের জন্য উপযুক্ত কিনা তা Kaiser – Meyer – Olkin (KMO) তথ্যজমান নির্ণয় করেও সিদ্ধান্ত নেয়া হয়। এখানে

$$KMO = \frac{\sum_{i \neq j} \sum^{P} r_{ij}^2}{\sum_{i \neq j} \sum^{P} r_{ij}^2 + \sum_{i \neq j} \sum^{P} a_{ij}^2} \qquad (৯.৩.৩)$$

যেখানে r_{ij} = i-তম ও j-তম চলকের মধ্যে সংশ্লেষাঙ্ক

a_{ij} = i-তম ও j-তম চলকের মধ্যে আংশিক সংশ্লেষাঙ্ক

Kaiser উল্লেখ করেছেন যে KMO = 0.90 হলে উৎপাদক বিশ্লেষণ ফলপ্রসূ; KMO = 0.80, উৎপাদক বিশ্লেষণ ভাল, KMO = 0.70, বিশ্লেষণ ভাল নয়; KMO = 0.60, বিশ্লেষণ করা যেতে পারে; KMO = 0.50, উৎপাদক বিশ্লেষণ হতে কোনো ভাল তথ্য পাওয়া যায় না; KMO < 0.50, উৎপাদক বিশ্লেষণ গ্রহণযোগ্য নয়।

এতক্ষণ উৎপাদক বিশ্লেষণের জন্য প্রাপ্ত সকল চলকের উপযুক্ততা নিয়ে আলোচনা করা হয়েছে। এই বিশ্লেষণে কোনো একটি বিশেষ চলক উপযুক্ত কিনা সে সম্পর্কে সিদ্ধান্ত নেয়ার জন্য Measure of Sampling Adequacy [MSA] নির্ণয় করা হয়। i-তম চলকের জন্য MSA_i নির্ণয় করার সূত্র হলো

$$MSA_i = \frac{\sum_{j \neq i} r_{ij}^2}{\sum_{j \neq i} r_{ij}^2 + \sum_{j \neq i} a_{ij}^2} \qquad (৯.৩.৪)$$

কোনো চলকের ক্ষেত্রে $MSA_i < 0.50$ হলে ঐ চলকটি বিশ্লেষণ হতে বাদ দেয়া উচিত। এখানে MSA_i-এর মান হলো Anti-Image Correlation Matrix-এর কৌণিক মানসমূহ (diagonal elements)। বাস্তব ক্ষেত্রে বিশ্লেষণের জন্য MSA ও MSA_i-এর মান নির্ণয় করার জন্য অনেক Computer program প্রয়োগ করা যায়। ফলে পরিসংখ্যানের ছাত্র বা গবেষক না হলেও এই বিশ্লেষণ যে কোনো গবেষক করতে পারেন।

আগেই উল্লেখ করা হয়েছে যে P চলক হতে m উৎপাদক চিহ্নিত করাই হলো উৎপাদক বিশ্লেষণের মূল উদ্দেশ্য। এই m-এর মান সম্পর্কে সিদ্ধান্ত নেয়ার ব্যাপারে Jeffers (1967), Jolliffe (1972, 1973) মতামত ব্যক্ত করেছেন। তাঁদের মতে $\lambda_i \leqslant 0.70$ হলে ঐগুলির প্রাসঙ্গিক উৎপাদক পরবর্তী বিশ্লেষণ হতে বাদ দেয়া যেতে পারে। অর্থাৎ ঐসমস্ত উৎপাদকই বিবেচিত হবে যেগুলি আইগেন মান $\lambda_i > 0.70$ এর প্রাসঙ্গিক। এখানে λ_i হলো R ম্যাট্রিক্স-এর i-তম আইগেন মান। অবশ্য কোনো কোনো গবেষকের মতে যে উৎপাদকগুলি চলকসমূহের ভেদের 90% ব্যাখ্যা করতে পারে ঐগুলিই চিহ্নিত উৎপাদক। এখানে j-তম উৎপাদক

$$\frac{\lambda_j}{\sum_{j=1}^{p} \lambda_j} \times 100\%$$

ভেদ ব্যাখ্যা করতে পারে। ফলে

$$\sum_{j}^{m} \frac{\lambda_j}{\Sigma \lambda_j} = 0.90$$

হলে m উৎপাদক উপাত্তের ভেদ ব্যাখ্যা করার জন্য যথেষ্ট। সাধারণত $\lambda_1 > \lambda_2 > \cdots > \lambda_p$ এবং λ_1-ই উপাত্তের ভেদের বৃহত্তর অংশ ব্যাখ্যা করে থাকে। ফলে λ_1-এর প্রাসঙ্গিক উৎপাদক ব্যবহার করে গবেষক উপাত্তের মধ্যে কোন চলকটি বা কোন

চলকগুলি অধিক প্রভাবশালী সে সম্পর্কে সিদ্ধান্ত নিতে পারেন। আগেই উল্লেখ করা হয়েছে যে প্রভাবশালী চলক চিহ্নিত করা হয় λ_{ij}-এর বৃহত্তম মানের ভিত্তিতে। এখানে λ_{ij} হলো i-তম আইগেন মানের ভিত্তিতে আইগেন ভেক্টরের j-তম মান

$$[\, i \neq j = 1, 2, \cdots, P \,]$$

উদাহরণ ৯.১

উদাহরণ ৬.১২-এ দেয়া উপাত্তের ভিত্তিতে Vertebrae-এর বৈশিষ্ট্য পর্যালোচনা করার জন্য কোন চলকগুলি গুরুত্বপূর্ণ তা পর্যবেক্ষণ করার জন্য উৎপাদক বিশ্লেষণ করা যাক।

উপরিউক্ত উদাহরণের ক্ষেত্রে নমুনা সংশ্লেষাঙ্ক ম্যাট্রিক্স R নিচে দেয়া হলো :

$R =$

x	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
x_1	1.00									
x_2	0.96	1.00								
x_3	0.96	0.99	1.00							
x_4	0.95	0.99	0.99	1.00						
x_5	0.75	0.66	0.67	0.67	1.00					
x_6	0.77	0.69	0.69	0.69	0.89	1.00				
x_7	0.70	0.82	0.83	0.85	0.51	0.49	1.00			
x_8	0.39	0.21	0.21	0.18	0.59	0.49	−0.03	1.00		
x_9	−0.22	−0.02	−0.01	0.04	−0.06	−0.12	0.36	−0.45	1.00	
x_{10}	−0.08	0.02	0.04	0.06	−0.13	−0.18	0.33	−0.31	0.41	1.00

এখানে $n = 62$ এবং $-2\, ln\lambda = ln \mid n \mid R \mid = 1223.84$ এবং $-2 ln\lambda$-এর বিন্যাস 45 স্বাধীনতার মাত্রাবিশিষ্ট χ^2 বিন্যাস $[P(\chi^2 > 1223.84) = 0.000]$। সুতরাং চলকসমূহ তাৎপর্যপূর্ণভাবে সংশ্লেষিত। এই বিশ্লেষণের ক্ষেত্রে KMO = 0.83 হওয়ায় উৎপাদক বিশ্লেষণ ভালভাবে করা যাবে। নিচে Anti-Image Correlation Matrix দেয়া হলো :

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
x_1	0.84									
x_2	-0.18	0.84								
x_3	-0.26	-0.67	0.87							
x_4	0.05	-0.49	-0.20	0.90						
x_5	-0.41	0.15	-0.07	0.10	0.78					
x_6	-0.11	0.11	0.10	-0.28	-0.61	0.86				
x_7	0.24	0.26	-0.31	-0.24	-0.16	0.10	0.89			
x_8	-0.30	0.04	0.13	-0.05	-0.33	0.15	-0.04	0.82		
x_9	0.64	-0.04	0.04	-0.31	-0.47	0.08	-0.10	0.08	0.43	
x_{10}	-0.09	0.21	-0.12	-0.06	0.06	0.12	-0.29	0.06	-0.09	0.70

উপরিউক্ত ম্যাট্রিক্সের কৌণিক মানসমূহ হতে লক্ষ্য করা যাচ্ছে যে, চলক x_9-এর ক্ষেত্রেই MSA-এর মান 0.50 অপেক্ষা কিছুটা ছোট। এই চলকটি বিশ্লেষণ হতে বাদ দেয়া যেতে পারতো। এখানে MSA = 0.43 খুব বেশি ছোট না হওয়ায় একে বিশ্লেষণ হতে বাদ না দিয়েই পরবর্তী বিশ্লেষণসমূহ করা হয়েছে।

এখন দশটি চলক নিয়েই উৎপাদক বিশ্লেষণ করা হলো। এই বিশ্লেষণের জন্য প্রধান উপাদান বিশ্লেষণ (principal component analysis) পদ্ধতি প্রয়োগ করা হয়েছে। বিশ্লেষণের প্রাথমিক ফলাফল সারণি ৯.১-এ উপস্থাপন করা হলো। লক্ষ্য করা যাচ্ছে যে প্রথম তিনটি আইগেন মান 0.70 অপেক্ষা বড় এবং প্রথম তিনটি উৎপাদক (factors) উপাত্তের ভেদের মোট 88.5% ব্যাখ্যা করতে পারে। সুতরাং এখানে তিনটি উৎপাদককে চিহ্নিত করা যেতে পারতো। কিন্তু প্রধান উপাদান বিশ্লেষণ পদ্ধতি প্রয়োগ করাতে $\lambda \geq 1$-এর প্রাসঙ্গিক দুটি উৎপাদক চিহ্নিত করা হয়েছে। এই দুটি উৎপাদকের উৎপাদক ভার ম্যাট্রিক্স সারণি ৯.২-এ উপস্থাপন করা হলো।

সারণি ৯.১ : উৎপাদক বিশ্লেষণের প্রাথমিক তথ্যজমানসমূহ ।

চলক	কম্যুনালিটি	উৎপাদক	আইগেন মান	ভেদের %	ভেদের ক্রমযোজিত %
x_1	1.000	1	5.894	58.9	58.9
x_2	1.000	2	2.147	21.5	80.4
x_3	1.000	3	0.806	8.1	88.5
x_4	1.000	4	0.647	6.5	95.0
x_5	1.000	5	0.297	3.0	97.9
x_6	1.000	6	0.110	1.1	99.0
x_7	1.000	7	0.086	0.9	99.9
x_8	1.000	8	0.008	0.1	100.0
x_9	1.000	9	0.002	0.0	100.0
x_{10}	1.000	10	0.001	0.0	100.0

লক্ষ্য করা যাচ্ছে যে, প্রথম উৎপাদক উপাত্তের ভেদের 58.9% ব্যাখ্যা করতে পারে এবং এই ভেদ ব্যাখ্যা করার জন্য চলক x_1, x_2, x_3, x_4, x_5, x_6 এবং x_7 প্রভাবশালী। এগুলির মধ্যে x_1, x_2 x_3 এবং x_4 খুবই প্রভাবশালী। অর্থাৎ Vertebrae-এর বৈশিষ্ট্য পর্যালোচনা করার জন্য Centrum-এর দৈর্ঘ্য (x_1), Anterior হতে নেয়া Centrum-এর প্রশস্ততা (x_2), Posterior হতে নেয়া Centrum-এর প্রশস্ততা (x_3) এবং Centrum-এর উচ্চতা (x_4) খুবই গুরুত্বপূর্ণ। দ্বিতীয় উৎপাদক চলকের ভেদের 21.5% ব্যাখ্যা করতে পারে এবং এই উৎপাদকের ক্ষেত্রে Posterior

সারণি ৯.২ : উৎপাদক ভার ম্যাট্রিক্স এবং নিরূপিত কম্যুনালিটি ।

চলক	উৎপাদক ভার [λ_{ij}]		কম্যুনালিটি
	উৎপাদক-1 (λ_{1j})	উৎপাদক-2 (λ_{2j})	[নিরূপিত]
x_1	0.9727	−0.0939	0.9550
x_2	0.9635	0.1356	0.9467
x_3	0.9664	0.1442	0.9547
x_4	0.9632	0.1850	0.9620

x_5	0.8270	−0.2615	0.7523
x_6	0.8333	−0.2736	0.7692
x_7	0.7944	0.5243	0.9060
x_8	0.3815	−0.7206	0.6648
x_9	−0.0520	0.7755	0.6041
x_{10}	−0.0192	0.7253	0.5264

Zygapophysis-এর দৈর্ঘ্য (x_8), Neural spine এবং Longitudinal axis-এর মধ্যে কোণ (x_9) ও Haemal spine এবং Longitudinal axis-এর মধ্যে কোণ (x_{10}) খুবই গুরুত্বপূর্ণ। এখানে চিহ্নিত উৎপাদক দুটি x_1 চলকের ভেদের 95.5%, x_2 চলকের ভেদের 94.67%, x_3 চলকের ভেদের 95.47%, x_4 চলকের ভেদের 96.2% ইত্যাদি ব্যাখ্যা করতে পারে। এই ব্যাখ্যা করা ভেদের পরিমাণকে বলা হয়

নিরূপিত কম্যুনালিটি $\left[\text{কম্যুনালিটি} = \sum_{j=1}^{m} \lambda_{ij}^2 \right]$

আগেই উল্লেখ করা হয়েছে যে, চলকগুলির সংশ্লেষণ ঐগুলির মধ্যে সাধারণ উৎপাদক থাকার কারণেই হয়। সুতরাং উৎপাদক এবং চলকের মধ্যে সংশ্লেষাঙ্ক চলকগুলির সংশ্লেষাঙ্ক নিরূপণ করার জন্য ব্যবহৃত হতে পারে। প্রধান উৎপাদক পদ্ধতি প্রয়োগ করে বিশ্লেষণ করাতে উৎপাদক ভারগুলি (factor loadings) উৎপাদক এবং চলকগুলির মধ্যে সংশ্লেষাঙ্ক নির্দেশ করে। ধরা ঐ সংশ্লেষাঙ্ক হলো r_{fi}। এখন i-তম ও j-তম চলকের সংশ্লেষাঙ্ক হলো

$$r_{ij} = \sum_{f=1}^{m} r_{fi}\, r_{fj} = \sum_{f=1}^{m} \lambda_{fi}\, \lambda_{fj}$$

এখানে r_{fi} = i-তম চলক ও f-তম উৎপাদকের মধ্যে সংশ্লেষাঙ্ক।

r_{fj} = j-তম চলক ও f-তম উৎপাদকের মধ্যে সংশ্লেষাঙ্ক।

আলোচিত উদাহরণের ক্ষেত্রে

$$r_{11} = \lambda_{11}\, \lambda_{11} + \lambda_{21}\, \lambda_{21} = 0.9550$$

$$r_{12} = \lambda_{11}\, \lambda_{12} + \lambda_{21}\, \lambda_{22} = 0.9245$$

r_{ij}-এর অন্যান্য মানগুলি নিচে উপস্থাপন করা হলো।

সারণি ৯.৩ : নিরূপিত সংশ্লেষাঙ্ক ম্যাট্রিক্স।

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
x_1	0.95									
x_2	0.92	0.95								
x_3	0.93	0.95	0.95							
x_4	0.92	0.95	0.96	0.96						
x_5	0.83	0.76	0.76	0.75	0.75					
x_6	0.84	0.76	0.76	0.75	0.76	0.77				
x_7	0.72	0.84	0.84	0.86	0.52	0.52	0.91			
x_8	0.44	0.27	0.26	0.23	0.50	0.51	−0.07	0.66		
x_9	−0.12	0.05	0.06	0.09	−0.24	−0.25	0.36	−0.58	0.60	
x_{10}	−0.09	0.08	0.09	0.11	−0.20	−0.21	0.36	−0.53	0.56	0.53

আগেই উল্লেখ করা হয়েছে যে প্রথম উৎপাদক চলকসমূহের ভেদের বৃহত্তম অংশ ব্যাখ্যা করে এবং এই উৎপাদক চলক x_1, x_2, x_3 ও x_4 দ্বারা বেশি প্রভাবিত। এই চলক চারটির মধ্যে x_3 ও x_4 বেশি সংশ্লেষিত। অতঃপর বেশি মাত্রায় সংশ্লেষিত চলক জোড়া হলো x_1 ও x_3।

উৎপাদক বিশ্লেষণ হতে চলকের গুরুত্ব নির্দেশ করার জন্য উৎপাদক ভার $[\lambda_{ij}]$ ব্যবহার করা হয়। i-তম উৎপাদকের ক্ষেত্রে j-তম চলক গুরুত্বপূর্ণ বিবেচিত হবে যদি λ_{ij} বড় হয়। সুতরাং উৎপাদক বিশ্লেষণ হতে সুফল পেতে হলে নিম্নলিখিত বিষয়গুলি বিবেচ্য [thurstone (1947)]।

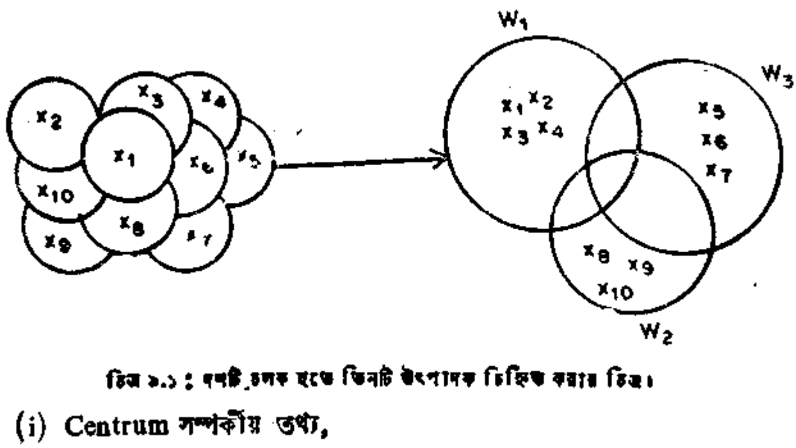
(i) উৎপাদক ভার ম্যাট্রিক্সের যে কোনো স্তম্ভের মানগুলির বেশিরভাগ মান শূন্যের কাছাকাছি হওয়া উচিত।

(ii) উৎপাদক ভার ম্যাট্রিক্সের যে কোনো সারির মানগুলির বেশিরভাগ মান শূন্যের কাছাকাছি হওয়া উচিত।

(iii) উৎপাদক ভার ম্যাট্রিক্সের যে কোনো দুটি স্তম্ভের একটির মান ছোট হলে অন্যটির মান বড় হওয়া উচিত।

উপরিউক্ত বিষয়গুলি লক্ষণীয় না হলে উৎপাদকের রোটেশন (rotation) করতে হয়। রোটেশন করার বিভিন্ন পদ্ধতি আছে এবং ঐ পদ্ধতিগুলি প্রয়োগ করার জন্য কম্পিউটার প্রোগ্রামও আছে। এখানে এ ব্যাপারে আলোচনা করা হলো না। ইচ্ছুক গবেষক বা ছাত্র/ছাত্রীগণ Dillon and Goldstein (1984)-এর বই পর্যালোচনা করতে পারেন।

রোটেশন করার উদ্দেশ্য হলো উৎপাদকসমূহ চিহ্নিত করতে যেন কোনো বিতর্কের সৃষ্টি না হয়। সাধারণত উৎপাদকের সাথে যে চলকগুলির সংশ্লেষণ বেশি সেগুলি দ্বারা উৎপাদক চিহ্নিত হয়। আলোচিত উদাহরণের ক্ষেত্রে লক্ষ্য করা যাচ্ছে যে x_1, x_2, x_3 এবং x_4 চলক প্রথম উৎপাদকের সাথে বেশি সংশ্লেষিত (সারণি ৯.২)। দ্বিতীয় উৎপাদকের ক্ষেত্রে চলক x_8, x_9 এবং x_{10} বেশি সংশ্লেষিত। প্রধান উৎপাদক পদ্ধতি প্রয়োগ করার কারণে এখানে আর কোনো উৎপাদক দেখানো হয়নি। প্রথম দুটি উৎপাদকই চলকের ভেদের 80.4% ব্যাখ্যা করতে পারে। তৃতীয় উৎপাদক চলকের ভেদের মান 8.1% ব্যাখ্যা করতে পারে বলে এটি তেমন গুরুত্বপূর্ণ নয়। যাহোক কোন কোন চলক দ্বারা উৎপাদকগুলি গঠিত হতে পারে তা নিচে চিত্র ৯.১-এ দেখানো হলো। লক্ষণীয় বিষয় হলো Vertebrae-এর বৈশিষ্ট্য পর্যালোচনা করার জন্য দশটি চলকের পরিবর্তে দুটি উৎপাদকই যথেষ্ট। এই দুটি উৎপাদক হলো



চিত্র ৯.১ : দশটি চলক হতে তিনটি উৎপাদক চিহ্নিত করার চিত্র।

(i) Centrum সম্পর্কীয় তথ্য,

(ii) Spine দ্বারা সৃষ্ট কোন অন্য একটি উৎপাদকও বিবেচিত হতে পারে তা।

(iii) Spine এবং Posterior face পর্যন্ত দূরত্ব।

দশম অধ্যায়

নির্ণায়ক বিশ্লেষণ
(Discriminant Analysis)

১০.১ সূচনা

পূর্বাভাস করা যায় এমন কতকগুলি চলক গুচ্ছের ভিত্তিতে এক বা একাধিক নমুনা বিন্দুকে পরস্পর বিসদৃশ (mutually exclusive) এবং সর্বসম্বলিত (exhaustive) গুচ্ছে শ্রেণী বিভক্ত করার পরিসংখ্যানিক পদ্ধতিকে নির্ণায়ক বিশ্লেষণ (Discriminant analysis) বলা হয়। এই বিশ্লেষণের ভিত্তি হলো একটি গুণগত (qualitative) চলকের ভিত্তিতে কতগুলি অনপেক্ষ চলকের (independent variables) মাধ্যমে নমুনা বিন্দুকে শ্রেণীবিভক্ত করা। এখানে গুণগত চলকটি নির্ভরশীল চলক হিসেবে ব্যবহৃত হয়। এই বিশ্লেষণের মূল উদ্দেশ্য হলো পরীক্ষাগারে কোনো জীব বা প্রাণীর কতগুলি বৈশিষ্ট্য পর্যবেক্ষণ করে ঐগুলিকে শ্রেণী বিভক্ত করা বা ঐ প্রাণীগুলি পর্যালোচিত বৈশিষ্ট্যের ভিত্তিতে পূর্বনির্ধারিত শ্রেণীতে অন্তর্ভুক্ত হওয়ার সঠিকতা পর্যালোচনা করা। যেমন, সারণি ১.১ (প্রথম খণ্ড)-এ দেয়া Slug-এর দৈহিক বৈশিষ্ট্যের পাঁচটি পরিমাপ Body length (x in mm), Body weight (x_2 in gm), Mantle length (x_3 in mm), Keel length (x_4 in mm) এবং Shell length (x_5 in mm) পর্যালোচনা করে Slug গুলিকে শ্রেণী বিভক্ত করার পরিসংখ্যানিক পদ্ধতি হলো নির্ণায়ক বিশ্লেষণ। এখানে Slug গুলিকে পূর্ব নির্ধারিত চারটি ভাগে বিভক্ত করা আছে। এই বিভক্তিকরণের জন্য একটি গুণগত নির্ভরশীল চলক (x_6) ব্যবহৃত হয়েছে। এখন চলক x_1, x_2, x_3, x_4 এবং x_5-এর ভিত্তিতে যে কোনো একটি Slug তার জন্য পূর্ব নির্ধারিত শ্রেণীতে অন্তর্ভুক্ত হয়েছে কিনা তা পর্যালোচনা করার পরিসংখ্যানিক পদ্ধতি হলো নির্ণায়ক বিশ্লেষণ।

১০.২ নির্ণায়ক বিশ্লেষণের বিভিন্ন ধাপ (Steps of Discriminant Analysis)

ধরা যাক কোনো গণসমষ্টি হতে n_1 এবং n_2 আকারের দুটি নমুনা চয়ন করা হয়েছে বা $n_1 + n_2 = n$ আকারের নমুনা উপাত্ত সংগ্রহ করে উপাত্তের কোনো একটি বৈশিষ্ট্য অনুসারে নমুনা বিন্দুগুলিকে দুটি ভাগে বিভক্ত করা হয়েছে, যেখানে প্রথম ভাগের

নমুনা বিন্দুর সংখ্যা n_1 এবং দ্বিতীয় ভাগের নমুনা বিন্দুর সংখ্যা n_2। প্রথম এবং দ্বিতীয় নমুনা হতে $x_1, x_2, \cdots, x_p$ চলকের মান নথিভুক্ত হয়েছে বিবেচনা করা যাক। ধরা যাক i-তম $[i = 1, 2]$ নমুনার ক্ষেত্রে j-তম $[i = 1, 2, \ldots, p]$ চলকের l-তম মান $(l = 1, 2, \ldots, n_1)$ হলো x_{ijl}। এখানে l-তম নমুনার ক্ষেত্রে j-তম নমুনা গড় হলো

$$\bar{x}_{ij} = \frac{1}{n_i} \sum_{l=1}^{n_i} x_{ijl} \qquad i = 1.2;\ j = 1, 2 \cdots, p$$

নমুনা ভেদাঙ্ক হলো

$$S_{ijj} = \frac{1}{n_i - 1} \sum_{l} (x_{ijl} - \bar{x}_{ij})^2$$

এবং নমুনাসহ ভেদাঙ্ক হলো

$$S_{ijj'} = \frac{1}{n_i - 1} \sum_{l} (x_{ijl} - \bar{x}_{ij})\,(x_{ij'l} - \bar{x}_{ij'}),\ j \neq j'$$

এখন প্রথম ও দ্বিতীয় নমুনার ক্ষেত্রে নমুনা গড় এবং নমুনা ভেদাঙ্কসহ ভেদাঙ্কগুলিকে যথাক্রমে ভেক্টর ও ম্যাট্রিক্স আকারে লেখা যায় :

$$\bar{X}_1 = \begin{bmatrix} \bar{x}_{11} \\ \bar{x}_{12} \\ \vdots \\ \bar{x}_{1p} \end{bmatrix}_{p \times 1}, \quad S_1 = \begin{bmatrix} S_{111} & S_{112} & S_{113} & \cdots & S_{11p} \\ S_{121} & S_{122} & S_{123} & \cdots & S_{12p} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ S_{1p1} & S_{1p2} & \cdots & \cdots & S_{1pp} \end{bmatrix}_{p \times p}$$

এবং

$$\bar{X}_2 = \begin{bmatrix} \bar{x}_{21} \\ \bar{x}_{22} \\ \vdots \\ \bar{x}_{2p} \end{bmatrix}, \quad S_2 = \begin{bmatrix} S_{211} & S_{212} & S_{213} & \cdots & S_{21p} \\ S_{221} & S_{222} & S_{223} & \cdots & S_{22p} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ S_{2p1} & S_{2p2} & \cdots & \cdots & S_{2pp} \end{bmatrix}_{p \times p}$$

এখানে $\bar{X}_1$ এবং $\bar{X}_2$ হলো যথাক্রমে

$$\mu_1 = \begin{pmatrix} \mu_{11} \\ \mu_{12} \\ \vdots \\ \mu_{1p} \end{pmatrix} \text{ এবং } \mu_2 = \begin{bmatrix} \mu_{21} \\ \mu_{22} \\ \vdots \\ \mu_{2p} \end{bmatrix}$$ -এর নিরূপক। আবার S_1 এবং S_2 হলো যথাক্রমে Σ_1 ও Σ_2-এর নিরূপক।

এই বিশ্লেষণের জন্য অনুমান করা হয় যে,

(i) p অনপেক্ষ চলকের বিন্যাস হলো বহুচলক পরিমিত বিন্যাস (Multivariate normal distribution)

(ii) $p \times p$ সহ ভেদাঙ্ক ম্যাট্রিক্স

$$\Sigma_1 = \Sigma_2 = \Sigma \text{ (ধরা যাক)}$$

অবশ্য শেষোক্ত অনুমান অনেক উপাত্তের ক্ষেত্রে বহাল নাও থাকতে পারে। এ সম্পর্কে নাস্তি কল্পনা

$$H_0 : \Sigma_1 = \Sigma_2 \tag{১০.২.১}$$

যাচাই করার জন্য Box-এর M-যাচাই পদ্ধতি প্রয়োগ করা হয়। এখানে

$$M = n - k)\, ln \mid S \mid - \sum_{i=1}^{k} (n_i - 1)\, ln\ S_i\ ;\ n = \sum_{i=1}^{k} n_i$$

$$S = \frac{1}{n-k} \sum_{i=1}^{k} (n_i - 1) S_i$$

Box (1949) দেখিয়েছেন যে MC^{-1}-এর বিন্যাস $\frac{1}{2}(k-1)p(p+1)$ স্বাধীনতার মাত্রাসহ χ^2 বিন্যাস, যেখানে

$$C^{-1} = 1 - \frac{2p^2 + 3p - 1}{6(p+1)(k-1)} \left[\sum_{i=1}^{k} \frac{1}{n_i - 1} - \frac{1}{n-k} \right]$$

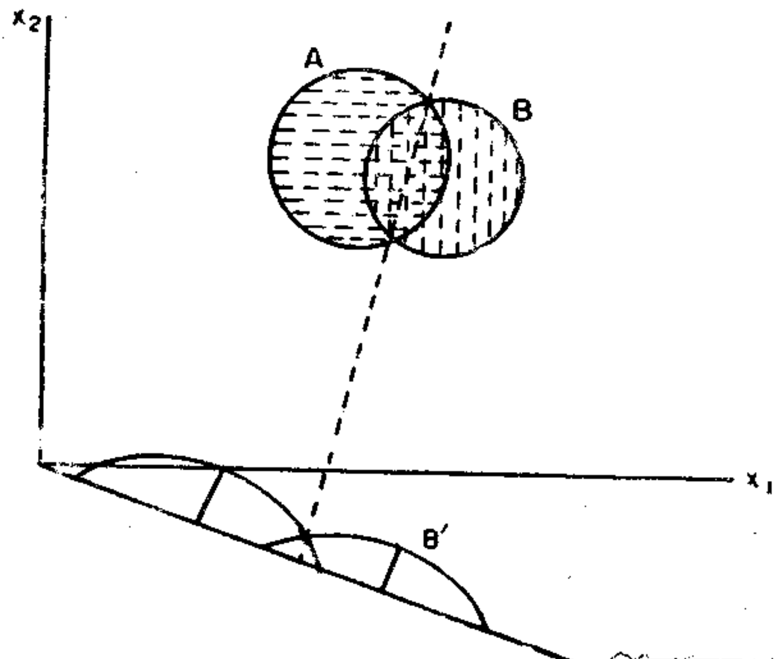
কোনো নমুনার ক্ষেত্রে $k = 2$ $[i = 1, 2]$ হলে

$$S = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2}{n_1 + n_2 - 2}$$

এই S হলো $\Sigma_1 = \Sigma_2 = \Sigma$-এর নিরূপক।

নাস্তি কল্পনা (১০.২.১) k নমুনার ক্ষেত্রেও প্রযোজ্য এবং এই নাস্তি কল্পনা যাচাই করার জন্য M যাচাই তথ্যজ্ঞান ব্যবহার করা যায়। অবশ্য নাস্তি কল্পনা বাতিল হলে নির্ণায়ক বিশ্লেষণ ক্ষতিগ্রস্ত হবে না। কারণ উপাত্ত বহুচলক পরিমিত বিন্যাস অনুসরণ না করলেও নাস্তি কল্পনা বাতিল হতে পারে। অপরপক্ষে উপাত্ত বহুচলক পরিমিত বিন্যাস অনুসরণ না করলেও রৈখিক এমনকি দ্বি-ঘাত নির্ণায়ক বিশ্লেষণ অর্থহীন হবে না (Huberty 1994)।

নির্ণায়ক বিশ্লেষণের কাজ হলো অনপেক্ষ চলক $x_1, x_2, \cdots, x_p$-এর একটি রৈখিক সংযোগ (linear combination) এমনভাবে নির্ণয় করা যেন ঐ রৈখিক সংযোগ পূর্ব চিহ্নিত গুচ্ছের নমুনা বিন্দুগুলির মধ্যে ন্যূনতম গুচ্ছায়ন বিচ্যুতিসহ পার্থক্য নির্দেশ করতে পারে। নিচে একটি চিত্রের সাহায্যে বিষয়টি ব্যাখ্যা করা যাক। ধরা যাক n আকারের নমুনা বিন্দুগুলির কিছু A গুচ্ছে এবং কিছু B গুচ্ছে



চিত্র ১০.১ : দুই গুচ্ছের জন্য নির্ণায়ক বিশ্লেষণের সচিত্র ব্যাখ্যা।

অন্তর্ভুক্ত। এই নমুনা বিন্দুগুলি হতে x_1 ও x_2 নামক দুটি চলকের মান নথিভুক্ত করা হয়েছে। চিত্রের উপরের অংশ A গুচ্ছের এবং B গুচ্ছের নমুনা বিন্দুগুলির x_1 ও x_2

এর মানের বিক্ষেপ চিত্র নির্দেশ করছে। A গুচ্ছের এবং B গুচ্ছের উপবৃত্ত যেখানে ছেদ করছে ঐ ছেদ বিন্দুগুলি দিয়ে যে রেখা অতিক্রম করে তা নতুন অক্ষরেখা y-তে প্রতিফলিত হয়। ঐ অক্ষে স্পর্শ করা ঐ রেখাটি A´ এবং B´ দুটি এক চলক বিন্যাসের সৃষ্টি করেছে বা চিত্রের নিম্নাংশে লক্ষণীয়। এখানে y-অক্ষ A গুচ্ছ এবং B গুচ্ছের মধ্যে পার্থক্য নির্দেশকারী নির্ণায়ক সাফল্যাঙ্ক (discriminant score) এবং এটি x_1 ও x_2 চলকের একটি রৈখিক সংযোগ। সুতরাং লক্ষ্য করা যাচ্ছে যে, দুই গুচ্ছের মধ্যে পার্থক্য নির্দেশ করার জন্য প্রাপ্ত চলকগুলির একটি রৈখিক ফাংশন নির্ণয় করা হলো নির্ণায়ক বিশ্লেষণের কাজ। ধরা যাক এরূপ একটি ফাংশন হলো

$$y_l = b_0 + b_1x_{1l} + b_2x_{2l} + \cdots + b_px_{pl} \qquad (১০.২.২)$$

এখানে y_l = l-তম নমুনা বিন্দুর নির্ণায়ক সাফল্যাঙ্ক এবং $b_1, b_2, \cdots, b_p$ হলো নির্ণায়ক ভর। অনেক সময় ফাংশন বিবেচনা করা হয়

$$y_l = b_1x_{1l} + b_2x_{2l} + \dots + b_px_{pl} \qquad (১০.২.৩)$$

এই গুচ্ছের মধ্যে পার্থক্য নির্দেশ করার জন্য নির্ণায়ক বিশ্লেষণ করা হলে b_1, $b_2, \cdots, b_p$-এর নিরূপক পাওয়ার সূত্র হলো

$$\hat{b} = \begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \vdots \\ \hat{b}_p \end{bmatrix} = S^{-1}[\bar{X}_1 - \bar{X}_2]$$

ধরা যাক $x_1, x_2, \cdots, x_p$ চলকের ভেক্টর হলো $X = [x_1\ x_2 \cdots x_p]'$। তাহলে কোনো নমুনা বিন্দু প্রথম গুচ্ছ A-এর অন্তর্ভুক্ত হবে, যদি

$$|\hat{b}'(X - \bar{X}_1)| \leqslant |\hat{b}'(X - \bar{X}_2)|$$

এবং তা দ্বিতীয় গুচ্ছ B-এর অন্তর্ভুক্ত হবে, যদি

$$|\hat{b}'(X - \bar{X}_2)| < |\hat{b}'(X - \bar{X}_1)|$$

এখানে X দ্বারা কোনো নমুনা বিন্দুর $x_1, x_2, \dots, x_p$ চলকের মানের একটি ভেক্টর বুঝানো হয়েছে এবং $\bar{X}_1$ and $\bar{X}_2$ হলো যথাক্রমে ঐ চলকগুলির প্রথম ও দ্বিতীয় গুচ্ছের ক্ষেত্রে প্রাপ্ত গড় মান।

উপরিউক্ত নিয়মে নমুনা বিন্দুগুলিকে শ্রেণীভুক্ত করার পদ্ধতিকে বলা হয় Fisher-এর নির্ণায়ক আইন। এই আইন Fisher-এর রৈখিক নির্ণায়ক ফাংশন-এর ভিত্তিতে প্রয়োগ করা হয়।

নির্ণায়ক ফাংশন-এর গড় মানকে বলা হয় গ্রুপ সেন্ট্রয়েড (Group centroid)। ধরা যাক i-তম গুচ্ছের [$i = 1, 2$] গ্রুপ সেন্ট্রয়েড হলো $\bar{Y}_1 = \hat{b}' \bar{X}_1$। এখন দুই গুচ্ছের মধ্যে নির্ণায়ক বিশ্লেষণ করা হলে দুই গুচ্ছের গ্রুপ সেন্ট্রয়েড-এর পার্থক্য হবে

$$\begin{aligned} \bar{Y}_1 - \bar{Y}_2 &= \hat{b}' \bar{X}_1 - \hat{b}' \bar{X}_2 \\ &= (\bar{X}_1 - \bar{X}_2)' S^{-1} (\bar{X}_1 - \bar{X}_2) \\ &= D^2 \end{aligned}$$

এই D^2 হলো Mahalanobis-এর দূরত্ব। এখন দুই গুচ্ছের মধ্যে তাৎপর্যপূর্ণ পার্থক্য আছে কিনা তা যাচাই করার জন্য নাস্তি কল্পনা

$$H_0 : \mu_1 = \mu_2 \qquad (১০.২.৪)$$

যাচাই করতে হয়। এখানে যাচাই তথ্যজমান হলো F, যেখানে

$$F = \frac{n_1 n_2}{n_1 + n_2} \cdot \frac{n_1 + n_2 - p - 1}{(n_1 + n_2 - 2)p} D^2 \qquad (১০.২.৫)$$

এই F-এর বিন্যাস হলো p এবং $(n_1 + n_2 - p - 1)$ স্বাধীনতার মাত্রাবিশিষ্ট ভেদাঙ্ক অনুপাত বিন্যাস। সুতরাং

$$F > F_{\alpha\,;\, p,\,(n_1 + n_2 - p - 1)}$$

হলে নাস্তি কল্পনা (১০.২.৪) α সংশয় মাত্রায় বাতিল বলে পরিগণিত হবে।

নাস্তি কল্পনা (১০.২.৪) বাতিল হলেই নির্ণায়ক বিশ্লেষণ অর্থবহ হয়। সেক্ষেত্রে দুই গুচ্ছের মধ্যে পার্থক্য নির্দেশকারী বিন্দু হলো

$$Y_C = \frac{\hat{b}' \bar{X}_1 + \hat{b}' \bar{X}_2}{2}\,; \text{ যদি } n_1 = n_2$$

কিন্তু $n_1 \neq n_2$ হলে

$$Y_C = \frac{n_2 \bar{Y}_1 + n_1 \bar{Y}_2}{n_1 + n_2}$$

বিশ্লেষণের এই পর্যায়ে $x_1, x_2, \cdots, x_p$ চলকগুলির মধ্যে কয়টি চলক দুটি গুচ্ছের মধ্যে পার্থক্য নির্দেশ করতে গুরুত্বপূর্ণ ভূমিকা রাখে তা পর্যালোচনা করা যায়। ধরা যাক $q < P$ চলক দুই গুচ্ছের মধ্যে পার্থক্য নির্দেশ করার জন্য গুরুত্বপূর্ণ। সেক্ষেত্রে এই q চলকের তাৎপর্যপূর্ণ গুরুত্ব যাচাই করার জন্য যাচাই তথ্যজমান হলো

$$Q = \frac{(n-P-1)C(D_P^2 - D_q^2)}{(P-q)(1+CD_q^2)} \qquad (১০.২.৬)$$

এখানে $D_P^2 = P$ চলকের ভিত্তিতে প্রাপ্ত Mahalanobis D^2

$D_q^2 = q$ চলকের ভিত্তিতে প্রাপ্ত Mahalanobis D^2

$$C = \frac{n_1 n_2}{(n_1+n_2)\ (n_1+n_2-2)}$$

এই Q-এর বিন্যাস হলো $(P-q)$ এবং (n_1+n_2-P-1) স্বাধীনতার মাত্রা-বিশিষ্ট F-বিন্যাস।

এতক্ষণ দুই গুচ্ছের মধ্যে পার্থক্য নির্দেশ করার জন্য নির্ণায়ক বিশ্লেষণ আলোচনা করা হয়েছে। এই বিশ্লেষণ k গুচ্ছের মধ্যে পার্থক্য নির্দেশ করার জন্য করা যায়। সেক্ষেত্রে মোট $(k-1)$ রৈখিক সংযোগ পাওয়া যায়। এই ফাংশন হলো

$$D_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p \qquad (১০.২.৭)$$

অথবা

$$D_i = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p \qquad (১০.২.৮)$$

$$i = 1,\ 2,\ \cdots,\ (k-1)$$

নিচে $\beta_0, \beta_1, \cdots, \beta_p$ নিরূপণের পদ্ধতি আলোচনা করা হলো। এই পরামান-গুলিকে বলা হয় নির্ণায়ক ভর। এগুলি গুচ্ছের মধ্যে পার্থক্য নির্দেশ করতে চলকের গুরুত্ব নির্দেশ করে [Blackith and Reyment (1971)]।

ধরা যাক কোনো গণসমষ্টি হতে n আকারের নমুনা চয়ন করা হয়েছে এবং এই নমুনাকে নমুনা বিন্দুর কোনো বৈশিষ্ট্য অনুসারে k গুচ্ছে বিভক্ত করা হয়েছে। ধরা যাক i-তম $[i=1, 2, \cdots, k]$ গুচ্ছের নমুনা বিন্দুর সংখ্যা n_i। প্রতি গুচ্ছের প্রতি নমুনা বিন্দু হতে P চলকের মান নথিভুক্ত করা যায়। যেমন সারণি ১.১ (প্রথম খণ্ড)-এ 50 Slug-এর প্রতিটি হতে পাঁচটি চলকের মান নথিভুক্ত করা হয়েছে এবং Slug গুলিকে চারটি প্রজাতিতে বিভক্ত করা হয়েছে। ধরা যাক i-তম গুচ্ছের j-তম চলকের l-তম মান হলো $[i=1, 2, \ldots, k;\ j=1, 2, \cdots, p;\ l=1, 2, \cdots, n_i]$ x_{ijl}। এই মানগুলিকে নিম্নরূপভাবে সাজিয়ে লেখা যায়।

সারণি ১০.১ : বিভিন্ন গুচ্ছভুক্ত নমুনা বিন্দুর বিভিন্ন চলকের মান $[x_{ijl}]$

গুচ্ছ	চলক	নমুনা বিন্দু					মোট $X_{ij.}$
	x_1	x_{111}	x_{112}	...	...	x_{11n_1}	$X_{11.}$
1	x_2	x_{121}	x_{122}	...	...	x_{12n_1}	$X_{12.}$
	$\vdots$						
	x_p	x_{1p1}	x_{1p2}	...	...	x_{1pn_1}	$X_{1p.}$
	x_1	x_{211}	x_{212}	...	...	x_{21n_2}	$X_{21.}$
2	$\vdots$						
	x_p	x_{2p1}	x_{2p2}	...	...	x_{2pn_2}	$X_{2p.}$
	$\vdots$						
	x_1	x_{k11}	x_{k12}	...	...	x_{k1n_k}	$X_{k1.}$
	x_2	x_{k21}	x_{k22}	...	...	x_{k2n_k}	$X_{k2.}$
k	$\vdots$						
	x_p	x_{kp1}	x_{kp2}	...	...	x_{kpn_k}	$X_{kp.}$

এই উপাত্তের ক্ষেত্রে

$$X_{.j.} = \sum_{i}^{k} X_{ij.}\,, \quad n = n_1 + n_2 + .. + n_k$$

মোট উপাত্তের ক্ষেত্রে সকল চলকের ভিত্তিতে ধরা যাক

$$T = \begin{bmatrix} SS(x_1) & SP(x_1x_2) & \cdots & SP(x_1x_p) \\ SP(x_2x_1) & SS(x_2) & \cdots & SP(x_2x_p) \\ \cdots & \cdots & \cdots & \cdots \\ SP(x_px_1) & SP(x_px_2) & \cdots & SS(x_p) \end{bmatrix}_{P \times P}$$

এখানে

$$SP(x_j\, x_{j'}) = \sum_{i}^{k} \sum_{l}^{n_j} x_{ijl}\, x_{ij'l} - \frac{X_{.j.}\, X_{.j'.}}{n}$$, j-এর সকল মানের জন্য ;

$j = 1, 2, .., P$ ।

আবার

$$B = \begin{bmatrix} SS(\bar{x}_1) & SP(\bar{x}_1\,\bar{x}_2) & \cdots & SP(\bar{x}_1\,\bar{x}_p) \\ SP(\bar{x}_2\bar{x}_p) & SS(\bar{x}_2) & \cdots & SP(\bar{x}_2\,\bar{x}_p) \\ \cdots & \cdots & \cdots & \cdots \\ SP(\bar{x}_p\,\bar{x}_1) & SP(\bar{x}_p\bar{x}_2) & \cdots & SS(\bar{x}_p) \end{bmatrix}_{P \times P}$$

$$SP(\bar{x}_j\,\bar{x}_{j'}) = \sum_{i=1}^{k} \frac{X_{ij.}\,X_{ij'.}}{n_i} - \frac{X_{.j.}\,X_{.j'.}}{n}$$, j-এর সকল মানের জন্য ; $j = 1,2, \ldots , P$ ।

ধরা যাক $W = T - B$ । এখন $W^{-1}B$ ম্যাট্রিক্স নির্ণয় করা যায়। এই ম্যাট্রিক্সের আইগেন মানসমূহ ভিন্ন ভিন্ন হলে মোট আইগেন মানের সংখ্যা হবে r, যেখানে r হলো $(k-1)$ ও P-এর মধ্যে যেটি ছোট $[r = \min(k-1, p)]$ এবং r নির্ণায়ক ফাংশন পাওয়া যাবে, যেখানে নির্ণায়ক সহগ বা নির্ণায়ক ভর হলো আইগেন মানসমূহের প্রাসঙ্গিক আইগেন ভেক্টর।

ধরা যাক $W^{-1}B$-এর S-তম আইগেন মান হলো λ_s $(S = 1,2,\ldots, r)$ । তাহলে S-তম নির্ণায়ক ফাংশন দ্বারা গুচ্ছসমূহের মধ্যে পার্থক্য নির্দেশ করতে উপাত্তের ভেদের কত অংশ ব্যাখ্যা করতে পারে তা নির্ণয় করা হয়

$$\frac{\lambda_s}{\sum_{S}^{r} \lambda_s}$$

দ্বারা। এই পর্যায়ে S-তম $[S = 1,2,\ldots,r]$ নির্ণায়ক ফাংশনের তাৎপর্য যাচাই করার তথ্যজমান হলো

$$V_s = [(n-1) - \tfrac{1}{2}(P+k)]\, ln(1-\lambda_s) \qquad (১০.২.৯)$$

এই V_s-এর বিন্যাস $(P + k - 2S)$ স্বাধীনতার মাত্রাসহ χ^2-বিন্যাস[Bartlett (1947)]। এছাড়া নির্ণায়ক ফাংশন দ্বারা নির্ণায়ক বিশ্লেষণ তাৎপর্যপূর্ণ কিনা তা যাচাই করার যাচাই তথ্যজমান হলো

$$V = -\{(n-1) - \tfrac{1}{2}(P+k)\, ln\ \Lambda$$, যেখানে

$$\Lambda = \prod_{S=1}^{r} (1+\lambda_s)^{-1}$$

এই V-এর বিন্যাস হলো $P(k-1)$ স্বাধীনতার মাত্রাসহ χ^2 বিন্যাস এবং $V > \chi^2_{0.05}$ হলে নির্ণায়ক বিশ্লেষণ তাৎপর্যপূর্ণ বলে বিবেচিত হবে।

নির্ণায়ক বিশ্লেষণ তাৎপর্যপূর্ণ বিবেচিত হওয়ার অর্থ হলো প্রথম ফাংশন তাৎপর্যপূর্ণ, যেখানে প্রথম ফাংশন $W^{-1}B$-এর বৃহত্তম আইগেন মানের (λ_1) সাথে সম্পর্কিত এবং এই ফাংশনের নির্ণায়ক সহগগুলি হলো λ_1-এর প্রাসঙ্গিক আইগেন ভেক্টর। দ্বিতীয় ফাংশনের তাৎপর্য যাচাই করার জন্য যাচাই তথ্যজমান হলো $(P-1)(k-2)$ স্বাধীনতার মাত্রাবিশিষ্ট χ^2-তথ্যজমান, যেখানে

$$\chi^2 = V - V_1$$

অনুরূপভাবে তৃতীয়, চতুর্থ এবং ... r-তম ফাংশনের তাৎপর্য যাচাই করার যাচাই তথ্যজমান ও ঐগুলির স্বাধীনতার মাত্রা নিচে যথাক্রমে দেয়া হলো:

ফাংশন	যাচাই তথ্যজমান	স্বাধীনতার মাত্রা
৩য়	$V - V_1 - V_2$	$(P-1)(k-3)$
৪র্থ	$V - V_1 - V_2 - V_3$	$(P-1)(k-4)$
...	...	
r-তম	$V - V_1 - V_2 - \cdots - V_{r-1}$	$(P-1)(k-r)$

এখানকার $V_1, V_2, .. V_{r-1}$ নির্ণয় করার জন্য সূত্র (১০.২.৯) প্রয়োগ করতে হয়।

S-তম নির্ণায়ক ফাংশন তাৎপর্যপূর্ণ হোক আর না হোক ঐ ফাংশনের জন্য নির্ণায়ক ভর হলো S-তম আইগেন মানের প্রাসঙ্গিক আইগেন ভেক্টর $[b_{Sj}\ ;\ S = 1,2, .., r\ ;\ j = 1,2, ..., P]$, এখানে b_{sj} হলো S-তম নির্ণায়ক ফাংশনের ক্ষেত্রে j-তম চলকের ভর যা গুচ্ছের মধ্যে পার্থক্য নির্দেশ করতে গুরুত্বপূর্ণ ভূমিকা পালন করে। এখানে নির্ণেয় S-তম ফাংশন হলো

$$\hat{D}_s = b_{s1}x_1 + b_{s2}x_2 + \dots + b_{sp}x_p \qquad (১০.২.১০)$$

এখন $S = 1, 2, \cdots, r$ যে কোনো নমুনা বিন্দুর $x_1, x_2, \ldots x_p$-এর মান বসিয়ে নির্ণায়ক সাফল্যাঙ্ক $\hat{D}_s$ নির্ণয় করা যায়। এই $\hat{D}_S$-এর মান প্রতি গুচ্ছের জন্য ভিন্ন ভিন্নভাবেও নির্ণয় করা যায় এবং গুচ্ছভুক্ত নির্ণেয় $\hat{D}_s$-এর মানের গড় নির্ণয় করা যায়। এই গড়কে বলা হয় গ্রুপ সেন্ট্রয়েড। কোনো ফাংশনের ক্ষেত্রে গ্রুপ সেন্ট্রয়েড্-এর মানের পার্থক্য যে দুই গুচ্ছের ক্ষেত্রে বেশি হবে ঐ ফাংশন সে দুই গুচ্ছের মধ্যে ভালভাবে পার্থক্য নির্দেশ করে। উপরে প্রাপ্ত ফাংশনের b_{sj}-এর মান পাওয়া

যায় যদি $x_1, x_2, \ldots, x_p$-এর প্রতিটি আদর্শায়িত হয় এবং সেক্ষেত্রে b_{gj}-কে বলা হয় আদর্শায়িত নির্ণায়ক সহগ [Standardized discriminant coefficients]। চলকগুলি আদর্শায়িত না হলে নির্ণেয় নির্ণায়ক ফাংশন হলো

$$\hat{D}_g = b_{g0} + b_{g1}x_1 + b_{g2}x_2 + \ldots + b_{gp}x_p \qquad (১০.২.১১)$$

এখানকার $b_{gj}(j = 0, 1, 2, \cdots, P)$-কে বলা হয় অ-আদর্শায়িত নির্ণায়ক সহগ (Unstandardized discriminant coefficients)।

অনেক সময় $x_1, x_2, \cdots, x_p$ চলকগুলি পরস্পর সংশ্লেষিত হয়। সেক্ষেত্রে নির্ণায়ক সহগ [ভর] চলকের সঠিক গুরুত্ব নির্দেশ করে না। এই সমস্যা দূর করার জন্য নির্ণায়ক ভার (discriminant loading) নির্ণয় করা হয়। এখানে নির্ণায়ক ভার হলো নির্ণায়ক সাফল্যাঙ্ক $[\hat{D}_g]$ এবং চলকের মানের সরল সংশ্লেষাঙ্ক (simple correlation coefficient)। এই সংশ্লেষাঙ্ক গুচ্ছের মধ্যে পার্থক্য নির্দেশ করতে চলকের গুরুত্বের একটি পরিমাপ। কোনো চলকের জন্য সংশ্লেষাঙ্ক বড় হলে ঐ চলক গুচ্ছের মধ্যে পার্থক্য নির্দেশ করতে বেশি গুরুত্বপূর্ণ।

উদাহরণ ১০.৯

সারণি ১.১ (প্রথম খণ্ড)-এ দেয়া উপাত্তের ভিত্তিতে চারপ্রকার Slug-এর মধ্যে পার্থক্য নির্দেশ করার জন্য নির্ণায়ক বিশ্লেষণ করা যাক।

উপরিউক্ত উপাত্তের ক্ষেত্রে চার গুচ্ছের চলকগুলির গড় সারণি ১০.২.১-এ ও পরিমিত ব্যবধান সারণি ১০.২.২-এ উপস্থাপন করা হলো।

সারণি ১০.২.১ : চার গুচ্ছ Slug-এর চলকগুলির গড়।

	Slug			
গড় চলক	M. Sowerbyi	M. Rusticus	M. Gagates	L. Tenellus
x_1	52.33	57.20	46.80	52.80
x_2	3.22	4.37	4.33	2.64
x_3	27.13	24.27	26.50	26.40
x_4	14.00	16.07	17.30	15.10
x_5	11.80	12.47	11.50	11.40

সারণি ১০.২.২ ঃ চার গুচ্ছ Slug-এর চলকগুলির পরিমিত ব্যবধান।

পরিমিত ব্যবধান	চলক	M. Sowerbyi	M. Rusticus	M. Gagates	L. Tenellus
	x_1	11.67	12.98	8.28	7.83
	x_2	1.67	2.13	2.44	1.06
	x_3	6.96	5.67	5.44	3.31
	x_4	3.48	4.38	5.29	2.64
	x_5	2.08	1.85	2.59	1.51

সারণি ১০.৩ : নির্ণায়ক বিশ্লেষণের প্রাথমিক ফলাফল।

নির্ণায়ক ফাংশন	আইগেন মান, λ_1	ব্যাখ্যা করা ভেদের শতকরা হার	Λ	χ^2	d.f.	P-মান
1	0.9886	74.73	0.3736	43.82	15	0.00
2	0.2937	22.20	0.7429	13.23	8	0.10
3	0.0405	3.06	0.9611	1.77	3	0.62

এই উপাত্ত চার গুচ্ছে বিভক্ত হওয়ার কারণে এখানে $r = 3$ [$r = \min(k-1, P) = \min(3,5)$; $k = 4$, $P = 5$]। অর্থাৎ তিনটি নির্ণায়ক ফাংশন পাওয়া যায়। প্রথম ফাংশন দ্বারা উপাত্তের 74.73% ভেদ করা যায় এবং এই ফাংশন পরিসংখ্যানিকভাবে তাৎপর্যপূর্ণ [$\chi^2 = 43.82$, $P = 0.00$]। দ্বিতীয় ফাংশন তাৎপর্যপূর্ণ না হলেও [$\chi^2 = 13.23$, $P = 0.10$] এটি উপাত্তের ভেদের 22.20% ব্যাখ্যা করতে পারে।

নির্ণায়ক ফাংশনগুলি কোন কোন Slug-এর ভালভাবে পার্থক্য নির্দেশ করতে পারে তা গ্রুপ সেন্ট্রয়েড হতে বুঝা যায়। নিচে তিনটি নির্ণায়ক ফাংশন হতে প্রাপ্ত গ্রুপ সেন্ট্রয়েড দেখানো হলো। লক্ষ্য করা যাচ্ছে যে প্রথম ফাংশন দ্বারা M. Gagates ও L. Tenellus-এর গ্রুপ সেন্ট্রয়েড-এর মানের দূরত্ব খুব বেশি। সুতরাং

সারণি ১০.৪ : নির্ণায়ক ফাংশন হতে প্রাপ্ত গ্রুপ সেন্ট্রয়েড।

Slug-এর প্রকার	ফাংশন প্রথম	দ্বিতীয়	তৃতীয়
M. Sowerbyi	−0.4362	0.2121	−0.2702
M. Rusticus	−0.0640	0.5719	0.2042
M. Gagates	1.7849	−0.3628	−0.0194
L. Tenellus	−1.0345	−0.8132	0.1184

প্রথম ফাংশন M. Gagates L. Tenellus-এর মধ্যে ভালভাবে পার্থক্য নির্দেশ করে। দ্বিতীয় ফাংশন এবং তৃতীয় ফাংশন তাৎপর্যপূর্ণ না হলেও ঐ দুটি যথাক্রমে M. Rusticus ও L. Tenellus এবং M. Sowerbyi ও M. Rusticus-এর মধ্যে পার্থক্য নির্দেশ করে। এই পার্থক্য নির্দেশ করার জন্য Slug-এর কোন বৈশিষ্ট্য গুরুত্বপূর্ণ ভূমিকা পালন করেছে তা নির্ণায়ক সহগ হতে লক্ষ্য করা যাক। সারণি ১০.৫-এ অ-আদর্শায়িত ও আদর্শায়িত নির্ণায়ক সহগগুলি উপস্থাপন করা হলো। লক্ষ্য করা যাচ্ছে যে, প্রথম ফাংশন M. Gagates ও L. Tenellus-এর মধ্যে পার্থক্য নির্দেশ করছে ভালভাবে। এই পার্থক্য নির্দেশ করার জন্য উক্ত দুই Slug-এর Body weight

সারণি ১০.৫ : নির্ণায়ক ফাংশন হতে নির্ণায়ক সহগসমূহ।

চলক	অ-আদর্শায়িত সহগ			আদর্শায়িত সহগ		
	ফাংশন					
	1	2	3	1	2	3
Body length	− 0.15	0.08	0.07	− 1.69	0.87	0.81
Body weight	1.09	1.20	− 0.09	2.07	2.27	− 0.15
Mantle length	− 0.06	− 0.14	− 0.10	− 0.33	− 0.81	− 0.58
Keel length	0.03	− 0.48	0.21	0.10	− 1.96	0.84
Shell length	− 0.08	− 0.15	− 0.01	− 0.17	− 0.30	− 0.03
Constant	6.39	4.59	− 3.89	—	—	—

গুরুত্বপূর্ণ ভূমিকা পালন করছে। দ্বিতীয় ফাংশন হতে বুঝা যাচ্ছে যে, M. Rusticus ও L. Tenellus-এর পার্থক্যও Body weight-এর জন্যই হচ্ছে। M. Sowerbyi ও M. Rusticus-এর মধ্যে পার্থক্য পরিলক্ষিত হচ্ছে মূলত Keel length-এর পার্থক্যের কারণে।

উপরিউক্ত বিশ্লেষণ করার জন্য STATGRAPHICS কম্পিউটার প্রোগ্রাম ব্যবহার করা হয়েছে। এই বিশ্লেষণ SPSS, STATISTICA বা অন্যান্য কম্পিউটার প্রোগ্রাম ব্যবহার করেও করা যায়।

গ্রন্থপঞ্জি

Bartlett, M. S. (1937) : Properties of Sufficiency and Statistical tests, Proc. Ray. Soc. Ser. A., 160, 268 – 282.

Bartlett, M. S. (1947) : Multivariate Analysis, J. Roy. Stats. Soc. B., 9, 176 – 179.

Belsley, D. A., Kuh, E. and Welsh, R. E. (1980) : Regression Diagnostics : Identifying Influential Data and Sources of Collinearity, John Wiley and Sons Inc., New York.

Bhuyan, K.C. (1995) : নির্ভরণ বিশ্লেষণ, বাংলা একাডেমী ঢাকা, বাংলাদেশ।

Bhuyan, K. C. (1995 – 96) : Fertility differentials according to females education, employment and family planning adoption in rural Bangladesh, Turk, Jour. Popn. Studies, 17 – 18, 21 – 39.

Bhuyan, K. C. and Nair G. A. (1995) : Structural changes in the relationships of body dimensions of four species of Pulmonate Slugs, Biom. Jour. 37, 1005 – 1015.

Bhuyan, K. C. (1996) : Fertility differentials according to Socio economic status and family planning adoption in rural Bangladesh, Sankhya, B, 58(2), 302 – 322.

Bhuyan, K. C. and Majumder, A. K. (1996) : Testing equality of regressions when error variances are heterogeneous to study production differentials of rice in Bangladesh, Biom. Jour., 38(6), 1 – 10.

Blackith, R. E. and Reyment, R. A. (1971) : Multivariate Morphometrics, Academic Press, New York.

Bliss, C.I. (1970) : Statistics in Biology, Vol. 2, McGraw-Hill Book Company, New York,

Blower, J. G., Cook, L. M. and Bishop, J. A. (1981) : Estimating the Size of Animal Populations, George Allen and Urwin, London.

Box, G. E. P. (1949) : A general distribution theory for a class of likelihood criteria, Biometrika, 36, 317 – 346.

Chatterjee, S. and Price, B. (1991) : Regression Analysis by Example, John Wiley and Sons. Inc, New York.

Cochran, W. G. (1954) : Some methods for strengthening the common χ^2-tests, Biometrics, 10, 417 – 451.

Cochran, W. G. and Cox, G. M. (1957) : Experimental Designs, John Wiley and sons Inc., New York.

Daniel, W. W. (1989) : Applied Nonparametric Statistics, 2nd ed., PWS Kent, Boston.

Dillon, D, R. and Goldstein, M. (1984) : Multivariate Analysis—Methods and Applications, John Wiley and Sons Inc, New York.

Draper, N. and Smith, H. (1981) : Applied Regression Analysis 2nd ed., John Wiley and Sons. Inc., New York.

Farrar, D. E. and Gluaber, R. R. (1967) : Multicollinearity in regression on analysis—The problem revisited. Review of Economics and statistics, 49, 92 – 107.

Federer. W. T. (1955) : Experimental Design— Theory and Applications, Macmillan and Co.

Fisher, R. A. (1921) : On the probable error of a coefficient of correlation deduced from a small sample, Metron, 1, 3 – 21.

Huberty, C. J. (1994) : Applied Discriminant Analysis, John Wiley and Sons Inc., New York.

Jeffers, J. N. R. (1967) : Two case studies the application of Principal component analysis, Appl. Statist, 16, 225 – 236.

John, P. W. M. (1971) : Statistical Design and Analysis of Experiments, MacMillan Company, New York.

Johnston, J. (1984) : Econometric Methods, 3rd ed., McGraw Hill Book Company, Singapore.

Jolliffe, I. T. (1972) : Discarding variables in Principal component analysis I. Artificial data, Appl. Statist., 21, 160 – 173.

Jolliffe, I. T. (1973) : Discarding variables in Principal component analysis II Read data, Appl. statist. 22, 21 – 31.

Kempthorne, O. (1952) : The Design and Analysis of Experiments, John Wiley and Sons Inc., New York.

Mallows C. L. (1673) : Some comments on Cp, Technometrics, 15, 661 – 675.

Nair, G, A., Mohamed. A. I. and Bhuyan, K. C. (1993) : Comparative effects of chemical Pesticides on Survival, body mass, respiration and heart rates of the pulmonate slugs, Milux Rusticus and Milax Sowerbyi. Jour Afr. Zool. 109, 141 – 149.

Neter, J. and Wasserman, W. (1974) : Applied Linear Statistical Model, IRWIN.

Osama, A. El-Silline (1997) : Morphological Study of the Candal Vertebrae of some Commercial Fishes Collected from Benghazi Coastal Water, Unpublished M. Sc. Thesis, Department of Zoology, Garyounis University, Libya.

Sokal, R. R. and Rohlf, F. (1981) : Biometry, 2nd ed., Freeman.

Theil, H. (1950) : A rank invariant method of linear and Polynomial regression analysis, III, Koninklijke Nederlandse Akademic Van watenschappen, Proceedings, A, 53, 18, 1397 – 1412.

Winer, B. J. (1962) : Statistical Principals in Experimental Designs, McGraw-Hill Book Company.

পরিশিষ্ট-এক

দৈব সংখ্যা*

10480	15011	01536	02011	81647	91646	69179	14194	62590
22368	46573	25595	85393	30995	89198	27982	53402	93965
24130	48360	22527	97265	76393	64809	15179	24830	49340
42167	93093	06243	61680	07856	16376	39440	53537	71341
37570	39975	81837	16656	06121	91782	60468	81305	49684
77921	06907	11008	42751	27756	53498	18602	70659	90655
99562	72905	56420	69994	98872	31016	71194	18738	44013
96301	91977	05463	07972	18876	20922	94595	56869	69014
89579	14342	63661	10281	17453	18103	57740	84378	25331
85475	36857	53342	53988	53060	59533	38867	62300	08158
28918	69578	88231	33276	70997	79936	56865	05859	90106
63553	40961	48235	03427	49626	69445	18663	72695	52180

09429	93969	52636	92737	88974	33488	36320	17617	30015
10365	61129	87529	85689	48237	52267	67689	93394	01511
07119	97336	71048	08178	77233	13916	47564	81056	97735
51085	12765	51821	51259	77452	16308	70756	92144	49442
02368	21382	52404	60268	89368	19885	55322	44819	01188
01011	54092	33362	94904	31273	04146	18594	29852	71585
52162	53916	46369	58585	23216	14513	83149	98736	23495
07056	97628	33787	09996	42691	06691	76988	13602	51851
48663	91245	85828	14346	09172	30168	90229	04734	59193
54164	58492	22421	74103	47070	25306	76468	26384	58151
32639	32363	05597	24200	13363	38005	94342	28728	3:806
29334	27001	87637	87308	58731	00256	45834	15398	46557

02488	33062	28834	07351	19731	92420	60952	61280	50001
81525	72295	04839	96423	24878	82651	66566	14778	76797
29676	20591	68086	26432	46901	20849	89768	81536	86645
00742	57392	39064	66432	84673	40027	32832	61362	98947
05366	04213	25669	26422	44407	44048	37937	63904	45766
91921	26418	64117	94305	26766	25940	39972	22209	71500
00582	04711	87917	77341	42206	35126	74087	99547	81817
00725	69884	62797	56170	86324	88072	76222	36086	84637
69011	65795	95876	55293	18988	27354	26575	08625	40801
25976	57948	29888	88604	67917	48708	18912	82271	65424
09763	83473	73577	12908	30883	18317	28290	35797	05998

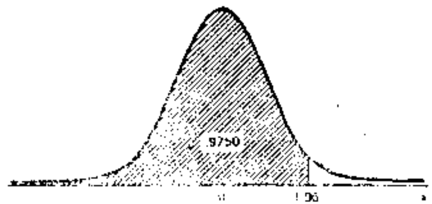
91567	42595	27958	30134	04024	86385	29880	99730	55536
17955	56349	90999	49127	20044	59931	06115	20542	18059
46503	18584	18845	49618	02304	51038	20655	58727	28168
92157	89634	94824	78171	84610	82834	09922	25417	44137
14577	62765	35605	81263	39667	47358	56873	56307	61607
98427	07523	33362	64270	01638	92477	66969	98420	04880
34914	63976	88720	82765	34476	17032	87589	40836	32427
70060	28277	39475	46473	23219	53416	94970	25832	69975
53976	54914	06990	67245	68350	82948	11398	42878	80287
76072	29515	40980	07391	58745	25774	22987	80059	39911
90725	52210	83974	29992	65831	38857	50490	83765	55657

64364	67412	33339	31926	14683	24413	59744	92351	97473
68962	00358	31662	25388	61642	34072	81249	35648	56891
95012	68379	93526	70765	10592	04542	76463	54328	02349
15664	10493	20492	38391	91132	21999	59516	81652	27195

[a]Reproduced with permission from Probability and Statistics in Engineering and Management Science by W.W. Hines and D.C. Montgomery, The Ronald Press, New York, 1972.

পরিশিষ্ট-দুই

সারণি : Normal Curve Areas $P(z \leqslant z_0)$ Entries in the body of the table are areas between $-\infty$ and z.



z	−0.09	−0.08	−0.07	−0.06	−0.05	−0.04	−0.03	−0.02	−0.01	0.00	z
−3.80	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	−3.80
−3.70	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	−3.70
−3.60	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0002	.0002	−3.60
−3.50	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	−3.50
−3.40	.0002	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	−3.40
−3.30	.0003	.0004	.0004	.0004	.0004	.0004	.0004	.0005	.0005	.0005	−3.30

৩২

সারণি : Nomal Curve Areas $P(z \leqslant z_0)$ Entries in the body of the table are areas between $-\infty$ and z.

− 3.20	.0005	.0005	.0005	.0006	.0006	.0006	.0006	.0006	.0007	.0007	− 3.20
− 3.10	.0007	.0007	.0008	.0008	.0008	.0008	.0009	.0009	.0009	.0001	− 3.10
− 3.00	.0010	.0010	.0011	.0011	.0011	.0012	.0012	.0013	.0013	.0003	− 3.00
− 2.90	.0014	.0014	.0015	.0015	.0016	.0016	.0017	.0018	.0018	.0019	− 2.90
− 2.80	.0019	.0020	.0021	.0021	.0022	.0023	.0023	.0024	.0025	.0026	− 2.80
− 2.70	.0026	.0027	.0028	.0029	.0030	.0031	.0032	.0033	.0034	.0035	− 2.70
− 2.60	.0036	.0037	.0038	.0039	.0040	.0041	.0043	.0044	.0045	.0047	− 2.60
− 2.50	.0048	.0049	.0051	.0052	.0054	.0055	.0057	.0059	.0060	.0062	− 2.50
− 2.40	.0064	.0066	.0068	.0069	.0071	.0073	.0075	.0078	.0080	.0082	− 2.40
− 2.30	.0084	.0087	.0089	.0091	.0094	.0096	.0099	.0102	.0104	.0107	− 2.30
− 2.20	.0110	.0113	.0116	.0119	.0122	.0125	.0129	.0132	.0136	.0139	− 2.20

−2.10	.0143	.0146	.0150	.0154	.0158	.0162	.0166	.0170	.0174	.0179	−2.10
−2.00	.0183	.0188	.0192	.0197	.0202	.0207	.0212	.0217	.0222	.0228	−2.00
−1.90	.0233	.0139	.0244	.0250	.0256	.0262	.0268	.0274	.0281	.0287	−1.90
−1.80	.0294	.0301	.0307	.0314	.0322	.0329	.0336	.0344	.0351	.0359	−1.80
−1.70	.0367	.0375	.0384	.0392	.0401	.0409	.0418	.0427	.0436	.0446	−1.70
−1.60	.0455	.0465	.0475	.0485	.0495	.0505	.0516	.0526	.0537	.0548	−1.60
−1.50	.0559	.0571	.0582	.0594	.0606	.0618	.0630	.0643	.0655	.0668	−1.50
−1.40	.0681	.0694	.0708	.0721	.0735	.0749	.0764	.0778	.0793	.0808	−1.40
−1.30	.0823	.0838	.0853	.0869	.0885	.0901	.0934	.0934	.0951	.0968	−1.30
−1.20	.0985	.1003	.1020	.1038	.1056	.1075	.1093	.1112	.1131	.1151	−1.20
−1.10	.1170	.1190	.1210	.1230	.1251	.1271	.1292	.1314	.1335	.1357	−1.10
−1.00	.1379	.1401	.1423	.1446	.1469	.1492	.1515	.1539	.1562	.1587	−1.00

−0.90	.1611	.1635	.1660	.1685	.1711	.1736	.1762	.1788	.1814	.1841	−0.90
−0.80	.1867	.1814	.1922	.1949	.1977	.2005	.2033	.2061	.2090	.2119	−0.80
−0.70	.2148	.2177	.2206	.2236	.2266	.2296	.2327	.2358	.2389	.2420	−0.70
−0.60	.2451	.2483	.2514	.2546	.2578	.2611	.2643	.2676	.2709	.2743	−0.60
−0.50	.2776	.2810	.2843	.2877	.2921	.2946	.2981	.3015	.3050	.3085	−0.50
−0.40	.3121	.3156	.3192	.3228	.3264	.3300	.3336	.3372	.3409	.3446	−0.40
−0.30	.3483	.3520	.3557	.3594	.3632	.3669	.3707	.3745	.3783	.3821	−0.30
−0.20	.3859	.3897	.3936	.3974	.4013	.4502	.4090	.4129	.4168	.4207	−0.20
−0.10	.4247	.4286	.4325	.4364	.4404	.4443	.4483	.4522	.4562	.4602	−0.10
−0.00	.4641	.4681	.4721	.4761	.4801	.4840	.4880	.4920	.4960	.5000	−0.00
0.00	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359	0.00
0.10	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753	0.10

0.20	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141	0.20
0.30	.6179	.6217	.6255	.6293	.6331	.6368	,6406	.6443	.6480	.6517	0.30
0.40	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879	0.40
0.50	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224	0.50
0.60	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7559	0.60
0.70	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852	0.70
0.80	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133	0.80
0.90	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389	0·90
1.00	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621	1,00
1.10	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830	1.10
1.20	.1849	.1869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015	1.20
1.30	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177	1.30

1.40	.9192	.9207	.9222	.9236	.9251	.9265	.9269	.9292	.9306	.9319	1.40
1.50	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441	1.50
1.60	.9452	.9463	.9474	.9484	.9495	.9505	.6515	.6525	.9535	.9545	1.60
1.70	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633	1.70
1.80	.9641	.9649	.9659	.9664	.9671	.9678	.9686	.9693	.9699	.9706	1.80
1.90	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767	1.90
2.00	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817	2.00
2.10	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857	2.10
2.20	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9860	2.20
2.30	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916	2.30
2.40	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936	2.40
2.50	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9954	.9952	2.50
2.60	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964	2.60

2.70	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974	2.70
2.80	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981	2.80
2.90	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986	2.90
3.00	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990	3.00
3.10	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993	3.10
3.20	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995	3.20
3.30	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997	3.30
3.40	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998	3.40
3.50	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	3.50
3.60	.9998	.9998	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	3.60
3.70	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	3.70
3.80	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	3.80

পরিশিষ্ট-তিন : Critical value of Spearman's Rank Correlation Coefficient.

	Level of significance for one-tailed test			
	0.05	0.025	0.01	0.005
	Level of significance for two-tailed test			
n	0.10	0.05	0.02	0.01
5	0.900	--	—	—
6	0.829	0.886	0.943	—
7	0.714	0.786	0.893	—
8	0.643	0.738	0.833	0.881
9	0.600	0.683	0.783	0.833
10	0.564	0.648	0.745	0.794
11	0.523	0.623	0.736	0.818
12	0.497	0.591	0.703	0.780
13	0.475	0.566	0.673	0.745
14	0.457	0.545	0.646	0.716
15	0.441	0.525	0.623	0.689
16	0.425	0.507	0.601	0.666
17	0.412	0.490	0.582	0.645
18	0.399	0.476	0.594	0.625
19	0.388	0.462	0.549	0.625
20	0.377	0.450	0.534	0.591
21	0.368	0.438	0.521	0.576
22	0.359	0.428	0.508	0.562

23	0.351	0.418	0.496	0.549
24	0.343	0.409	0.485	0.527
25	0.336	0.400	0.475	0.526
26	0.329	0.392	0.465	0.515
27	0.323	0.385	0.456	0.505
28	0.317	0.377	0.448	0.496
29	0.311	0.370	0.440	0.484
30	0.305	0.364	0.432	0.478

৩৩।

পরিশিষ্ট-চার : Mann-Whitney U-test Values (two-tailed test) P = 0.05

n_1 \ n_2	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
							0	0	0	0	1	1	1	1	1	2	2	2	2
[illegible]				0	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8
4			0	1	2	3	4	4	5	6	7	8	9	10	11	11	12	13	13
5		0	1	2	3	5	6	7	8	9	11	12	13	14	15	17	18	19	20
[illegible]		1	2	3	5	6	8	10	11	13	14	16	17	19	21	22	24	25	27
[illegible]		1	3	5	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34
8	0	2	4	6	8	10	13	15	17	19	22	24	26	29	31	34	36	38	41
9	0	2	4	7	10	12	15	17	20	23	26	28	31	34	37	39	42	45	48
10	0	3	5	8	11	14	17	20	23	26	29	33	36	39	42	45	48	52	55
11	0	3	6	9	13	16	19	23	26	30	33	37	40	44	47	51	55	58	62
12	1	4	7	11	14	18	22	26	29	33	37	41	45	49	53	57	61	65	69

13	1	4	8	12	16	20	24	28	33	37	41	45	50	54	59	63	67	72	76
14	1	5	9	13	17	22	26	31	36	40	45	50	55	59	64	67	74	78	83
15	1	5	10	14	19	24	29	34	39	44	49	54	59	64	70	75	80	85	90
16	1	6	11	15	21	26	31	37	42	47	53	59	64	70	75	81	86	92	98
17	2	6	11	17	22	28	34	39	45	51	57	63	67	75	81	87	93	99	105
18	2	7	12	18	24	30	36	42	48	55	61	67	74	80	86	93	99	106	112
19	2	7	13	19	25	32	38	45	52	58	65	72	78	85	92	99	106	113	119
20	2	8	13	20	27	34	41	48	55	62	69	76	83	90	98	105	112	119	127

n_1 and n_2 are the number of observations in each sample.