ইউনিট

৭

অপরামাত্রিক পরীক্ষা
Non-Parametric Test

ভূমিকা

পরিসাংখ্যিক কল্পনা (Hypothesis) যাচাইয়ের ক্ষেত্রে সাধারণত দুটি পদ্ধতি অনুসরণ করা হয়। এদের একটি হল অপরামিতিক যাচাই পদ্ধতি। যখন সমগ্রকের বিন্যাস গঠন অজানা থাকে এবং পরামিতি নির্দিষ্ট কিংবা অনির্দিষ্ট থাকে, তখন বিন্যাসের গঠন সম্পর্কে একটি কল্পনা করে তা যাচাই করা হয়। কল্পনা যাচাইয়ের এই পদ্ধতিকে বলা হয় অপরামিতিক যাচাই পদ্ধতি (Non-Parametric Test Method)। এ ধরনের যাচাই পদ্ধতির মধ্যে কাইবর্গ যাচাই (χ^2 Test), টি (t), ম্যান হুইটনি যাচাই অন্যতম।

ইউনিট সমাপ্তির সময়	ইউনিট সমাপ্তির সর্বোচ্চ সময় ০২ সপ্তাহ

এ ইউনিটের পাঠসমূহ

পাঠ ৭.১	ঃ	অপরামাত্রিক পরীক্ষা
পাঠ ৭.২	ঃ	চিহ্ন ঘটিত পরীক্ষা
পাঠ ৭.৩	ঃ	ম্যান হুইটনি পরীক্ষা
পাঠ ৭.৪	ঃ	ক্রুসক্যাল ওয়ালিস এইচ পরীক্ষা, অনুক্রমিক ঘটন পরীক্ষা
পাঠ ৭.৫	ঃ	স্পিয়ারম্যানের গুণাক্রমিক সংস্রব পরীক্ষা

পাঠ ৭.১ অপরামাত্রিক পরীক্ষা
Non- Parametric Test

উদ্দেশ্য

এ পাঠ শেষে আপনি-

- অপরামাত্রিক পরীক্ষা সংজ্ঞা বলতে পারবেন।
- অপরামাত্রিক পরীক্ষার সুবিধা অসুবিধা ব্যাখ্যা করতে পারবেন।
- অপরামাত্রিক পরীক্ষার ব্যবহার বাস্তব জীবনে প্রয়োগ করতে পারবেন।

অপরামাত্রিক বা অপরামিতি পরীক্ষার সংজ্ঞা
Definition of Non- Parametric Tests

পরামাত্রিক পরীক্ষায় কোন নমুনা যে সমগ্রক হতে নির্বাচন করা হয় তার কোন বৈশিষ্ট্য বা পরামিতি সম্পর্কে অনুমান করা হয়। এক্ষেত্রে সমগ্রক কতকগুলো বৈশিষ্ট্য মেনে চলে। সেগুলো হলো:

(ক) **বন্টনের স্বাভাবিকতা (Normality of distribution):** যে সমগ্রক হতে নমুনা নির্বাচন করা হয় সে সমগ্রকের নিবেশন পরিমিত হতে হবে।

(খ) **ভেদাংকের সমরূপতা (Homogenity of Variances):** যে সকল নমুনার উপর পর্যবেক্ষণ করা হয়েছে সেগুলোর ভেদাংক একই রকম হবে।

(গ) **অবিচ্ছিন্নতা (Continuity):** এক্ষেত্রে যে চলকটি পরিমাপ করা হয় উহা অবিচ্ছিন্ন প্রকৃতির হতে হবে।

উপরের শর্ত বা বৈশিষ্ট্য বিদ্যমান ধরে সমগ্রকের পরামিতির অনুমান যাচাই করাই পরামাত্রিক পরীক্ষা।

যেমন: Z পরীক্ষা, t পরীক্ষা এবং F পরীক্ষা।

অন্যদিকে অপরমাত্রিক পরীক্ষায় এসব শর্ত বিশেষ করে বন্টনের স্বাভাবিকতা পূরণ করতে হয় না। এক্ষেত্রে কোন নমুনা যে সমগ্রক হতে নির্বাচন করা হয় সে সমগ্রকের সুনির্দিষ্ট বৈশিষ্ট্য সম্পর্কে কল্পনা করা হয় না। সাধারণত: ক্রমবোধক (Ordinal) এবং নামীয় (Cardinal) মাপকে পরিমাপকৃত তথ্যের ক্ষেত্রে অনুমান পরীক্ষা করাকে অপরামাত্রিক পরীক্ষা বলে। যেমন: চিহ্ন ঘটিত পরীক্ষা (Sign Test), ম্যান-হুইটনী ইউ পরীক্ষা (Mann-Whitney U Test), ক্রুসক্যাল-ওয়ালিশ এইচ পরীক্ষা (Kruskal Wallis H Test) ইত্যাদি।

অপরামাত্রিক পরীক্ষার সুবিধাসমূহ
Advantages of Non-Parametric Tests

(ক) যদি নমুনার আকার খুব ছোট (সাধারণত $n \leq 6$) হয় এবং নিবেশন সম্পর্কে সঠিক ধারণা না থাকে সেক্ষেত্রে অনুমান যাচাইয়ের একমাত্র উপায় হল অপরামাত্রিক পরীক্ষা।

(খ) যে নিবেশন হতে নমুনা সংগ্রহ করা হয় সেই নিবেশনের প্রকৃতি জানা না থাকলেও অপরামাত্রিক পরীক্ষণের মাধ্যমে সম্ভাবনা নির্ণয় করা যায়।

(গ) অপরামাত্রিক পরীক্ষার মাধ্যমে ক্রমবোধক এবং নামীয় মাপকে পরিমাপকৃত তথ্যের নির্ভরশীলতা যাচাই করা যায়।

(ঘ) বিভিন্ন সমগ্রক হতে আলাদাভাবে সংগৃহীত নমুনার ক্ষেত্রে অপরামাত্রিক পরীক্ষা ব্যবহার করা যায়।

(ঙ) অপরামাত্রিক পরীক্ষা তুলনামূলকভাবে পরামাত্রিক পরীক্ষার চেয়ে সহজ এবং জটিলতা কম।

অপরামাত্রিক পরীক্ষার অসুবিধাসমূহ
Disadvantages of Non-Parametric Tests

(ক) যেসব তথ্যে পরামাত্রিক পরিসংখ্যানিক মডেলের সকল বৈশিষ্ট্য বিদ্যমান, সেক্ষেত্রে অপরামাত্রিক পরীক্ষা ব্যবহার করলে তথ্যের অপচয় হবে। এক্ষেত্রে পরামাত্রিক পরীক্ষার চেয়ে অপরামাত্রিক কম শক্তি সম্পন্ন।

(খ) অপরামাত্রিক পরীক্ষা সুষম নয়।

(গ) সংযোজন সম্পর্কে একটি বিশেষ ধারণা ছাড়া অপরামাত্রিক পরীক্ষার ক্ষেত্রে ভেদাংক বিশ্লেষণের জন্য সংযোগফল পরিমাপ করা যায় না।

(ঘ) অপরামাত্রিক পরীক্ষা সাধারণত আচরণ বিজ্ঞানের ক্ষেত্রে ব্যবহৃত হয় বিধায় পরিসংখ্যানে ইহার সীমিত ব্যবহার।

অপরামাত্রিক পরীক্ষা ব্যবহার

Uses of Non-Parametric Tests

অপরামাত্রিক পরীক্ষার বহুল ব্যবহার রয়েছে। নিম্নে উল্লেখ করা হলো:

(ক) যুগল নমুনার পার্থক্যের চিহ্ন জানা থাকলে অপরামাত্রিক পরীক্ষা ব্যবহার করা হয়,

(খ) দুই বা ততোধিক অনপেক্ষ নমুনা একই সমগ্রক হতে নেওয়া হয়েছে কিনা যাচাইয়ে,

(গ) কোন পরীক্ষণের নমুনা দৈবচয়িত কিনা যাচাইযে,

(ঘ) দুটি চলকের মধ্যকার সম্পর্কের পরীক্ষা যাচাইয়ে,

(ঙ) সাধারণত সমগ্রকের কোন বৈশিষ্ট্য বা পরামিতির উপর অনুমান ব্যতীত অন্য কোন বিষয়ের অনুমান যাচাইয়ে অপরামাত্রিক পরীক্ষা ব্যবহার করা হয়।

সারসংক্ষেপ

যদি নমুনার আকার খুব ছোট (সাধারণত $n \leq 6$) হয় এবং নিবেশন সম্পর্কে সঠিক ধারণা না থাকে সেক্ষেত্রে অনুমান যাচাইয়ের একমাত্র উপায় হল অপরামাত্রিক পরীক্ষা। অপরামাত্রিক পরীক্ষা সুষম নয়। যুগল নমুনার পার্থক্যের চিহ্ন জানা থাকলে অপরামাত্রিক পরীক্ষা ব্যবহার করা হয়।

পাঠ ৭.২ চিহ্ন ঘটিত পরীক্ষা
Sign Tests

উদ্দেশ্য

এ পাঠ শেষে আপনি-

- চিহ্ন ঘটিত পরীক্ষার সংজ্ঞা লিখতে পারবেন।
- চিহ্ন ঘটিত পরীক্ষা প্রয়োগ করতে পারবেন।

চিহ্নঘটিত পরীক্ষা
Sign Tests

এই পদ্ধতির নামানুসারে বুঝা যায় যে, এক্ষেত্রে সংখ্যাত্মক পরিমাপ ব্যবহার না করে বরং চিহ্ন (যোগ বা বিয়োগ) ব্যবহার করা হয়। যে সমস্ত গবেষনায় তথ্যকে সংখ্যাত্মকভাবে প্রকাশ করা যায় না সেক্ষেত্রে এ পদ্ধতি খুবই উপযোগী। এই পরীক্ষা সাধারণত সেসব ক্ষেত্রে ব্যবহার করা হয় যেখানে তথ্য যুগলভাবে পর্যবেক্ষণ করা হয় এবং দুটো জিনিসের মধ্যে তুলনা করা হয়।

ধরি তুলনাধীন দুটি চলক A ও B এবং তাদের পরিমাপের প্রতীক x এবং y। এক্ষেত্রে n সংখ্যক জোড়া মান বিবেচনা করা হল।

$\therefore$ পর্যবেক্ষণসমূহ : $(x_1y_1)\ (x_2y_2)\ , \ldots\ldots\ldots\ldots.,\ (x_ny_n)$

এবং উহাদের পার্থক্যসমূহ: $(x_1\text{-}y_1)\ (x_2\text{-}y_2)\ , \ldots\ldots\ldots\ldots.,\ (x_n\text{-}y_n)$

চিহ্নঘটিত পরীক্ষার ভিত্তি হল পার্থক্যগুলোর চিহ্নসমূহ। যদি জোড়াটি সমযোগী হয় তাহলে পার্থক্য হবে শূন্য। শূন্য পার্থক্য বিশিষ্ট জোড়াগুলোকে গণনা হতে বাদ দিতে হয় এবং সেই অনুসারে নমুনার আকার (n) হ্রাস পায়। যদি নমুনার আকার ছোট হয় তাহলে দ্বিপদী বিন্যাসের সাহায্যে সম্ভাবনা নির্ণয় করা হয়।

এক্ষেত্রে নাস্তি অনুমান, $H_o : \pi = \frac{1}{2}$

ধরি ধনাত্মক চিহ্নযুক্ত জোড়ের সংখ্যা r এবং π গুরুত্বের স্তর।

এখন, n, r এবং π এর ভিত্তিতে দ্বিপদী বিন্যাসের সাহায্যে সম্ভাবনা নির্ণয় করা হয়।

সিদ্ধান্ত :

একদিক বিশিষ্ট পরীক্ষা:

ক. ডানদিক বিশিষ্ট পরীক্ষা: $P(x \geq r) \leq \alpha$ হলে H_o বর্জনীয় হবে।

খ. ডানদিক বিশিষ্ট পরীক্ষা: $P(x \leq r) \leq \alpha$ হলে H_o বর্জনীয় হবে।

দুই দিক বিশিষ্ট পরীক্ষা: $P(x \geq r) \leq \frac{\alpha}{2}$ হলে H_o বর্জনীয় হবে।

* যদি নমুনার আকার বড় হয় তাহলে দ্বিপদী বিন্যাসের পরিবর্তে z test এর মাধ্যমে অনুমান যাচাই করা হয়। আবার মধ্যমার পরীক্ষাও z test এর সাহায্যে করা যায়।

এক্ষেত্রে, $z = \frac{x-\mu}{\sigma}$

$$= \frac{x-\frac{n}{2}}{\frac{\sqrt{n}}{2}}$$

$$= \frac{\frac{2x-n}{2}}{\frac{\sqrt{n}}{2}}$$

$$= \frac{2x-n}{\sqrt{n}}$$

এখানে, $\mu = n\pi = \frac{n}{2}$

$$\sigma = \sqrt{n\pi(1-\pi)}$$

$$= \sqrt{n\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)}$$

$$= \sqrt{\frac{n}{4}}$$

$$= \sqrt{\frac{n}{4}}$$

$$= \frac{\sqrt{n}}{2}$$

তবে continuity correction ব্যবহার করা হলে–

$$z = \frac{2(x \pm 0.5)-n}{\sqrt{n}}$$

$$= \frac{2x \pm 1 - n}{\sqrt{n}}$$

যদি $x < \frac{n}{2}$ হয় তাহলে +1 এবং $x > \frac{n}{2}$ হলে -1 ব্যবহার করা হবে।

উদাহরণ: একটি প্রতিষ্ঠান হতে 17 জোড়া কর্মচারীকে বাছাই করা হল। প্রত্যেক জোড়ার একজনকে **১**ম পদ্ধতিতে এবং অপরজনকে ২য় পদ্ধতিতে প্রশিক্ষণ দেয়া হল এবং প্রশিক্ষণ শেষে পরীক্ষা নেয়া হল। পরীক্ষার ফলাফল নিম্নে দেওয়া হলো–

জোড়া	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
১ম পদ্ধতি	65	68	70	63	73	75	72	78	64	73	79	80	67	64	62	74	82
২য় পদ্ধতি	63	68	68	60	72	75	73	70	66	70	77	78	63	65	60	74	78

চিহ্নঘটিত পরীক্ষা ব্যবহার করে (ক) পদ্ধতি দুটি সমান কার্যকরী (খ) **১**ম পদ্ধতি ২য় পদ্ধতি অপেক্ষা অধিক কার্যকরী কিনা 5% তাৎপর্য স্তরে পরীক্ষা করুন।

সমাধান : (ক) নাস্তি অনুমান, $H_o : \pi = 0.5$ অর্থাৎ পদ্ধতি দুটি সমান কার্যকরী।

বিকল্প অনুমান, $H_A : \pi \neq 0.5$ অর্থাৎ পদ্ধতি দুটি সমান কার্যকরী নয়।

ধরি **১**ম পদ্ধতির ফলাফল X_1 এবং ২য় পদ্ধতি ফলাফল X_2 ।

অতএব ফলাফলের মধ্যকার পার্থক্য $(x_1\text{-}x_2)$ নিম্নে দেওয়া হলো:

2, 0, 2, 3, 1, 0, -1, 8, -2, 3, 2, 2, 4, -1, 2, 0, 4

যেহেতু 17 জোড়া তথ্যে পার্থক্যের মধ্যে 11টি ধনাত্মক এবং ৩টি শূন্য।

অতএব r = 11 এবং n = 17-3 = 14।

এখন দ্বিপদী বিন্যাসের সাহায্যে সম্ভাবনা,

$$P(X \geq 11) = {}^{14}C_{11}(0.5)^{11}(0.5)^{3} + {}^{14}C_{12}(0.5)^{12}(0.5)^{2} + {}^{14}C_{13}(0.5)^{13}(0.5)^{1} + {}^{14}C_{14}(0.5)^{14}(0.5)^{0}$$

$$= 0.022217 + 0.005554 + 0.000854 + 0.000061$$

$$= 0.0287$$

যেহেতু H_A: $\pi \neq 0.5$। অতএব পরীক্ষাটি দুই দিক বিশিষ্ট।
দেওয়া আছে–
গুরুত্বের স্তর, $\alpha = 5\% = 0.05$

$$\therefore \frac{\alpha}{2} = \frac{.05}{2}$$
$$= 0.025$$

যেহেতু $P(x \geq 11) > \frac{\alpha}{2}$

অতএব নাস্তি অনুমান (H_o) গ্রহণীয়। অর্থাৎ পদ্ধতি দুটি সমান কার্যকরী।

(খ) নাস্তি অনুমান, $H_o : \pi = 0.5$ অর্থাৎ ১ম পদ্ধতি ২য় পদ্ধতি অপেক্ষা অধিক কার্যকরী নয়।
বিকল্প অনুমান, $H_A : \pi > 0.5$ অর্থাৎ ১ম পদ্ধতি ২য় পদ্ধতি অপেক্ষা অধিক কার্যকরী।
ক অংশ অনুসারে $P(x \geq 11) = 0.0287$
যেহেতু H_A: $\pi > 0.5$। অতএব পরীক্ষাটি একদিক (ডানদিক) বিশিষ্ট।
দেওয়া আছে গুরুত্বের স্তর, $\alpha = 5\% = 0.05$
যেহেতু $P(x \geq 11) < \alpha$
অতএব নাস্তি অনুমান (H_o) বর্জনীয়। অর্থাৎ ১ম পদ্ধতি ২য় পদ্ধতির চেয়ে অধিক কার্যকরী।

সারসংক্ষেপ

চিহ্নঘটিত পরীক্ষা পদ্ধতির নামানুসারে বুঝা যায় যে, এক্ষেত্রে সংখ্যাত্মক পরিমাপ ব্যবহার না করে বরং চিহ্ন (যোগ বা বিয়োগ) ব্যবহার করা হয়। যে সমস্ত গবেষনায় তথ্যকে সংখ্যাত্মকভাবে প্রকাশ করা যায় না সেক্ষেত্রে এ পদ্ধতি খুবই উপযোগ। এই পরীক্ষা সাধারণত সেসব ক্ষেত্রে ব্যবহার করা হয় যেখানে তথ্য যুগলভাবে পর্যবেক্ষণ করা হয় এবং দুটো জিনিসের মধ্যে তুলনা করা হয়।

পাঠ ৭.৩ ম্যান হুইটনি ইউ পরীক্ষা
Mann - Whitney U Test

উদ্দেশ্য

এ পাঠ শেষে আপনি-

- ম্যান হুইটনি ইউ পরীক্ষার সংজ্ঞা বলতে পারবেন।
- ম্যান হুইটনি ইউ পরীক্ষা প্রয়োগ করতে পারবেন।

ম্যান - হুইটনি ইউ পরীক্ষা
Mann - Whitney U Test

দুটি অনপেক্ষ নমুনার মধ্যকার পার্থক্য পরীক্ষা করার জন্য এবং দুটি নমুনা যে দুটি সমগ্রক হতে নেওয়া হয়েছে তাদের চলকের বিন্যাস একই কিনা পরীক্ষা করার জন্য সর্বপ্রথম ব্যবহৃত পরীক্ষা হল ম্যান-হুইটনি U পরীক্ষা। উক্ত পরীক্ষার ধাপসমূহ–

ধাপ-১ : দুটি নমুনা তথ্যগুলোকে একত্র করে ছোট থেকে বড় ক্রমানুসারে সাজিয়ে ক্রমিক মান নির্ণয় করতে হবে।

ধাপ -২ : প্রত্যেক নমুনার ক্রমিক মান আলাদাভাবে উপস্থাপন করতে হবে। অত:পর প্রত্যেক নমুনার ক্রমিক মানের সমষ্টি নির্ণয় করতে হবে।

ধাপ - ৩: ম্যান-হুইটনি U নমুনাজ মান নির্ণয় করতে হবে। এর সূত্র:

$$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$$

এখানে n_1 = ১ম নমুনার আকার

n_2 = ২য় নমুনার আকার

R_1 = ১ম নমুনার ক্রমিক মানের সমষ্টি

ধাপ - ৪: যদি প্রত্যেক নমুনার আকার কমপক্ষে ৮ হয় তাহলে U এর বিন্যাস পরিমিত বিন্যাসে রূপান্তরিত হবে।

এক্ষেত্রে $$z = \frac{U - \mu_u}{\sigma_u}$$

এখানে, $$\mu_u = \frac{n_1 n_2}{2}$$

$$\sigma_u^2 = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}$$

ধাপ - ৫: সিদ্ধান্তঃ

একপার্শ্বিক পরীক্ষাঃ

ক. ডানদিক বিশিষ্ট পরীক্ষা : $z_{\text{নির্ণীত}} < -z_\alpha$ হলে Ho বর্জনীয় হবে।

খ. বামদিক বিশিষ্ট পরীক্ষা : $z_{\text{নির্ণীত}} > z_\alpha$ হলে Ho বর্জনীয় হবে।

দ্বিপার্শ্বিক পরীক্ষাঃ

$z_{\text{নির্ণীত}} < -z_{\alpha/2}$ অথবা z নির্ণীত $> z_{\alpha/2}$ হলে Ho বর্জনীয় হবে।

উদাহরণ: দুটি প্রতিষ্ঠান হতে নির্বাচিত দুটি নমুনার দ্রব্যগুলোর শক্তি নিম্নরূপ:

প্রতিষ্ঠান - ১	প্রতিষ্ঠান - ২
18.3, 16.4, 22.7, 17.8, 18.9, 25.3, 16.1, 24.2	12.6, 14.1, 20.5, 10.7, 15.9, 19.6, 12.9, 15.2, 11.8, 147

দুটি নমুনার শক্তির মধ্যে পার্থক্য আছে কিনা ৫% গুরুত্বের স্তরে পরীক্ষা করুন।

সমাধান: নাস্তি অনুমান, H_o : দুটি নমুনার শক্তির মধ্যে পার্থক্য নেই।
বিকল্প অনুমান, H_A : দুটি নমুনার শক্তি মধ্যে পার্থক্য আছে।
নিম্নে ম্যান-হুইটনি U পরীক্ষার মাধ্যমে অনুমান যাচাই করা হল–

ধাপ -১: নিম্নে নমুনা দুটির তথ্যগুলো একত্রে ছোট থেকে বড় ক্রমানুসারে সাজিয়ে ক্রমিক মান নির্ণয় করা হলো–

তথ্য	10.7	11.8	12.6	12.9	14.1	14.7	15.2	15.9	16.1	16.4	17.8	18.3	18.9	19.6	20.5	22.7	24.2	25.3
ক্রমিক মান	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18

ধাপ-২: নিম্নে প্রত্যেক নমুনার ক্রমিক মান আলাদাভাবে উপস্থাপন করে তাদের সমষ্টি বের করা হলো–

প্রতিষ্ঠান-১	
দ্রব্য	ক্রমিক মান
18.3	12
16.4	10
22.7	16
17.8	11
18.9	13
25.3	18
16.1	9
24.2	17
	মোট = 106

প্রতিষ্ঠান-২	
দ্রব্য	ক্রমিক মান
12.6	3
14.1	5
20.5	15
10.7	1
15.9	8
19.6	14
12.9	4
15.2	7
11.8	2
14.7	6
	মোট = 65

এখানে $n_1 = 8$ এবং $n_2 = 10$ । তাদের ক্রমিক মানের সমষ্টি যথাক্রমে $R_1 = 106$ এবং $R_2 = 65$।
ধাপ-৩: ম্যান-হুইটনি U এর নমুনাজ মান,

$$U = n_1n_2 + \frac{n_1(n_1+1)}{2} - R_1$$
$$= (8)(10) + \frac{8(8+1)}{2} - 106$$
$$= 80 + 36 - 106$$
$$= 116 - 106 = 10$$

ধাপ-৪: আমরা জানি $z = \frac{U - \mu_u}{\sigma_u}$

এখানে, $\mu_u = \frac{n_1 n_2}{2}$

$$= \frac{(8)(10)}{2}$$

$$= \frac{80}{2}$$

$$= 40$$

এবং $\sigma_u^2 = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}$

$$= \frac{(8)(10)(8+10+1)}{12}$$

$$= \frac{(80)(19)}{12}$$

$$= 126.67$$

$\therefore \sigma_u = \sqrt{126.67}$

$$= 11.25$$

$\therefore z = \frac{10-40}{11.25}$

$$= -2.67$$

$\therefore$ z এর নির্ণীত মান = - 2.67

ধাপ-৫: এখানে গুরুত্বের স্তর, $\alpha = 5\% = 0.05$

যেহেতু পরীক্ষাটি দ্বিপার্শ্বিক।

$$\therefore \frac{\alpha}{2} = \frac{.05}{2} = 0.025$$

$\therefore$ z-এর তাত্ত্বিক মান $\pm$ 1.96

ধাপ-৬ : যেহেতু z এর নির্ণীত মান উহার তাত্ত্বিক মানদ্বয়ের মধ্যে নেই। অতএব নাস্তি অনুমান বর্জনীয়। অর্থাৎ দুটি নমুনার শক্তির মধ্যে পার্থক্য আছে।

সারসংক্ষেপ

দুটি অনপেক্ষ নমুনার মধ্যকার পার্থক্য পরীক্ষা করার জন্য এবং দুটি নমুনা যে দুটি সমগ্রক হতে নেওয়া হয়েছে তাদের চলকের বিন্যাস একই কিনা পরীক্ষা করার জন্য সর্বপ্রথম ব্যবহৃত পরীক্ষা হল ম্যান-হুইটনি U পরীক্ষা। এর সূত্র:

$$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$$

পাঠ ৭.৪ ক্রুসক্যাল ওয়ালিস এইচ এবং অনুক্রমিক ঘটন পরীক্ষা
Kruskal - Wallis H Test and Run Test

উদ্দেশ্য

এ পাঠ শেষে আপনি-

- ক্রুসক্যাল ওয়ালিস এইচ পরীক্ষার সংজ্ঞা বলতে পারবেন।
- ক্রুসক্যাল ওয়ালিস এইচ পরীক্ষার প্রয়োগ করতে পারবেন।
- অনুক্রমিক ঘটন পরীক্ষার সংজ্ঞা লিখতে পারবেন।
- অনুক্রমিক ঘটন পরীক্ষা প্রয়োগ করতে পারবেন।

ক্রুসক্যাল-ওয়ালিস H পরীক্ষা
Kruskal - Wallis H Test

দুটি নমুনা একই জাতীয় সমগ্রক হতে নেওয়া হয়েছে কিনা পরীক্ষার ক্ষেত্রে ম্যান-হুইটনি U পরীক্ষা ব্যবহৃত হয়। অন্যদিকে যে পদ্ধতির সাহায্যে k সংখ্যক (দুয়ের অধিক) নমুনা একই সমগ্রক নেওয়া হয়েছে কিনা যাচাই করা হয় তাই ক্রুসক্যাল – ওয়ালিস H পরীক্ষা। একে সংক্ষেপে H পরীক্ষা বলা হয়।

ধরি k সংখ্যক নমুনার আকার সমূহ $n_1, n_2, ..., n_k$ এবং তাদের ক্রমিক মানের সমষ্টি সমূহ যথাক্রমে $R_1, R_2, ..., R_k$।

∴ ক্রুসক্যাল -ওয়ালিস H পরীক্ষার সূত্র–

$$H = \frac{12}{n(n+1)} \sum_{j=1}^{k} \frac{R_j^2}{n_j} - 3(n+1)$$

এখানে k = নমুনার সংখ্যা।

n_j= j তম নমুনায় নমুনার আকার

$n = n_1 + n_2 + ... + n_k$

R_j = j তম নমুনার ক্রমিক মানের সমষ্টি

সিদ্ধান্ত : যদি $H > \chi^2_{(k-1).\alpha}$ হয় তাহলে নাস্তি অনুমান, H_o বর্জনীয় হবে। অন্যথায় নাস্তি অনুমান গ্রহণীয় হবে। এখানে k-1 হচ্ছে স্বাধীনতার মাত্রা এবং α হচ্ছে গুরুত্বের স্তর।

বি:দ্র: যদি নমুনায় অন্তর্ভূক্ত তথ্যের মধ্যে টাই থাকে তাহলে উপরের নিয়ে প্রাপ্ত H এর মান যা হওয়া উচিত তার চেয়ে কম হয়। এক্ষেত্রে সংশোধিত H নির্ণয় করা হয়। এর সূত্র–

$$H_c = \frac{H}{1 - \frac{\Sigma(T^3 - T)}{n^3 - n}}$$

এখান T = কোন তথ্যের টাই হবার সংখ্য

উদাহরণ: তিনটি নমুনার তথ্য নিম্নে দেয়া হলোঃ

১ম নমুনা	35, 24, 30, 43
২য় নমুনা	44, 42, 46
৩য় নমুনা	28, 15, 23, 38, 18

নমুনাগুলো একই সমগ্রক হতে নেওয়া হয়েছে কিনা (ক) 5% এবং (খ) 1% গুরুত্বের স্তরে পরীক্ষা করুন।

সমাধান: নাস্তি অনুমান, H_o : নমুনাগুলো একই সমগ্রক হতে নেওয়া হয়েছে।

বিকল্প অনুমান, H_A : নমুনাগুলো একই সমগ্রক হতে নেওয়া হয় নাই।

নিম্নে ক্রুসক্যাল-ওয়ালিস H পরীক্ষার সাহায্যে অনুমান যাচাই করা হলোঃ–

ধাপ-১ঃ প্রদত্ত তথ্যসমূহকে নিম্নে উচ্চ ক্রমানুসারে সাজিয়ে ক্রমিক মান (Rank) নির্ণয় করা হলো–

তথ্য	15	18	23	24	28	30	35	38	42	43	44	46
ক্রমিক মান	1	2	3	4	5	6	7	8	9	10	11	12

ধাপ-২: নিম্নে প্রত্যেক নমুনার ক্রমিকমান আলাদাভাবে উপস্থাপন করে তাদের সমষ্টি বের করা হলো–

নমুনা	তথ্য	সমষ্টি (R_j)
১ম	7, 4, 6, 10	27
২য়	11, 9, 12	32
৩য়	5, 1, 3, 8, 2	19

এখানে $n_1 = 4$, $n_2 = 3$ এবং $n_3 = 5$

$R_1 = 27$, $R_2 = 32$ এবং $R_3 = 19$

$\therefore n = n_1 + n_2 + n_3 = 4 + 3 + 5 = 12$

ধাপ-৩ : ক্রুসক্যাল-ওয়ালিস H এর নমুনাজ মান,

$$H = \frac{12}{n(n+1)}\sum_{j=1}^{k}\frac{R_j^2}{n_j} - 3(n+1)$$

$$= \frac{12}{12(12+1)}\left[\frac{(27)^2}{4} + \frac{(32)^2}{3} + \frac{(19)^2}{5}\right] - 3(12+1)$$

$$= \frac{12}{156}[182.25 + 341.33 + 72.20] - 39$$

$$= \frac{12}{156}[595.78] - 39$$

$= 45.43-39$

$= 6.83$

ধাপ-৪: সিদ্ধান্তঃ

এখানে স্বাধীনতার মাত্রা, $v = k-1 = 3-1 = 2$ | $\therefore$ নমুনার সংখ্যা, $k = 3$

(ক) গুরুত্বের স্তর, $\alpha = 5\% = 0.05$

$\therefore \chi^2$ এর তাত্ত্বিকমান, $\chi^2_{2,0.05} = 5.991$

যেহেতু $H > \chi^2_{2,0.05}$। অতএব নাস্তি অনুমান বর্জনীয়। অর্থাৎ নমুনাগুলো একই সংমগ্রক হতে নেওয়া হয় নাই।

(খ) গুরুত্বের স্তর, $\alpha = 1\% = 0.01$

$\therefore \chi^2$ এর তাত্ত্বিকমান, $\chi^2_{2,0.01} = 9.21$

যেহেতু $H < \chi^2_{2,0.05}$। অতএব নাস্তি অনুমান গ্রহণীয়। অর্থাৎ নমুনাগুলো একই সংমগ্রক হতে নেওয়া হয়েছে।

অনুক্রমিক ঘটন পরীক্ষা

Run Test

দৈব নমুনায়নের জন্য কোন পরিসংখ্যানিক পরীক্ষা ইতিপূর্বে আলোচিত হয় নাই। কোন নমুনার দৈবায়ন যে পরিসংখ্যানিক পরীক্ষার সাহায্যে যাচাই করা হয় তাকে অনুক্রমিক ঘটন পরীক্ষা (Run Test) বলে। যেমনঃ কোন টিকেট কাউন্টারে পুরুষ ও মহিলার অবস্থান দৈবচয়িত কিনা পরীক্ষাটি অনুক্রমিক ঘটন পরীক্ষার সাহায্যে যাচাই করা হয়।

পরীক্ষার ধাপসমূহ:

১ম ধাপ ঃ অনুক্রমের (Run) সংখ্যা নির্ণয় করতে হবে। অনুক্রম হচ্ছে সমজাতীয় চিহ্নের ধারাবাহিক অনুগমন যেগুলোর পূর্বে বা পরে ভিন্ন প্রকারের চিহ্ন থাকে। যেমন ঃ একটি মুদ্রা 10 নিক্ষেপে ফলাফলগুলো পাওয়া গেল-

H H H T T H H T T T

এক্ষেত্রে অনুক্রমের সংখ্যা, R = 4, H-এর সংখ্যা, n_1 = 5 এবং T এর সংখ্যা, n_2 = 5।

২য় ধাপ ঃ যদি n_1 এবং n_2 এর সংখ্যা 20 বা তার কম হয় তাহলে n_1 ও n_2 এর মান এবং গুরুত্বের স্তরের সাহায্যে সারণী হতে R এর তাত্ত্বিক মান নির্ণয় করতে হবে।

যদি R এর পর্যবেক্ষিত মান R এর টেবিলের মানদ্বয়ের ছোট মানটির সমান বা কম অথবা বড় মানটি সমান বা বেশি হলে নাস্তি অনুমান বর্জণীয় হবে।

যদি n_1 বা n_2 এর সংখ্যা 20 এর অধিক হয় তাহলে নিম্নের সূত্রে সাহায্যে z এর মান নির্ণয় করতে হবে।

$$z = \frac{R-\mu_r}{\sigma_r}$$

এখানে $$\mu_r = \frac{2n_1n_2}{n}+1$$

$$\sigma r = \sqrt{\frac{2n_1n_2(2n_1n_2-n)}{n^2(n-1)}}$$

যদি $r \geq 35$ হয় তাহলে–

$$z = \frac{R+h-\mu_r}{\sigma_r}$$

এক্ষেত্রে $R < \mu_r$ হলে h = 0.5 হবে এবং $R > \mu_r$ হলে h = -0.5 হবে।

উদাহরণ: এক সারি বৃক্ষ রোগ নির্ণয়ের উদ্দেশ্যে পরীক্ষা করা হলো। রোগমুক্ত ও রোগযুক্ত বৃক্ষের অনুক্রম নিম্নরূপ–

H H H H I H I I I H I I I H H H H H I I I I

রোগমুক্ত বৃক্ষের অবস্থান দৈবচয়িত কিনা 5% গুরুত্বের স্তরে পরীক্ষা করুন।

সমাধানঃ নাস্তি অনুমান, H_o : রোগমুক্ত বৃক্ষের অবস্থান দৈবচয়িত।

বিকল্প অনুমান, H_A : রোগমুক্ত বৃক্ষের অবস্থান দৈবচয়িত নয়।

ধাপ-১: দেওয়া আছে, রোগমুক্ত ও রোগযুক্ত বৃক্ষের অনুক্রম–

H H H H I H I I I H I I I H H H H H I I I I

এখানে অনুক্রমের সংখ্যা, R = ৮; রোগমুক্ত বৃক্ষ (H) এর সংখ্যা, n_1 = 11; রোগযুক্ত বৃক্ষ (I) এর সংখ্যা, n_2 = 11 এবং n = 11+11=22

ধাপ-২: যেহেতু $n_1 = 11 < 20$ এবং $n_2 = 11 < 20$

দেয়া আছে গুরুত্বের স্তর, $\alpha = 5\% = 0.05$

$\therefore$ R -এর তাত্ত্বিক মান = 7 এবং 17

ধাপ-৩ : যেহেতু R এর মান (8) উহার তাত্ত্বিক মান 7 অপেক্ষা বড় কিন্তু 17 অপেক্ষা ছোট। অতএব নাস্তি অনুমান গ্রহণীয়। অর্থাৎ রোগমুক্ত বৃক্ষের অবস্থান দৈবচয়িত।

সারসংক্ষেপ

দুটি নমুনা একই জাতীয় সমগ্রক হতে নেওয়া হয়েছে কিনা পরীক্ষার ক্ষেত্রে ম্যান-হুইটনি U পরীক্ষা ব্যবহৃত হয়। অন্যদিকে যে পদ্ধতির সাহায্যে k সংখ্যক (দুয়ের অধিক) নমুনা একই সমগ্রক নেওয়া হয়েছে কিনা যাচাই করা হয় তাই ক্রুসক্যাল – ওয়ালিস H পরীক্ষা। ক্রুসক্যাল -ওয়ালিস H পরীক্ষার সূত্রঃ $H = \frac{12}{n(n+1)} \sum_{j=1}^{k} \frac{R_j^2}{n_j} - 3(n+1)$ । এছাড়া দৈব নমুনায়নের জন্য কোন পরিসংখ্যানিক পরীক্ষা ইতিপূর্বে আলোচিত হয় নাই। কোন নমুনার দৈবায়ন যে পরিসংখ্যানিক পরীক্ষার সাহায্যে যাচাই করা হয় তাকে অনুক্রমিক ঘটন পরীক্ষা (Run Test) বলে। যদি n_1 বা n_2 এর সংখ্যা 20 এর অধিক হয় তাহলে নিম্নের সূত্রে সাহায্যে z এর মান নির্ণয় করতে হবে। এর সূত্রঃ $z = \frac{R - \mu_r}{\sigma_r}$

পাঠ ৭.৫ স্পিয়ারম্যানের গুণাক্রমিক সংস্রব পরীক্ষা
Spearman's Rank Correlation Test

উদ্দেশ্য

এ পাঠ শেষে আপনি-

- স্পিয়ারম্যানের গুণাক্রমিক সংস্রব পরীক্ষার সংজ্ঞা লিখতে পারবেন।
- স্পিয়ারম্যানের গুণাক্রমিক সংস্রব পরীক্ষার প্রয়োগ করতে পারবেন।

স্পিয়ারম্যানের গুণাক্রমিক সংস্রব পরীক্ষা
Spearman's Rank Correlation Test

দুটি চলকের মধ্যকার সহ-সম্পর্ক পরীক্ষা করার জন্য অপরামাত্রিক পরীক্ষা ব্যবহার করা হয়। যদি চলকদ্বয়ের পরিমাপ সংখ্যাত্মকভাবে প্রকাশ করা না যায়, বরং র‍্যাংকিং (গুণাক্রমিক) মান জানা থাকে, তাহলে যে পরীক্ষার সাহায্যে চলকদ্বয়ের মধ্যকার সহসম্পর্ক যাচাই করা হয় তাই স্পিয়ারম্যানের গুণানুক্রমিক সংস্রব পরীক্ষা। এর ধাপসমূহ–

ধাপ-১ ঃ প্রথমে গুণাক্রমিক সংশ্লেষ (R) নির্ণয় করতে হবে। এটি নির্ণয়ের সূত্র নিম্নরূপ–

$$R = 1 - \frac{6\Sigma d^2}{n(n^2-1)}$$

এখানে

$d = R_1 - R_2$
R_1 = ১ম চলকের ক্রম (Rank)
R_2 = ২য় চলকের ক্রম (Rank)
n = প্রত্যেক চলকের সংখ্যা

ধাপ-২ ঃ এখন গুরুত্বের স্তর (α) এবং n এর সাহায্যে R এর তাত্ত্বিক মান সারণী হতে নির্ণয় করতে হবে।

ধাপ-৩ ঃ যদি R এর নির্ণীত মান উহার তাত্ত্বিক মান অপেক্ষা বড় হয় তাহলে নাস্তি অনুমান (H_o) বর্জনীয় হবে। অন্যথায় গ্রহণীয় হবে।

ধাপ-৪ ঃ যদি $n > 50$ হয় তাহলে নমুনাজমান z হবে–

$$z = R\sqrt{n-1}$$

যদি z এর নির্ণীত মান উহার তাত্ত্বিক মান অপেক্ষা বড় হয় তাহলে নাস্তি অনুমান (H_o) বর্জনীয় হবে।

উদাহরণ: দুজন বিচারক কর্তৃক কোন একটি প্রতিযোগিতায় অংশগ্রহণকারীদের র‍্যাংকিং (গুণানুক্রমিক) মান নিম্নরূপঃ

১ম বিচারক	1	2	3	4	5	6	7	8	9	10
২য় বিচারক	1	4	3	2	7	10	6	9	8	5

বিচারকদ্বয়ের রায়ের মধ্যে সামঞ্জস্য আছে কিনা 5% গুরুত্বের স্তরে পরীক্ষা করুন।

সমাধানঃ নাস্তি অনুমান, H_o : বিচারকদ্বয়ের রায়ের মধ্যে সামঞ্জস্যতা আছে।
বিকল্প অনুমান, H_A : বিচারকদ্বয়ের রায়ের মধ্যে সামঞ্জস্যতা নেই।

নিম্নে স্পিয়ারম্যানের গুণানুক্রমিক সংশ্লেষ পরীক্ষার সাহায্যে অনুমান যাচাই করা হলো–

সারণী

১ম বিচারকের গুণাক্রমিক মান (R_1)	২য় বিচারকের গুণাক্রমিক মান (R_2)	$d = R_1 - R_2$	d^2
1	1	0	0
2	4	-2	4
3	3	0	0
4	2	2	4
5	7	-2	4
6	10	-4	16
7	6	1	1
8	9	-1	1
9	8	1	1
10	5	5	25
			$\Sigma d^2 = 56$

ধাপ-১ ঃ গুণানুক্রমিক সংশ্লেষ,

$$R = 1 - \frac{6\Sigma d^2}{n(n^2-1)}$$

$$= 1 - \frac{6(56)}{10(10^2-1)}$$

$$= 1 - \frac{336}{990}$$

$$= 0.66$$

ধাপ-২ ঃ এখানে $n = 10$

দেয়া আছে গুরুত্বের স্তর, $\alpha = 5\% = 0.05$

$\therefore$ R-এর তাত্ত্বিক মান $= 0.648$

ধাপ-৩: যেহেতু R এর নির্ণীত মান (0.66) তার তাত্ত্বিক মান (0.648) অপেক্ষা বড়। অতএব নাস্তি অনুমান (H_o) বর্জনীয়। অর্থাৎ বিচারকদ্বয়ের রায়ের মধ্যে সামঞ্জস্যতা নেই।

সারসংক্ষেপ

দুটি চলকের মধ্যকার সহ-সম্পর্ক পরীক্ষা করার জন্য অপরামাত্রিক পরীক্ষা ব্যবহার করা হয়। যদি চলকদ্বয়ের পরিমাপ সংখ্যাত্মকভাবে প্রকাশ করা না যায়, বরং র‍্যাংকিং (গুণাক্রমিক) মান জানা থাকে, তাহলে যে পরীক্ষার সাহায্যে চলকদ্বয়ের মধ্যকার সহসম্পর্ক যাচাই করা হয় তাই স্পিয়ারম্যানের গুণানুক্রমিক সংস্রব পরীক্ষা। এর সূত্র:

$$R = 1 - \frac{6\Sigma d^2}{n(n^2-1)}$$

পাঠোত্তর মূল্যায়ন

রচনামূলক প্রশ্ন

১. অপরামাত্রিক পরীক্ষা কাকে বলে।
২. অপরামাত্রিক পরীক্ষার সুবিধা ও অসুবিধাগুলো লিখুন।
৩. অপরামাত্রিক পরীক্ষার ব্যবহার লিখুন।
৪. বিভিন্ন অপরামাত্রিক পরীক্ষার পদ্ধতিসমূহ বর্ণনা করুন।
৫. একটি প্রতিষ্ঠান হতে 17 জোড়া কর্মচারীকে বাছাই করা হল। প্রত্যেক জোড়ার একজনকে ১ম পদ্ধতিতে এবং অপরজনকে ২য় পদ্ধতিতে প্রশিক্ষণ দেয়া হল এবং প্রশিক্ষণ শেষে পরীক্ষা নেয়া হল। পরীক্ষার ফলাফল নিম্নে দেওয়া হলো–

জোড়া	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
১ম পদ্ধতি	65	68	70	63	73	75	72	78	64	73	79	80	67	64	62	74	82
২য় পদ্ধতি	63	68	68	60	72	75	73	70	66	70	77	78	63	65	60	74	78

চিহ্ন ঘটিত পরীক্ষা ব্যবহার করে (ক) পদ্ধতি দুটি সমান কার্যকরী (খ) ১ম পদ্ধতি ২য় পদ্ধতি অপেক্ষা অধিক কার্যকরী কিনা 5% তাৎপর্য স্তরে পরীক্ষা করুন।

৬. দুটি প্রতিষ্ঠান হতে নির্বাচিত দুটি নমুনার দ্রব্যগুলোর শক্তি নিম্নরূপঃ

প্রতিষ্ঠান - ১	প্রতিষ্ঠান - ২
18.3, 16.4, 22.7, 17.8, 18.9, 25.3, 16.1, 24.2	12.6, 14.1, 20.5, 10.7, 15.9, 19.6, 12.9, 15.2, 11.8, 147

দুটি নমুনার শক্তির মধ্যে পার্থক্য আছে কিনা ৫% গুরুত্বের স্তরে পরীক্ষা করুন।

৭. তিনটি নমুনার তথ্য নিম্নে দেয়া হলোঃ

১ম নমুনা	35, 24, 30, 43
২য় নমুনা	44, 42, 46
৩য় নমুনা	28, 15, 23, 38, 18

নমুনাগুলো একই সমগ্রক হতে নেওয়া হয়েছে কিনা (ক) 5% এবং (খ) 1% গুরুত্বের স্তরে পরীক্ষা করুন।

৮. এক সারি বৃক্ষ রোগ নির্ণয়ের উদ্দেশ্যে পরীক্ষা করা হলো। রোগমুক্ত ও রোগযুক্ত বৃক্ষের অনুক্রম নিম্নরূপ–

H H H H I H I I I H I I I H H H H H I I I I

রোগমুক্ত বৃক্ষের অবস্থান দৈবচয়িত কিনা 5% গুরুত্বের স্তরে পরীক্ষা করুন।

৯. দুজন বিচারক কর্তৃক কোন একটি প্রতিযোগিতায় অংশগ্রহণকারীদের র‍্যাংকিং (গুণানুক্রমিক) মান নিম্নরূপঃ

১ম বিচারক	1	2	3	4	5	6	7	8	9	10
২য় বিচারক	1	4	3	2	7	10	6	9	8	5

বিচারকদ্বয়ের রায়ের মধ্যে সামঞ্জস্য আছে কিনা 5% গুরুত্বের স্তরে পরীক্ষা করুন।

রেফারেন্স (References)

1. S.P.Gupta and M.P.Gupta (2023), Business Statistics, S Chand & Sons, New Delhi, India.
2. Richard I. Levin and D. S. Rubin (2023), Business Statistics, Prentice Hall Inc. New Delhi, India.
3. Murray R Spigel and Larry Stephens (2023), "Theory and Problems of Statistics: Schaum's Outline Series." McGraw Hill, New Delhi, India.
4. খন্দকার মোঃ সাদেকুর রহমান কাজল (২০২৪), অর্থনীতির জন্য পরিসংখ্যান, সমন্বয় পাবলিকেশন্স, ঢাকা।
5. মোঃ আব্দুল আজিজ(২০২৪), অর্থনীতির জন্য পরিসংখ্যান, দি এনজেল পাবলিকেশন্স, ঢাকা।
6. ড. নুর ইসলাম, আবুল খায়ের (২০২৪), দি ইউনাইটেড পাবলিশার্স, ঢাকা।