



B. L. D. E. Association's

S. B. ARTS & K. C. P. SCIENCE COLLEGE, VIJAYAPUR

First INTERNAL TEST 20 23 -2024

Name / ಹೆಸರು: Suchitra Gargu Chavan Roll No./ಸ್ಥಳಾಂಕ: 157 Date/ದಿನಾಂಕ: 11-07-2024

Subject/ವಿಷಯ: Statistics Semester/ಸೆಮಿಸ್ಟರ್: 4th Exam Seat No./ಪರೀಕ್ಷೆ ಸ್ಥಳಾಂಕ: U15KM2250247

Invigilator's Signature / ಕೊಠಡಿ ಮೇಲ್ವಿಚಾರಕರ ಸಹಿ Signature of Evaluator / ಪರೀಕ್ಷಕರ ಸಹಿ Marks obtained / ಪಡೆದ ಅಂಕಗಳು

30/30

15/15 Marks
12/10/24

I

1. Define order statistic?

→ Let $x_1, x_2, \dots, x_n$ be observation of random sample from continuous type distribution, then $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ is said to be order statistics

where $X_{(1)}$ is the smallest order statistics

$X_{(2)}$ is the second smallest order statistics

$X_{(n)}$ is the largest order statistics

2. Define parameter and statistics.

→ parameter: The statistical constant of the population is called parameter.

Statistics: The function of sample value is called Statistics.

3. Define estimator and estimate.

→ Estimator: A function of a sample observation $T = T(x_1, x_2, \dots, x_n)$ with unknown estimator value is called estimator.

estimate: The particular value of estimator is called estimator.

5. Define consistent estimator

→ The estimator $T_n = T(x_1, x_2, \dots, x_n)$ for a random sample of size n is said to be consistent estimator of θ with the parametric space if T_n converges θ in probability.

4. State Cramer Rao Inequality

→ It provides lower bound variance of unbiased estimator is called Cramer Rao Inequality.

II.

7. If T is consistent estimator of θ then show that T^2 is consistent estimator of θ^2 .

→ T is consistent estimator of θ

$$T \xrightarrow{P} \theta$$

$$P\{|t - \theta| < \varepsilon\} \rightarrow 1 \text{ as } n \rightarrow \infty$$

$$P\{-\varepsilon < t - \theta < \varepsilon\} \rightarrow 1$$

- Adding θ on both sides.

$$P\{-(\varepsilon + \theta) < t - \theta + \theta < (\varepsilon + \theta)\} \rightarrow 1$$

$$P\{-(\varepsilon + \theta) < t < (\varepsilon + \theta)\} \rightarrow 1$$

- Squaring on both sides.

$$P\{-(\varepsilon + \theta)^2 < t^2 < (\varepsilon + \theta)^2\}$$

$$P\{-(\varepsilon^2 + \theta^2 + 2\varepsilon\theta) < t^2 < (\varepsilon^2 + \theta^2 + 2\varepsilon\theta)\} \rightarrow 1$$

- Subtracting θ^2 on both sides

$$P\{-(\varepsilon^2 + \theta^2 + 2\varepsilon\theta - \theta^2) < t^2 - \theta^2 < (\varepsilon^2 + \theta^2 + 2\varepsilon\theta - \theta^2)\} \rightarrow 1$$

$$P\{-(\varepsilon^2 + 2\varepsilon\theta) < t^2 - \theta^2 < (\varepsilon^2 + 2\varepsilon\theta)\} \rightarrow 1$$

ε is very small.

So,

$$\varepsilon^2 + 2\varepsilon\theta = \varepsilon_1$$

$$P\{-\varepsilon_1 < t^2 - \theta^2 < \varepsilon_1\} \rightarrow 1 \text{ as } n \rightarrow \infty$$

$$P\{|t^2 - \theta^2| < \varepsilon_1\}$$

$\therefore t^2$ is consistent estimator of θ^2 .

Find the sufficient statistic for θ in $f(x, \theta) = \frac{e^{-\theta} \theta^x}{x!}$

$$f(x, \theta) = \frac{e^{-\theta} \theta^x}{x!}$$

$$L = \prod_{i=1}^n \frac{e^{-\theta} \theta^{x_i}}{x_i!}$$

$$L = \frac{e^{-n\theta} \theta^{\sum x_i}}{\prod_{i=1}^n x_i!} \quad \text{--- (1)}$$

WKT:-

$$L = g\left[\sum x_i\right] h(x)$$

$\therefore \sum x_i$ is the sufficient estimator.

Equation (1) becomes,

$$L = e^{-n\theta} \theta^{\sum x_i} \frac{1}{\prod_{i=1}^n x_i!}$$

Here,

$$g(\sum x_i) = e^{-n\theta} \theta^{\sum x_i}$$

$$h(x) = \frac{1}{\prod_{i=1}^n x_i!}$$

$$L = [g(\sum x_i) h(x)]$$

$\therefore \sum x_i$ is the sufficient estimator.

Hence By Neyman's Factorization theorem

$L = [g(\sum x_i) h(x)]$ is sufficient estimator of θ .

III.

11. State and prove Neyman-FactORIZATION theorem.
 Statement:- Let $x_1, x_2, \dots, x_n$ be a random sample from θ and $T = t(x_1, x_2, \dots, x_n)$ be a statistic, then t is sufficient for θ , if the likelihood function of x can be expressed as.

$$L = \{g(T, \theta) h(x)\}$$

where,

$g(T, \theta)$ is the function of T & θ .

$h(x)$ is the independent of θ .

Proof:-

Case (i)

Suppose that T is sufficient of θ
 i.e.,

$$P(X|T) = h(x)$$

$$\frac{P(X=x|T=t)}{P(T=t)} = h(x)$$

$$\therefore P(A/B) = \frac{P(A \cap B)}{P(B)}$$

$$\frac{P(X=x \cap T=t)}{P(T=t)} = h(x)$$

$$P(X=x \cap T=t) = P(T=t) h(x)$$

$$P(X=x) = P(T=t) h(x)$$

$$P(x_1, x_2, \dots, x_n = x) = P(T=t) h(x)$$

$$P(x_1=x_1, x_2=x_2, \dots, x_n=x_n) = P(T=t) h(x)$$

$$P(x_1=x_1, x_2=x_2, \dots, x_n=x_n) = P(T=t) h(x)$$

$$P(x_1=x_1) P(x_2=x_2) \dots P(x_n=x_n) = P(T=t) h(x)$$

$$\prod_{i=1}^n P(x_i) = P(T=t) h(x)$$

$$L = g(T, \theta) h(x)$$

Case (ii)

Suppose that $L = g(x, \theta) h(x)$

$$P(X=x) = g(x, \theta) h(x)$$

$$\frac{P(X=x/T=t)}{g(t, \theta)} = h(x)$$

$$\frac{P(X=x/T=t)}{P(T=t)} = h(x).$$

$$P(X|T) = h(x).$$

$\therefore T$ is the sufficient estimator of θ .

I.

4. state Cramer Rao Inequality.

→ It provides the lower bound for variance of unbiased estimator, then it is said to be

Cramer Rao Inequality

Name / ಹೆಸರು: Vijayalaxmi, Hadimani, Roll No./ ಸ್ಥಳಾಂಕ: 160, Date/ ದಿನಾಂಕ: 11-7-2024Subject/ ವಿಷಯ: Statistics, Semester/ ಸಮಿಸ್ತರ: IV, Exam Seat No / ಪರೀಕ್ಷೆ ಸ್ಥಳಾಂಕ: _____UUCMS No.: U15KM2230072

Invigilator's Signature / ಕೊಠಡಿ ಮೇಲ್ವಿಚಾರಕರ ಸಹಿ Signature of Evaluator / ಪರೀಕ್ಷಕರ ಸಹಿ Marks obtained/ ಪಡೆದ ಅಂಕಗಳು

30/30

Q-I

15/15 Shokee
12/07/24

1) Order Statistic :- Let $x_1, x_2, \dots, x_n$ be a random sample observations from continuous type distribution then $x_1, x_2, \dots, x_n$ are called order statistic. where,

x_1 :- Smallest value in $x_1, x_2, \dots, x_n$

x_2 :- next larger value in $x_1, x_2, \dots, x_n$

x_n :- highest value in $x_1, x_2, \dots, x_n$.

2) Parameter :- It is a statistical constant of the Population is called Parameter.

Statistic :- It is a function of sample values is called Statistic.

3) Estimator :- The Function of sample observation $T_n = T(x_1, x_2, \dots, x_n)$ which estimate the unknown value of the parameter is called Estimator.

Estimate :- The Particular value of the estimator is called Estimate.

4) Cramer-Rao Inequality :-
T is an ^{unbiased} estimator of θ then
 $Var(T)$ is $Var(T) = \frac{1}{I(\theta)}$

where, $\theta \rightarrow$ Population Parameter

$I(\theta) \rightarrow$ Fisher's information

then $I(\theta) = -E \left[\frac{\partial^2 \log L}{\partial \theta^2} \right]$

L is likelihood function

L = Product of $f(x, \theta)$

5) Consistent estimator :- ^{an estimator.} $T_n = T(x_1, x_2, \dots, x_n)$ be a random samples based on population size 'n'. is said to be consistent estimator then $T \xrightarrow{P} \theta$ as $n \rightarrow \infty$.
 T converges θ in Probability.

Q-II

7) Given $T \xrightarrow{P} \theta$ as $n \rightarrow \infty$
 i.e. $P\{|T - \theta| < \epsilon\} \rightarrow 1$ as $n \rightarrow \infty$
 $P\{-\epsilon < T - \theta < \epsilon\} \rightarrow 1$ as $n \rightarrow \infty$

Adding θ both side.

$$P\{-\epsilon + \theta < T < \epsilon + \theta\} \rightarrow 1 \text{ as } n \rightarrow \infty$$

$P\{\text{squaring on both side}\}$

$$P\{-(\epsilon + \theta)^2 < T^2 < (\epsilon + \theta)^2\} \rightarrow 1 \text{ as } n \rightarrow \infty$$

$$P\{-(\epsilon^2 + \theta^2 + 2\epsilon\theta) < T^2 < (\epsilon^2 + \theta^2 + 2\epsilon\theta)\} \rightarrow 1 \text{ as } n \rightarrow \infty$$

$P\{\text{Subtract } \theta^2 \text{ on both side}\}$

$$P\{-(\epsilon^2 + 2\epsilon\theta) < T^2 - \theta^2 < (\epsilon^2 + 2\epsilon\theta)\} \rightarrow 1 \text{ as } n \rightarrow \infty$$

Here ϵ is very small $\therefore \epsilon^2 + 2\epsilon\theta = \epsilon_1$

$$\therefore P\{-\epsilon_1 < T^2 - \theta^2 < \epsilon_1\} \rightarrow 1 \text{ as } n \rightarrow \infty$$

$$P\{|T^2 - \theta^2| < \epsilon_1\} \rightarrow 1 \text{ as } n \rightarrow \infty$$

$\therefore T^2$ is a consistent estimator of θ^2

9) Given $f(x, \theta) = \frac{e^{-\theta} \theta^x}{x!}$

$$\text{likelihood fn } L = \prod_{i=1}^n \frac{e^{-\theta} \theta^{x_i}}{x_i!}$$

$$= \frac{e^{-n\theta} \theta^{\sum x_i}}{\prod_{i=1}^n x_i!} \quad \text{--- (1)}$$

WKT $\neq$ if

$$L = g[t(x) \cdot h(x)]$$

where $t(x)$ is consistent estimator.

$$\text{eqn (1) becomes, } L = \frac{e^{-n\theta} \theta^{\sum x_i}}{\prod_{i=1}^n x_i!}$$

$$L = g(\sum x_i) \cdot h(x) \quad \because \left[\text{Here } g(\sum x_i) = e^{-n\theta} \theta^{\sum x_i} \right]$$

then $\sum x_i$ is consistent estimator

hence by the Neyman-Fisher Factorization theorem

$\sum_{i=1}^n x_i$ is consistent estimator for θ .

Q-III

11) Neyman-Factorization Theorem.

Statement:- Let $x_1, x_2, \dots, x_n$ be the random sample from $f(x, \theta)$ and $T = T(x_1, x_2, \dots, x_n)$ be a statistic. Then T is sufficient for θ if likelihood function of $x_1, x_2, \dots, x_n$ can be expressed as $L(x_1, x_2, \dots, x_n) = g(T, \theta) \cdot h(x)$ where $g(T, \theta)$ is the function of T and θ and $h(x)$ is independent of θ .

Proof:- [Discrete case]

Case 1:- Suppose that T is sufficient for θ .

$$P(X/T) = h(x)$$

$$P(X=x/T=t) = h(x)$$

$$\frac{P(X=x \cap T=t)}{P(T=t)} = h(x)$$

$$P(X=x \cap T=t) = P(T=t) \cdot h(x)$$

$$P(X=x) = P(T=t) \cdot h(x)$$

$$P(x_1, x_2, \dots, x_n = x) = P(T=t) \cdot h(x)$$

$$P(x_1=x_1, x_2=x_2, \dots, x_n=x_n) = P(T=t) \cdot h(x)$$

$$P(x_1=x_1), P(x_2=x_2), \dots, P(x_n=x_n) = P(T=t) \cdot h(x)$$

$$\prod_{i=1}^n P(x_i) = P(T=t) \cdot h(x)$$

$$L = g(T, \theta) \cdot h(x)$$

Case 2:- Suppose that $L = g(T, \theta) \cdot h(x)$

$$P(X=x) = g(T, \theta) \cdot h(x)$$

$$\frac{P(X=x)}{g(T, \theta)} = h(x)$$

$$P$$

$$\frac{P(X=x/T=t)}{P(T=t)} = h(x)$$

$$P(X/T) = h(x)$$

∴ T is sufficient for θ

10



Name / ಹೆಸರು: Mehboob Makandar Roll No. / ಸ್ಥಳಾಂಕ: 09 Date / ದಿನಾಂಕ: 08-8-2024
 Subject / ವಿಷಯ: Statistics Semester / ಸೆಮಿಸ್ಟರ್: IV Exam Seat No. / ಪರೀಕ್ಷೆ ಸ್ಥಳಾಂಕ: UISKMA22S0107

Supervisor's Signature / ಕುರಿತು ಮೇಲ್ವಿಚಾರಕರ ಸಹಿ Signature of Evaluator / ಪರೀಕ್ಷಕರ ಸಹಿ Marks obtained / ಪಡೆದ ಅಂಕಗಳು

30 / 30

15/15
 18/24

I

1) Type I Error :- The error of rejecting null hypothesis (H_0) when it is actually true.

2) Type 2 Error :- The error of accepting null hypothesis (H_0) when it is actually not true.

3) Null hypothesis :- It is a hypothesis which is tested for possible rejection under the assumption that it is true. Denoted by H_0 .

4) Alternative hypothesis :- A hypothesis which is complementary to null hypothesis. Denoted by H_1 .

5) Simple hypothesis :- It is a statistical hypothesis which specifies the population completely then it is termed as simple hypothesis.

6) Composite hypothesis :- A hypothesis which does not completely specify the population is termed as composite hypothesis.

7) 95% Confidence Limits are
 Lower Confidence Limits (LCL) : $\bar{x} - (z_{\alpha/2} \sigma / \sqrt{n})$
 Upper Confidence Limits (UCL) : $\bar{x} + (z_{\alpha/2} \sigma / \sqrt{n})$

Where $z_{\alpha/2} = 1.96$

Confidence interval for mean is $[\bar{x} - (z_{\alpha/2} \sigma / \sqrt{n}), \bar{x} + (z_{\alpha/2} \sigma / \sqrt{n})]$

5) 99% Confidence Limits are
Lower Confidence Limits (LCL) = $\bar{x} - (z_{\alpha/2} \cdot \sigma / \sqrt{n})$

Upper Confidence Limits (UCL) = $\bar{x} + (z_{\alpha/2} \cdot \sigma / \sqrt{n})$

Where $z_{\alpha/2}$

II

8) Let $x_1, x_2, \dots, x_n$ be a random sample of size n drawn from a population with pdf $f(x, \theta) = (1+\theta)x^\theta, 0 < x < 1, 0 \leq \theta$.

First population moment about origin is

$$\begin{aligned} \mu'_1 &= \int_0^1 x f(x) dx \\ &= \int_0^1 x (1+\theta) x^\theta dx \\ &= (1+\theta) \int_0^1 x^{\theta+1} dx \\ &= (1+\theta) \left[\frac{x^{\theta+2}}{\theta+2} \right]_0^1 \\ &= \frac{1+\theta}{\theta+2} (1-0) \end{aligned}$$

$$\mu'_1 = \frac{1+\theta}{\theta+2} \quad \text{--- (1)}$$

First sample moments about origin.

$$m'_1 = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}$$

By equating population moment about Sample moment

$$\frac{1+\theta}{\theta+2} = \bar{x}$$

$$1+0 = x(2+0)$$

$$1+0 = 2x + 0x$$

$$1-2x = 0x - 0$$

$$1-2x = 0(x-1)$$

$$0' = \frac{1-2x}{x-1} \quad \hat{\theta} = \frac{2x-1}{1-x}$$

9) The given prob. function.

$$f(x, \theta) = \frac{1+\theta}{\theta}, 0 \leq x \leq \theta$$

$H_0: \theta = 1$ against $H_1: \theta = 2$

The Likelihood func L_0 under H_0 is

$$L_0 = \frac{1}{\theta} = 1$$

The Likelihood func L_1 under H_1 is

$$L_1 = \frac{1}{\theta} = \frac{1}{2}$$

Critical region: $\{w: 0.5 \leq x\} = \{w: 0.5 \leq x \leq \theta\}$

Acceptance region: $\{\bar{w}: 0 \leq x < 0.5\}$

Type I error

$$\alpha = \int_w L_0 dx = \int_{0.5}^{\theta} L_0 dx = \int_{0.5}^{\theta} 1 dx = [x]_{0.5}^{\theta}$$

$$\alpha = 0.5$$

Type II error

$$\beta = \int_w L_1 dx = \int_{0.5}^{\theta} L_1 dx = \int_{0.5}^{\theta} \frac{1}{2} dx$$

$$\frac{1}{2} [x]_{0.5}^{\theta} = \frac{1}{2} \times 0.5$$

$$\beta = 0.25$$

Power of Test: $1 - \beta$

$$= 1 - 0.25 = 0.75$$

III

1. Let x_1 be no of observations possessing an attribute A in one sample of size n_1 and Let x_2 be no of observations possessing an attribute A in other sample of size n_2 then,

$$p_1 = \frac{x_1}{n_1}, \quad p_2 = \frac{x_2}{n_2}$$

$$q_1 = 1 - p_1, \quad q_2 = 1 - p_2$$

Let P_1 be the first population proportion

Let P_2 be the second

Then $Q_1 = 1 - P_1$ and $Q_2 = 1 - P_2$

For large sample

$$p_1 \sim N\left(P_1, \frac{P_1 Q_1}{n_1}\right), \quad p_2 \sim N\left(P_2, \frac{P_2 Q_2}{n_2}\right)$$

Then

$$p_1 - p_2 \sim N\left(P_1 - P_2, \frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}\right)$$

But P_1, P_2, Q_1 and Q_2 are unknown

They are replaced by corresponding

sample values p_1, p_2, q_1, q_2 respectively

Then

$$p_1 - p_2 \sim N\left(P_1 - P_2, \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}\right)$$

SNV is (standard normal variable).

$$Z = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}} \sim N(0, 1)$$

By Probability Curve,

$$P[-Z_{\alpha/2} < Z < Z_{\alpha/2}] = 1 - \alpha$$

$$P \left[-Z_{\alpha/2} < \frac{(P_1 - P_2) - (P_1 - P_2)}{\sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}}} < Z_{\alpha/2} \right] = 1 - \alpha$$

$$P \left[-Z_{\alpha/2} \sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}} < (P_1 - P_2) - (P_1 - P_2) < Z_{\alpha/2} \sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}} \right]$$

$$= 1 - \alpha$$

$$P \left[-\frac{(P_1 - P_2) - Z_{\alpha/2} \sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}}}{(P_1 - P_2) + Z_{\alpha/2} \sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}}} < - \right]$$

$$P \left[(P_1 - P_2) + Z_{\alpha/2} \sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}} > P_1 - P_2 > \right]$$

$$(P_1 - P_2) - Z_{\alpha/2} \sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}} \right] = 1 - \alpha$$

$$P \left[(P_1 - P_2) - Z_{\alpha/2} \sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}} < P_1 - P_2 < \right]$$

$$(P_1 - P_2) + Z_{\alpha/2} \sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}} \right] = 1 - \alpha$$

$\therefore 100(1 - \alpha)\%$ confidence interval for the difference of two population proportions for large samples is

$$\left[(P_1 - P_2) - Z_{\alpha/2} \sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}}, (P_1 - P_2) + Z_{\alpha/2} \sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}} \right]$$

And confidence limits are.

$$LCL = (p_1 - p_2) - Z_{\alpha/2} \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

$$UCL = (p_1 - p_2) + Z_{\alpha/2} \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

~~$$LCL = (p_1 - p_2) - Z_{\alpha/2} \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$~~

$LCL =$

$$= (p_1 - p_2) - Z_{\alpha/2} \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

$$LCL = (p_1 - p_2) - Z_{\alpha/2} \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

$$UCL = (p_1 - p_2) + Z_{\alpha/2} \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

$$LCL = (p_1 - p_2) - Z_{\alpha/2} \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

$$UCL = (p_1 - p_2) + Z_{\alpha/2} \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

$$LCL = (p_1 - p_2) - Z_{\alpha/2} \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

for large samples is difference of two population proportions confidence interval for the

$$LCL = (p_1 - p_2) - Z_{\alpha/2} \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

Name / ಹೆಸರು: Aradi S. Hosamani Roll No./ಸ್ಥಳಾಂಕ: 136 Date/ದಿನಾಂಕ: 08-08-2024Subject/ವಿಷಯ: Statistics Semester/ಸೆಮಿಸ್ಟರ್: IV Exam Seat No / ಪರೀಕ್ಷೆ ಸ್ಥಳಾಂಕ: U15KM2280154
(Statistical Inference - I)

Invigilator's Signature / ಕೊಠಡಿ ಮೇಲ್ವಿಚಾರಕರ ಸಹಿ Signature of Evaluator / ಪರೀಕ್ಷಕರ ಸಹಿ Marks obtained / ಪಡೆದ ಅಂಕಗಳು

30/30
15/15
D. Lakshmi
09/08/24

1] Type I Error:
The error of rejecting null hypothesis if it is actually true.

2] Type II Error:
The error of accepting null hypothesis if it is actually not true.

2] Null hypothesis:
The hypothesis which is tested for possible rejection hypothesis under the assumption that it is true, is called null hypothesis.

2] Alternative Hypothesis:
The hypothesis which is complementary to the null hypothesis is called Alternative hypothesis.

3] Simple Hypothesis:
It is statistical hypothesis specified the population completely it is termed as simple hypothesis.

2] Composite Hypothesis:
A hypothesis which does not completely specify the population it is termed as composite hypothesis.

4] 95% confidence limit:
Lower confidence limit (LL) = $\bar{x} - \left(Z_{1/2} \cdot \frac{\sigma}{\sqrt{n}} \right)$
Upper confidence limit (UL) = $\bar{x} + \left(Z_{1/2} \cdot \frac{\sigma}{\sqrt{n}} \right)$
where $Z_{1/2} = 1.96$
confidence interval for mean:
 $\left[\bar{x} - \left(Z_{1/2} \cdot \frac{\sigma}{\sqrt{n}} \right), \bar{x} + \left(Z_{1/2} \cdot \frac{\sigma}{\sqrt{n}} \right) \right]$

Q9. confidence limit for proportion.

5] the point estimation in general does not coincide with the true value of the parameter. It is preferred to estimate to obtain a large interval which may be expected to cover the true value of the parameter with some defined prob. or confidence such an interval is called confidence interval & this process is called confidence interval.

8] let $x_1, x_2, \dots, x_n$ be a random sample of size n drawn from the population with pdf $f(x, \theta) = (1+\theta)x^\theta$, $0 \leq x \leq 1$, $0 \leq \theta$

the likelihood function of $x_1, x_2, \dots, x_n$

$$L = (1+\theta)^n$$

$$L = (1+\theta)x_1^\theta (1+\theta)x_2^\theta \dots (1+\theta)x_n^\theta$$

First population mean for origin is

$$\mu_1' = \int x f(x) dx$$

$$= \int_0^1 x (1+\theta) x^\theta dx$$

$$= \int_0^1 x^{\theta+1} (1+\theta) dx$$

$$= (1+\theta) \int_0^1 x^{\theta+1} dx$$

$$= (1+\theta) \left[\frac{x^{\theta+2}}{\theta+2} \right]_0^1$$

$$= \frac{(1+\theta)}{\theta+2} \left[x^{\theta+2} \right]_0^1$$

$$= \frac{(1+\theta)}{\theta+2} [1-0]$$

$$\mu_1' = \frac{1+\theta}{\theta+2}$$

First sample moment of origin is given by

$$m_1' = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}$$

$$u' = m'$$

$$\frac{1+0}{2+0} = \bar{x}$$

$$1+0 = \bar{x}(2+0)$$

$$1+0 = 2\bar{x} + \bar{x}0$$

$$1 - 2\bar{x} = \bar{x}0 - 0$$

$$1 - 2\bar{x} = 0\bar{x}(\bar{x}-1)$$

$$\hat{\theta} = \frac{1-2\bar{x}}{\bar{x}-1}$$

9] The given probability is

$$f(x, \theta) = \begin{cases} (1+\theta) \frac{1}{\theta} & 0 \leq x \leq \theta \\ 0 & \text{o.w.} \end{cases}$$

$$H_0: \theta = 1 \quad \text{against} \quad H_1: \theta = 2$$

The likelihood function of L_0 under the H_0 is

$$L_0 = \frac{1}{\theta} = \frac{1}{1} = 1$$

The likelihood function of L_1 under the H_1 is

$$L_1 = \frac{1}{\theta} = \frac{1}{2}$$

$$\text{Critical region} = \{w: 0.5 \geq x\}$$

$$= \{w: 0.5 \leq x \leq \theta\}$$

$$\text{Acceptance region} = \{w: 0.5 < x \leq \theta\}$$

Type I:

$$\alpha = \int_{\omega} f_0(x) dx$$

$$\alpha = \int_{0.5}^{\theta} 1 dx = \left[x \right]_{0.5}^{\theta} = \theta - 0.5 = 2 - 0.5 = 1.5$$

$$\boxed{\alpha = 0.5}$$

Type II:

$$\beta = \int_{\omega} f_1(x) dx = \int_{\omega} \frac{1}{2} dx$$

$$= \frac{1}{2} [x]_0^{\theta} = \frac{1}{2} (0.5) = 0.25$$

$$\boxed{\beta = 0.25}$$

power of test

$$1 - \beta = 1 - 0.25$$

$$1 - \beta = 0.75$$

Critical region = $\{w : 1 \leq x \leq 1.5\}$

Acceptance region = $\{w : x < 1 \text{ or } x \geq 1.5\}$

acceptance region = $\{w : x < 1 \text{ or } x \geq 1.5\}$ $0 \leq x < 1$

Type I:

$$\alpha = \int_{0.5}^{1.5} f(x) dx$$

$$= \int_{0.5}^{1.5} 1 dx$$

$$= [x]_{0.5}^{1.5} = [1.5 - 0.5]$$

$$\alpha = 0.5$$

$$\beta = 1 - \int_{0.5}^{1.5} f(x) dx$$

$$= 1 - \int_{0.5}^{1.5} \frac{1}{2} dx$$

$$= 1 - \frac{1}{2} [x]_{0.5}^{1.5}$$

$$= 1 - \frac{1}{2} [1.5 - 0.5] = 1 - \frac{1}{2} (1) = 1 - 0.5$$

$$= 1 - 0.5 = 0.5$$

$$\beta = 0.5$$

$$\beta = 0.5$$

$$\beta = 0.5$$

power of test

$$1 - \beta = 1 - 0.25$$

$$= 0.75$$

$$= 0.75$$

$$1 - \beta = 1 - 0.75$$

$$= 0.25$$

11

12) let $x_1, x_2, \dots, x_n$ be a random sample of size n drawn from the prob. distn with pdf

$$f(x; \theta) = dF(x) = \theta e^{-x\theta}$$

The likelihood Function of $x_1, x_2, \dots, x_n$

$$L = dF(x_1) \cdot dF(x_2) \cdot \dots \cdot dF(x_n)$$

$$= \theta e^{-x_1\theta} \cdot \theta e^{-x_2\theta} \cdot \dots \cdot \theta e^{-x_n\theta}$$

$$L = \theta^n e^{-\theta \sum x_i}$$

Taking log on both side

$$\log L = \log (\theta^n e^{-\theta \sum x_i})$$

$$\log L = n \log \theta - \theta \sum x_i$$

differentiating partially w.r.t θ

$$\frac{\partial \log L}{\partial \theta} = n \left(\frac{1}{\theta} - \bar{x} \right)$$

Again diff w.r.t θ

$$\frac{\partial^2 \log L}{\partial \theta^2} = n \left(\frac{-1}{\theta^2} \right) = \frac{-n}{\theta^2}$$

$$V \left(\frac{\partial \log L}{\partial \theta} \right) = E \left(\frac{\partial^2 \log L}{\partial \theta^2} \right)$$

$$= E \left(\frac{-n}{\theta^2} \right)$$

$$= \frac{n}{\theta^2}$$

$$Z = \frac{\frac{\partial \log L}{\partial \theta}}{\sqrt{V \frac{\partial \log L}{\partial \theta}}} = \frac{n \left(\frac{1}{\theta} - \bar{x} \right)}{\sqrt{n/\theta^2}}$$

$$Z = \sqrt{n} (1 - \theta \bar{x}) \sim N(0, 1)$$

$\mathbb{P} =$

$$\mathbb{P} [|z| \leq \lambda_2]$$

$$\mathbb{P} [-\lambda_2 \leq z \leq \lambda_2]$$

$$\mathbb{P} [-\lambda_2 \leq \sqrt{n} (1 - \theta \bar{x}) \leq \lambda_2]$$

$$\mathbb{P} \left[-\frac{\lambda_2}{\sqrt{n}} \leq (1 - \theta \bar{x}) \leq \frac{\lambda_2}{\sqrt{n}} \right]$$

$$\mathbb{P} \left[1.96 \leq \sqrt{n} (1 - \theta \bar{x}) \leq 1.96 \right]$$

$$\mathbb{P} \left[\frac{1.96}{\sqrt{n}} \leq 1 - \theta \bar{x} \leq \frac{1.96}{\sqrt{n}} \right]$$

$$\mathbb{P} \left[\frac{1.96}{\sqrt{n}} - 1 \leq 0 \leq 1.96 - \frac{1.96}{\sqrt{n}} \right]$$

$$\mathbb{P} \left[\frac{1.96 + 1}{\sqrt{n}} \geq 0 \geq \frac{1.96}{\sqrt{n}} + \frac{1}{\sqrt{n}} \right]$$

$$\mathbb{P} \left[\frac{1 + 1.96}{\sqrt{n}} \geq 0 \geq \frac{1.96 - 1}{\sqrt{n}} \right]$$

$$\mathbb{P} \left[\left(\frac{1 + 1.96}{\sqrt{n}} \right) / \bar{x} \geq 0 \geq \left(\frac{1.96 - 1}{\sqrt{n}} \right) / \bar{x} \right]$$

$$\mathbb{P} = \left(\frac{1 \pm 1.96}{\sqrt{n}} \right) / \bar{x}$$

Hence central confidence limits for large sample with 95% confidence is

$$\hat{\theta} = \left(\frac{1 \pm 1.96}{\sqrt{n}} \right) / \bar{x}$$

B.L.D.E. Association's

**S. B. Arts & K.C.P. Science College,
VIJAYAPUR- 586 103.**



ASSIGNMENT

For B.A./ B.Sc.[✓]IV..... Semester
2023 - 2024

Name of the Student Alfiya. Choudhary

Roll No. 105 R.C.U. Seat No. U15KM22S0124

Subject Statistical Interface - I

Assignment No.	Date	Marks Assigned	Marks Obtained	Name and Signature of Teacher	Remarks
1	02/08/24	05	05	J.D. Horaker's Place	Good
2					
3					
4					

SECTION-A

Answer any five questions.

1. Define estimator and estimate.

Estimator: the function of sample observations $T_n = T(x_1, x_2, \dots, x_n)$ which estimates the unknown value of parameter is called estimator.

Estimate: the particular value of the estimator is called estimate.

2. Define consistent estimator.

An estimator $T_n = T(x_1, x_2, \dots, x_n)$ based on a random sample of size n is said to be consistent estimator of θ in the parametric space if T_n converges θ in probability i.e., if $T_n \xrightarrow{P} \theta$ as $n \rightarrow \infty$

In other words, T_n is consistent estimator of θ if for every $\epsilon > 0$, $\eta > 0$, there exists a positive integer n such that

$$P\{|T_n - \theta| < \epsilon\} \rightarrow 1 \text{ as } n \rightarrow \infty$$

$$P\{|T_n - \theta| < \epsilon\} \geq 1 - \eta \quad \forall n \text{ where } \theta \text{ is the parameter.}$$

3. Define null and alternative hypothesis.

Null hypothesis is a hypothesis which is tested for possible rejection under the assumption that it is true. Denoted by H_0 .

Alternative hypothesis is a hypothesis which is complementary to null hypothesis is called Alternative hypothesis. Denoted by H_1 .

4. Define most power full test.

Most Powerfull Test (MP-test):

Let us consider a problem of testing simple null hypothesis i.e., $H_0: \theta = \theta_0$ against alternative hypothesis $H_1: \theta \neq \theta_1$. The critical region W is the most powerful rejection region of size α for testing $H_0: \theta = \theta_0$ against $H_1: \theta = \theta_1$

$$P[\text{Type-I error}] = \alpha = P[X \in W / H_0 \text{ is true}] = \int_W L_0 dz \rightarrow (1)$$

where L_0 is likelihood function under H_0

$$\text{and } P[X \in W / H_1 \text{ is true}] \geq P[X \in W_1 / H_1 \text{ is true}]$$

$$1 - \beta \geq 1 - \beta_1 \rightarrow (2)$$

for every other critical region, W_1 satisfying eqⁿ (1)

5. State properties of method of moment estimator.

Properties

i) The first and second order derivatives $\frac{\partial}{\partial \theta} \log L$ and $\frac{\partial^2}{\partial \theta^2} \log L$ exists and are continuous function of θ (parameter). On the range for almost all x . $\left| \frac{\partial}{\partial \theta} \log L \right| < F_1(x)$ and $\left| \frac{\partial^2}{\partial \theta^2} \log L \right| < F_2(x)$ $\forall \theta$ where $F_1(x)$ and $F_2(x)$ are integrable f^2 s over $(-\infty, \infty)$.

ii) The third order derivative $\frac{\partial^3}{\partial \theta^3} \log L$ exists such that $\left| \frac{\partial^3}{\partial \theta^3} \log L \right| < M(x)$ where $M(x)$ is given by $E[M(x)] < k$ where k is some +ve quantity.

iii) For every θ in k $E \left[\frac{-\partial^2}{\partial \theta^2} \log L \right] = \int_{-\infty}^{\infty} \left(\frac{-\partial^2}{\partial \theta^2} \log L \right) L \cdot dx = I\theta$ is finite and non zero.

SECTION-B

Answer any four questions.

1) If $x_1, x_2, \dots, x_n$ is a random sample from $N(0, \sigma^2)$ distribution, then show that $\frac{1}{n} \sum x_i^2$ is an unbiased estimator of σ^2 .

Given $E(x_i) = 0$, $V(x_i) = \sigma^2$ $\forall i = 1, 2, \dots, n$

$$\text{WKT } V(x_i) = E(x_i^2) - [E(x_i)]^2$$

$$\Rightarrow E(x_i^2) = V(x_i) + [E(x_i)]^2$$

$$= \sigma^2 + 0^2$$

$$\Rightarrow E(x_i^2) = \sigma^2$$

$$E \left(\frac{1}{n} \sum_{i=1}^n x_i^2 \right) = \frac{1}{n} E \left(\sum_{i=1}^n x_i^2 \right)$$

$$= \frac{1}{n} \sum_{i=1}^n E(x_i^2)$$

$$= \frac{1}{n} \sum_{i=1}^n (\sigma^2)$$

$$= \frac{n\sigma^2}{n}$$

$$\therefore E \left(\frac{1}{n} \sum_{i=1}^n x_i^2 \right) = \sigma^2$$

$\therefore \frac{1}{n} \sum x_i^2$ is an unbiased estimator of $\sigma^2(x)$

2> Find the sufficient statistic for θ in $f(x, \theta) = \frac{e^{-\theta} \theta^x}{x!}$, $x = 0, 1, 2, \dots, \infty$

Given $f(x, \theta) = \frac{e^{-\theta} \theta^x}{x!}$

Likelihood function is

$$L = \prod_{i=1}^n \frac{e^{-\theta} \theta^{x_i}}{x_i!}$$

$$L = \frac{e^{-n\theta} \theta^{\sum x_i}}{\prod_{i=1}^n x_i!} \rightarrow \textcircled{1}$$

WKT if $L = g(t(x), h(x))$

then $t(x)$ is sufficient statistic

eqⁿ $\textcircled{1}$ becomes

$$L = e^{-n\theta} \theta^{\sum x_i} \cdot \frac{1}{\prod_{i=1}^n x_i!}$$

$$L = g(\sum x_i, h(x)) \quad \left[\begin{array}{l} \therefore g(\sum x_i) = e^{-n\theta} \theta^{\sum x_i} \\ h(x) = \frac{1}{\prod_{i=1}^n x_i!} \end{array} \right]$$

$\therefore \sum x_i$ is sufficient statistic

Hence by Neyman Factorization Theorem

$t = \sum_{i=1}^n x_i$ is sufficient statistic for θ

3> Find the method of moment estimator for ' θ ' is

$$f(x, \theta) = \begin{cases} (1+\theta)x^\theta & ; 0 < x < 1, \theta > 0 \\ 0 & ; \text{otherwise} \end{cases}$$

Given $f(x, \theta) = (1+\theta)x^\theta$; $0 < x < 1$; $\theta > 0$

$$E(x) = \int_0^1 x \cdot f(x) dx$$

$$= \int_0^1 x \cdot (1+\theta)x^\theta dx$$

$$\Rightarrow (1+\theta) \int_0^1 x^{1+\theta} dx$$

$$= (1+\theta) \left[\frac{x^{2+\theta}}{2+\theta} \right]_0^1$$

$$= \frac{1+\theta}{2+\theta} [1-0]$$

$$E(x) = \frac{\theta+1}{\theta+2}$$

Value of parameter (θ) can be calculated as

$$\bar{x} = \frac{\theta+1}{\theta+2}$$

$$(\theta+2)\bar{x} = \theta+1$$

$$\theta\bar{x} + 2\bar{x} = \theta+1$$

$$(\theta\bar{x} - \theta) = (-2\bar{x} + 1)$$

$$\theta(\bar{x} - 1) = (-2\bar{x} + 1)$$

$$\theta = \frac{-2\bar{x} + 1}{\bar{x} - 1}$$

Moment of Estimator of $\theta = \frac{-2\bar{x} + 1}{\bar{x} - 1}$

- 4) Find the value of α , β and power of testing using C.R. = $\{x \geq 0.5\}$ for testing $H_0: \theta = 1$ against $H_1: \theta = 2$ based on single observation from

$$f(x, \theta) = \begin{cases} 1/\theta & ; 0 < x < \theta \\ 0 & ; \text{o.w.} \end{cases}$$

$$\text{Given } f(x, \theta) = \begin{cases} 1/\theta & ; 0 < x < \theta \\ 0 & ; \text{o.w.} \end{cases}$$

$H_0: \theta = 1$ against $H_1: \theta = 2$

$$\text{C.R.} = \{x \geq 0.5\}$$

$$W = 0.5 \leq x \leq 1$$

$$\bar{W} = 0 \leq x \leq 0.5$$

~~under~~ The likelihood f_0 under H_0 is

$$L_0 = f(x) = \frac{1}{\theta}$$

$$L_0 = 1/1 = 1$$

The likelihood f_1 under H_1 is

$$L_1 = f(x, \theta) = \frac{1}{\theta}$$

$$L_1 = 1/2$$

$$(\because H_1: \theta = 2)$$

Size of Type-I Error

$$\alpha = P[\text{Type-I Error}]$$

$$= P[\text{reject } H_0 / H_0 \text{ true}]$$

$$= P(X \in W / H_0)$$

$$= \int_W L_0 dx$$

$$= \int_{0.5}^1 1 dx$$

$$= [x]_{0.5}^1$$

$$= 1 - 0.5$$

$$\boxed{\alpha = 0.5}$$

Size of Type-II Error

$$\beta = P[\text{Type-II Error}]$$

$$> P[\text{Accept } H_0 / H_2 \text{ false}]$$

$$> P(X \in \bar{W} / H_2)$$

$$= \int_W L_2 dx$$

$$= \int_0^{0.5} \frac{1}{2} dx$$

$$= \left[\frac{1}{2} x \right]_0^{0.5}$$

$$\beta = 0.25$$

$$\text{Power of test} = 1 - \beta = 1 - 0.25 = 0.75$$

SECTION-C

Answer the following

- 1) Define MVUE. If $x_1, x_2, \dots, x_n$ is a random sample from poisson distribution with parameter ' μ '. Show that sample mean and sample variance and unbiased estimator of μ .

MVUE : Minimum Variance Unbiased Estimation.

Given

A random sample $x_1, x_2, \dots, x_n$ is drawn from poisson distribution with parameter ' μ '

i.e., $X \sim P(X, \mu)$

$$P(X, \mu) = \frac{e^{-\mu} \mu^x}{x!}$$

$$E(X) = \text{Mean} = \mu, \quad V(X) = \text{Variance} = \mu$$

Now,

$$\begin{aligned} E(X) &= E\left(\frac{1}{n} \sum_{i=1}^n x_i\right) \\ &= \frac{1}{n} \left[\sum_{i=1}^n x_i \right] \\ &= \frac{1}{n} \left(\sum_{i=1}^n \mu \right) \\ &= \frac{1}{n} \times n \mu \end{aligned}$$

$$E(X) = \mu$$

Mean is unbiased estimator of μ .

$$\begin{aligned} E[V(X)] &= E\left[\frac{1}{n} \left(\sum_{i=1}^n (x_i - \bar{x}) \right)^2\right] \\ &= \frac{1}{n} \left[E\left(\sum_{i=1}^n x_i^2 + \sum_{i=1}^n \bar{x}^2 - 2\bar{x} \sum_{i=1}^n x_i \right) \right] \end{aligned}$$

$$E[V(X)] = \frac{1}{n} \left[\sum E(x_i^2) + \sum E(\bar{x}^2) - 2\bar{x} \sum E(x_i) \right] \rightarrow (1)$$

WKT

$$V(X) = E(x_i^2) - E(\bar{x})^2$$

$$\Rightarrow E(x_i^2) = V(X) + E(\bar{x})^2$$

$$\Rightarrow E(x_i^2) = \mu + \mu^2$$

The above eq: (1) becomes

$$\begin{aligned} E[V(X)] &= \frac{1}{n} \left[\sum (\mu + \mu^2) + \sum \mu^2 - 2\mu \sum \mu \right] \quad (\because \bar{x} = \mu) \\ &= \frac{1}{n} (n\mu + n\mu^2 + n\mu^2 - 2n\mu^2) \end{aligned}$$

$$E[V(X)] = \mu$$

Variance is unbiased estimator of μ .

2) Obtain MLE of θ from the following pdf

i) $f(x, \theta) = \theta^x (1-\theta)^{1-x}, \quad x=0, 1$

ii) $f(x, \theta) = \frac{e^{-\theta} \theta^x}{x!}, \quad x=0, 1, 2, \dots, \infty; \theta > 0$

$$i) \quad f(x, \theta) = \theta^x (1-\theta)^{1-x}, \quad x=0,1$$

$$L = \prod_{i=1}^n \theta^{x_i} (1-\theta)^{1-x_i}$$

$$= \theta^{x_1} [1-\theta]^{1-x_1} \cdot \theta^{x_2} [1-\theta]^{1-x_2} \dots \theta^{x_n} [1-\theta]^{1-x_n}$$

$$L = \theta^{\sum x_i} [1-\theta]^{n-\sum x_i}$$

Apply log on both sides

$$\log L = \sum x_i \log \theta + (n - \sum x_i) \log (1-\theta)$$

Partially differentiating w.r.t θ and equating equal to zero. $\left(\because \frac{\partial \log L}{\partial \theta} = 0 \right)$

$$0 = \frac{\sum x_i}{\theta} + \frac{(n - \sum x_i)(-1)}{(1-\theta)} = 0$$

$$\Rightarrow \frac{\sum x_i}{\theta} + \frac{(n - \sum x_i)(-1)}{(1-\theta)} = 0$$

$$\Rightarrow \frac{\sum x_i}{\theta} + \frac{n}{1-\theta} - \frac{\sum x_i(-1)}{1-\theta} = 0$$

$$\Rightarrow \frac{\sum x_i(1-\theta) + n\theta + \theta \sum x_i}{\theta(1-\theta)} = 0$$

$$\Rightarrow \sum x_i - \theta \sum x_i + n\theta + \theta \sum x_i = 0$$

$$\Rightarrow \sum x_i = n\theta = 0$$

$$\Rightarrow \theta = \frac{\sum x_i}{n}$$

$$\therefore \hat{\theta} = \bar{x}$$

$$ii) \quad f(x, \theta) = \frac{e^{-\theta} \theta^x}{x!}, \quad x=0,1,2,\dots,n$$

$$L = \prod_{i=1}^n \frac{e^{-\theta} \theta^{x_i}}{x_i!}$$

$$L = \frac{e^{-n\theta} \theta^{\sum x_i}}{\sum x_i!}$$

Applying log on both sides

$$\log L = -n\theta + \sum x_i \log \theta - \log \sum (x_i)$$

Partially differentiating w.r.t θ and equating equal to zero

$$\therefore \frac{\partial \log L}{\partial \theta} = 0$$

$$0 = -n + \frac{\sum x_i}{\theta}$$

$$\theta = \frac{\sum x_i}{n}$$

$$\therefore \hat{\theta} = \bar{x}$$

Done
2/02/2024



B. L. D. E. Association's

S. B. ARTS & K. C. P. SCIENCE COLLEGE, VIJAYAPUR

IInd INTERNAL TEST 2023 -2024

26

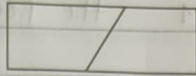
30

Name / ಹೆಸರು: Akshata A khavkar

Roll No./ಸ್ಥಳಾಂಕ: 94 Date / ದಿನಾಂಕ: 7-8-2024

Subject/ವಿಷಯ: Probability distribution Semester/ಸೆಮಿಸ್ಟರ್: IInd Exam Seat No / ಪರೀಕ್ಷೆ ಸ್ಥಳಾಂಕ: 035KM2350147

Evaluator's Signature / ಕೊರಡಿ ಮೇಲ್ವಿಚಾರಕರ ಸಹಿ Signature of Evaluator / ಪರೀಕ್ಷಕರ ಸಹಿ Marks obtained / ಪಡೆದ ಅಂಕಗಳು



13/15

Signature
08/08/2024

1) The recurrence relation of Binomial and Poisson distribution are

$$B.D :- M_{r+1} = pq (nr M_r + \frac{d}{dr} M_r)$$

$$P.D \rightarrow M_{r+1} = \lambda M_r + \frac{d}{dr} M_r$$

4) the four arithmetic operations are

addition (+)

subtraction (-)

multiplication (x)

division (/)

2) When 3 coins are tossed simultaneously then the outcomes are

{HHH, HTH, THT, TTT, HTT, TTH, HHT, THH}

6) numeric (10, 5, 55)

integral (1L, 55L)

complex $9 + 3i$

character it is exiting task

3) R Programming was introduced by statistician at bell laboratory

5 Marks

1 the properties of random variable

expectations

$$E(x) = a$$

$$E(x) = \sum x P(x)$$

$$E(x) = \sum a P(x)$$

$$= a \sum P(x)$$

$$= a (1) \quad (\because \sum P(x) = 1)$$

$$E(x) = a$$

$$E(ax) = a E(x)$$

$$E(ax) = \sum ax \cdot P(x)$$

$$= ax \cdot P(x)$$

$$\sum x \cdot P(x) = E(x)$$

$$= a E(x)$$

$$E(ax+b) = a E(x) + b$$

$$= \sum ax \cdot P(x) + \sum b P(x)$$

$$= a E(x) + b \sum P(x)$$

$$= a E(x) + b (1)$$

$$E(ax+b) = a E(x) + b$$

Variance :-

$$1) V(x) \geq 0$$

$$V(x) = E(x^2) - (E(x))^2$$

$$V(x) = E(x - E(x))^2$$

$$V(a) = E(a) [E(a)]^2$$

$$= E(a - a)^2$$

$$V(a) = 0$$

$$2) V(ax) = a^2 V(x)$$

$$V(ax) = E(x^2) - (E(x))^2$$

$$V(x) = E(x - E(x))^2$$

$$= a^2 E(x - E(x))^2$$

$$V(x) = a^2 V(x)$$

$$3) \quad V(ax+b) = a^2 V(x)$$

$$V(x) = E(x - E(x))^2$$

$$V(ax+b) = E((a+b) - E(ax+b))^2$$

$$= E(ax + b - E(ax) - b)^2$$

$$= E(ax - E(ax))^2$$

$$= a^2 E(x - E(x))^2$$

$$= a^2 V(x)$$

$$2) \quad (i) \quad E(x/n) = p$$

$$= \frac{1}{n} E(x)$$

$$= \frac{1}{n} np$$

$$= p$$

$$ii) \quad E\left[\frac{x}{n} - p\right]^2 = \frac{pq}{n}$$

$$= E\left[\frac{x}{n} - E\left(\frac{x}{n}\right)\right]^2$$

$$= \text{var}\left(\frac{x}{n}\right) = \frac{1}{n^2} \text{var}(x)$$

$$= \frac{1}{n^2} npq$$

$$= \frac{pq}{n} \quad \text{--- (i)}$$

from eq (i)

$$= E\left(\frac{x}{n} - p\right)^2 = E\left(\frac{x}{n} - E\left(\frac{x}{n}\right)\right)^2$$

$$= E\left(\frac{x}{n} - p\right)^2 = \frac{pq}{n}$$

$$iii) \quad E\left(\frac{x - np}{npq}\right)^2 = 1$$

$$E(x) = np$$

$$V(x) = npq$$

$$\Rightarrow \frac{E(x^2) - (E(x))^2}{V(x)}$$

$$= \frac{V(x)}{V(x)}$$

$$= 1 //$$

$$1) \quad P(x, 1) = P(x-x) = \begin{cases} \frac{e^{-1} 1^x}{x!} & \text{if } 0 \leq x < \infty \\ \text{other wise,} & \end{cases}$$

$$\mu_1^1 = \text{mean} = E(x) = \sum_{x=0}^{\infty} x P(x)$$

$$\begin{aligned} & \sum_{x=0}^{\infty} \frac{x \cdot e^{-1} 1^x}{x!} \\ &= 1e^{-1} \left\{ \sum_{x=0}^{\infty} \frac{x \cdot 1 \cdot 1^{x-1}}{x(x-1)!} \right\} \\ &= 1e^{-1} \left\{ \sum_{x=0}^{\infty} \frac{1^{x-1}}{(x-1)!} \right\} \\ &= 1e^{-1} (1 + 1 + 1^2 + 1^3 + \dots) \\ &= 1e^{-1} e^1 \\ &\Rightarrow 1, \end{aligned}$$

Hence the mean of the Poisson distribution is 1

$$\begin{aligned} \mu_2^1 \quad E(x^2) &= \sum_{x=0}^{\infty} x^2 P(x) \\ &= \sum_{x=0}^{\infty} \{x(x-1) + x\} \frac{e^{-1} 1^x}{x!} \\ &= e^{-1} \sum_{x=0}^{\infty} \frac{1^x}{x(x-1)(x-2)!} \\ &= e^{-1} \sum_{x=0}^{\infty} \frac{1^{x-2} 1^2}{(x-2)!} + 1 \\ &= e^{-1} 1^2 = \left[\sum_{x=2}^{\infty} \frac{1^{x-2} 1^2}{(x-2)!} \right] + 1 \\ &= e^{-1} 1, \quad e^{-1} 1 + 1 \\ &\Rightarrow 1^2 + 1, \end{aligned}$$

$\therefore$ variance of the Poisson distribution is $1^2 + 1 = 2$



B. L. D. E. Association's

S. B. ARTS & K. C. P. SCIENCE COLLEGE, VIJAYAPUR

INTERNAL TEST 20 -20

30
30

ಹೆಸರು: Pranav M. Hegde Roll No./ಸ್ಥಳಾಂಕ: 104 Date/ದಿನಾಂಕ: 07-08-2024

ತ/ವಿಷಯ: Probability & Distribution Semester/ಸೆಮಿಸ್ಟರ್: _____ Exam Seat No / ಪರೀಕ್ಷೆ ಸ್ಥಳಾಂಕ: V15K17350155

Signature / ಕೊರವಿ ಮೇಲ್ವಿಚಾರಕರ ಸಹಿ Signature of Evaluator / ಪರೀಕ್ಷಕರ ಸಹಿ Marks obtained / ಪಡೆದ ಅಂಕಗಳು

15
15

Pranav
08/08/2024

Recurrence relation of B.D =

$$\mu_{n+1} = p\mu_n + q\mu_{n-1} + \frac{d\mu_n}{dx}$$

Recurrence relation of Poisson distribution =

$$\mu_{n+1} = \lambda \mu_n + \frac{d\mu_n}{d\lambda}$$

If 3 coins are tossed

Total no. of outcomes are $= 2^3 = 8$

Sample space = {HHH, TTT, HHT, THH, THT, HTH, TTH, HTT}

Robert gentleman and Ross Thak

4 basic Arithmetic operations are

Addition

Subtraction

Multiplication

Division

Numeric (10, 5, 5, 5)

Integer (11, 55)

Complex (9+3i)

Character is sitting false

II

1) a) $E(a) = a$

$$E(1) = \sum x_i \cdot P(x_i)$$

$$E(a) = \sum a_i \cdot P(a_i)$$

$$E(a) = a \cdot \sum P(a_i)$$

$$E(a) = a \cdot 1$$

$$E(a) = a$$

$$[\because P(x) = 1]$$

b) $E(ax) = a E(x)$

$$E(x) = \sum x_i \cdot P(x_i)$$

$$E(ax) = \sum (ax)_i \cdot P(x_i)$$

$$E(ax) = \sum a x_i \cdot P(x_i)$$

$$E(ax) = a \sum x_i \cdot P(x_i)$$

$$E(ax) = a E(x)$$

c) $E(ax+b) = a E(x) + b$

$$E(x) = \sum x_i \cdot P(x_i)$$

$$E(ax+b) = \sum (ax+b)_i \cdot P(x_i)$$

$$= \sum a x_i \cdot P(x_i) + \sum b \cdot P(x_i)$$

$$= a \sum x_i \cdot P(x_i) + b \sum P(x_i)$$

$$E(ax+b) = a E(x) + b$$

d) $V(a) = 0$

$$V(x) = E(x^2) - (E(x))^2$$

$$V(a) = E(a^2) - (E(a))^2$$

$$V(a) = E(a-a)^2$$

$$V(a) = E(0)$$

$$V(a) = 0$$

e) $V(ax) = a^2 V(x)$

$$V(x) = E(x^2) - (E(x))^2$$

$$V(ax) = E((ax)^2) - (E(ax))^2$$

$$= a^2 E(x^2) - (a E(x))^2$$

$$V(ax) = a^2 V(x)$$

$$i) E\left(\frac{x}{n}\right) = p$$

$$E(x) = np$$

$$V(x) = npq$$

$$E(x) = \frac{1}{n} = p$$

$$np = \frac{1}{n} = p$$

$$\frac{np}{n} = p$$

$$E\left(\frac{x}{n} - p\right)^2 = \frac{pq}{n}$$

$$\therefore V(x) = \frac{1}{n^2}$$

$$E\left(\left(\frac{x}{n} - E\left(\frac{x}{n}\right)\right)^2\right)$$

$$\therefore V(x) = \frac{npq}{n^2} = \frac{pq}{n}$$

$$= V\left(\frac{x}{n}\right)$$

$$= \frac{npq}{n^2}$$

$$= \frac{pq}{n}$$

$$ii) E\left(\frac{x - npq}{npq}\right)^2 = 1$$

$$\therefore npq = E(x)$$

$$= E\left(\frac{x - E(x)}{V(x)}\right)^2$$

$$\therefore npq = V(x)$$

$$= \frac{E(x^2) - (E(x))^2}{V(x)}$$

$$= \frac{V(x)}{V(x)} = 1$$

III

1) A Random variable 'x' follows poisson distribution. if it assumes only negative value the probability mass function is given by

$$P(x, \lambda) = P(x=x) = \begin{cases} \frac{e^{-\lambda} \lambda^x}{x!}, & x=0, 1, 2, 3, \dots \end{cases}$$

Here ' λ ' is parameter of poisson distribution we shall have notation $(X \sim P) \lambda$, x is random variable and parameter λ

$$\text{Mean} = E(x) = \sum_{x=0}^{\infty} x \cdot P(x)$$

$$= \sum_{x=0}^{\infty} x \cdot \frac{e^{-\lambda} \lambda^x}{x!}$$

$$= \lambda e^{-\lambda} \sum_{x=0}^{\infty} \left[\frac{x \lambda^{x-1}}{x(x-1)!} \right]$$

10
$$= \lambda e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^{x-1}}{(x-1)!}$$

$$= \lambda e^{-\lambda} \sum_{x=0}^{\infty} \left[1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \dots \right]$$

$$= \lambda e^{-\lambda} \sum_{x=0}^{\infty} e^{\lambda}$$

$$= \lambda e^{-\lambda} e^{\lambda}$$

$\therefore E(x) = \lambda$ -- mean of poisson distribution.

$$\text{Variance} = E(x^2) = \sum_{x=0}^{\infty} x^2 \cdot P(x)$$

$$= \sum_{x=0}^{\infty} \{x(x-1) + x\} \cdot \frac{e^{-\lambda} \lambda^x}{x!}$$

$$= \sum_{x=0}^{\infty} \{x(x-1) \sum_{x=0}^{\infty} x \cdot \frac{e^{-\lambda} \lambda^x}{x!}\}$$

$$= e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda x^{x-1} (x(x-1) + x)}{x!}$$

$$= e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda x^{x-1} \lambda^x}{x(x-1)(x-2)!} + 1$$

$$= e^{-\lambda} \lambda^2 \sum_{x=0}^{\infty} \frac{\lambda^{x-1}}{(x-2)!} + 1$$

$$= e^{-\lambda} \lambda^2 e^{\lambda} + 1$$

$E(x^2) = \lambda^2 + \lambda$ - Variance of Poisson distribution

$$V(x) = E(x^2) - (E(x))^2$$

$$= \lambda^2 + \lambda - \lambda^2$$

$$= \lambda$$

B.L.D.E. Association's

**S. B. Arts & K.C.P. Science College,
VIJAYAPUR- 586 103.**



ASSIGNMENT

For B.A./ B.Sc.II..... Semester
20 23 - 20 24

Name of the Student SUNITA - B. BIRADAR

Roll No. 149 163 R.C.U. Seat No. UISKM2380&16

Subject Probability and distribution

Assignment No.	Date	Marks Assigned	Marks Obtained	Name and Signature of Teacher	Remarks
1	02/08/24	05	05	S.B. Naganoor <u>S.B. Naganoor</u>	Good
2					
3					
4					

1) Define conditional probability

The probability of occurrence of one event under the condition that another event has already occurred is called conditional probability

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

2) With usual notations prove that $P(\bar{A}) = 1 - P(A)$.

proof = Since $\bar{A}$ event of A taking probability both sides

$$P(A \cup \bar{A}) = P(S)$$

$$P(A) + P(\bar{A}) = P(S)$$

$$P(A) + P(\bar{A}) = 1$$

$$P(\bar{A}) = 1 - P(A)$$

3) Define random variable.

Random variable is a function which assigns a real number to every sample point in the sample space.

4) $V(x) = 9$, $E(x^2) = 25$. Find $E(x)$

$$E(x) = \sum x \cdot p(x)$$

$$V(x) = E(x - E(x))^2$$

$$= E(x^2) - (E(x))^2$$

$$9 = 25 - (E(x))^2$$

$$(E(x))^2 = 25 - 9$$

$$(E(x))^2 = 16$$

$$E(x) = \sqrt{16}$$

$$E(x) = 4$$

1) State and prove multiplication theorem of probability.

For two events A & B

$$P(A \cap B) = P(A) \cdot P(B/A), P(A) > 0$$

$$P(A \cap B) = P(B) \cdot P(A/B), P(B) > 0$$

where $P(B/A)$ represents conditional probability of occurrence B when the event A has already happened and where $P(A/B)$ is the conditional probability occurrence of A given that B is already happened

In the equal notation we have

$$P(B/A), P(A/B)$$

$$P(A) = \frac{n(A)}{n(S)}, P(B) = \frac{n(B)}{n(S)}, P(A \cap B) = \frac{n(A \cap B)}{n(S)} \quad \text{--- (1)}$$

for the conditional event B/A the favorable outcomes must be one of the sample points of B i.e. for the event A/B the sample space is B & out of the $n(B)$ sample points, $n(A \cap B)$ portion is to the occurrence of the event A hence

$$P(A/B) = \frac{n(A \cap B)}{n(B)}$$

Now eqn (1) the 3rd term becomes

$$P(A \cap B) = \frac{P(A/B) \cdot n(B)}{n(S)}$$

rewriting eqn (1)

$$P(A \cap B) = \frac{n(B)}{n(S)} \cdot \frac{n(A \cap B)}{n(B)}$$

$$P(A \cap B) = \frac{n(B)}{n(S)} \cdot P(A/B)$$

$$P(A \cap B) = P(B) \cdot P(A/B) \quad \text{--- (2)}$$

Similarly we get from eqn (1) $P(A \cap B) = \frac{n(A)}{n(S)} \cdot \frac{n(A \cap B)}{n(B)}$

$$= P(A) \cdot P(B/A) \quad \text{--- (3)}$$

from eqn (2) & (3) we get the result.

write axiomatic definitions of probability of an event with usual notations and prove that

$$i) P(\emptyset) = 0.$$

proof = By definition of probability

$$P(\emptyset) = \frac{n(\emptyset)}{n(S)} = \frac{0}{1} = 0$$

$$\boxed{\therefore P(\emptyset) = 0}$$

90) $0 \leq P(A) \leq 1$

proof - By definition of probability of an event A is $P(A) = \frac{m}{n}$

the least and highest possible values of m are zero and n

i.e. : $0 \leq m \leq n$ (divide by n)

$$\frac{0}{n} \leq \frac{m}{n} \leq \frac{n}{n}$$

$$0 \leq P(A) \leq 1$$

3) Define the terms

91) probability mass function (pmf) =

let ' x ' is a discrete random variable and $p(x)$ be a function such that $p(x) = P(X=x)$ then $p(x)$ is called probability mass function

$$\sum P(x) = 1$$

92) probability density function (pdf)

let ' x ' be a continuous random variable a function $f(x)$ is said to be probability density function (pdf) of the continuous random variable ' x ' if the following conditions are satisfied

$$\int_{-\infty}^{\infty} f(x) \cdot dx = 1$$

1) State and prove Bayes theorem of probability

Let event $E_1, E_2, \dots, E_n$ be a set of events associated with a sample space S , where all the events $E_1, E_2, \dots, E_n$ have non zero probability occurrence and they form partition of S .

Let A, B event associated with S according to the Bayes theorem

$$P\left(\frac{E_i}{A}\right) = \frac{P(E_i) \cdot P(A/E_i)}{\sum_{k=1}^n P(E_k) \cdot P(A/E_k)}$$

According to the conditional probability

$$P\left(\frac{E_i}{A}\right) = \frac{P(E_i \cap A)}{P(A)} \quad \text{--- (1)}$$

Multiplication theorem of probability

$$P(E_i \cap A) = P(E_i) \cdot P\left(\frac{A}{E_i}\right) \quad \text{--- (2)}$$

By using total probability

$$P(A) = \sum_{k=1}^n P(E_k) \cdot P(A/E_k) \quad \text{--- (3)}$$

Eq (3) & (2) & (1) we get

$$P\left(\frac{E_i}{A}\right) = \frac{P(E_i) \cdot P(A/E_i)}{\sum_{k=1}^n P(E_k) \cdot P(A/E_k)}$$

2) Find the value of n for the following p.d.f and hence find mean

$$f(x) = \begin{cases} \frac{k(x+4)}{8} & 0 < x < 4 \\ 0 & \text{o.w} \end{cases}$$

$$\int_{-\infty}^{\infty} f(x) \cdot dx = 1$$

$$\int_0^4 \frac{k}{8}(x+4) dx = 1$$

$$\frac{k}{8} \left[\int_0^4 x \cdot dx + \int_0^4 4 dx \right] =$$

$$\frac{k}{8} \int_0^4 (x+4) dx = 1$$

$$\frac{k}{8} \left[\int_0^4 x dx + \int_0^4 4 dx \right] = 1$$

$$\frac{k}{8} \left[\frac{x^2}{2} + [4x]_0^4 \right] = 1$$

$$\frac{k}{8} \left[\frac{(4)^2}{2} - 0 + [4(4) - 0] \right] = 1$$

$$\frac{k}{8} \left[\frac{16 - 0}{2} + [16 - 0] \right] = 1$$

$$\frac{k}{8} \left[\frac{16}{2} + 16 \right] = 1$$

$$\frac{k}{8} [8 + 16] = 1$$

$$\frac{k}{8} (24) = 1$$

$$\frac{24k}{8} = 1 \quad \text{or} \quad 24k = 8$$

$$3k = 1$$

$$k = \frac{1}{3}$$

$$k = \frac{8}{24}$$

$$k = \frac{1}{3}$$

8) obtain mean and variance of Binomial distribution
Moments of Binomial distribution

* 1st moment about origin μ'_1 $I(\mu'_1)$

$$\mu'_1 = E(x) = \sum_{x=0}^n x \cdot p(x)$$

$$\sum_{x=0}^n x \binom{n}{x} p^x q^{n-x}$$

$$\sum_{x=0}^n x \cdot \frac{n!}{x!} \binom{n-1}{x-1} p^x q^{n-x}$$

$$\sum_{x=0}^n n \binom{n-1}{x-1} p^{x-1} q^{n-x}$$

$$np \sum_{x=1}^n \binom{n-1}{x-1} p^{x-1} q^{n-x}$$

$$= np \sum_{x=1}^n (p+q)^{n-1}$$

$$\mu'_1 = E(x) = np$$

* 2nd moment about origin μ'_2

$$\mu'_2 = E(x^2) = \sum_{x=0}^n x^2 p(x)$$

$$\sum_{x=0}^n [x(x-1) + x] p(x)$$

$$\sum_{x=0}^n x(x-1) p(x) + \sum_{x=0}^n x p(x)$$

$$\sum_{x=0}^n x(x-1) \binom{n}{x} p^x q^{n-x} + np$$

$$\sum_{x=0}^n x(x-1) \left[\frac{n}{x} \frac{n-1}{x-1} \binom{n-2}{x-2} \right] p^{x-2} \cdot p^2 q^{n-1} + np$$

$$= n(n-1)p^2 \sum_{x=2}^n \binom{n-2}{x-2} p^{x-2} q^{n-2} + np$$

$$\mu'_2 = n(n-1)p^2 + np$$

* III = μ'_3

$$\mu'_3 = \sum_{x=0}^n x^3 p(x)$$

$$\sum_{x=0}^n \{x(x-1)(x-2) + 3x(x-1) + x\} p(x)$$

$$\sum_{x=0}^n x(x-1)(x-2) p(x) + 3 \sum_{x=0}^n [x(x-1) + x] p(x)$$

$$\sum_{x=0}^n x(x-1)(x-2) \binom{n}{x} p^x q^{n-x} + 3n(n-1)p^2 + np$$

$$\sum_{x=0}^n x(x-1)(x-2) \left[\frac{n(n-1)(n-2)}{x(x-1)(x-2)} \binom{n-3}{x-3} p^{x-3} p^3 q^{n-x} + 3 \right] +$$

$$3n(n-1)p^2 + np$$

$$= n(n-1)(n-2)p^3 \sum_{x=3}^n \binom{n-3}{x-3} p^{x-3} q^{n-3} + 3[n(n-1)p^2 + np]$$

$$= \left\{ n(n-1)(n-2)p^3 \sum_{x=3}^n p(x) + 3n(n-1)p^2 + np \right\}$$

$$\boxed{\mu'_3 = n(n-1)(n-2)p^3 + 3n(n-1)p^2 + np}$$

* Similarly

* $\mu_4' = \mu_4'$

$$\mu_4' = E(x^4)$$

$$= E(x^4 p(x))$$

$$x^4 = x(x-1)(x-2)(x-3) + 6x(x-1)(x-2) + 7x(x-1) + x$$

$$\mu_4' = n(n-1)(n-2)(n-3)p^4 + 6n(n-1)(n-2)p^3 + 7n(n-1)p^2 + np$$

Moment generating function of Binomial distribution

Let $x \sim B(n, p)$ then

$$M_x(t) = E(e^{tx}) = \sum_{x=0}^n e^{tx} p(x) = \sum_{x=0}^n e^{tx} \binom{n}{x} p^x q^{n-x}$$

$$= \sum_{x=0}^n \binom{n}{x} (pe^t)^x q^{n-x}$$

$$M_x(t) = (q + pe^t)^n$$

* Mean

$$M_x(t) = (q + pe^t)^n$$

$$\text{Mean} = \frac{d}{dt} M_x(t) = \frac{d}{dt} (q + pe^t)^n \Big|_{t=0}$$

$$= n(q + pe^t)^{n-1} \cdot pe^t \Big|_{t=0}$$

$$= n(q + pe^0)^{n-1} pe^0$$

$$= np(q + p)^{n-1}$$

$$\boxed{\frac{d}{dt} M_x(t) = np} \quad [e^0 = 1]$$

* Variance

$$\text{var} = E(x^2) - (E(x))^2$$

$$= \frac{d^2}{dt^2} [M_x(t)] - \left[\frac{d}{dt} M_x(t) \right]^2 \Big|_{t=0}$$

$$= \frac{d}{dt} \left[\frac{d}{dt} M_x(t) \right] - n^2 p^2$$

$$= \frac{d}{dt} [n \cdot p e^t (q + p e^t)^{n-1}]_{t=0} - n^2 \cdot p^2$$

$$= n p e^t (q + p e^t)^{n-1} + e^t (n-1) (q + p e^t)^{n-2} (p e^t) - n^2 p^2$$

$$= (1 + (n-1)p) - n^2 p^2$$

$$= n p + n p^2 (n-1) - n^2 p^2$$

$$= n p + n^2 p^2 - n p^2 - n^2 p^2$$

$$= n p (1-p)$$

$$\boxed{\text{Var}(X) = n p q}$$

PART - A

I

- 1) Define standard normal distribution.
A special normal distribution where the mean is 0 and the standard deviation is 1.
* It is also called as z-distribution.
- 2) Define 'R' programming.
→ R is a programming language for statistical computing and data visualisation.
* It has been adopted in the fields of data mining, bioinformatics, and data analysis.
- 3) write M.G.F of Binomial variate
 $M_X(t) = (q + pe^t)^n$
This is the M.G.F of Binomial variate.

PART B

4) If x is a random variable a and b are constants prove that

$$\begin{aligned} \text{i) } E(ax+b) &= aE(x) + b \\ \text{ii) } V(ax+b) &= a^2 V(x) \\ \text{iii) } E(ax+b) &= aE(x) + b \\ \text{Proof: } E(x) &= \sum x \cdot p(x) \\ E(ax+b) &= \sum (ax+b) \cdot p(x) \\ &= \sum x \cdot p(x) + \sum b p(x) \\ &= a \sum x \cdot p(x) + b \sum p(x) \\ &= aE(x) + b(1) \\ &= aE(x) + b, \end{aligned}$$

$$\begin{aligned} \text{iv) } V(ax+b) &= a^2 V(x) \\ V(x) &= E(x - E(x))^2 \\ V(ax+b) &= E((ax+b) - E(ax+b))^2 \\ &= E(ax+b - E(ax) - b)^2 \\ &= E(ax - E(ax))^2 \end{aligned}$$

$$= \sigma^2 E(x - E(x))^2$$

$$= \sigma^2 V(x).$$

Hence proved

$$\Rightarrow \text{For B.D } p_T p(x+1) = \left(\frac{n-x}{x+1} \right) \frac{p}{q} p(x).$$

We know that

$$X \sim B(n, p)$$

The pmf of Binomial distribution

$$p(x) = \binom{n}{x} p^x q^{n-x} \quad \text{--- (1)}$$

Similarly

$$p(x+1) = \binom{n}{x+1} p^{x+1} q^{n-(x+1)} \quad \text{--- (2)}$$

Dividing Eq (2) by (1)

$$\frac{p(x+1)}{p(x)} = \frac{\binom{n}{x+1} p^{x+1} q^{n-x-1}}{\binom{n}{x} p^x q^{n-x}}$$

$$= \left(\frac{n!}{(n-x-1)!(x+1)!} \right) p$$

$$\frac{n!}{(n-x)!x!} q$$

$$= \frac{(n-x)(n-x-1)!x!p}{(n-x-1)!(x+1)x!q}$$

$$\frac{p(x+1)}{p(x)} = \frac{(n-x)p}{(x+1)q}$$

$$p(x+1) = \left(\frac{n-x}{x+1} \right) \frac{p}{q} p(x).$$

3) Explain 'R' as a calculator.

R can be used as a powerful calculator by entering equations directly at the prompt in the command console. Simply type your arithmetic expression and press Enter.

- * R will evaluate the expressions and respond with the result. While this is a simple interaction interface there could be problems if you are not careful. R.
- * R will normally execute your arithmetic expression by evaluating each item from left to right but some operators have precedence in the order of evaluation.

Simple arithmetic Expressions:

- The operators R uses for basic arithmetic are
- + Addition
 - Subtraction
 - x Multiplication
 - / Division
 - ^ Exponentiation

Some arithmetic Expressions with example

$4 + 8$ will return the result 12

5×14 will return the result 70

$7 / 4$ will return the result 1.75

4^3 will return the result 64

$6 - 3$ will return the result 3

R operators classification

1. Arithmetic
2. Relational
3. Logical
4. Assignment
5. Special operators

4) Show that for normal distribution mean, median, and mode are coincide

i.e. Mean = Median = Mode.

$$\text{Mean} = E(x) = \int_{-\infty}^{\infty} x f(x) dx$$

$$= \int_{-\infty}^{\infty} x \cdot \frac{e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}} dx$$

$$= \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \sigma\sqrt{2\pi}$$

$$\text{i.e. } \frac{x-\mu}{\sigma} = z \Rightarrow x = \sigma z + \mu = \frac{dx}{dz} = \sigma$$

$$\int_{-\infty}^{\infty} (\sigma z + \mu) \frac{e^{-\frac{1}{2}z^2}}{\sigma\sqrt{2\pi}} \sigma dz$$

$$\int_{-\infty}^{\infty} \sigma z \frac{e^{-\frac{1}{2}z^2}}{\sigma\sqrt{2\pi}} dz + \int_{-\infty}^{\infty} \mu \frac{e^{-\frac{1}{2}z^2}}{\sigma\sqrt{2\pi}} dz$$

$$= \int_{-\infty}^{\infty} z \cdot \frac{e^{-\frac{1}{2}z^2}}{\sqrt{2\pi}} dz + \int_{-\infty}^{\infty} \mu \frac{e^{-\frac{1}{2}z^2}}{\sqrt{2\pi}} dz$$

$$= 0 + \frac{\mu}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2} dz$$

$$= \frac{\mu}{\sqrt{2\pi}} \cdot \sqrt{2\pi}$$

$$\therefore E(x) = \mu$$

$\because z \cdot \frac{e^{-\frac{1}{2}z^2}}{\sqrt{2\pi}} dz$ is odd function

$\therefore$ It is equal to zero

$\Rightarrow$ Median:

$$\int_{-\infty}^M f(x) dx = \frac{1}{2}$$

$$\int_{-\infty}^M \frac{e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}} dx = \frac{1}{2}$$

$$\text{But } \int_{-\infty}^{\infty} f(x) dx = 1$$

$$\Rightarrow \int_{-\infty}^{\infty} f(x) dx = \int_{-\infty}^{\infty} f(x) dx$$

$$\int_{-\infty}^{\infty} f(x) dx = \int_{-\infty}^{\infty} \frac{e^{-\frac{1}{2} \left(\frac{x-\mu}{\sigma}\right)^2}}{\sigma \sqrt{2\pi}} dx = \frac{1}{2} \quad \text{--- (1)}$$

Subst, $\frac{x-\mu}{\sigma} = z \Rightarrow x = \mu + \sigma z, \frac{dx}{dz} = \sigma$

$$\Rightarrow dx = \sigma dz$$

$$\int_{-\infty}^{\infty} \frac{e^{-\frac{1}{2} z^2}}{\sigma \sqrt{2\pi}} \sigma dz = \frac{1}{2} \quad \text{--- (2)}$$

$$\int_{-\infty}^{\infty} \frac{e^{-\frac{1}{2} \left(\frac{x-\mu}{\sigma}\right)^2}}{\sigma \sqrt{2\pi}} dx = 1$$

$$\Rightarrow 2 \cdot \int_0^{\infty} \frac{e^{-\frac{1}{2} z^2}}{\sqrt{2\pi}} dz = 1$$

$$\int_0^{\infty} \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}} dz = \frac{1}{2} \quad \text{--- (3)}$$

from (2) & (3)

$$\frac{\mu - \mu}{\sigma} = 0$$

$$\boxed{\mu = \mu}$$

Mode :-

Mode = The value of 'x' for $f(x)$ is maximum
the value of x for $\frac{e^{-\frac{1}{2} \left(\frac{x-\mu}{\sigma}\right)^2}}{\sigma \sqrt{2\pi}}$ is maximum

Set the first order differentiation,

$$\frac{d}{dx} \left(\frac{e^{-\frac{1}{2} \left(\frac{x-\mu}{\sigma}\right)^2}}{\sigma \sqrt{2\pi}} \right) = 0$$

$$\frac{d}{dx} \left(\frac{e^{-\frac{1}{2} \left(\frac{x-\mu}{\sigma}\right)^2}}{\sigma \sqrt{2\pi}} \right) = 0$$

$$= \frac{-\sqrt{2\pi} e^{-\frac{1}{2} \left(\frac{x-\mu}{\sigma}\right)^2} \frac{1}{\sigma} \left(\frac{x-\mu}{\sigma}\right)}{(-\sqrt{2\pi})^2} = 0$$

$$= \frac{e^{-\frac{1}{2} \left(\frac{x-\mu}{\sigma}\right)^2} \cdot \left(\frac{x-\mu}{\sigma}\right)}{\sigma \sqrt{2\pi}} = 0$$

$$= e^{-\frac{1}{2} \left(\frac{x-\mu}{\sigma}\right)^2} \cdot \left(\frac{x-\mu}{\sigma}\right) = 0$$

$$= \frac{x-\mu}{\sigma} = 0$$

$$= x-\mu = 0$$

$$\boxed{i: x=\mu}$$

~~Errechnet~~
02/08/2024

B.L.D.E. Association's

**S. B. Arts & K.C.P. Science College,
VIJAYAPUR- 586 103.**



ASSIGNMENT

For B.A./ B.Sc.II..... Semester
2023 - 2024

Name of the Student BHAYYASHREE, BIRADAR

Roll No. 129 R.C.U. Seat No. U15KM2350184

Subject Probability and Distribution

Assignment No.	Date	Marks Assigned	Marks Obtained	Name and Signature of Teacher	Remarks
1	02/08/24	05	05	S.B. Naganoor <u>S.B. Naganoor</u>	Good
2					
3					
4					

② Define conditional probability.

The probability occurrence of one event under the condition that another event has already occurred is called conditional probability.

$$\text{eg :- } P(A/B) = \frac{P(A \cap B)}{P(B)}$$

$P(A/B)$ represents conditional probability of occurrence A when the event A is under the already happen.

② With usual notations prove that $P(\bar{A}) = 1 - P(A)$
proof = since A event of A taking probability both sides

$$P(A \cup A) = P(S)$$

$$P(A) + P(\bar{A}) = P(S)$$

$$P(A) + P(\bar{A}) = 1$$

$$P(\bar{A}) = 1 - P(A)$$

③ Define random variable.

Random variable is a function which assigns a real number to every sample point in the sample space.

eg :- 2 fair coins are tossed

$$S = \{HH, TH, HT, TT\} \quad n=4$$

④ $V(x) = 9$, $E(x^2) = 25$, Find $E(x)$.

$$V(x) = E(x)^2 - (E(x^2))$$

$$9 = E(x)^2 - 25$$

$$E(x)^2 = 25 - 9$$

$$E(x)^2 = 16$$

$$9 = 25 - (E(x))^2$$

$$(E(x))^2 = 25 - 9$$

$$(E(x))^2 = 16$$

$$E(x) = \sqrt{16}$$

$$E(x) = 4$$

II (2) state and prove multiplication theorem of probability.

If A and B are the two events with +ve probabilities $\{P(A) \neq 0, P(B) \neq 0\}$ Then A and B are independent if and only if $P(A \cap B) = P(A) \cdot P(B)$.

we have,

$$\left. \begin{aligned} P(A \cap B) &= P(A) \cdot P(B/A), P(A) > 0, P(A) \neq 0 \\ P(A \cap B) &= P(B) \cdot P(A/B), P(B) > 0, P(B) \neq 0 \end{aligned} \right\} \textcircled{1}$$

If A and B are independent, Then $P(A/B) = P(A)$ and $P(B/A) = P(B) \rightarrow \textcircled{2}$

From eqn (1) and (2) we get

$$\frac{P(A \cap B)}{P(B)} = P(A)$$

$$\frac{P(A \cap B)}{P(A)} = P(B)$$

$$P(A \cap B) = P(A) \cdot P(B)$$

$$P(A \cap B) = P(B) \cdot P(A)$$

Hence proved,

(20) write axiomatic definitions of probability of an event with usual notations and prove that

(i) $P(\emptyset) = 0$

By defⁿ of probability

$$P(\emptyset) = \frac{n(\emptyset)}{n(S)} = \frac{0}{1} = 0$$

$$\boxed{P(\emptyset) = 0}$$

(ii) $0 \leq P(A) \leq 1$

By defⁿ of probability of an event A is $P(A) = m/n$ [fxy(xy) = fxc . fxy]

The least and highest possible value of 'm' are zero and 'n'

i.e. ; $0 \leq m \leq n$ (divide by n)

$$\frac{0}{n} < \frac{m}{n} < \frac{n}{n}$$

$$0 \leq p(A) \leq 1$$

③ Define the terms.

① Distribution function:—

Let x is a discrete random variable and $p(x)$ be a function such that $P(x) = P(X=x)$. Then $P(x)$ is called probability mass function.

i) $P(x) \geq 0 \forall x$

ii) $\sum P(x) = 1$

② probability density function (pdf) :-

Let x be a continuous random variable a function $f(x)$ is said to be probability d.f of continuous random variable x if the following conditions are satisfied.

i) $f(x) \geq 0 \forall x$

ii) $\int_{-\infty}^{\infty} f(x) \cdot dx = 1$

III 1) state and prove Baye's theorem of probability.

Let event $E_1, E_2, \dots, E_n$ are be a set of events associated with a sample space S , where all the events $E_1, E_2, \dots, E_n$ have non-zero probability occurrence and they form partition of S .

Let A & B event associated with S according to the Baye's theorem

$$P(E_i/A) = \frac{P(E_i) \cdot P(A/E_i)}{\sum_{k=1}^n P(E_k) \cdot P(A/E_k)}$$

According to the conditional probability

$$P\left(\frac{E_i}{A}\right) = \frac{P(E_i \cap A)}{P(A)} \rightarrow (1)$$

Multiplication theorem of probability

$$P(E_i \cap A) = P(E_i) \cdot P\left(\frac{A}{E_i}\right) \rightarrow (2)$$

By using total probability

$$P(A) = \sum_{k=1}^n P(E_k) \cdot P\left(\frac{A}{E_k}\right) \rightarrow (3)$$

Put (3) & (2) in (1) we get

$$P\left(\frac{E_i}{A}\right) = \frac{P(E_i) \cdot P\left(\frac{A}{E_i}\right)}{\sum_{k=1}^n P(E_k) \cdot P\left(\frac{A}{E_k}\right)}$$

(2) Find the value of k for the following p.d.f and hence find mean.

$$f(x) = \begin{cases} \frac{k(x+4)}{8} & 0 < x < 1 \\ 0 & \text{o.w} \end{cases}$$

$$\int_{-\infty}^{\infty} f(x) \cdot dx = 1$$

$$\int_0^4 k(x+4) dx = 1$$

$$\frac{k}{8} \left[\int_0^4 x \cdot dx + \int_0^4 4 dx \right] = 1$$

$$\frac{k}{8} \int_0^4 (x+4) \cdot dx = 1$$

$$\frac{k}{8} \left[\int_0^4 x \cdot dx + \int_0^4 4 dx \right] = 1$$

$$\frac{k}{8} \left[\frac{x^2}{2} \right]_0^4 + [4x]_0^4 = 1$$

$$\frac{k}{8} \left[\frac{(4)^2}{2} - 0 \right] + [4(4) - 0] = 1$$

$$\frac{k}{8} \left[\frac{16}{2} - 0 \right] + [16 - 0] = 1$$

$$\frac{k}{8} \left[\frac{16}{2} + 16 \right] = 1$$

$$\frac{k}{8} [8+16] = 1$$

$$\frac{k}{8} [24] = 1$$

$$\frac{24k}{8} = 1 \quad \text{or} \quad 24k = 8$$

$$3k = 1$$

$$k = \frac{1}{3}$$

$$k = \frac{8}{24}$$

$$k = \frac{1}{3}$$

Moment of Binomial distribution

* 1st moment about origin $\Pi (u'_1)$

$$\begin{aligned} u'_1 = E(x) &= \sum x \cdot p(x) \\ &= \sum_{x=0}^n x \binom{n}{x} p^x q^{n-x} \\ &= \sum_{x=0}^n x \cdot \frac{n}{x} \binom{n-x}{x-1} p^x q^{n-x} \\ &= \sum_{x=0}^n n \binom{n-x}{x-1} p^{x-1} q^{n-x} \\ &= np \sum_{x=1}^n \binom{n-1}{x-1} p^{x-1} q^{n-x} \\ &= np \sum_{x=1}^n (p+q)^{n-1} \end{aligned}$$

$$u'_1 = E(x) = np$$

* 2nd moment of about origin $\Pi (u'_2)$

$$\begin{aligned} u'_2 = E(x^2) &= \sum_{x=0}^n x^2 p(x) \\ &= \sum_{x=0}^n [x(x-1) + x] p(x) \\ &= \sum_{x=0}^n x(x-1) p(x) + \sum_{x=0}^n x \cdot p(x) \\ &= \sum_{x=0}^n x(x-1) \binom{n}{x} p^x q^{n-x} + np \\ &= \sum_{x=0}^n \left[\frac{n}{x} \frac{(n-1)}{(x-1)} \binom{n-2}{x-2} p^{x-2} \cdot p^2 q^{n-1} + np \right] \end{aligned}$$

$$= n(n-1) p^2 \sum_{x=2}^n \binom{n-2}{x-2} p^{x-2} 2^{n-2} + np$$

$$= \mu'_2 = n(n-1) p^2 + np$$

*

$$\text{III} = \mu'_3$$

$$\mu'_3 = \sum_{x=0}^n x^3 p(x)$$

$$= \sum_{x=0}^n \{x(x-1)(x-2) + 3x(x-1) + x\} p(x)$$

$$= \sum_{x=0}^n x(x-1)(x-2) p(x) + 3 \sum_{x=0}^n [x(x-1) + x] p(x)$$

$$= \sum_{x=0}^n x(x-1)(x-2) \binom{n}{x} p^x 2^{n-x} + 3n(n-1) p^2 + np$$

$$= \sum_{x=0}^n x(x-1)(x-2) \left[\frac{n(n-1)(n-2)}{x(x-1)(x-2)} \binom{n-3}{x-3} p^{x-3} p^3 2^{n-3} + 3 \right]$$

$$+ 3n(n-1) p^2 + np$$

$$= n(n-1)(n-2) p^3 \sum_{x=3}^n \binom{n-3}{x-3} p^{x-3} 2^{n-3} + 3[n(n-1) p^2 + np]$$

$$= \{n(n-1)(n-2) p^3 \sum_{x=3}^n \binom{n-3}{x-3} p(x) + 3n(n-1) p^2 + np\}$$

$$= [n(n-1)(n-2) p^3 + 3n(n-1) p^2 + np]$$

$$= \mu'_3 = n(n-1)(n-2) p^3 + 3n(n-1) p^2 + np$$

Similarly,

*

$$\text{IV} = \mu'_4$$

$$\mu'_4 = \sum (x^4)$$

$$= \sum x^4 p(x)$$

$$x^4 = x(x-1)(x-2)(x-3) + 6x(x-1)(x-2) + 7x(x-1) + x$$

$$\mu'_4 = n(n-1)(n-2)(n-3) p^4 + 6n(n-1)(n-2) p^3 + 7n(n-1) p^2 + np$$

* Mean :-

$$M_x(t) = (2 + pe^t)^n$$

$$\text{Mean} = \frac{d}{dt} M_x(t) = \frac{d}{dt} (2 + pe^t)^n \Big|_{t=0}$$

$$= n (2 + pe^t)^{n-1} \cdot pe^t \Big|_{t=0}$$

$$= n (2 + pe^0)^{n-1} \cdot pe^0$$

$$= np (2 + p)^{n-1} \quad \because [e^0 = 1]$$

$$\frac{d}{dt} M_x(t) = np$$

* Variance :-

$$\text{Var} = E(x^2) - (E(x))^2$$

$$= d^2 \left[\frac{M_x(t)}{dt^2} - \left[\frac{d}{dt} M_x(t) \right]^2 \right] \Big|_{t=0}$$

$$= \frac{d}{dt} \left[\frac{d}{dt} M_x(t) \right] - n^2 p^2$$

$$= \frac{d}{dt} [n \cdot pe^t (2 + pe^t)^{n-1}]_{t=0} - n^2 p^2$$

$$= np \cdot e^t (2 + pe^t)^{n-1} + e^t (n-1) (p + pe^t)^{n-2} (pe^t) - n^2 p^2$$

$$= [1 + (n-1)p] - n^2 p^2$$

$$= np + np^2 (n-1) - n^2 p^2$$

$$= np + n^2 p^2 - np^2 - n^2 p^2$$

$$= np (1 - p)$$

$$\text{Var}(x) = np(1-p)$$

PART - A

I

- 1) Define standard normal distribution.
→ A special normal distribution where the mean is '0' and the standard deviation is 1.
* It is also called as z-distribution.
- 2) Define 'R' programming.
→ R is a programming language for statistical computing and data visualization.
* It has been adopted in the fields of data mining, bioinformatics and data analysis.
- 3) What is command and line environment in R?
→ The R environment is the software used to run R code. R code submitted to the R environment either by typing it directly at the command line, by cutting-and-pasting from a text file containing R code, or by specifying an entire file of R code to run.
- 4) write M.G.F of Binomial variate.
→ $M_x(t) = (2 + pe^t)^n$
This is the MGF of Binomial distribution.

PART - B

1. If x , is a random variable a and b are constants. prove that.
 - i) $E(ax+b) = a E(x) + b$
 - ii) $V(ax+b) = a^2 V(x)$
- i) $E(ax+b) = a E(x) + b$
proof :- $E(x) = \sum x \cdot p(x)$
 $E(ax+b) = \sum (ax+b) \cdot p(x)$
 $= \sum ax \cdot p(x) + \sum b p(x)$
 $= a \sum x \cdot p(x) + b \sum p(x)$
 $= a E(x) + b (1)$
 $= a E(x) + b,$

$$ii) V(ax+b) = a^2 V(x)$$

$$V(x) = E(x^2) - (E(x))^2$$

$$V(ax+b) = E(ax+b) - E(ax+b)^2$$

$$= E(ax+b - E(ax+b))^2$$

$$= E(ax - E(ax))^2$$

$$= a^2 E(x - E(x))^2$$

$$= a^2 V(x)$$

Hence proved,

$$2. \text{ For B.D p.T } p(x+1) = \left(\frac{n-x}{x+1} \right) \frac{p}{q} p(x)$$

→ We know that,

$$x \sim B(n, p)$$

The pmf of Binomial distribution

$$p(x) = \binom{n}{x} p^x q^{n-x} \rightarrow (i)$$

Similarly

$$p(x+1) = \binom{n}{x+1} p^{x+1} q^{n-(x+1)} \rightarrow (2)$$

Dividing eqn (2) by (i)

$$\frac{p(x+1)}{p(x)} = \frac{\binom{n}{x+1} p^{x+1} q^{n-x-1}}{\binom{n}{x} p^x q^{n-x}}$$

$$= \left(\frac{n!}{(n-x-1)!(x+1)!} \right) p$$

$$\frac{n!}{(n-x)! x!} q$$

$$= \frac{(n-x)(n-x-1)! x! p}{(n-x-1)!(x+1)x! q}$$

$$\frac{p(x+1)}{p(x)} = \frac{(n-x)p}{(x+1)q}$$

$$p(x+1) = \left(\frac{n-x}{x+1} \right) \frac{p}{q} p(x)$$

3. Explain 'R' as a calculator.

→ R can be used as a powerful calculator by entering equations directly at the prompt in the command console. Simply type your arithmetic expression and press enter.

* R will evaluate the expressions and respond with the R result. While this is a simple interaction interface, there could be problems if you are not careful.

* R will normally execute your arithmetic expression by evaluating each item from left to right but some operators have precedence in the order of evaluation.

Simple arithmetic expressions:

The operators R uses for basic arithmetic are

- + Addition
- Subtraction
- x Multiplication
- / Division
- ^ exponentiation

Some arithmetic expressions with example

$4 + 8$ will return the result 12

5×14 will return the result 70

$7/4$ will return the result 1.75

4^3 will return the result 64

$6 - 3$ will return the result 3

R operators classifications

1. Arithmetic
2. Relational
3. Logical
4. Assignment
5. Special operators

~~Downloaded from~~
20/10/2024



B. L. D. E. Association's

S. B. ARTS & K. C. P. SCIENCE COLLEGE, VIJAYAPUR

First

INTERNAL TEST 2023-2024

e/ಹೆಸರು: Rajeshwari Pujari Roll No./ಸ್ಥಳಾಂಕ: 29 Date/ದಿನಾಂಕ: 09-07-2024

Subject/ವಿಷಯ: Statistics paper - II Semester/ಸೆಮಿಸ್ಟರ್: VI Exam Seat No / ಪರೀಕ್ಷೆ ಸ್ಥಳಾಂಕ: ULSKM2130325

Signature of Evaluator / ಪರೀಕ್ಷಕರ ಸಹಿ

Marks obtained / ಪಡೆದ ಅಂಕಗಳು

--	--

14
15

28
30

Pranadabhi
10/07/2024

UMP test:- The critical region of ω which is a function of the mean and the parameter of sampling test.

LRT: The likelihood ratio test is defined as the variance of normal distribution and mean is μ is called LRT.

pdf of normal distribution its mean is 0 and variance is 1.

ASN function:- The sample size 'n' of the H_0 sequential equation of the random sampling with population is true density function.

i) If $\sum Z_i \geq \log A$ where reject the H_0

ii) If $\sum Z_i \leq \log B$ where accepting the H_0 / reject H_1

iii) If $\log B \sum Z_i \leq \log A$ and $\sum Z_i \leq \log$

let $x_1, x_2, \dots, x_n$ be a random sample from $f(x|\theta) = \theta x^{\theta-1}$, $0 < x < 1$, $\theta > 0$

UMP test of size α for testing of hypothesis $H_0: \theta \leq \theta_0$ v/s $H_1: \theta > \theta_0$

consider the test

$$f(x|\theta) = \theta x^{\theta-1}$$

$$\frac{f(x|\theta_0)}{f(x|\theta_1)} = \frac{\theta_0 x^{\theta_0-1}}{\theta_1 x^{\theta_1-1}}$$

$$L(x) = \prod_{i=1}^n \theta_0^{\frac{1}{n}} x^{\theta_0-1} \leq x^{\sum \theta_0}$$

$$L(x) = \prod_{i=1}^n \theta_0^{\frac{1}{n}} x^{\theta_0-1} \leq x^{\sum \theta_0}$$

$$L(x) = \prod_{i=1}^n \theta_0^{\frac{1}{n}} x^{\theta_0-1} \leq x^{\sum \theta_0}$$

The statistical test is of the form $\phi(x)$ with $\phi(x) = 1$ if $x \geq c$ and $\phi(x) = 0$ otherwise.

If $\theta_0 = 1$ then $\phi(x) = 1$ if $x \geq c$ and $\phi(x) = 0$ otherwise.

If $\theta_0 = 2$ then $\phi(x) = 1$ if $x \geq c$ and $\phi(x) = 0$ otherwise.

If $\theta_0 = 3$ then $\phi(x) = 1$ if $x \geq c$ and $\phi(x) = 0$ otherwise.

The given pdf is $f(x, \theta) = \theta e^{-\theta x}$

for $\theta_2 > \theta_1$

Then likelihood function is
$$\frac{f(x, \theta_1)}{f(x, \theta_2)} = \frac{\theta_2}{\theta_1} (\theta_1 - \theta_2)^{-\sum x_i}$$

$\frac{f(x, \theta_1)}{f(x, \theta_2)} (\theta_1 < \theta_2) \sum x_i$ is increase in $T(x) = \sum x_i$

So $\{ \exp(\theta) : \theta > 0 \}$ possess MLR in $T(x) = \sum x_i$
 $\therefore$ UMP Test is
$$\phi(x) = \begin{cases} 1 & \sum x_i > c \\ 0 & \sum x_i \leq c \end{cases}$$

where c is determined as α constant term such that

$\alpha \in H_0 : \phi(x)$

$P(\sum x_i) > c / H_0 = \alpha$

since $x_i \sim \exp(\theta)$

$\Rightarrow T = \sum x_i \sim \text{gamma}$

II

~~X~~ Likelihood Ratio Test for the variance of N.D.

Let $x_1, x_2, \dots, x_n$ be the random variable of normal distribution with mean μ of the variance of σ^2 , we have the normal distribution μ, σ^2 are

II

Q. The likelihood function

The pdf of $(0, \theta)$ is

$$f(x|\theta) = \frac{\theta}{\pi} \left(\frac{1}{\theta^2 + x^2} \right) \quad \theta > 0, \theta_1 < \theta_2$$

The likelihood function is

$$L(x) = \frac{f(x|\theta_2)}{f(x|\theta_1)} = \frac{\theta_2 / \pi (\theta_2^2 + x^2)}{\theta_1 / \pi (\theta_1^2 + x^2)}$$

$$= \frac{\theta_2 (\theta_1^2 + x^2)}{\theta_1 (\theta_2^2 + x^2)}$$

$$= \frac{\theta_2 (1 + \theta_1^2/x^2)}{\theta_1 (1 + \theta_2^2/x^2)}$$

when $x \rightarrow \pm \infty$

$$\text{Then } T(x) = \frac{\theta_2}{\theta_1} > 1$$

at $x = 0$

$$T(x) = \frac{\theta_1}{\theta_2} < 1$$

$\therefore T(x)$ is not possess of MLR property

we have

$$d_m = \frac{L_{1,m}}{L_{0,m}} = \frac{f(x_1, \theta_1)}{f(x_1, \theta_0)}$$

$$d_m = \frac{L(\theta_1, x_1, \dots, x_n | H_1)}{L(\theta_0, x_1, \dots, x_n | H_0)} = \frac{L_{1,m}}{L_{0,m}}$$

$$= \frac{\theta_1^{\sum x_i} (1-\theta_1)^{n-\sum x_i}}{\theta_0^{\sum x_i} (1-\theta_0)^{n-\sum x_i}}$$

$$= \left(\frac{\theta_1}{\theta_0} \right)^{\sum x_i} \left(\frac{1-\theta_1}{1-\theta_0} \right)^{n-\sum x_i} \rightarrow \textcircled{1}$$

$$d_m = \left(\frac{\theta_1}{\theta_0} \right)^{\sum x_i} \left(\frac{1-\theta_1}{1-\theta_0} \right)^m \left(\frac{1-\theta_1}{1-\theta_0} \right)^{-\sum x_i}$$

log on both side

$$\log d_m = \sum x_i \log \left(\frac{\theta_1}{\theta_0} \right) + (m - \sum x_i) \log \left(\frac{1-\theta_1}{1-\theta_0} \right)$$

$$= \sum x_i \log \left(\frac{\theta_1}{\theta_0} \right) + m \log \left(\frac{1-\theta_1}{1-\theta_0} \right) - \sum x_i \log \left(\frac{1-\theta_1}{1-\theta_0} \right)$$

$$\log d_m = z_i = \sum x_i \left[\log \left(\frac{\theta_1}{\theta_0} \right) \left(\frac{1-\theta_1}{1-\theta_0} \right) \right] + m \log \left(\frac{1-\theta_1}{1-\theta_0} \right)$$

Hence SPRT for testing $H_0: \theta = \theta_0$ v/s $H_1: \theta = \theta_1$ is given by

1) Reject H_0 if $\sum x_i \geq \log A$

$$\sum x_i \left[\log \left(\frac{\theta_1}{\theta_0} \right) \left(\frac{1-\theta_1}{1-\theta_0} \right) \right] + m \log \left(\frac{1-\theta_1}{1-\theta_0} \right) \geq \log \left(\frac{1-\beta}{\alpha} \right)$$

$$\sum x_i \geq \frac{\log(1-\beta/\alpha)}{\log \left(\frac{\theta_1}{\theta_0} \right) \left(\frac{1-\theta_1}{1-\theta_0} \right)}$$

ii) Reject H_0 if $\sum x_i \leq \log B$

$$\sum x_i \left[\log \left(\frac{\theta_1}{\theta_0} \right) \left(\frac{1-\theta_1}{1-\theta_0} \right) \right] + m \log \left(\frac{1-\theta_1}{1-\theta_0} \right) \leq \log \left(\frac{\beta}{1-\alpha} \right)$$

$$\sum x_i \leq \frac{\log(\beta/(1-\alpha)) - m \log((1-\theta_1)/(1-\theta_0))}{\log(\theta_1/\theta_0) - ((1-\theta_1)/(1-\theta_0))}$$

iii) If $\log B \leq \sum x_i \leq \log A$

$$b \leq \sum x_i \leq a$$

OC function is given by

$$l(\theta) = \frac{A^{h(\theta)} - 1}{A^{h(\theta)} - B^{h(\theta)}}$$

$$E(x) = E \left(\frac{f(x, \theta)}{f(x, \theta_0)} \right)^{h(\theta)}, \quad f(x, \theta) = 1$$

$$\sum_{x=0}^{\infty} \left[\left(\frac{\theta_1}{\theta_0} \right)^x \left(\frac{1-\theta_1}{1-\theta_0} \right)^{1-x} \right]^{h(\theta)} \theta^x (1-\theta)^{1-x} = 1$$

$$\Rightarrow \left(\frac{1-\theta_1}{1-\theta_0} \right)^{h(\theta)} (1-\theta)^1 + \left(\frac{\theta_1}{\theta_0} \right)^{h(\theta)} \theta^1 = 1 \Rightarrow$$

$$\left(\frac{1-\theta_1}{1-\theta_0} \right)^{h(\theta)} = \theta \left(\frac{1-\theta_1}{1-\theta_0} \right)^{h(\theta)} + \left(\frac{\theta_1}{\theta_0} \right)^{h(\theta)} \theta$$

$$E(x) = \theta \log \left[\frac{\theta_1 (1-\theta_1)}{\theta_0 (1-\theta_0)} \right] + \log \left(\frac{1-\theta_1}{1-\theta_0} \right)$$

ASNPs given by

$$F(n) = l(\theta) \frac{\log B + (1-\theta_1) \log(A)}{E(x)}$$

Here

$$l(\theta) = \frac{[(1-\beta)(\alpha)]^h - 1}{\{[(1-\beta)/\alpha]^h - [\beta/(1-\alpha)]^h\}}$$

$$\log B = \log \left(\frac{\beta}{1-\alpha} \right)$$

$$\log A = \log \left(\frac{1-\beta}{\alpha} \right)$$

$$e(x) = \theta \log \left[\frac{\theta_1 (1-\theta_0)}{\theta_0 (1-\theta_1)} \right] + \log \left(\frac{1-\theta_1}{1-\theta_0} \right)$$

Substitutes these values in (4)
we get ASN function



B. L. D. E. Association's
S. B. ARTS & K. C. P. SCIENCE COLLEGE, VIJAYAPUR

I

INTERNAL TEST 20 -20

Student's Name: Sangeeta S. Kolakar Roll No./ಸ್ಥಳಾಂಕ: Date/ದಿನಾಂಕ: 9-7-2024

Subject/ವಿಷಯ: Statistics Semester/ಸೆಮಿಸ್ಟರ್: IV Exam Seat No./ಪರೀಕ್ಷೆ ಸ್ಥಳಾಂಕ: VISKM2150358

Signature of Student/ಕೊರಗಿದ ಮೇಲ್ವಿಚಾರಕರ ಸಹಿ Signature of Evaluator/ಪರೀಕ್ಷಕರ ಸಹಿ Marks obtained/ಪಡೆದ ಅಂಕಗಳು

15/15
30/30

① UMP test:- The critical region ω is the UMP of critical region size α for testing $H_0: \theta = \theta_0$ against $H_1: \theta = \theta_1$.

if $p(x \in \omega | H_0) = \int L_0 dx$ — (1)
and $p(x \in \omega | H_1) \geq p(x \in \omega | H_0)$ whatever the critical region, equation (1) satisfied

② ASN function:- A sample size n in the sequential testing in random variable which can be determined in terms of the density function $f(x, \theta)$

The ASN function of the $H_0: \theta = \theta_0$ vs $H_1: \theta = \theta_1$ is given by

$$E(n) = \frac{L(\theta) \cdot \log B + [1 - L(\theta)] \log A}{E(Z)}$$

$$\text{where } Z = \log \left[\frac{f(x, \theta_1)}{f(x, \theta_0)} \right]$$

$$A = \frac{1 - \beta}{\alpha}$$

$$B = \frac{\beta}{1 - \alpha}$$

③ The pdf of normal distribution its mean 0 and variance σ^2 , $N(0, \sigma^2)$

$$= \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x_i - \mu}{\sigma} \right)^2}$$

$$= \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x_i}{\sigma} \right)^2}$$

④ The OC function $L(\theta)$ is defined as the $L(\theta) = p(\text{accepting } H_0: \theta = \theta_0)$ where θ is the parameter of true value of the parameter and since the power function is $p(\theta) = p(\text{rejecting } H_0: \theta = \theta_0)$ where θ is the parameter.

⑥ Condition of SPRT :-

$\lambda_n \geq A$ is determined by the process in terms of the

$\lambda_n \leq B$ we can determine the procedure in terms

$\lambda_n \leq B$ or $B < \lambda_n < A$ is continued is given by $A = \frac{1-\beta}{\alpha}$ and $B = \frac{\beta}{1-\alpha}$

Q II

③ The Cauchy's $C(0, \theta)$ is, the

The likelihood ratio function $\theta \geq 0, \theta_1 < \theta_2$

The pdf of the $f(x, \theta) = \frac{\theta}{\pi} \left(\frac{1}{x^2 + \theta^2} \right)$

$$L(\theta) = \frac{\prod_{i=1}^n f(x_i, \theta_2)}{\prod_{i=1}^n f(x_i, \theta_1)} = \frac{\frac{\theta_2^n}{\pi^n} \left(\frac{1}{x^2 + \theta_2^2} \right)}{\frac{\theta_1^n}{\pi^n} \left(\frac{1}{x^2 + \theta_1^2} \right)}$$

$$= \frac{\theta_2^n}{\theta_1^n} \left(\frac{x^2 + \theta_1^2}{x^2 + \theta_2^2} \right)$$

$$= \left(\frac{\theta_2}{\theta_1} \right)^n \left[\frac{x^2 + \theta_1^2}{x^2 + \theta_2^2} \right]$$

$$= \left(\frac{\theta_2}{\theta_1} \right)^n \left[\frac{1 + x^2/\theta_1^2}{1 + x^2/\theta_2^2} \right]$$

where $x = \pm \infty$

then $\frac{\theta_2}{\theta_1} \geq 1$

at $x = 0$

$\frac{\theta_1}{\theta_2} < 1$

The Cauchy's method not possesses the MLR property

Given the exponential distribution,
the pdf = $\theta x^{\theta-1}$ $0 < x < 1$ $\theta > 0$

The

$$\begin{aligned} L(\theta) &= \prod_{i=1}^n f(x_i, \theta) \\ &= \prod_{i=1}^n \theta x_i^{\theta-1} \\ &= \theta^n e^{-\theta \sum x_i} \\ &= (\theta)^n e^{-\theta T(x)} \end{aligned}$$

The MLR property of the family distribution is
 $H(\theta) = e^{\theta \sum x_i} \cdot V(x)$

Here, $H(\theta) = \theta^n$

$\theta \cdot \theta(\theta) = \theta$

$T(x) = \sum x_i$

$V(x) = 1$

$x \sim$ exponential distribution

$\sum x_i \sim$ Gamma distribution

$\sum x_i \sim \ell(n, \theta)$

For $f(x, \theta)$ consider,

The pdf of the distribution $f(x, \theta) = \theta x^{\theta-1}$

The likelihood ratio function is the

$$L(\theta) = \frac{\prod_{i=1}^n f(x_i, \theta_2)}{\prod_{i=1}^n f(x_i, \theta_1)} = \frac{\theta_2 \prod_{i=1}^n x_i^{\theta_2-1}}{\theta_1 \prod_{i=1}^n x_i^{\theta_1-1}}$$

$$= \frac{\theta_2}{\theta_1} \left(\prod_{i=1}^n x_i \right)^{\theta_2 - \theta_1}$$

$$\left(\frac{\theta_2}{\theta_1}\right) \left(\prod_{i=1}^n x_i\right)^{\theta_2 - \theta_1}$$

where $L(\theta)$ is an increase of the function

$$T(x) = \prod_{i=1}^n x_i (\sum x_i)$$

The UMP test size α is the

$$\phi(x) = \begin{cases} 1 & \sum \log x_i > c \\ 0 & \text{o.w.} \end{cases}$$

$$\text{Then, } U = \sum \log x_i = \alpha$$

$$U = -\sum \log x_i = \alpha$$

$$\phi(U) = \begin{cases} 1 & \sum \log x_i < c^* \\ 0 & \text{o.w.} \end{cases}$$

The c^* constant is the;

Take the transformation of y .

$$E_{H_0} \phi(x) = \alpha$$

$$P[U < c^* / \theta = \theta_0]$$

$$g(y) = \sum \log x_i \left| \frac{dx}{dy} \right|$$

$$g(y) = \sum \log x_i \quad g(y) = \theta x e^{-(\theta-1)y} e^{-y}$$

$$= \theta x e^{-y\theta + y - y}$$

$$= \theta x e^{-y\theta}$$

$$U = \sum \log x_i = \alpha$$

$$P[U < c^*]$$

$x \sim$ The Gamma distribution (n, θ)



B. L. D. E. Association's

S. B. ARTS & K. C. P. SCIENCE COLLEGE, VIJAYAPUR

INTERNAL TEST 20 23 - 2024

Sangeeta S. Kolakar

Roll No./ಸ್ಥಳಾಂಕ:

Date/ದಿನಾಂಕ: 5-8-2024

Statistical
Inference - II

Semester/ಸೆಮಿಸ್ಟರ್: II

Exam Seat No / ಪರೀಕ್ಷೆ ಸ್ಥಳಾಂಕ:

UISKM2150958

Signature / ಕೊರತೆ ಮೆಲ್ವಿಚಾರಕರ ಸಹ

Signature of Evaluator / ಪರೀಕ್ಷಕರ ಸಹ

Marks obtained / ಪಡೆದ ಅಂಕಗಳು

30/30

15/15

06/08/2024

2 marks questions

1) Define non parametric test

Non parametric test is a type of hypothesis test that does not require any assumptions above the underlying distribution of the data.

2) Write the demerits of non-parametric test.

- * Non-parametric can be used only if measurement are nominal or ordinal even in that case of a parameter test exist, it is more powerful than the Non parametric test.
- * N-p test are designed test statistical hypothesis only and not for estimating the parameter.

3) State the Kolmogorov smirnov test

Let $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ be a random statistics in a random sample of size n from continuous distribution function F . Define the following distⁿ

$$F_n(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ \frac{1}{n} & \text{if } X_{(1)} \leq x \leq X_{(2)} \\ \frac{2}{n} & \text{if } X_{(2)} \leq x \leq X_{(3)} \\ \vdots & \vdots \\ 1 & \text{if } x \geq X_{(n)} \end{cases}$$

4) Distinguish between parametric & non parametric test

Parametric	Non parametric
parametric assume a specific distribution	No distributional assumptions
It's require numerical data	can use numerical or ordinal data
more powerful	Less powerful
Allow for hypothesis testing & confidence intervals	Primarily used for hypothesis testing
Ex :- t test, ANOVA	Ex :- U test, H test

Write the test

- (5) Mann-Whitney U test is a non-parametric test that is used to compare two population means that comes from the same population.

$$U = n_1 n_2 + \frac{n_2(n_2+1)}{2} - \sum_{i=1}^{n_1} R_i$$

Q II

(1) Advantages :-

- * NP methods are readily comprehensible very simple and easy to apply and do not require complicated sampling theory
- * No assumption about the form of frequency function of the parent population from which the sampling is done
- * No parametric technique will apply to the data which are measured in nominal scale where NP methods exist to deal with such data
- * Since the socio-economic data are not in general normal distributed NP tests have found applications in psychometry, sociology and education statistics

* Disadvantages :-

- (5)
- * NP-test can be used only if the measurements are nominal or ordinal. Even in that case a parametric test exists, it is more powerful than the NP-test

In other words if all the assumption of a statistical model are satisfied by the data and if the measurements are of required strength then the NP-test are wasteful of time and data.

→ NP-tests are designed to test statistical hypotheses only and not for estimating the parameters.

01/02
Kruskal Walli's Test :- It was introduced by W. H. Kruskal & W. A. Walli's with one sample problem of variance (ANOVA) (F-test) when the normality & equality of variance assumptions are violated.

* K. W. Test is also called as H-test.

Assumptions:-

- * K-W (H) test has atleast 3 independently drawn random sample
- * Each sample has at least 5 observation i.e. $n_i \geq 5$

Null Hypothesis H_0 : The k-independent samples have been drawn from population which are identically distributed.

$$\text{i.e. } H_0: \mu_1 = \mu_2 = \dots = \mu_k$$

Alternative Hypothesis H_1 : The k-independent samples have been drawn from population which are not identically distributed.

i.e. H_1 : at least one of the μ_i is different from others.

$$\mu_1 \neq \mu_2 = \mu_3 = \dots = \mu_k$$

$$\mu_1 = \mu_2 \neq \mu_3 = \mu_4 = \dots = \mu_k$$

Or $\mu_1 \neq \mu_2 \neq \mu_3 = \dots = \mu_k$

$$\mu_1 \neq \mu_2 \neq \mu_3 \neq \dots \neq \mu_k$$

K. W. test is an extension of the mann whitney U-test to the situation where more than two population are considered & is based on the ranks of the sample observations.

Q NO

III

②

method - A 99 64 101 85 79 88 99 95 90 100

method - B 83 102 125 61 91 86 94 89 93 7

method - C 89 98 56 105 87 90 87 101 76 8

use $K-S(1)$ test to determine at 5% l.g.s if the 3 methods are equally effective.

Step 1:- Null hypothesis $H_0: \mu_A = \mu_B = \mu_C$

i.e. 3 methods are equally effective

Alternative Hypothesis H_1 : At least two of μ are

i.e. the 3 methods are not equally effective

Step 2:-

A	99	64	101	85	79	88	99	95	90	100
Rank	24	3	25.5	8	6	11	22	20	25.5	2
B	83	102	125	61	91	96	94	89	93	7
Rank	7	28	30	2	17	21	19	13	18	4
C	89	98	56	105	87	90	87	101	76	89
Rank	13	23	1	29	95	15.5	15	26.5	5	12

$$n_1 = 10, n_2 = 10, n_3 = 10, n = 30$$

$$T_1 = 24 + 3 + 25.5 + 8 + 6 + 11 + 22 + 20 + 25.5 + 2 = 161$$

$$T_2 = 7 + 28 + 30 + 2 + 17 + 21 + 19 + 13 + 18 + 4 = 159$$

$$T_3 = 13 + 23 + 1 + 29 + 95 + 15.5 + 15 + 26.5 + 5 + 12 = 145$$

Step 3:- Then the test statistic is

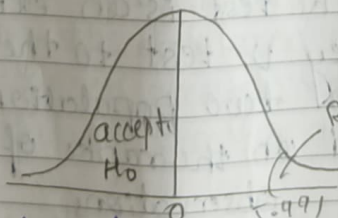
$$H = \frac{12}{n(n+1)} \left[\frac{\sum T_i^2}{n_i} - 3(n+1) \right]$$

$$= \frac{12}{30(31)} \left[\frac{(161)^2}{10} + \frac{(159)^2}{10} + \frac{(145)^2}{10} - 3(31) \right]$$

$$= 0.196$$

Step 4:- df = k-1 = 3-1 = 2

the total value $\chi^2(0.05)$ for 2df is 5.99



Since the calculated H value $0.196 < 5.99$

so it is not significant & we fail to reject H_0



ಹೆಸರು:

Pooja. v. Kolakar

Roll No./ಸ್ಥಳಾಂಕ:

Date/ದಿನಾಂಕ: 5-8-2024

ವಿಷಯ:

Statistical inference

Semester/ಸೆಮಿಸ್ಟರ್: II

Exam Seat No / ಪರೀಕ್ಷೆ ಸ್ಥಳಾಂಕ: 015KM21S0365

Signature / ಕೊಠಡಿ ಮೇಲ್ವಿಚಾರಕರ ಸಹಿ

Signature of Evaluator / ಪರೀಕ್ಷಕರ ಸಹಿ

Marks obtained / ಪಡೆದ ಅಂಕಗಳು

20/30

15/15

Suragender

2 marks questions:-

Define non-parametric test:-

Non-parametric test is a type of hypothesis test that does not require any assumptions about the underlying distributions of the data.

The demerits of non-parametric test

Demerits:-

Non-parametric can be used only if measurement are nominal or ordinal. Even in that case if a parameter test exist, it is more powerful than the Non-parametric test.

Non-parametric test are designed test statistical hypothesis only and not for estimating the parameter.

State the Kolmogorov-Smirnov test:-

Let $x_{(1)}, x_{(2)}, \dots, x_{(n)}$ be a random sample of size 'n' from continuous distribution function F. Define the following distribution.

$$F_n(x) = \begin{cases} 0 & \text{if } x \leq x_{(1)} \\ \frac{1}{n} & \text{if } x_{(1)} < x \leq x_{(2)} \\ \frac{2}{n} & \text{if } x_{(2)} < x \leq x_{(3)} \\ \vdots & \vdots \\ 1 & \text{if } x \geq x_{(n)} \end{cases}$$

where F_n is an empirical distribution.

5] write the test statistic of Mann-Whitney test
 → Mann-Whitney test is a non-parametric test that is used to compare two population means that comes from the same population.

$$u = n_1 n_2 + \frac{n_2(n_2+1)}{2} - \sum_{i=1}^{n_1+1} R_i$$

3]	parametric	Non-parametric
(*)	Parametric assume a specific distribution.	(*) No distributional assumptions.
(*)	It's require numerical data.	(*) can use numerical or ordinal data.
(*)	more powerful.	(*) Less powerful.
(*)	Allow for hypothesis testing and confidence intervals. Ex: t-test, ANOVA	(*) primary used for hypothesis testing. Ex: U-test, H test.

- Q11
- 1] Advantages of Non-parametric test:-
- (*) Np methods are readily comprehensible, very simple and easy to apply and don't require complicated sampling theory.
 - (*) No assumption about this is about made the form of frequency, function of the parent population from which the sampling is done.
 - (*) No parametric technique will apply to the data which are measured in nominal scale while Np methods exist to deal with such that data.
 - (*) Since, the social economic data are not in general normally distributed Np tests have found application in psychometry, sociology and education statistics.

Drawbacks:-

Np test can be used only if the measurement are nominal or ordinal. Even in that case if a parametric test exists, it is more powerful than Np-test.

In other words if all the assumption of a Statistical model are satisfied by the data and if the measurement are of required strength then the Np-test are wasteful of time and data.

Np-tests are designed to test Statistical hypotheses only and not for estimating the parameters.

Kruskal Walli's test:- It was introduced by W. H. Kruskal and W. A. Walli's with one sample problem, K.W Test is Np and alternative test of analysis of variance (ANOVA) (F-test) when the normality and equality of variance assumptions are isolated, K.W Test is also called as H-test.

Assumptions:- 1] K-W(H) test has at least 3 independently drawn random sample.

2] Each sample has at least 5 observation i.e. $n_1 \geq 5$

Null hypothesis:- H_0 the k-independent samples have been drawn from population which are identically distributed.
i.e. $H_0: \mu_1 = \mu_2 = \dots = \mu_k$

alternative hypothesis:- H_1 the k-independent samples have been drawn from population which are not identically distributed.
i.e. H_1 : at least one of the μ is different from others.

$\mu_1 \neq \mu_2 = \mu_3 = \dots = \mu_k$ or $\mu_1 \neq \mu_2 \neq \mu_3 = \dots = \mu_k$
 $\mu_1 = \mu_2 \neq \mu_3 = \mu_4 = \dots = \mu_k$ or $\mu_1 \neq \mu_2 \neq \mu_3 \neq \dots \neq \mu_k$

K.W test is an extension of the mann whitney U-test to the situation where more than two populations are considered and is based on the ranks of the sample observations.

QNO 11) 27.

method-A	99	64	101	85	79	88	97	95	90
method-B	83	102	105	61	91	86	94	89	93
method-C	89	98	56	105	87	90	87	101	76

use k:- w(u) test to determine at 5% l. of s if the 3 methods are equally effective.

Step 1:- Null hypothesis H_0 :- $\mu_A = \mu_B = \mu_C$ i.e. 3 methods are equally effective
Alternative hypothesis H_1 :- At least two of μ are different i.e. the 3 methods are not equally effective.

<u>Step 2:-</u> A	99	64	101	85	79	88	97	95	90
Ranks	24	3	25.5	8	6	11	22	20	15.5
B	83	102	105	61	91	86	94	89	93
Ranks	7	28	30	2	17	21	19	13	18
C	89	98	56	105	87	90	87	101	76
Ranks	13	23	1	29	95	15.5	95	26.5	5

$$n_1 = 10; n_2 = 10, n_3 = 10, n = 30.$$

$$T_1 = 24 + 3 + 25.5 + 8 + 6 + 11 + 22 + 20 + 15.5 + 25 = 161$$

$$T_2 = 7 + 28 + 30 + 2 + 17 + 21 + 19 + 13 + 18 + 4 = 159$$

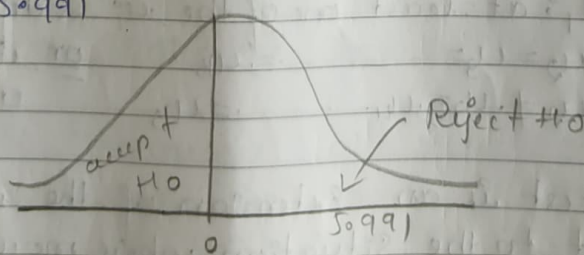
$$T_3 = 13 + 23 + 1 + 29 + 95 + 15.5 + 95 + 26.5 + 5 + 13 = 145$$

Step 3:- then the test statistic is $H = \frac{12}{n(n+1)} \frac{\sum T_i^2}{n_i} - 3(n+1)$

$$= \frac{12}{30(31)} \left[\frac{(161)^2}{10} + \frac{(159)^2}{10} + \frac{(145)^2}{10} \right] - 3(31)$$

$$= 0.196$$

Step 4:- $df = k-1 = 3-1 = 2$, the total value $\chi^2(0.05)$ for 2 df is 5.991



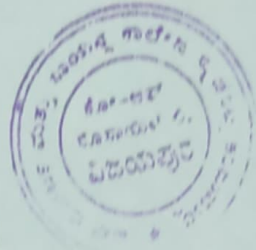
Since,

the calculated H value $0.196 < 5.991$

So it is not significant & we fail to reject H_0

B.L.D.E. Association's

**S. B. Arts & K.C.P. Science College,
VIJAYAPUR- 586 103.**



ASSIGNMENT

* For B.A./ B.Sc.^{VI}..... Semester
20²³ - 20²⁴

Name of the Student PREMA MAGADHART

Roll No. 293 R.C.U. Seat No. U1SKN2160111

Subject STATISTICAL INFERENCE II

Assignment No.	Date	Marks Assigned	Marks Obtained	Name and Signature of Teacher	Remarks
1	12/08/24	05	05	S.B. Naganoor Pranganee	Good
2					
3					
4					

① Define the following term

(a) UMP test (Uniformly most powerful):

The critical region ω is the UMP critical region of size α (and the corresponding test UMP of level α) for testing $H_0: \theta = \theta_0$ v/s $H_1: \theta \neq \theta_0$ in $H_1: \theta = \theta_1 \neq \theta_0$

(b) Wald's SPRT (Sequential probability ratio Test):

To test the hypothesis $H_0: \theta = \theta_0$ v/s $H_1: \theta = \theta_1$ (alternative hypothesis) for a distribution with probability density function $f(x; \theta)$ for any (+)ve integer 'n' the likelihood function of a sample $x_1, x_2, \dots, x_n$ from the population with pdf $f(x; \theta)$ is given by

$$L_{1n} = \prod_{i=1}^n f(x_i; \theta_1), \text{ when } H_1 \text{ is true}$$

$$L_{0n} = \prod_{i=1}^n f(x_i; \theta_0), \text{ when } H_0 \text{ is true.}$$

and likelihood ratio L_n is given by

$$L_n = \frac{L_{1n}}{L_{0n}} = \frac{\prod_{i=1}^n f(x_i; \theta_1)}{\prod_{i=1}^n f(x_i; \theta_0)} \quad n=1, 2, 3, \dots$$

(c) ASN Function (Average sample Number):

The sample size 'n' in sequential testing is a random variable which can be determined in terms of the density function $f(x; \theta)$

The ASN Function for the SPRT testing $H_0: \theta = \theta_0$ v/s $H_1: \theta = \theta_1$ is given by

$$E(n) = \frac{L(\theta) \log B + [1 - L(\theta)] \log A}{E(z)}$$

where,

$$z = \log \left[\frac{f(x_i; \theta_1)}{f(x_i; \theta_0)} \right]$$

$$A = \frac{1 - \beta}{\alpha}, \quad B = \frac{\beta}{1 - \alpha}$$

(d) OC Function (Operating Characteristic Function):

OC function $L(\theta)$ is defined as the probability that will be accepting H_0 , when θ is the true

value of the parameter,

$$L(\theta) = 1 - \alpha$$

where α is probability of rejecting H_0

The OC function of a SPRT for testing $H_0: \theta = \theta_0$ vs $H_1: \theta = \theta_1$ in a sampling from the population with pdf $f(x; \theta)$ is given by,

$$L(\theta) = \frac{A^{h(\theta)} - 1}{A^{h(\theta)} - B^{h(\theta)}}$$

For each value of $h(\theta) \neq 0$ is determine such that

$$E \left[\frac{f(x; \theta_1)}{f(x; \theta_0)} \right]^{h(\theta)} = 1$$

(e) define non-parametric test:

The nonparametric test is defined as the hypothesis test which is not based on underlying assumptions. i.e. It does not require population's distribution to be denoted by specific parameters.

(f) write the test statistics of mann-whitney U-test?

Mann-whitney 'U' test is a non-parametric test that is used to compare two population means that come from the same population. to test whether two means are equal or not.

The mann-whitney 'U' test is an alternative for the independent sample's t-test in case of a non-normality. It is used for equal sample sizes and also to test the median of two populations. this test is usually used for ordinal data.

$$U = n_1 n_2 + \frac{n_2(n_2+1)}{2} - \sum_{i=1}^{n_1} R_i$$

(g) state the kolmogorov-smirnov test is

Kolmogorov-smirnov test is another that can be adapted to the two sample problem.

In the two sample case the comparison is made between the empirical distribution function of the two samples.

The order statistics corresponding to random sample of size m & n from continuous population F_X & F_Y are.

$$x_{(1)}, x_{(2)} \dots x_{(n)} \text{ \& } y_1, y_2 \dots y_n$$

Then their respective empirical distribution function $S_m(x)$ & $S_n(x)$ are defined as.

$$S_m(x) = \begin{cases} 0 & \text{if } x < x_{(1)} \\ k/m & \text{if } x_k \leq x < x_{(k+1)} \\ 1 & \text{if } x \geq x_{(m)} \end{cases}$$

for $k = 1, 2, \dots, m$

$$S_n(x) = \begin{cases} 0 & \text{if } x < y_{(1)} \\ k/n & \text{if } y_k \leq x < y_{(k+1)} \\ 1 & \text{if } x \geq y_n \end{cases}$$

for $k = 1, 2, \dots, (n-1)$

② Write the property of MLR (Monotone Likelihood Ratio)

The MLR property is a property of the ratio of two probability density function.

The distribution of $f(x)$ and $g(x)$ accept the property if for every $x_1 < x_2$

$$\frac{f(x_1)}{g(x_1)} \leq \frac{f(x_2)}{g(x_2)}$$

The ratio is non decreasing in the argument x if $f'(x) > 0$

i.e. the functions are first differentiable, then quick

and write this property as $\frac{\partial}{\partial x} \left(\frac{f(x)}{g(x)} \right) \geq 0$

The MLR doesn't hold then theUMP test also doesn't hold the family of distribution.

One parameter exponential has monotone likelihood ratio
 $P(x; \theta) = u(\theta) v(x)$

③ find the SPRT of Binomial distribution.

We have $d_m = \frac{L_m}{L_0}$

$$d_m = \frac{L(x_1, \dots, x_m / H_1)}{L(x_1, \dots, x_m / H_0)}$$

$$= \frac{\theta_1^{\sum x_i} (1-\theta_1)^{m-\sum x_i}}{\theta_0^{\sum x_i} (1-\theta_0)^{m-\sum x_i}}$$

$$= \left(\frac{\theta_1}{\theta_0}\right)^{\sum x_i} \left(\frac{1-\theta_1}{1-\theta_0}\right)^{m-\sum x_i}$$

Taking log on both side,

$$\log d_m = \sum x_i \log \left(\frac{\theta_1}{\theta_0}\right) + (m - \sum x_i) \log \left(\frac{1-\theta_1}{1-\theta_0}\right)$$

$$= \sum x_i \log \left(\frac{\theta_1}{\theta_0}\right) + m \log \left(\frac{1-\theta_1}{1-\theta_0}\right) - \sum x_i \log \left(\frac{1-\theta_1}{1-\theta_0}\right)$$

$$= \sum x_i \left[\log \frac{\theta_1}{\theta_0} \cdot \frac{1-\theta_0}{1-\theta_1} \right] + m \log \left(\frac{1-\theta_1}{1-\theta_0}\right)$$

Hence SPRT for testing $H_0: \theta = \theta_0$ vs $H_1: \theta = \theta_1$ is given by Reject H_0 if $\log d_m \geq \log A$

$$= \sum x_i \left[\log \frac{\theta_1}{\theta_0} \cdot \frac{1-\theta_0}{1-\theta_1} \right] + m \log \left(\frac{1-\theta_1}{1-\theta_0}\right) \geq \log \left(\frac{1-\theta_1}{1-\theta_0}\right)$$

$$\sum x_i \geq \frac{\log \left(\frac{1-\theta_1}{1-\theta_0}\right)}{\log \left(\frac{\theta_1}{\theta_0}\right) \cdot \frac{1-\theta_0}{1-\theta_1}} - m \log \left(\frac{1-\theta_1}{1-\theta_0}\right)$$

If Reject H_0 , if $\sum x_i \leq \log B$

$$= \sum x_i \left[\log \frac{\theta_1}{\theta_0} \cdot \frac{1-\theta_0}{1-\theta_1} \right] + m \log \left(\frac{1-\theta_1}{1-\theta_0}\right) \leq \log \left(\frac{\theta_1}{\theta_0}\right) \cdot \frac{1-\theta_0}{1-\theta_1}$$

$$= \sum x_i \leq \frac{\log \left(\frac{\theta_1}{\theta_0}\right) \cdot \frac{1-\theta_0}{1-\theta_1}}{\log \left(\frac{\theta_1}{\theta_0}\right) \cdot \frac{1-\theta_0}{1-\theta_1}} - m \log \left(\frac{1-\theta_1}{1-\theta_0}\right)$$

$$\left[\log \frac{\theta_1}{\theta_0} \cdot \frac{1-\theta_0}{1-\theta_1} \right]$$

$$\text{if } \log B \leq \sum z_i \leq \log A$$

The OC function is given by $L(\theta) = \frac{A^{h(\theta)} - 1}{A^{h(\theta)} - B^{h(\theta)}}$

where for each value of θ $h(\theta) \neq 0$ is to be determined

$$E \left(\frac{f(x, \theta_1)}{f(x, \theta_0)} \right)^{h(\theta)}$$

$$\sum_{x=0}^1 \left[\frac{f(x, \theta_1)}{f(x, \theta_0)} \right]^{h(\theta)} f(x, \theta) = 1$$

$$\sum_{x=0}^1 \left[\left(\frac{\theta_1}{\theta_0} \right)^x \left(\frac{1-\theta_1}{1-\theta_0} \right)^{1-x} \right]^{h(\theta)} \theta^x (1-\theta)^{1-x} = 1$$

$$\left[\left(\frac{1-\theta_1}{1-\theta_0} \right)^{h(\theta)} (1-\theta) + \left(\frac{\theta_1}{\theta_0} \right)^{h(\theta)} \theta \right] = 1$$

$$\theta \left[\left(\frac{\theta_1}{\theta_0} \right)^{h(\theta)} - \left(\frac{1-\theta_1}{1-\theta_0} \right)^{h(\theta)} \right] = 1 - \left(\frac{1-\theta_1}{1-\theta_0} \right)^{h(\theta)}$$

$$\theta = \frac{1 - \left(\frac{1-\theta_1}{1-\theta_0} \right)^{h(\theta)}}{\left(\frac{\theta_1}{\theta_0} \right)^{h(\theta)} - \left(\frac{1-\theta_1}{1-\theta_0} \right)^{h(\theta)}}$$

$$\left(\frac{\theta_1}{\theta_0} \right)^{h(\theta)} - \left(\frac{1-\theta_1}{1-\theta_0} \right)^{h(\theta)}$$

$$= \theta(h) \text{ say}$$

$$\theta = h(\theta)$$

$$L(\theta) = \frac{A^{h(\theta)} - 1}{A^{h(\theta)} - B^{h(\theta)}}$$

$$A = \frac{1-\beta}{\alpha}, \quad B = \frac{\beta}{1-\alpha}$$

$$h(\theta) = 1 - \left(\frac{1-\theta_1}{1-\theta_0} \right)^{h(\theta)}$$

$$\left(\frac{\theta_1}{\theta_0} \right)^{h(\theta)} - \left(\frac{1-\theta_1}{1-\theta_0} \right)^{h(\theta)}$$

ASN function

$$E(n) = \frac{L(\theta) \log B + (1-L(\theta)) \log A}{E(z)}$$

$$z = \log \left[\frac{f(x_1, \theta_1)}{f(x_1, \theta_0)} \right]$$

$$E(z) = \sum_{x=0}^{\infty} \left[\log \left(\frac{f(x, \theta_1)}{f(x, \theta_0)} \right) \right] f(x, \theta)$$

$$= \sum_{x=0}^{\infty} \left[\log \left(\frac{\theta_1}{\theta_0} \right) \right]^x \left(\frac{1-\theta_1}{1-\theta_0} \right)^{1-x} \theta^x (1-\theta)^{1-x} \rightarrow \text{from } \textcircled{2}$$

$$= (1-\theta) \log \left(\frac{1-\theta_1}{1-\theta_0} \right) + \theta \log \left(\frac{\theta_1}{\theta_0} \right)$$

$$= \log \left(\frac{1-\theta_1}{1-\theta_0} \right) - \theta \log \left(\frac{1-\theta_1}{1-\theta_0} \right) + \theta \log \left(\frac{\theta_1}{\theta_0} \right)$$

$$E(z) = \theta \left[\log \left(\frac{\theta_1}{\theta_0} \right) \right] + \log \left(\frac{1-\theta_1}{1-\theta_0} \right)$$

④ Find the OC Function of Poisson distribution.

→ OC function of Poisson distribution.

Let

$$L(\lambda) = \frac{A^{h(\lambda)} - 1}{A^{h(\lambda)} - B^{h(\lambda)}}$$

$$\text{where } A = \frac{1-\beta}{\alpha} \quad \& \quad B = \frac{\beta}{1-\alpha}$$

$h(\lambda)$ is given by,

$$E \left[\frac{f(x_1, \lambda_1)}{f(x_1, \lambda_0)} \right]^{h(\lambda)} = 1$$

$$E \left[\frac{\frac{e^{-\lambda_1} \lambda_1^{x_1} / x_1!}{e^{-\lambda_0} \lambda_0^{x_1} / x_1!}} \right]^{h(\lambda)} = 1$$

$$E \left[e^{(\lambda_0 - \lambda_1) \left(\frac{\lambda_1}{\lambda_0} \right)^x} \right]^{h(\lambda)} = 1$$

$$e^{(\lambda_0 - \lambda_1) h(\lambda)} E \left[\left(\frac{\lambda_1}{\lambda_0} \right)^x \right]^{h(\lambda)} = 1$$

$$e^{(\lambda_0 - \lambda_1) h(\lambda)} \sum_{x=0}^{\infty} \frac{e^{-\lambda} \lambda^x}{x!} \left(\frac{\lambda_1}{\lambda_0} \right)^{x h(\lambda)} = 1$$

$$\sum_{x=0}^{\infty} \frac{d^x}{x!} \left(\frac{d_1}{d_0} \right)^{x h(d)} = e^d e^{(d_1 - d_0) h(d)} \quad E(X) = \sum x P(X)$$

$$\Rightarrow \sum_{x=0}^{\infty} \left(d \left(\frac{d_1}{d_0} \right)^{h(d)} \right)^x = e^{d + d_1 - d_0 h(d)}$$

$$e^{d \left(\frac{d_1}{d_0} \right)^{h(d)}} = e^{d + (d_1 - d_0) h(d)}$$

$$\Rightarrow d \left(\frac{d_1}{d_0} \right)^{h(d)} = d + (d_1 - d_0) h(d) \quad \left\{ \frac{d_0^x}{x!} = e^d \right.$$

$$d \left(\left(\frac{d_1}{d_0} \right)^{h(d)} - 1 \right) = (d_1 - d_0) h(d)$$

$$d = \frac{(d_1 - d_0) h(d)}{\left(\frac{d_1}{d_0} \right)^{h(d)} - 1}$$

5) state kolmogorov smirnov two sample test.
 kolmogorov smirnov test is another that can be adapted to the two sample problem.

In the two sample case the comparison is made between the empirical distribution function of the two sample.

The order statistics corresponding to two random sample of size m & n from continuous population F_X & F_Y are

$$x_1, x_2, \dots, x_m \text{ \& \; } y_1, y_2, \dots, y_n$$

Then their respective empirical distribution function $S_m(x)$ & $S_n(x)$ are defined by.

$$S_m(x) = \begin{cases} 0 & \text{if } x < x_{(1)} \\ k/m & \text{if } x_k \leq x < x_{k+1} \\ 1 & \text{if } x \geq x_{(m)} \end{cases}$$

(for $k = 1, \dots, m$)

$$S_n(x) = \begin{cases} 0 & \text{if } x < y_{(1)} \\ i/n & \text{if } y_i \leq x < y_{i+1} \\ 1 & \text{if } x \geq y_{(n)} \end{cases}$$

(for $i = 1, \dots, n$)

In a combined order arrangement of the $m+n$ sample observations $S_m(x)$ & $S_n(x)$ are respective proportions of x & y observation which don't exceed the specified value of x .

If the null hypothesis H_0 is $H_0: F_Y(x) = F_X(x) \forall x$ is true,

The population distributions are identical.

The two sided Kolmogorov-Smirnov two sample test criteria denoted by $D_{m,n}$ is based on the maximum absolute difference between the two empirical distributions

$$D_{m,n} = \max_x |S_m(x) - S_n(x)|$$

Since here only the magnitudes & not the directions of the deviations are considered,

$D_{m,n}$ is appropriate general two sided alternative,

$$H_0: F_Y(x) \neq F_X(x) \text{ for some } x$$

The rejection region is in the upper tail $D_{m,n} \geq c_\alpha$ where

$$P(D_{m,n} \geq c_\alpha | H_0) \leq \alpha$$

because of the Glivenko-Cantelli theorem the test is consistent for this alternative.

Then, the p-value is $P(D_{m,n} \geq D_0 | H_0)$

where D_0 is the observed value of the two sample Kolmogorov-Smirnov test statistic & as with one sample Kolmogorov-Smirnov $D_{m,n}$ is completely distribution free for any continuous population distribution

since order is preserved under the monotone transformation.

i.e. if we $Z = F(X)$ for the common continuous CDF we have

$S_m(z) = S_m(x)$ & $S_n(z) = S_n(x)$ where random variable Z corresponding to X has uniform distribution on the unit interval.

6

find the Ump test for Exponential distribution,

pdf of exponential distribution

$$f(x; \theta) = \theta e^{-\theta x}$$

$$f(x, \theta) = \prod_{i=1}^n f(x_i, \theta)$$

$$= \prod_{i=1}^n \theta e^{-\theta x_i}$$

$$f(x, \theta) = (\theta)^n e^{-\theta \sum x_i}$$

Then the likelihood function is,

$$L = \frac{f(x, \theta_2)}{f(x, \theta_1)}$$

$$= \frac{(\theta_2)^n e^{-\theta_2 \sum x_i}}{(\theta_1)^n e^{-\theta_1 \sum x_i}}$$

$$= \left(\frac{\theta_2}{\theta_1} \right)^n e^{-(\theta_2 - \theta_1) \sum x_i}$$

$\frac{f(x, \theta_2)}{f(x, \theta_1)}$ is increasing in $T(x) = \sum x_i$

so $\{ \text{EXP}(\theta), \theta > 0 \}$ possy MLF in $T(x) = \sum x_i$

$$\text{Ump test is, } \phi(x) = \begin{cases} 1 & \sum x_i > c \\ 0 & \sum x_i \leq c \end{cases}$$

where 'c' is constant term and it is determined such that

$$E_{H_0} \phi(x) = \alpha$$

$$P(\sum x_i > c / H_0) = \alpha$$

Since,

$$x_i \sim \text{EXP}(\theta)$$

$$\sum x_i \sim \text{Gamma}(n, \theta)$$

where 'c' is obtained from table of incomplete gamma distribution.

(8)

Construct SPRT for testing $H_0: \lambda = \lambda_0$ vs $H_1: \lambda = \lambda_1$ for poisson distribution.

→

 $X \sim P(\lambda)$

$$\text{then, } f(x) = \frac{e^{-\lambda} \lambda^x}{x!} \quad \lambda = \lambda_0 \text{ or } \lambda_1$$

TO test $H_0: \lambda = \lambda_0$ vs $H_1: \lambda = \lambda_1$

$$L = \prod_{i=1}^n f(x_i; \lambda)$$

$$= \prod_{i=1}^n \frac{e^{-\lambda} \lambda^{x_i}}{x_i!}$$

$$= \frac{e^{-n\lambda} \lambda^{\sum x_i}}{\prod_{i=1}^n x_i!}$$

$$L_{\lambda_0} = \frac{e^{-n\lambda_0} \lambda_0^{\sum x_i}}{\prod_{i=1}^n x_i!}$$

$$L_{\lambda_1} = \frac{e^{-n\lambda_1} \lambda_1^{\sum x_i}}{\prod_{i=1}^n x_i!}$$

$$h_n(x) = \frac{L_{\lambda_1}}{L_{\lambda_0}}$$

$$= \frac{e^{-n\lambda_1} \lambda_1^{\sum x_i}}{e^{-n\lambda_0} \lambda_0^{\sum x_i}} \times \frac{\prod_{i=1}^n x_i!}{\prod_{i=1}^n x_i!}$$

$$= \frac{e^{-n\lambda_1} \lambda_1^{\sum x_i}}{e^{-n\lambda_0} \lambda_0^{\sum x_i}}$$

$$= e^{-n(\lambda_1 - \lambda_0)} \left(\frac{\lambda_1}{\lambda_0} \right)^{\sum x_i}$$

applying log on both side

$$\log h_n(x) = -n(\lambda_1 - \lambda_0) \log e + \sum x_i \log \left(\frac{\lambda_1}{\lambda_0} \right)$$

$$= -n(\lambda_1 - \lambda_0) + \sum_{i=1}^n x_i \log \left(\frac{\lambda_1}{\lambda_0} \right)$$

By SPRT continue sampling is $B < h_n(x) < A$

$$\log B < \log h_n(x) < \log A$$

$$\Rightarrow \log B < -n(\lambda_1 - \lambda_0) + \sum_{i=1}^n x_i \log \left(\frac{\lambda_1}{\lambda_0} \right) < \log A$$

$$\Rightarrow \log B + n(\lambda_1 - \lambda_0) < \sum_{i=1}^n x_i \log \left(\frac{\lambda_1}{\lambda_0} \right) < \log A + n(\lambda_1 - \lambda_0)$$

$$\frac{\log B + n(\delta_1 - \delta_0)}{\log\left(\frac{\delta_1}{\delta_0}\right)} < \sum_{i=1}^n x_i < \frac{\log A + n(\delta_1 - \delta_0)}{\log\left(\frac{\delta_1}{\delta_0}\right)}$$

con (i) If

$\delta_1 > \delta_0$ From (i) continue sampling if

$$\frac{\log B + n(\delta_1 - \delta_0)}{\log\left(\frac{\delta_1}{\delta_0}\right)} < \sum_{i=1}^n x_i < \frac{\log A + n(\delta_1 - \delta_0)}{\log\left(\frac{\delta_1}{\delta_0}\right)}$$

$$\text{i.e. } a_n < \sum_{i=1}^n x_i < b_n$$

Reject H_0 , if $\sum_{i=1}^n x_i \geq b_n$

Accept H_0 , if $\sum_{i=1}^n x_i \leq a_n$

con (ii) if $\delta_1 < \delta_0$ From (i) continue sampling if

$$\frac{\log A + n(\delta_1 - \delta_0)}{\log\left(\frac{\delta_1}{\delta_0}\right)} < \sum_{i=1}^n x_i < \frac{\log B + n(\delta_1 - \delta_0)}{\log\left(\frac{\delta_1}{\delta_0}\right)}$$

$$\text{i.e. } b_n < \sum_{i=1}^n x_i < a_n$$

Reject H_0 , if $\sum_{i=1}^n x_i \leq b_n$

Accept H_0 , if $\sum_{i=1}^n x_i \geq a_n$

9) Construct SPRT for testing $H_0: \mu = \mu_0$ v/s $H_1: \mu = \mu_1$
for normal distribution $N(\mu, \sigma^2)$ where σ^2 is known.
Let $x_n \sim N(\mu, \sigma^2)$

$$\text{then } f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

To test $H_0: \mu = \mu_0$ v/s $H_1: \mu = \mu_1$, σ^2 is known

$$L = \prod_{i=1}^n f(x_i; \mu, \sigma^2)$$

$$= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x_i - \mu}{\sigma}\right)^2}$$

$$= \left(\frac{1}{\sigma\sqrt{2\pi}} \right)^n e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2}$$

$$L_0 = \left(\frac{1}{\sigma\sqrt{2\pi}} \right)^n e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu_0)^2}$$

$$L_1 = \left(\frac{1}{\sigma\sqrt{2\pi}} \right)^n e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu_1)^2}$$

$$d_n(x) = \frac{L_1}{L_0}$$

$$= \left(\frac{1}{\sigma\sqrt{2\pi}} \right)^n \cancel{e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu_0)^2}} \cancel{\left(\frac{1}{\sigma\sqrt{2\pi}} \right)^n e^{\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu_0)^2}}$$

$$d_n(x) = \frac{1}{e^{\frac{1}{2\sigma^2} \left[\sum_{i=1}^n (x_i - \mu_0)^2 - \sum_{i=1}^n (x_i - \mu_1)^2 \right]}}$$

applying log on both side

$$\log d_n(x) = \frac{1}{2\sigma^2} \left[\sum_{i=1}^n (x_i - \mu_0)^2 - \sum_{i=1}^n (x_i - \mu_1)^2 \right] \log e$$

$$= \frac{1}{2\sigma^2} \left[\sum_{i=1}^n (x_i^2 + \mu_0^2 - 2x_i\mu_0) - \sum_{i=1}^n (x_i^2 + \mu_1^2 - 2x_i\mu_1) \right]$$

$$= \frac{1}{2\sigma^2} \left[\sum_{i=1}^n x_i^2 + n\mu_0^2 - 2\mu_0 \sum_{i=1}^n x_i - \sum_{i=1}^n x_i^2 - n\mu_1^2 + 2\mu_1 \sum_{i=1}^n x_i \right]$$

$$= \frac{1}{2\sigma^2} \left[\cancel{n(\mu_0^2 - \mu_1^2)} + 2 \sum_{i=1}^n x_i (\mu_1 - \mu_0) \right]$$

$$\log d_n(x) = \frac{1}{2\sigma^2} \left[2 \sum_{i=1}^n x_i (\mu_1 - \mu_0) - n(\mu_1^2 - \mu_0^2) \right]$$

By SPRT continue sampling if $B < d_n(x) < A$

$$\log B < \log d_n(x) < \log A$$

$$\log B < \frac{1}{2\sigma^2} \left[2 \sum_{i=1}^n x_i (\mu_1 - \mu_0) - n(\mu_1^2 - \mu_0^2) \right] < \log A$$

$$2\sigma^2 \log B < 2 \sum_{i=1}^n x_i (\mu_1 - \mu_0) - n(\mu_1^2 - \mu_0^2) < 2\sigma^2 \log A$$

$$2\sigma^2 \log B + n(\mu_1^2 - \mu_0^2) < \sum_{i=1}^n x_i (\mu_1 - \mu_0) < 2\sigma^2 \log A + n(\mu_1^2 - \mu_0^2)$$

$$\frac{2\sigma^2 \log B + n(\mu_1^2 - \mu_0^2)}{2(\mu_1 - \mu_0)} < \sum_{i=1}^n x_i < \frac{2\sigma^2 \log A + n(\mu_1^2 - \mu_0^2)}{2(\mu_1 - \mu_0)}$$

case (i)

If $\mu_1 > \mu_0$ from (i) continue sampling if

$$\frac{2\sigma^2 \log B + n(\mu_1^2 - \mu_0^2)}{2(\mu_1 - \mu_0)} < \sum_{i=1}^n x_i < \frac{2\sigma^2 \log A + n(\mu_1^2 - \mu_0^2)}{2(\mu_1 - \mu_0)}$$

i.e. $a_n < \sum_{i=1}^n x_i < b_n$

Reject H_0 , if $\sum_{i=1}^n x_i \geq b_n$

Accept H_0 , if $\sum_{i=1}^n x_i \leq a_n$

case (ii)

If $\mu_1 < \mu_0$ from (i) continue sampling if,

$$\frac{2\sigma^2 \log A + n(\mu_1^2 - \mu_0^2)}{2(\mu_1 - \mu_0)} < \sum_{i=1}^n x_i < \frac{2\sigma^2 \log B + n(\mu_1^2 - \mu_0^2)}{2(\mu_1 - \mu_0)}$$

i.e. $a_n < \sum_{i=1}^n x_i < b_n$

Reject H_0 , if $\sum_{i=1}^n x_i \leq a_n$

Accept H_0 , if $\sum_{i=1}^n x_i \geq b_n$

(10)

Test for the variance of normal distribution.

Let $x_1, x_2, \dots, x_n$ be a random sample of size n from $N(\mu, \sigma^2)$ where both μ & σ^2 are unknown

we want to test hypothesis $H_0: \sigma^2 = \sigma_0^2$

N/S alternative hypothesis $H_1: \sigma^2 \neq \sigma_0^2$

Then, the parameter space (1) is

$$\Theta = \{(\mu, \sigma^2) : -\infty < \mu < \infty, \sigma^2 > 0\}$$

&

$$\Theta_0 = \{(\mu, \sigma^2) : -\infty < \mu < \infty, \sigma^2 = \sigma_0^2\}$$

The likelihood funⁿ of the sample observation is,

$$L = \prod_{i=1}^n \left(\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2} \left(\frac{x_i - \mu}{\sigma} \right)^2} \right)$$

$$= \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^n e^{-\frac{1}{2\sigma^2} \sum (x_i - \mu)^2} \rightarrow (1)$$

Then the max likelihood estimate (MLE) is given by,

$$\hat{\mu} = \bar{x} \text{ \& } \hat{\sigma}^2 = s^2 \rightarrow (2)$$

eq-n (2) \& eq-n (1) become

$$L(1) = \max L = \left(\frac{1}{\sqrt{2\pi}s^2} \right)^{n/2} e^{-n/2}$$

In subspace Θ_0 we have only one variable parameter

$$L(\Theta_0) = \left(\frac{1}{\sqrt{2\pi}\sigma_0^2} \right)^{n/2} e^{-\frac{1}{2\sigma_0^2} \sum (x_i - \mu)^2}$$

The MLE for μ is given by $\frac{\partial \log L}{\partial \mu} = 0$

$$L(\Theta_0) = \left(\frac{1}{\sqrt{2\pi}\sigma_0^2} \right)^{n/2} e^{-\frac{1}{2\sigma_0^2} \sum (x_i - \bar{x})^2}$$

$$= \left(\frac{1}{\sqrt{2\pi}\sigma_0^2} \right)^{n/2} e^{-\frac{ns^2}{2\sigma_0^2}}$$

The test statistic is given by

$$\lambda = \frac{L(\Theta_0)}{L(\Theta)} = \frac{\left(\frac{1}{\sqrt{2\pi}\sigma_0^2} \right)^{n/2} e^{-\frac{ns^2}{2\sigma_0^2}}}{\left(\frac{1}{\sqrt{2\pi}s^2} \right)^{n/2} e^{-n/2}}$$

$$= \left(\frac{s^2}{\sigma_0^2} \right)^{n/2} e^{-1/2 \left(\frac{ns^2}{\sigma_0^2} \right)}$$

We have under H_0 , the statistic

$$\chi^2 = \frac{ns^2}{\sigma_0^2}$$

It follows χ^2 distribution with $(n-1)$ d.f.

$$\lambda = \left(\frac{\chi^2}{n} \right)^{n/2} e^{-1/2 \chi^2}$$

The critical region for testing H_0 is given by

$$1 \leq k$$

$$\left(\frac{\chi^2}{n}\right)^{n/2} e^{-1/2(\chi^2-n)} \leq k$$

$$(\chi^2)^{n/2} e^{-1/2\chi^2 n/2} \leq kn^{n/2}$$

$$(\chi^2)^{n/2} e^{-1/2\chi^2} \leq k(n\bar{c})^{n/2}$$

$$(\chi^2)^{n/2} e^{-1/2\chi^2} \leq k_1$$

where k_1 is determined by,

$$P(\chi^2 < k_1') = \alpha/2 \quad \& \quad P(\chi^2 < k_1') = 1 - \alpha/2$$

$$(k_1')^{n/2} e^{-1/2 k_1'} = (k_{\alpha/2})^{n/2} e^{-1/2 k_{\alpha/2}}$$

Here,

$$k_1' = \chi_{n-1}^2, 1 - \alpha/2$$

&

$$k_{\alpha/2} = \chi_{n-1}^2, \alpha/2$$

Lower $(1 - \alpha/2)$ Point & Upper $(\alpha/2)$ Point of the χ^2 distribution with $(n-1)$ d.f.

~~Don't know~~
12/08/2024



B. L. D. E. Association's

S. B. ARTS & K. C. P. SCIENCE COLLEGE, VIJAYAPUR

Ist

INTERNAL TEST 2023-2024

Name / ಹೆಸರು: Pooja V. Kolakar Roll No. / ಸ್ಥಳಾಂಕ: _____ Date / ದಿನಾಂಕ: 12-7-2024Subject / ವಿಷಯ: Sampling distribution Semester / ಸೆಮಿಸ್ಟರ್: VI Exam Seat No / ಪರೀಕ್ಷೆ ಸ್ಥಳಾಂಕ: 015KM2150365

(P-II)

Invigilator's Signature / ಕೋಶದ ಮೇಲ್ವಿಚಾರಕರ ಸಹಿ Signature of Evaluator / ಪರೀಕ್ಷಕರ ಸಹಿ Marks obtained / ಪಡೆದ ಅಂಕಗಳು

30/30

(15/15)

Total 137/140

I

1] Populations- It may be defined as the group of the people, events, and objects is known as the "population".

(*) Sample- Finite subset of population is called as "Sample".

2] the methods of collecting SRS namely, are follows.

(*) Interview method.

(*) Mailed question nair methods.

3] Advantages of stratified random sample.

(*) The stratified verification makes an administrative convenience.

(*) The stratified verification is possible for makes a each and every unit.

4] Proportional Allocation:-

Allocation consists 'n' units is divided into 'k' stratum, of the homogeneous can be allocated. in 'n' number of size such that $N = \sum_{i=1}^k k_i$, of the simple allocation of n_i 's are $n_i = (1, 2, \dots, k)$ of the allocation is called proportional allocation.

5] the methods of collecting SRS namely, are follows:-

(*) Lottery method.

(*) Random number table.

[I] Statement:-

7] In sample random sampling without replacement, In sample mean is unbiased estimator of population mean, i.e. $E(\bar{y}_n) = (\bar{y}_N)$ Proof:- let the consider $y_1, y_2, \dots, y_N$ are the N observations of the population Mean.

(*) let the $y_1, y_2, \dots, y_n$ are the ^{size} n observations of same mean.

$$\Rightarrow \text{Population Mean } (\bar{Y}_N) = \frac{\sum_{i=1}^N y_i}{N}$$

$$\Rightarrow \text{Sample Mean } (\bar{y}_n) = \frac{\sum_{i=1}^n y_i}{n} \quad \text{--- (1)}$$

Now,

eqn (1) becomes

$$(\bar{y}_n) = \frac{\sum_{i=1}^n y_i}{n}$$

$$\bar{y}_n = \frac{\sum_{i=1}^n a_i y_i}{n}$$

$$\left\{ \because y_i = a_i y_i \right.$$

taking expectations on both sides, we get

$$E(\bar{y}_n) = E\left(\frac{\sum_{i=1}^n a_i y_i}{n}\right)$$

$$= \frac{1}{n} E\left(\sum_{i=1}^n a_i y_i\right)$$

$$= \frac{1}{n} E(a_i y_i) \quad \text{--- (2)}$$

then,

$$E(a_i) = 1 \cdot p(a=0) + 0 \cdot p(a=1)$$

$$= \frac{1 \cdot n}{N} + 0 \cdot \left(\frac{n}{N}\right)$$

$$= \frac{n}{N} \quad \text{--- (3)}$$

$$\left\{ \begin{array}{l} \because a_i = n/N \\ p = n/N \\ q = 1-p \end{array} \right.$$

Now, we substitute the eqn (3) in eqn (2).

$$E(\bar{y}_n) = \frac{1}{n} E\left(\frac{n}{N}\right) \sum_{i=1}^n y_i$$

$$= \frac{1}{n} \cdot \frac{n}{N} \sum_{i=1}^n y_i$$

$$= \frac{1}{N} \sum_{i=1}^n y_i$$

Hence, the proof $E(\bar{y}_n) = (\bar{Y}_N)$, the sample mean is unbiased estimator of population mean in simple random sample without replacement.

with usual notation pr. $v(\bar{y}_n) = \left(\frac{1}{n} - \frac{1}{N} \right) s^2$

let us consider the $y_1, y_2, \dots, y_N$ are the observations of size 'N' of the population mean, and let us consider $y_1, y_2, \dots, y_n$ are the observations of 'n' of the sample size 'n' of the sample mean.

population mean $(\bar{y}_N) = \frac{\sum_{i=1}^N y_i}{N}$

Sample mean $(\bar{y}_n) = \frac{\sum_{i=1}^n y_i}{n}$

population mean square $(\bar{y}_N)^2 = \frac{1}{N-1} \sum (y_i - \bar{y}_N)^2$

Sample mean square $(\bar{y}_n)^2 = \frac{1}{n-1} \sum (y_i - \bar{y}_n)^2$

$$v(\bar{y}_n) = E(\bar{y}_n^2) - [E(\bar{y}_n)]^2$$

$$= E(y^2) - \bar{y}_N^2 \quad \text{--- (3)}$$

$$E(y^2) = E \left[\sum_{i=1}^n \bar{y}_N^2 \right]$$

$$= \frac{1}{n} E \left[\sum_{i=1}^n (\sum \bar{y}_N^2) \right]$$

$$= \frac{1}{n} \sum_{i=1}^n E(y^2) + E \sum_{i=1}^n y_i y_i \quad \text{--- (2)}$$

$$= \frac{n}{n} \sum \bar{y}_i^2 \quad \text{--- (3)}$$

$$s^2 = \frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{y}_N)^2$$

$$s^2 (N-1) = \sum_{i=1}^N y_i^2 - \sum_{i=1}^N \bar{y}_N^2$$

$$s^2 (N-1) + N \bar{y}_N^2 = \sum y_i^2 \quad \text{--- (4)}$$

Substituting eqn (4) in eqn (3)

$$E(\bar{y}^2) = \frac{n}{N} [s^2 (N-1) + N \bar{y}_N^2]$$

$$E(\bar{y}^2) = n \left[\left(1 - \frac{1}{N} \right) s^2 + \bar{y}_N^2 \right] \quad \text{--- (5)}$$

$$E(y_i y_j) = E(a_i a_j y_i y_j)$$

$$= \frac{n(n-1)}{N(N-1)}$$

(6)

$$\Rightarrow \sum y_i^2 = \sum_{i=1}^N y_i^2 + \sum y_i y_j$$

$$\Rightarrow \sum y_i y_j = \sum_{i=1}^N y_i^2 - \sum y_i^2$$

$$\sum y_i y_j = N \bar{y}^2 - \sum y_i^2$$

$$E \sum y_i y_j = N \bar{y}^2 - S^2(N-1) + N \bar{y}^2 \quad (7)$$

$$\text{eqn (7) in eqn (6)} \quad \frac{n(n-1)}{N(N-1)} \sum y_i^2 - S^2(N-1) + N \bar{y}^2$$

Substitute eqn (7) & eqn (8) in eqn (2)

$$E(\sum \sum y_i y_j) = \frac{n(n-1)}{N(N-1)} \sum (N \bar{y}^2 - S^2(N-1) + N \bar{y}^2)$$

$$= \frac{n(n-1)}{N(N-1)} \sum (N \bar{y}^2(N-1) - (N-1) S^2)$$

$$= \frac{n(n-1)}{N(N-1)} \sum (\bar{y}^2(N-1) - S^2)$$

$$= n(n-1) \left[\bar{y}^2 - \frac{S^2}{N} \right] \quad (8)$$

Substitute eqn (8) in eqn (2)

$$E(y^2) = \frac{1}{n^2} \sum E(y^2) + \sum_{i \neq j} E(\sum y_i y_j)$$

$$= \frac{1}{n^2} \left[n(1 - \frac{1}{N}) S^2 + \bar{y}^2 \right] + \sum n(n-1) \left[\bar{y}^2 - \frac{S^2}{N} \right]$$

$$= \frac{n}{n^2} \left(\frac{S^2 - S^2}{N} \right) + \bar{y}^2 + (n-1) \left(\bar{y}^2 - \frac{S^2}{N} \right)$$

$$= \frac{1}{n} \left(\frac{S^2 - S^2}{N} \right) + \bar{y}^2 + n \bar{y}^2 - \bar{y}^2 - \frac{n S^2}{N} + \frac{S^2}{N}$$

$$= \frac{1}{n} \left[S^2 + n \bar{y}^2 - \frac{n S^2}{N} \right]$$

$$= \left(\frac{1}{n} - \frac{1}{N} \right) S^2 + \bar{y}^2$$

$$V(\bar{y}_n) = E(y^2) - E(y)^2$$

$$= \left(\frac{1}{n} - \frac{1}{N} \right)^2 + \frac{1}{N} y_N^2 - y_N^2$$

$$V(\bar{y}_n) = \left(\frac{1}{n} - \frac{1}{N} \right) S^2 \quad \text{or} \quad V(\bar{y}_n) = \left(\frac{N-n}{Nn} \right) S^2$$

Q No 11] SRS: The technique of drawing a sample in such a way each unit of the population, an equal and independent chance being included in the sample.

(*) methods of selection SRS:-

(i) Lottery method:- A ticket or chits mainly associated with each unit of the population those each sampling unit as its identification mark 1 to N. The processor of selecting an individual is simple. All the tickets or chits are placed in a container special device in which a throwing or reshuffling possible before each draw, draw of tickets or chits may be continued until a sample of the required size of the sample is obtained.

This processor of numbering tickets or chits selecting one after reshuffling becomes when the population size is large. It may be rather difficult to achieve a independent of population. Generally this method is used in place of simple random sampling.

5] Random number table method:-

A random number table on arrangement of digits 0 to 9 in either a linear or rectangular pattern, where each position is filled with one of these digits a table of random number is obtained that all random numbers table.

Common random number tables are

1] Triplets random number table.

2] Fisher's & Yates's random number table.

3] Kendall's random number table.

This method of selecting the random sample consisting a following steps.

- 1] Identify the N units in the population numbered from 1 to N .
- 2] Select at random any page at the random number table & pick up the number any row or column diagonal at random.
- 3] The population units corresponding to the numbers selected in step two constitute the random sample.

Q No 5

4] Stratified Random Sampling:-

Consisting of N units is divided into k relatively homogeneous mutually disjoint subgroups termed as strata of size $N_1, N_2, \dots, N_k$ such that $N = \sum_{i=1}^k N_i$ of simple random sampling without.

Replacement of size n_i ($i=1, 2, \dots, k$) is drawn from each of the stratum respectively, such that $n = \sum_{i=1}^k n_i$ the sample termed as str of size n and the technique of drawing such sample is called "Stratified Random Sampling".

Name / ಹೆಸರು : Prerna Magadha Roll No./ಸ್ಥಳಾಂಕ : _____ Date/ದಿನಾಂಕ : 12-7-2024Subject/ವಿಷಯ : Sampling distribution. (P-II) Semester/ಸೆಮಿಸ್ಟರ್ : VI Exam Seat No / ಪರೀಕ್ಷೆ ಸ್ಥಳಾಂಕ : _____

Invigilator's Signature / ಕೊಠಡಿ ಮೇಲ್ವಿಚಾರಕರ ಸಹಿ Signature of Evaluator / ಪರೀಕ್ಷಕರ ಸಹಿ Marks obtained / ಪಡೆದ ಅಂಕಗಳು

30/30

15/15

13/7/24

Q NO I

- ① population :- It is defined as the total number of people, objects etc.
 sample :- It is a finite number of the population is called sample.

- ③ The methods of collecting SRs
 * Lottery method
 * Random number table method.

- ⑤ Advantages of stratified random sample.
 * The stratified verification makes an administrative convenience.
 * The stratified verification is possible for make each and every unit

- ⑥ Proportional Allocation :-
 The allocation of total number consisting

Allocation consists 'n' Units is divided into k stratum. of the Homogeneity can be allocated in 'n' number of size such that $N = \sum_{i=1}^k n_i$, of the simple allocation of n_i 's are $n_i = (1, 2, \dots, k)$ of the allocation is called proportional allocation.

- ③ ① Interview method
 ② mailed Questionnaire method.

NO II

⑦

In simple random sampling without replacement of sample mean is unbiased estimator of pop-n mean

$$E(\bar{y}_n) = \bar{y}_n$$

Let $y_1, y_2, \dots, y_n$ is a random sample of size 'n' ^{population}

the
let $y_1, y_2, \dots, y_n$ is a sample of size 'n' drawn from pop sample.

$$\text{Population mean} = \frac{\sum_{i=1}^N y_i}{N}$$

$$\text{sample mean} = \frac{\sum_{i=1}^n y_i}{n}$$

$$E(\bar{y}_n) = \sum E(y_i)$$

$$E(y_i) = \sum E(a_i) y_i \rightarrow (1)$$

$$\begin{aligned} E(a_i) &= 1P(a=1) + 0P(a=0) \\ &= P(a=1) + P(a=0) \\ &= \left(\frac{n}{N}\right) + \left(1 - \frac{n}{N}\right) \end{aligned}$$

$$\left. \begin{aligned} P &= \frac{n}{N} \\ q &= 1 - P \\ &= 1 - \frac{n}{N} \end{aligned} \right\}$$

$$E(a_i) = \frac{n}{N} \rightarrow (2)$$

$$E(\bar{y}_n) = \sum_{i=1}^n E(a_i) y_i$$

$$= \frac{1}{n} E(a_i) \sum_{i=1}^n y_i$$

$$= \frac{1}{n} \cdot \frac{n}{N} \sum_{i=1}^n y_i$$

$$= \sum_{i=1}^n \frac{y_i}{N}$$

$$E(\bar{y}_n) = \bar{y}_N$$

$$V(\bar{y}_n) = \left(\frac{1}{n} - \frac{1}{N} \right) S^2$$

let us consider $y_1, y_2, y_3, \dots, y_n$ are the observation of population.

let us consider $y_1, y_2, \dots, y_n$ be a sample observation of the population

$$\text{population mean} = \frac{\sum_{i=1}^N y_i}{N}$$

$$\text{sample mean} = \frac{\sum_{i=1}^n y_i}{n}$$

$$\text{population mean square} = \frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{y}_N)^2$$

$$\text{sample mean square} = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y}_n)^2$$

$$V(\bar{y}_n) = E(y^2) - [E(y)]^2$$

$$V(\bar{y}_n) = E(y^2) - \bar{y}_N \rightarrow (1)$$

$$E(y^2) = E\left[\sum_{i=1}^n \bar{y}_N^2\right]$$

$$= \frac{1}{n} E(k \bar{y}^2)$$

$$= \frac{1}{n} \sum_{i=1}^n E(y_i^2) + E\left[\sum_{i=1}^n y_i y_j\right] \rightarrow (2) \quad \left\{ \begin{array}{l} a = n \\ n \end{array} \right.$$

$$E(\bar{y}^2) = \frac{n}{N} \sum \bar{y}_i^2 \rightarrow (3)$$

$$s^2 = \frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{y}_N)^2$$

$$s^2(N-1) = \sum_{i=1}^N y_i^2 - \sum_{i=1}^N \bar{y}_N^2$$

$$s^2(N-1) + N \bar{y}_N^2 = \sum y_i^2 \rightarrow (4)$$

substitute eq-n (4) in eq-n (3)

$$E(\bar{y}^2) = \frac{n}{N} [s^2(N-1) + N \bar{y}_N^2]$$

$$E(\bar{y}^2) = n \left[\left(1 - \frac{1}{N}\right) s^2 + \bar{y}_N^2 \right] \rightarrow (5)$$

$$E(y_i y_j) = E(a_i a_j y_i y_j) \rightarrow (6)$$

$$= \frac{n(n-1) \sum y_i y_j}{N(N-1)}$$

$$\sum y_i^2 = \sum_{i=1}^N y_i^2 + \sum_{i \neq j} y_i y_j$$

$$\sum_{i \neq j} y_i y_j = \sum_{i=1}^N y_i^2 - \sum y_i^2$$

$$\sum_{i \neq j} y_i y_j = N \bar{y}^2 - \sum y_i^2$$

$$E \sum y_i y_j = N \bar{y}^2 - s^2(N-1) + N \bar{y}^2 \rightarrow (7)$$

eq-n (7) in eq-n (6)

$$E(\sum y_i y_j) = \frac{n(n-1) \sum y_i y_j}{N(N-1)} = \frac{n(n-1) (N \bar{y}^2 - s^2(N-1) + N \bar{y}^2)}{N(N-1)}$$

substitute eq-n (7) & eq-n (5) in eq-n (2) = n(n-1)

$$E(y^2) = \frac{1}{n} E(\sum y_i^2) + \sum_{i \neq j} E(y_i y_j)^2$$

$$= \frac{1}{n} [n(1 - \frac{1}{N}) s^2 +$$

eq-n (7) in eq-n (6)

$$E(\sum y_i y_j) = \frac{n(n-1) \sum (N \bar{y}^2 - s^2(N-1) + N \bar{y}^2)}{N(N-1)}$$

$$= \frac{n(n-1) \sum [N \bar{y}^2(N-1) - (N-1)s^2]}{N(N-1)}$$

$$= \frac{n(n-1) - N(N-1) \sum [\bar{y}^2 - \frac{s^2}{N}]}{N(N-1)}$$

$$= n(n-1) [\sum \bar{y}^2 - \frac{s^2}{N}] \rightarrow (8)$$

substitute eq-n (8) in eq-n (2)

$$E(y^2) = \frac{1}{n^2} \sum E(y^2) + \sum_{i \neq j} E(y_i y_j)$$

$$= \frac{1}{n^2} [n(1 - \frac{1}{N}) s^2 + \bar{y}^2 N] + \sum n(n-1) [\bar{y}^2 - \frac{s^2}{N}]$$

$$= \frac{n}{n^2} (\frac{s^2 - s^2}{N}) + \bar{y}^2 N + (n-1) (\bar{y}^2 N - \frac{s^2}{N})$$

$$= \frac{1}{n} (\frac{s^2 - s^2}{N}) + \bar{y}^2 N + n \bar{y}^2 N - \bar{y}^2 N - \frac{n s^2}{N} + \frac{s^2}{N}$$

$$= \frac{1}{n} [s^2 + n \bar{y}^2 N - \frac{n s^2}{N}]$$

$$= (\frac{1}{n} - \frac{1}{N}) s^2 + \bar{y}^2 N$$

$$V(\bar{y}_n) = E(y^2) - [E(y)]^2$$

$$= \left(\frac{1}{n} - \frac{1}{N}\right) \frac{\sum y_i^2}{N} - \bar{y}^2$$

$$V(\bar{y}_n) = \left(\frac{1}{n} - \frac{1}{N}\right) s^2 //$$

$$\text{or } V(\bar{y}_n) = \left(\frac{N-n}{Nn}\right) s^2 //$$

Q No 15 SRS : The technique of drawing a sample in such a way each unit of the popⁿ has an equal and independent chance of being included in the sample.

(8) method of selecting SRS is

(i) Lottery method is

A ticket or chit mainly associated with each unit of the population those each sampling unit & its identification mark 1 to N. The process of selecting an individual is simple. All the tickets or chits are placed in a container or spherical device. In which a throwing or reshuffling possible before each draw. Draw of tickets or chits may be continued until a sample of the required size of the sample is obtained.

This process of numbering tickets or chits selecting one after reshuffling becomes when the population size is large. It may be rather difficult to achieve a thorough shuffling in practice. This method is quite independent of population size. Generally this method is used in place of SRS.

(ii) Random number table method is

A random number table on a sequence of digits 0 to 9 in either a linear or rectangular pattern, where each position is filled with one of these digits a table of random number is obtained that are random numbers table.

Common random number tables are

- ① Triplet random number table
- ② Fishers & Yates random number table
- ③ Kindolt's random number table

This method of selecting the random sample consisting a following steps.

- ① Identify the N units in the population numbers from 1 to N
- ② Select at random any page at the random number table & pick up the number any row or column diagonal at random.
- ③ The population units corresponding to the numbers selected in step too constitute the random sample.

Q No I

④ stratified random sampling:-

Consisting of N units is divided into k relatively homogeneous mutually disjoint sub groups termed as strata of size $N_1, N_2, \dots, N_k$ such that $N = \sum_{i=1}^k N_i$ of simple random sampling without replacement of size $n_i (i=1, 2, \dots, k)$ is drawn from each of the stratum respectively such that $n = \sum_{i=1}^k n_i$. The sample termed as str of size n and the technique of drawing such sample is called stratified random sampling.



B. L. D. E. Association's

S. B. ARTS & K. C. P. SCIENCE COLLEGE, VIJAYAPUR

IInd. INTERNAL TEST 2023 -2024

Name / ಹೆಸರು : Sangeeta S. Kolakar Roll No./ಸ್ಥಳಾಂಕ : _____ Date / ದಿನಾಂಕ : 8-8-2024

Subject/ವಿಷಯ : Sampling methods Semester/ಸೆಮಿಸ್ಟರ್ : VI Exam Seat No / ಪರೀಕ್ಷೆ ಸ್ಥಳಾಂಕ : VI SKM 2160358
and official statistics and Econometrics

Invigilator's Signature / ಕೊಠಡಿ ಮೇಲ್ವಿಚಾರಕರ ಸಹಿ Signature of Evaluator / ಪರೀಕ್ಷಕರ ಸಹಿ Marks obtained / ಪಡೆದ ಅಂಕಗಳು

(30/30)

(15/15)

Moskeri
8/8/24

Q1

(b) The methods of collection of official statistics

- * ~~Surveys~~
- * ~~Census~~

(a) Consisting N units is divided into k relatively homogeneous mutually disjoint of subgroups is known as the strata

Sampling interval :-

Sampling interval is a fixed time period or distance between consecutive observations

$$SI = \frac{N}{n}$$

(a) CSO (Central Statistical Organization :- Central Statistical Organization is a government agency for responsible of collecting analysing disseminating of statistical data and information of the country

(c) Advantages of SRS (Stratified random sample)

- * Stratification is a designed for a administrative convenience
- * Stratification is a natural characteristic helps to improve sampling design
- * Stratification makes possible use of sampling design in different strata

(f) Econometrics

Econometrics is branch of economics which deals with the quantitative data of economic behaviour is called as the econometrics.

Q II

(a) * Collection of data relative to industries and crop estimator

* collection of data on different socio-economic condition such as literacy, birth and death rate, density, sex ratio etc.

* It coordinates on the survey of collection of data on agricultural sector

* Conducting an annual survey of the industries

* Deciding the type of the particular round of the industry

* These are the functions of the NSS.

(b) The functions of the CSO

* It provides a technical leadership in building up the statistical systematic sample of the country

* For compilation and releasing of Statistical account industry every year

* It organizes of the conference of the central and state organization

* periodic economic census

- * production of the report on annual survey of the industry
- * Survey take to be taken by in NSSO

Q III

(1)

Statement:- variance of estimated mean

$$V(\bar{y}_{sys}) = \frac{(N-1)S^2}{N} - \frac{(n-1)k}{N} S_{wsy}^2$$

$$= \frac{(N-1)S^2}{N} - \frac{n-1}{N} S_{wsy}^2$$

where $S_{wsy}^2 = \frac{1}{k(n-1)} \sum_{i=1}^k \sum_{j=1}^n (y_{ij} - \bar{y}_{i.})^2$

S_{wsy}^2 is the mean square among the units which lie within a same systematic sample

Proof:- Let population consist N units
let sample of size n drawn from the population such that $N = nk$

$\therefore$ where k is sampling interval
 $\therefore$ i th systematic sample mean is,

$$\bar{y}_{i.} = \bar{y}_{sys} = \frac{1}{n} \sum_{j=1}^n y_{ij}$$

$$V(\bar{y}_{sys}) = \frac{1}{k} \sum_{i=1}^k (\bar{y}_{i.} - \bar{y}_{..})^2$$

population mean $\bar{y}_{..} = \frac{1}{k} \sum_{i=1}^k \bar{y}_{i.}$

population mean square $S^2 = \frac{1}{N-1} \sum_{i=1}^k \sum_{j=1}^n (y_{ij} - \bar{y}_{..})^2$

$$(N-1)S^2 = \sum_i \sum_j (y_{ij} - \bar{y}_{..})^2$$

$$= \sum_i \sum_j (y_{ij} - \bar{y}_{i.} + \bar{y}_{i.} - \bar{y}_{..})^2$$

$$\sum_i \sum_j (y_{ij} - \bar{y}_{i.})^2 + \sum_i \sum_j (\bar{y}_{i.} - \bar{y}_{..})^2 + 2 \sum_i \sum_j (y_{ij} - \bar{y}_{i.})(\bar{y}_{i.} - \bar{y}_{..})$$

$$= (N-1)S^2 = \sum_i \sum_j (y_{ij} - \bar{y}_{i.})^2 + \sum_i n(\bar{y}_{i.} - \bar{y}_{..})^2 + 0$$

Given $S^2_{wsy} = \frac{1}{k(n-1)} \sum_i \sum_j (y_{ij} - \bar{y}_{i.})^2$

$$S^2_{wsy} k(n-1) = \sum_i \sum_j (y_{ij} - \bar{y}_{i.})^2$$

eqn (1) becomes

$$(N-1)S^2 = k(n-1)S^2_{wsy} + nk \text{ var}(\bar{y}_{sys})$$

$$nk \text{ var}(\bar{y}_{sys}) = (N-1)S^2 - k(n-1)S^2_{wsy}$$

$$\text{var}(\bar{y}_{sys}) = \frac{(N-1)S^2}{nK} - \frac{k(n-1)S^2_{wsy}}{nK} \quad (\because nk = nK)$$

$$\boxed{\text{var}(\bar{y}_{sys}) = \frac{(N-1)S^2}{N} - \frac{k(n-1)}{N} S^2_{wsy}}$$

or

$$\text{var}(\bar{y}_{sys}) = \frac{(N-1)}{nK} S^2 - \frac{k(n-1)}{N} S^2_{wsy}$$

10

$$\text{var}(\bar{y}_{sys}) = \frac{(N-1)}{N^2} S^2 - \frac{(n-1)}{N} S^2_{wsy}$$

(b) * collection of data on regular basis by official agencies through censuses and surveys

* Collection of data by regulatory agencies for administrative uses through statistical returns (e.g. Review of Industrial Disputes in India).

2



B. L. D. E. Association's

S. B. ARTS & K. C. P. SCIENCE COLLEGE, VIJAYAPUR

Second INTERNAL TEST 2023 -2024

name / ಹೆಸರು: Rajeshwari M. puja Roll No./ಸ್ಥಳಾಂಕ: 29 Date/ದಿನಾಂಕ: 08-08-2024
 Subject/ವಿಷಯ: Statistics paper-II Semester/ಸೆಮಿಸ್ಟರ್: VI Exam Seat No / ಪರೀಕ್ಷೆ ಸ್ಥಳಾಂಕ: U15K1M2150325

Evaluator's Signature / ಕೊಠಡಿ ಮೇಲ್ವಿಚಾರಕರ ಸಹಿ Signature of Evaluator / ಪರೀಕ್ಷಕರ ಸಹಿ Marks obtained / ಪಡೆದ ಅಂಕಗಳು

30/30

15/15 Date 8/8/24

Advantages of Stratified Random sample.

- * Stratification may be designed for administrative convenience.
- * Stratification by ~~natural~~ characteristics helps in improving the sampling design.
- * Stratification brings a gain in the precision in estimation of a characteristics of a popⁿ.

- * sense
- * survey.

Systematic random sample commonly used method of the compute and update list of in population is available.

Stratification may be designed if administrative convenience.

economic is branch of economics dealing with quantitative analysis of economics.

or

Functions of NSS

Collection of data relating to industries & crop estimation

collection of data on different socio-economic condition such as literacy birth & death density test saturation etc.

- * It co-ordinative survey of collect data form agriculture sector.
- * conducting a Annual survey of industries on yearly basis
- * deciding the topic of the particular random survey.

2) Functions of CSO.

- * It provide technical leadership in building up statistical system of country.
- * For compilation and release of national account statistics every year.
- * It organise of conference of central & state organization
- * publication reports and Annual survey of industries (ASI)
- * surveys to be taken by in NSSO
- * periodic economic census.

III

Statement:- variance of estimated mean.

$$V(\bar{y}_{sys}) = \left(\frac{N-1}{N} \right) s^2 - (n-1)k / Ns^2_{sys}$$

$$\text{when } s^2_{sys} = \frac{1}{k(n-1)} \sum \sum (y_{ij} - \bar{y}_{i.})^2$$

s^2_{sys} is the mean square among the units which the written some systematic sample

Proof:- let population consist of N units
 let the systematic sample of size n be drawn from population such that $N = nk$
 where

$$k = \text{is sampling mean is}$$

$$\bar{y}_i = \bar{y}_{sys} = \frac{1}{n} \sum y_{ij}$$

$$V(\bar{y}_{sys}) = \frac{1}{k} \sum \sum (\bar{y}_{..} = \frac{1}{k}$$

pop^{en} mean $\bar{y}_{..} = 1/k \sum \bar{y}_{i.}$

pop^m mean square $s^2 = 1/(N-1) \cdot \sum \sum (y_{ij} - \bar{y}_{..})^2$

$$(N-1)s^2 = \sum \sum (y_{ij} - \bar{y}_{i.} + \bar{y}_{i.} - \bar{y}_{..})^2$$

$$= \sum \sum (y_{ij} - \bar{y}_{i.})^2 + \sum \sum (\bar{y}_{i.} - \bar{y}_{..})^2 + 2 \sum \sum (\bar{y}_{i.} - \bar{y}_{..})(y_{ij} - \bar{y}_{i.})$$

$$(N-1)s^2 = \sum \sum (y_{ij} - \bar{y}_{i.})^2 + \sum n(\bar{y}_{i.} - \bar{y}_{..})^2 + 0 \rightarrow (1)$$

eqⁿ (1) becomes.

$$(N-1)s^2 = k(n-1)s^2_{wy} + nk \text{var}(\bar{y}_{sys})$$

$$nk \text{var}(\bar{y}_{sys}) = (N-1)s^2 - k(n-1)s^2_{wy}$$

$$\text{var}(\bar{y}_{sys}) = \frac{(N-1)s^2}{nk} - \frac{k(n-1)}{nk} s^2_{wy}$$

$$\text{var}(\bar{y}_{sys}) = \frac{(N-1)s^2}{N} - \frac{k(n-1)}{N} s^2_{wy}$$

$$\text{var}(\bar{y}_{sys}) = \frac{(N-1)s^2}{N} - \frac{(n-1)}{N} s^2_{wy}$$

- 1) CSO was set up on 2nd may 1951. It comes under the ministry of statistics and programme implementation. It holds a prominent in Indian statistical system. It co-ordinates the statistical activities in the country & enforces statistics standards. It is headed by director general assisted by five additional director general.

2) * collection of the data on regular basis by official agencies through censuses and surveys

* collection data of regularity agencies for administrative uses through statistical returns.

B.L.D.E. Association's

**S. B. Arts & K.C.P. Science College,
VIJAYAPUR- 586 103.**



ASSIGNMENT

For B.A./ B.Sc. VI-(P-II) Semester
20 - 20

Name of the Student Sakshi

Roll No. _____ R.C.U. Seat No. U15KM2130420

Subject Sampling Method, Official statistics & econometrics

Assignment No.	Date	Marks Assigned	Marks Obtained	Name and Signature of Teacher	Remarks
1	12/08/24	05	05	st. J.D. Horakmi <u>Delonki</u>	Good.
2					
3					
4					

① Define simple random sample?
⇒ the technique of drawing a sample in such a way each unit of the population has an equal and independent chance of being included in the sample.

② Define poplⁿ and sample?

① Population :- All objects under consideration is called population.

② Sample :- finite sub set of population is called sample.

③ Write the merits of census enumeration?

① the results are not accurate

② the data are collected each & every unit

③ It provides the detailed study of all the units of poplⁿ

④ Mention the principles of sample survey?

① Statement of objectives

② Definition of the poplⁿ to be studied

③ Determination of sampling frame & sampling units

④ selection of proper sampling design

⑤ Organisation of field

⑥ Summary & Analysis of data.

⑤ Write the advantages of stratified random sample?

① stratified may be designed for administrative convenience

② stratification by natural characters helps in improving the sampling design.

③ stratification makes it possible to use different sampling designs in different strata.

⑥ Define stratified random sample?

→ Consisting of a popⁿ is divided into k relatively homogeneous mutually disjoint subgroups termed as strata of size $N_1, N_2, \dots, N_k$ such that $N = \sum N_i$. A SRSWOR of size n_i ($i=1, 2, \dots, k$) is drawn from each of the stratum respectively such that $n = \sum n_i$. The sample termed as stratified random sample. A special case of size n & the technique of drawing such sample is called SRS.

⑦ Define proportional allocation?

→ Let the popⁿ consisting of N units is divided into k strata relatively homogeneous mutually disjoint strata of size $N_1, N_2, \dots, N_k$ such that $N = \sum N_i$. Let sample allocations of n_i 's ($i=1, 2, \dots, k$) to various strata is called proportional allocation.

⑧ Define optimum allocation?

→ The allocation of n_i 's various strata in accordance with the below principle (i) known as optimum allocation. Thus in optimum allocation n_i 's are to be obtained such that

- (i) $V_{\text{as}}(y_{st})$ is minimum for fixed sample size n
- (ii) $V_{\text{as}}(y_{st})$ is minimum for fixed cost
- (iii) total cost is minimum for fixed value of various y_{st}

⑨ What are the methods of collection of SRS?

- ① Television method
- ② Mailed questionnaire method

⑩ Define cost function?

→ In stratified sampling cost may vary from stratum to stratum. The simplest form of cost function in stratified sampling is

$$C = a + \sum c_i n_i$$

where a = overhead cost

C_i = cost per unit in the i^{th} stratum
 n_i = particular sampling size.

5 Marks

① Explain sampling error & non-sampling error

⇒ Sampling error :- Sampling errors have their origin in sampling and arise due to fact that only a part of the poplⁿ has been used to estimate poplⁿ parameters and draw the inferences about the poplⁿ.

Sampling error due to following reasons

- ① Faulty selection of sample
- ② Substitution of the random sample number
- ③ Faulty demarcation of sampling units
- ④ Constant error due to improper choice of the stratified for estimate the poplⁿ parameters

Non-sampling error :-

A non-sampling error is an error that results during data collection causing the data to differ from the true values.

Types of Non-sampling error.

① Non-Response error

② Response error

③ Tabulation error.

④

② Explain the steps involved in the sample survey?

⇒ ① Statement of objectives :- The first step is to define in clear and terms the objectives of the survey. The survey should take care that these objectives are dealt with available resources in terms of money, man power & time limit required for the available of the results for the

Survey.

- ② Defining the poplⁿ to be studied :- the poplⁿ from which the sample is chosen should be defined in clear and unambiguous terms.
- ③ Determination of sampling frame of sampling units :- the poplⁿ must be capable of division into what are called sampling units for the purposes of the sample selection. The sampling units must cover the entire poplⁿ and they must be distinct, unambiguous and non-overlapping.
- ④ selection of proper sampling design :- select proper sampling design to give estimation so for collection of this less cost.

③ In SRSWOR, show that mean is unbiased estimator of poplⁿ mean.

⇒ Proof :- $E(\bar{y}_n) = \bar{y}_N$

Let poplⁿ size be N
observations are $y_1, y_2, \dots, y_N$

$$\text{popl}^n \text{ mean } (\bar{y}_N) = \frac{\sum_{i=1}^N y_i}{N}$$

Let the sample size be n and obsⁿ are $y_1, y_2, \dots, y_n$

$$\text{sample mean } \bar{y}_n = \frac{\sum_{i=1}^n y_i}{n}$$

$$\bar{y}_n = \frac{\sum_{i=1}^n a_i y_i}{n} \quad y_i = a_i y_i$$

Taking expectation on both sides

$$E(\bar{y}_n) = E\left(\frac{\sum_{i=1}^n a_i y_i}{n}\right)$$

$$= \frac{1}{n} E(\sum_{i=1}^n a_i y_i)$$

$$= \frac{1}{n} E(a_i) y_i$$

$$\text{where } a_i = \begin{cases} 1 \\ 0 \end{cases}$$

$$\begin{aligned} E(a_i) &= 1p(a=1) + 0p(a=0) \\ &= p(a=1) + p(a=0) \\ &= 1 \cdot \frac{n}{N} + 0 \left(1 - \frac{n}{N}\right) \end{aligned}$$

$$E(a_i) = \frac{n}{N}$$

By definition

$$\begin{aligned} E(\bar{y}_n) &= \frac{1}{n} \sum_{i=1}^n y_i \\ &= \frac{1}{n} \sum_{i=1}^n y_i \end{aligned}$$

$$E(\bar{y}_n) = \frac{\sum y_i}{n}$$

$$E(\bar{y}_n) = \bar{y}_N$$

⑤ Insite notation $V(\bar{y}_{st}) = \frac{1}{N^2} \sum N_i (N_i - n_i) \frac{s_i^2}{n_i}$

⇒ Proof ÷ let poplⁿ size be N divided into k strata of size $N_1, N_2, \dots, N_k$ for each strata SRSWOR of sizes $n_1, n_2, \dots, n_k$ drawn respectively such that $n = \sum n_i$

since $\bar{y}_{Ni} = \frac{1}{N_i} \sum y_{ij}$ i^{th} poplⁿ stratum mean

$\bar{y}_{ni} = \frac{1}{n_i} \sum y_{ij}$ i^{th} sample stratum mean

poplⁿ mean

$$\begin{aligned} \bar{y}_N &= \frac{1}{N} \sum_{i=1}^k \left(\sum_{j=1}^{N_i} y_{ij} \right) \\ &= \frac{1}{N} \sum_{i=1}^k N_i \bar{y}_{Ni} \end{aligned}$$

$s_i^2 = \frac{1}{N_i - 1} \sum_{j=1}^{N_i} (y_{ij} - \bar{y}_{Ni})^2$ i^{th} stratum poplⁿ mean square.

$$s_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{ni})^2 \quad \text{ith stratum sample mean square}$$

We have

$$\bar{y}_{st} = \frac{1}{N} \sum N_i \bar{y}_{ni}$$

Taking variance on both sides

$$v(\bar{y}_{st}) = \text{var}\left(\frac{1}{N} \sum N_i \bar{y}_{ni}\right)$$

$$= \text{var}\left(\frac{\sum N_i}{N} \bar{y}_{ni}\right)$$

$$= \text{var}\left(\sum p_i \bar{y}_{ni}\right) \quad \because p_i = \frac{N_i}{N}$$

$$= \sum_{i=1}^K p_i^2 \text{var}(\bar{y}_{ni})$$

$$= \sum_{i=1}^K p_i^2 \left(\frac{1}{n_i} - \frac{1}{N_i}\right) s_i^2$$

$$= \sum_{i=1}^K \left(\frac{1}{n_i} - \frac{1}{N_i}\right) p_i^2 s_i^2$$

$$= \sum_{i=1}^K \left(\frac{1}{n_i} - \frac{1}{N_i}\right) \frac{N_i^2}{N^2} \cdot s_i^2$$

$$= \sum_{i=1}^K \left(\frac{N_i^2 - n_i N_i}{N_i^2 n_i}\right) \frac{N_i^2}{N^2} \cdot s_i^2$$

$$= \sum_{i=1}^K \left(\frac{N_i - n_i}{N_i}\right) \frac{N_i^2}{N^2} \cdot \frac{s_i^2}{n_i}$$

$$\therefore \text{var}(\bar{y}_{st}) = \frac{1}{N^2} \sum N_i (N_i - n_i) \frac{s_i^2}{n_i}$$

10 Marks

(i) With usual notations P.T

$$v(\bar{y}_n) = \left(\frac{1}{n} - \frac{1}{N}\right) s^2$$

proof:— let the considered $y_1, y_2, \dots, y_N$ are the poplⁿ obsⁿ of size N $y_1, y_2, \dots, y_n$ sample obsⁿ of size n

poplⁿ mean $(\bar{y}_N) = \frac{\sum y_i}{N}$

sample mean $(\bar{y}_n) = \frac{\sum y_i}{n}$

poplⁿ mean square $s^2 = \frac{1}{N-1} \sum (y_i - \bar{y}_N)^2$

Sample mean square $s^2 = \frac{1}{n-1} \sum (y_i - \bar{y}_n)^2$

Now we have

$$\begin{aligned} V(\bar{y}_n) &= E(\bar{y}_n)^2 - [E(\bar{y}_n)]^2 \\ &= E(\bar{y}_n)^2 - (\bar{y}_N)^2 \rightarrow (1) \end{aligned}$$

Now,

$$\begin{aligned} V(\bar{y}_n) &= E(\bar{y}_n^2) - \\ E(\bar{y}_n^2) &= E\left(\frac{1}{n} \sum y_i\right)^2 \\ &= \frac{1}{n^2} E\left(\sum y_i\right)^2 \end{aligned}$$

$$= \frac{1}{n^2} E\left(\sum y_i + \sum y_i y_j\right)$$

$$E(\bar{y}_n^2) = \frac{1}{n^2} \left[E(\sum y_i^2) + E(\sum y_i y_j) \right] \rightarrow (2)$$

Now consider,

$$\begin{aligned} E(\sum y_i^2) &= E(\sum a_i y_i^2) \\ &= \sum E(a_i) y_i^2 \end{aligned}$$

we have

$$E(a_i) = \frac{n}{N}$$

$$= \frac{n}{N} y_i^2$$

$$E(y_i^2) = \frac{n}{N} \sum y_i^2 \rightarrow (3)$$

WKT

$$s^2 = \frac{1}{N-1} (\sum y_i - \sum \bar{y}_N)^2 \rightarrow *$$

$$(\sum y_i - \bar{y}_N)^2 = \sum y_i^2 - 2\sum y_i \bar{y}_N + N \bar{y}_N^2$$

$$E \sum y_i^2 = (\sum y_i - \bar{y}_N)^2 + N \bar{y}_N^2$$

$$\sum y_i^2 = (N-1)S^2 + N \bar{y}_N^2 \rightarrow (4)$$

From (3) & (4)

$$E(\sum y_i^2) = \frac{n}{N} [(N-1)S^2 + N \bar{y}_N^2]$$

$$= n \left[\left(1 - \frac{1}{N}\right) S^2 + \bar{y}_N^2 \right] \rightarrow (5)$$

Now,

$$E(\sum y_i y_j) = E(\sum a_i a_j y_i y_j)$$

$$= \sum E(a_i a_j) y_i y_j$$

WKT

$$E(a_i a_j) = \frac{n(n-1)}{N(N-1)}$$

$$= \frac{\sum n(n-1)}{N(N-1)} y_i y_j$$

$$E(\sum y_i y_j) = \frac{n(n-1)}{N(N-1)} \sum y_i y_j \rightarrow (6)$$

But

$$\sum (y_i^2) = \sum y_i^2 + \sum y_i y_j$$

$$\sum y_i y_j = (\sum y_i)^2 - \sum y_i^2$$

$$= N(\bar{y}_N)^2 - \sum y_i^2$$

$$\sum y_i y_j = N \bar{y}_N^2 - ((N-1)S^2 + N \bar{y}_N^2)$$

Eqⁿ (6) becomes

$$E(\sum y_i y_j) = \frac{n(n-1)}{N(N-1)} [N \bar{y}_N^2 - (N-1)S^2 - N \bar{y}_N^2]$$

$$= \frac{n(n-1)}{N(N-1)} [N \bar{y}_N^2 (N-1) - (N-1)S^2]$$

$$= \frac{n(n-1)}{N(N-1)} [(N-1) (N \bar{y}_N - s^2)]$$

$$= n(n-1) \left(\bar{y}_N^2 - \frac{s^2}{N} \right) \rightarrow (7)$$

from eqⁿ (5) & (7) in (2)

$$E(\bar{y}_n)^2 = \frac{1}{n^2} \left[n \left(\frac{1}{n} - \frac{1}{N} \right) s^2 + \bar{y}_N^2 + n(n-1) \left(\bar{y}_N^2 - \frac{s^2}{N} \right) \right]$$

$$= \frac{n}{n^2} \left\{ s^2 - \frac{s^2}{N} + \bar{y}_N^2 + (n-1) \bar{y}_N^2 - (n-1) \frac{s^2}{N} \right\}$$

$$= \frac{1}{n} \left\{ s^2 - \frac{s^2}{N} + \bar{y}_N^2 + n \bar{y}_N^2 - \bar{y}_N^2 - \frac{n s^2}{N} + \frac{s^2}{N} \right\}$$

$$= \frac{1}{n} \left\{ s^2 + n \bar{y}_N^2 - \frac{n s^2}{N} \right\}$$

$$= \frac{1}{n} \left\{ s^2 \left(\frac{1}{n} - \frac{1}{N} \right) + n \bar{y}_N^2 \right\}$$

$$= \frac{s^2}{n} \left(\frac{1}{n} - \frac{1}{N} \right) + \bar{y}_N^2$$

$$E(\bar{y}_n^2) = s^2 \left(\frac{1}{n} - \frac{1}{N} \right) + \bar{y}_N^2 \rightarrow (8)$$

$$v(\bar{y}_n) = E(\bar{y}_n^2) - [E(\bar{y}_n)]^2$$

$$= s^2 \left(\frac{1}{n} - \frac{1}{N} \right) + \bar{y}_N^2 - \bar{y}_N^2$$

$$v(\bar{y}_n) = \left(\frac{1}{n} - \frac{1}{N} \right) s^2$$

② show that $v(\bar{y}_n) > v(\bar{y}_n)_{prop} \geq v(\bar{y}_n)_{Ney}$

$$\text{proof: } v(\bar{y}_n)_{SRS} = \left(\frac{1}{n} - \frac{1}{N} \right) s^2 \rightarrow (1)$$

$$v(\bar{y}_n)_{prop} = \left(\frac{1}{n} - \frac{1}{N} \right) s p^0 s_i \rightarrow (2)$$

$$v(\bar{y}_n)_{Ney} = \frac{1}{n} \left[\sum p_i s_i^2 - \frac{1}{N} \left(\sum p_i s_i \right)^2 \right] \rightarrow (3)$$

$$s^2 = \frac{1}{N-1} \sum (y_{ij} - \bar{y}_N)^2$$

$$(N-1)s^2 = \sum (y_{ij} - \bar{y}_{Ni} + \bar{y}_{Ni} - \bar{y}_N)^2$$

$$(N-1)s^2 = \sum (y_{ij} - \bar{y}_{Ni})^2 + \sum (\bar{y}_{Ni} - \bar{y}_N)^2 + 2 \sum (y_{ij} - \bar{y}_{Ni})(\bar{y}_{Ni} - \bar{y}_N)$$

we have

$$s_i^2 = \frac{1}{N_i - 1} \sum (y_{ij} - \bar{y}_{Ni})^2$$

$$\Rightarrow \sum (y_{ij} - \bar{y}_{Ni})^2 = (N-1)s^2$$

above eqⁿ becomes

$$(N-1)s^2 = \sum (N_i - 1)s_i^2 + \sum N_i (\bar{y}_{Ni} - \bar{y}_N)^2 + 0$$

if N and N_i is very large
then we can take $(N-1) \approx N$
 $(N_i - 1) \approx N_i$

$$Ns^2 = \sum N_i s_i^2 + \sum N_i (\bar{y}_{Ni} - \bar{y}_N)^2$$

$$s^2 = \frac{\sum N_i s_i^2}{N} + \frac{\sum N_i (\bar{y}_{Ni} - \bar{y}_N)^2}{N}$$

$$s^2 = \sum P_i s_i^2 + \sum P_i (\bar{y}_{Ni} - \bar{y}_N)^2$$

Consider eqⁿ ①

$$\begin{aligned} \text{var}(\bar{y}_N)_{\text{SRS}} &= \left(\frac{1}{n} - \frac{1}{N} \right) s^2 \\ &= \left(\frac{1}{n} - \frac{1}{N} \right) \left[\sum P_i s_i^2 + \sum P_i (\bar{y}_{Ni} - \bar{y}_N)^2 \right] \end{aligned}$$

$$v(\bar{y}_N) = \left(\frac{1}{n} - \frac{1}{N} \right) \sum P_i s_i^2 + \left(\frac{1}{n} - \frac{1}{N} \right) \sum P_i (\bar{y}_{Ni} - \bar{y}_N)^2$$

$$v(\bar{y}_N)_{\text{SRS}} = \text{var}(\bar{y}_{\text{st}})_{\text{prop}} + \text{positive quantity}$$

$$\text{var}(\bar{y}_n)_{\text{SRS}} > \text{var}(\bar{y}_{st})_{\text{prop}}$$

$$\text{consider } \text{var}(\bar{y}_{st})_{\text{prop}} - \text{var}(\bar{y}_{st})_{\text{Ney}}$$

$$\left(\frac{1}{n} - \frac{1}{N} \right) \sum p_i s_i^2 - \left[\frac{1}{n} \left(\sum p_i s_i \right)^2 - \frac{1}{N} \left(\sum p_i s_i \right)^2 \right]$$

$$= \frac{1}{n} \sum p_i s_i^2 - \frac{1}{N} \sum p_i s_i^2 - \frac{1}{n} \left(\sum p_i s_i \right)^2 + \frac{1}{N} \left(\sum p_i s_i \right)^2$$

$$= \frac{1}{n} \left[\sum p_i s_i^2 - \frac{\left(\sum p_i s_i \right)^2}{N} \right] + \text{positive quantity}$$

$$\text{var}(\bar{y}_{st})_{\text{prop}} - \text{var}(\bar{y}_{st})_{\text{Ney}} > 0$$

$$\text{var}(\bar{y}_{st})_{\text{prop}} \geq \text{var}(\bar{y}_{st})_{\text{Ney}} \rightarrow \textcircled{B}$$

$$\therefore \text{var}(\bar{y}_n) \geq \text{var}(\bar{y}_{st})_{\text{prop}} \geq \text{var}(\bar{y}_{st})_{\text{Ney}}$$

③ Write the functions of NSS.

① Collection of data relating to industries and crop estimation

② Collection of data on different socio-economic condition such as literacy birth & death rate, density, sex-ratio etc.

③ It co-ordinate survey to collect data from agriculture sector

④ Conducting an annual survey of industries on a yearly basis.

⑤ Deciding the topic of the particular round of survey.

- ④ Write the functions of CSO
- ① It provides technical leadership in building up the statistical system of country
 - ② It co-ordinates statistical activities of the states & union territories also it ensures timely compilation of requisite data
 - ③ For compilation & release of national account statistics every year
 - ④ It organizes conferences of central & state organizations
 - ⑤ Publication reports on annual survey of industries (ASI)
 - ⑥ Surveys to be taken by in NSSO
 - ⑦ Periodic economic census.

Durakoti
12/08/2024