QP code 65116

SOLUTION – SET I

Q1a 1 – (b), 2 – (a), 3 – (a), 4 – (b), 5 – (a).

b) 1) Symmetric, 2) Discrete 3) k population means
4) Maximum 5) Non-parametric

c) 1 A variable whose range is finite or countably infinite is known as a discrete random variable
Ex - Number of children in a family
2 Mean Median & Mode coincide at the centre of the density curve.
3. S.S = {HH HT TH TT}
4 - F distribution is positively skewed.
5. α = Type I error = Reject H_0 | H_0 is true.

Q2 a For a r.v X $f(x)$ is called pdf if $f(x) \geq 0$
& $\int_{-\infty}^{\infty} f(x)\,dx = 1$

$F(x)$ = cumulative distribution function = $P(X \leq x)$
$= \int_{-\infty}^{x} f(x)\,dx$.

b $X \sim \chi^2_n$ $f(x) = \dfrac{x^{n/2-1} e^{-x/2}}{2^{n/2-1}\,\Gamma(n/2)}$

If X & Y are independent χ^2 variates with n & p d.f
then $X+Y \sim \chi^2_{(n+p)}$; $E(X) = n$ $V(X) = 2n$

c If $Y \sim N(0,1)$ and $Z \sim \chi^2_n$
then $X = \dfrac{Y}{\sqrt{\frac{Z}{n}}} \sim t_n$

$f(t) = \dfrac{\Gamma\left(\frac{n+1}{2}\right)\left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}}}{\Gamma(n/2)\sqrt{n\pi}}$

1) Variance of t variate is > 1
2) The curve never touches X axis

(2)

e One Sided Hypothesis — Two sided Hypothesis
Parameter lies on ± sides — Parameter lies on both sides
Ex $H_1: \mu > 65$ — $H_1: \mu \neq 65$
L.O.S = α — L.O.S = $\frac{\alpha}{2}$

f $H_0: \mu = 12.5$ $H_1: \mu \neq 12.5$
$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$ Rej H_0 if $|z| > z_{\alpha/2}$
$2.82 > 1.96$
$\Rightarrow$ Reject H_0.

q4a In one way ANOVA we get whether there is statistically significant difference in more than 2 groups but the pairwise differences cannot be studied. This is done by post hoc methods such as Least square difference, Duncan's method, Duanett's method, Tukey's Honest square differences.

b For Non parametric tests - the underlying form of the density function is unknown, they are used for qualitative data or ranked data. Also when the sample size is small NPT are used. These are based on order statistics / ranks - Sign, Run tests & Wilcoxon's Test, Kruskal Wallis are some examples.

c

Mann Whitney U Test	Kruskal Wallis Test
Used for Ordinal data	Used for Ordinal data
Comparing Median in two groups	Comparing Median in more than 2 groups
Equivalent to Unpaired t test	Equivalent to one Way ANOVA

d Chi square test of association are used to test the relationship between categorical Variables. For example Area & support to political parties may be associated. The above testing can be done by the χ^2 test.

e Run test is used to determine for serial randomness whether or not observations occur in a sequence in time or over space. Used More often in geographic studies for testing randomness of observations.

(3)

object or individual. For example, effectiveness of coaching program on students performance, effectiveness of drug on patients record observations of the same person & the difference in the observations d is calculated as $X-Y$ & the respective test statistic $\bar{d}$ & $t = \bar{d} / s_d/\sqrt{n}$ is evaluated.

c | Type I error | Type II error
Reject H_0 | H_0 is True | Accept H_0 | H_1 is true
Denoted by α | Denoted by β.
Person is found guilty even when he is innocent | Person is not found guilty even when he did crime

d | Large Sample Test $H_0 : \mu = \mu_0$

$H_1 : \mu > \mu_0$

Test statistic $= t = \frac{\bar{X} - \mu}{S/\sqrt{n}}$ $n+n-1$

S = sample std dev

Reject H_0 if $t_{cal} > t_{\alpha, n-1}$

e | ANOVA is a technique used to test the equality of two or more population means by examining the variances of samples. It determines whether the differences between samples are due to random error or whether due to treatment effects. Certain assumptions need to be satisfied for ANOVA. It is based on comparing variance between samples to variance within sample. Ex. Average performance of different types of cars.

f | $d = X - Y$ $\bar{d} = 8.5$ $S_d^2 = 2.5$ $t_{cal} = 16.99$.
$t_{tab} = 2.82$ $\therefore$ Reject H_0.

3.4 a | Nonparametric test are used when the underlying form of the density function is not known. Generally used for small samples & when the data is qualitative or ranked in nature. These are based on order statistics & ranks. Sign Test, Run test, Wilcoxon's Test are some examples of NPT.

(9)

b Wilcoxon's signed rank test is a Non parametric test which considers magnitude & direction into consideration. The data are assigned ranks

$W = Min(W+, W-)$

Wcal is compared with the table value of W

$W_{cal} < W_{Tab} \Rightarrow H_0$ is Rejected

c Test statistic for Mann Whitney U Test-

$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$

$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$

R_1 = Sum of ranks for group 1

R_2 = Sum " " for group 2

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

~~With respective means test statistics are evaluated for groups~~

d In one way ANOVA we get if there is statistically significant difference in more than 2 groups but does not give pairwise differences. This is done by post hoc comparisons such as Least sq difference Duncuan's method, Dunnett's method & Tukey's Honest square difference.

e Chisquare test of association are used to test if there is a relationship between categorical variables. Ex - Gender of applicants & their success for selection (are they associated) $\chi^2 = \sum \frac{(O-E)^2}{E}$

f $\chi^2_{cal} = 0.83$ $\chi^2_{Tab} = 3.84$

$\chi^2_{cal} < \chi^2_{Tab} \Rightarrow$ Accept H_0

10 a pmf $P(x)$ is such that $P(x) \geq 0$ and $\sum P(x) = 1$

b One sided Hypothesis is the one where the population parameter lies entirely on one side

ex. $H_0: \mu \geq 65$ or $H_0: \mu < 65$.

For a Two sided Hypothesis the parameter can lie on

(5)

either side $H_0: \mu \neq 65$

c Sign Test is a Non parametric Test used to test if Median of two groups is same or not. The number of +ve & -ve sign of the differences is taken minimum for further calculations is taken. The table value is determined accordingly. H_0 is accepted or rejected.

d Non parametric Test — Parametric Test

Underlying distribution is not known — Underlying distribution is known

Used for small sample size — Used for large samples

Used for qualitative data — Used for quantitative data

e $E(X) = np = 3 \times 0.5 = 1.5$ $V(X) = npq = 3 \times 0.5 \times 0.5 = 0.75$