Computer Oriented Statistical Techniques

Time : 2½ Hrs.] Prelim Question Paper Solution **[Marks : 75**

Q.1 Attempt any THREE of the following : [15]

Q.1(a) Use the following frequency distribution of weekly wages to find the arithmetic mean of wage of employees at P & R company. [5]

Weekly Wage ($)	Number of employees
250.00 – 259.99	8
260.00 – 269.99	10
270.00 – 279.99	16
280.00 – 289.99	14
290.00 – 299.99	10
300.00 – 309.99	5
310.00 – 319.99	2

Ans.: X = class mark or mid value

$u = \frac{X - a}{c}$, where a = 284.995 and c = class width = 10

Class-interval	f	X	u	fu
250.00 – 259.99	8	254.995	–3	–24
260.00 – 269.99	10	264.995	–2	–20
270.00 – 279.99	16	274.995	–1	–16
280.00 – 289.99	14	284.995	0	0
290.00 – 299.99	10	294.995	1	10
300.00 – 309.99	5	304.995	2	10
310.00 – 319.99	2	314.995	3	6
	Σf = N = 65			Σfu = –34

$\bar{u} = \frac{\sum fu}{N} = -0.5231$; $\bar{X} = c\bar{u} + a =$ **279.764**

Q.1(b) Find the median TV viewing time of the following data for the 400 students. [5]

Viewing time (minutes)	Number of students
300 – 399	14
400 – 499	46
500 – 599	58
600 – 699	76
700 – 799	68
800 – 899	62
900 – 999	48
1000 – 1099	22
1100 – 1199	6

Ans.: N = Σf = 400 ; N/2 = 200

The median class is 700 – 799

$$\text{Median} = L_1 + \left\{\frac{\left(\frac{N}{2} - \left(\sum f\right)_1\right)}{f_{median}}\right\} c = 699.5 + \left[\left(\frac{200 - 194}{68}\right) \times 100\right]$$

= **708.3235**

Q.1(c) The final grades in Mathematics of students of a class are recorded in the following table. Find the mode. **[5]**

Grades in Mathematics	Number of students
50 - 54	1
55 - 59	2
60 - 64	11
65 - 69	10
70 - 74	12
75 - 79	21
80 - 84	6
85 - 89	9
90 - 94	4
95 - 99	4

Ans.: The modal class is 75 – 79

$$\text{Mode} = L_1 + \left\{\left(\frac{\Delta_1}{\Delta_1 + \Delta_2}\right) c\right\} = 74.5 + \left[\left(\frac{9}{9+15}\right)\times 5\right] = \mathbf{76.375}$$

Q.1(d) Cities A, B and C are equidistant from each other. A motorist travels from A to B at 30 mph, from B to C at 40 mph and from C to A at 50 mph. Determine his average speed. **[5]**

Ans.: $\frac{1}{H} = \frac{1}{N}\sum\left(\frac{1}{X}\right) = \frac{1}{3}\left(\frac{1}{30}+\frac{1}{40}+\frac{1}{50}\right) = \frac{47}{1800}$

$\therefore$ H = Harmonic Mean = $\frac{1800}{47}$ = **38.2979**

Q.1(e) Find the 10 – 90 percentile range of the height of students from the following data. **[5]**

Height (inches)	Number of students
60 - 62	5
63 - 65	18
66 - 68	42
69 - 71	27
72 - 74	8

Ans.: N = 100, 10N/100 = 10, 90N/100 = 90

Class for 10th Percentile is 63 – 65

$$P_{10} = L_1 + \left\{\frac{c}{f}\left(\frac{10N}{100} - \text{c.f.}\right)\right\} = 62.5 + \left\{\frac{3}{18}(10-5)\right\} = \mathbf{63.3333}$$

Class for 90th Percentile is 69 – 71

$$P_{90} = L_1 + \left\{\frac{c}{f}\left(\frac{10N}{100} - \text{c.f.}\right)\right\} = 68.5 + \left\{\frac{3}{27}(90-65)\right\} = \mathbf{71.2778}$$

10 – 90 percentile range = $P_{90} - P_{10}$ = **7.9445**

Q.1(f) Find the standard deviation for the following data of daily wages of workers. **[5]**

Daily wages (in Rs.)	number of workers
200 – 220	24
220 – 240	32
240 – 260	40
260 – 280	17
280 – 300	7

Ans.: X = mid value or class mark

$u = \frac{X-a}{c}$, where a = 67 and c = class width = 3

Class-interval	f	X	u	fu	fu^2
60 – 62	5	61	–2	–10	20
63 – 65	18	64	–1	–18	18
66 – 68	42	67	0	0	0
69 – 71	27	70	1	27	27
72 – 74	8	73	2	16	32
	$\Sigma f = N = 100$			$\Sigma fu = 15$	$\Sigma fu^2 = 97$

$\bar{u} = \frac{\sum fu}{N} = 0.15$

Standard Deviation = $c\sqrt{\left(\frac{\sum f u^2}{N}\right) - \bar{u}^2}$ = **2.9202**

Q.2 Attempt any THREE of the following: [15]

Q.2(a) A population consists of three numbers 3, 7, and 15. Consider all possible samples of size 2 that can be drawn with replacement from this population. Show that average of the sample mean is the population mean. [5]

Ans.: μ = population mean = (3 + 7 + 17)/3 = 27/3 = **9**

Possible Samples	Sample Mean ($\bar{X}$)
(3, 3)	3
(3, 7)	5
(3, 17)	10
(7, 3)	5
(7, 7)	7
(7, 17)	12
(17, 3)	10
(17, 7)	12
(17, 17)	17
Total	**81**

Average of $\bar{X}$ values = $\frac{81}{9}$ = **9**

Q.2(b) Find the first four moments about mean for the following data. [5]

X	12	14	16	18	20	22
f	1	4	6	10	7	2

Ans.: $m_r' = \frac{\sum f(X - A)^r}{N}$; put A = 0 $\therefore m_r' = \frac{\sum f X^r}{N}$

m_1' = 17.6, m_2' = 315.7333, m_3' = 5763.2 and m_4' = 106862.9333

$m_1 = 0$

$m_2 = m_2' - (m_1')^2$ = **5.9733**

$m_3 = m_3' - 3\, m_2'm_1' + 2\,(m_1')^3$ = **–3.9662**

$m_4 = m_4' - 4\, m_3'm_1' + 6\, m_2'(m_1')^2 - 3\,(m_1')^4$ = **89.1625**

Q.2(c) For a data Q_1 = 268.25, Q_3 = 290.75, P_{10} = 258.125, P_{90} = 301 and median = 279. Find (a) Quartile Coefficient of Skewness. (b) 10 – 90 Percentile Coefficient of Skewness. [5]

Ans.: Given: Q_1 = 268.25, Q_3 = 290.75, P_{10} = 258.125, P_{90} = 301 & median = 279.

(a) Quartile Coefficient of Skewness = $\frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1}$ = **0.0444**

(b) 10 – 90 Percentile Coefficient of Skewness = $\frac{P_{90} - 2P_{50} + P_{10}}{P_{90} - P_{10}}$ = **0.0262**

Q.2(d) In how many ways 5 boys and 2 girls can be arranged in a row such that the girls occupy the ends. **[5]**

Ans.: N(S) = 7!

Two girls can be arranged in 2! ways and 5 boys in 5! ways.

$\therefore$ N(A) = 2!×5!

$\therefore$ $P(A) = \frac{2!5!}{7!}$ = **0.0476**

Q.2(e) If an unbiased die is rolled once, find the probability distribution of "number on the uppermost face of the die". Also find the mean of the probability distribution. **[5]**

Ans.: Let X = number on the uppermost face of the die.

x	1	2	3	4	5	6
p(x)	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Mean = E(X) = Σ x p(x) = **3.5**

Q.2(f) Find the probability that in 120 tosses of a fair coin proportion of heads is (a) between 40% & 60% (b) more than 5/8 **[5]**

Ans.: Given: N = 120, p = P(H) = 0.5, q = 0.5

$\mu = Np = 60$, $\sigma^2 = Npq = 30$

(a) Prob(proportion of heads is between 40% & 60%)

= P(number of heads between 48 & 72)

= P(47.5 < X < 72.5)

$= P\left(\frac{47.5-60}{\sqrt{30}} < Z < \frac{72.5-60}{\sqrt{30}}\right)$

= P(– 2.28 < Z < 2.28) = 2×0.4887 = **0.9774**

(b) P(proportion of heads is more than 5/8)

= P(number of heads is more than 75)

= P(74.5 < X)

$= P\left(\frac{74.5-60}{\sqrt{30}} < Z\right)$

= P(2.65 < Z)

= 0.5 – 0.4960 = **0.0040**

Q.3 Attempt any THREE of the following: **[15]**

Q.3(a) A random sample of 50 mathematics grades out of a total of 200, showed a mean of 75 and a standard deviation of 10. Find 95% confidence limits for mean of 200 grades. **[5]**

Ans.: Given: $N_P = 200$, N = 50, $\bar{X} = 75$ and s = 10.

$\sigma_{\bar{X}} = \frac{s}{\sqrt{N}} \times \sqrt{\frac{N_p - N}{N_p - 1}}$ = **1.2278**

For 95% confidence limits, table value using Normal distribution = 1.96 95% confidence limits for μ

: $\bar{X} \pm$ (table value) $\sigma_{\bar{X}}$

: 75 ± (1.96×1.2278) i.e. **(72.5936, 77.4064)**

Q.3(b) A sample of 150 light bulbs of brand A showed a mean life time of 1400 hours with a standard deviation of 120 hours. A sample of 200 light bulbs of brand B showed a mean life time of 1200 hours with a standard deviation of 80 hours. Find 98% confidence limits for the difference of mean life time of the populations of brands A & B. **[5]**

Ans.: Given: $N_1 = 150$, $\bar{X}_1 = 1400$, $s_1 = 120$, $N_2 = 200$, $\bar{X}_2 = 1200$, $s_2 = 80$.

$$\sigma_{\bar{X}_1-\bar{X}_2} = \sqrt{\frac{s_1^2}{N_1}+\frac{s_2^2}{N_2}} = \mathbf{11.3137}$$

For 98% confidence limits, table value using Normal distribution = 2.33 Confidence limits for difference of population means ($\mu_1-\mu_2$)

: $(\bar{X}_1-\bar{X}_2) \pm$ (table value) $\sigma_{\bar{X}_1-\bar{X}_2}$

: $(1400-1200) \pm (2.33\times11.3137)$ i.e. **(173.6391, 226.3609)**

Q.3(c) The standard deviation of the lifetimes of a sample of 200 electric bulbs was computed to be 100 hours. Find 95% confidence limits for the standard deviation of all such electric bulbs. [5]

Ans.: Given: N = 200, s = 100

$$\sigma_s = \frac{s}{\sqrt{2N}} = \mathbf{5}$$

For 95% confidence limits, table value using Normal distribution = 1.96 95% confidence limits for σ

: $s \pm$ (table value) σ_s

: $100 \pm (1.96\times5)$ i.e. **(90.2, 109.8)**

Q.3(d) The breaking strength of cables produced by a manufacturer has a mean of 1800 pounds (lb) and a standard deviation of 100 lb. By a new technique in the manufacturing process, it is claimed that the breaking strength can be increased. To test this claim, a sample of 50 cables is tested and it is found that the mean breaking strength is 1850 lb. Can we support the claim at the 0.01 significance level? [5]

Ans.: Given: N = 50 (large sample), $\bar{X} = 1850$, $\sigma = 100$

H_0: $\mu = 1800$ and H_1: $\mu > 1800$ (One-tailed Test)

Level of significance = 0.01. Table value using Normal distribution = 2.33

$$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{N}} = \mathbf{3.5355 > 2.33}$$

$\therefore$ Reject H_0 i.e. there is significant increase in the breaking strengh.

Q.3(e) A pair of dice is tossed 100 times and it is observed that 7 appeared 23 times. Test the hypothesis that the dice are fair (i.e. not loaded) by using a two-tailed test at 5% significance level. [5]

Ans.: H_0: The dice are fair and H_1: The dice are not fair

Under H_0, P(sum = 7) = P{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)} = 6/36 = 1/6

Given: N = 100, P = 0.23

Level of significance = 0.05. Table value using Normal distribution = 1.96

$$Z = \frac{P-p}{\sqrt{\frac{pq}{N}}} = \frac{0.23-\frac{1}{6}}{\sqrt{\frac{(1/6)(5/6)}{100}}} = \mathbf{1.6994 < 1.96}$$

$\therefore$ Accept H_0 i.e. the dice are fair.

Q.3(f) A sample of 100 electric light bulbs produced by manufacturer A showed a mean lifetime of 1190 h with a standard deviation of 90 h. A sample of 75 bulbs produced by manufacturer B showed a mean lifetime of 1230 h with a standard deviation of 120 h. Is there a difference between the mean lifetime of the two brands of bulbs at significance level of 0.01? [5]

Ans.: Given: $N_1 = 100$, $\bar{X}_1 = 1190$, $s_1 = 90$, $N_2 = 75$, $\bar{X}_2 = 1230$, $s_2 = 120$

H_0: $\mu_1 = \mu_2$ and H_1: $\mu_1 \neq \mu_2$ (Two-tailed Test)

Level of significance = 0.01. Table value using Normal distribution = 2.58

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}} = -\ \mathbf{2.420 \text{ which lies between } -2.58\ \&\ +2.58}$$

∴ Accept H_0 i.e. There is no significant difference between the performances of two types of bulbs.

Q.4 Attempt any THREE of the following: [15]

Q.4(a) A sample of 12 measurements of the breaking strength of cotton threads gave a mean of 7.38 g and a standard deviation of 1.24 g. Find 99% confidence limits for the actual breaking strength. [5]

Ans.: Given: N = 12 (small sample), $\bar{X} = 7.38$, s = 1.24

For t-distribution, d.f. = N – 1 = 11, Table value = 2.72

99% confidence limits for μ

$$: \bar{X} \pm \left(2.72 \times \frac{s}{\sqrt{N-1}}\right) \text{ i.e. } 7.38 \pm \left(2.72 \times \frac{1.24}{\sqrt{11}}\right) \text{ i.e. } \mathbf{(6.3631,\ 8.3969)}$$

Q.4(b) The mean life time of electric bulbs produced by a company has in the past been 1120 h. A sample of 8 electric bulbs recently chosen from the supply of newly produced bulbs showed a mean life time of 1070 h with a standard deviation of 125 h. Test the hypothesis that the mean life time has not changed, using significance level of 0.01. [5]

Ans.: Given: N = 8 (small sample), $\bar{X} = 1070$, s = 125

H_0: $\mu = 1120$ and H_1: $\mu \neq 1120$ (Two-tailed Test)

Level of significance = 0.01

For t-distribution, d.f. = N – 1 = 7, Table value = 3.50

$$t = \frac{\bar{X} - \mu}{s / \sqrt{N-1}} = -\ \mathbf{1.0583 \text{ which lies between } -3.50\ \&\ +3.50}$$

∴ Accept H_0 i.e. the mean life time has not changed.

Q.4(c) In the past, the standard deviation of weights of certain 40.0-ounce package filled by a machine was 0.25 ounce (oz). A random sample of 20 packages showed a standard deviation of 0.32 oz. At 0.05 significance level is the apparent increase in variability significant? [5]

Ans.: Given: N = 20, s = 0.32

H_0: $\sigma = 0.25$ and H_1: $\sigma > 0.25$ (One-tailed Test)

For chi-square distribution, d.f. = N – 1 = 19

Table value of chi-square distribution are $\chi^2_{0.95} = 30.1$

$$\chi^2 = \frac{Ns^2}{\sigma^2} = \mathbf{32.768 > 30.1}$$

∴ Reject H_0 i.e. there is increase in variability.

Q.4(d) The number of books borrowed from a public library during a particular week is given in the following table. [5]

Day	Mon	Tue	Wed	Thu	Fri
Number of books borrowed	132	108	120	114	146

Test the hypothesis that the number of books borrowed does not depend on the day of the week, using significance level of 0.01.

Ans.: This is chi-square Test of goodness of fit.

H_0: The number of books borrowed does not depend on the day of the week
Level of significance = 0.01.
For chi-square, d.f. = number of classes – 1 = 5 – 1 = 4
Table value = $\chi^2_{0.99}$ = 13.3
Under H_0: $P_1 = P_2 = ... = P_5 = 1/5$
N = 620 $\therefore$ e = 620/5 = 124

$$\chi^2 = \sum\frac{(o-e)^2}{e} = \mathbf{7.4194 < 13.3}$$

$\therefore$ Accept H_0 i.e. The number of books borrowed does not depend on the day of the week.

Q.4(e) In his experiment with peas, Mendel observed that 315 were round & yellow, 108 were round & green, 101 were wrinkled & yellow and 36 were wrinkled & green. According to his theory, the numbers should be in the proportion 9:3:3:1. Is there any evidence to doubt his theory at 0.05 significance level? **[5]**

Ans.: This is chi-square Test of goodness of fit.
H_0: The numbers are in the proportion 9:3:3:1
Level of significance = 0.05.
For chi-square, d.f. = number of classes – 1 = 4 – 1 = 3
Table value = $\chi^2_{0.95}$ = 7.81
Under H_0: $P_1 = 9/16$, $P_2 = P_3 = 3/16$ and $P_4 = 1/16$
N = 560 $\therefore e_1 = 315$, $e_2 = e_3 = 105$ and $e_4 = 35$

$$\chi^2 = \sum\frac{(o-e)^2}{e} = \mathbf{0.2667 < 7.81}$$

$\therefore$ Accept H_0 i.e. the numbers are in the proportion 9:3:3:1.

Q.4(f) The following table shows the number of students in class A & B, who passed and failed in an examination. Using Yates' correction at 1% significance level, test the hypothesis that there is no difference between class A & class B. **[5]**

	Passed	**Failed**
Class A	**72**	**17**
Class B	**64**	**23**

Ans.: This is chi-square test of independence of attributes.
H_0: *Given attributes are independent*
Level of significance = 0.01
For chi-square distribution, d.f. = 1, Table value = 6.63

	Passed	Failed	Row Totals
Class A	$a_1 = 72$	$a_2 = 17$	$N_A = 89$
Class B	$b_1 = 64$	$b_2 = 23$	$N_B = 87$
Column Totals	$N_1 = 136$	$N_2 = 40$	N = 176

$$\chi^2 = \frac{N\left(|a_1b_2 - a_2b_1| - \frac{N}{2}\right)^2}{N_1N_2N_AN_B} \quad \underline{\mathbf{0.9627 < 6.63}}$$

$\therefore$ Accept H_0 i.e. There is no difference between class *A* & class B.

Q.5 Attempt any THREE of the following: **[15]**

Q.5(a) A temperature of 100 degrees Celsius (^{0}C) corresponds to 212 degrees Fahrenheit (^{0}F), while a temperature of 0^{0}C corresponds to 32^{0}F. Assuming that a linear relationship exists between Celsius and Fahrenheit temperatures, find the equation connecting Celsius and Fahrenheit temperatures. Also find the Fahrenheit value for ^{0}C = 200. **[5]**

Ans.: There is a linear relation between *C* and F.

Let Y = F and X = C

Slope = m = $\frac{Y_2 - Y_1}{X_2 - X_1} = \frac{212-32}{100-0}$ = **1.8**

Equation: $(Y - Y_1) = m(X - X_1)$ i.e. (Y – 32) = 1.8(X – 0)

i.e. Y = 32 + 1.8X or (^{0}F) = 32 + 1.8(^{0}C)

For ^{0}C = 200, ^{0}F = **392**

Q.5(b) The following table gives the values of pressure (P) corresponding to various values of the volume (V). The relation between P & V is PV^{α} = C where α & C are constants. **[5]**

V	54.3	61.8	72.4	88.7	118.6	194.0
P	61.2	49.2	37.6	28.4	19.2	10.1

Find the values of α & C. Write equation connecting P & V.

Ans.: The relation between P & V is PV^{α} = C where α & C are constants.

Taking log to the base e on both the sides we get

$\ln P = \ln C + (-\alpha) \ln V$

Let $Y = \ln P$, $a_0 = \ln C$, $a_1 = -\alpha$ and $X = \ln V$

∴ equation is $Y = a_0 + a_1X$

The values of a_0 & a_1 are obtained from the following equations

$\Sigma Y = a_0N + a_1\Sigma X$

$\Sigma XY = a_0\Sigma X + a_1\Sigma X^2$

This gives

$$a_0 = \frac{(\sum Y)(\sum X^2) - (\sum X)(\sum XY)}{N\sum X^2 - (\sum X)^2} = \mathbf{4.2039}$$

$$a_1 = \frac{N\sum XY - (\sum X)(\sum Y)}{N\sum X^2 - (\sum X)^2} = -\ \mathbf{1.4058}$$

∴ α = **<u>1.4058</u>** and $C = e^{4.2039}$ = **66.9469**

∴ $PV^{1.4058} = 66.9469$

Q.5(c) Fit a least square parabola $Y = a_0 + a_1X + a_2X^2$ to the following data. **[5]**

X	0	1	2	3	4	5	6
Y	2.4	2.1	3.2	5.6	9.3	14.6	21.9

Ans.: The least square parabola is $Y = a_0 + a_1X + a_2X^2$

The values of a_0, a_1 and a_2 are obtained by solving the following equations.

$\Sigma Y = Na_0 + a_1\,\Sigma X + a_2\,\Sigma X^2$

$\Sigma XY = a_0\,\Sigma X + a_1\,\Sigma X^2 + a_2\,\Sigma X^3$

$\Sigma X^2Y = a_0\,\Sigma X^2 + a_1\,\Sigma X^3 + a_2\,\Sigma X^4$

Solving them we get $\mathbf{a_0 = 2.5095}$, $\mathbf{a_1 = -\ 1.2}$ **and** $\mathbf{a_2 = 0.7333}$

∴ Equation of parabola is $2.5095 - 1.2X + 0.7333X^2$

Q.5(d) The following table shows the respective heights (in inches) X & Y of a sample of 12 fathers and their oldest sons. **[5]**

X	65	63	67	64	68	62	70	66	68	67	69	71
Y	68	66	68	65	69	66	68	65	71	67	68	70

Find the correlation coefficient between X and Y.

Ans.: $$r = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{N\sum X^2 - (\sum X)^2}\sqrt{N\sum Y^2 - (\sum Y)^2}} = \mathbf{0.7027}$$

Q.5(e) A correlation coefficient based on a sample of size 18 was computed to be 0.32. Can we conclude at 5% significance level that the corresponding population correlation coefficient differs from zero? **[5]**

Ans.: Given: N = 18, r = 0.32

H_0: $\rho = 0$ and H_1: $\rho > 0$ (One-tailed Test)

Level of significance = 0.05

For t-distribution, d.f. = N – 2 = 16, Table value = 1.75

$$t = \frac{r\sqrt{N-2}}{\sqrt{1-r^2}} = \mathbf{1.3510 < 1.75}$$

$\therefore$ Accept H_0 i.e. population correlation coefficient is zero.

Q.5(f) 12 pairs of (X, Y) values gave the regression equation of Y on X as Y = 35.82 + 0.476 X. The standard error of estimate is 1.28 and the standard deviation of X is 2.66. At the 0.05 significance level, test the null hypothesis that the population regression coefficient of the regression equation is 0.18 versus the alternative hypothesis that it exceeds 0.18. **[5]**

Ans.: Given: N = 12, $a_1 = 0.476$, $S_{Y.X} = 1.28$, $S_X = 2.66$

H_0: $A_1 = 0.18$ and H_1: $A_1 > 0.18$ (One-tailed Test)

Level of significance = 0.05

For t-distribution, d.f. = N – 2 = 10, Table value = 1.81

$$t = \frac{(a_1 - A_1)\sqrt{N-2}}{S_{Y.X}/S_X} = \mathbf{1.9336 > 1.81}$$

$\therefore$ Reject H_0 i.e. the population regression coefficient of the regression equation exceeds 0.18.

❑ ❑ ❑ ❑ ❑