(i)

Q. 1 Suppose 2 machines produce nails which are on average 10 inches long. A sample of 11 nails is selected from each machine.

Machine A : 6, 8, 8, 10, 10, 10, 10, 10, 12, 12, 14

Machine B : 6, 6, 6, 8, 8, 10, 12, 12, 14, 14, 14

Which machine is better than the other?

A.

| A(x) | B(y) | $x-\bar{x}$ | $(x-\bar{x})^2$ | $y-\bar{y}$ | $(y-\bar{y})^2$ |
|---|---|---|---|---|---|
| 6 | 6 | -4 | 16 | -4 | 16 |
| 8 | 6 | -2 | 4 | -4 | 16 |
| 8 | 6 | -2 | 4 | -4 | 16 |
| 10 | 8 | 0 | 0 | -2 | 4 |
| 10 | 8 | 0 | 0 | -2 | 4 |
| 10 | 10 | 0 | 0 | 0 | 0 |
| 10 | 12 | 0 | 0 | 2 | 4 |
| 10 | 12 | 0 | 0 | 2 | 4 |
| 12 | 14 | 2 | 4 | 4 | 16 |
| 12 | 14 | 2 | 4 | 4 | 16 |
| 14 | 14 | 4 | 16 | 4 | 16 |
| 110 | 110 | 0 | 48 | 0 | 112 |

Mean of Machine :-

$(\bar{x})$ A = 10

$(\bar{y})$ B = 10

$$S_1^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2 = \frac{1}{10}\times 48 = 4.8$$

$$S_2^2 = 11.2$$

$$S_1 = \sqrt{4.8} = 2.19$$

$$S_2 = \sqrt{11.2} = 3.34$$

$\therefore$ $S_2 > S_1$ i.e Standard deviation of Machine A is less than that of machine B and their means are equal. So machine A is better than machine B.

Q.1 (ii) In a study on effectiveness of a medine over a group of patients, the following results were obtained:

| | | | | | |
|---|---|---|---|---|---|
| Percentage of relief | 0-20 | 20-40 | 40-60 | 60-80 | 80-100 |
| No. of patients | 10 | 10 | 25 | 15 | 40 |

Find the mean, median, quartile deviation and standard deviation.

A.

| Class | $x$ | $f$ | $xf$ | c.f | $x^2$ | $fx^2$ |
|---|---|---|---|---|---|---|
| 0-20 | 10 | 10 | 100 | 10 | 100 | 1000 |
| 20-40 | 30 | 10 | 300 | 20 | 900 | 9000 |
| 40-60 | 50 | 25 | 1250 | 45 | 2500 | 62500 |
| 60-80 | 70 | 15 | 1050 | 60 | 4900 | 73500 |
| 80-100 | 90 | 40 | 3600 | 100 | 8100 | 324000 |
| | | 100 | 6300 | | | 470,000 |

(i) Mean $= \frac{\Sigma xf}{\Sigma f} = \frac{6300}{100} = 63$

(ii) Median $= l + \left(\frac{\frac{n}{2} - F}{f_m}\right) i$

Median Class $= 60-80$ $\because \frac{n}{2} = \frac{\Sigma f}{2} = \frac{100}{2} = 50$

$l = 60$ ; $\frac{n}{2} = 50$ ; $F = 45$ ; $f_m = 15$ ; $i = 20$

$\therefore$ Median $= 60 + \frac{50-45}{15} \times 20$

$= 66.67$

(iii) Quartile Deviation $= \frac{Q_3 - Q_1}{2} = \frac{87.5 - 44}{2} = 21.75$

$Q_1 = l + \left(\frac{\frac{n}{4} - F}{f_m}\right) \times i = 40 + \frac{25-20}{25} \times 20 = 44$

$Q_3 = l + \left(\frac{\frac{3n}{4} - F}{f_m}\right) \times i = 80 + \frac{75-60}{40} \times 20 = 87.5$

(iv) Standard Deviation $= \sqrt{\dfrac{\Sigma f x^2}{\Sigma f} - \left(\dfrac{\Sigma f x_i}{\Sigma f}\right)^2}$

$$= \sqrt{\frac{470000}{100} - \left(\frac{6300}{100}\right)^2}$$

$$= \sqrt{731}$$

$$= 27.037$$

Q.2(i) The two random variables X and Y have the joint probability density $f(x,y) = \begin{cases} 6(1-x-y) & x,y>0;\ x+y<1 \\ 0 & \text{otherwise} \end{cases}$

Find the marginal distribution of X and Y. Test whether X and Y are independent.

A. (i) Marginal distribution of X and Y :-

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,.\,dy\,dx = \int_0^1\int_0^1 6(1-x-y)\,dy\,dx$$

$$= \int_0^1 \left[6y - 6xy - 3y^2\right]_0^1 dx$$

$$= \int_0^1 (3-6x).dx$$

$$= \left[3x - 3x^2\right]_0^1$$

$$= 0$$

A.

$$f(x) = \int_0^1 6(1-x-y)\,dy = \left[6y - 6xy - 3y^2\right]_0^1 = 3-6x$$

$$f(y) = \int_0^1 6(1-x-y)\,dx = \left[6x - 3x^2 - 6xy\right]_0^1 = 3-6y$$

(ii) $f(x)\cdot f(y) = (3-6x)(3-6y)$

$$f(x)\,f(y) \neq f(x,y)$$

Therefore X and Y are dependent variables

Q.2 (ii) Two RV having joint pdf $f(x,y) = \begin{cases} \frac{xy}{96} & 0 < x < 4;\ 1 < y < 5 \\ 0 & \text{otherwise.} \end{cases}$

(i) Find Marginal densities

(ii) Find $E[X]$, $E[Y]$, $E[XY]$, $E[2X+3Y]$, $V(X)$, $V(Y)$, $Cov(X,Y)$, $r(X,Y)$

A. (i) Marginal densities of X & Y:

$$f(x) = \int_1^5 \frac{xy}{96}\cdot dy = \left[\frac{xy^2}{192}\right]_1^5 = \frac{x}{8}$$

$$f(y) = \int_0^4 \frac{xy}{96}\cdot dx = \left[\frac{x^2y}{192}\right]_0^4 = \frac{y}{12}$$

(ii) $$E[X] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\cdot f(x,y)\cdot dx\,dy = \int_0^4\int_1^5 x\cdot\frac{xy}{96}\,dx\,dy = \frac{8}{3}$$

$$E[Y] = \int_0^4\int_1^5 y\cdot\frac{xy}{96}\,dx\,dy = \frac{31}{9}$$

$$E[XY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\,f(x,y)\,dx\,dy = \int_0^4\int_1^5 xy\cdot\frac{xy}{96}\,dx\,dy = \frac{248}{27}$$

$$E[2X+3Y] = 2E[X] + 3E[Y] = 2\cdot\frac{8}{3} + 3\cdot\frac{31}{9} = \frac{47}{3}$$

$$V(X) = E[X^2] - (E[X])^2 = \int_1^5\int_0^4 x^2\,\frac{xy}{96}\,dx\,dy - \left(\frac{8}{3}\right)^2 = 8 - \left(\frac{8}{3}\right)^2 = \frac{8}{9}$$

$$V(Y) = E[Y^2] - (E[Y])^2 = \int_1^5\int_0^4 y^2\,\frac{xy}{96}\,dx\,dy - \left(\frac{31}{9}\right)^2 = 13 - \left(\frac{31}{9}\right)^2 = \frac{92}{81}$$

$$Cov(X,Y) = E[XY] - E[X]E[Y] = \frac{248}{27} - \frac{8}{3}\cdot\frac{31}{9} = 0.$$

$r(X,Y) = 0$ $\quad\because\quad r(X,Y) = \frac{Cov(X,Y)}{\sigma_X\cdot\sigma_Y} = 0$

Q.3 From the following data calculate the coefficient of rank correlation between X and Y

X: 36 56 20 65 42 33 44 50 15 60

Y: 50 35 70 25 58 75 60 45 80 38

A:

| X | Y | $R_x$ | $R_y$ | $d_i = R_x - R_y$ | $d_i^2$ |
|---|---|---|---|---|---|
| 36 | 50 | 7 | 6 | 1 | 1 |
| 56 | 35 | 3 | 9 | −6 | 36 |
| 20 | 70 | 9 | 3 | 6 | 36 |
| 65 | 25 | 1 | 10 | −9 | 81 |
| 42 | 58 | 6 | 5 | 1 | 1 |
| 33 | 75 | 8 | 2 | 6 | 36 |
| 44 | 60 | 5 | 4 | 1 | 1 |
| 50 | 45 | 4 | 7 | −3 | 9 |
| 15 | 80 | 10 | 1 | 9 | 81 |
| 60 | 38 | 2 | 8 | −6 | 36 |
| | | | | | 318 |

Coefficient of rank correlation $(r) = 1 - \dfrac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$

$\sum d_i^2 = 318$

$n = 10$

$\therefore \; r = 1 - \dfrac{6 \times 318}{10 \times 99} = -0.927$

Q.4 (i) The heights of ten men taken at random from a normal population were 163, 164, 165, 167, 168, 169, 169, 170, 172, 171 cms respectively. Do the data support the hypothesis of a population mean of 166.1 cms?

| $x$ | $x-\bar{x}$ | $(x-\bar{x})^2$ |
|---|---|---|
| 163 | -4.8 | 23.04 |
| 164 | -3.8 | 14.44 |
| 165 | -2.8 | 7.84 |
| 167 | -0.8 | 0.64 |
| 168 | 0.2 | 0.04 |
| 169 | 1.2 | 1.44 |
| 169 | 1.2 | 1.44 |
| 170 | 2.2 | 4.84 |
| 172 | 4.2 | 17.64 |
| 171 | 3.2 | 10.24 |
| 1678 | | 81.6 |

$$\text{Mean} = \bar{x} = \frac{1678}{10} = 167.8$$

$$s^2 = \frac{1}{n-1} \Sigma (x_i - \bar{x})^2 = \frac{1}{9} \times 81.6$$

$$s^2 = 9.067$$

$$s = 3.011$$

Null Hypothesis :- $H_0: \mu = 166.1$ cms

Alternative Hypothesis :- $H_1: \mu \neq 166.1$ cms

Test statistic : $t = \dfrac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \dfrac{167.8 - 166.1}{\frac{3.011}{\sqrt{10}}} = \cancel{16.186}$

$|t| = 1.785$

Tabulated value of $t$ for $(10-1)$ d.f at 5% LOS is 2.26

i.e $|t| < t_{0.05}$

$\Rightarrow$ Ho accepted

i.e The data support the assumption of mean height of 166.1 cms in the men.

Q.4 (ii) 15.5% of random sample of 1600 undergraduates were smokers whereas 20% of a random sample of 900 graduates were smokers in a state. Can we conclude that less number of undergraduates are smokers than the post graduates at 1% LOS?

$P_1 = \frac{15.5}{100} = 0.155$ ; $n_1 = 1600$

$P_2 = \frac{20}{100} = 0.2$ ; $n_2 = 900$

$p = \dfrac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = 0.1712$ ; $q = 1 - p = 0.8288$

Test statistic :- $z = \dfrac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = -2.867$

$|z| = 2.867$

Null Hypothesis :- $H_0 : p_1 = p_2$

Alternative Hypothesis :- $H_1 : p_1 < p_2$ (one-tailed test)

Tabulated value of 'Z' for 1% LOS = 2.33

$|z| > |Z_{0.01}|$

$\Rightarrow$ $H_0$ rejected

Hence, we can conclude that less number of undergraduates are smokers than the post graduates at 1% LOS.

Q.5 (i) A random sample is selected from each of 3 makes of ropes (Type 1, Type 2 and Type 3) and their breaking strength (in certain units) are measured with the results in the following table. Test whether the breaking strengths of the ropes differ significantly at 5% level of significance.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Type 1 : | 70 | 72 | 75 | 80 | 83 | | |
| Type 2 : | 60 | 65 | 57 | 84 | 87 | 73 | |
| Type 3 : | 100 | 110 | 108 | 112 | 113 | 120 | 107 |

[n = 3 ; N = 18]

| Sources of variation | Degrees of Freedom | Sum of squares | Mean Square | F-ratio |
|---|---|---|---|---|
| Between ropes | $n-1$ | $Q_1$ | $M_1 = \dfrac{Q_1}{n-1}$ | $F = \dfrac{M_1}{M_2}$ $(M_1 > M_2)$ |
| Error | $N-n$ | $Q_2$ | $M_2 = \dfrac{Q_2}{N-n}$ | (or) $\dfrac{M_2}{M_1}$ $(M_2 > M_1)$ |
| Total | $N-1$ | $Q$ | | |

$H_0 : \mu_1 = \mu_2 = \mu_3$ (i.e ropes do not differ)

| | | | | | | | | Total |
|---|---|---|---|---|---|---|---|---|
| I (−80) : | −10 | −8 | −5 | 0 | 3 | | | $T_1 = -20$ |
| II (−80) : | −20 | −15 | −23 | 4 | 7 | −7 | | $T_2 = -54$ |
| III (−80) : | 20 | 30 | 28 | 32 | 33 | 40 | 27 | $T_3 = 210$ |

Sum of total observations $T = T_1 + T_2 + T_3 = 136$

Total number of observations $N = 18$

Correction Factor $\frac{T^2}{N} = \frac{(136)^2}{18} = 1027.56$

(Q) TSS = $[(-10)^2 + (-8)^2 + (-5)^2 + (0)^2 + (3)^2 + \dots (27)^2] - 1027.56$

$= 7992 - 1027.56 = 6964.44$

$Q_1 = \frac{\Sigma T_i^2}{n_i} - \frac{\Sigma T_i}{N} = \frac{(20)^2}{5} + \frac{(54)^2}{6} + \frac{(210)^2}{7} - 1027.56$

$= 5838.44$

$Q_2 = Q - Q_1 = 6964.44 - 5838.44 = 1126$

ANOVA TABLE

| Source of variation | Degrees of Freedom | Sum of Squares | Mean square | F-ratio |
|---|---|---|---|---|
| Between ropes | 2 | 5838.44 | 2999.22 | F = 38.89 |
| Error | 15 | 1126 | 75.07 | |
| Total | 17 | 6964.44 | | |

The table value for $F_{(2,15)}$ at 5% level of significance is 3.68

$F_{calc}$ (38.89) > $F_{0.05}$

$\Rightarrow$ $H_0$ rejected

Hence, the breaking strengths of the ropes differ significantly

Q.5 (ii) Four doctors each tests four treatments for a certain disease and observe the number of days each patient takes to recover. The results are as follows (recovery time in days)

| | Treatments | | | |
|---|---|---|---|---|
| Doctors | 1 | 2 | 3 | 4 |
| A | 10 | 14 | 19 | 20 |
| B | 11 | 15 | 17 | 21 |
| C | 9 | 12 | 16 | 19 |
| D | 8 | 13 | 17 | 20 |

Discuss the difference between (i) doctors and (ii) treatments.

A: $H_0$: (i) No significant difference between doctors
(ii) No significant difference between treatments

$H_1$: There is a significant difference between doctors and treatments

[Subtracting 15 from all elements]

| | Treatments | | | | | |
|---|---|---|---|---|---|---|
| Doctors | 1 | 2 | 3 | 4 | Total | |
| A | −5 | −1 | 4 | 5 | 3 | $(r_1)$ |
| B | −4 | 0 | 2 | 6 | 4 | $(r_2)$ |
| C | −6 | −3 | 1 | 4 | −4 | $(r_3)$ |
| D | −7 | −2 | 2 | 5 | −2 | $(r_4)$ |
| Total | −22 $(c_1)$ | −6 $(c_2)$ | 9 $(c_3)$ | 20 $(c_4)$ | 1 | |

Sum of total observations $T = 1$

Total number of observations $N = 16$

Correction Factor $= \frac{T^2}{N} = \frac{1}{16} = 0.0625$

$SSC = \frac{C_1^2 + C_2^2 + C_3^2 + C_4^2}{4} = 250.25$

$d.f = C-1 = 4-1 = 3$

$MSC = \frac{SSC}{d.f} = 83.4167$

$SSR = \frac{r_1^2 + r_2^2 + r_3^2 + r_4^2}{4} = 11.25$

$d.f = r-1 = 4-1 = 3$

$MSR = \frac{SSR}{d.f} = 3.75$

$TSS = [(-5)^2 + (-1)^2 + (4)^2 + (5)^2 + \cdots + (2)^2 + (5)^2] - 0.0625$

$= 266.9375$

$SSE = TSS - (SSC + SSR) = 5.4375$

$d.f = (c-1)(r-1) = 9$

$MSE = \frac{SSE}{d.f} = 0.604167$

$F_C = \frac{MSC}{MSE} = 138.069$

$F_R = \frac{MSR}{MSE} = 6.2069$

$F_{0.05}(3, 9) = 3.86$

$F_R > F_{0.05}$ & $F_C >> F_{0.05}$ $\therefore$ Ho rejected in both (i) & (ii)

Difference between :-

(i) doctors :- Significant

(ii) treatments :- Highly significant

| Source of Variation | Sum of Squares | Degrees of Freedom | Mean Squares | F-ratio |
|---|---|---|---|---|
| Column Treatments | 250·25 | 3 | 83·4167 | 138·069 |
| Row Treatments | 11·25 | 3 | 3·75 | 6·2069 |
| Error (or) Residual | 5·4375 | 9 | 0·604167 | |