Assignment 1 Questions

Q1) For a given data set calculate less than more than and relative frequency.

) For a given data set calculate less than more than and relative frequency.

| Marks | No. of Students |
|---|---|
| 0–10 | 2 |
| 10–20 | 4 |
| 20–30 | 8 |
| 30–40 | 14 |
| 40–50 | 20 |
| 50–60 | 15 |
| 60–70 | 8 |
| 70–80 | 5 |
| 80–90 | 4 |

Q2) Draw a histogram for the following data:

| Daily Wages (in Rs.) | Number of Workers |
|---|---|
| 0-50 | 8 |
| 50-100 | 16 |
| 100-150 | 27 |
| 150-200 | 19 |
| 200-250 | 10 |
| 250-300 | 6 |

Q3) The marks of 30 students in statistics are given below:
10, 12, 25, 32, 27, 32, 38, 43, 39, 55, 29, 38, 57, 08, 06, 13, 27, 25, 29, 53, 55, 45, 35, 48, 47, 59, 15, 19, 48, 55
Classify the above data by taking a suitable class interval.

Q4) Present the data in a table form.

In a sample study about coffee habits in two towns A and B, the following information given.
Town A: Females were 40%, total coffee drinkers were 45% and female non-coffee drinkers were 20%.
Town B: Males were 55%, male non-coffee drinkers were 30% and female coffee drinkers were 15%.

Q5) 30 pairs of values of two variable X and Y are give below. Form a two-way table
Take class intervals of X as 10 to20, 20 to 30 etc. And Y as 100 to 200, 200 to 300 etc.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| x | 14 | 20 | 33 | 25 | 41 | 18 | 24 | 29 | 38 | 45 |
| y | 148 | 242 | 296 | 312 | 518 | 196 | 214 | 340 | 492 | 568 |
| x | 23 | 32 | 37 | 19 | 28 | 34 | 38 | 29 | 44 | 40 |
| y | 282 | 400 | 288 | 292 | 431 | 440 | 500 | 512 | 415 | 514 |
| x | 22 | 39 | 43 | 44 | 12 | 27 | 39 | 38 | 17 | 26 |
| y | 282 | 481 | 516 | 598 | 122 | 200 | 451 | 387 | 245 | 413 |

Name: Sukant Thombare
Roll NO: 21102A0037 (PIPN A

Q1

$$i = \frac{L - S}{1 + 3.222 \log_{10} n}$$

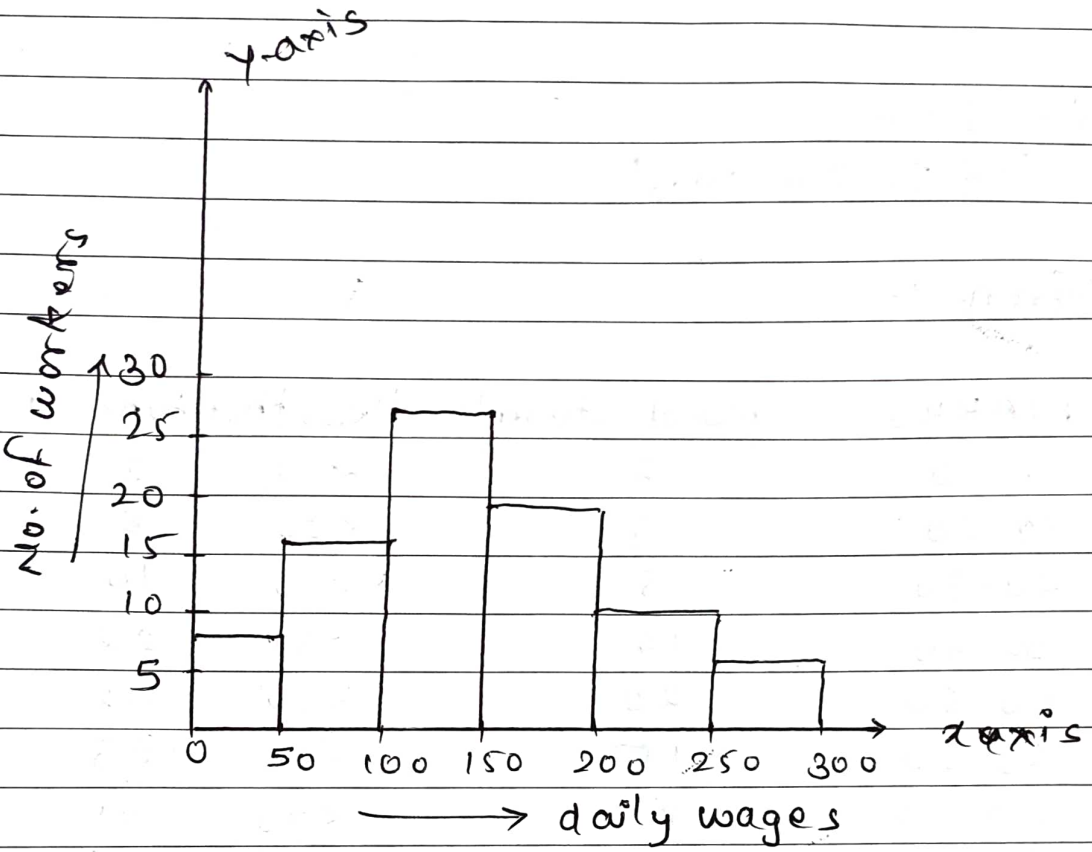
Given :-

| MARKS | No. of students | less than type | | more than type | |
|---|---|---|---|---|---|
| 0-10 | 2 | < 10 | 2 | ≥ 0 | 80 |
| 10-20 | 4 | < 20 | 6 | ≥ 10 | 73 |
| 20-30 | 8 | < 30 | 14 | ≥ 20 | 74 |
| 30-40 | 14 | < 40 | 28 | ≥ 30 | 66 |
| 40-50 | 20 | < 50 | 48 | ≥ 40 | 52 |
| 50-60 | 15 | < 60 | 63 | ≥ 50 | 32 |
| 60-70 | 8 | < 70 | 71 | ≥ 60 | 17 |
| 70-80 | 5 | < 80 | 76 | ≥ 70 | 9 |
| 80-90 | 4 | < 90 | 80 | ≥ 80 | 4 |

relative.

2.5%.
5%.
10%.
17.5%.
25%.
18.75%.
10%.
6.25%.
5%.

Q.2




Q.3 $i = \frac{L - S}{1 + 3.222 \log_{10} n}$

$i = \frac{59 - 06}{1 + 3.222 \log_{10} 30} = 9.20 \approx 9$

| Classes | Tally marks | No. of students |
|---|---|---|
| 6 - 15 | 𝍸 | 5 |
| 15 - 24 | II | 2 |
| 24 - 33 | 𝍸 III | 8 |
| 33 - 42 | IIII | 4 |
| 42 - 51 | 𝍸 | 5 |
| 51 - 60 | 𝍸 I | 6 |
| | | 30 |

Q4]

| | Town A | | | Town B | | |
|---|---|---|---|---|---|---|
| | Men | female | Total | Men | female | Total |
| Coffee | 25 | 20 | 45 | 25 | 15 | 40 |
| | | | | | | 6 |
| Non-coffee | 35 | 20 | 55 | 30 | 30 | 60 |
| Total | 60 | 40 | 100 | 55 | 45 | 100 |

Q.5 For x only

$$i = \frac{L + S}{1 + 3.322 \log_{10} n} = \frac{598 - 12}{1 + 3.22}$$

| Y \ X | 0-20 | 20-30 | 30-40 | 40-50 | ~~50-60~~ |
|---|---|---|---|---|---|
| 100-200 | III | | | | |
| 200-300 | II | IIII | II | | |
| 300-400 | | II | I | | |
| 400-500 | | II | IIII | I | |
| 500-600 | | I | I | IIII | |

## QA Assignment 2

Q1

Following table gives the production of food grain in India in million tonnes.

| | *Production in m. tonnes years* | | | | | |
|---|---|---|---|---|---|---|
| *Food Grains* | 1977–78 | 1978–79 | 1979–80 | 1980–81 | 1981–82 | 1982–83 |
| Rice | 52.7 | 53.8 | 42.3 | 53.2 | 54.0 | 58.0 |
| Wheat | 31.7 | 35.5 | 31.8 | 36.5 | 36.0 | 39.5 |
| Coarse grain | 30.0 | 30.4 | 27.0 | 28.5 | 30.5 | 31.0 |
| Pulses | 12.0 | 12.2 | 8.0 | 11.7 | 11.5 | 13.0 |
| Total | 126.4 | 131.9 | 109.1 | 129.9 | 132.0 | 141.5 |

Q2 **Example:** Find the mean median mode of the following data.

| **Class Interval** | 0 - 10 | 10 - 20 | 20 - 30 | 30 - 40 | 40 - 50 |
|---|---|---|---|---|---|
| **Frequency (fi)** | 9 | 13 | 8 | 15 | 10 |

**Q3 Construct a bivariate frequency distribution table of the marks obtained by students in English (X) and statistics (Y).**

| **Marks in statistics(X)** | 37 | 20 | 46 | 28 | 35 | 26 | 41 | 48 | 32 | 23 | 20 | 39 | 47 | 33 | 27 | 26 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Marks in English(Y) | 30 | 32 | 41 | 33 | 29 | 43 | 30 | 21 | 44 | 38 | 47 | 24 | 32 | 21 | 20 | 21 |

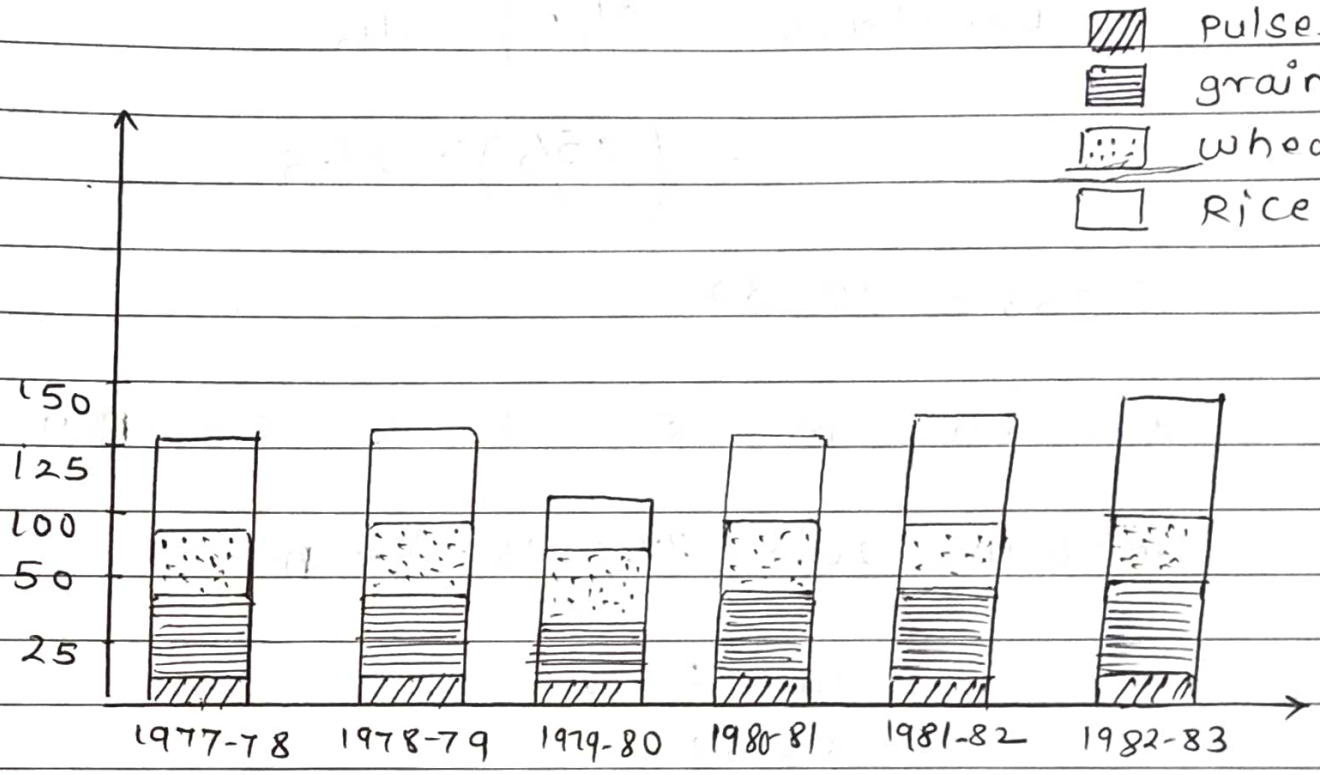
Q4) Explain the following types of tables with examples.
Simple and Complex Tables [Simple Table: One-way table. Complex Table: Two-way table, and manifold tabulation]

Q5) The following numbers represent exam scores in an accounting class: 78, 93, 85, 81, 73, 96, 72, 86, 90, 85 ? draw a stem-and-leaf diagram from this data.

Name: Sukant Thombare
Roll No: 21102A0037 CRPN A

# Assignment 2

Q.1 Interpret the following data using subdivide bar graph.




Q.2

$$\text{mean} = \frac{\Sigma fm}{\Sigma f}$$

| CI | f | m | fm |
|---|---|---|---|
| 0-10 | 9 | 5 | 45 |
| 10-20 | 13 | 15 | 195 |
| 20-30 | 8 | 25 | 200 |
| 30-40 | 15 | 35 | 525 |
| 40-50 | 10 | 45 | 450 |
| Total | 55 | | 1415 |

$$\text{mean} = \frac{1415}{55} = 25.72$$

Q.2 median :

$$\text{median class} = \left(N/2\right)^{th} \text{ value}$$

$$= \left(55/2\right) = 27.5$$

Class = 20 - 30

$l = 20 \quad m = 15 \quad f = 8 \quad c = 11$

$$\text{median} = 20 + \left(\frac{27.5 - 15}{8}\right) \times 11$$

$$= 20 + 15.62$$

$$= 35.62$$

mode :

$$\text{mode} = lb + \left[\frac{D1}{D1 + D2}\right] \times C$$

model class = 30 - 40

$lb = 30, \quad D_1 = 15 - 10 = 5 \quad D_2 = 15 - 8 = 7$

$$\text{model} = 30 + \left[\frac{5}{5+7}\right] \times 10$$

$$= 30 + 0.416$$

$$= 30.416$$

Q.3 for x

$$i = \frac{48-20}{1+3.222\log_{10}16} = \frac{28}{4.879} = 5.738 \approx 6 \quad K=5$$

for y

$$i = \frac{47-20}{1+3.222\log_{10}16} = \frac{27}{4.879} = 5.533 \approx 6$$

| x / y | 20-26 | 26-32 | 32-38 | 38-44 | 44-50 | |
|---|---|---|---|---|---|---|
| 20-26 | | II | I | I | I | 5 |
| 26-32 | | | II | I | | 3 |
| 32-38 | I | I | | | I | 3 |
| 38-44 | I | I | | | | 3 |
| 44-50 | I | | I | | | 2 |
| | 3 | 4 | 4 | 2 | 3 | 16 |

Q.4 simple table : one-way table

In simple table, single characteristic is used to present the data.

It is simplest type of table and is often the data. refered to as first order table or one way table.

They are used to show univariate frequency distribution because they examine only one variable.

e.g Grades & no. of students

| Grade | No. of students |
|---|---|
| A | 10 |
| B | 20 |
| C | 30 |
| D | 5 |

Complex table: Two way table

Complex table displays data in accordance with two or more characteristic of certain phenomenon that are interrelated to each other.

eg

| Sections | No. of student | | Total |
|---|---|---|---|
| | Morning | Afternoon | |
| A | 12 | 28 | 40 |
| B | 17 | 8 | 25 |
| C | 25 | 20 | 45 |
| D | 28 | 12 | 40 |

Manifold Tabulation:

It is a table that explains more that 3 characteristic of data.

These tables offer information on a wide range of phenomenon that are interrelated with each other.

This is most complicated type of table.

Q.5

78, 93, 85, 81, 73, 96, 72, 86, 90, 85

| Stem | leaf |
|---|---|
| 7 | 8, 3, 2 |
| 8 | 5, 1, 6, 5 |
| 9 | 3, 6, 0 |

# Assignment

Q.1 A] Let say patient has a selfesteem score of 76 what would be prediction depression score.

B] Suppose patient has a depression score of 11 what would be predicted self esteem score.

| Depression (x) | 10 | 12 | 19 | 4 | 25 | 15 | 21 | 7 |
|---|---|---|---|---|---|---|---|---|
| selfesteem | 104 | 100 | 98 | 150 | 75 | 105 | 82 | 133 |

$\Rightarrow$

| Depression x selfesteem | Depression$^2$ | selfesteem$^2$ |
|---|---|---|
| 1040 | 100 | 10816 |
| 1200 | 144 | 10000 |
| 1862 | 361 | 9604 |
| 600 | 16 | 22500 |
| 1875 | 625 | 5625 |
| 1722 | 441 | 6724 |
| 931 | 49 | 17689 |
| 1575 | 225 | 11025 |
| 10805 | 1961 | 93983 |

$n = 8$

$$r = \frac{n\Sigma xy - \Sigma x \Sigma y}{\sqrt{[n\Sigma x^2 - (\Sigma x)^2][n\Sigma y^2 - (\Sigma y)^2]}}$$

$$= \frac{8(10805) - 113 \times 847}{\sqrt{[8(1961) - (113)^2][8(93983) - (847)^2}}$$

$$= \frac{86440 - 95711}{\sqrt{(15688 - 12769)(751864 - 717409)}}$$

$$= \frac{-9271}{} = -0.924$$

Mean : $(\bar{x}) = \frac{10+12+19+4+25+15+21+7}{8} = 14.125$

$(\bar{y}) = \frac{104+100+98+150+75+105+82+133}{8}$

$= 105.87$

SD : $x = \sqrt{\frac{\Sigma(x-\bar{x})^2}{(n-1)}} = 7.219$

for $y = \sqrt{\frac{\Sigma(y-\bar{y})^2}{n-1}} = 24.804$

$x' = \left[\frac{r_{xy} \times \sigma_x}{\sigma_y}\right](y-\bar{y}) + \bar{x}$

$= \left[\frac{-0.924 \times 7.219}{24.804}\right](76-105.87) + 14.125$

$= 22.159$

$y' = \left[\frac{r_{xy} \times \sigma_y}{\sigma_x}\right](x-\bar{x})\Big] + \bar{y}$

$= \left[\frac{-0.924 \times 24.804}{7.219}\right](11-14.125)\Big] + 105.87$

$= \left[(-3.17480) - (3.125)\right] + 105.87$

$y' = 115.797$

# Assignment : 2

Q.1 Estimate the yield when rainfall is 29 cm and the rainfall when the yield is 600 kg

| | Yield in kg (x) | Rainfall in cm (x) |
|---|---|---|
| mean | 508 | 26.7 |
| SD | 36.8 | 4.6 |
| r | 0.52 | |

find the a] Regression line y on x

$y - \bar{y} = b_{yx}(x - \bar{x})$ $\quad b_{yx} = r \times \frac{\sigma_y}{\sigma_x}$

b] Regression line x on y

$x - \bar{x} = b_{xy}(y - \bar{y})$ $\quad b_{xy} = r \times \frac{\sigma_x}{\sigma_y}$

⟹ Regression line y on x

$$y - 508 = 0.52 \times \frac{36.8}{4.6}(29 - 26.7)$$

$$y - 508 = \frac{104}{25}(23) \qquad y - 508 = 0.52 \times \frac{36.8}{4.6}(x - 26.7$$

$$y = 9.568 + 508$$

$$y = 517.568 \text{ kg}$$

$$\boxed{y = 4.16x + 396.92}$$

$$y = 517.568 \text{ kg}$$

Regression line x on y

$$x - 4.6 = 0.52 \times \frac{4.6}{36.8}(y - 508)$$

$$\boxed{x = 0.065y - 6.346}$$

$$x = 32.654 \text{ kg}$$

Q.2 Estimate age of husband when the age estimate age of wife when the husband age

| (x) Age of husband | (y) Age of wife | $xy$ | $x^2$ | $y^2$ |
|---|---|---|---|---|
| 25 | 18 | 450 | 625 | |
| 22 | 15 | 330 | 484 | |
| 28 | 20 | 560 | 784 | |
| 26 | 17 | 442 | 676 | |
| 35 | 22 | 770 | 1225 | |
| 20 | 14 | 280 | | |
| 22 | 16 | 352 | | |
| 40 | 21 | 840 | | |
| 20 | 15 | 300 | | |
| 18 | 14 | 252 | | |
| 256 | 172 | 4576 | 7002 | |

$$r = \frac{10 \times (4576) - (256 \times 172)}{\sqrt{(10 \times (7002) - (256)^2]\,(10 \times 3036 - (172)^2)}}$$

$$= \frac{1728}{\sqrt{(70020 - 65536)(30360 - 29585)}}$$

$$= 0.9263$$

Mean : $(\bar{x}) = 25.6$

$(\bar{y}) = 17.2$

SD $x = 7.058$

SD $y = 2.936$

$$x' = \left[\frac{r_{xy} \times \sigma x}{\sigma y}\right](y - \bar{y}) + \bar{x}$$

$$= \left[\frac{0.9263 \times 7.058}{2.936}\right](19 - 17.2) + 25.6$$

$$x' = [29.359]$$

$$y' = \left[\frac{r_{xy} \times \sigma y}{\sigma x}\right](x - \bar{x}) + \bar{y}$$

$$= \left[\frac{0.9263 \times 2.936}{7.058}\right][30 - 25.6] + 17.2$$

$$y' = 18.895$$

# Assignment 6

For a given dataset calculate the regress sum and developed a multiple linear regression model

| $x_1$ | $x_2$ | $y$ | $x_1^2$ | $x_2^2$ | $x_1y$ | $x_2y$ | $x_1x_2$ |
|---|---|---|---|---|---|---|---|
| 3 | 8 | -3.7 | 9 | 64 | -11.1 | -29.6 | 24 |
| 4 | 5 | 3.5 | 16 | 25 | 14 | 17.5 | 20 |
| 5 | 7 | 2.5 | 25 | 49 | 12.5 | 17.5 | 35 |
| 6 | 3 | 11.5 | 36 | 9 | 69 | 34.5 | 18 |
| 2 | 1 | 5.7 | 4 | 1 | 11.4 | 5.7 | 2 |
| 20 | 24 | 19.5 | 400 | 576 | 390 | 468 | 480 |
| | | | 490 | 724 | 48 | | |

$\Sigma x_1 = 20$

$\Sigma x_2 = 24$

$\Sigma y = 19.5$

$\Sigma x_1^2 = 90$

$\Sigma x_2^2 = 148$

$\Sigma x_1 y = 95.8$

$\Sigma x_2 y = 45.6$

$\Sigma x_1 x_2 = 99$

$$\Sigma x_1^2 = [90 - ((20)^2/5)] = 10$$

$$\Sigma x_2^2 = [148 - ((24)^2/5)] = 32.8$$

$$\Sigma x_1 y = [95.8 - ((20 \times 19.5)/5] = 17.8$$

$$\Sigma x_2 y = [45.6 - ((24 \times 19.5)/5)) = -48$$

$$\Sigma x_1 x_2 = [99 - ((20 \times 24)/5)] = 3$$

$b_0, b_1, b_2$

$$b_1 = \frac{32.8 \times 17.8 - (3 \times -48)}{[10 \times 32.8) - (3)^2} = 2.2816$$

$$b_2 = \frac{(10 \times -48) - (3 \times 17.8)}{(10 \times 32.8) - (3)^2} = -1.6721$$

$$b_0 = \bar{y} - b_1 \bar{x}_1 - b_2 \bar{x}_2$$

$$= 3.9 - (2.2816)(4) - (-1.6721)(4.8)$$

$$b_0 = 2.79968$$

$$\hat{y} = b_0 + b_1 * x_1 + b_2 * x_2$$

$$\hat{y} = 2.7996 + 2.2816 x_1 - 1.6721 x_2$$

# Assignment 8

Q 1

| | $Y$ | $x_1$ | $x_2$ | $x_1^2$ | $x_2^2$ | $x_2 Y$ | $x_1 x_2$ |
|---|---|---|---|---|---|---|---|
| | 2.45 | 84 | 15 | 7056 | 225 | 36.75 | 1260 |
| | 1.77 | 66 | 8 | 4356 | 64 | 13.76 | 528 |
| | 2.37 | 68 | 46 | 4624 | 2116 | 109.02 | 3128 |
| | 2.23 | 65 | 24 | 4725 | 576 | 53.52 | 1560 |
| | 1.92 | 69 | 12 | 4761 | 144 | 23.04 | 828 |
| | 1.99 | 72 | 25 | 5184 | 625 | 49.75 | 1800 |
| | 1.99 | 63 | 45 | 3869 | 2025 | 89.55 | 2835 |
| | 2.35 | 56 | 72 | 3136 | 5184 | 169.2 | 4032 |
| Total | 17.02 | 543 | 247 | 37311 | 10959 | 544. | 15971 |

$$\Sigma x_1^2 = 37311 - (43)^2/8 \qquad \Sigma x_2^2 = 10959 - (247)^2/8$$

$$= 454.875$$

$$\Sigma x_1^2 = 454.875 \qquad \Sigma x_2^2 = 3332.875$$

$$\Sigma x_1 Y = 1158.16 - \frac{543 \times 17.02}{8} \qquad \Sigma x_2 Y = 544.59 - \frac{247 \times 17.02}{8}$$

$$\Sigma x_1 Y = 2.9275 \qquad \Sigma x_2 Y = 19.0975$$

$$\Sigma x_1 x_2 = 15971 - \frac{543 \times 247}{8} = -794.125$$

$$b_1 = \frac{(333.2875 \times 2.9275) - (-794.125) \times 19.0975)}{454.875 \times 3332.875 - (-794.125)^2}$$

$$b_1 = 0.0281$$

$$b_2 = 0.0184$$

$$b_0 = -0.1026$$

$$\hat{y} = -0.1026 + 0.0281 x_1 + 0.0124 x_2$$

| | $\hat{y}$ | $(y-\hat{y})^2$ | $(y-\bar{y})^2$ | $(\hat{y}-\bar{y})^2$ |
|---|---|---|---|---|
| | 2.3838 | 0.0044 | 0.109 | 0.0457 |
| | 1.7912 | 0.0051 | 0.1661 | 0.1131 |
| | 2.3186 | 0.0076 | 0.0588 | 0.0565 |
| | 1.9615 | 0.0721 | 0.0105 | 0.0276 |
| | 1.9251 | 0.0002 | 0.043 | 0.0909 |
| | 2.1706 | 0.0326 | 0.0189 | 0.0018 |
| | 2.1657 | 0.0308 | 0.0189 | 0.0015 |
| | 2.3038 | 0.0021 | 0.0495 | 0.0311 |
| Total | 17.02 | 0.1497 | 0.4697 | 0.3182 |

$R^2 = \frac{0.3182}{0.4697} = 0.6774$

$h_0 \Rightarrow \partial = b_0\ b_1 \& b_2 = 0$

$h_a \Rightarrow b_1 \& b_2 \neq 0$

$df_1 = k = 2$ $\quad df_2 = n-k-1 = 8-2-1 = 5$

$MSR = \frac{0.3182}{2} = 0.1591$

$MSE = \frac{0.1497}{5} = 0.0299$

$f\ test = \frac{MSR}{MSE} = \frac{0.1591}{0.0299} = 5.3139$

$\therefore$ for 2,5 df and $\alpha = 0.05$ f test value is 5.7861

$\therefore$ f test cal $<$ f test table

$h_0$ is accepted

# Multiple Linear Regression

| | y | $x_1$ | $x_2$ | $x_1^2$ | $x_2^2$ | $x_1y$ | $x_2y$ | $x_1x_2$ |
|---|---|---|---|---|---|---|---|---|
| | 140 | 60 | 22 | 3600 | 484 | 8400 | 3080 | 1320 |
| | 155 | 62 | 25 | 3844 | 625 | 9610 | 3875 | 1550 |
| | 159 | 67 | 24 | 4489 | 576 | 10653 | 3816 | 1608 |
| | 179 | 70 | 20 | 4900 | 400 | 12530 | 3580 | 1400 |
| | 192 | 71 | 15 | 5041 | 225 | 13632 | 2880 | 1065 |
| | 200 | 72 | 14 | 5184 | 196 | 14400 | 2800 | 1008 |
| | 212 | 75 | 14 | 5625 | 196 | 15900 | 2968 | 1050 |
| | 215 | 78 | 11 | 6084 | 121 | 16770 | 2365 | 858 |
| Mean | 181.5 | 69.375 | 18.125 | | | | | |
| Sum | 1452 | 555 | 145 | Sum | 38767 | 2823 | 101895 | 25364 | 9859 |
| | | | | Reg Sums | 263.875 | 194.875 | 1162.5 | -953.5 | -200.375 |

**Step 3: Calculate $b_0$, $b_1$, and $b_2$**

- The formula to calculate $b_1$ is: $[(\Sigma x_2^2)(\Sigma x_1 y) - (\Sigma x_1 x_2)(\Sigma x_2 y)] \; / \; [(\Sigma x_1^2)(\Sigma x_2^2) - (\Sigma x_1 x_2)^2]$
- Thus, $b_1 = [(194.875)(1162.5) - (-200.375)(-953.5)] \; / \; [(263.875)(194.875) - (-200.375)^2] =$ **3.148**
- The formula to calculate $b_2$ is: $[(\Sigma x_1^2)(\Sigma x_2 y) - (\Sigma x_1 x_2)(\Sigma x_1 y)] \; / \; [(\Sigma x_1^2)(\Sigma x_2^2) - (\Sigma x_1 x_2)^2]$
- Thus, $b_2 = [(263.875)(-953.5) - (-200.375)(1152.5)] \; / \; [(263.875)(194.875) - (-200.375)^2] =$ **-1.656**

The formula to calculate $b_0$ is: $y - b_1X_1 - b_2X_2$

- Thus, $b_0 = 181.5 - 3.148(69.375) - (-1.656)(18.125) =$ **-6.867**

**Step 5: Place $b_0$, $b_1$, and $b_2$ in the estimated linear regression equation.**

- The estimated linear regression equation is: $\hat{y} = b_0 + b_1 * x_1 + b_2 * x_2$

$$\hat{y} = -6.867 + 3.148x_1 - 1.656x_2$$

We often use three different **sum of squares** values to measure how well the regression line actually fits the data:

**1. Sum of Squares Total (SST) –** The sum of squared differences between individual data points ($y_i$) and the mean of the response variable ($\bar{y}$).

- $SST = \Sigma(y_i - \bar{y})^2$

**2. Sum of Squares Regression (SSR)** – The sum of squared differences between predicted data points ($\hat{y}_i$) and the mean of the response variable($\bar{y}$).

- $SSR = \Sigma(\hat{y}_i - \bar{y})^2$

**3. Sum of Squares Error (SSE)** – The sum of squared differences between predicted data points ($\hat{y}_i$) and observed data points ($y_i$).

- $SSE = \Sigma(\hat{y}_i - y_i)^2$

The following relationship exists between these three measures:

**SST = SSR + SSE**

## R-squared = SSR / SST