Program No :: Part B1

Write R Program to create a vector containing following 8 values and perform the following operations

4 3 0 5 2 9 4 5

a. Find mean, median, mode
b. Find the range
c. Find the 35th and 78th percentile
d. Find the variance and standard deviation
e. Find the interquartile range
f. Find the z-score for each value

x	$x-\mu$	$x-\mu^2$
0	-4	16
2	-2	4
3	-1	1
4	0	0
4	0	0
5	1	1
5	1	1
9	5	25

$\Sigma x = 32$ $\Sigma(x-\mu)^2 = 48$

a. Mean $= \frac{\Sigma x}{N} = \frac{32}{8} = \underline{\underline{4}}$

Median $= \left[\left(n/2\right)^{th} \text{ term} + \frac{(n+1)}{2}^{th} \text{ term} \right] / 2$

$= \frac{4+4}{2} = \frac{8}{2} = \underline{\underline{4}}$

Mode $= \underline{\underline{4, 5}}$

b. Range = Highest value - lowest value

$= 9 - 0$

$= \underline{\underline{9}}$

c. 35 th Percentile $= \frac{P}{100} \times N$

$P_{35} = \frac{35}{100} \times 8$

$= 0.35 \times 8 = 2.8$

$2.8 + 1 = 3.8$ = 3 rd term

$P_{35} = \underline{\underline{3}}$

78 th Percentile $= \frac{P}{100} \times N$

$P_{78} = \frac{78}{100} \times 8$

$= 0.78 \times 8 = 6.24$

$= 6.24 + 1 = 7.24$ = 7 th term

$P_{78} = \underline{\underline{5}}$

d. Variance of Sample :

$$s^2 = \frac{\Sigma(x-\mu)^2}{n-1} = \frac{48}{8-1} = \frac{48}{7} = \underline{\underline{6.8571}}$$

Standard Deviation

$$s = \sqrt{\frac{\Sigma(x-\mu)^2}{n-1}} = \sqrt{\frac{48}{7}} = \underline{\underline{2.6186}}$$

e. Interquartile Range = $Q_3 - Q_1$

$$Q_1 = P_{25} = \frac{25}{100} \times 8 = 0.25 \times 8 = 2$$

$$\frac{2\text{nd term} + 3\text{rd term}}{2} = \frac{2+3}{2} = \frac{5}{2} = 2.5$$

$$Q_1 = 2.5$$

$$Q_3 = P_{75} = \frac{75}{100} \times 8 = 0.75 \times 8 = 6$$

$$\frac{6\text{th term} + 7\text{th term}}{2} = \frac{5+5}{2} = \frac{10}{2} = 5$$

$$Q_3 = \underline{\underline{5}}$$

Interquartile Range = $Q_3 - Q_1 = 5 - 2.5 = \underline{\underline{2.5}}$

f. Z score = $\frac{x-\mu}{\sigma}$

$x = 0$, $\frac{0-4}{2.61} = \frac{-4}{2.61} = \underline{\underline{-1.532}}$

$x = 2$, $2-4/2.61 = -2/2.61 = \underline{\underline{-0.7662}}$

$x = 3$; $\frac{3-4}{2.61} = \frac{-1}{2.61} = \frac{-1}{2.61} = -0.3831$

$x = 4$; $\frac{4-4}{2.61} = \frac{0}{2.61} = 0$

$x = 4$; $\frac{4-4}{2.61} = \frac{0}{2.61} = 0$

$x = 5$; $\frac{5-4}{2.61} = \frac{1}{2.61} = 0.3831$

$x = 5$; $\frac{5-4}{2.61} = \frac{1}{2.61} = 0.3831$

$x = 9$; $\frac{9-4}{2.61} = \frac{5}{2.61} = 1.9157$

Program No : Part B2

Write R script to find the correlation coefficient and type of correlation between advertisement expenses and sales volume using Karl Pearson's coefficient of correlation method (Direct method)

Firm	1	2	3	4	5	6	7	8	9	10
Advertisement Exp (in lakhs)	11	13	14	16	16	15	15	14	13	13
Sales volume (Rs in lakhs)	50	50	55	60	65	65	65	60	60	50

Calculation of Karl Pearson's coefficient of correlation

Firm	x	y	$x = x - \bar{x}$	x^2	$y = y - \bar{y}$	y^2	xy
1	11	50	-3	9	-8	64	24
2	13	50	-1	1	-8	64	8
3	14	55	0	0	-3	9	0
4	16	60	2	4	2	4	4
5	16	65	2	4	7	49	14
6	15	65	1	1	7	49	7
7	15	65	1	1	7	49	7
8	14	60	0	0	2	4	0
9	13	60	-1	1	2	4	-2
10	13	50	-1	1	-8	64	8
	140 Σx	580 Σy		22 Σx^2		360 Σy^2	70 Σxy

$$\bar{x} = \frac{\Sigma x}{n} = \frac{140}{10} = \underline{\underline{14}}$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{580}{10} = \underline{\underline{58}}$$

correlation coefficient $= r = \dfrac{\Sigma xy}{\sqrt{\Sigma x^2 - \Sigma y^2}}$

$$= \frac{70}{\sqrt{22 \times 360}}$$

$$= \frac{70}{\sqrt{7920}}$$

$$= \underline{\underline{0.7865}}$$

There exist a positive correlation of higher degree between advertisement expenses and sales volume

Program No : Part B3

Compute the regression equation of y on x from the following data.

x	y	xy	x^2	y^2
2	18	36	4	324
4	12	48	16	144
5	10	50	25	100
6	8	48	36	64
8	7	56	64	49
11	5	55	121	25
$\Sigma x = 36$	$\Sigma y = 60$	$\Sigma xy = 293$	$\Sigma x^2 = 266$	$\Sigma y^2 = 706$

Y on X

$\Sigma y = Na + b\Sigma x$

$60 = 6a + 36b \rightarrow 1$

$\Sigma xy = a\Sigma x + b\Sigma x^2$

$293 = 36a + 266b \rightarrow 2$

From 1 and 2 $\Rightarrow$ $6a + 36b = 60 \quad \times 36$

$36a + 266b = 293 \quad \times 293$

$\Rightarrow$ $\cancel{216a} + 1296b = 2160$

$\cancel{216a} + 1596b = 1758$

$(-) \quad (-) \quad (-)$

$-300b = 402$

$b = \dfrac{402}{-300}$

$b = \underline{\underline{-1.34}}$

$\bar{x} = \dfrac{\Sigma x}{n} = \dfrac{36}{6} = \underline{\underline{6}}$ $\quad \bar{y} = \dfrac{\Sigma y}{n} = \dfrac{60}{6} = \underline{\underline{10}}$

$$b_{yx} = \frac{n\Sigma xy - \Sigma x \Sigma y}{n\Sigma x^2 - (\Sigma x)^2} = \frac{6 \times 293 - 36 \times 60}{6 \times 266 - (36)^2} = \frac{-402}{300}$$

$$= \underline{\underline{-1.34}}$$

Regression equation of y on x is,

$$(y - \bar{y}) = b_{yx}(x - \bar{x})$$

$$(y - 10) = -1.34(x - 6)$$

$$y = -1.34x + 8.04 + 10$$

$$y = -1.34x + 18.04$$

Program No : Part B4

The time taken by a large group of students to compute a piece of homework, T minutes are normally distributed with a mean of 53 minutes and standard deviation of 65. Find the probability that the time taken by a random student from the group to complete this homework will be less than 60 minutes. Write R script to find the probability that the time taken by a random student from the group to complete this homework

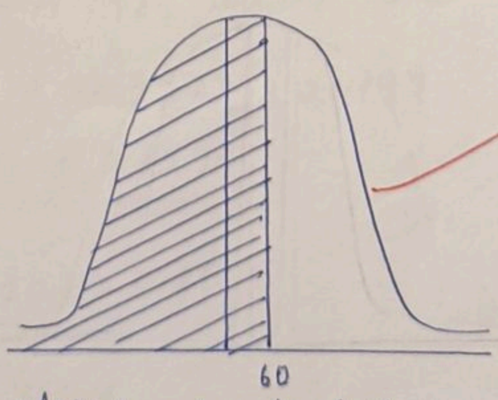
a) will be less than 60 minutes

b) Between 50 and 80 minutes

$\mu = 57 \quad \sigma = 6.5 \quad P(x<60) = ? \quad P(50<x<80) = ?$

$$P(x<60) = z = \frac{x-\mu}{\sigma} = \frac{60-57}{6.5} = \frac{3}{6.5} = 0.4615$$

The probability associated with $Z = 0.4615$ is 0.1772



Adding the above probability value with

0.5 give the solution i.e $p(x<60) = 0.1772 + 0.5$

$= 0.6772$

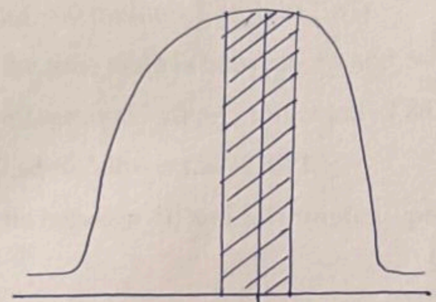
$\therefore p(x<60) = 0.6772$

b. $P(50<x<80) = z = \dfrac{x-\mu}{\sigma} = \dfrac{50-57}{6.5} = -1.0769$

The probability associated with $z = -1.0769$ is 0.3577

$Z = \dfrac{x-\mu}{\sigma} = \dfrac{80-57}{6.5} = 3.5384$

The probability associated with $Z = 3.5384$ is 0.4998



57 80

$\mu = 57$

Adding the above probability values i.e. 0.3577 and 0.4998, give the solution i.e

$p(50<x<80) = 0.3577 + 0.499$

$= 0.8575$

$P(50<x<80) = 0.8575$

Program No : Part B5

Write a script to perform the following using binomial distribution

i. If $n=4$ and $p=0.10$, find $P(x=3)$

ii. If $n=12$ and $p=0.45$, find $P(5<=x<=7)$

i. $n=4$, $p=0.10$, $q=1-p=1-0.10=0.9$, $x=3$

Binomial distribution $p(x) = {}^{n}C_{x} \cdot p^{x} \cdot q^{n-x}$

$$= {}^{4}C_{3} \times (0.10)^{3} \times (0.5)^{4-3}$$

$$= \frac{4!}{1!\,3!} \times (0.10)^{3} \times (0.9)^{1}$$

$$= 4 \times 0.001 \times 0.9 = \underline{0.0036}$$

$$P(x=3) = 0.0036$$

ii. $n=12$ $p=0.45$ $q=0.55$ $x=5,6,7$

Binomial distribution $P(x=5) = {}^{n}C_{x}\, p^{x} \cdot q^{n-x}$

$$= {}^{12}C_{5} \times (0.45)^{5} \times (0.55)^{7}$$

$$= \frac{12!}{7!\,5!} \times (0.45)^{5} \times (0.55)^{7}$$

$$= 792 \times 0.01845 \times 0.015224$$

$$= \underline{0.22245}$$

Binomial distribution $= p(x=6) = {}^{n}C_{x} \cdot p^{x} \cdot q^{n-x}$

$= {}^{12}C_6 \times (0.45)^6 \times (0.55)^6$

$= \dfrac{12!}{6!\,6!} \times (0.45)^6 \times (0.55)^6$

$= 924 \times 0.008303 \times 0.02768$

$= \underline{\underline{0.21236}}$

Binomial distribution $P(x=7) = {}^nC_x \cdot p^x \cdot q^{n-x}$

$= {}^{12}C_7 \times (0.45)^7 \times 0.55)^5$

$= \dfrac{12!}{5!\,7!} \times (0.45)^7 \times (0.55)^5$

$= 792 \times 0.00373 \times 0.05073$

$= \underline{\underline{0.14865}}$

$P(5 <= x <= 7) = P(x=5) + P(x=6) + P(x=7)$

$= 0.22245 + 0.21236 + 0.14865$

$= \underline{\underline{0.58346}}$

$P(5 <= x <= 7) = \underline{\underline{0.58346}}$

Program No :: Part B6

Perform the following using uniform distribution between 200 and 240

i. $P(x > 230)$

ii $P(205 \le x \le 220)$

$a = 200 \quad b = 240$

$$\text{Mean} = \mu = \frac{a+b}{2} = \frac{200+240}{2} = \frac{440}{2} = \underline{\underline{220}}$$

$$\text{Standard Deviation} = \sigma = \frac{b-a}{\sqrt{12}} = \frac{240-200}{\sqrt{12}} = \frac{40}{3.4641}$$

$$= \underline{\underline{11.5470}}$$

$$\text{Height} = f(x) = \frac{1}{(b-a)} = \frac{1}{(240-200)} = \frac{1}{40} = \underline{\underline{0.025}}$$

i. $P(x > 230)$

$x_1 = 230 \quad x_2 = 240$

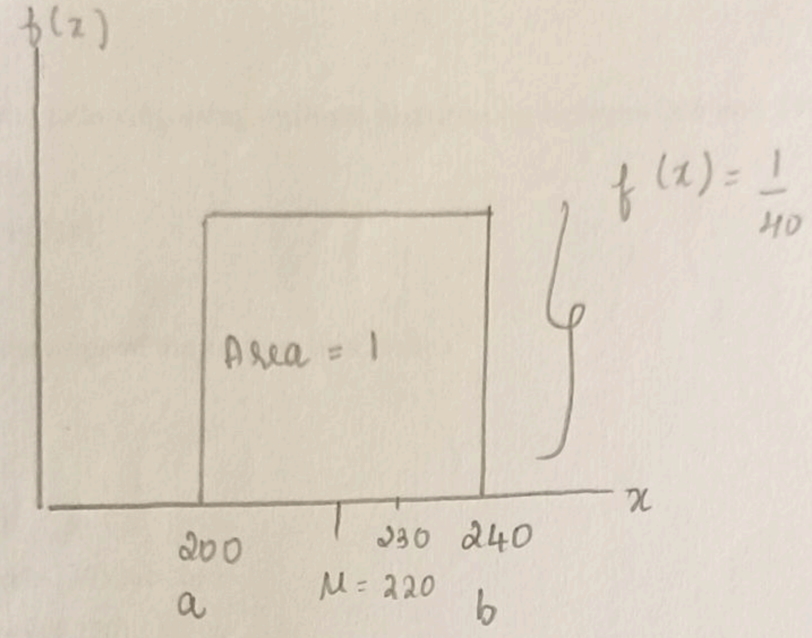
$$P(x > 230) = \frac{x_2 - x_1}{b-a} = \frac{240-230}{240-200} = \frac{10}{40} = \underline{\underline{0.25}}$$

$$P(x > 230) = \underline{\underline{0.25}}$$

ii. $P(205 \le x \le 220)$

$x_1 = 205, \quad x_2 = 220$

$$P(205 \le x \le 220) = \frac{x_2 - x_1}{b-a} = \frac{220-205}{240-200} = \frac{15}{40} = \underline{\underline{0.375}}$$



$\sigma = 11.5470$

Program No : Part B7

Following are the scores of max vertical jumps before and after the training program. Test whether the training program is helpful to the student (Use Paired t-test)

Player	Max Vertical Jump: Before Training program	After Training program	d	d - d'	(d-d')^2
1	22	24	-2	-1.05	1.1025
2	20	22	-2	-1.05	1.1025
3	19	19	0	0.95	0.9025
4	24	22	2	2.95	8.7025
5	25	28	-3	2.05	4.0205
6	25	26	-1	-0.05	0.0025
7	28	28	0	0.95	0.9025
8	22	24	-2	-1.05	1.1025
9	30	30	0	0.95	0.9025
10	27	29	-2	-1.05	1.1025
11	24	25	-1	-0.05	0.0025
12	18	20	-2	-1.05	1.1025
13	16	17	-1	-0.05	0.0025
14	19	18	1	1.95	3.8025
15	19	18	1	1.95	3.8025
16	28	28	0	0.95	0.9025
17	24	26	-2	-1.05	1.1025
18	25	27	-2	-1.05	1.1025
19	25	27	-2	-1.05	1.1025
20	23	24	-1	-0.05	0.025
			= -19		= 32.95

$H_0 : D = 0$ or $(\mu_1 - \mu_2) = 0)$ The two population mean is equal

$H_1 : D < 0$ or $(\mu_1 - \mu_2) < 0$ The two population mean are less than 0

$$d' = \frac{\Sigma d}{n} = \frac{-19}{20} = -0.95$$

$$\sigma^2 = \frac{(d-d')^2}{n-1} = \frac{32.95}{19} = 1.7342$$

$$\sigma = \sqrt{1.7342} = 1.31689$$

$$t \text{ test} = t = \frac{d^2 - D}{\frac{sd}{\sqrt{n}}}$$

$$= \frac{-0.95 - 0}{\frac{1.31689}{4.4721}} = \frac{-0.95}{0.2944}$$

$$= -3.2269$$

VALUED

To Calculate the critical t value $df = n-1 = 20-1 = 19$

Alpha $= \alpha = 0.01$ (99% confidence level)

$$t_{\alpha, n-1} = t_{0.01, 19} = -2.539$$

The observed t value is -3.2269 and critical t value is -2.539, i.e. $-3.2269 < -2.539$

$\therefore$ We reject the Null Hypothesis.

Program No :- Part B8

A company has three manufacturing plants, and company officials want to determine whether there is difference in the average age of workers at the three locations. The following data are the ages of five randomly selected workers at each plant. Perform a one-way ANOVA to determine whether there is a significant difference in the mean ages of the workers at three plants. Use $\alpha = 0.01$. Write R script for the above problem.

Plant (Employee Age)

1	2	3
29	32	25
27	33	24
30	31	24
27	34	25
28	30	25

$T_j : T_1 = 141$ $\quad T_2 = 160$ $\quad T_3 = 123$

$n_j : n_1 = 5$ $\quad n_2 = 5$ $\quad n_3 = 5$

$\bar{x}_j : \bar{x}_1 = 28.2$ $\quad \bar{x}_2 = 32$ $\quad \bar{x}_3 = 24.6$

$T = 141 + 160 + 123 = 424$

$N = 5 + 5 + 5 = 15$

$\bar{x} = 424 / 15 = 28.2666$

$H_0 :: \mu_1 = \mu_2 = \mu_3$

H_1 :: At least one of the mean is different from others

$$SSC = \sum_{j=1}^{c} n_j (\bar{x}_j - \bar{x})^2$$

$$= [5(28.2 - 28.26666)^2 + 5(32 - 28.2666)^2 + 5(24.6 - 28.2666)^2]$$

$$= [5(0.0044) + 5(13.9382) + 5(13.4439)]$$

$$= [0.022 + 69.691 + 67.2195]$$

$$= \underline{\underline{136.9325}}$$

$$SSE = \sum_{i=1}^{n} \sum_{j=1}^{c} (\bar{x}_{ij} - \bar{x}_j)^2$$

$$= (29 - 28.2)^2 + (27 - 28.2)^2 + (30 - 28.2)^2 + (27 - 28.2)^2 + (28 - 28.2)^2 + (32 - 32)^2 + (33 - 32)^2 + (31 - 32)^2 + (34 - 32)^2 + (30 - 32)^2 + (25 - 24.6)^2 + (24 - 24.6)^2 + (24 - 24.6)^2 + (25 - 24.6)^2 + (25 - 24.6)^2$$

$$= 0.64 + 1.44 + 3.24 + 1.44 + 0.04 + 0 + 1 + 1 + 4 + 4 + 0.16 + 0.36 + 0.36 + 0.16 + 0.16$$

$$= \underline{\underline{18}}$$

$$SST = SSC + SSE$$

$$= 136.9325 + 18$$

$$= \underline{\underline{154.9325}}$$

$$SST = \sum_{i=1}^{j} \sum_{j=1}^{c} (x_{ij} - \bar{x})^2$$

$$= (29-28.2666)^2 + (27-28.2666)^2 + (30-28.2666)^2 + (27-28.2666)^2 + (28-28.2666)^2 + (32-28.2666)^2 + (33-28.2666)^2 + (31-28.2666)^2 + (34-28.2666)^2 + (30-28.2666)^2 + (25-28.2666)^2 + (24-28.2666)^2 + (24-28.2666)^2 + (25-28.2666)^2 + (25-28.2666)^2$$

$$= 0.5378 + 1.6042 + 3.0046 + 1.6042 + 0.0710 + 13.938 + 22.4050 + 7.4714 + 32.8718 + 3.0046 + 10.6706 + 18.2038 + 18.2038 + 10.6706 + 10.6706$$

$$= \underline{154.9329}$$

$df_C = C - 1 = 3 - 1 = 2$

$df_E = N - C = 15 - 3 = 12$

$df_T = N - 1 = 15 - 1 = 14$

Source of variance	SS	df	MS	F
Between (SSC)	136.9334	2	68.4667	65.6446
Error	18	12	1.5	
Total	154.933	14	45.6444	

$$MSC = \frac{SSC}{df_C} = \frac{136.9334}{2} = \underline{68.4667}$$

$$MSE = \frac{SSE}{df_E} = \frac{18}{12} = \underline{1.5}$$

$$MSE = \frac{SSE}{df_E} = \frac{18}{12} = \underline{1.5}$$

$$F = \frac{MSC}{MSE} = \frac{68.4667}{1.5} = \underline{45.6444}$$

Critical F value

$df_C = 2$

$df_E = 12$

$= \underline{6.93}$

The decision is the reject NULL hypothesis because observed F value of 45.6444 is greater than the critical F value 6.93