22MCA0139 Digital Assignment 1 RAJAT SINGH

Foundations of Data Science

Q1) From the following data obtain multiple correlation coefficient value $R_{1.23}$, $R_{2.13}$

x_1: 65 72 54 68 55 59 78 58 57 51

x_2: 56 58 48 61 50 51 55 48 52 42

x_3: 9 11 8 13 10 8 11 16 11 7

Ans

X_1	X_2	X_3	X_1^2	X_2^2	X_3^2	X_1X_2	X_1X_3	X_2X_3
65	56	9	4225	3136	81	3640	585	504
72	58	11	5184	3364	121	4176	792	638
54	48	8	2916	2304	64	2592	432	384
68	61	13	4624	3721	169	4148	884	793
55	50	10	3025	2500	100	2750	550	506
59	51	8	3481	2601	64	3009	472	408
78	55	11	6084	3025	121	4290	858	605
58	48	10	3364	2304	100	2784	580	480
57	52	11	3249	2704	121	2964	627	572
51	42	7	2601	1764	49	2142	357	294
			38753	27423	990	32495	6137	5178

$$r_{12} = \frac{N(\Sigma X_1X_2) - (\Sigma X_1)(\Sigma X_2)}{\sqrt{\{N(\Sigma X_1^2) - (\Sigma X_1)^2\}\{N(\Sigma X_1^2 - \Sigma X_2^2)\}}}$$

$$= \frac{10 \times 32495 - 617 \times 521}{\sqrt{\{10 \times 38753 - 617 \times 617\}\{10 \times 27423 - 521 \times 521\}}}$$

$$= \frac{3493}{\sqrt{6841 \times 2789}} = 0.80$$

$$r_{13} = \frac{(10 \times 6137) - 617 \times 98}{\sqrt{\{10 \times 38753 - 617 \times 617\}\{10 \times 990 - 98 \times 98\}}}$$

$$= \frac{904}{\sqrt{423}} = 0.64$$

$$r_{23} = \frac{10 \times 5178 - 521 \times 98}{\sqrt{\{10 \times 27423 - 521 \times 521\}\{10 \times 990 - 98 \times 98\}}}$$

$$= \frac{722}{808.59} = 0.79$$

$$R^2_{1.23} = \frac{r_{12}^2 + r_{13}^2 - 2 r_{12} r_{13} r_{23}}{1 - r_{23}^2}$$

$$= \frac{(0.80)^2 + (0.64)^2 - 2 \times 0.8 \times 0.64 \times 0.79}{1 - (0.79)^2}$$

$$= 0.63$$

$$\Rightarrow R_{1.23} = 0.79$$

$$R_{2.13}^2 = \frac{r_{12}^2 + r_{13}^2 - 2 r_{12} r_{13} r_{23}}{1 - r_{13}^2}$$

$$= \frac{(0.8)^2 + (0.74)^2 - 2 \times 0.8 \times 0.64 \times 0.79}{1 - 0.64^2}$$

$$= \frac{0.45}{0.51} = 0.88$$

$$\Rightarrow R_{2.13} = 0.94$$

---

2) If $r_{12} = 0.7$, $r_{13} = 0.74$, $r_{23} = 0.54$, calc multiple correlation coefficient $R_{2.13}$

Ans

$$R_{2.13} = \frac{r_{12}^2 + r_{13}^2 - 2 r_{12} r_{13} r_{23}}{1 - r_{13}^2}$$

$$= \frac{(0.7)^2 + (0.54)^2 - 2 (0.7) (0.74) (0.54)}{1 - (0.74)^2}$$

$$= \frac{0.22216}{0.4524} = 0.4910$$

Q3 Discuss the various properties of Partial & multiple correlation.

Correlation

Correlation is a process to establish a relationship between 2 variables. In statistics under relation and functions, methods of correlation summarize the relationship between 2 variables in a single unitless number called the correlation coefficient. The correlation coefficient represented by 'r', ranges from [-1 to +1].

Types of Correlation

The 3 classes of correlation are:

→ positive, negative & no correlation

→ Linear & non-linear correlation

→ Simple, multiple & partial correlation.

Properties of Multiple Correlation

The following are some properties of multiple correlation

→ Multiple correlation coefficient is degree of association between observed value of dependent variable & its estimate obtained by multiple regression.

→ Multiple correlation coefficient lies between 0 & 1.

→ If multiple correlation coefficient is 1, then association is perfect & multiple regression equation may said to be perfect prediction formula.

Q4) Bring out the procedure to draw the box and whisker plot for the given dataset and briefly explain its properties with examples.

Ans

Step1

Order the dataset from small to large.

Step2

Determine Median,
if n = even then take Avg of 2 middle values

Step3

Divide the dataset into 2 halfs using the median.

Step4

Determine median of two halves which is Q_1 and Q_3 (Quartiles)

Step5

Find $IQR = Q_3 - Q_1$

Step 6:

Draw box from Q_1 to Q_3 with vertical line inside the box of the median.

Step 7:

Draw whiskers at end of box with smallest and largest values.

For example, we have,

65, 72, 59, 68, 55, 59, 78, 58, 57, 51

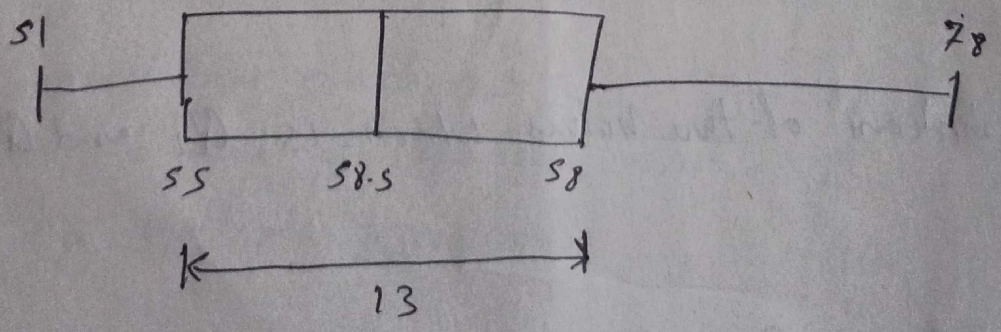
Sorted: 51, 54, 55, 57, 58, 59, 65, 68, 72, 78

$$\text{Median} = \frac{58+59}{2} = 58.5$$

$Q_1 = 55$

$Q_3 = 68$

$IQR = 68 - 55 = 13$



Box & Whisker Plot

Q5) Using any popular Data Visualization tools available like Tableau, MS-PowerBI or STATCRAFT bring and keep panels on data set through visuals in terms of chart graph plot provide demo also.

Ans

We use a dataset of a cricketer's scores over a period of time. We want to visualize this stat.
We use Python and Matplotlib to do this.

Code:

```
import matplotlib.pyplot as plt
import numpy as np

days = ~~np.arrange(1, )~~ range(1,14)

score = [25, 27, 42, 20, 78, 72, 50, 55, 31, 100,
         161, 2, 94]
```

+ Code + Markdown | Run All Clear Outputs of All Cells Restart Interrupt | Variables Outline ...

```
import matplotlib.pyplot as plt
import numpy as np
days = range(1,14)
score = [25,27,42,20,78,77,50,55,31,100,161,2,94]
plt.plot(days, score)
plt.xlabel("Day Number")
plt.ylabel("Score")
plt.title("Batsman Performance Graph")
plt.show()
```

[3] ✓ 0.1s