VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

Probability and Statistics

Assignment-1

NAME : Mohammed Kaif K
REG.NO : 23MIS0258
SUBJECT CODE : IMAT202P
SUBJECT TITLE : Probability And Statistics Lab
LAB SLOT : L35+L36

Instructions:

- Use A4 size papers to answer the questions and upload (scanned copy without any shadows on it) the same.
- Every scanned page must contain your registered number and Name.
- Answer all questions theoretically and write R-programming for the same.

<table>
<tr><th>S. No.</th><th>Question</th></tr>
<tr><td>1.</td><td>Find mean, median ,mode (theoretically) and also verify the same with R programming for the following frequency distribution
<table>
<tr><td>Age group (in years)</td><td>20-25</td><td>25-30</td><td>30-35</td><td>35-40</td><td>40-45</td><td>45-50</td><td>50-55</td><td>55-60</td><td>60-65</td></tr>
<tr><td>No of Members</td><td>30</td><td>160</td><td>210</td><td>180</td><td>145</td><td>105</td><td>70</td><td>60</td><td>40</td></tr>
</table></td></tr>
<tr><td>2.</td><td>Calculate Quartile deviation and Mean deviation from mean and standard deviation (theoretically) and also verify the same with R programming
for the following data:
<table>
<tr><td>Marks</td><td>0-10</td><td>10-20</td><td>20-30</td><td>30-40</td><td>40-50</td><td>50-60</td><td>60-70</td></tr>
<tr><td>No.of Students</td><td>6</td><td>5</td><td>8</td><td>15</td><td>7</td><td>6</td><td>3</td></tr>
</table></td></tr>
</table>

Theoretical Method:

Assignment - 1

Name: Mohammed Kaif K

Reg no: 23MIS0258

① Mean, Median and Mode for age groups distribution

Age groups :-	20-25	25-30	30-35	35-40	40-45	45-50	50-55	55-60	60-65
Mid points :-	22.5	27.5	32.5	37.5	42.5	47.5	52.5	57.5	62.5
Frequency :-	30	160	210	180	145	105	70	60	40

Total Members :- 1000

mean :-

$$\text{mean} = \frac{\Sigma(\text{mid} \times \text{freq})}{\Sigma \text{freq}} = \frac{22.5\times30 + 27.5\times160 + 32.5\times210 + 37.5\times180 + 42.5\times145 + 47.5\times105 + 52.5\times70 + 57.5\times60 + 62.5\times40}{1000}$$

$$= \frac{39425}{1000} = 39.425$$

Median :-

cumm freq: 30, 190, 400, 580, 725, 830, 900, 960, 1000

Median grp is 30 - 35

$$\text{median} = L + \left(\frac{N/2 - C.F}{f}\right) \times w = 30 + \left(\frac{500 - 190}{210}\right) \times 5$$

$$= 30 + \left(\frac{310}{210}\right) \times 5 = 30 + 7.38 = 37.38$$

Mode :-

Modal group is 30 - 35 (highest frequency is 210)

$$\text{mode} = L + \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right) \times w = 30 + \left(\frac{210 - 160}{2\times210 - 160 - 180}\right) \times 5$$

$$= 30 + \left(\frac{50}{80}\right) \times 5 = 30 + 3.125 = 33.125$$

2. Quartile Deviation, Mean Deviation from mean & Standard Deviation for marks Distribution.

marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70
mid	5	15	25	35	45	55	65
Freq	6	5	8	15	7	6	3

Total stud : 50

Mean:

$$\text{mean} = \frac{\Sigma(\text{mid} \times \text{freq})}{\Sigma \text{freq}} = \frac{5\times6 + 15\times5 + 25\times8 + 35\times15 + 45\times7 + 55\times6 + 65\times3}{50}$$

$$= \frac{1670}{50} = 33.4$$

Quartile Deviation:

cum freq: 6, 11, 19, 34, 41, 47, 50

Q. group : 20-30 (where cummaltive freq reaches 12.5)

$$Q_1 = L + \left(\frac{N/4 - C.F}{f}\right) \times w = 20 + \left(\frac{12.5 - 11}{8}\right) \times 10 = 20 + \left(\frac{1.5}{8}\right) \times 10$$

$$= 20 + 1.875 = 21.875$$

Q_3 group : 40-50 (where C.F reaches 37.5)

$$Q_3 = 40 + \left(\frac{37.5 - 34}{7}\right) \times 10 = 45$$

$$\text{Quartile Deviation} = \frac{Q_3 - Q_1}{2} = \frac{45 - 21.875}{2}$$

$$= \frac{23.125}{2} = \boxed{11.5625}$$

Mean Deviation from Mean.

$$\text{Mean Deviation} = \frac{\Sigma |\text{mid} - \text{mean}| \times \text{freq}}{\Sigma \text{freq}}$$

$$= \frac{|5-33.4| \times 6 + |15-33.4| \times 5 + |25-33.4| \times 8 + |35-33.4| \times 15 + |45-33.4| \times 7 + |53-33.4| \times 6 + |65-33.4| \times 3}{50}$$

$$= \frac{170.4 + 92 + 67.2 + 24 + 81.2 + 129.6 + 94.3}{50}$$

$$= \frac{659.2}{50} = 13.184$$

Standard Deviation:-

$$\text{Variance} = \frac{\Sigma (\text{mid} - \text{mean})^2 \times f}{\Sigma f} = \frac{(5-33.4)^2 \times 6 + (15-32.4)^2 \times 5 + (25-33.4)^2 \times 8 + (35-33.4)^2 \times 15 + (45-33.4)^2 \times 7 + (55-33.4)^2 \times 6 + (65-33.4)^2 \times 3}{50}$$

$$= \frac{13871.2}{50} = 277.42$$

$$\text{Standard Deviation} = \sqrt{277.42}$$

$$= 16.65653$$

R program code and output:

1.

Code:

```
age_intervals <- c(22.5, 27.5, 32.5, 37.5, 42.5, 47.5, 52.5, 57.5, 62.5)
age_counts <- c(30, 160, 210, 180, 145, 105, 70, 60, 40)

total_count <- sum(age_counts)
average_age <- sum(age_intervals * age_counts) / total_count

all_ages <- rep(age_intervals, age_counts)
middle_value <- median(all_ages)

most_common_age <- age_intervals[which.max(age_counts)]

average_age
middle_value
most_common_age
```

Output:

```
[Previously saved workspace restored]

> age_intervals <- c(22.5, 27.5, 32.5, 37.5, 42.5, 47.5, 52.5, 57.5, 62.5)
> age_counts <- c(30, 160, 210, 180, 145, 105, 70, 60, 40)
>
> total_count <- sum(age_counts)
> average_age <- sum(age_intervals * age_counts) / total_count
>
> all_ages <- rep(age_intervals, age_counts)
> middle_value <- median(all_ages)
>
> most_common_age <- age_intervals[which.max(age_counts)]
>
> average_age
[1] 39.425
> middle_value
[1] 37.5
> most_common_age
[1] 32.5
>
```

2.

Code:

```
marks <- c(5, 15, 25, 35, 45, 55, 65)

students <- c(6, 5, 8, 15, 7, 6, 3)

expanded_data <- rep(marks, students)

Q1 <- quantile(expanded_data, 0.25)

Q3 <- quantile(expanded_data, 0.75)

quartile_deviation <- (Q3 - Q1) / 2


mean_value <- mean(expanded_data)

mean_deviation <- mean(abs(expanded_data - mean_value))

population_standard_deviation <- sqrt(sum((expanded_data - mean_value)^2) /
length(expanded_data))


cat("Q1:", Q1, "\n")

cat("Q3:", Q3, "\n")

cat("Quartile Deviation:", quartile_deviation, "\n")

cat("Mean Value:", mean_value, "\n")

cat("Mean Deviation:", mean_deviation, "\n")

cat("Standard Deviation:", population_standard_deviation, "\n")
```

Output:

```
> marks <- c(5, 15, 25, 35, 45, 55, 65)
> students <- c(6, 5, 8, 15, 7, 6, 3)
> expanded_data <- rep(marks, students)
>
> Q1 <- quantile(expanded_data, 0.25)
> Q3 <- quantile(expanded_data, 0.75)
> quartile_deviation <- (Q3 - Q1) / 2
>
> mean_value <- mean(expanded_data)
> mean_deviation <- mean(abs(expanded_data - mean_value))
> population_standard_deviation <- sqrt(sum((expanded_data - mean_value)^2) / l$
>
> cat("Q1:", Q1, "\n")
Q1: 25
> cat("Q3:", Q3, "\n")
Q3: 45
> cat("Quartile Deviation:", quartile_deviation, "\n")
Quartile Deviation: 10
> cat("Mean Value:", mean_value, "\n")
Mean Value: 33.4
> cat("Mean Deviation:", mean_deviation, "\n")
Mean Deviation: 13.184
> cat("Standard Deviation:", population_standard_deviation, "\n")
Standard Deviation: 16.65653
>
```