Assignment 2-Probability and Statistics

Probability and statitics (National University of Computer and Emerging Sciences)

Q No 1: With reference to the Q No 1 of the Assignment no 1.

    i.   Find its mean, median, mode and Geometric Mean.

ANS:

Data: 33,31,28,15,17,17,16,18,16,18,25,31,27,30,29,33,28,20,22,23,30,36,37,27,33,28,31,29,32 31,22,25,35,19,21,24,30,26,37,27,31,29,34,21,23,25,35,26,38,27.

Class	Frequency	Mid(x)	C.B	Cumulative Frequency	F.x	F.log(x)
15-18	7	16.5	14.5-18.5	7	115.5	8.52
19-22	6	20.5	18.5-22.5	13	123	7.87
23-26	8	24.5	22.5-26.5	21	196	11.11
27-30	13	28.5	26.5-30.5	34	370.5	18.91
31-34	10	32.5	30.5-34.5	44	325	15.12
35-38	6	36.5	34.5-38.5	50	219	9.37
	$\sum f = 50$				$\sum f(x) = 1349$	$\sum f\log(x) = 70.9$

➢ Mean = $\dfrac{\sum f(x)}{\sum f}$ = 1349/50 = 26.98

➢ Median = $l + \dfrac{h}{f}\left(\dfrac{n}{2} - c\right)$ = $26.5 + \dfrac{4}{13}\left(\dfrac{50}{2} - 21\right)$ = 27.73

➢ Mode = $\left(f_m - f_{m-1}\right) + \left(¿¿ m - f_{m+1}\right) \times h = 26.5 + \dfrac{13 - 8}{(13 - 8) + (13 - 10)} \times 4$ = 27.83

$$l + \dfrac{f_m - f_{m-1}}{¿}$$

➢ Log G.M = $\dfrac{\sum f\log(x)}{\sum f}$ = $\dfrac{70.9}{50}$ = 10 x 1.418 = 14.18

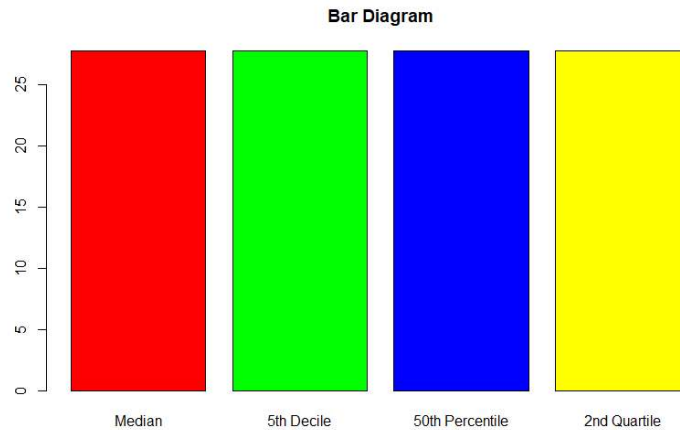   **ii.   Show the relation among median, fifth decile, 50th percentile and second quartile.**

ANS:

➢ Median = 27.73.

➢ 5th decile = $l + \dfrac{h}{f}\left(\dfrac{5 \times n}{10} - c\right)$ = $26.5 + \dfrac{4}{13}\left(\dfrac{5 \times 50}{10} - 21\right)$ = 27.73

- 50th percentile= $l+\dfrac{h}{f}\left(\dfrac{50\times n}{100}-c\right)$ = $26.5+\dfrac{4}{13}\left(\dfrac{50\times50}{100}-21\right)$ = 27.73

- Second quartile = $l+\dfrac{h}{f}\left(\dfrac{2\times n}{4}-c\right)$ = $26.5+\dfrac{4}{13}\left(\dfrac{2\times50}{4}-21\right)$ = 27.73

- Median = 5th decile = 50th percentile = 2nd quartile.

iii. Represent part(ii) graphically and compare with part(ii).

Bar Diagram

iv. Calculate the absolute and relative measure of dispersion using semi interquartile range.

ANS:

- 3rd Quartile = $l+\dfrac{h}{f}\left(\dfrac{3\times n}{4}-c\right)$ = $22.5+0.5\left(\dfrac{3\times50}{4}-34\right)$ = 24.25

- 1st Quartile = $l+\dfrac{h}{f}\left(\dfrac{1\times n}{4}-c\right)$ = $14.5+0.5\left(\dfrac{1\times50}{4}-7\right)$ = 17.63

- **Semi-interquartile range =** $\dfrac{24.25-17.63}{2}$ = **3.31**

v. What is the Co-efficient of Skewness? Also interpret your result.

ANS:

Frequency	MidPoint(x)	F(x-mean)^2
7	16.5	768.81
6	20.25	251.94
8	24.5	49.2
13	28.5	30.04
10	32.5	304.704

6	36.5	543.78
∑f=50		∑f(x-mean)^2

Std = $\sqrt{\dfrac{\sum f(x-mean)2}{\sum f}}$ = 6.24

- **Formula 1:**

  $S_k = \dfrac{3(mean-median)}{Std} = \dfrac{3(26.98-27.73)}{6.24}$ = -0.36

- **Formula 2:**

  $S_k = \dfrac{mean-mode}{Std} = \dfrac{26.98-27.83}{6.24}$ = -0.136

- **Formula 3:**

  $S_k = \dfrac{Q3+Q1-2(Q2)}{Q3-Q1} = \dfrac{24.25+17.63-2(27.72)}{24.25-17.63}$ = -2.04

vi. **What can you conclude about the shape of this data? Explain your reasoning.**

vii. **Repeat part(i) in R.**

  a. **Mean: 26.92**

  data <-
  c(33,31,28,15,17,17,16,18,16,18,25,31,27,30,29,33,28,20,22,23,30,36,37,
  27,33,28,31,29,32,31,22,25,35,19,21,24,30,26,37,27,31,29,34,21,23,25,3
  5,26,38,27)
   mean(data)

  b. **Median: 27.5**

  data <-
  c(33,31,28,15,17,17,16,18,16,18,25,31,27,30,29,33,28,20,22,23,30,36,37,
  27,33,28,31,29,32,31,22,25,35,19,21,24,30,26,37,27,31,29,34,21,23,25,3
  5,26,38,27)

  median(data)

  c. **Mode: 31**

  data <-
  c(33,31,28,15,17,17,16,18,16,18,25,31,27,30,29,33,28,20,22,23,30,36,37,
  27,33,28,31,29,32,31,22,25,35,19,21,24,30,26,37,27,31,29,34,21,23,25,3
  5,26,38,27)

  getmode <- function(v) {

  +   uniqv <- unique(v)

```
+    uniqv[which.max(tabulate(match(v, uniqv)))]

+ }

> getmode(data)
```

d. **GM:**

```
data <-
c(33,31,28,15,17,17,16,18,16,18,25,31,27,30,29,33,28,20,22,23,30,36,37,
27,33,28,31,29,32,31,22,25,35,19,21,24,30,26,37,27,31,29,34,21,23,25,3
5,26,38,27)
```

geometric.mean(data)

Q No 2: The following scores represent the final examination grade for an elementary statistics course:

23 60 79 32 57 74 52 70 82 36 80 77 81 95 41 65 92 85 55 76 52 10 64 75 78 25 80 98 81 67 41 71 83 54 64 72 88 62 74 43 60 78 89 76 84 48 84 90 15 79 34 67 17 82 69 74 63 80 85 61

i. **Find the median, mode, Quartile and Std Deviation.**

- Mean = $\dfrac{\sum f(x)}{\sum f}$ = 3929/60 = 65.48

- Median = $\dfrac{71+72}{2}$ = 71.5

- Mode = 74.80

- 1st Quartile = $\dfrac{54+55.5}{2}$ = 54.75

- 2nd Quartile = $\dfrac{71+72}{2}$ = 71.5

- 3rd Quartile = $\dfrac{80+80.5}{2}$ = 80.25

- Std= $\sqrt{\dfrac{\sum f(x-mean)2}{\sum f}}$

  $\sum(x\text{-mean})^2$=3078.03 + 2548.23 + 2350.31 + 1804.55 + 1638.63 + 1120.91 + 990.09 + 869.07+599.27+599.27+505.35+305.55+181.71+181.71+131.79+ 109.83+71.91+30.03+30.03+20.07+12.11+6.15+2.19+2.19+0.23+2.31+2.31 +12.39+20.43+30.47+42.51+72.59+72.59+72.59+90.63+110.67+110.67 +132.71+156.76+156.76+182.79+182.79+210.83+210.83+210.83+240.87+ 240.87+272.91+272.91+306.95+342.99+342.99+381.03+381.03+507.15+553.19 601.23+703.31+871.43+1057.55 = 26350.96

  $\sqrt{\dfrac{26350.96}{60}}$ = 20.96

ii. **Make a grouped frequency distribution of this data and again calculate mean, median, mode, Quartiles and Std Dev.**

ANS:

Number of Class = k=1+3.3log(60) = 7

Range = h =98-10 / 7 = 13

Class	Frequency	MidPoints(x)	C.B	F(x)	Cumulative Frequency	F(x-mean)^2
10-22	3	16	9.5-22.5	48	3	7449.09
23-35	4	29	22.5-35.5	116	7	5425.796
36-48	5	42	35.5-48.5	210	12	2839.34
49-61	8	55	48.5-61.5	440	20	938.31
62-74	14	68	61.5-74.5	952	34	65.92
75-87	20	81	74.5-87.5	1620	54	4602.58
88-100	6	94	87.5-100.5	564	60	4761.29
	$\sum$f=60			$\sum$f(x)=3950		$\sum$=26082.33

➢ Mean = $\dfrac{\sum f(x)}{\sum f}$ = 3950/60 = 65.83

➢ Median = $l+\dfrac{h}{f}\left(\dfrac{n}{2}-c\right)$ = 70.76

- Mode = $\dfrac{(f_m - f_{m-1}) + (¿¿ m - f_{m+1}) \times h = 78.4}{l + \dfrac{f_m - f_{m-1}}{¿}}$

- 1st Quartile = $l + \dfrac{h}{f}\left(\dfrac{1 \times n}{4} - c\right)$ = 53.375

- 2nd Quartile = $l + \dfrac{h}{f}\left(\dfrac{2 \times n}{4} - c\right)$ = 70.76

- 3rd Quartile = $l + \dfrac{h}{f}\left(\dfrac{3 \times n}{4} - c\right)$ = 81.65

- Std=20.84

iii. **Compare your results of part (i) and part(ii), also comment on your results.**

Mean(i)=65.58,(ii)= 65.83 difference of 0.52

Mode(i)=74.80,(ii)=78.4 .

Median(i)=71.5,(ii)= 70.76 difference of 0.74

Std(i)=20.96,(ii)= 20.84 difference of 0.12

Q1,Q2,Q3 have very little difference.

Overall results are very close to each other.

iv. **Find the mean, median, mode and std deviation through R.**
   a. Mean: 65.4833

data <-c(23 60 79 32 57 74 52 70 82 36 80 77 81 95 41 65 92 85 55 76 52 10 64 75 78 25 80 98 81 67 41 71 83 54 64 72 88 62 74 43 60 78 89 76 84 48 84 90 15 79 34 67 17 82 69 74 63 80 85 61)

mean(data)
   b. Median: 71.5
   data <-c(23 60 79 32 57 74 52 70 82 36 80 77 81 95 41 65 92 85 55 76 52 10 64 75 78 25 80 98 81 67 41 71 83 54 64 72 88 62 74 43 60 78 89 76 84 48 84 90 15 79 34 67 17 82 69 74 63 80 85 61)

median(data)

   c. Mode: 74

data <-c(23 60 79 32 57 74 52 70 82 36 80 77 81 95 41 65 92 85 55 76 52 10 64 75 78 25 80 98 81 67 41 71 83 54 64 72 88 62 74 43 60 78 89 76 84 48 84 90 15 79 34 67 17 82 69 74 63 80 85 61)

getmode <- function(v) {

+   uniqv <- unique(v)

+   uniqv[which.max(tabulate(match(v, uniqv)))]

+ }

> getmode(data)

d. Std: 21.133

data <-c(23 60 79 32 57 74 52 70 82 36 80 77 81 95 41 65 92 85 55 76 52 10 64 75 78 25 80 98 81 67 41 71 83 54 64 72 88 62 74 43 60 78 89 76 84 48 84 90 15 79 34 67 17 82 69 74 63 80 85 61)

sd(data);

v. Define Skewness of the data set:
Skewness = mean – mode =65.83 – 78.4 =-12.57

Q NO 3: A study of the effect of smoking on sleep patterns is conducted. The measure observed is the time, in minutes, that it takes to fall sleep.These data are obtained:

Smokers = 69.3 56.0 22.1 47.6 53.2 5.2.7 34.4 60.2 43.8 23.2 13.8

Non-Smokers = 28.6 25.1 26.4 34.9 28.8 28.4 38.5 30.2 30.6 31.8 41.6 21.1 36.0 37.9 13.9

i. By calculating mean for each data, which group do you think have better asleep time taken and why?
Smokers: Mean=43.7
Non-Smokers: Mean=30.25
As mean of non-smokers is less than the smokers so they have better asleep time taken.

ii. **Calculate std dev, variance and co-efficient of variations for each group and compare them.**
Std Dev(Smokers)= 16.21
Variance(Smokers) =262.67
Co-efficient of variance(Smokers) = 37.1

Std Dev(Non-Smokers)= 6.89
Variance(Non-Smokers)= 47.54
Co-efficient of variance(Non-Smokers)= 22.78

Smokers have greater std dev, variance and co-efficient than non smokers.

iii. **Which of them have greater relative dispersion and why?**
Smokers have greater value of dispersion as smokers have greater mean and std dev.

iv. **Do part(iii) in R.**
Smokers
a. Std: 16.93

data <-c(69.3 56.0 22.1 47.6 53.2 5.2.7 34.4 60.2 43.8 23.2 13.8)

sd(data)

b. Variance: 286.55
data <-c(69.3 56.0 22.1 47.6 53.2 5.2.7 34.4 60.2 43.8 23.2 13.8)

var(data)

c. Co-efficient: 38.7
data <-c(69.3 56.0 22.1 47.6 53.2 5.2.7 34.4 60.2 43.8 23.2 13.8)
sd(data)/ var(data) * 100

Non-Smokers
d. Std: 7.14

data <-c(28.6 25.1 26.4 34.9 28.8 28.4 38.5 30.2 30.6 31.8 41.6 21.1 36.0 37.9 13.9)

sd(data)

e. Variance: 50.95
data <-c(28.6 25.1 26.4 34.9 28.8 28.4 38.5 30.2 30.6 31.8 41.6 21.1 36.0 37.9 13.9)

var(data)

f. Co-efficient: 23.6

data <-c(28.6 25.1 26.4 34.9 28.8 28.4 38.5 30.2 30.6 31.8 41.6 21.1
36.0 37.9 13.9)

sd(data)/ var(data) * 100

Q.NO.4
Source(a)

Frequency	MidPoints(x)	C.B	F(x)	Cumulative Frequency	F(x-mean)^2
40	1010	1000-1020	40400	40	99600.4
96	1030	1020-1040	98880	136	85824.96
364	1050	1040-1060	382200	500	37675.64
372	1070	1060-1080	398040	872	37947.72
85	1090	1080-1100	92650	957	77010.85
43	1110	1100-1120	47730	1000	107930.43
∑f=1000			∑f(x)=1059900		∑=44390

➢ Mean = $\dfrac{\sum f(x)}{\sum f}$ = 1060

➢ 1st Quartile = $l+\dfrac{h}{f}\left(\dfrac{1 \times n}{4}-c\right)$ = 1046.26

➢ 3rd Quartile = $l+\dfrac{h}{f}\left(\dfrac{3 \times n}{4}-c\right)$ = 1073.44

Source(b)

Frequency	MidPoints(x)	C.B	F(x)	Cumulative Frequency	F(x-mean)^2
339	1035	1030-1040	350865	339	21595.68
136	1045	1040-1050	142120	475	31256.28
25	1055	1050-1060	26375	500	665.64
20	1065	1060-1070	21300	520	468.51
130	1075	1070-1080	139750	650	28629.33
350	1085	1080-1090	379750	1000	215958.90
∑f=1000			∑f(x)=106016		∑=491574

			0		

- Mean = $\dfrac{\sum f(x)}{\sum f}$ = 1060

- 1st Quartile = $l + \dfrac{h}{f}\left(\dfrac{1 \times n}{4} - c\right)$ = 1037.37

- 3rd Quartile = $l + \dfrac{h}{f}\left(\dfrac{3 \times n}{4} - c\right)$ = 1082.86