NAME JALAL KHAN

ID 14885

Paper BASIC STATISTICS

DATE 26/9/2020

Scanned with CamScanner

Question 1

(a)

No. of children (x)	No. of families (f)	f(x)	1/x	f(1/x)	log(x)	f(log x)
1	4	4	1.08	4	0	0
2	13	26	0.50	6.5	0.3	3.9
3	9	27	0.33	2.97	0.477	4.29
4	4	16	0.25	1	0.6	2.4
5	1	5	0.2	0.2	0.69	0.69
	31	78		14.67		11.28

$$AM = \frac{\sum f(x)}{\sum f} = \frac{78}{31} = 2.5161$$

$$GM = \text{Antilog}\left[\frac{\sum f(\log x)}{\sum f}\right] = \text{Antilog}\left[\frac{11.28}{31}\right] = 2.311$$

$$HM = \frac{\sum f}{\sum f(1/x)} = \frac{31}{14.67} = 2.11$$

Scanned with CamScanner

Q1

(b)

DATE

marks	C.B	f	X	f(x)	1/x	f(1/x)	log(x)	f(log(x))
0-9	-0.05-9.5	2	4.5	8.5	0.22	0.44	0.65	1.3
10-19	9.5-19.5	31	14.5	449.5	0.06	1.86	1.61	49.91
20-29	19.5-29.5	73	24.5	1788.5	0.04	2.92	1.38	100.74
30-39	29.5-39.5	85	34.5	2932.5	0.028	2.38	1.53	130.05
40-49	39.5-49.5	28	44.5	1246	0.022	0.616	1.65	46.2
		219		6425		8.216		328.2

$$AM = \frac{\Sigma f(x)}{\Sigma f} = \frac{6425}{219} = 29.33$$

$$GM = \text{Antilog}\left[\frac{\Sigma f(\log x)}{\Sigma f}\right] = \text{Antilog}\left[\frac{328.2}{219}\right] = 31.52$$

$$HM = \frac{\Sigma f}{\Sigma f(1/x)} = \frac{219}{8.216} = 26.655$$

Scanned with CamScanner

QUESTION 2

(a)

x	f	C.F
1	4	4
2	13	17
3	9	26
4	4	30
5	1	31
	31	

$$\text{Median} = \left(\frac{N+1}{2}\right)^{th} = \frac{32}{2} = 16^{th}$$

$\boxed{\text{Median} = 2}$

$\boxed{\text{Mode} = 2}$

↳ Because it has highest frequency.

Scanned with CamScanner

Question 2

(b)

Marks	C.B	f	C.F
0-9	-0.5-9.5	2	2
10-19	9.5-19.5	31	33
20-29	19.5-29.5	73	106
30-39	29.5-39.5	85	191
40-49	39.5-49.5	28	219.
		219	

$\Rightarrow$ Median $= l + \frac{N/2 - m}{f} \times c$

$\because l = 29.5, N = 219, m = 106, c = 10, f = 85$

Median $= 29.5 + \frac{(219/2) - 106}{85} \times 10$

Median $= 29.91 \simeq 30.$

$\Rightarrow$ Mode $= l + h\left[(f_1 - f_0)/(2f_1 - f_0 - f_2)\right]$

$= 29.5 + 10\left[(85-73)/(2(85)-73-28)\right]$

$\boxed{\text{Mode} = 31.23}$

Scanned with CamScanner

QUESTION 3 (a)

DATE

Marks	C.B	f	C.F	
0-9	-0.5-9.5	2	2	
10-19	9.5-19.5	31	33	
20-29	19.5-29.5	71	106	→ Q_1 class
30-39	29.5-39.5	85	191	→ Q_3 class.
40-49	39.5-49.5	28	219.	

Q_1 class $= \frac{N}{4} = \frac{219}{4} = 54.75$

Q_3 class $= 3\left[\frac{N}{4}\right] = 3 \times 54.75 = 164.25$.

$$Q_1 = l_1 + \frac{N/4 - m_1}{f_1} \times c_1$$

$$= 19.5 + \frac{54.75 - 33}{71} \times 10 \qquad \boxed{Q_1 = 22.56}$$

$$Q_3 = l_3 + \frac{3(N/4) - m_3}{f_3} \times 10$$

$$= 29.5 + \frac{164.25 - 106}{85} \times 10 \qquad \boxed{Q_3 = 36.35}$$

$IQR = 36.35 - 22.56 = 13.79.$

Semi $IQR = 13.79/2 = 6.895.$

Scanned with CamScanner

Question 3

(b)

C.B	X	f	$\bar{x}$	$x-\bar{x}$	$(x-\bar{x})^2$	$f(x-\bar{x})^2$
-0.5 - 9.5	4.5	2	29.33	-24.83	616.52	1233.04
9.5 - 19.5	14.5	31	29.33	-14.83	219.92	6817.52
19.5 - 29.5	24.5	73	29.33	-4.83	23.32	1702.36
29.5 - 39.5	34.5	85	29.33	5.17	26.73	2272.05
39.5 - 49.5	44.5	28	29.33	15.17	230.12	6443.36
		219				18468.33

$$\text{Variance} = \sigma^2 = \frac{\Sigma f(x-\bar{x})^2}{\Sigma f}$$

$$\sigma^2 = \frac{18468.33}{219}$$

$$\boxed{\sigma^2 = 84.33}$$

$$\text{Standard deviation} = \sigma = \sqrt{\sigma^2}$$

$$= \sqrt{84.33}$$

$$\boxed{\sigma = 9.18}$$

Scanned with CamScanner

Q4)

A)Range & Quartile Range & Semi Inter Quartile Range

Ans)

Range: The **Range is the** difference between the lowest and highest values.

Example: In {4, 6, 9, 3, 7} the lowest value is 3, and the highest is 9. So the **range** is 9 − 3 = 6.

Quartile Range: The interquartile **range (IQR)** is a measure of variability, based on dividing a data set into **quartiles**. **Quartiles** divide a rank-ordered data set into four equal parts. The values that divide each part are called the first, second, and third **quartiles**; and they are denoted by Q1, Q2, and Q3, respectively.

Example:

5, 7, 4, 4, 6, 2, 8
Put them in order: 2, 4, 4, 5, 6, 7, 8. Cut the list into quarters: And the result is: **Quartile** 1 (Q1) = 4. **Quartile** 2 (Q2), which is also the Median, = 5.

Semi inter quartile range:

The semi-interquartile range is a measure of spread or dispersion. It is computed as one half the difference between the 75th percentile [often called (Q3)] and the 25th percentile (Q1). The formula for semi-interquartile range is therefore: (Q3-Q1)/2.

The semi-interquartile range is half of the difference between the upper quartile and the lower quartile. In the previous example, the quartiles were Q 1 = 4 and. Q 3 = 11 . The semi-interquartile range is: 1 2 (Q 3 − Q 1)

Scanned with CamScanner

Example:

Question: Find the Quartile Deviation for the following set of data:
{490, 540, 590, 600, 620, 650, 680, 770, 830, 840, 890, 900}

Step 1: Find the **first** quartile, Q_1.
This is the median of the lower half of the set {490, 540, 590, 600, 620, 650}.
$Q_1 = (590 + 600) / 2 = 595$.

Step 2: Find the **third** quartile, Q_3.
This is the median of the upper half of the set {680, 770, 830, 840, 890, 900}.
$Q_3 = (830 + 840) / 2 = 835$.

Step 3: Subtract Step 1 from Step 2.
835 – 595 = 240.

Step 4: Divide by 2. 240 / 2 = 120

QB) Variance & Standard Deviation &Coefficient of variation?

Ans

Variance: Variance (σ^2) in statistics is a measurement of the spread between numbers in a data set. That is, it measures how far each number in the set is from the mean and therefore from every other number in the set.

Example:

for the numbers 1, 2, and 3 the mean is 2 and the **variance** is 0.667.

Standard Deviation: The **standard deviation** is a statistic that measures the dispersion of a dataset relative to its mean and is calculated as the square root of the variance. The **standard deviation** is calculated as the square root of variance by determining each data point's **deviation** relative to the mean.

Example: 5, 15, 15, 14, 16 and 2, 7, 14, 22, 30. However, the second is clearly more spread out.

Scanned with CamScanner

Coefficient of variation:

The coefficient of variation (relative standard deviation) is a statistical measure of the dispersion of data points around the mean. The metric is commonly used to compare the data dispersion between distinct series of data. Unlike the standard deviation that must always be considered in the context of the mean of the data, the coefficient of variation provides a relatively simple and quick tool to compare different data series.

Example:

Fred wants to find a new investment for his portfolio. He is looking for a safe investment that provides stable returns. He considers the following options for investment:

- **Stocks:** Fred was offered stock of ABC Corp. It is a mature company with strong operational and financial performance. The volatility of the stock is 10% and the expected return is 14%.
- **ETFs:** Another option is an Exchange-Traded Fund (ETF) which tracks the performance of the S&P 500 index. The ETF offers an expected return of 13% with a volatility of 7%.
- **Bonds:** Bonds with excellent credit ratings offer an expected return of 3% with 2% volatility.

Scanned with CamScanner