ASSIGNMENT

SPRING-2020

Program: B.B.A / MMC **Name: Haris khan**

Course Title: BASIC STATISTICS / STATISTICS-I **ID: 13825**

Dated: 27th June ,2020

Instructor: Raza Ahmed Khan

Question # 01

a

Number of children per family x	Number of families f	$f \cdot x$	$f \cdot \log x$	f/x
1	4	4	$4 \log 1 = 0$	$4/1 = 4$
2	13	26	$13 \log 2 = 3.913$	$13/2 = 6.5$
3	9	27	$9 \times \log 3 = 4.294$	$9/3 = 3$
4	4	16	$4 \times \log 4 = 2.408$	$4/4 = 1$
5	1	5	$1 \times \log 5 = 0.699$	$1/5 = 0.2$
	$\Sigma f = 31$	$\Sigma fx = 78$	$\Sigma f \log x = 11.314$	$\Sigma f/x = 14.7$

Arithematic Mean

$$A.M = \frac{\Sigma fx}{\Sigma f} = \frac{78}{31}$$

$$\boxed{A.M = 2.516}$$

Geometrik Mean

$$G.M = \text{Antilog}\left(\frac{\Sigma (f \log x)}{\Sigma f}\right)$$

$$G.M = \text{Antilog}\left(\frac{11.314}{31}\right)$$

$$G.M = \text{Antilog}(0.365)$$

$$\boxed{G.M = 2.317}$$

Hormonic Mean

$$H.M = \frac{\Sigma f}{\Sigma (f/x)}$$

$$H.M = \frac{31.}{14.7}$$

$$\boxed{H.M = 2.1088}$$

logical Relationship of A.M, G.M, H.M

$$A.M \geq G.M \geq H.M$$

$$2.586 > 2.317 > 2.1088$$

logical Relationship of A.M, G.M, H.M Justified.

Q1 b

Class interval Mark x	f	Midpoint m	$f.m$	$f \log m$	f/m
0 – 9	2	4.5	$2 \times 4.5 = 9$	$2\log 4.5 = 1.306$	0.444
10 – 19	31	14.5	$31 \times 14.5 = 449.5$	$31 \log 14.5 = 36.002$	2.138
20 – 29	73	24.5	$73 \times 24.5 = 1788.5$	$73 \log 24.5 = 101.409$	2.98
30 – 39	85	34.5	$85 \times 34.5 = 2932.5$	$85 \log 34.5 = 130.715$	2.464
40 – 49	28	44.5	$28 \times 44.5 = 1246$	$28 \log 44.5 = 46.154$	0.629
	$\Sigma f = 219$		$\Sigma fm = 6425.5$	$\Sigma(f \log m) = 315.586$	$\Sigma f/m = 8.655$

$$A.M = \frac{\Sigma fm}{\Sigma f} = \frac{6425.5}{219} = 29.34$$

$$\boxed{A.M = 29.34}$$

$$G.M = \text{Antilog}\left(\frac{\Sigma(f \times \log m)}{\Sigma f}\right)$$

$$G.M = \text{Antilog}\left(\frac{315.586}{219}\right)$$

$$G.M = \text{Antilog}(1.441)$$

$$\boxed{G.M = 27.607}$$

(4)

$$H.M = \frac{\Sigma f}{\Sigma (f/m)} = \frac{219}{8.655}$$

$$\boxed{H.M = 25.303}$$

logical Relationship of A.M, G.M and H.M

$$A.M \geq G.M \geq H.M$$

$$\boxed{29.34 > 27.67 > 25.303}$$

logical Relationship Justified

Scanned with CamScanner

Q 2

a

x	f	c.f
1	4	4
2	13	17
3	9	26
4	4	30
5	1	31
	$\Sigma f = 31$	

$n = \Sigma f = 31$

n is odd

Median is the central value

Position of central value $= \frac{n+1}{2} = \frac{31+1}{2}$

$= 16^{th}$ term

from the above table we can see that the 16^{th} term is 2.

So

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

Mode

Mode is the most frequent term from the table we can see that the term having the ~~large~~ Highest frequency of 13 is 2.

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

Q3 (b)

Class interval	f	Class Boundaries	cf	Mid Point
0 – 9	2	-0.5 – 9.5	2	4.5
10 – 19	31	9.5 – 19.5	33	14.5
20 – 29	73 f_0	19.5 – 29.5	106	24.5
30 – 39	85 f_1	29.5 – 39.5	191	34.5
40 – 49	28 f_2	39.5 – 49.5	219	44.5

Median lies Here

← Mode also lies here

$n = \Sigma f = 219$

Median

$\left(\frac{n+1}{2}\right)$th term OR $\frac{n}{2}$ th term

$\left(\frac{219+1}{2}\right)$th term

$\frac{219}{2}$ th term

109.5th term

$\left(\frac{220}{2}\right)$th term

110th term

$L = 29.5$, $h = 10$

$f = 85$, $\Sigma f = 219$

$C = 106$

$$\text{Median} = L + \frac{h}{f}\left(\frac{\Sigma f}{2} - C\right)$$

$$\text{Median} = 29.5 + \frac{10}{85}\left(\frac{219}{2} - 106\right)$$

[redacted]

8

Median $= 29.5 + 0.1176(3.5)$

Median $= 29.5 + 0.4118$

Median $= 29.9118$

$\boxed{\text{Median} \simeq 30}$

$L_1 = 29.5$ $f_1 = 85$

$i = 10$ $f_2 = 28$

$f_0 = 73$

$$\text{Mode} = L_1 + \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right) i$$

$$\text{Mode} = 29.5 + \left(\frac{85-73}{2(85)-73-28}\right) \times 10$$

$$\text{Mode} = 29 + \left(\frac{12}{69} \times 10\right)$$

Mode $= 29 + 1.739$

$\boxed{\text{Mode} = 30.74}$

Q3. a

x	f	cf
1	4	4
2	13	17
3	9	26
4	4	30
5	1	31

$\Sigma f = 31$

$N = \Sigma f$.

for Q_1

$Q_1 = \left(\frac{N+1}{4}\right)^{th}$ term

$Q_1 = \left(\frac{31+1}{4}\right)^{th}$ term

$Q_1 = 8^{th}$ term

8^{th} term is 2

so $\boxed{Q_1 = 2}$

for Q_3

$Q_3 = 3\left(\frac{N+1}{4}\right)^{th\ term} = 3\left(\frac{31+1}{4}\right)^{th}$ term

$Q_3 = 3(8) = 24^{th}$ term

24th term is 3

So $\boxed{Q_3 = 3}$

Quartile Range

$Q.R = Q_3 - Q_1$

$Q.R = 3 - 2$

$\boxed{Q.R = 1}$

Semi Inter Quartile Range

$$S.I.Q.R = \frac{Q_3 - Q_1}{2}$$

$$S.I.Q.R = \frac{3-2}{2} = 1/2$$

$\boxed{S.I.Q.R = 1/2}$

Q3 b

x	f	fx	$\bar{x}$	$x-\bar{x}$	$(x-\bar{x})^2$	$f(x-\bar{x})^2$
1	4	4	2.516	-1.516	2.298256	9.193024
2	13	26	2.516	-0.516	0.266256	3.461328
3	9	27	2.516	0.484	0.234256	2.108304
4	4	16	2.516	1.484	2.202256	8.809024
5	1	5	2.516	2.484	6.17	6.17
	$\Sigma f=31$	$\Sigma fx=78$				

$$\Sigma f(x-\bar{x})^2 = 29.74168$$

$$\text{Variance} = \frac{\Sigma f(x-\bar{x})^2}{\Sigma f - 1} = \frac{29.74168}{31-1}$$

$$\boxed{\text{Variance} = 0.99139}$$

$$C.V = \frac{\sqrt{\text{Variance}}}{\bar{x}} \times 100$$

$$C.V = \frac{0.99139}{2.516} \times 100$$

$$\boxed{C.V = 39.40}$$

Question No 4:

Question No: 04

Write down the short notes on the followings:

- **Range:**
  - it is the difference between the highest observation in series and the lowest observation in series. It is an important measure of dispersion. However, for less number of observations its value is more accurate, i.e. the use of range be limited to a maximum of ten observations. The standard deviation is used when a more precise measure is desired.

- **Quartile Range:**
  - A quartile is a statistical term describing a division of observations into four defined intervals based upon the values of the data and how they compare to the entire set of observations.
  - The quartile measures the spread of values above and below the mean by dividing the distribution into four groups.
  - Key takes are:
    - The quartile measures the spread of values above and below the mean by dividing the distribution into four groups. A quartile divides data into three points a lower quartile, median, and upper quartile to form four groups of the data set. Quartiles are used to calculate the interquartile range, which is a measure of variability around the median.
  - How Quartiles Work:
    - Just like the median divides the data into half so that 50% of the measurement lies below the median and 50% lies above it, the quartile breaks down the data into quarters so that 25% of the measurement are less than the lower quartile, 50% are less than the mean, and 75% are less than the upper quartile.
    - A quartile divides data into three points a lower quartile, median, and upper quartile to form four groups of the data set. The lower quartile or first quartile is denoted as Q1 and is the middle number that falls between the smallest value of the data set and the median. The second quartile, Q2, is also the median. The upper or third quartile, denoted as Q3, is the central point that lies between the median and the highest number of the distribution.
  - Each quartile contains 25% of the total observations. Generally, the data is arranged from smallest to largest:
    - First quartile: the lowest 25% of numbers
    - Second quartile: between 25.1% and 50% (up to the median)
    - Third quartile: 51% to 75% (above the median)
    - Fourth quartile: the highest 25% of numbers.

- **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.

- **Variance:** is measures of variability. The variance is computed as the average squared deviation of each number from its mean. Calculating the variance is an important part of many statistical applications and analyses. It is the first step in calculating the standard deviation.
  - **Formula is:**
    - The formula (in summation notation) for the variance in a population is:

$$\sigma^2 = \frac{\sum (X-\mu)^2}{N}$$

where μ is the mean and N is the number of scores.

    - When the variance is computed in a sample, the statistic:

$$S^2 = \frac{\sum (X-M)^2}{N}$$

(where M is the mean of the sample) can be used. S^2 is a biased estimate of σ^2, however.

  - By far the most common formula for computing variance in a sample is:

$$s^2 = \frac{\sum (X-M)^2}{N-1}$$

- **Standard Deviation:**
  - The standard deviation is a reference value that measures the dispersion in the data. It is best viewed as an index that is defined by the formula. The smaller the value of the standard deviation, the better the quality, because the distribution is more closely compacted around the central value.

- **Coefficient of Variation:**
  - The coefficient of variation (CV) is a statistical measure of the dispersion of data points in a data series around the mean. The coefficient of variation represents the ratio of the standard deviation to the mean, and it is a useful statistic for comparing the degree of variation from one data series to another, even if the means are drastically different from one another.

- In finance, the coefficient of variation allows investors to determine how much volatility, or risk, is assumed in comparison to the amount of return expected from investments.
- The lower the ratio of the standard deviation to mean return, the better risk-return trade-off.
- Formula;

$$CV = \sigma / \mu$$

where:

σ=standard deviation

μ=mean

- if the expected return in the denominator of the coefficient of variation formula is negative or zero, the result could be misleading.

::::END::::