

Program: - BBA/MMC
 Paper: - Basic Statistics
 Instructor: - Raza Ahmad P
 Student Name: - M. Jawar Khan Afridi
 Student ID: - 12967

Q.1)
 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	$9 \log 3 = 4.244$	$9/3 = 3$
4	4	16	$4 \log 4 = 2.408$	$4/4 = 1$
5	1	5	$1 \log 5 = 0.699$	$1/5 = 0.2$
		$\Sigma f = 31$	$\Sigma f \log x = 11.314$	$\Sigma f/x = 14.7$

Arithmetic mean

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

$$A.M = 2.516$$

Geometric mean

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

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

ID: 12967

Q 1a

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

$$G.M = 2.317$$

Harmonic Mean

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

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

$$H.M = 2.1088$$

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

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

$$2.516 > 2.317 > 2.1088$$

Hence Proved

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

is justified.

Class Interval Marks	Frequency f	mid-point	F.m	F x log m	f/m
0 - 9	2	4.5	2 x 4.5 = 9	2 x log 4.5 = 1.344	2/4.5 = 0.444
10 - 19	31	14.5	31 x 14.5 = 449.5	31 x log 14.5 = 3.602	31/14.5 = 2.138
20 - 29	73	24.5	73 x 24.5 = 1786.5	73 x log 24.5 = 4.049	73/24.5 = 2.978
30 - 39	88	34.5	88 x 34.5 = 3036	88 x log 34.5 = 4.535	88/34.5 = 2.551
40 - 49	28	44.5	28 x 44.5 = 1246	28 x log 44.5 = 4.618	28/44.5 = 0.629
	$\Sigma f = 219$		$\Sigma f m = 6125.5$	$\Sigma (f \times \log m) = 315.586$	$\Sigma (f/m) = 8.655$

Arithmetic mean

$$A.M = \frac{\Sigma (f m)}{\Sigma f} = \frac{6125.5}{219} = 29.0340$$

$A.M = 29.0340$

Q1(b)

Page 4

12967

Geometric mean

$$G.M = \text{Anti log} \left(\frac{\sum f \log m}{\sum f} \right)$$

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

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

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

$$G.M = 27.607$$

Harmonic mean

$$H.M = \frac{\sum f}{\sum (f/m)}$$

$$H.M = \frac{219}{8.655}$$

$$H.M = 25.303$$

Logical Relationship of AM, GM, HM
 $AM > GM > HM$ | $29.240 > 27.607 > 25.303$

ID: 12967

Page (5)

Q20A

x	f	C.F
1	4	4
2	13	17
3	9	26
4	4	30
5	1	31
<hr/>		
	$\Sigma f = 31$	

$$n = \Sigma f = 31$$

Median is the central value

$$\begin{aligned} \text{Position of central value} &= \frac{n+1}{2} = \frac{31+1}{2} \\ &= 16^{\text{th}} \text{ term} \end{aligned}$$

From the above table we can see that 16th term is 2

So the median will

$$\text{Median} = 2$$

mode

mode is the most frequent term from the table we can see that the term having the highest frequency of 3 is 2

So mode = 2

Question # 2 (b) (B)

Class Interval	f	Class Boundaries	cf	Mid-point
0-9	2	0.5-9.5	2	4.5
10-19	37	9.5-19.5	33	14.5
20-29	338	19.5-29.5	106	24.5
30-39	851	29.5-39.5	191	34.5
40-49	285	39.5-49.5	219	44.5

medicines
mode too

median

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

$$\frac{2}{2}$$

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

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

Amu

or

$$\frac{n}{2} \text{th term}$$

$$\frac{2}{2}$$

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

$$109.5 \text{th term}$$

110th term

2000

Date 6 19967

Pg 7

$$\text{median} = L + h \left(\frac{\sum f - c}{f} \right)$$

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

$$L = 29.5 \quad h = 10 \quad f = 85$$

$$\sum f = 219 \quad c = 106$$

$$\text{median} = 29.5 + 0.1176(35)$$

$$\text{median} = 29.5 + 0.4118$$

$$\text{median} = 29.9118$$

$$\text{median} = 30$$

median
3 ms

$$L = 29.5 \quad f_1 = 85$$
$$i = 10 \quad F_2 = 28$$
$$f_0 = 73$$

$$\text{mode} = L + \left(\frac{F_1 - F_0}{2F_1 - f_0 - F_2} \right) i$$

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

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

$$\text{mode} = 29 + 17.39$$

Page 8 Pg 8

Page 6

12967

Page 8 Pg 8

mode 30.74.

12967 ppg 9

Q 3(a)

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

$$\Sigma f = 31$$

$$n = \Sigma f$$

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

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

$$Q_1 = 8^{\text{th}} \text{ term}$$

8th term is 2

$$\text{So } Q_1 = 2$$

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

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

12967

(10)

Page 10

24th term is 3
So $Q_3 = 3$

Quantile Range

$$Q.R = Q_3 - Q_1$$

$$Q.R = 3 - 2$$

$$Q.R = 1$$

Semi Inter Quantile Range

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

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

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

(11)

12967

7

Q36

x	f	fx	\bar{x}	$x - \bar{x}$	$(x - \bar{x})^2$
1	4	4	2.516	-1.516	2.298256
2	13	26	2.516	-0.516	0.266256
3	9	27	2.516	0.484	0.234256
4	4	16	2.516	1.484	2.202256
5	1	5	2.516	2.484	6.17
$\Sigma f = 37$		$\Sigma fx = 78$			

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

$$= \frac{29.74168}{37 - 1}$$

$$\text{variance} = 0.97135$$

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

$$C.V = \frac{0.99135}{2.516} \times 100$$

$$C.V = 39.40$$

$$\begin{array}{r} f(x - \bar{x})^2 \\ 9 \cdot 193024 \\ 3 \cdot 4616328 \\ 2 \cdot 108304 \\ 8 \cdot 809024 \\ 6 \cdot 17 \end{array}$$

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

Q#4 write a short note on the following

1 Range

2 Quartile Range

3 Semi Inter Quartile Range

4 Variance

5 Standard Deviation

6 Coefficient of varianction

Range

Let's start with the range because it is the most straightforward measure of variability to calculate and the simplest to understand. The range of a dataset is the difference between the largest and smallest values in that dataset. For example, in the two datasets below, dataset 1 has a range of $20 - 38 = 18$ while dataset 2 has a range of $11 - 52 = 41$. Dataset 2 has a broader range and, hence, more variability than dataset 1.

Dataset 1	Dataset 2
20	11
21	16
22	19
25	23
26	25
29	32
33	39
34	46
38	52

While the range is easy to understand, it is based on only the two most extreme values in the dataset, which makes it very susceptible to [outliers](#). If one of those numbers is unusually high or low, it affects the entire range even if it is atypical.

Additionally, the size of the dataset affects the range. In general, you are less likely to observe extreme values. However, as you increase the [sample](#) size, you have more opportunities to obtain these extreme values. Consequently, when you draw random samples from the same population, the range tends to increase as the sample size increases. Consequently, use the range to compare variability only when the sample sizes are similar.

The Interquartile Range (IQR) . . . and other Percentiles

The interquartile range is the middle half of the data. To visualize it, think about the [median](#) value that splits the dataset in half. Similarly, you can divide the data into quarters. [Statisticians](#) refer to these quarters as quartiles and denote them from low to high as Q1, Q2, and Q3. The lowest quartile (Q1) contains the quarter of the dataset with the smallest values. The upper quartile (Q4) contains the quarter of the dataset with the highest values. The interquartile range is the middle half of the data that

is in between the upper and lower quartiles. In other words, the interquartile range includes the 50% of data points that fall between Q1 and Q3.

The interquartile range is a robust measure of variability in a similar manner that the median is a robust measure of central tendency. Neither measure is influenced dramatically by outliers because they don't depend on every value. Additionally, the interquartile range is excellent for [skewed](#) distributions, just like the median. As you'll learn, when you have a normal distribution, the standard deviation tells you the percentage of observations that fall specific distances from the mean. However, this doesn't work for skewed distributions, and the IQR is a great alternative.

I've divided the dataset below into quartiles. The interquartile range (IQR) extends from the low end of Q2 to

Variance

Variance is the average squared difference of the values from the mean. Unlike the previous measures of variability, the variance includes all values in the calculation by comparing each value to the mean. To calculate this statistic, you calculate a set of squared differences between the data points and the mean, sum them, and then divide by the number of observations. Hence, it's the average squared difference.

There are two formulas for the variance depending on whether you are calculating the variance for an entire population or using a sample to [estimate](#) the population variance. The equations are below, and then I work through an example in a table to help bring it to life.

Population variance

The formula for the variance of an entire population is the following:

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

In the equation, σ^2 is the population [parameter](#) for the variance, μ is the parameter for the population mean, and N is the number of data points, which should include the entire population.

Sample variance

To use a sample to estimate the variance for a population, use the following formula. Using the previous equation with sample data tends to underestimate the variability. Because it's usually impossible to measure an entire population, [statisticians](#) use the equation for sample variances much more frequently.

Standard Deviation

The standard deviation is the standard or typical difference between each data point and the mean. When the values in a dataset are grouped closer together, you have a smaller standard deviation. On the other hand, when the values are spread out more, the standard deviation is larger because the standard distance is greater.

Conveniently, the standard deviation uses the original units of the data, which makes interpretation easier. Consequently, the standard deviation is the most widely used measure of variability. For example, in the pizza delivery example, a standard deviation of 5 indicates that the typical delivery time is plus or minus 5 minutes from the mean. It's often reported along with the mean: 20 minutes (s.d. 5).

The standard deviation is just the square root of the variance. Recall that the variance is in squared units. Hence, the square root returns the value to the natural units. The symbol for the standard deviation as a population parameter is σ while s represents it as a sample estimate. To calculate the standard deviation, calculate the variance as shown above, and then take the square root of it. Voila! You have the standard deviation!

Coefficient of variation

The coefficient of variation (CV) is a measure of relative variability. It is the [ratio](#) of the [standard deviation](#) to the [mean \(average\)](#). For example, the expression "The standard deviation is 15% of the mean" is a CV.

The CV is particularly useful when you want to compare results from two different surveys or tests that have different measures or values. For example, if you are comparing the results from two tests that have different scoring mechanisms. If sample A has a CV of 12% and sample B has a CV of 25%, you would say that sample B has more variation, relative to its mean.

Formula

The **formula for the coefficient of variation** is:

Coefficient of Variation = (Standard Deviation / Mean) * 100.

In symbols: $CV = (SD / \bar{X}) * 100$.

Multiplying the coefficient by 100 is an optional step to get a percentage, as opposed to a decimal.

Coefficient of Variation Example

A researcher is comparing two multiple-choice tests with different conditions. In the first test, a typical multiple-choice test is administered. In the second test, alternative choices (i.e. incorrect answers) are randomly assigned to test takers. The results from the two tests are:

	Regular Test	Randomized Answers
Mean	59.9	44.8
SD	10.2	12.7

Trying to compare the two test results is challenging. Comparing standard deviations doesn't really work, because the *means* are also different. Calculation using the formula $CV=(SD/Mean)*100$ helps to make sense of the data:

	Regular Test	Randomized Answers
Mean	59.9	44.8
SD	10.2	12.7
CV	17.03	28.35

Looking at the standard deviations of 10.2 and 12.7, you might think that the tests have similar results. However, when you adjust for the difference in the means, the results have more significance:

Regular test: CV = 17.03

Randomized answers: CV = 28.35

The coefficient of variation can also be used to compare **variability** between different measures. For example, you can compare IQ scores to scores on the Woodcock-Johnson III Tests of Cognitive Abilities.

Note: The Coefficient of Variation should only be used to compare positive data on a [ratio scale](#). The CV has little or no meaning for measurements on an [interval scale](#). Examples of interval scales include temperatures in Celsius or Fahrenheit, while the Kelvin scale is a ratio scale that starts at zero and cannot, by definition, take on a negative value (0 degrees Kelvin is the absence of heat).

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.

- Q1 is the "middle" value in the *first* half of the rank-ordered data set.
- Q2 is the [median](#) value in the set.
- Q3 is the "middle" value in the *second* half of the rank-ordered data set.

The interquartile range is equal to Q3 minus Q1.

For example, consider the following numbers: 1, 3, 4, 5, 5, 6, 7, 11. Q1 is the middle value in the first half of the data set. Since there are an even number of data points in the first half of the data set, the middle value is the average of the two middle values; that is, $Q1 = (3 + 4)/2$ or $Q1 = 3.5$. Q3 is the middle value in the second half of the data set. Again, since the second half of the data set has an

even number of observations, the middle value is the average of the two middle values; that is, $Q3 = (6 + 7)/2$ or $Q3 = 6.5$. The interquartile range is $Q3$ minus $Q1$, so $IQR = 6.5 - 3.5 = 3$.

Semi Inter Quartile Range

Briefly, the **semi-interquartile range** is a measure of the dispersion or spread of a variable; it is the distance between the 1st **quartile** and the 3rd **quartile**, halved. It is common to describe a variable using a measure of central tendency, or average, most commonly the mean or median.

The **interquartile range** is equal to $Q3$ minus $Q1$. For **example**, consider the following numbers: 1, 3, 4, 5, 5, 6, 7, 11. $Q1$ is the middle value in the first half of the data set. ... The **interquartile range** is $Q3$ minus $Q1$, so $IQR = 6.5 - 3.5 = 3$

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.

Featured snippet from the web

Unlike range and quartiles, the **variance** combines all the values in a data set to produce a measure of spread. ... It is calculated as the average squared deviation of each number from the mean of a data set. For **example**, for the numbers 1, 2, and 3 the mean is 2 and the **variance** is 0.667