Name: Chandrakant B Thakur

Statistics Assignment 2

1. How can we figure out what the interquartile range is?

Interquartile Range (IQR)
The quartiles of a ranked set of data values are three points which divide the data into exactly four equal parts, each part comprising of quarter data.

1. Q1 is defined as the middle number between the smallest number and the median of the data set.
2. Q2 is the median of the data.
3. Q3 is the middle value between the median and the highest value of the data set.

The interquartile range IQR tells us the range

where the bulk of the values lie. The interquartile

range is calculated by subtracting the first quartile

from the third quartile.

IQR = Q3 - Q1

Uses
1. Unlike range, IQR tells where the majority of data lies and is thus preferred over range.
2. IQR can be used to identify outliers in a data set.
3. Gives the central tendency of the data.

2. What exactly is the value of the 5-number theory?

The 5 number summary is an exploratory data analysis tool that provides insight into the distribution of values for one variable. Collectively, this set of statistics describes where data values occur, their central tendency, variability, and the general shape of their distribution.

The five number summary provides this information using various descriptive statistics. These statistics are all order statistics—each one describes where a particular value falls in the distribution. The five statistics in this summary are the following, from highest to lowest data values:

- Highest value in the dataset.
- Third quartile (Q3)—greater than 75% of the values in the dataset
- Median or second quartile (Q2)—splits the dataset in half.
- First quartile (Q1)—greater than 25% of the values.
- Lowest value in the dataset.

3. What is the relationship between standard deviation and variance?

Variance and Standard deviation Relationship

Variance is equal to the average squared deviations from the mean, while standard deviation is the number's square root. Also, the standard deviation is a square root of variance. Both measures exhibit variability in distribution, but their units vary: Standard deviation is expressed in the same units as the original values, whereas the variance is expressed in squared units.

Formulas:

The variance of a set of *n* equally likely values can be written as:

$$\operatorname{Var}(X) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2.$$

The standard deviation is the square root of the variance:

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2}.$$

4. What does the difference between variance and standard deviation mean?

Standard deviation and variance are statistical measures of dispersion of data, i.e., they represent how much variation there is from the average, or to what extent the values typically "deviate" from the mean (average). A variance or standard deviation of zero indicates that all the values are identical.

Variance is the mean of the squares of the deviations (i.e., difference in values from the mean), and the standard deviation is the square root of that variance. Standard deviation is used to identify outliers in the data.

Standard Deviation	Variance	
Mathematical Formula	Square root of Variance	Average of the squares of deviations of each value from the mean in a sample.
Symbol	Greek letter sigma - σ	No dedicated symbol; expressed in terms of standard deviation or other values.
Values in relation to given data set	Same scale as values in the given data set; therefore,	Scale larger than the values in the given data set; not

Name: Chandrakant B Thakur

Standard Deviation	Variance	
	expressed in the same units.	expressed in the same unit as the values themselves.
Are Values Negative or Positive?	Always non-negative	Always non-negative
Real World Application	Population sampling; identifying outliers	Statistical formulas, finance.

5. When is it appropriate to refer to a skewed data distribution?

A skewed distribution is asymmetric, meaning it has a long "tail", and there is no value that gives us a mirror image. Skewness is a number that measures the asymmetry of a skewed distribution. A symmetric distribution has zero skewness, but zero skewness does not imply a symmetric distribution.

Skewness is a numerical measure of the asymmetry of a skewed distribution. We can measure skew for both unimodal (one mode) and multimodal (more than one mode) data sets.

A symmetric distribution has the same mean and median, and it also has zero skewness. However, zero skewness does not always mean that a distribution is symmetric.

A symmetric unimodal distribution has the same mean, median, and mode, and it also has zero skewness.

There are two basic types of skewed distributions: positively skewed (right-skewed) and negatively skewed (left-skewed). The words right and left tell you which direction the long tail of the distribution is pointing.