STATISTICS WORKSHEET-4

Ques1:

The Central Limit Theorem is a statistical theory that states that if we take a sufficiently large sample size from population with a finite level of variance, the mean of samples from the population will be roughly equal to the population.

CLT is important because it assumes the sample distribution of mean will be normal most of the time.

Ques2:

Sampling is a process in statistical analysis where we select a subset of the population and do a research on it.

There are two types of sampling:

- Probability sampling
- Non-probability sampling

Ques3:

Type I error occurs when a null hypothesis is true but we reject it and type II error occurs when a null hypothesis is false but we fails to reject (means accept) it.

Type I error is better than the type II error.

Ques4:

Normal distribution is a probability bell curve. It works only on numerical data not on categorical data. In normal distribution mean is zero and standard deviation.

Ques5:

Covariance: A systematic relationship between a pair of random variables where in a change in one variable reciprocal by an equivalent change in another variable. Covariance can be positive, negative or zero. Range is between minus infinity to plus infinity.

Correlation: It measure which determines the change in one variable due to change in other variable. It can be positive correlation or negative correlated. Range is between -1 to +1.

Ques6:

In Univariate, one variable is analysed at a time.

In Bivariate, two variables are analysed together and examine the association between them.

And in Multivariate, there are more than two variables are analyse at a time.

Ques7:

In sensitivity we determine how the target variables are affected based on change in other variables known as input variable.

We calculate sensitivity by

$$\text{Sensitivity} = \frac{TP}{TP+FN}$$

Ques8:

Making a guess based on assumption without scientific proof or explaining the situation based on reasonable assumption.

H0 is the null hypothesis, which means that our assumption doesn't change over the period of time.

H1 is the alternative hypothesis, which means that our assumption does change over the period of time.

For two tail test H0 and H1 is:

$H0 : \mu = \mu 0$

$H1 : \mu \neq \mu 0$

Ques9:

Quantitaive data are the numeric variables (e.g. height, weight, temperature etc). Quantitative data are the categorical variables. Qualitative data are represented by a name, symbol or a number codes.

Ques10:

Range is the difference between maximum and minimum values.

Interquartile range is the difference between third quartile (Q3) and first quartile (Q1).

Ques11:

Bell shape curve is a type of graph that is use to visualize the distribution of a variable, also known as normal distribution. This means that the data is proper and have no skewness in it.

Ques12:

Interquartile range is use to find the outliers.

Ques13:

The p-value describes how likely you are to have found particular set of observation if null hypothesis were true.

Ques14:

The binomial distribution with parameters n and p is the discrete probability distribution of the number of success in a sequence of n independent variable. The formula is

$$P(x) = nCx\ p^x\ q^{n-x}$$

Ques15:

ANOVA is used to compare difference of means among more than two groups. In this technique we check the variations. ANOVA compares the amount of variation between the groups with the amount of variation within the groups.

It's applications are:

- Null hypothesis, typically is that all means are equal.
- The independent variables are categorical.
- Dependent variables are continuous.