# STATISTICS WORKSHEET-4

1. The central limit theorem states that if we have a population with mean µ and standard deviation σ and take sufficiently large random samples from the population with replacement, then the distribution of the sample mean is asymptotically normal.

   This is useful since the researcher never knows which mean in the sampling distribution corresponds to the population mean, but by taking numerous random samples from a population, the sample means will cluster together, allowing the researcher to obtain a very accurate estimate of the population mean.

2. When we conduct research about a group of people, it's rarely possible to collect data from every person in that group. Instead, you select a sample. The sample is the group of individuals who will actually participate in the research.
   To draw valid conclusions from your results, you have to carefully decide how you will select a sample that is representative of the group as a whole. There are two types of sampling methods:
   **Probability** sampling involves random selection, allowing you to make strong statistical inferences about the whole group.
   **Non-probability** sampling involves non-random selection based on convenience or other criteria, allowing you to easily collect data.
3. In statistics, a Type I error is a false positive conclusion, while a Type II error is a false negative conclusion.
   The probability of making a Type I error is the significance level, or alpha (α), while the probability of making a Type II error is beta (β). These risks can be minimized through careful planning in your study design.
   You decide to get tested for COVID-19 based on mild symptoms. There are two errors that could potentially occur:
   **Type I error** (false positive): the test result says you have coronavirus, but you actually don't.
   **Type II error** (false negative): the test result says you don't have coronavirus, but you actually do.

4. Normal distribution, also known as the Gaussian distribution, is a probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean.
   In graphical form, the normal distribution appears as a "bell curve".

5. Covariance and correlation are two terms that are opposed and are both used in statistics and regression analysis. Covariance shows you how the two variables differ, whereas correlation shows you how the two variables are related. Here, in this tutorial, you will explore covariance and correlation, which will help you understand the difference between covariance and correlation
6. • Univariate statistics summarize only one variable at a time.
   • Bivariate statistics compare two variables.
   • Multivariate statistics compare more than two variables.

**7.** Sensitivity: The sensitivity of a test is its ability to determine the patient cases correctly. To estimate it, we should calculate the proportion of true positive in patient cases. Mathematically, this can be stated as:

**Sensitivity=TP/TP+FN**

**8.** Hypothesis testing is formulated in terms of two hypotheses:
• H0: the null hypothesis;
• H1: the alternate hypothesis.
The hypothesis we want to test is if H1 is “likely” true. So, there are two possible outcomes: Reject H0 and accept H1 because of sufficient evidence in the sample in favor or H1; • Do not reject H0 because of insufficient evidence to support H1.

**9.** Quantitative data are measures of values or counts and are expressed as numbers.They are data about numeric variables (e.g. how many; how much; or how often).
Qualitative data are measures of 'types' and may be represented by a name, symbol, or a number code.It is about categorical variables (e.g. what type).

**10.** The range is the difference between the largest and smallest values in a data set.
**11.** A bell curve is a graph depicting the normal distribution, which has a shape reminiscent of a bell. The top of the curve shows the mean, mode, and median of the data collected. Its standard deviation depicts the bell curve's relative width around the mean.
**12.** Z-Score method
**13.** The p-value is a number, calculated from a statistical test, that describes how likely you are to have found a particular set of observations if the null hypothesis were true. P-values are used in hypothesis testing to help decide whether to reject the null hypothesis.
**14.** P (X) = nCxp xq n-x
P = binomial probability
x = number of times for a specific outcome within n trials
{n Cx } = number of combinations
p = probability of success on a single trial
q = probability of failure on a single trial
n = number of trials

**15.** ANOVA checks the impact of one or more factors by comparing the means of different samples. We can use ANOVA to prove/disprove if all the medication treatments were equally effective or not. Another measure to compare the samples is called a t-test. When we have only two samples, t-test and ANOVA give the same results.