# ASSIGNMENT-3

## STATISTICS

## ANSWER SHEET

---

**1. Which of the following is the correct formula for total variation?**

a) Total Variation = Residual Variation – Regression Variation
b) Total Variation = Residual Variation + Regression Variation
c) Total Variation = Residual Variation * Regression Variation
d) All of the mentioned

**B**

**2. Collection of exchangeable binary outcomes for the same covariate data are called outcomes.**

a) random
b) direct
c) binomial
d) none of the mentioned

**C**

**3. How many outcomes are possible with Bernoulli trial?**

a) 2
b) 3
c) 4
d) None of the mentioned

**A**

**4. If Ho is true and we reject it is called**

a) Type-I error
b) Type-II error
c) Standard error
d) Sampling error

**A**

**5. Level of significance is also called:**

a) Power of the test
b) Size of the test
c) Level of confidence
d) Confidence coefficient

**B**

**6. The chance of rejecting a true hypothesis decreases when sample size is:**

a) Decrease

b) Increase

c) Both of them

d) None

B

**7. Which of the following testing is concerned with making decisions using data?**

a) Probability

b) Hypothesis

c) Causal

d) None of the mentioned

B

**8. What is the purpose of multiple testing in statistical inference?**

a) Minimize errors

b) Minimize false positives

c) Minimize false negatives

d) All of the mentioned

D

**9. Normalized data are centred at and have units equal to standard deviations of the original data**

a) 0

b) 5

c) 1

d) 10

A

**10. What Is Bayes' Theorem?**

the Bayes theorem establishes the contingent likelihood of an outcome A in the presence of an earlier occurrence B. The Bayes Rule or Bayes Law are other names for the Bayes Theorem. It is a technique for estimating an event's likelihood based on the prevalence of previous events.

It is applied in the transition probability calculation. Based on the hypothesis, the Bayes theorem measures probability.

The Bayes theorem asserts that the combination of the probability of B given A and the chance of A determines the contingent likelihood of an outcome A given the presence of some other event B.

$$\mathbf{P(A \mid B) = P(B \mid A) / P(A)P(B)}$$

The likelihood that something will be happening is expressed here as P(A), where P(A) = how probable something will happen (based on prior information).

**11. What is z-score?**

A z-score measures the distance between a data point and the mean using standard deviations. Z-scores can be positive or negative. The sign tells you whether the observation is above or below the mean. For example, a z-score of +2 indicates that the data point falls two standard deviations above the mean, while a -2 signifies it is two standard deviations below the mean. A z-score of zero equals the mean. Standardizing the raw data by transforming them into z-scores provides the following benefits:

**(a)** Understand where a data point fits into a distribution.
**(b)** Compare observations between dissimilar variables.
**(c)** Identify outliers.
**(d)** Calculate probabilities and percentiles using the standard normal distribution.

**12. What is t-test?**

T Test is one of the foundational statistical tests. It is used to compare the means of two groups and determine if the difference is statistically significant. It is a very common test often used in data and statistical analysis.

Let's suppose, you have measurements from two different groups, say, you are measuring the marks scored by students from a class you attended special coaching versus those who did not. And you want to determine if the scores of those who attending the coaching is significantly higher than those who did not. You can use the T-Test to determine this.
This test assumes that the test statistic, which is often the 'sample mean', should follow a normal distribution. Let's look at the types of T-Tests.

## types of T-Test -

**One Sample T-Test**
Tests the null hypothesis that the population mean is equal to a specified value `μ` based on a sample mean `x̄`.

**Two Sample Independent T-Test**
Tests the null hypothesis that two sample means x̄1 and x̄2 are equal. Compares the means of two numeric variables obtained from two independent groups.

**Two Sample Dependent T-Test (aka Paired T-Test)**
Compare the means of two numeric variables of same size where the observations from the two variables are paired.

**13. What is percentile?**
A percentile is a measure used in statistics indicating the value below which a given percentage of observations in a group of observations fall. For example, the 20th percentile is the value (or score) below which 20% of the observations may be found.

**14. What is ANOVA?**
Analysis of variance (ANOVA) tests the hypothesis that the means of two or more populations are equal. ANOVAs assess the importance of one or more factors by comparing the response variable means at the different factor levels. The null hypothesis states that all population means (factor level means) are equal while the alternative hypothesis states that at least one is different.

**15. How can ANOVA help?**

The mean values can be assessed just by looking at data. But ANOVA does more than only comparing means.

Even though the mean values of various groups appear to be different, this could be due to a sampling error rather than the effect of the independent variable on the dependent variable. If it is due to sampling error, the difference between the group means is meaningless. ANOVA helps to find out if the difference in the mean values is statistically significant.

ANOVA also indirectly reveals if an independent variable is influencing the dependent variable. For example, in the above blood sugar level experiment, suppose ANOVA finds that group means are not statistically significant, and the difference between group means is only due to sampling error. This result infers that the type of medication (independent variable) is not a significant factor that influences the blood sugar level.