# STATISTICS WORKSHEET-3

Q1 to Q9 have only one correct answer. Choose the correct option to answer your question.

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

## Ans= b

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

a) random

b) direct

c) binomial

d) none of the mentioned

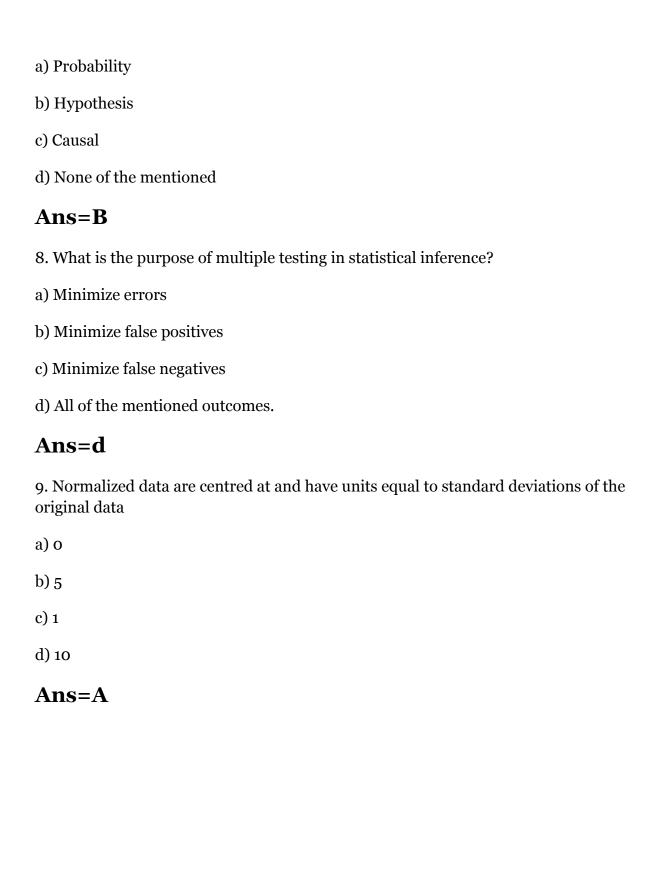## Ans= C

3. How many outcomes are possible with Bernoulli trial?

a) 2

b) 3

c) 4

d) None of the mentioned

## Ans= 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

**Ans=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

**Ans=A**

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

a) Decrease

b) Increase

c) Both of them

d) None

**Ans=A**

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

a) Probability

b) Hypothesis

c) Causal

d) None of the mentioned

## Ans=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 outcomes.

## Ans=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

## Ans=A

Q10and Q15 are subjective answer type questions, Answer them in your own words briefly.

## 10. What Is Bayes' Theorem?

BT states that the conditional probability of an event, based on the occurrence of another event, is equal to the likelihood of the second event given the first event multiplied by the probability of the first event.

BT provides a way to calculate the probability of a hypothesis based on its prior probability, the probabilities of observing various data given the hypothesis, and the observed data itself.

FORMULA=$P(Ei|A)=P(Ei\cap A)P(A)$.

## 11. What is z-score?

Z-score indicates how much a given value differs from the standard deviation. The Z-score, or standard score, is the number of standard deviations a given data point lies above or below mean. Standard deviation is essentially a reflection of the amount of variability within a given data set.

a z-score is the number of standard deviation from the mean value of the reference population (a population whose known values have been recorded, like in these charts the CDC compiles about people's weights). For example:

- A z-score of 1 is 1standard deviation above the mean.
- A score of 2 is 2 standard deviation above the mean.
- A score of -1.8 is -1.8 standard deviation *below* the mean.

A z-score tells you where the score lies on a normal distribution curve. A z-score of zero tells you the values is exactly average while a score of +3 tells you that the value is much higher than average.

12. What is t-test?

In statistics, the t-statistic is the ratio of the departure of the estimated value of a parameter from its hypothesized value to its standard error. It is used in hypothesis testing via Student's t-test. The t-statistic is used in a t-test to determine whether to support or reject the null hypothesis

A t-test can only be used when comparing the means of two groups (a.k.a. pairwise comparison). If you want to compare more than two groups, or if you want to do multiple pairwise comparisons, use an ANOVA test or a post-hoc test.

The t-test is a parametric test of difference, meaning that it makes the same assumptions about your data as other parametric tests. The t-test assumes your data:

1. are independent
2. are (approximately) normally distributed.
3. have a similar amount of variance within each group being compared (a.k.a. homogeneity of variance)

If your data do not fit these assumptions, you can try a nonparametric alternative to the t-test, such as the Wilcoxon Signed-Rank test for data with unequal variances.

13. What is percentile?

In statistics, a percentile is a term that describes how a score compares to other scores from the same set. While there is no universal definition of percentile, it is commonly expressed as the percentage of values in a set of data scores that fall below a given value.

$Px = X(n+1)/100$

Px=value at which x percentage of data lie below that value

n=total number of observation.

14. What is ANOVA?

ANOVA stands for Analysis of Variance. One-Way Analysis of Variance tells you if there are any statistical differences between the means of three or more independent groups.

Use of ANOVA:-

You might use Analysis of Variance (ANOVA) as a marketer, when you want to test a particular hypothesis. You would use ANOVA to help you understand how your different groups respond, with a null hypothesis for the test that the means of the different groups are equal. If there is a statistically significant result, then it means that the two populations are unequal (or different).

15. How can ANOVA help?

The one-way ANOVA can help you know whether or not there are significant differences between the means of your independent variables (such as the first example: age, sex, income). When you understand how each independent variable's mean is different from the others, you can begin to understand which of them has a connection to your dependent variable (landing page clicks), and begin to learn what is driving that behavior.