WORKSHEET 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?
b) Total Variation = Residual Variation + Regression Variation

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

3. How many outcomes are possible with Bernoulli trial?
a) 2

4. If Ho is true and we reject it is called
a) Type-I error

5. Level of significance is also called: a) Power of the test

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

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

8. What is the purpose of multiple testing in statistical inference?
d) All of the mentioned

9. Normalized data are centred at and have units equal to standard deviations of the original data
a) 0

Q10 and Q15 are subjective answer type questions, Answer them in your own words briefly.

10. What Is Bayes' Theorem?

Bayes Theorem provides a principled way for calculating a conditional probability. It is a deceptively simple calculation, although it can be used to easily calculate the conditional probability of events where intuition often fails.

Although it is a powerful tool in the field of probability, Bayes Theorem is also widely used in the field of machine learning. Including its use in a probability framework for fitting a model to a training dataset, referred to as maximum a posteriori or MAP for short, and in developing

models for classification predictive modeling problems such as the Bayes Optimal Classifier and Naive Bayes.

11. What is z-score?

A z-score can be placed on a **normal distribution** curve. Z-scores range from -3 standard deviations (which would fall to the far left of the normal distribution curve) up to +3 standard deviations (which would fall to the far right of the normal distribution curve). In order to use a z-score, you need to know the mean μ and also the population standard deviation σ.

12. What is t-test?

A t-test is a type of inferential statistic used to determine if there is a significant difference between the means of two groups, which may be related to certain features. It is mostly used when the data sets, like the data set recorded as the outcome from flipping a coin 100 times, would follow a normal distribution and may have unknown variances. A t-test is used as a hypothesis testing tool, which allows testing of an assumption applicable to a population.

13. What is percentile?

In statistics, a **percentile** (or a **centile**) is a score *below which* a given percentage of scores in its frequency distribution falls (exclusive definition) or a score *at or below which* a given percentage falls.

14. What is ANOVA?

An **ANOVA** test is a way to find out if survey or experiment results are significant. In other words, they help you to figure out if you need to reject the null hypothesis or accept the alternate hypothesis.

15. How can ANOVA help?

ANOVA, or its non-parametric counterparts, allow you **to** determine if differences in mean values between three or more groups are by chance or if they are indeed significantly different. **ANOVA** is particularly useful when analyzing the multi-item scales common in market research.