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

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

3. How many outcomes are possible with Bernoulli trial?
   a) 2
   b) 3
   c) 4
   d) None of the mentioned

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

5. Level of significance is also called:
   a) Power of the test
   b) Size of the test
   c) Level of confidence
   d) Confidence coefficient

6. The chance of rejecting a true hypothesis decreases when sample size is:
   a) Decrease
   b) Increase
   c) Both of them
   d) None

7. Which of the following testing is concerned with making decisions using data?
   a) Probability
   b) Hypothesis
   c) Causal
   d) None of the mentioned

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

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

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

10. What Is Bayes' Theorem?
11. What is z-score?
12. What is t-test?
13. What is percentile?
14. What is ANOVA?
15. How can ANOVA help?

10 answer: Bayes theorem 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.

11 answer: z score standardization
This technique consists of subtracting the mean of the column from each value in a column, and then dividing the result by the standard deviation of the column. The formula to achieve this is the following:
The result of standardization is that the features will be rescaled so that they'll have the properties of a standard normal distribution, as follows:
$\mu=0$
$\sigma=1$
μ is the mean and σ is the standard deviation from the mean.
In summary, the z score (also called the **standard score**) represents the number of standard deviations with which the value of an observation point or data differ than the mean value of what is observed .
12 answer: A t-test is a statistical test that is used to compare the means of two groups. It is often used in hypothesis testing to determine whether a process or treatment actually has an effect on the population of interest, or whether two groups are different from one another.
13 answer: In statistics, a percent**ile**(or a centile) is a score below which a given percentage of scores in its frequency distribution fall (exclusive definition) or a score at or below which a given percentage fall (inclusive definition)
14 answer: **Analysis of variance** (**ANOVA**) is an analysis tool used in statistics that splits an observed aggregate variability found inside a data set into two parts: systematic factors and random factors. The systematic factors have a statistical influence on the given data set, while the random factors do not.

15 answer: 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 nul hypothesis or accept the alternate hypothesis. Basically, you're testing groups **to** see if there's a difference between them.