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
Answer: b) Total Variation = Residual Variation + Regression Variation

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
Answer: c) binomial

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

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
Answer: a) Type-I 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
Answer: b) Size of the test

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

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

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
Answer: 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
Answer: a) 0

Q10and Q15 are subjective answer type questions, Answer them in your own words briefly.
10. What Is Bayes' Theorem?
Answer: Bayes' Theorem 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.
Formula: $P(A|B)=P(B)P(A \cap B)=P(B)P(A) \cdot P(B|A)$
where:P(A)= The probability of A occurringP(B)= The probability of B occurringP(A|B)=The proba bility of A given BP(B|A)= The probability of B given AP(A∩B))= The probability of both A and B occurring

11. What is z-score?
Answer: a z-score (also called a *standard score*) gives you an idea of how far from the mean a data point is. But more technically it's a measure of how many standard deviations below or above the population mean a raw score is. 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. Z-scores are a way to **compare results** to a "normal" population. Results from tests or surveys have thousands of possible results and units; those results can often seem meaningless. For example, knowing that someone's weight is 150 pounds might be good information, but if you want to compare it to the "average" person's weight, looking at a vast table of data can be overwhelming (especially if some weights are recorded in kilograms). A z-score can tell you **where** that person's weight is **compared to the average population's** mean weight.

Formula: $\mathbf{z = (x - \mu) / \sigma}$

12. What is t-test?
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. What is percentile?
Answer: 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.

14. What is ANOVA?
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. Analysts use the ANOVA test to determine the influence that independent variables have on the dependent variable in a regression study.
The formula for Analysis of Variance is: ANOVA coefficient, F= Mean sum of squares between the groups (MSB)/ Mean squares of errors (MSE). Therefore **F = MSB/MSE**

15. How can ANOVA help?
Answer: ANOVA is helpful for **testing three or more variables**. It is similar to multiple two-sample t-tests. However, it results in fewer type I errors and is appropriate for a range of issues. ANOVA groups differences by comparing the means of each group and includes spreading out the variance into diverse sources.