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?

A theory that describes the probability of an event occurring based on conditions that might be related to the event, i.e. it tells us how to calculate conditional probability.

- P(A|B) – the probability of event A occurring, given event B has occurred
- P(B|A) – the probability of event B occurring, given event A has occurred
- P(A) – the probability of event A
- P(B) – the probability of event B

P(A|B) = P(A)/ P(B). P(A|B)/ P(B)

11. What is z-score?

a z-score is just the number of standard deviations a data point is from the mean of a sample distribution. It is expressed as z = (X - mu)/sigma where X is the data point coordinate, mu is the mean of the distribution and sigma is the sample standard deviation. Works so long as you have >30 data points.

12. What is t-test?

The t-distribution is also symmetric and mound shaped but its exact shape depends on the sample size. If the sample size is large (think 1,000 plus) then the z and t distributions are essentially the same. With smaller samples the t distributions has “fatter tails” and is more shallow and flat (compared to z).

The t distribution is appropriate with continuous data when you do not know the population standard deviation (most real world cases).

In short the t-score, just like a z-score, will tell you how many standard deviations something is from the mean. It’s just that the t and z distributions are different for small samples.

13. What is percentile?

A percentile is a comparison score between a particular score and the scores of the rest of a group. It shows the percentage of scores that a particular score surpassed. For example, if you score 75 points on a test, and are ranked in the 85 th percentile, it means that the score 75 is higher than 85% of the scores.

14. What is ANOVA?

Analysis of Variance (ANOVA) is a statistical formula used to compare variances across the means (or average) of different groups. A range of scenarios use it to determine if there is any difference between the means of different groups.

15. How can ANOVA help?

We use ANOVA as a means to compare several groups mean values between themselfs. ANOVA stands f or analysis of variance, and in the process partitions the variances due to group memberships, and c ompare those to total variability. The statistics of interest is the F-value, together with associate p-value.