ASSIGNMENT-3

STATISTICS

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

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

 a) random

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:

 c) Level of confidence

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

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. The statement of Bayes Theorem is as follows: Let E1,E2,E3,...,EnE1,E2,E3,...,En be a set of events associated with a sample space S, where all events E1,E2,E3,...,EnE1,E2,E3,...,En have non-zero probability of occurrence and they form a partition of S. Let A be any event which occurs with E1orE2orE3...orEnE1orE2orE3...orEn, then according to Bayes Theorem, P(Ei|A)=P(Ei)P(A|Ei)∑nk=1P(Ek)P(A|Ek),i=1,2,3,...,nP(Ei|A)=P(Ei)P(A|Ei)∑k=1nP(Ek)P(A|Ek),i=1,2,3,...,n

● **Here Eii ∩ Ejj = ф, where i ≠ j. (i.e) They are mutually exhaustive events**

 ● **The union of all the events of the partition, should give the sample space.**

● **0 ≤ P(Eii) ≤ 1**

11. What is z-score?

   Answer: A z-score describes the position of a raw score in terms of its distance from the mean, when measured in standard deviation units. The z-score is positive if the value lies above the mean, and negative if it lies below the mean. 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.

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. 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

13. What is percentile?

Answer: 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 85th percentile, it means that the score 75 is higher than 85% of the scores

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 t- and z-test methods developed in the 20th century were used for statistical analysis until 1918, when Ronald Fisher created the analysis of variance method.12 ANOVA is also called the Fisher analysis of variance, and it is the extension of the t- and z-tests. The term became well-known in 1925, after appearing in Fisher's book, "Statistical Methods for Research Workers."3 It was employed in experimental psychology and later expanded to subjects that were more complex.

15. How can ANOVA help?

Answer: There are many types of ANOVA

1. One-way ANOVA

2. Two-way ANOVA

3. MANOVA

The different types of ANOVA are:

1. ANOVA is helpful for testing three or more variables

2. ANOVA gives most significant and influential factors

3. ANOVA also gives the most contributing factors.