STATISTICS WORKSHEET-3

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:

b) Size 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

10. What Is Bayes' Theorem?

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.

$$P(A|B) = P(B|A) * P(A)/P(B)$$

where, P(A|B) — is the probability of A given that B has already happened.

P(B|A) — is the probability of B given that A has already happened. It looks circular and arbitrary now but we will see why it works shortly.

P(A) — is the unconditional probability of A occurring.

P(B) — is the unconditional probability of B occurring.

P(A|B) is an example of a conditional probability — one that measure probability over only certain states of the world (states where B has occurred). P(A) is an example of an unconditional probability and is measured over all states of the world.

Bayes' theorem describes the probability of occurrence of an event related to any condition. It is also considered for the case of conditional probability. Bayes theorem is also known as the formula for the probability of "causes". For example: While calculating the probability of taking a blue ball from the second bag out of three different bags of balls, where each bag contains three different colour balls viz. red, blue, black. In this case, the probability of occurrence of an event is calculated depending on other conditions is known as conditional probability.

Application of Bayes' theorem in Artificial intelligence:

Following are some applications of Bayes' theorem:

* It is used to calculate the next step of the robot when the already executed step is given.

* Bayes' theorem is helpful in weather forecasting.

11. What is z-score?

Z-score is a statistical measurement that describes a value's relationship to the mean of a group of values. Z-score is measured in terms of standard deviations from the mean. If a Z-score is 0, it indicates that the data point's score is identical to the mean score. A Z-score of 1.0 would indicate a value that is one standard deviation from the mean. Z-scores may be positive or negative, with a positive value indicating the score is above the mean and a negative score indicating it is below the mean.

A measure of how many standard deviations below or above the population mean a raw score is called z score. It will be positive if the value lies above the mean and negative if it lies below the mean. It is also known as standard score. It indicates how many standard deviations an entity is, from the mean. In order to use a z-score, the mean μ and also the population standard deviation σ should be known. A z score helps to calculate the probability of a score occurring within a standard normal distribution. It also enables us to compare two scores that are from different samples. A table for the values of ϕ, indicating the values of the cumulative distribution function of the normal distribution is termed as a z score table.

The equation is given by $z = (x - \mu)/ \sigma$.

μ = mean

σ = standard deviation

x = test value

With multiple samples and want to describe the standard deviation of those sample means, the following formula is used:

$z = (x - \mu)/ (\sigma/\sqrt{n})$

Interpretation

1. If a z-score is equal to -1, then it denotes an element, which is 1 standard deviation less than the mean.

2. If a z score is less than 0, then it denotes an element less than the mean.

3. If a z score is greater than 0, then it denotes an element greater than the mean.

4. If the z score is equal to 0, then it denotes an element equal to the mean.

5. If the z score is equal to 1, it denotes an element, which is 1 standard deviation greater than the mean; a z score equal to 2 signifies 2 standard deviations greater than the mean; etc.

12. What is t-test?

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.

t test example:
Whether the mean petal length of iris flowers differs according to their species. The two different species of irises growing in a garden and measure 25 petals of each species. You can test the difference between these two groups using a t test and null and alterative hypotheses.

- The null hypothesis (H0) is that the true difference between these group means is zero.
- The alternate hypothesis (Ha) is that the true difference is different from zero.

A t test can only be used when comparing the means of two groups. The t test is a parametric test of difference, meaning that it makes the same assumptions about the data as other parametric tests. The t test assumes the data to be:

1. are independent
2. are (approximately) normally distributed
3. have a similar amount of variance within each group being compared (a.k.a. homogeneity of variance)

13. What is percentile?

The *n*th percentile of a set of data is the value at which *n* percent of the data is below it. Percentile is a score below which a given percentage k of scores in its frequency distribution falls or a score at or below which a given percentage falls.

For example, the 50th percentile (the median) is the score below which 50% of the scores in the distribution are found or at or below which 50% of the scores are found. Percentiles are expressed in the same unit of measurement as the input scores; for

example, if the scores refer to human weight, the corresponding percentiles will be expressed in kilograms or pounds.

Percentiles can be calculated using the formula n = (P/100) x N, where P = percentile, N = number of values in a data set (sorted from smallest to largest), and n = ordinal rank of a given value.

For example, a student taking a difficult exam might earn a score of 75 percent. This means that he correctly answered every three out of four questions. A student who scores in the 75th percentile, however, has obtained a different result. This percentile means that the student earned a higher score than 75 percent of the other students who took the exam. In other words, the percentage score reflects how well the student did on the exam itself; the percentile score reflects how well he did in comparison to other students.

14. What is ANOVA?

An ANOVA test is a type of statistical test used to determine if there is a statistically significant difference between two or more categorical groups by testing for differences of means using variance.

Analysis of variance (ANOVA) is a statistical technique that is used to check if the means of two or more groups are significantly different from each other. ANOVA checks the impact of one or more factors by comparing the means of different samples.

Types of ANOVA Tests

ONE-WAY ANOVA:

A one-way ANOVA (analysis of variance) has one categorical independent variable (also known as a factor) and a normally distributed continuous (i.e., interval or ratio level) dependent variable.

The independent variable divides cases into two or more mutually exclusive levels, categories, or groups.

The one-way ANOVA test for differences in the means of the dependent variable is broken down by the levels of the independent variable.

An example of a one-way ANOVA includes testing a therapeutic intervention (CBT, medication, placebo) on the incidence of depression in a clinical sample.

TWO-WAY ANOVA:

A two-way ANOVA (analysis of variance) has two or more categorical independent variables (also known as a factor), and a normally distributed continuous (i.e., interval or ratio level) dependent variable.

The independent variables divide cases into two or more mutually exclusive levels, categories, or groups. A two-way ANOVA is also called a factorial ANOVA.

An example of a factorial ANOVAs include testing the effects of social contact (high, medium, low), job status (employed, self-employed, unemployed, retired), and family history (no family history, some family history) on the incidence of depression in a population.

15. How can ANOVA help?

ANOVA can help to identify the sources of variation in a data set. This can help to improve the accuracy of data predictions and analyses. Additionally, ANOVA can help to identify relationships between different variables in a data set. This information can be used to improve data models and predictions. ANOVA can also be used to partition the total variance into between-groups and within-groups variance which is helpful with data.