WORKSHEET SET 3

STATISTICS WORKSHEET 3

OBJECTIVE TYPE QUESTIONS

Q1. B) TOTAL VARIATION = RESIDUAL VARIATION + REGRESSION VARIATION

Q2. C) BINOMIAL

Q3. A) 2

Q4. A) TYPE–I ERROR

Q5. B) SIZE OF THE TEST

Q6. B) INCREASED

Q7. B) HYPOTHESIS

Q8. D) ALL OF THE MENTIONED

Q9. A) 0

SUBJECTIVE TYPE QUESTIONS

Q10. <u>BAYE'S THEOREM</u>

It is named after **REVERAND THOMAS BAYES**

Bayes theorem states that the conditional probability of an event **A** given the occurrence of another event **B**, is equal to the product of the likelihood of **B** given **A** and the probability of **A**.

MATHEMATICAL FORMULA

Let E1, E2, E3, ..., En be a set of events associated with a sample space S, where all events E1, E2, E3, ..., En have non-zero probability of occurrence and they form a partition of S. Let A be any event which occurs with E1 or E2 or E3...or En,

then according to **Bayes Theorem**,

P(Ei | A) = { P(Ei)*P(A | Ei)/ $\sum$ P(Ek)* P(A | Ek) }, for any k = 1,2,3,...,n

APPLICATIONS

- To define the accuracy of medical test results by considering how likely any given person is to have a disease.
- Used to find the reverse probabilities if we know the conditional probability of an event.
- It can be used in the condition while answering the probabilistic queries conditioned on the piece of evidence.

EXAMPLE

A bag I contains 4 white and 6 black balls while another Bag II contains 4 white and 3 black balls. One ball is drawn at random from one of the bags, and it is found to be black. Find the probability that it was drawn from Bag I.

SOLUTION

Let E1 be the event of choosing bag I, E2 the event of choosing bag II, and A be the event of drawing a black ball.

Then,

P(E1) = P(E2) = 1/2
Also, P(A | E1) = P(drawing a black ball from Bag I) = 6/10 = 3/5

P(A | E2) = P(drawing a black ball from Bag II) = 3/7

By using Bayes' theorem, the probability of drawing a black ball from bag I out of two bags,

P(E1|A) = [P(E1)*P(A | E1)/ { P(E1)*P(A | E1)+P(E2)*P(A | E2) }]

= (0.5*0.6) / { ((0.5)*(0.6)) + ((0.5)*(3/7)) }
= 7/12

Q11.   Z SCORE

A measure of how many standard deviations below or above the population mean a raw score is called **z score**. It is also known as **standard score**. It indicates how many standard deviations an entity is away from the mean.

It can be **zero, negative, or positive**.

If a z-score is 0, it denotes that the data point's score is identical to the mean score.

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

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

FORMULA

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

where μ = mean, σ = standard deviation, x = test value

When we have multiple samples and want to describe the standard deviation of those sample means,

we use the following formula:

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

APPLICATION

- 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.
- z score is used in medical field to find how a certain new born baby's weight compares to the mean weight of all babies.
- It can be used to find how a certain shoe size compares to the mean population size.

Q12. t -TEST

A t-Test is a statistical method of comparing the means of two samples gathered from either the same group or different categories.

It is aimed at hypothesis testing, which is used to test a hypothesis pertaining to a given population.

It is the difference between population means and a hypothesized value.

The testing uses randomly selected samples from the two categories or groups.

EXPLAINATION

A T-test studies a set of data gathered from two similar or different groups to determine the probability of the difference in the result than what is usually obtained.

The accuracy of the test depends on various factors, including the distribution patterns used and the variants influencing the collected samples.

Depending on the parameters, the test is conducted, and a T-value is obtained as the statistical inference of the probability of the usual resultant being driven by chance.

For example, if one wishes to figure out if the mean of the length of petals of a flower belonging to two different species is the same, a T-test can be done. The user can select petals randomly from two other species of that flower and come to a standard conclusion. The inference obtained in the process indicates the probability of the mean differences to have happened by chance.

The final **T-test interpretation** could be obtained in either of the two ways:

- A null hypothesis signifies that the difference between the means is zero and where both the means are shown as equal.
- An alternate hypothesis implies the difference between the means is different from zero. This hypothesis rejects the null hypothesis, indicating that the data set is quite accurate and not by chance.

This T-test, however, is only valid and should be done when the mean or average of only two categories or groups needs to be compared. If the number of comparisons to be made are more than two, conducting this is not recommended.

ASSUMPTIONS

- The measurement scale used for such hypothesis testing follows a set of continuous or ordinal patterns.
- The accounted parameters and variants influencing the samples and surrounding the groups are based on the standard consideration.
- The tests are completely based on random sampling. As no individuality is maintained in the samples, the reliability is often questioned.
- When the data is plotted with respect to the T-test distribution, it should follow a normal distribution and bring about a bell-curved graph.
- For a clearer bell curve, the sample size needs to be bigger.
- The variance should be such that the standard deviations of the samples are almost equal.

TYPES OF t-TEST

The type of T-test to be conducted is decided by whether the samples to be analysed are from the same category or distinct categories.

- **One sample t-Test**
  While performing this test, the mean or average of one group is compared against the set average, which is either the theoretical value or means of the population. For example, a teacher wishes to figure out the average height of the students of class 5 and compare the same against a set value of more than 45 kgs.
- **Two sample t-Test**
  This is the test conducted when samples from two different groups, species, or populations are studied and compared. It is also known as an independent T-test. For example, if a teacher wants to compare the height of male students and female students in class 5, she would use the independent two-sample test.
- **Paired t-Test**
  This hypothesis testing is conducted when two groups belong to the same population or group. The groups are studied either at two different times or under two varied conditions.
- **Equal Variance t-Test**
  This test is conducted when the sample size in each group or population is the same or the variance of the two data sets is similar. It is also referred to as pooled T-test.
- **Unequal Variance t-Test**
  The unequal variance testing is used when the variance and the number of samples in each group are different. It is often referred to as Welch's test

APPLICATION

- The test is useful when comparing population age, length of crops from two different species, student grades, etc.

Q13.　PERCENTILE

A percentile is a measure used in statistics indicating the value below which a given percentage of observations in a group of observations fall.

Percentile, a measure used in statistics, always has a number next to it – it indicates that the person or thing being measured or evaluated is at the top of that number in percentage terms.

EXAMPLES

- Imagine a country has just 100 people, and Mr. Brown is at the 42nd percentile regarding physical strength. This means that there are 42 people physically weaker than him.
- The 20th percentile is the value (or score) below which 20% of the observations may be found.

The 25th percentile is also known as the first quartile (Q1), the 50th percentile as the median or second quartile (Q2), and the 75th percentile as the third quartile (Q3).

The range of values containing the central half of the observations is called the interquartile range: that is, the range between the 25th and 75th percentiles.

APPLICATIONS

- Doctors and health care professionals commonly use children's and infants' height and weight to assess their growth in comparison to the averages and percentiles found in growth charts.
- When setting a speed limit on road traffic, the 85th percentile traffic speed is commonly used – it is also used to determine whether the limit has been set either too low or too high.
- It is used with the median value to report data that are markedly non-normally distributed.

Q14.   ANOVA

ANOVA stands for **Analysis Of Variance**. **Ronald Fisher** created this method. ANOVA is also called the **Fisher analysis of variance**, and it is the **extension of the t- and z-tests**.

It 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. Another key part of ANOVA is that it 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.

FORMULA

F = MST / MSE

Where F = ANOVA coefficient,

MST = Mean sum of squares due to treatment

MSE = Mean sum of squares due to error

If no true variance exists between the groups, the ANOVA's F-ratio should equal close to 1.

TYPES OF ANOVA TEST

There are different types of ANOVA tests. Most common are

- **One-Way ANOVA**

  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 is also called a factorial ANOVA.
  An example of a factorial ANOVA includes 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.

  The difference between these two types depends on the number of independent variables in your test.

Q15. HOW CAN ANOVA HELP?

An ANOVA examines the relationship between a categorical and a numeric variable by judging the differences between two or more means. This analysis gives a p-value to decide whether the relationship is vital or not.
The ANOVA test allows a comparison of more than two groups at the same time to determine whether a relationship exists between them.
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.

APPLICATION

- The ANOVA test is the initial step in analysing factors that affect a given data set. Once the test is finished, an analyst performs additional testing on the methodical factors that measurably contribute to the data set's inconsistency.
- The analyst utilizes the ANOVA test results in an f-test to generate additional data that aligns with the proposed regression models.
- The result of the ANOVA formula, the F statistic (also called the F-ratio), allows for the analysis of multiple groups of data to determine the variability between samples and within samples.
- If no real difference exists between the tested groups, which is called the null hypothesis, the result of the ANOVA's F-ratio statistic will be close to 1.
- One can apply ANOVA when the data needs to be experimental.
- One should use it for small samples. And if they want to perform ANOVA for a large number of experimental designs, they should use the same sample size with various factors.
- One can test two or more variables with ANOVA. The results of ANOVA are quite similar to type I errors.
- The ANOVA is employed with test groups, subjects and within groups.
- In the field of business application, the marketing experts can test the two different marketing strategies of the business to see that one strategy is better than the other one in terms of cost efficiency and time efficiency.

There are different types of ANOVA tests. And these tests depend on a number of factors.