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?
   1. Total Variation = Residual Variation – Regression Variation

b) Total Variation = Residual Variation + Regression Variation

1. Total Variation = Residual Variation \* Regression Variation
2. All of the mentioned
3. Collection of exchangeable binary outcomes for the same covariate data are called outcomes.
   1. random
   2. direct

c) binomial

d) none of the mentioned

1. How many outcomes are possible with Bernoulli trial?

a) 2

1. 3
2. 4



1. None of the mentioned
2. If Ho is true and we reject it is called
   1. Type-I error
   2. Type-II error
   3. Standard error
   4. Sampling error
3. Level of significance is also called:

a) Power of the test

1. Size of the test
2. Level of confidence
3. Confidence coefficient
4. The chance of rejecting a true hypothesis decreases when sample size is:
   1. Decrease

b) Increase

1. Both of them
2. None
3. Which of the following testing is concerned with making decisions using data?
   1. Probability

b) Hypothesis

1. Causal
2. None of the mentioned
3. What is the purpose of multiple testing in statistical inference?
   1. Minimize errors
   2. Minimize false positives
   3. Minimize false negatives

d) All of the mentioned

1. Normalized data are centred at and have units equal to standard deviations of the original data

a) 0

1. 5
2. 1
3. 10

# Q10and Q15 are subjective answer type questions, Answer them in your own words briefly.

1. What Is Bayes' Theorem?

Ans: 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.

In other words, Bayes Theorem is the extension of Conditional probability. Conditional probability helps us to determine the probability of A given B, denoted by P(A|B). So, Bayes' theorem says if we know P(A|B) then we can determine P(B|A), given that P(A) and P(B) are known to us.

Some advantages to using Bayesian analysis include the following: It provides a natural and principled way of combining prior information with data, within a solid decision theoretical framework. You can incorporate past information about a parameter and form a prior distribution for future analysis.

Bayes' Theorem is the most important concept in Data Science. It is most widely used in Machine Learning as a classifier that makes use of Naive Bayes' Classifier. It has also emerged as an advanced algorithm for the development of Bayesian Neural Networks.

1. What is z-score?

Ans: **Z-score is also known as standard score** gives us an idea of how far a data point is from the mean. It indicates how many standard deviations an element is from the mean. Hence, Z-Score is measured in terms of standard deviation from the mean.

For example, a standard deviation of 2 indicates the value is 2 standard deviations away from the mean. In order to use a z-score, we need to know the population mean (μ) and also the population standard deviation (σ).

The formula is given as: *A z-score can be calculated using the following formula.*

#### z = (X – μ) / σ

#### where,  z = Z-Score,  X = The value of the element,  μ = The population mean, and  σ = The population standard deviation

### **Interpretation of Z-score**

* An element having a z-score less than 0 represents that the element is less than the mean.
* An element having a z-score greater than 0 represents that the element is greater than the mean.
* An element having a z-score equal to 0 represents that the element is equal to the mean.
* An element having a z-score equal to 1 represents that the element is 1 standard deviation greater than the mean; a z-score equal to 2, 2 standard deviations greater than the mean, and so on.
* An element having a z-score equal to -1 represents that the element is 1 standard deviation less than the mean; a z-score equal to -2, 2 standard deviations less than the mean, and so on.

#### Lightbox

If the number of elements in a given set is large, then about 68% of the elements have a z-score between -1 and 1; about 95% have a z-score between -2 and 2; about 99% have a z-score between -3 and 3. This is known as the Empirical Rule or the 68-95-99.7 Rule.

1. What is t-test?

Ans: 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.

1. What is percentile?

Ans: Percentiles are used in statistics to give you a number that describes the value that a given percent of the values are lower than.

Or we can say a value on a scale of 100 that indicates the percent of a distribution that is equal to or below it a score in the 95th percentile.

Percentile is NOT Percentage. So what is the percentile value of 4 in the above dataset? - ( number of value below 4/ total number) \* 100. (6/8) \*100 = 75 percentile.

1. What is ANOVA?

Ans: Analysis of variance, or ANOVA, is a statistical method that separates observed variance data into different components to use for additional tests. A one-way ANOVA is used for three or more groups of data, to gain information about the relationship between the dependent and independent variables.

It 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 Anova is :

F= MST/MSE

Where,

F means ANOVA Coefficient,

MST means Mean sum of squares due to treatment  
MSE means Mean sum of square due to error.

1. How can ANOVA help?

Ans: 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. It is employed with subjects, test groups, between groups and within groups.

ANOVA, or its non-parametric counterparts, allow you to determine if differences in mean values between three or more groups are by chance or if they are indeed significantly different. ANOVA is particularly useful when analyzing the multi-item scales common in market research.

