WORKSHEET-III

Q.1: ans. *– b.*

Q.2: ans. *– c.*

Q.3: ans. *– a.*

Q.4: ans. *– a.*

Q.5: ans. *– b.*

Q.6: ans. *– b.*

Q.7: ans. *– b.*

Q.8: ans. *– a.*

Q.9: ans. *– a.*

Q.10: ans. *– Bayes' Theorem, named after 18th-century British mathematician Thomas Bayes, is a mathematical formula for determining conditional probability. Conditional probability is the likelihood of an outcome occurring, based on a previous outcome having occurred in similar circumstances. Bayes' theorem provides a way to revise existing predictions or theories (update probabilities) given new or additional evidence.*

In finance, Bayes' Theorem can be used to rate the risk of lending money to potential borrowers. The theorem is also called Bayes' Rule or Bayes' Law and is the foundation of the field of Bayesian statistics.

Applications of Bayes' Theorem are widespread and not limited to the financial realm. For example, Bayes' theorem can be used to determine the accuracy of medical test results by taking into consideration how likely any given person is to have a disease and the general accuracy of the test. Bayes' theorem relies on incorporating prior probability distributions in order to generate posterior probabilities.

Prior probability, in Bayesian statistical inference, is the probability of an event occurring before new data is collected. In other words, it represents the best rational assessment of the probability of a particular outcome based on current knowledge before an experiment is performed.

Posterior probability is the revised probability of an event occurring after taking into consideration the new information. Posterior probability is calculated by updating the prior probability using Bayes' theorem. In statistical terms, the posterior probability is the probability of event A occurring given that event B has occurred.

Q.11: ans. *– Z-score indicates how much a given value differs from the standard deviation. The Z-score, or standard score, is the number of standard deviations a given data point lies above or below mean. Standard deviation is essentially a reflection of the amount of variability within a given data set.*

WORKSHEET-III

Etymology. From the use of "Z" in Z distribution, another name for normal distribution. 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 isA 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. he basic z score formula for a sample is:

$z = (x - \mu) / \sigma$

For example, let's say you have a test score of 190. The test has a mean (μ) of 150 and a standard deviation (σ) of 25. Assuming a normal distribution, your z score would be:

- $z = (x - \mu) / \sigma$
- $= (190 - 150) / 25 = 1.6.$

The z score tells you how many standard deviations from the mean your score is. In this example, your score is 1.6 standard deviations above the mean.

$$z_i = \frac{x_i - \bar{x}}{s}$$

You may also see the z score formula shown to the left. This is exactly the same formula as z = x – μ / σ, except that x̄ (the sample mean) is used instead of μ (the population mean) and s (the sample standard deviation) is used instead of σ (the population standard deviation). However, the steps for solving it are exactly the same. When you have multiple samples and want to describe the standard deviation of those sample means (the standard error), you would use this z score formula:

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

This z-score will tell you how many standard errors there are between the sample mean and the population mean.

Example problem: In general, the mean height of women is 65" with a standard deviation of 3.5". What is the probability of finding a random sample of 50 women with a mean height of 70", assuming the heights are normally distributed?

WORKSHEET-III

- *$z = (x – \mu) / (\sigma / \sqrt{n})$*
- *$= (70 – 65) / (3.5/\sqrt{50}) = 5 / 0.495 = 10.1$*

The key here is that we're dealing with a sampling distribution of means, so we know we have to include the standard error in the formula. We also know that 99% of values fall within 3 standard deviations from the mean in a normal probability distribution (see 68 95 99.7 rule). Therefore, there's less than 1% probability that any sample of women will have a mean height of 70".

***Q.12: ans**. – A **t-test** is an inferential statistic used to determine if there is a significant difference between the means of two groups and how they are related. T-tests are used when the data sets follow a normal distribution and have unknown variances, like the data set recorded from flipping a coin 100 times. he t-test is a test used for hypothesis testing in statistics and uses the t-statistic, the t-distribution values, and the degrees of freedom to determine statistical significance. A t-test compares the average values of two data sets and determines if they came from the same population. In the above examples, a sample of students from class A and a sample of students from class B would not likely have the same mean and standard deviation. Similarly, samples taken from the placebo-fed control group and those taken from the drug prescribed group should have a slightly different mean and standard deviation.*

Mathematically, the t-test takes a sample from each of the two sets and establishes the problem statement. It assumes a null hypothesis that the two means are equal.

Using the formulas, values are calculated and compared against the standard values. The assumed null hypothesis is accepted or rejected accordingly. If the null hypothesis qualifies to be rejected, it indicates that data readings are strong and are probably not due to chance.

The t-test is just one of many tests used for this purpose. Statisticians use additional tests other than the t-test to examine more variables and larger sample sizes. For a large sample size, statisticians use a z-test. Other testing options include the chi-square test and the f-test.

Using a T-Test

Consider that a drug manufacturer tests a new medicine. Following standard procedure, the drug is given to one group of patients and a placebo to another group called the control group. The placebo is a substance with no therapeutic value and serves as a benchmark to measure how the other group, administered the actual drug, responds.

After the drug trial, the members of the placebo-fed control group reported an increase in average life expectancy of three years, while the members of the group who are prescribed the new drug reported an increase in average life expectancy of four years.

Initial observation indicates that the drug is working. However, it is also possible that the observation may be due to chance. A t-test can be used to determine if the results are correct and applicable to the entire population.

WORKSHEET-III

Four assumptions are made while using a t-test. The data collected must follow a continuous or ordinal scale, such as the scores for an IQ test, the data is collected from a randomly selected portion of the total population, the data will result in a normal distribution of a bell-shaped curve, and equal or homogenous variance exists when the standard variations are equal.

T-Test Formula

Calculating a t-test requires three fundamental data values. They include the difference between the mean values from each data set, or the mean difference, the standard deviation of each group, and the number of data values of each group.

This comparison helps to determine the effect of chance on the difference, and whether the difference is outside that chance range. The t-test questions whether the difference between the groups represents a true difference in the study or merely a random difference.

The t-test produces two values as its output: t-value and degrees of freedom. The t-value, or t-score, is a ratio of the difference between the mean of the two sample sets and the variation that exists within the sample sets.

The numerator value is the difference between the mean of the two sample sets. The denominator is the variation that exists within the sample sets and is a measurement of the dispersion or variability.

This calculated t-value is then compared against a value obtained from a critical value table called the T-distribution table. Higher values of the t-score indicate that a large difference exists between the two sample sets. The smaller the t-value, the more similarity exists between the two sample sets.

Q.13: ans. *– Your percentile is 70 It means 70% students of the total students appeared for the Exam are Behind you. Therefore Your calculated Approximate rank is 262340. However this is not your final rank.*

Q.14: ans. *– 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

The ANOVA test is the initial step in analyzing 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 ANOVA test allows a comparison of more than two groups at the same time to determine whether a relationship exists between them. 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. The distribution of all possible values of the F statistic is the F-distribution. This is actually a group of distribution functions, with two characteristic numbers, called the numerator degrees of freedom and the denominator degrees of freedom.

Q.15: ans.- *A researcher might, for example, test students from multiple colleges to see if students from one of the colleges consistently outperform students from the other colleges. In a business application, an R&D researcher might test two different processes of creating a product to see if one process is better than the other in terms of cost efficiency.*

*The type of **ANOVA** test used depends on a number of factors. It is applied when data needs to be experimental. Analysis of variance is employed if there is no access to statistical software resulting in computing **ANOVA** by hand. It is simple to use and best suited for small samples. With many experimental designs, the sample sizes have to be the same for the various factor level combinations.*

***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.*

One-Way ANOVA Versus Two-Way ANOVA

*There are two main types of **ANOVA**: one-way (or unidirectional) and two-way. There also variations of **ANOVA**. For example, **MANOVA** (multivariate **ANOVA**) differs from **ANOVA** as the former tests for multiple dependent variables simultaneously while the latter assesses only one dependent variable at a time. One-way or two-way refers to the number of independent variables in your analysis of variance test. A one-way **ANOVA** evaluates the impact of a sole factor on a sole response variable. It determines whether all the samples are the same. The one-way **ANOVA** is used to determine whether there are any statistically significant differences between the means of three or more independent (unrelated) groups.*

*A two-way **ANOVA** is an extension of the one-way **ANOVA**. With a one-way, you have one independent variable affecting a dependent variable. With a two-way **ANOVA**, there are two independents. For example, a two-way **ANOVA** allows a company to compare worker productivity based on two independent variables, such as salary and skill set. It is utilized to observe the interaction between the two factors and tests the effect of two factors at the same time.*

.