STATISTICS WORKSHEET – 3

1. Question -1 - Answer - **B**
2. Question -2 - Answer - **C**
3. Question -3 - Answer - **A**
4. Question -4 - Answer - **A**
5. Question -5 - Answer - **A**
6. Question -6 - Answer - **A**
7. Question -7 - Answer - **B**
8. Question -8 - Answer - **D**
9. Question -9 - Answer – **A**

10. 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.

•Bayes' Theorem allows you to update the predicted probabilities of an event by incorporating new information.

•Bayes' Theorem was named after 18th-century mathematician Thomas Bayes.

- It is often employed in finance in calculating or updating risk evaluation.
- The theorem has become a useful element in the implementation of machine learning.
- The theorem was unused for two centuries because of the high volume of calculation capacity required to execute its transactions.

11. 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.

The Z score Formula: One Sample

The 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:

$z = (x – \mu) / \sigma$

$= (190 – 150) / 25 = 1.6.$

12. 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.

- 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 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.

- The 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.
- The 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 is the final statistical measure for determining differences between two means that may or may not be related. The testing uses randomly selected samples from the two categories or groups. It is a statistical method in which samples are chosen randomly, and there is no perfect normal distribution.

13. A percentile is a measure used in statistics indicating the value below which a given percentage of observations in a group of observations fall. For example, the 20th percentile is the value below which 20% of the observations may be found.

The term percentile and the related term percentile rank are often used in the reporting of scores from norm-referenced tests. For example, if a score is at the 86th percentile, where 86 is the percentile rank, it is equal to the value

below which 86% of the observations may be found. In contrast, if it is in the 86th percentile, the score is at or below the value of which 86% of the observations may be found. Every score is in the 100th percentile.
"Percentile" is in everyday use, but there is no universal definition for it. The most common definition of a percentile is a number where a certain percentage of scores fall below that number. You might know that you scored 67 out of 90 on a test. But that figure has no real meaning unless you know what percentile you fall into. If you know that your score is in the 90th percentile, that means you scored better than 90% of people who took the test.

14. 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. 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.
An ANOVA test is a way to find out if survey or experiment results are significant. In other words, they help you to figure out if you need to reject the null hypothesis or accept the alternate hypothesis.
Basically, you're testing groups to see if there's a difference between them.
Examples of when you might want to test different groups:

- A group of psychiatric patients are trying three different therapies: counseling, medication and biofeedback. You want to see if one therapy is better than the others.
- A manufacturer has two different processes to make light bulbs. They want to know if one process is better than the other.
- Students from different colleges take the same exam. You want to see if one college outperforms the other.

The Formula for ANOVA is:

$$F = \frac{MSE}{MST}$$

Where:

F= ANOVA coefficient

MST= Mean sum of squares due to treatment

MSE= Mean sum of squares due to error

15. ANOVA is helpful for testing three or more variables. It is similar to multiple two-sample t-tests. However, it results in fewer types 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 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.