STATISTICS WORKSHEET 3

1. B
2. C
3. A
4. A
5. A
6. B
7. B
8. D
9. A
10. Bayes theorem is one of the most popular machine learning concepts that helps to calculate the probability of occurring one event with uncertain knowledge while other one has already occurred.
    $P(X \mid Y) = [P(Y \mid X) * P(X)] / P(Y)$

11. 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).
    $z = (X – \mu) / \sigma$
12. A t-test is a type of inferential statistic used to determine if there is a significant difference between the means of two groups, which may be related in certain features.
    t = variance between group/ variance within group
    If t-value is large => the two groups belong to different groups.
    If t-value is small => the two groups belong to same group.

13. percentiles are the values below which a certain percentage of the data in a data set is found. If you want to know where you stand compared to the rest of the crowd, you need a statistic that reports relative standing, and that statistic is called a percentile.
    Percentile = (n/N) * 100
    n = ordinal rank of given value

N = total number of values in data set
P = percentile

14. 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.
    Another measure to compare the samples is called a t-test. When we have only two samples, t-test and ANOVA give the same results. However, using a t-test would not be reliable in cases where there are more than 2 samples.
15. ANOVA is a method to determine if the mean of groups are different. In inferential statistics, we use samples to infer properties of populations. Statistical tests like ANOVA help us justify if sample results are applicable to populations.
    ANOVA is preferred for three or more groups.
    ANOVA can used in feature selection process of machine learning. The features can be compared by performing an ANOVA test and similar ones can be eliminated from the feature set.