STATISTICS WORKSHEET- 6

1. D
2. A
3. A
4. C
5. C
6. A
7. C
8. B
9. B
10.

In the univariate case, box-plots do provide some information that the histogram does not. That is, it typically provides the median, 25th and 75th percentile, min/max that is not an outlier and explicitly separates the points that are considered outliers. In box plots we can better visualize the outliers with respect to the inferential statistics of the feature. Histograms only give a meaure of desity of the feature values

11.

The metrics are chosen on terms of nature of the problem. Classification , Regression and unsupervised learning all have different metrics. Also based on the problem given to decide if we want specificity or sensitivity also where and how the results would would be applied in real word. To know distribution of the target variable.

12.

Statistical significance can be accessed using hypothesis testing: – Stating a null hypothesis which is usually the opposite of what we wish to test. we choose a suitable statistical test and statistics used to reject the null hypothesis and choose a critical region for the statistics to lie in that is extreme enough for the null hypothesis to be rejected (p-value)
We then calculate the observed test statistics from the data and check whether it lies in the critical region. There are multiple test we performed based on the nature of the problem and features of our dataset.

13.

Types of distribution that are non-Gaussian or non-log normal are the skewed distributions, discrete distributions and binomial distribution.

14.

When there are way to many outliers, in those cases if we use mean. We will be way off as mean is drastically affected by outliers. Thus, in such cases it is preferable to use mean as the metric for central tendency than mean. . Another time when we usually prefer the median over the mean (or mode) is when our data is skewed

15.

A likelihood function takes the data set as a given, and represents the likeliness of different parameters for your distribution. The Likelihood function gives us an idea of how well the data summarizes these parameters.