Worksheet set 6 Statistics

10. What is the difference between a boxplot and histogram?
Ans
A histogram is a type of bar chart that graphically displays the frequencies of a data set. Similar to a bar chart, a histogram plots the frequency, or raw count, on the Y-axis (vertical) and the variable being measured on the X-axis (horizontal).

A box plot, also called a box-and-whisker plot, is a chart that graphically represents the five most important descriptive values for a data set. These values include the minimum value, the first quartile, the median, the third quartile, and the maximum value.

Although histograms and box plots are collectively part of the chart aid category, they do represent very different types of charts. Both charts effectively represent different data sets; however, in certain situations, one chart may be superior to the other in achieving the goal of identifying variances among data. The type of chart aid chosen depends on the type of data collected, rough analysis of data trends, and project goals.

A histogram is highly useful when wide variances exist among the observed frequencies for a particular data set. As seen in the two graphs to the left, the histogram shows that there are three peaks within the data, indicating it is tri-modal (three commonly recurring groups of numbers). This is important because to improve processes, it is critical to understand what is causing these three modes. Had this data simply been graphed using a box plot, the values would average one another out, causing the distribution to look roughly normal.

Another instance when a histogram is preferable over a box plot is when there is very little variance among the observed frequencies. The histogram displayed to the right shows that there is little variance across the groups of data; however, when the same data points are graphed on a box plot, the distribution looks roughly normal with a high portion of the values falling below six.

11. How to select metrics?

12. How do you assess the statistical significance of an insight?
Ans- To assess statistical significance, you would use hypothesis testing. The null hypothesis and alternate hypothesis would be stated first. Second, you'd calculate the p-value, which is the likelihood of getting the test's observed findings if the null hypothesis is true. Finally, you would select the threshold of significance (alpha) and reject the null hypothesis if the p-value is smaller than the alpha — in other words, the result is statistically significant.

13. Give examples of data that doesnot have a Gaussian distribution, nor log-normal.
14. Give an example where the median is a better measure than the mean.
15. What is the Likelihood?
Ans- Likelihood function is a fundamental concept in statistical inference. It indicates how likely a particular population is to produce an observed sample. Let P(X; T) be the distribution of a random vector X, where T is the vector of parameters of the distribution.