STATISTICS WORKSHEET- 6 Answers

1) a) The outcome from the roll of a die
2) a) Discrete
3) a) pdf
4) c) mean
5) a) variance
6) a) variance
7) c) 0 and 1
8) b) bootstrap
9) b) summarized

10) **What is the difference between a boxplot and histogram?**

Answer= A histogram is a graph used to represent the frequency distribution of a few data points of one variable. Histograms often classify data into various “bins” or “range groups” and count how many data points belong to each of those bins.
and A box plot is a graphical rendition of statistical data based on the minimum, first quartile, median, third quartile, and maximum. The term "box plot" comes from the fact that the graph looks like a rectangle with lines extending from the top and bottom.
the diffrence between two are clearly seen via both this definitaiton Histograms are a special kind of bar graph that shows a bar for a range of data values instead of a single value. A box plot is a data display that draws a box over a number line to show the interquartile range of the data. The 'whiskers' of a box plot show the least and greatest values in the data set.

11) **How to select metrics ?**

Answer= Good metrics are important to your company growth and objectives. Your key metrics should always be closely tied to your primary objective. The key point is to choose metrics that clearly indicate where you are now in relation to your goals.

12) **How do you assess the statistical significance of an insight?**

Answer= Statistical significance is often calculated with statistical hypothesis testing, which tests the validity of a hypothesis by figuring out the probability that your results have happened by chance.if the pvalue falls below the significance level, then the result is statistically significant. The p-value is a function of the means and standard deviations of the data samples.

13) **Give examples of data that doesnot have a Gaussian distribution, nor log-normal?**

Answer= Any distribution of money or value will be non--Gaussian. For example: distributions of income; distributions of house prices;

distributions of bets placed on a sporting event. These distributions cannot have negative values and will usually have extended right hand tails.

14) **Give an example where the median is a better measure than the mean?**
Answer= Income is the classic example of when to use the median instead of the mean because its distribution tends to be skewed. The median indicates that half of all incomes fall below and half are above it.The median provides a helpful measure of the centre of a dataset. By comparing the median to the mean, you can get an idea of the distribution of a dataset. When the mean and the median are the same, the dataset is more or less evenly distributed from the lowest to highest values.

15) **What is the Likelihood?**

Answer=Likelihood is a strange concept in that it is not a probability but is proportional to a probability. The likelihood of a hypothesis (H) given some data (D) is the probability of obtaining D given that H is true multiplied by an arbitrary positive constant K: $L(H) = K \times P(D|H)$.