Statistics Worsksheet 6

Q.1) D -> All

Q.2) A -> Discrete

Q.3) A -> Pdf

Q.4) C -> Mean

Q.5) A -> Variance

Q.6) A -> Variance

Q.7) C -> 0 and 1

Q.8) B -> Bootstrap

Q.9) B -> Summarized

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

Answer : Both histograms and box plots are used to explore and present the data in an easy and understandable manner. Histograms are preferred to determine the underlying probability distribution of a data. Box plots on the other hand are more useful when comparing between several data sets. They are less detailed than histograms and take up less space.

Q. 11) How to select metrics?

Answer : Good metrics have three key attributes: their data are consistent, cheap, and quick to collect. All outcomes need to be precisely defined and clearly prioritized.
The common framework used is SMART (specific, measurable, achievable, relevant, and time-bound)

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

Answer : We perform hypothesis testing to determine statistical significance. First, we would state the null hypothesis and alternative hypothesis. Second, we would calculate the p-value, the probability of obtaining the observed results of a test assuming that the null hypothesis is true. Statistical significance is calculated using a p-value, which tells us the probability of our result being observed, given that a certain statement (the null hypothesis) is true.

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

Answer : Any type of categorical data won't have a gaussian distribution or lognormal distribution. Exponential distributions — eg. the amount of time that a car battery lasts or the amount of time until an earthquake occurs.

Q. 14) Give an example where the median is a better measure than the mean.

Answer : When there are a number of outliers that positively or negatively skew the data shows median is a better measure than the mean.

Q. 15) What is the Likelihood?

Answer : Likelihood refers to how well a sample provides support for particular values of a parameter in a model.