STATISTICS WORKSHEET- 6

Ans1: All of the mentioned

Ans2: Discrete

Ans3: pdf

Ans4: mean

Ans5: variance

Ans6: none of the mentioned

Ans7: 0 and 1

Ans8: Bootstrap

Ans9: summarized

Ans10: Both Histogram and box plots are used to explore and present the data in an easy and understandable manner. Histogram are to determine the underlying probability distribution of a data. Histograms are a special kind of a bar graph that shows a bar for a range of data values instead of a single value. Box plots are more useful when comparing between several data sets. They are less detailed than histograms and take up less space. Boxplot is a data display that draws a box over a number line to show the interquartile range of data. The 'whiskers' of a boxplot show the least and greatest values in the data set.

Ans11: Statistics, in general is the method of collection of data, tabulation and interpretation of numerical data. A metric is an observed value of a certain quantity at a given point in time. Metrics are measures of quantitative assessment commonly used for comparing and tracking performance or production. Metrics can be used in a variety of scenarios. Metrics are numbers that tell you important information about a process under question. Metrics tell you accurate

measurements about how the process is functioning and provide base for you to suggest improvements. To select metrics first we need to prioritize objectives, examine which metric consistently predicts their achievement and identify which activities influence predictors, in that order. And continuously re-evaluate this process to keep up with the times.

Ans12: 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. Stating a null hypothesis which is usually the opposite of what we wish to test. Then we choose a suitable statistical test and statistics used to reject null hypothesis. We choose a critical region for the statistics to lie In that is extreme enough for the null hypothesis to be rejected. We calculated the observed test statistics from the data and check whether it lies in the critical region.

Ans13: Any distribution of money or value will be non-Gaussian distribution. For ex: distribution of house prices, income etc. these distributions cannot have negative values and will usually have extended right hand tails. A log- normal distribution is a continuous distribution of random variable whose natural logarithm is normally distributed. Exponential distributions do not have a Gaussian distribution nor a log- normal distribution. Any data with data type is categorical will not have these distributions as well. for example: time until the next earthquake, duration of a phone car.

Ans14: In statistics, mean, median and mode are the three most common measures of central tendency. Each one calculates the central point using a different method. Choosing the best measure of central tendency depends on the type of data we have. Median is a better measure than the mean as median better represents the central tendency for the skewed distribution. Outliers and skewed data have a smaller effect on the mean as compared to median. For ex: imagine we have the median dataset below and find that the median is 46. However, we discover data entry errors and need to change four values, which

are shaded in the median fixed dataset. We make them all significantly higher so that we now have a skewed distribution with large outliers. The mean depends on all values in the datasets while the median does not. Consequently, when some of the values are more extreme, the effect on the median is smaller with other types of changes, the median can change. When you have a skewed distribution median is a better measure of central tendency than the mean.

Ans15: The likelihood function often simply called as likelihood is the joint probability of the observed data viewed as a function of the parameters of the chosen statistical model. The likelihood function, parameterized by a possibly multivariate parameter, is usually defined differently for discrete and continuous Probability distribution.

.