STATISTICS WORKSHEET-6

ANSWERS:

1. d) All of the mentioned
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. 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.

11. 1. Use standards. I prefer metrics that have been tested by others;
    2. Measure yourself the way your customer measures you
    3. Only measure metrics that have an owner

12. You would perform hypothesis testing to determine statistical significance. First, you would state the null hypothesis and alternative hypothesis. Second, you would calculate the p-value, the probability of obtaining the observed results of a test assuming that the null hypothesis is true. Last, you would set the level of the significance (alpha) and if the p-value is less than the alpha, you would reject the null — in other words, the result is statistically significant.

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

14. When there are a number of outliers that positively or negatively skew the data. These data are based on the U.S. household income for 2006. Income is the classic example of when to use the median instead of the mean because its distribution tends to be skewed.

15. Likelihood refers to possibility. It shows how much possibility there is for something to happen. The word likelihood indicates the meaning of 'being likely' as in the expression 'in all likelihood'.