STATISTICS WORKSHEET-4

Answer 1 - d) All of the mentioned

Answer 2 - a) Discrete

Answer 3 - a) pdf

Answer 4 - c) mean

Answer 5 - a) variance

Answer 6 - a) variance

Answer 7 - c) 0 and 1

Answer 8 - b) bootstrap

Answer 9 - b) summarized

Answer 10 - 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.

Answer 11 -

- Classification. This algorithm will predict data type from defined data arrays. For example, it may respond with yes/no/not sure.
- Regression. The algorithm will predict some values. For example, weather forecast for tomorrow.
- Ranking. The model will predict an order of items.

Answer 12 – 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.

Answer 13 – Many random variables have distributions that are asymptotically Gaussian but may be significantly non-Gaussian for small numbers. For example the Poisson Distribution, which describes (among other things) the number of unlikely events occurring after providing a sufficient opportunity for a few events to occur. It is pretty non-Gaussian unless the mean number of events is very large. The mathematical form of the distribution is still Poisson, but a histogram of the number of events after many trials with a large average number of events eventually looks fairly Gaussian.

Answer 14 – 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 27581, and half are above it. For these data, the mean overestimates where most household incomes fall.

Answer 15 – The likelihood is the probability that a particular outcome is observed when the true value of the parameter is , equivalent to the probability mass on ; it is not a probability density over the parameter . The likelihood, , should not be confused with , which is the posterior probability of given the data .