STATISTICS WORKSHEET- 6 answers

Q1 – 9

1. d) All of the mentioned
2. c) Continuous
3. d) all of the mentioned
4. c) mean
5. a) variance
6. d) none of the mentioned
7. c) 0 and 1
8. b) bootstrap
9. b) summarized

Q10 – 15

10. A box plot is a form of data visualization that is used to display several aspects of a data set. A segmented box represents the median, upper quartile (Q3), and lower quartile (Q1). A quartile is a segment of data that composes 25% of observed responses, while the median is the data value that represents the middle data value. Lower and higher limits are displayed at the end of whiskers, which are lone segments that extend from the main box. Boxplots may also depict values that are far outside of the normal range of responses (referred to as outliers).

A histogram is a graphical representation of the spread of data points. Values are broken up into ranges on the x-axis, and the number of responses is denoted on the y-axis. The total number of responses that occur within each range is depicted as a bar graph. Histograms are helpful in determining how data is distributed.

12. Two questions arise about any hypothesized relationship between two variables:

1) what is the probability that the relationship exists;
2) if it does, how strong is the relationship

There are two types of tools that are used to address these questions: the first is addressed by tests for statistical significance; and the second is addressed by Measures of Association.

Tests for statistical significance are used to address the question: what is the probability that what we think is a relationship between two variables is really just a chance occurrence?

If we selected many samples from the same population, would we still find the same relationship between these two variables in every sample? If we could do a census of the population, would we also find that this relationship exists in the population from which the sample was drawn? Or is our finding due only to random chance?

Tests for statistical significance tell us what the probability is that the relationship we think we have found is due only to random chance. They tell us what the probability is that we would be making an error if we assume that we have found that a relationship exists.

We can never be completely 100% certain that a relationship exists between two variables. There are too many sources of error to be controlled, for example, sampling error, researcher bias, problems with reliability and validity, simple mistakes, etc.

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.

15. The likelihood function (often simply called the likelihood) represents the probability of random variable realizations conditional on particular values of the statistical parameters. Thus, when evaluated on a given sample, the likelihood function indicates which parameter values are more *likely* than others, in the sense that they would have made the observed data more probable