**STATISTICS WORKSHEET- 6**

**Q1 to Q9 have only one correct answer. Choose the correct option to answer your question.**

**1. Which of the following can be considered as random variable?**

d) All of the mentioned

**2. Which of the following random variable that take on only a countable number of possibilities?**

a) Discrete

**3. Which of the following function is associated with a continuous random variable?**

a) pdf

**4. The expected value or \_\_\_\_\_\_\_ of a random variable is the center of its distribution.**

c) mean

**5. Which of the following of a random variable is not a measure of spread?**

c) empirical mean

**6. The \_\_\_\_\_\_\_\_\_ of the Chi-squared distribution is twice the degrees of freedom.**

b) standard deviation

**7. The beta distribution is the default prior for parameters between \_\_\_\_\_\_\_\_\_\_\_\_**

c) 0 and 1

**8. Which of the following tool is used for constructing confidence intervals and calculating standard errors for difficult statistics?**

b) bootstrap

**9. Data that summarize all observations in a category are called \_\_\_\_\_\_\_\_\_\_ data.**

b) summarized

**Q10and Q15 are subjective answer type questions, Answer them in your own words briefly.**

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

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. How to select metrics?**

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

**13. Give examples of data that does not have a Gaussian distribution, nor log-normal.**

* Allocation of wealth among individuals
* Values of oil reserves among oil fields (many small ones, a small number of large ones)

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

Income is the classic example of when to use the median instead of the mean because its distribution tends to be skewed.

**15. What is the Likelihood?**

The likelihood function represents the probability of random variable realizations conditional on particular values of the statistical parameters.