WORKSHEET

**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?

a) The outcome from the roll of a die

b) The outcome of flip of a coin

c) The outcome of exam

d) All of the mentioned

Answer: D

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

a) Discrete

b) Non Discrete

c) Continuous

d) All of the mentioned

Answer: A

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

a) pdf

b) pmv

c) pmf

d) all of the mentioned

Answer: A

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

a) mode

b) median

c) mean

d) bayesian inference

Answer: C

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

a) variance

b) standard deviation

c) empirical mean

d) all of the mentioned

Answer: A

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

a) variance

b) standard deviation

c) mode

d) none of the mentioned

Answer: A

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

a) 0 and 10

b) 1 and 2

c) 0 and 1

d) None of the mentioned

Answer: C

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

a) baggyer

b) bootstrap

c) jacknife

d) none of the mentioned

Answer: B

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

a) frequency

b) summarized

c) raw

d) none of the mentioned

Answer: B

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

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

Answer: Histograms are preferred to determine the underlying [probability distribution](https://citoolkit.com/articles/probability-distributions/) of a data. Box plots on the other hand are more useful when comparing between several data sets.

Although histograms are better in displaying the [distribution](https://citoolkit.com/articles/probability-distributions/) of data, you can use a box plot to tell if the distribution is symmetric or skewed.

11. How to select metrics?

Answer:

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

Answer: Statistical significance can be accessed using hypothesis testing:  
– Stating a null hypothesis which is usually the opposite of what we wish to test (classifiers A and B perform equivalently, Treatment A is equal of treatment B)  
– Then, we choose a suitable statistical test and statistics used to reject the null hypothesis  
– Also, we choose a critical region for the statistics to lie in that is extreme enough for the null hypothesis to be rejected (p-value)  
– We calculate the observed test statistics from the data and check whether it lies in the critical region

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

Answer:

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

Answer: The median is generally considered to be the best representative of the central location of the data. The more skewed the distribution, the greater the difference between the median and mean, and the greater emphasis should be placed on using the median as opposed to the mean

15. What is the Likelihood?

Answer: The likelihood function (often simply called the likelihood) measures the goodness of fit of a statistical model to a sample of data for given values of the unknown parameters. It is formed from the joint probability distribution of the sample, but viewed and used as a function of the parameters only, thus treating the random variables as fixed at the observed values.