# 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?

**a) variance**

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

**a) variance**

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?

***ANS-***

**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. Value

### 11. How to select metrics?

ANS-

Here's what you should do: prioritize objectives, examine which metric consistently predicts their achievement, and identify which activities influence predictors, in that order. And continuously re-evaluate this process to keep up with the times.

1. Define your primary objective
2. Choose your metric(s) - determine cause and effect
3. Create relevant activities
4. Evaluate periodically

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

**ANS-**

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.

### 13. Give examples of data that doesnot have a Gaussian distribution, nor log-normal.

**ANS-**

Exponential distributions do not have a log-normal distribution or a Gaussian distribution. In fact, any type of data that is categorical will not have these distributions as well.

Example: Duration of a phone car, time until the next earthquake, etc.

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

**ANS-**

**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.

## 15. What is the Likelihood?

**ANS-**

The **likelihood function** (often simply called the **likelihood**) represents the probability of random variable realizations conditional on particular values of the statistical parameters.[1] 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