STATISTICS WORKSHEET- 6

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

# Answers are Highlighted in green

1. Which of the following can be considered as random variable?
   1. The outcome from the roll of a die
   2. The outcome of flip of a coin
   3. The outcome of exam
   4. All of the mentioned
2. Which of the following random variable that take on only a countable number of possibilities?
   1. Discrete
   2. Non Discrete
   3. Continuous
   4. All of the mentioned
3. Which of the following function is associated with a continuous random variable?
   1. pdf
   2. pmv
   3. pmf
   4. all of the mentioned
4. The expected value or of a random variable is the center of its distribution.
   1. mode
   2. median
   3. mean
   4. bayesian inference
5. Which of the following of a random variable is not a measure of spread?
   1. variance
   2. standard deviation
   3. empirical mean
   4. all of the mentioned
6. The of the Chi-squared distribution is twice the degrees of freedom.
   1. variance
   2. standard deviation
   3. mode
   4. none of the mentioned
7. The beta distribution is the default prior for parameters between
   1. 0 and 10
   2. 1 and 2
   3. 0 and 1
   4. None of the mentioned
8. Which of the following tool is used for constructing confidence intervals and calculating standard errors for difficult statistics?
   1. baggyer
   2. bootstrap
   3. jacknife
   4. none of the mentioned
9. Data that summarize all observations in a category are called data.
   1. frequency
   2. summarized
   3. raw
   4. none of the mentioned

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

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

**Answer:**

|  |  |
| --- | --- |
| **BOX PLOT** | **HISTOGRAM** |
| A box plot is a chart that graphically represents the five most important descriptive values for a data set. These values include the minimum value, the first quartile, the median, the third quartile, and the maximum value | A [histogram](https://www.brighthubpm.com/six-sigma/13307-what-is-a-histogram/) is a type of bar chart that graphically displays the frequencies of a data set. |
| Box plots are more useful when comparing between several data sets | Histograms are preferred to determine the underlying probability distribution of a data |
|  |  |

1. How to select metrics?

**Answer:**

The type of metric to use depends on the statistical nature of the problem at hand, the distribution to which the features belong, and the type of feature we are dealing with

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

**Answer:**

To assess statistical significance, you would use hypothesis testing. The null hypothesis and alternate hypothesis would be stated first. Secondly, you would 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

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

**Answer:**

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

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

**Answer:**

* Customer satisfaction survey
* Income level research

1. What is the Likelihood?

**Answer:**

This describes the plausibility, under a certain statistical model (the null hypothesis in hypothesis testing), of a certain parameter value after observing a particular outcome.

