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
ANS: 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
ANS: 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
ANS: 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
ANS: 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
ANS: 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
ANS: D

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
ANS: 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
ANS: B

9. Data that summarize all observations in a category are called _____ data.
a) frequency
b) summarized
c) raw

d) none of the mentioned
ANS: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?

ANS: Histograms are a special kind of bar graph that shows a bar for a range of data values instead of a single value. In a histogram, bars represent ranges instead of individual values. These bars are called bi ns, and they are presented continuously with no spaces between them. Each bar represents a range of data points, and the height of the bar tells us how many data points are in that range.

A box plot is a data display that draws a box over a number line to show the interquartile range of the d ata. The 'whiskers' of a box plot show the least and greatest values in the data set.They have 5 vertical l ines. The lines farthest on the left and right tell the least and greatest values of the data set. The line in the middle is the median. The other two lines are called the lower quartile and upper quartile. The low

er quartile line is on the left of the median, and it tells us that onequarter of the data points are less tha n or equal to the lower quartile.

11. How to select metrics?
ANS: There are two main types of metrics:-
1-*Leading indicators* measure the activities necessary to achieve your goals. Think of them as inputs such as activity per sales rep.

2- *Lagging indicators* measure the actual results - they show whether or not you hit your goals. Think of them as outputs such as revenue closed

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: Many random variables have distributions that are *asymptotically* Gaussian but may be significantly non-Gaussian for small numbers. example is the location of the centers of raindrop ripples on a pond; they are not uniformly spaced in (say) the east-west direction, but they are uniformly distributed.
The simplest example is the distribution of numbers that show up on the top of a fair die after a large number of throws. Each number from 1 to 6 will occur with approximately equal frequency. Increasing the number of throws will not tend to produce a bell-shaped histogram , in fact the fractional occurrence will approach a constant 1/6 over the possible numbers.

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

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.