STATISTICS WORKSHEET- 6

1. Which of the following can be considered as random variable?
Ans: d) All of the mentioned

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

3. Which of the following function is associated with a continuous random variable?
Ans: a) pdf

4. The expected value or _____ of a random variable is the center of its distribution.
Ans: c) mean

5. Which of the following of a random variable is not a measure of spread?
Ans: a) variance

6. The _____ of the Chi-squared distribution is twice the degrees of freedom.
Ans: a) variance

7. The beta distribution is the default prior for parameters between _____
Ans: c) 0 and 1

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

9. Data that summarize all observations in a category are called _____ data.
Ans: b) summarized

10. What is the difference between a boxplot and histogram?
Ans: Histograms and box plots are very similar in that they both help to visualize and describe numeric data. Although histograms are better in determining the underlying distribution of the data, box plots allow you to compare multiple data sets better than histograms as they are less detailed and take up less space.

11. How to select metrics?
Ans: The key point is to choose metrics that clearly indicate where you are now in relation to your goals. Good metrics can be improved. Good metrics measure progress, which means there needs to be room for improvement. For example, reducing churn by 0.8% or increasing your activation rate by 3%.

12. How do you assess the statistical significance of an insight?
Ans: Statistical significance is often calculated with statistical hypothesis testing, which tests the validity of a hypothesis by figuring out the probability that your results have happened by chance.

13. Give examples of data that does not 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.

15. What is the Likelihood?

Ans: The likelihood is the probability that a particular outcome is observed when the true value of the parameter is , equivalent to the probability mass on ; it is not a probability density over the parameter . The likelihood, , should not be confused with , which is the posterior probability of given the data .