STATISTICS WORKSHEET

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. (c) Continuous

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 centre of its distribution.

Ans. (b) median

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

Ans. (b) standard deviation

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

Ans. (b) standard deviation

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. Boxplot: Boxplot is a graphically represent of five number of summaries.
Histogram: Histogram is a one type of bar chart which is represent data distribution.

11. How to select metrics?

Ans. Metric selecting on basis of your model and it's data types. What you want to predict from model then you select metrics depend upon your models.
Ex. If your model is regression model then most probability you select R-square score, MSE etc.

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 alternative hypothesis would be start first. Second, you get the test and calculate p-value, which is null

hypothesis is true. Finally, you select threshold of significance (α) and reject the null hypothesis, if p-value is smaller than α in other words, result is statistically significant.

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

Ans. Any type of categorical data would not a gaussian or log-normal distribution.
Exp.- Amount of time that car battery lasts.

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

Ans. When data distribution trend to be skewed then best use median instead of mean.
Exp. Find average income of a country.

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.