STATISTICS WORKSHEET
ASSIGNMENT-6(Submitted by : Anshul Dubey)
Internship-28

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: ALL of the mentioned

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: Discrete

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: pdf

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: MEAN

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: empirical deviation

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: variance

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: 0 and 1

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: bootstrap

9. Data that summarize all observations in a category are called _____ data.
a) frequency
b) summarized
c) raw
d) none of the mentioned

ANS: 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: 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: Based on Regression or Classification, we can select the metric to be used example r squared for regression and accuracy for classification.

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

Ans: Statistical significance can be accessed using hypothesis testing: – Stating a null hypothesis which is usually the opposite of what we wish to test (classifiers A and B perform equivalently, Treatment A is equal of treatment B)

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

Ans:
• Life data analysis(helps to measure time to failure rate).
• Analyse the lifetime of dental and medical implants.

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

Ans: If data contains outliers such as the 1000 for 1 student's marks out of 100, the mean would be dominated by the outlier value rather than the actual correct values. But median won't be affected by it.

15. What is the Likelihood?

Ans: The likelihood function (often simply called the likelihood) measures the goodness of fit of a statistical model to a sample of data for given values of the unknown parameters