FLIP ROBO

STATISTICS WORKSHEET- 6

(ASSIGNMENT - Worksheet6 Answers – Vivek Kumar Sahu – Internship 35)

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

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

3. Which of the following function is associated with a continuous random variable?
   a) pdf
   b) pmv
   c) pmf
   d) all of the mentioned

4. The expected value or _____ of a random variable is the center of its distribution.
   a) mode
   b) median
   c) mean
   d) bayesian inference

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

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

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

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

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

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. While boxplots and histograms are visualizations used to show the distribution of the data, they communicate information differently.

Histograms are bar charts that show the frequency of a numerical variable's values and are used to approximate the probability distribution of the given variable. It allows you to quickly understand the shape of the distribution, the variation, and potential outliers.

Boxplots communicate different aspects of the distribution of data. While you can't see the shape of the distribution through a box plot, you can gather other information like the quartiles, the range, and outliers. Boxplots are especially useful when you want to compare multiple charts at the same time because they take up less space than histograms.

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)

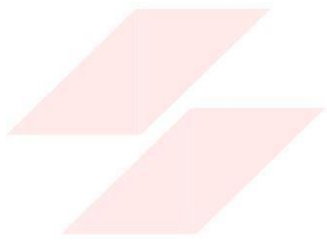13. Give examples of data that doesnot have a Gaussian distribution, nor log-normal.
 Ans.
   • Life data analysis(helps to measure time to failure rate).
   • Analyse the lifetime of dental and medical implants
   • Any type of categorical data won't have a gaussian distribution or lognormal distribution. Exponential distributions‐eg. the amount of time that a car battery lasts or the amount of time until an earthquake occurs.

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