

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

Answer:-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

Answer:-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

Answer:- C

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

Answer :- 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

Answer :- 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

Answer :- A

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

Answer: - 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) jackknife
 - d) none of the mentioned

Answer: - B

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

Answer:- B

Q10 and Q15 are subjective answer type questions, Answer them in your own words briefly.

10. What is the difference between a boxplot and histogram?

Answer:- 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?

Answer:- KEY STEPS TO SELECTING EVALUATION METRICS

1. Classification. This algorithm will predict data type from defined data arrays. For example, it may respond with yes/no/not sure.
2. Regression. The algorithm will predict some values. For example, weather forecast for tomorrow.
3. Ranking. The model will predict an order of items.

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

Answer:- **Steps in Testing for Statistical Significance**

1. State the Research Hypothesis.
2. State the Null Hypothesis.
3. Select a probability of error level (alpha level)
4. Select and compute the test for statistical significance.
5. Interpret the results.

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

Answer:-

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 call, time until the next earthquake, etc.

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

Answer:-

The mean is used for normal distributions. The median is generally used for skewed distributions. The mean is not a robust tool since it is largely influenced by outliers. The median is better suited for skewed distributions to derive at central tendency since it is much more robust and sensible.

Example:- Income is the classic example of when to use the median instead of the mean because its distribution tends to be skewed. The median indicates that half of all incomes fall below 27581, and half are above it. For these data, the mean overestimates where most household incomes fall.

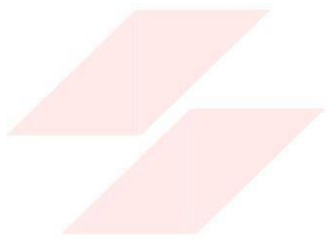
15. What is the Likelihood?

Answer:-

Likelihood is a confusing term. Likelihood is not a probability, but is proportional to a probability; the two terms can't be used interchangeably. Dissecting likelihood as a concept and understand its importance in machine learning.

Definition: - Likelihood, being the outcome of a likelihood function thus defined, describes the plausibility, under a certain statistical model (the null hypothesis in hypothesis testing), of a certain parameter value after observing a particular outcome.

"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."



FLIP ROBO