

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
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) jackknife
 - 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?

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- Choosing an appropriate metric is challenging generally in applied machine learning, but is particularly difficult for imbalanced classification problems. Firstly, because most of the standard metrics that are widely used assume a balanced class distribution, and because typically not all classes, and therefore, not all prediction errors, are equal for imbalanced classification.

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

Answer- 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.

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 car, time until the next earthquake, etc.

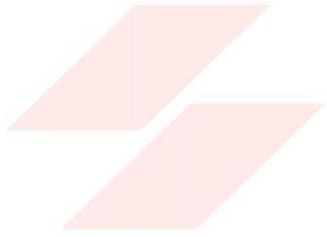
14. Give an example where the median is a better measure than the mean.

Answer- 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?

Answer- Likelihood is a strange concept in that it is not a probability but is proportional to a probability. The likelihood of a hypothesis (H) given some data (D) is the probability of obtaining D given that H is true multiplied by an arbitrary positive constant K:

$$L(H) = K \times P(D|H)$$



FLIP ROBO