

Submitted by: Monika Singh

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

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?

Ans:

In the univariate case, box-plots do provide some information that the histogram does not. That is, it typically provides the median, 25th and 75th percentile, min/max that is not an outlier and explicitly separates the points that are considered outliers. In box plots we can better visualize the outliers with respect to the inferential statistics of the feature. Histograms only give a measure of density of the feature values.

11. How to select metrics?

Ans:

The metrics are chosen on terms of nature of the problem. Classification, Regression and unsupervised learning all have different metrics. Also based on the problem given to decide if we want specificity or sensitivity also where and how the results would be applied in real word. To know distribution of the target variable.

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

Ans:

Hypothesis testing is a method for gaining access to statistical significance. Stating a null hypothesis that is typically the opposite of the hypothesis that we are testing. We select an appropriate statistical test, statistics that are used to reject the null hypothesis, and a critical region within which the statistics must reside in order for the null hypothesis to be rejected (p-value). The observed test statistics are then calculated from the data, and their location in the crucial zone is subsequently determined. Based on the features of our dataset and the nature of the issue, we performed a number of tests.

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

Ans:

The data that does not have a Gaussian distribution, nor log-normal are the skewed distributions, discrete distributions and binomial distribution.

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

Ans:

The mean is used for normal distributions. But the median is generally used for skewed distributions. When there are far too many outliers, using the mean is appropriate. We will be way off as mean is drastically affected by outliers. Thus, using median rather than mean as the metric for central tendency in these situations is better. When our data is skewed, we once again frequently choose the median over the mean.

There are several rules and regulations for various workplaces. Some of them use the median income as a reference, while others base their decisions on the average salary. Employees can learn their career's average salary by using the median wage.

15. What is the Likelihood?

Ans:

The likelihood of various distributional parameters is represented by a likelihood function, which takes the data set as a given. We can get a sense of how effectively the data summarizes these parameters using the Likelihood function.