

WORKSHEET

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

Ans:- 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

Ans:- a) 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:- a) 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:- c) 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:- a) variance

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:- a) 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:- c) 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) jackknife
- d) none of the mentioned

Ans:- b) 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:- b) summarized

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:- Both histograms and box plots are used to explore and present the data in an easy and understandable manner.

Histograms are a special kind of bar graph that shows a bar for a range of data values instead of a single value. Histograms are preferred to determine the underlying probability distribution of a data.

Histogram - gives only the count.

A box plot is a data display that draws a box over a number line to show the interquartile range of the data. The 'whiskers' of a box plot show the least and greatest values in the data set.

Box plots on the other hand are more useful when comparing between several data sets. They are less detailed than histograms and take up less space. Box plot - gives the quartiles and indicate the median data to compare easily.

11. How to select metrics?

The key point is to choose metrics that clearly indicate where you are now in relation to your goals. Good metrics can be improved. Good metrics measure progress, which means there needs to be room for improvement. For example, reducing churn by 0.8% or increasing your activation rate by 3%..

Similar to vectors, you can use the square brackets [] to select one or multiple elements from a matrix. Whereas vectors have one dimension, matrices have two dimensions. You should therefore use a comma to separate the rows you want to select from the columns.

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

Ans:- To assess statistical significance, you would use hypothesis testing. The null hypothesis and alternate hypothesis would be stated first. Second, you'd calculate the p-value, which is the likelihood of getting the test's observed findings if the null hypothesis is true. Finally, you would select the threshold of significance (alpha) and reject the null hypothesis if the p-value is smaller than the alpha — in other words, the result is statistically significant.

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 doesnot have a Gaussian distribution, nor log-normal.

Ans:- Any distribution of money or value will be non--Gaussian. For example: distributions of income; distributions of house prices; distributions of bets placed on a sporting event. These distributions cannot have negative values and will usually have extended right hand tails.

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.

Ans:- 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.

15. What is the Likelihood?

Ans:- Likelihood is a strange concept, in that it is not a probability, but it is proportional to a probability. The likelihood of a hypothesis (H) given some data (D) is proportional to the probability of obtaining D given that H is true, multiplied by an arbitrary positive constant K. In other words, $L(H) = K \times P(D|H)$.

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 .