

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) All of above

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) 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

Answer:- a) pdf (Probability Density Function)

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) 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

Answer:- 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

Answer:- 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

Answer:- c) 0 to 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

Answer:- 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

Answer:- 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?

Answer:- Histogram & box plot are very similar as they both help to visualize & describe the numeric data. Although histograms are better in determining the underlying distribution of the data,.

Although histograms are better in displaying the distribution of data, you can use a box plot to tell if the distribution is symmetric or skewed. In a symmetric distribution, the mean and median are nearly the same, and the two whiskers has almost the same length.

Box plots allow you to compare multiple data sets better than histograms as they are less detailed and take up less space. They are useful to find the outliers present in dataset. It's a chart that graphically represents the five most important descriptive values for a data set. These values include the minimum value, the first quartile, the median, the third quartile, and the maximum value.

11. How to select metrics?

Answer:- 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.

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

Answer:- 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.

13. Give examples of data that doesnot 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. 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, 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.

