

ASSIGNMENT

STATISTICS

Worksheet: -6

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: - b) The outcome of flip of a coin

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

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: - a) 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: - b) standard deviation

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 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

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

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

Answer: -

BOX PLOT	HISTOGRAM
<ol style="list-style-type: none">1. Box plots are graphical representations for the frequency of numeric data values.2. Box plots are more useful when comparing between several data sets. They are less detailed than histograms and take up less space.3. Box plot - gives the quartiles and indicate the median data to compare easily	<ol style="list-style-type: none">1. Histograms are also graphical representations for the frequency of numeric data values.2. Histograms are preferred to determine the underlying probability distribution of a data.3. Histogram - gives only the count

11. How to select metrics?

Answer: - Metrics are measures of quantitative assessment commonly used for assessing, comparing, and tracking performance or production. Generally, a group of metrics will typically be used to build a dashboard that management or analysts review on a regular basis to maintain performance assessments, opinions, and business strategies.

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

Answer: - Statistical significance is the likelihood that a relationship between two or more variables in an analysis is not purely coincidental, but is actually caused by another factor. In other words, statistical significance is a way of mathematically proving that a certain statistic is reliable.

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.

Here, a “hypothesis” is an assumption or belief about the relationship between your datasets. The result of a hypothesis test allows us to see whether this assumption holds under scrutiny or not. A standard hypothesis test relies on two hypotheses.

Null hypothesis: The default assumption of a statistical test that you’re attempting to disprove (e.g., an increase in cost won’t affect the number of purchases).

Alternative hypothesis: An alternate theory that contradicts your null hypothesis (e.g., an increase in cost will reduce the number of purchases). This is the hypothesis you hope to prove.

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 represents the probability of random variable realizations conditional on particular values of the statistical parameters. Thus, when evaluated on a given sample, the likelihood function indicates which parameter values are more likely than others, in the sense that they would have made the observed data more probable.