

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 : 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**

Answer: 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: 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: 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: empirical mean

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: standard deviation

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: 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: 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: summarized

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

Answer : Boxplot : A box plot is a chart that shows data from a five-number summary including one of the measures of central tendency. It does not show the distribution in particular as much as a stem and leaf plot or histogram does. But it is primarily used to indicate a distribution is skewed or not and if there are potential unusual observations (also called outliers) present in the data set. Boxplots are also very beneficial when large numbers of data sets are involved or compared.

Histogram : A histogram is the graphical representation of data where data is grouped into continuous number ranges and each range corresponds to a vertical bar.

The horizontal axis displays the number range.

The vertical axis (frequency) represents the amount of data that is present in each range.

The number ranges depend upon the data that is being used.

11. How to select metrics?

Answer : 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%. One exception to this might be customer satisfaction - if you're already at 100%, your team will be focused on maintaining that level instead of improving it.

There are two main types of metrics. Leading indicators measure the activities necessary to achieve your goals. Think of them as inputs such as activity per sales rep. Lagging indicators measure the actual results - they show whether or not you hit your goals. Think of them as outputs such as revenue closed.

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

Answer :

1) State the Research Hypothesis : A research hypothesis states the expected relationship between two variables. It may be stated in general terms, or it may include dimensions of direction and magnitude

2) State the Null Hypothesis : A null hypothesis usually states that there is no relationship between the two variables

3) Select a probability of error level (alpha level) : Researchers generally specify the probability of committing a Type I error that they are willing to accept, i.e., the value of alpha. In the social sciences, most researchers select an $\alpha = .05$. This means that they are willing to accept a probability of 5% of making a Type I error, of assuming a relationship between two variables exists when it really does not. In research involving public health, however, an alpha of .01 is not unusual. Researchers do not want to have a probability of being wrong more than 0.1% of the time, or one time in a thousand.

4) Select and compute the test for statistical significance

5) Interpret the results

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

Answer : Any type of categorical data won't have a gaussian distribution or lognormal distribution.

Exponential distributions - eg. the amount of time that a car battery lasts or the amount of time until an earthquake occurs.

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

Answer : The median is described as the numeric value separating the higher half of a sample, a population, or a probability distribution, from the lower half.

The mean is used for normal distributions. The median is generally used for skewed distributions. The mean is not a robust tool since it is largely influenced by outliers. The median is better suited for skewed distributions to derive at central tendency since it is much more robust and sensible.

15. What is the Likelihood?

Answer : The term Likelihood refers to the process of determining the best data distribution given a specific situation in the data.

When calculating the probability of a given outcome, you assume the model's parameters are reliable.

However, when you calculate the likelihood, you're attempting to determine whether the parameters in a model can be trusted based on the sample data you have observed.

Likelihood is the hypothetical probability that an event that has already occurred would yield a specific outcome. The concept differs from that of a probability in that a probability refers to the occurrence of future events, while a likelihood refers to past events with known outcomes.

