

STATISTICS WORKSHEET- 6

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-1 d) All of the mentioned

QUESTION-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-2 a) Discrete

QUESTION- 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-3 a) pdf

QUESTION- 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- 4 c) mean

QUESTION- 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- 5 a) variance

QUESTION- 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-6 a) variance

QUESTION- 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- 7 c) 0 and 1

QUESTION- 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- 8 b) bootstrap

QUESTION-9 Data that summarize all observations in a category are called _____ data.

- a) frequency
- b) summarized
- c) raw
- d) none of the mentioned

ANSWER-9 b) summarized

QUESTION-10 What is the difference between a boxplot and histogram?

ANSWER-10 Histograms are a special kind of bar graph that shows a bar for a range of data values instead of a single value. 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.

QUESTION-11 How to select metrics?

ANSWER 11- KEY STEPS TO SELECTING EVALUATION METRICS

1. **Classification.** This algorithm will predict data type from defined data arrays. For example, it may respond with yes/no/not sure.
2. **Regression.** The algorithm will predict some values. For example, weather forecast for tomorrow.
3. **Ranking.** The model will predict an order of items.

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

ANSWER 12- 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.

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

ANSWER- 13 Any type of categorical data won't have a gaussian distribution or lognormal distribution. Exponential distributions - e.g., the amount of time that a car battery lasts or the amount of time until an earthquake occurs.

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

ANSWER-14 The median better represents the central tendency for the skewed distribution. These data are based on the U.S. household income for 2006. Income is the classic example of when to use the median instead of the mean because its distribution tends to be skewed.

QUESTION-15. What is the Likelihood?

ANSWER- 15 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.