Statistics 4

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
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
3. Which of the following function is associated with a continuous random variable?
   a) pdf
   b) pmv
   c) pmf
   d) all of the mentioned
4. The expected value or _____ of a random variable is the center of its distribution.
   a) mode
   b) median
   c) mean
   d) bayesian inference
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
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

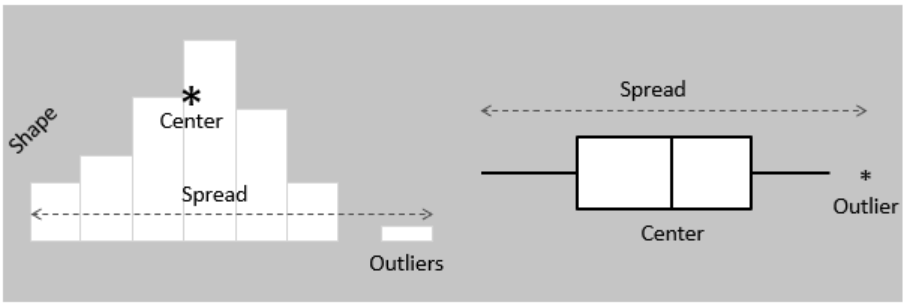
7. The beta distribution is the default prior for parameters between _____
   a) 0 and 10
   b) 1 and 2
   c) 0 and 1
   d) None of the mentioned
8. Which of the following tool is used for constructing confidence intervals and calculating standard errors for difficult statistics?
   a) baggyer
   b) bootstrap
   c) jacknife
   d) none of the mentioned

9. Data that summarize all observations in a category are called _____ data.
   a) frequency
   b) summarized
   c) raw
   d) none of the mentioned

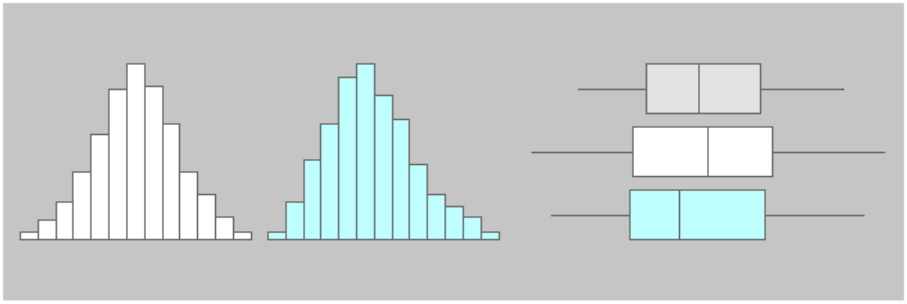
10. What is the difference between a boxplot and histogram?
11. How to select metrics?
12. How do you assess the statistical significance of an insight?
13. Give examples of data that doesnot have a Gaussian distribution, nor log-normal.
14. Give an example where the median is a better measure than the mean.
15. What is the Likelihood?

1. Answer: c)
2. Answer: c)
3. Answer: a)
4. Answer: c)
5. Answer: b)
6. Answer: d)
7. Answer: c)
8. Answer: b)
9. Answer: b)
10. Answer:

Histograms and box plots are graphical representations for the frequency of numeric data values. They aim to describe the data and explore the central tendency and variability before using advanced statistical analysis techniques. Both histograms and box plots allow to visually assess the central tendency, the amount of variation in the data as well as the presence of gaps, outliers or unusual data points.



Both histograms and box plots are used to explore and present the data in an easy and understandable manner. Histograms are preferred to determine the underlying probability distribution of a data. 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.



11. Answer:

12. Answer:
    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.
    The testing part of hypothesis tests allows us to determine which theory, the null or alternative, is better supported by data. There are many hypothesis testing methodologies, and one of the most common ones is the Z-test, which is what we’ll discuss here.

    But, before we get to the Z-test, it is important for us to visit some other statistical concepts the Z-test relies on.
13. 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. Answer:
    Income is the classic example of when to use the median instead of the mean because its distribution tends to be skewed.
15. Answer:
    The likelihood function (often simply called the 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. .