

**STATISTICS WORKSHEET- 6**

Q1 to Q9 have only one correct answer. Choose the correct option to answer your question. **correct answer highlighted**

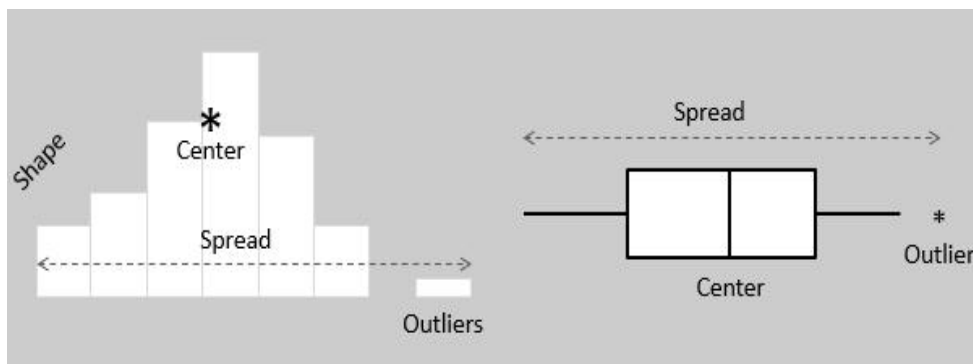
1. Which of the following can be considered as random variable?
    - a) The outcome from the roll of a die
    - b) The outcome of flip of a coin
    - c) The outcome of exam
    - d) All of the mentioned**
  2. Which of the following random variable that take on only a countable number of possibilities?
    - 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) jackknife
    - d) none of the mentioned
-

9. Data that summarize all observations in a category are called \_\_\_\_\_ data.
- frequency
  - summarized
  - raw
  - none of the mentioned

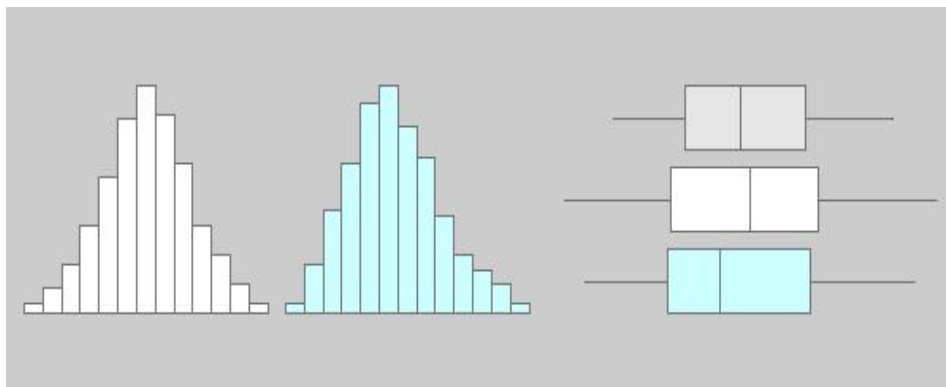
**Q10 and Q15 are subjective answer type questions, Answer them in your own words briefly.**

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

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 are ideal to represent moderate to large amount of data. They may not accurately display the distribution shape if the data size is too small. In practice, a sample size of at least 30 data values would be sufficient for both tools.



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.



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. How to select metrics?**

**KEY STEPS TO SELECTING EVALUATION METRICS**

**Classification.** This algorithm will predict data type from defined data arrays. For example, it may respond with yes/no/not sure.

**Regression.** The algorithm will predict some values. For example, weather forecast for tomorrow.

**Ranking.** The model will predict an order of items.

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

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

**13. Give examples of data that doesnot have a Gaussian distribution, nor log-normal**

Any distribution of money or value will be non--Gaussian. For example: distributions of income; distributions of house prices; distributions of bets placed on a sporting event. These distributions cannot have negative values and will usually have extended right hand tails.

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

It is best to use the median when the distribution is either *skewed* or there are outliers present. When a distribution is skewed, the median does a better job of describing the center of the distribution than the mean.

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?**

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