

## STATISTICS WORKSHEET- 6

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

**1. Which of the following can be considered as random variable?**

d) All of the mentioned

**2. Which of the following random variable that take on only a countable number of possibilities?**

a) Discrete

**3. Which of the following function is associated with a continuous random variable?**

a) pdf

**4. The expected value or \_\_\_\_\_ of a random variable is the center of its distribution.**

c) mean

**5. Which of the following of a random variable is not a measure of spread?**

a) variance

**6. The \_\_\_\_\_ of the Chi-squared distribution is twice the degrees of freedom.**

b) standard deviation

**7. The beta distribution is the default prior for parameters between \_\_\_\_\_**

c) 0 and 1

**8. Which of the following tool is used for constructing confidence intervals and calculating standard errors for difficult statistics?**

b) bootstrap

**9. Data that summarize all observations in a category are called \_\_\_\_\_ data.**

b) summarized

**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. In this article, we will further discuss the similarities and differences between these two 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.

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. How to select 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.

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

Researchers use a measurement known as the p-value to determine statistical significance: **if the p-value falls below the significance level, then the result is statistically significant.** The p-value is a function of the means and standard deviations of the data samples

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

data that does not have a Gaussian distribution, nor log-normal. **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.

**Exponential distributions** do not have a log-normal distribution or a Gaussian distribution.

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

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.**

**Mean :** The mean is used for normal distributions. The mean is not a robust tool since it is largely influenced by outliers. The mean is the arithmetic average of a set of numbers, or distribution. It is the most commonly used measure of central tendency of a set of numbers.

**Median:** The median is generally used for skewed distributions. The median is better suited for skewed distributions to derive at central tendency since it is much more robust and sensible 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.

### 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

There didn't seem much likelihood of it happening. There is every likelihood that sanctions will work. **If something is a likelihood, it is likely to happen.** The likelihood is that your child will not develop diabetes.