<u>STATISTICS WORKSHEET- 6</u>

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.
        a) variance

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

Q10and 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.

11. How to select metrics?
        Good metrics are important to your company growth and objectives. Your key metrics should always be closely tied to your primary objective. A good metric example might be month-on-month revenue growth or LTV:CAC ratio. 'Important' is somewhat subjective since growth for one company may be centered around revenue while another company may focus

more on user growth. The key point is to choose metrics that clearly indicate where you are now in relation to your goals.

1. Good metrics can be improved. 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.

2. Good metrics inspire action. When your metrics are important and can be improved, you and your team will immediately know what to do or what questions to ask. For example, why has our conversion rate dropped? Did we make site changes or test a new acquisition channel? Why is churn increasing? By asking questions you can determine possible causes and work to resolve them right away.

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

1. Create a null hypothesis.
2. Create an alternative hypothesis.
3. Determine the significance level.
4. Decide on the type of test you'll use.
5. Perform a power analysis to find out your sample size.
6. Calculate the standard deviation.
7. Use the standard error formula.
8. Determine the t-score.
9. Find the degrees of freedom.
10. Use a t-table.

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

There are many data types that follow a non-normal distribution by nature. Examples include: Weibull distribution, found with life data such as survival times of a product. Log-normal distribution, found with length data such as heights.

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

The median is usually preferred in these situations because the value of the mean can be distorted by the outliers. However, it will depend on how influential the outliers are. If they do not significantly distort the mean, using the mean as the measure of central tendency will usually be preferred.

15. What is the Likelihood?

In statistics, the likelihood function (often simply called the likelihood) measures the goodness of fit of a statistical model to a sample of data for given values of the unknown parameters. ... But in both frequentist and Bayesian statistics, the likelihood function plays a fundamental role.