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?
   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

   Ans- (d)

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

   Ans- (a)

3. Which of the following function is associated with a continuous random variable?
   a) pdf
   b) pmv
   c) pmf
   d) all of the mentioned

   Ans- (a)

4. The expected value or_____of a random variable is the center of its distribution.
   a) mode
   b) median
   c) mean
   d) bayesian inference

   Ans-(c)

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

   Ans- (A)

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

   Ans- (A)

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

Ans- (c)

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

   Ans-(b)

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

   Ans- (b)

Q10and Q15 are subjective answer type questions, Answer them in your own words briefly.

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

Ans- **Boxplots** may also depict values that are far outside of the normal range of responses (referred to as outliers). **A histogram** is a graphical representation of the spread of data points.

In descriptive statistics, **a box plot** is an effective graphical representation of five-number summary, while **histogram** is a type of bar chart to visualize the distribution of a dataset.

Histograms indicate the whole frequency distribution of a variable, whereas the **boxplot** summarises its most prominent features. These features include median and spread as well as the extent and nature of departures from symmetry, and the possible presence of observations having extreme values (outliers).

11. How to select metrics?

Ans- **Choosing the right metrics**

You may be thinking: what the heck is a 'good' metric? You see, all metrics are certainly not created equal. We've all been guilty of slipping into the trap of vanity metrics at some point,but we'll try and help you avoid that mistake.

'Good' can be broadly defined as metrics that show if you're achieving your objectives (the ones you prioritized before). Fundamentally, good metrics have three characteristics.

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

2. **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 youractivation 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.

3. **Good metrics inspire action.** When your metrics are important and can be improved, youand 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?

Ans- Calculating the statistical significance is rather extensive if you calculate it by handand this is why it's typically calculated using a calculator. When you calculate it by hand,however, it will help you more fully understand the concept. Here are the steps for calculating statistical significance:

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.

Ans- 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.

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

Ans- When you have a symmetrical distribution for continuous data, the mean, median, andmode are equal. In this case, analysts tend to use the mean because it includes all of the data in the calculations. However, if you have a skewed distribution, the median is often the bestmeasure of central tendency.
When you have ordinal data, the median or mode is usually the best choice. For categoricaldata, you have to use the mode.

In cases where you are deciding between the mean and median as the better measure of central tendency, you are also determining which types of statistical hypothesis tests are appropriate for your data—if that is your ultimate goal. I have written an article that discusseswhen to use parametric (mean) and nonparametric (median) hypothesis testsalongwith the advantages and disadvantages of each type.

15. What is the Likelihood?

Ans- 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. It is formed from the joint probability distribution of the sample, but viewed andused as a function of the parameters only, thus treating the random variables as fixed at the observed values