FLIP ROBO

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

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

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

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

Histogram and Box Plot Comparison are as follows :-

Even while histograms and box plots are both considered chart aids, they are actually completely distinct kinds of charts. Both charts successfully depict various sets of data; but, in certain cases, one chart may be more beneficial than the other in recognizing data variations. The kind of data gathered, a preliminary study of data patterns, and the project objectives significantly influence the type of chart aid that is used.

When there are significant differences in the observed frequencies for a specific data set, a histogram is quite helpful. The histogram displays that there are, as seen in the two graphs to the left, the data contains three peaks, suggesting that it is tri-modal (three commonly recurring groups of numbers). This is significant because it is essential to comprehend the root causes of these three modes in order to enhance procedures. The values would average out if this data had merely been plotted using a box plot, making the distribution seem to be fairly normal. When there is little variation in the recorded frequencies, a histogram is further preferred over a box plot. The histogram on the right demonstrates that there is minimal variation across the data groups, but when the same data points are plotted on a box plot, the distribution seems to be nearly normal and a significant proportion of the values fall below six.

The final set of graphs shows how a box plot can be more useful than a histogram. This occurs when there is moderate variation among the observed frequencies, which causes the histogram to look ragged and non-symmetrical due to the way the data is grouped. This may lead one to assume the data is slightly skewed. However, when a box plot is used to graph the same data points, the chart indicates a perfect normal distribution.

11. How to select metrics?

Metrics that demonstrate whether you are reaching your goals can be used to define "good" broadly (the ones you prioritised before). Essentially, effective metrics have three features.

1. Good metrics are crucial to the growth and goals of your business. Your core purpose should always be directly related to your key performance indicators. A excellent metric illustration would be the LTV:CAC ratio or month-over-month sales increase. The term "important" is rather arbitrary because one firm's growth may be driven mostly by revenue while another company may place more emphasis on user growth. The important thing is to select indicators that make it obvious where you stand right now in regard to your objectives.
2. Good metrics can be enhanced. Good metrics track progress, which implies that there is always space for improvement. For example, you may reduce churn by

0.8% while increasing activation by 3%. One possible exception is client satisfaction; if you're currently at 100%, your staff will be focused on sustaining that level rather than enhancing it.

3. Good metrics motivate people to take action. When your metrics are essential 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 decreased? Did we make site updates or try a new acquisition channel? Why is turnover increasing? By asking questions, you may identify potential reasons and try to remedy problems right away.

    i.   Create a null hypothesis.

    ii.   Create an alternative hypothesis.

    iii.   Determine the significance level.

    iv.   Decide on the type of test you'll use.

    v.   Perform a power analysis to find out your sample size.

    vi.   Calculate the standard deviation.

    vii.   Use the standard error formula.

    viii.   Determine the t-score.

    ix.   Find the degrees of freedom. 10.Use a t-table.

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

Determining statistical significance is time-consuming if done by hand, which is why it is usually done via a calculator. When you compute it by hand, though, you will have a better understanding of the concept. The steps for calculating statistical significance are as follows:

13. Give examples of data that does not have a Gaussian distribution, nor log-normal.

A gaussian or lognormal distribution will not exist for any form of categorical data. Exponential distributions - for example, how long a car battery lasts or how long it takes for an earthquake to begin.

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

The mean, median, and mode of a symmetrical distribution for continuous data are all equal. Analysts often use the mean in this scenario since it incorporates all of the data into the computations. Yet, if your distribution is skewed, the median is frequently the best estimate of central tendency.

When dealing with ordinal data, the median or mode is typically the best option. The mode must be used for categorical data.

While picking between the mean and median as the best measure of central tendency, you are also considering which statistical hypothesis tests are acceptable for your data—if that is your final aim. I wrote an essay about when to employ parametric

(mean) and nonparametric (median) hypothesis testing, as well as the benefits and drawbacks of each.

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