ASSIGNMENT-6

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.

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?

a) baggyer

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?

11. How to select metrics?

Selecting the right metrics to measure the performance of a model or system is an important step in the process of evaluating and improving it. Here are a few general guidelines to help you select appropriate metrics: 1. Define clear goals: Before selecting metrics, it's important to have a clear understanding of the goals of the model or system. This will help you identify which metrics are most relevant for measuring its performance. 2. Identify the relevant stakeholders: Consider who will be using the model or system, and what information they need to make decisions. Select metrics that will provide them with the information they need. 3. Use domain-specific metrics: Different domains have different characteristics and requirements, so it's important to use metrics that are specific to the domain in question. 4. Balance accuracy and interpretability: A good metric should be both accurate and interpretable. A metric that is highly accurate but difficult to understand is not very useful.

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

Assessing the statistical significance of an insight involves determining the probability that the insight is due to chance rather than a real effect. Here are a few common methods for assessing statistical significance: 1. P-value: A p-value is the probability of observing a test statistic as extreme or more extreme than the one observed, assuming the null hypothesis is true. A common threshold for statistical significance is a p-value of less than 0.05, meaning that there is less than a 5% chance that the result is due to chance. 2. Confidence intervals: A confidence interval is a range of values that is likely to contain the true value of a

population parameter with a certain level of confidence. If a confidence interval does not include zero, this indicates that the result is statistically significant. 3. Effect size: An effect size is a quantitative measure of the strength of the association between two variables. An effect size of zero indicates no association, while larger values indicate stronger associations. It can be used to assess the practical significance of a result

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

There are many examples of data that does not have a Gaussian (also known as normal) distribution or a log-normal distribution.

Here are a few examples: 1. Binomial distribution: This is a discrete probability distribution that describes the number of successful outcomes in a fixed number of Bernoulli trials. For example, the number of heads in 10 coin flips. 2. Poisson distribution: This is a discrete probability distribution that describes the number of events that occur in a fixed interval of time or space, given that the average rate of events is constant. For example, the number of phone calls received by a call center in a minute. 3. Exponential distribution: This is a continuous probability distribution that describes the time between events in a Poisson process, where events occur independently at a constant average rate. For example, the time between customer arrivals at a store.

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

The median is a better measure than the mean in situations where the data has extreme values, also known as outliers. These outliers can greatly affect the mean by pulling it away from the center of the distribution, making it a less representative measure of the typical or central tendency of the data. For example, consider the income of residents in a city. The mean income might be affected by a few residents who have extremely high incomes (e.g. CEOs, celebrities, etc.). These high incomes can greatly increase the mean, making it a less representative measure of the typical income in the city. In contrast, the median income, which is the value that separates the lower half from the upper half of the data, will be less affected by the extreme values and more representative of the typical income

15. What is the Likelihood?

In statistics, likelihood is a function that quantifies the probability of obtaining a certain set of observations given a set of parameters for a statistical model. It is used in statistical inference to estimate the parameters of a model and to compare different models. The likelihood is calculated by multiplying the probability of each observation given the current set of parameters. The parameters that maximize the likelihood are considered the most likely or the best estimates of the true parameters of the model.