In [ ]:
```
#STATISTICS WORKSHEET-1
```

1.Bernoulli random variables take (only) the values 1 and 0. a) True b) False

a.true

2.Which of the following theorem states that the distribution of averages of iid variables, properly normalized, becomes that of a standard normal as the sample size increases? a) Central Limit Theorem b) Central Mean Theorem c) Centroid Limit Theorem d) All of the mentioned

a. central limit theorem

3. Which of the following is incorrect with respect to use of Poisson distribution? a) Modeling event/time data b) Modeling bounded count data c) Modeling contingency tables d) All of the mentioned

b) Modeling bounded count data

4. Point out the correct statement. a) The exponent of a normally distributed random variables follows what is called the log-normal distribution b) Sums of normally distributed random variables are again normally distributed even if the variables are dependent c) The square of a standard normal random variable follows what is called chi-squared distribution d) All of the mentioned

c) The square of a standard normal random variable follows what is called chi-squared distribution

5. __ random variables are used to model rates. a) Empirical b) Binomial c) Poisson d) All of the mentioned

c) Poisson

6. .10 Usually replacing the standard error by its estimated value does change the CLT. a) True b) False

b) False

7. Which of the following testing is concerned with making decisions using data? a) Probability b) Hypothesis c) Causal d) None of the mentioned

b) Hypothesis

8.Normalized data are centered at__and have units equal to standard deviations of the original data. a) 0 b) 5 c) 1 d) 10

a) 0

9. Which of the following statement is incorrect with respect to outliers? a) Outliers can have varying degrees of influence b) Outliers can be the result of spurious or real processes c) Outliers cannot conform to the regression relationship d) None of the mentioned

c) Outliers cannot conform to the regression relationship

10. What do you understand by the term Normal Distribution?

Think of a normal distribution like a special pattern you see when you look at a graph of something, like people's heights or test scores.

1. In the Middle - The most common value is right in the middle of the graph. It's like the top of a hill.
2. Balanced Sides - If you go to the left or right from the middle, things are pretty similar. It's like taking a few steps away from the top of the hill, and you're still on a similar level.
3. Most Things Near the Middle - About 68% of the things you're looking at (like people or test scores) are close to that middle spot.
4. Spread-Out Measure - There's a thing called standard deviation. It measures how spread out things are. In a normal distribution, you can use this to see how much things spread away from the middle.
5. Bell Shape - The whole graph looks like a bell. It goes up in the middle and then slopes down on both sides.

1. How do you handle missing data? What imputation techniques do you recommend?

Handling missing data is crucial in data analysis, and various imputation techniques can be employed. One commonly used imputation technique is mean imputation.

Here's how mean imputation works:

1.Mean Calculation: For each variable with missing data, calculate the mean (average) of the observed values in that variable.

2.Replace Missing Values: Substitute the missing values in each variable with its mean.

While mean imputation is simple and easy to implement, it does have some drawbacks. It may not be the best choice if the missing data is not missing completely at random or if there are a significant number of missing values. In such cases, more advanced imputation methods, like multiple imputation or regression imputation, may be considered.

1. What is A/B testing?

A/B testing is like trying out two different things to see which one people like or respond to better. Imagine you have two versions of a website, an email, or an ad, and you want to know which version works best. You randomly show one version to some people and the other version to another group. Then, you compare how people react to each version. The one that gets a better response is likely the winner, and that's the version you might choose to use. It's a way of figuring out what people prefer or what works more effectively by experimenting with different options. Here's how A/B testing works: 1 Objective - Clearly define the goal of the test. 2 Variations - Create two versions, A (control) and B (variation). 3.Random Assignment - Randomly assign users to Group A or B. 4.Implementation - Show A to Group A and B to Group B. 5.Data Collection - Track relevant metrics for both groups. 6.Statistical Analysis - Analyze data for statistical significance. 7.Decision Making - Choose the version that performs better.

1. Is mean imputation of missing data acceptable practice?

Mean imputation is a common and simple technique for handling missing data, but its acceptability depends on the context and the nature of the data. Here are some considerations:

Pros:

1-Simplicity: Mean imputation is easy to understand and implement.

2-Preservation of Sample Size: It allows you to keep all available data, preventing a reduction in sample size.

Cons:

1.Bias: Mean imputation can introduce bias, especially if the missing data are not missing completely at random. It assumes that the missing values have the same mean as the observed values, which may not be true.

2.Underestimation of Variability: It tends to underestimate the variability in the data because it doesn't account for the uncertainty introduced by imputing missing values.

3.Impact on Relationships: If missing data are related to other variables, mean imputation can distort relationships and correlations.

4.Not Suitable for Categorical Data: Mean imputation is not suitable for categorical data as it relies on calculating means.

1. What is linear regression in statistics?

Linear regression is like trying to find the best straight line that explains the relationship between two things. For example, you might use it to understand how the amount of time someone studies ((x)) relates to the grades they get ((y)).

Imagine drawing a straight line on a graph. The line has a slope (steepness) and an intercept (where it crosses the y-axis). In simple terms, linear regression helps us figure out the best values for the slope and intercept so that the line fits the data points (grades and study time) as closely as possible.

Once we have this line, we can use it to make predictions. If we know how much time someone spends studying ((x)), we can predict what their likely grade ((y)) might be based on the relationship we found.

1. What are the various branches of statistics?

1-Descriptive Stats: Tells you what's happening in your data.

2-Inferential Stats: Helps you guess about a big group based on a small group.

3-Biostatistics: Uses stats to study health and biology.

4-Econometrics: Merges economics and stats to understand money and the economy.

5-Social Stats: Figures out what people generally do or think.

6-Psychometrics: Measures things in psychology, like intelligence or personality.

7-Statistical Genetics: Studies how genes affect traits and health.

8-Environmental Stats: Looks at numbers to understand nature, like pollution or weather.

9-Business Stats: Helps make decisions in business using data.

10-Spatial Stats: Deals with data that has a location, like crime on a map.

In [ ]: