3. Find the IQR for the given dataset and explain its significance.

A) IQR found in excel, IQR = 5.5

IQR = Q3 - Q1

The interquartile range is useful because it tells you how spread out the middle 50 percent of your data is. It gives you the range of values between the 25th percentile and the 75th percentile.

The IQR is also useful as it can be used to identify outliers.

Lower bound = Q1 – 1.5*IQR

Upper Bound = Q3 + 1.5*IQR

If values are less than the lower bound and or greater than Upper Bound it is an outlier

7. Compute the Z-scores for each value in the dataset and explain its significance in data standardization.

A) Z-standardization is a statistical procedure used to make data points from different datasets comparable. In this procedure, each data point is converted into a z-score. A z-score indicates how many standard deviations a data point is from the mean of the dataset.

Section 2: Inferential Statistics

12) Why inferential statistics?

A) While descriptive statistics summarize the characteristics of a data set, Inferential statistics helps you make predictions or generalizations about a population based on a sample of data.

Explain the difference between Correlation and Causation with an example?

A) Correlation is when two variables are related, while causation is when one variable directly causes another.

Examples

- **Ice cream sales and pool drownings**: Both increase during the summer, but ice cream sales don't cause drownings

Explanation

- **Correlation**

A statistical measure of how two variables are related. It can be positive, negative, or zero.

- **Causation**

A cause-and-effect relationship between variables. It's a stronger statement than correlation.

13) Population vs. Sample:

Why do we need sampling? Provide a real-world example

A) A population is the entire group that you want to draw conclusions about.

A sample is the specific group that you will collect data from. The size of the sample is always less than the total size of the population.

In research, a population doesn't always refer to people. It can mean a group containing elements of anything you want to study, such as objects, events, organizations, countries, species, organisms, etc.

Example : Imagine a pharmaceutical company testing a new drug for diabetes.

- **Population**: All diabetes patients worldwide.
- **Sample**: A selected group of 5,000 patients from different regions.

14) Hypothesis Testing Concepts:

Define Null Hypothesis, Alternate Hypothesis, Significance Level (α), and P-value

A) **1. Null Hypothesis (H0)**

- The null hypothesis states that there is no effect, no difference, or no relationship between variables in a study.
- It represents the default assumption that nothing significant is happening.

2. Alternate Hypothesis (HA or H1)

- The alternative hypothesis states that there is a significant effect, difference, or relationship.
- It challenges the null hypothesis and is what researchers aim to prove.

3. Significance Level (α)

- The significance level is the probability of rejecting the null hypothesis when it is actually true (Type I error).
- **Common values**: 0.05 (5%) or 0.01 (1%).

4. P-value

- The p-value is the probability of obtaining the observed data (or more extreme results) if the null hypothesis is true.
- **Decision Rule**:
    - If p-value < α, reject H0 (there is significant evidence to support HA).
    - If p-value > α, fail to reject H0 (not enough evidence to support HA).

15. Z-test Calculation:

A Z-score describes your deviation from the mean in units of standard deviation. It is not explicit as to whether you accept or reject your null hypothesis

Calculated Z value is 6.324

A Z-score of 6.32 is extremely high, meaning the sample mean (25) is far from the population mean (22).

16. P-value Computation for Z-test:

P value from online score calculator, P value < .00001

Since p-value < 0.05, we reject the null hypothesis, this confirms that the sample mean (25) is significantly different from the population mean (22)

17)

T calculated value = 1.334764

Two tailed test because its to test difference(in both directions)

T critical value from table = 1.943

Here T cal < T critical so failed to reject null hypothesis , that means there is no significant difference

20) The t-statistic was calculated as 5.0340 (assumed equal variance) which suggests a significant difference between the means of the two groups.

T-critical from table was 2.306

T-stat> T-critical so rejected Null hypothesis

Descriptive vs. Inferential Statistics

- **Descriptive Statistics**: Summarizes and organizes data, providing measures like mean, median, standard deviation, and variance.
- **Inferential Statistics**: Uses sample data to make generalizations about a population such as hypothesis testing, confidence intervals, and regression analysis.

Descriptive Statistics Examples

- **Business Analytics**: Summarizing customer purchase behavior (e.g., average spending per customer).
- **Healthcare**: Reporting average patient recovery times from a treatment.
- **Education**: Calculating the average test scores of students in different schools.

Inferential Statistics Examples

- **Market Research**: Using sample surveys to predict national customer preferences.
- **Medical Studies**: Testing new drug effectiveness on a sample before applying it to the entire population.
- **A/B Testing in Marketing**: Comparing two ad campaigns to determine which performs better based on a sample.