**Weekly Assignment**

**Statistics**

**Descriptive Statistics:**

**1.What is the purpose of descriptive statistics?**

Descriptive statistics aim to summarize and present key features of a dataset in a meaningful way. The primary purposes include simplifying large amounts of data, providing insights into central tendencies and variabilities, and facilitating a clearer understanding of the data's characteristics.

**2.Can you explain the difference between mean, median, and mode?**

**Mean:** The mean is the average of all values in a dataset. It is calculated by summing up all values and dividing by the total number of observations. For example, if you have test scores of 80, 90, and 100, the mean is (80 + 90 + 100) / 3 = 90.

**Median:** The median is the middle value in a dataset when it is sorted in ascending or descending order. If there is an even number of observations, the median is the average of the two middle values. For instance, in the dataset 10, 20, 30, the median is 20.

**Mode:** The mode is the most frequently occurring value in a dataset. It's possible to have multiple modes or none at all. In the dataset 5, 10, 10, 15, the mode is 10.

**3.How do you interpret the standard deviation of a dataset?**

Standard deviation measures the amount of variation or dispersion in a set of values. A higher standard deviation indicates greater variability, while a lower standard deviation suggests that values are closer to the mean.

For example, if you have two datasets of test scores:

Dataset A: 70, 75, 80, 85, 90 (low standard deviation)

Dataset B: 60, 70, 80, 90, 100 (high standard deviation)

In Dataset A, scores are relatively close to the mean (80), resulting in a lower standard deviation. In Dataset B, scores vary more widely from the mean, leading to a higher standard deviation.

**4.Describe the concept of skewness in statistics.**

Skewness measures the asymmetry of a probability distribution. A distribution can be positively skewed (tail on the right) or negatively skewed (tail on the left).

**Positive Skewness:** The majority of the values are concentrated on the left side of the mean, and the tail extends to the right. For example, if analyzing income data, a positive skewness might indicate a few individuals with extremely high incomes.

**Negative Skewness:** The bulk of the values are concentrated on the right side of the mean, with the tail extending to the left. In financial data, a negative skewness could suggest a few stocks with exceptionally low returns.

**Inferential Statistics:**

**5.What is the main goal of inferential statistics?**

The main goal of inferential statistics is to make predictions or inferences about a population based on a sample of data from that population. It involves generalizing insights gained from a subset (sample) to the entire group (population) and making informed decisions or predictions about unknown characteristics of the population.

**6.Explain the difference between a population and a sample.**

**Population:** It refers to the entire group that is the subject of the study. For example, if you are studying the heights of all students in a school, the population is the heights of every student in that school.

**Sample:** It is a subset of the population. Taking a sample is often more practical than studying the entire population. In the height example, you might measure the heights of 100 students as a sample to draw conclusions about the entire student population.

**7.What is a confidence interval, and how is it useful in inferential statistics?**

A confidence interval is a range of values constructed from a sample of data to estimate an unknown population parameter. It provides a range within which we are reasonably confident the true parameter lies. For instance, if you calculate a 95% confidence interval for the average height of students and get a result of 160 cm to 170 cm, you can be 95% confident that the true average height of all students falls within this range.

**8.Define p-value**

The p-value is a measure that helps you assess the evidence against a null hypothesis in a hypothesis test. In practical terms:

A small p-value (typically ≤ 0.05) suggests that you can reject the null hypothesis, indicating that there is enough evidence to support the alternative hypothesis.

A large p-value suggests that you do not have enough evidence to reject the null hypothesis.

For example, if you conduct a hypothesis test on whether a new drug is effective, a low p-value indicates that the drug has a significant effect. This information is crucial in making decisions based on statistical evidence.