STATISTICS TEST

Descriptive Statistics:

1. **What is the purpose of descriptive statistics ?**
   Descriptive statistics are used to organize, summarize, and present data in an informative way. The main purpose is to simplify large amounts of data in a meaningful and understandable manner. It helps researchers, analysts, and decision-makers understand the basic features of the data without drawing any conclusions beyond the data itself. Common descriptive statistics include measures of central tendency (like mean, median, and mode), measures of dispersion (like range, variance, and standard deviation), and graphical representations such as histograms and pie charts. These statistics provide a foundation for more complex analysis by offering a snapshot of the data.

2. **Can you explain the difference between mean, median and mode?**
   The mean, median, and mode are all measures of central tendency, but they capture different aspects of a dataset. The **mean** is the arithmetic average, calculated by adding all the values and dividing by the number of values. It is sensitive to outliers, which can skew the mean. The **median** is the middle value when the data is arranged in ascending or descending order; if there is an even number of observations, the median is the average of the two middle numbers. The **mode** is the value that appears most frequently in the dataset. A dataset can have more than one mode (bimodal or multimodal) or no mode at all. While the mean gives an overall average, the median provides a better measure of central tendency in skewed distributions, and the mode is useful for identifying the most common value.

3. **How do you interpret the standard deviation of a dataset ?**
   The standard deviation is a measure of how spread out the values in a dataset are around the mean. A small standard deviation indicates that the data points are close to the mean, suggesting low variability and a more consistent dataset. In contrast, a large standard deviation means the data points are spread out over a wider range of values, indicating greater variability. For example, in a dataset of student test scores, a low standard deviation means most students scored similarly, whereas a high standard deviation means there were significant differences in scores. In a normal distribution, about 68% of the values fall within one standard deviation of the mean, 95% within two, and 99.7% within three.

4. **Describe the concept of skewness in statistics ?**
   Skewness measures the asymmetry of a data distribution around its mean. If a distribution is **symmetrical**, it has zero skewness. When a distribution is **positively skewed** (right-skewed), the tail on the right side is longer or fatter than the left side, indicating that a few high values are pulling the mean to the right. In contrast, a **negatively skewed** (left-skewed) distribution has a longer or fatter left tail, indicating that a few low values are pulling the mean to the left. Skewness helps identify the direction and extent of deviation from a normal distribution, which is important when selecting statistical methods, as many assume normally distributed data.

Inferential Statistics :

5. **What is the main goal of inferential statistics ?**
   Inferential statistics aim to make conclusions, predictions, or decisions about a population based on data obtained from a sample. Since it is often impractical or impossible to collect data from every member of a population, inferential techniques allow us to use a representative sample to infer characteristics or test hypotheses about the larger group. This involves estimating population parameters, testing hypotheses, and making predictions with a certain level of confidence. Common methods include confidence intervals, hypothesis testing, regression analysis, and analysis of variance (ANOVA). Inferential statistics help researchers generalize findings and make data-driven decisions beyond the immediate dataset.

6. **Explain the difference between a population and a sample**
   In statistics, a **population** refers to the entire group of individuals, items, or data that you are interested in studying. It can be finite or infinite, and it includes all possible observations relevant to a particular study. A **sample**, on the other hand, is a subset of the population selected for analysis. The sample should ideally be representative of the population to allow for valid inferences. For example, if you want to study the average height of college students in a country, the population is all college students, while the sample might be 500 students selected from various universities. Sampling allows researchers to conduct studies more efficiently and cost-effectively.

7. **What is a confidence interval, and how is it useful in inferential statistics ?**
   A confidence interval is a range of values, derived from sample data, that is likely to contain the true population parameter (such as the mean or proportion) with a certain level of confidence. For example, a 95% confidence interval means that if we were to take 100 different samples and compute a confidence interval for each, about 95 of them would contain the true population parameter. Confidence intervals are useful because they provide an estimate of uncertainty around the sample statistic. They help researchers express the reliability of their estimates and guide decision-making, especially when comparing groups or testing hypotheses.

8. **Define p-value**
   The **p-value** (probability value) is a measure used in hypothesis testing to determine the strength of the evidence against the null hypothesis. It represents the probability of obtaining test results at least as extreme as the observed results, assuming that the null hypothesis is true. A small p-value (typically less than 0.05) indicates strong evidence against the null hypothesis and suggests that the observed effect is statistically significant. In contrast, a large p-value indicates weak evidence, and we fail to reject the null hypothesis. P-values help researchers decide whether the results of an experiment or study are likely to have occurred by chance.