Compulsory Task 1

1. State whether the mean, median, or mode would be useful in the following scenarios:

a. You are doing population statistics. You are asked to give an estimate of the typical income of a single person in the country. There is one snag: wealth distribution is out of whack, and 10% of the population holds 70% of the nation's wealth.

In this scenario, the median would be a more useful measurement since wealth is highly skewed. The mean is distorted to higher values by the presence of a small group of individuals holding a significant portion of the wealth. On the other hand, the median provides a better representation of the typical income for a person in a country. The mode would not be useful here as there might be a clear mode income data.

b. You are running a restaurant, and you are reviewing your menu. You have a list of all orders over the last six months. You are trying to find out which item you should keep based on what customers seem to like the most.

In this scenario, the mode would be the most useful measurement as it identifies the most frequent ordered dishes. Additionally, it provides insights into customer preferences by indicating the items that the mean and median do not directly reveal. While the mean and median can help identify the most popular dish, the mode offers a more comprehensive understanding of customer preferences.

c. You have been buying electricity once a month for the first six months of the year. You are trying to budget your electricity for the rest of the year and therefore need to estimate how much you will spend for the remainder of the year.

In this scenario, the mean would be the most useful measurement. You are looking to average your monthly expenditure over time, which makes the mean the most appropriate measure for estimating future spending.

d. You work in healthcare insurance. You are asked to provide an estimate of the typical amount of money spent on healthcare. This is taking into account the fact that there are a few people who spend a large amount of money on medical healthcare due to major issues.

In this scenario, the median would be the most useful measurement. Few people have extremely high healthcare costs, and their mean would be skewed. The median provides a better representation of the typical healthcare spending of the majority of people. The mode would not be useful in this context.

2. Variance vs Standard Deviation

a. You are choosing a new Internet provider. You find two providers with the same mean speed, but you want to have a more stable connection. You get a list of all reported speeds over the last month and are trying to find the provider that doesn't move too much from the mean value.

Standard deviation is the appropriate measure to use.

Standard Deviation:
Directly measures how much speeds deviate from the mean in the same units as the data (e.g., Mbps).

Easier to interpret: A smaller standard deviation means speeds cluster closer to the mean (more stability).

Variance:
Measures spread as the average squared deviations from the mean, resulting in squared units (e.g., $Mbps^2$). Less intuitive for decision-making since it's not in the original units.

b. You are going on holiday to Mauritius. You need to find a shuttle from the airport to your hotel, but you are worried about being overcharged or undercharged (being undercharged might mean that you get unreliable transport). You get a list of all available shuttle service prices and need to find out which services, if any, are overcharging or undercharging.

Standard deviation is the appropriate measure to use.

Reason: Standard deviation helps identify values that are far from the average (outliers). This is useful in detecting prices that are unusually high or low, helping identify overcharging or undercharging.