1. **What are the measures of central tendency, and how do they differ from each other?**

Answer: The three measures of central tendency are **Mean**, **Median** and **Mode** and are defined as below:

Mean is the arithmetic average of the given dataset.
Median is the middle value of the sorted dataset and where sample size is even, it is the average of the two middle values.
Mode is the value in the dataset with most frequency.

While Mean is best for symmetric distributions without outliers, Median is better for skewed distributions/those with outliers and Mode is useful for categorical data to identify the most common value in a dataset.

2. **How do you interpret the standard deviation in the context of data variability?**

Answer: In the context of data variability, **Variance/Standard Deviation** is a measure of dispersion/spread of values in a dataset, ie. how much the values vary from **Mean** in a given dataset. **Variance** is calculated as the average of squared differences between values and **Mean** and **Standard Deviation** is the square root of **Variance.** A higher value of Variance/Standard Deviation indicates higher dispersion/spread of values in the given dataset.

3. **What is a box plot, and what information can you extract from it?**

Answer: A box plot is a graphical representation of the distribution of a dataset. It displays the **Median, Quartiles** and **potential outliers.** The box shows **IQR (Inter Quartile Range)** which is the range between 25th percentile (Q1/first quartile) and 75th percentile (Q3/third quartile) of values. It also includes a line indicating the **Median.** The outer "whiskers" extend to the smallest and largest values within 1.5 times the IQR from the quartiles (Q1&Q3). Points outside this range are deemed as potential outliers.

From a box plot, we can extract information about the dataset's **central tendency, spread, symmetry** as well as **identify outliers.**

4. **Explain the significance of the interquartile range (IQR) and how it is used to detect outliers.**

Answer: The **interquartile range (IQR)** is the **range between 25th percentile (Q1/first quartile)** and **75th percentile (Q3/third quartile)** of values in a dataset. That is, the spread of middle 50% values in the dataset. Any values in the dataset lying **outside 1.5 times** the IQR from Q1 and Q3 can be deemed as outliers.

5. How Do Maximum Likelihood Estimators (MLE) Work?

Answer: **Maximum likelihood estimation (MLE)** is a method for estimating the parameters of an assumed probability distribution, for a given observed data. It is achieved by maximizing a likelihood function so that the probability of the observed data is highest. We first define the **likelihood function L(θ)**, where θ is the parameter being estimated. This is typically the product of individual probabilities for independent observations. It is often substituted with **log-likelihood function l(θ)** for simplicity, which is the natural logarithm of the likelihood function. The next step is to find the parameter value θ, that maximizes the log-likelihood function. Where likelihood function is differentiable, the derivative test can be applied to determine **θ**.