Write a python program to compute Central Tendency Measures: Mean, Median, Mode, Measure of Dispersion: Variance, Standard Deviation

What is Central Tendency?

Central tendency is a statistical concept that refers to the tendency of data to cluster around a central value or a typical value.

It identifies a single value as representative of an entire data distribution. In other words, the central tendency is a way of describing the centre or midpoint of a dataset. The three most common measures of central tendency are:

- Mean
- Median
- Mode

Mean

The mean, also known as the average, is calculated by summing up all the values in a dataset and dividing the sum by the total number of values. It is sensitive to extreme values, making it susceptible to outliers.

Median

The median is the middle value in a dataset when the values are arranged in ascending or descending order. If there is an even number of values, the median is the average of the two middle values. Unlike the mean, the median is less affected by outliers.

Mode

The mode is the value that occurs most frequently in a dataset. It is particularly useful for categorical data but can also be applied to numerical data. A dataset may have one mode (unimodal), two modes (bimodal), or more than two modes (multimodal).

Note: *The choice of which central tendency measure to use depends on the properties of the data. For example, the mean is best for symmetric distributions, while the median is better for skewed distributions with outliers. The mode is useful for categorical data.*

Let's consider a dataset of daily temperatures recorded over a week: 22°C, 23°C, 21°C, 25°C, 22°C, 24°C, and 20°C.

- **Mean:** (22 + 23 + 21 + 25 + 22 + 24 + 20) / 7 = 21.86°C
- **Median:** Arranging the temperatures in ascending order: 20°C, 21°C, 22°C, 22°C, 23°C, 24°C, 25°C. The median is 22°C.
- **Mode:** The mode is 22°C as it occurs most frequently in the dataset.

What is Dispersion?

Dispersion, also known as variability or spread, measures the extent to which individual data points deviate from the central value. It provides information about the spread or distribution of data points in a dataset.

Common measures of dispersion include

- Range
- Variance
- Standard Deviation

Range

In statistics, the range refers to the difference between the highest and lowest values in a dataset. It provides a simple measure of variability, indicating the spread of data points. The range is calculated by subtracting the lowest value from the highest value.

For example, in a dataset {4, 6, 9, 3, 7}, the range is 9 - 3 = 6.

Variance

Variance is a statistical measure that quantifies the amount of variation or dispersion of a set of numbers from their mean value. Specifically, variance is defined as the expected value of the squared deviation from the mean. It is calculated by:

1. Finding the mean (average) of the data set.
2. Subtracting the mean from each data point to get the deviations from the mean.
3. Squaring each of the deviations.
4. Calculating the average of the squared deviations. This is the variance.

Standard Deviation

Standard deviation is a measure of the amount of variation or dispersion of a set of values from the mean value. It is calculated as the square root of the variance, which is the average squared deviation from the mean.

Examples for Dispersion

Let's consider the same dataset of daily temperatures recorded over a week: 22°C, 23°C, 21°C, 25°C, 22°C, 24°C, and 20°C.

Range: Maximum temperature - Minimum temperature = 25°C - 20°C = 5°C

Variance: Variance = (Sum of squared differences from the mean) / (Number of data points)

Mean = 21.86 °C

Sum of squared differences from the mean = $(22 - 21.86)^2 + (23 - 21.86)^2 + (21 - 21.86)^2 + (25 - 21.86)^2 + (22 - 21.86)^2 + (24 - 21.86)^2 + (20 - 21.86)^2$

$= (0.14)^2 + (1.14)^2 + (-0.86)^2 + (3.14)^2 + (0.14)^2 + (2.14)^2 + (-1.86)^2$

= 0.0196 + 1.2996 + 0.7396 + 9.8596 + 0.0196 + 4.5796 + 3.4596

= 19.0772

Thus, Variance = 19.0772 / 7 ≈ 2.725 °C

Standard Deviation: Take the square root of the variance to get the standard deviation.

Thus, Standard Deviation ≈ √2.725 ≈ 1.65 °C

Program:

```
import numpy as np
from scipy import stats
data = [12, 15, 18, 20, 22, 25, 30, 32, 35, 40]
mean = np.mean(data)
median = np.median(data)
mode = stats.mode(data)
variance = np.var(data)
std_deviation = np.std(data)
print("Central Tendency Measures:")
print("Mean:", mean)
print("Median:", median)
print("Mode:", mode)
print("\nMeasures of Dispersion:")
print("Variance:", variance)
print("Standard Deviation:", std_deviation)
```

Output:

Central Tendency Measures:

Mean: 24.9

Median: 23.5

Mode: ModeResult(mode=12, count=1)

Measures of Dispersion:

Variance: 75.09

Standard Deviation: 8.665448632355973