**Central Tendency (Measure of Centre)**

The central tendency concept is that one single value can best describe the data. **Mean, median,** and **mode** are the three important parameters in statistics. Essentially, all three of them refer to a single aspect called the **Central Tendency**.

**Mean ( 𝜇 )**

The mean is equal to the sum of the values in the dataset divided by the number of values.

Table

Description automatically generated

One of the major disadvantages of using mean rather than using median or mode is, the mean is particularly sensitive to the effect of extreme values. Extreme values are also called **outliers**.

### Median

The median is the middle score for a dataset that has been sorted from small to large. Outliers less affect the median.

Even if you only had 10 scores? In this case, we simply have to take the middle two scores and average the result.

**Mode**

The mode is the most frequent score in a dataset. It represents the highest bar in a histogram or bar chart.

### Calculate Mean, Median and Mode with Python

We can easily calculate mean, median and mode values with python. We use the **numpy** library for the **mean** and **median**, and the **SciPy** library for the **mode**.

import numpy as np

from scipy import stats

salary = [102, 33, 26, 27, 30, 25, 33, 33, 24]

mean\_salary = np.mean(salary)

print("mean:", mean\_salary)

median\_salary = np.median(salary)

print("median:", median\_salary)

mode\_salary = stats.mode(salary)

print("mode:", mode\_salary)

**Dispersion (Measure of Spread)**

Dispersion is a way to explain how a dataset is distributed.

When a dataset has a small value, the values in the dataset are tightly clustered; when it is large the items in the set are widely scattered.

The first population is much more dispersed than the second population, however, the mean value for both populations is the same. Therefore, we can say a dispersion explains something more than the central tendency does.

Chart, histogram

Description automatically generated

Graphical user interface

Description automatically generated with medium confidence

### Range

The main advantage of the range is that it is easy to calculate. There are many disadvantages, on the other hand. It is highly susceptible to extreme values and does not use all the observations in a dataset.

𝑅𝑎𝑛𝑔𝑒=𝑀𝑎𝑥𝑖𝑚𝑢𝑚𝑉𝑎𝑙𝑢𝑒−𝑀𝑖𝑛𝑖𝑚𝑢𝑚𝑉𝑎𝑙𝑢𝑒

### Standard Deviation(𝜎-sigma(Σ, σ, ς-the letter s))

Standard deviation measures the spread around the mean.

It is also expressed as the square root of variance.

Variance(𝜎2) is defined as the average of the squared differences from the mean.

𝜇(müu)=mean

𝜎(sigma)=Standart Deviation

𝜎2 (variance)

Graphical user interface, text, application

Description automatically generated

Text

Description automatically generated

A picture containing shape

Description automatically generated

Text

Description automatically generated with medium confidence

Text, letter

Description automatically generated

The standard deviation is also useful when comparing the spread of two different datasets that have the same mean.

Chart, histogram

Description automatically generated

### Calculate Range, Variance and Standard Deviation with Python

import numpy as np

salary = [102, 33, 26, 27, 30, 25, 33, 33, 24]

print("Range: ", (np.max(salary)-np.min(salary)))

print("Variance: ", (np.var(salary)))

print("Std: ", (np.std(salary)))

Range: 78

Variance: 539.5555555555555

Std: 23.22833518691246

**Inter Quartile Range (IQR)**

**Inter Quartile Range(IQR)** is the **difference between Q3 and Q1**.

Outlier is, any data point more than 1.5 IQR below the Q1 or above the Q3.

Outliers: (Q1 - 1.5 \* IQR) or (Q3 + 1.5 \* IQR)

Chart, box and whisker chart

Description automatically generated

### Practice IQR

number\_list = [1, 5, 10, 15, 40]

Now we will try to find which numbers on our list are the outlier.

We have the following summary:

minimum number = 1

maximum number =40

median=10

Q1 = 5

Q3 = 15

IQR = Q3-Q1

IQR= 15-5 = 10

Therefore, (1.5 \* IQR) = 15

To determine if there are any outliers, we must consider the numbers that are 1.5\*IQR beyond the quartiles.

Q1 – (1.5 \* IQR) = 5-15 = -10

Q3 + (1.5 \* IQR) = 15+15 = 30

The last number in our list is 40. And it is outside of the interval from (–10) to (30), therefore 40 is an outlier. The rest of the numbers in the list are not outliers.