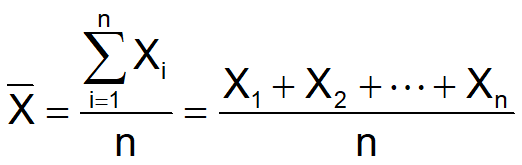
**Chapter 3: Numerical Descriptive Measures**

**Numerical Descriptive Measures for A Sample**

1. **Central tendency**

* **Sample Mean** = sum of values divided by the number of values.



**580VN: Menu - 6 - 1 - Data - AC - OPTN - 2**

**570VN: Mode - 3 - 1 - Data - AC - Shift - 1 - 4 - 2**

* **Sample Median:** is the “middle” number (50% above, 50% below).

1. Rank the data set in increasing order.

2. Median position = position in the ordered data.

* If the number of values is odd, the median is the middle number.
* If the number of values is even, the median is the average of the two middle numbers.

**580VN: Menu - 6 - 1 - Data - AC - OPTN - 2**

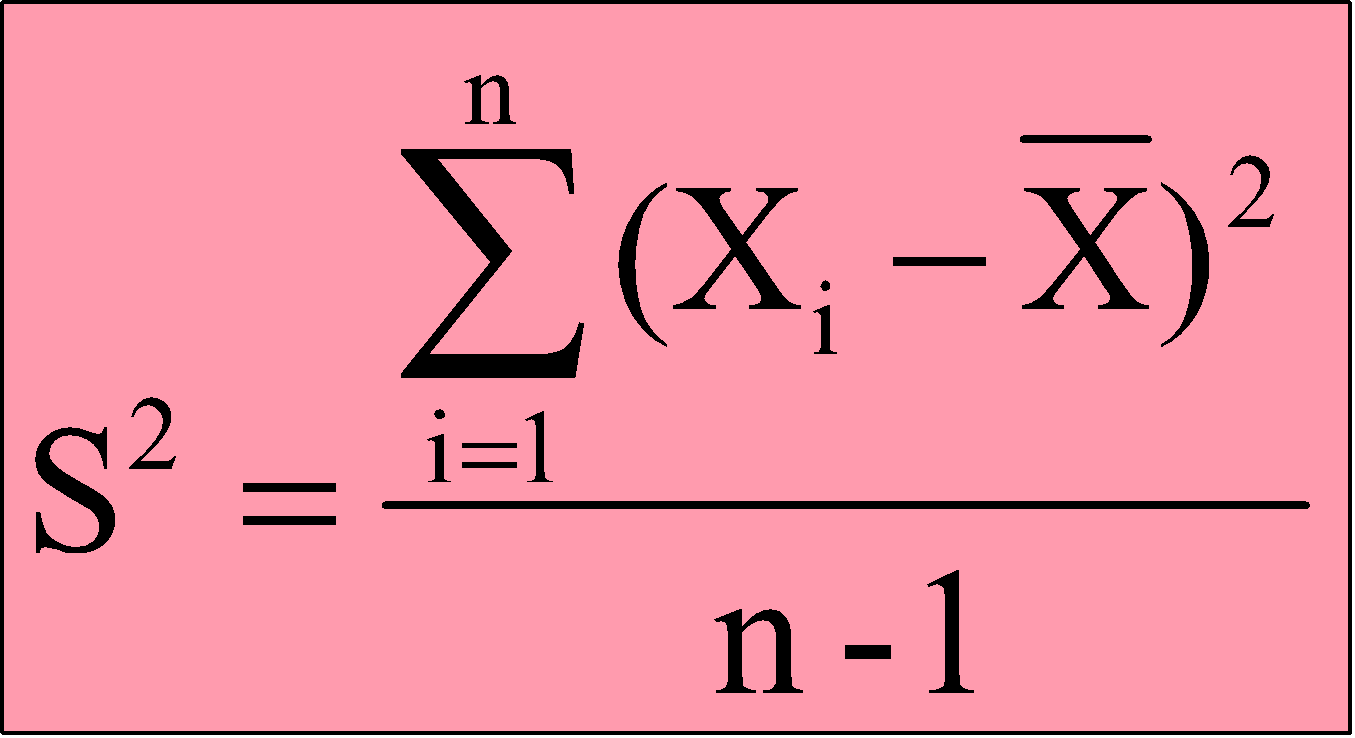
* **Sample Mode:**

Value that occurs most often.

There may be no mode OR there may be several modes.

1. **Variation**

* **Sample Range** =
* **Sample Variance:** Average (approximately) of squared deviations of values from the mean.

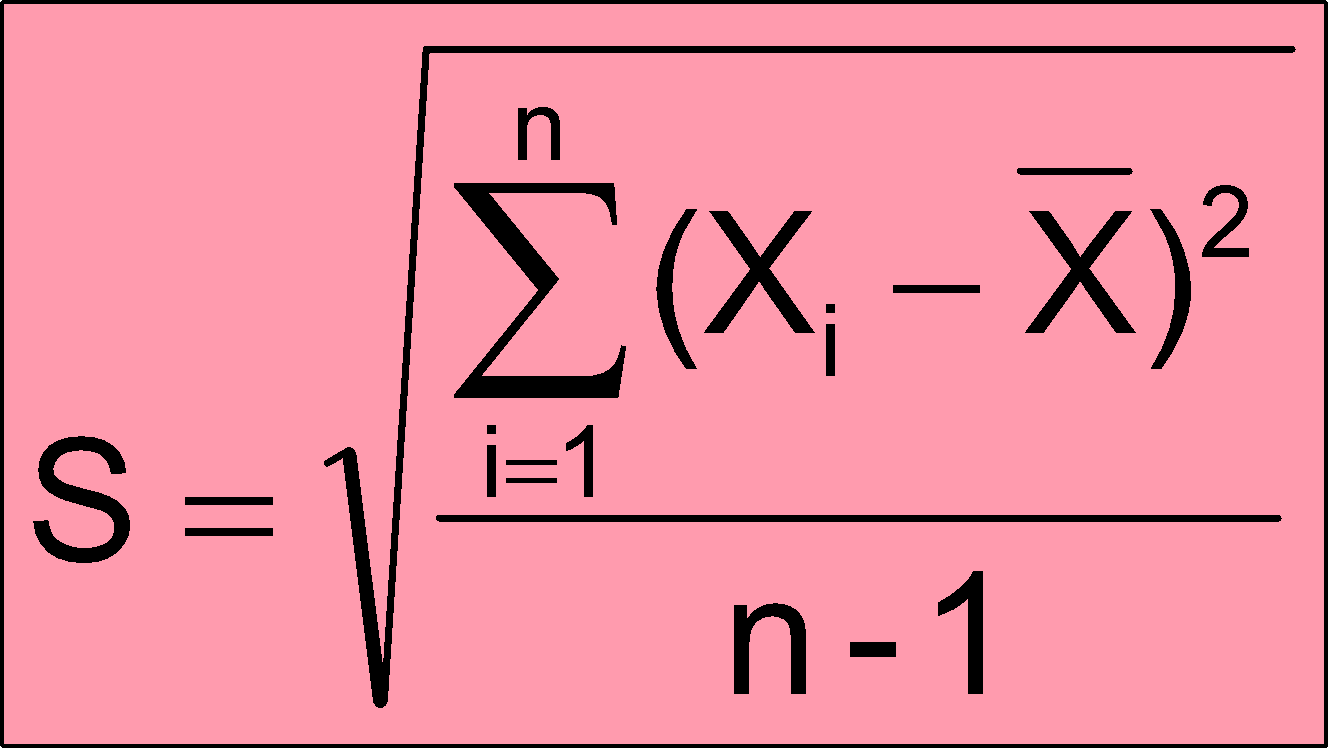
****

**Note: sample variance s^2 # population variance sigma^2**

**580VN: Menu - 6 - 1 - Data - AC - OPTN - 2**

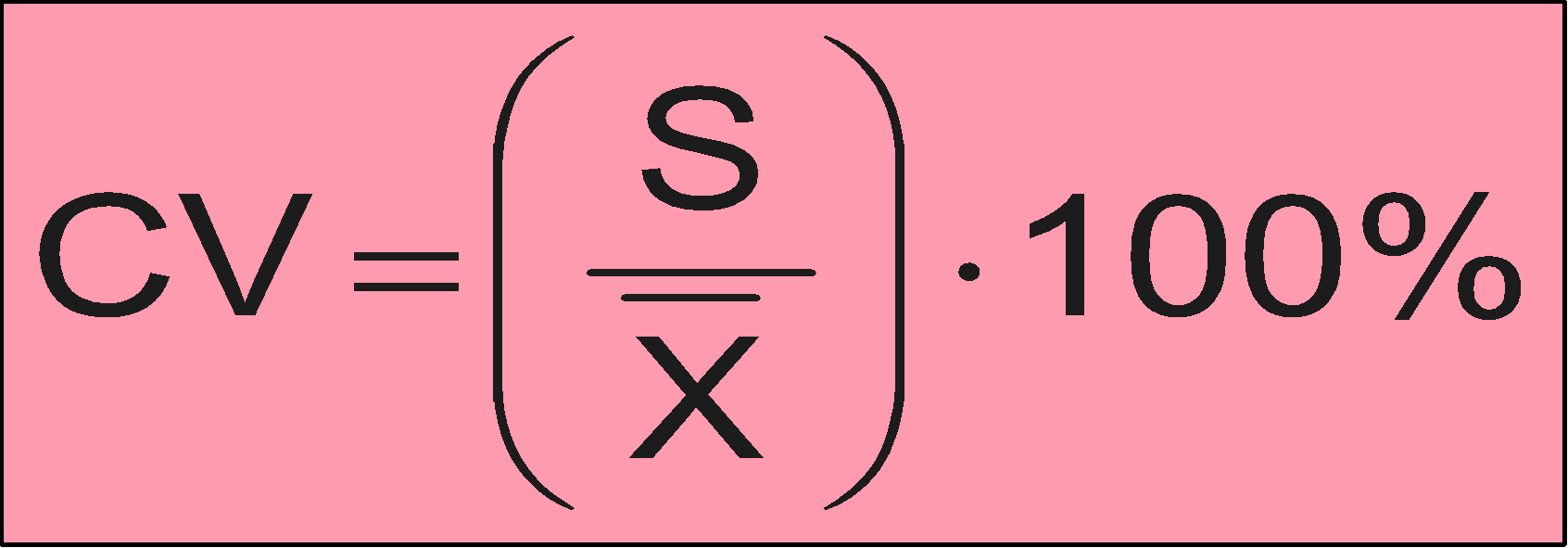
**570VN: Mode - 3 - 1 - Data - AC - Shift - 1 - 4 - 4 - ^2**

* **Sample Standard Deviation:** Is the square root of the variance.

****

**580VN: Menu - 6 - 1 - Data - AC - OPTN - 2**

**570VN: Mode - 3 - 1 - Data - AC - Shift - 1 - 4 - 4**

* **Sample Coefficient of Variation:** Measures relative variation. Always in percentage (%).
* ****

1. **Shape of a Distribution**

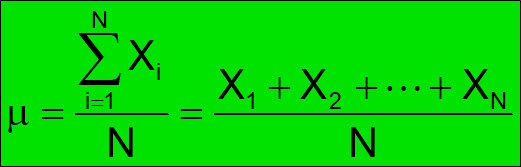
* **Skewness**

**Chart

Description automatically generated**

**Numerical Descriptive Measures for A Population**

1. **Population Mean**

****

**N: population size**

**Sample mean = Population mean**

**580VN: Menu - 6 - 1 - Data - AC - OPTN - 2**

**570VN: Mode - 3 - 1 - Data - AC - Shift - 1 - 4 - 2**

1. **Population Variance**



**580VN: Menu - 6 - 1 - Data - AC - OPTN - 2**

**570VN: Mode - 3 - 1 - Data - AC - Shift - 1 - 4 - 3 - ^2**

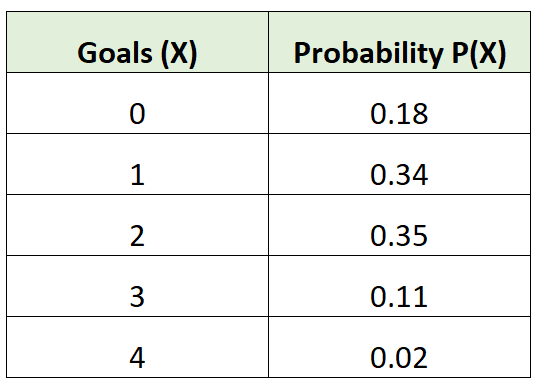
1. **Population Standard Deviation**



**580VN: Menu - 6 - 1 - Data - AC - OPTN - 2**

**570VN: Mode - 3 - 1 - Data - AC - Shift - 1 - 4 – 3**

1. **Probability distribution:** A table that displays the probability that a random variable takes on certain values.

→ Probability distribution → Population

**Box Plots**

**Table

Description automatically generated with medium confidence**

**Quartiles** split the ranked data into **4 segments** with an **equal number of values** per segment.

The first quartile, **Q1**, is the value for which 25% of the values are smaller and 75% are larger.

**Q2** is the same as the median (50% of the values are smaller and 50% are larger).

Only 25% of the values are greater than **the third quartile**.

**Find a quartile** by determining the value in the appropriate position in the ranked data, where:

First quartile position: ranked value.

Second quartile position: ranked value.

Third quartile position: ranked value.

where n is the number of observed values.

The **Interquartile Range** (**IQR**) is **Q3 – Q1** and measures the spread in the **middle 50% of the data**.

An **outlier** is an observation that is numerically distant from the rest of the data (< Q1 - 1.5 \* IQR or > Q3 + 1.5 \* IQR).

E.g. 1 1 2 3 4 1 2 3 4 1 100 → sorted data: 1 1 1 1 2 2 3 3 4 4 100

n = 11

Q1 position = (11+1)/4 = 3 → Q1 = x3 = 1

Q3 position = 3\*(11+1)/4 = 9 → Q3 = x9 = 4

IQR = Q3 - Q1 = 4 - 1 = 3

< Q1 - 1.5\*IQR = 1 - 1.5\*3 = -3.5 → There is no number < -3.5

> Q3 + 1.5\*IQR = 4 + 1.5\*3 = 8.5 → There is 1 number (100) > 8.5→ Outlier: 100