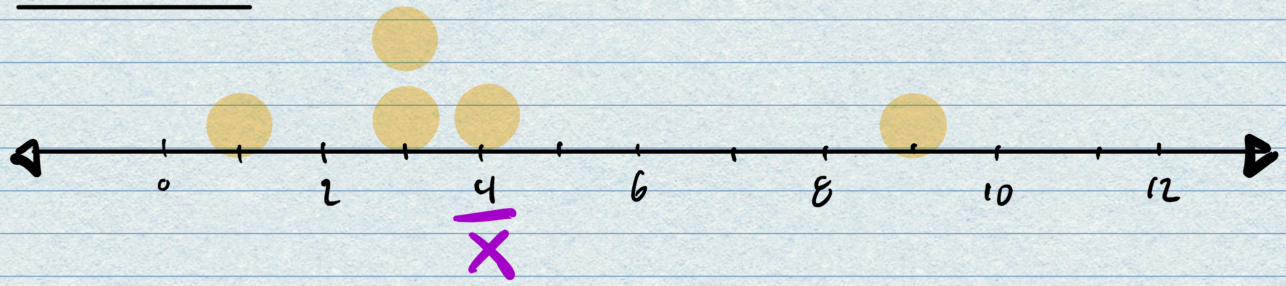
# §2.3 MEASURES OF VARIATION/VARIABILITY

---

3 WAYS: (1) RANGE ✓
(2) VARIANCE
(3) STANDARD DEVIATION

## (2) VARIANCE




DATA: 1 3 3 4 9

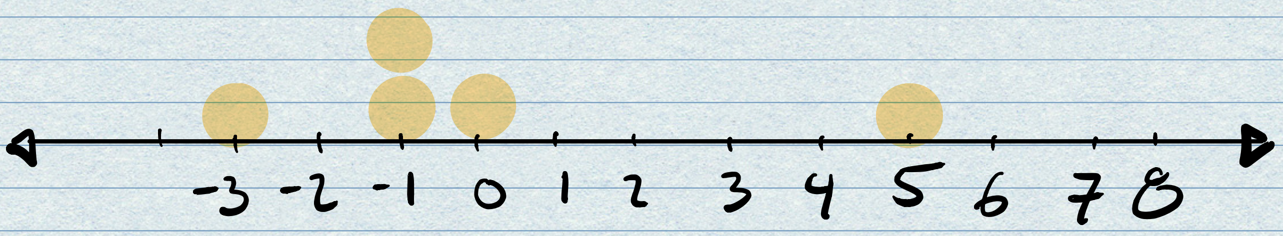
$$\text{MEAN } \bar{x} = \frac{\sum x_i}{n} = \frac{1+3+3+4+9}{5} = 4$$

**Def:** GIVEN A PIECE OF DATA $x$ TAKEN FROM A SET OF DATA WITH MEAN $\bar{x}$ ($\mu$) THE DEVIATION FROM THE MEAN IS $(x - \bar{x})$.

| $x_i$ | $x_i - \bar{x}$ |
|---|---|
| 1 | -3 |
| 3 | -1 |

← DATA VALUE 1 IS 3 BELOW THE MEAN ($\mu$)

| | | |
|---|---|---|
| 3 | -1 | (4) |
| 4 | 0 | |
| 9 | 5 | ← 9 IS 5 ABOVE THE MEAN |




NOW THE CENTER IS 0

$\Rightarrow$ MEAN OF DEVIATIONS FROM THE MEAN IS 0

WE TAKE A PSUEDO-MEAN OF THE SQUARES OF THE DEVIATIONS FROM THE MEAN. (VARIANCE)

| $x_i$ | $x_i - \bar{x}$ | $(x_i - \bar{x})^2$ |
|---|---|---|
| 1 | -3 | $(-3)^2 = 9$ |
| 3 | -1 | $(-1)^2 = 1$ |
| 3 | -1 | $(-1)^2 = 1$ |
| 4 | 0 | $(0)^2 = 0$ |
| 9 | + 5 | $(5)^2 = 25$ + |
| | 0 | 36 ← $\Sigma (x - \bar{x})^2$ |

(ALWAYS!)

**Def:** VARIANCE FOR A POPULATION

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N}$$

IN OUR EXAMPLE, ASSUMING THE DATA COMES FROM A POPULATION (ALL INDIVIDUALS OF INTEREST)

VARIANCE $\sigma^2 = \frac{36}{5} = \boxed{7.2}$

**Def:** VARIANCE FOR A SAMPLE

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

ex. IN OUR EXAMPLE, ASSUMING THE DATA COMES FROM A SAMPLE (SOME INDIVIDUALS OF INTEREST)

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{36}{5-1} = \boxed{9}$$

Q: WHY DO WE DIVIDE BY $n-1$ WHEN CALCULATING THE VARIANCE OF A SAMPLE, $s^2$ ?

THE VARIANCE OF A SAMPLE, $s^2$.

- THE POPULATION HAS MORE VARIATION THAN THE SAMPLE.
- WE WANT THE SAMPLE VARIANCE $s^2$ TO BE A GOOD ESTIMATE FOR THE POPULATION VARIANCE $\sigma^2$, WE MAKE $s^2$ A LITTLE BIT LARGER BY DIVIDING BY A SLIGHTLY SMALLER DENOMINATOR $(n-1)$.

$$\left( \frac{a}{n} < \frac{a}{n-1} \quad \text{SMALLER DENOM.} \Rightarrow \text{BIGGER NUMBER.} \right)$$

DATA WITH A LARGER VARIANCE HAS MORE VARIATION THAN DATA WITH A SMALLER VARIANCE.

NOTE: THE UNIT FOR VARIANCE $(\sigma^2, s^2)$ IS THE SQUARE OF THE UNIT OF OUR MEASUREMENTS.

ex. MASS OF A SAMPLE OF 6 COFFEE BEANS IN GRAMS

7.1 6.2 5.8 9.1 8.8 7.0

COMPUTE THE VARIANCE:

| $x_i$ | $x_i - \bar{x}$ | $(x_i - \bar{x})^2$ |
|---|---|---|

| | | | |
|---|---|---|---|
| 7.1 g | 7.1 - 7.43 | = -.33 g | .1089 $g^2$ |
| 6.2 g | 6.2 - 7.43 | = -1.23 g | 1.5129 $g^2$ |
| 5.8 g | 5.8 - 7.43 | = -1.63 g | 2.6569 $g^2$ |
| 9.1 g | 9.1 - 7.43 | = 1.67 g | 2.7889 $g^2$ |
| 8.8 g | 8.8 - 7.43 | = 1.37 g | 1.8769 $g^2$ |
| 7.0 g | 7.0 - 7.43 | = -.43 g | .1849 $g^2$ |

$$\bar{x} = \frac{7.1 + 6.2 + \ldots + 7.0}{6} = \frac{44.6}{6} \approx 7.43 \text{ g}$$

SAMPLE VARIANCE $s^2 = \dfrac{\sum (x_i - \bar{x})^2}{n-1}$

$$s^2 = \frac{9.1294 \text{ g}^2}{6-1} = \boxed{1.8259} \text{ g}^2$$ (UNIT)

TO OBTAIN A MEASUREMENT OF THE VARIATION OF THE DATA WITH A MORE MEANINGFUL UNIT WE TAKE THE SQUARE-ROOT OF THE VARIANCE.

<u>Def</u>: THE STANDARD DEVIATION OF A DATA SET

(POPULATION) $\sigma = \sqrt{\sigma^2}$

(SAMPLE) $s = \sqrt{s^2}$

ex. $s^2 = 1.8259 \text{ g}^2$ (SAMPLE VARIANCE)

SAMPLE STAND. DEV. $s = \sqrt{s^2}$

$$s = \sqrt{1.8259 \, g^2} = 1.3513 \; g$$ (GRAMS)

| SAMPLE STATISTICS | POPULATION PARAMETERS |
|---|---|
| SAMPLE SIZE $n$ | POPULATION SIZE $N$ |
| SAMPLE MEAN $\bar{x}$ | POPULATION MEAN $\mu$ |
| $\bar{x} = \dfrac{\sum x_i}{n}$ | $\mu = \dfrac{\sum x_i}{N}$ |
| SAMPLE VARIANCE $s^2$ | POPULATION VARIANCE $\sigma^2$ |
| $s^2 = \dfrac{\sum (x_i - \bar{x})^2}{n-1}$ | $\sigma^2 = \dfrac{\sum (x_i - \mu)^2}{N}$ |
| SAMPLE STANDARD DEVIATION $s$ | POPULATION STANDARD DEVIATION |
| $s = \sqrt{s^2}$ | $\sigma = \sqrt{\sigma^2}$ |
| (LOWER CASE) | |

" COMPUTING FORMULA"

ALTERNATIVE FORMULAS:

$$s^2 = \frac{\sum (x_i^2) - \frac{1}{n} \left( \sum x_i \right)^2}{n-1}$$

$$s^2 = \sum (x_i^2) - \frac{1}{n} \left( \sum x_i \right)^2$$

$$\sigma^2 = \frac{\ldots}{N}$$

<u>ex.</u> POPULATION : 3.1 , 4.8 , 5.1 , 2.6 , 4.4

FIND VARIANCE & STAND. DEV.

$$\mu = \frac{\sum x_i}{N} = \frac{3.1 + \ldots + 4.4}{5} = \frac{20}{5} = 4$$

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N} = \frac{4.78}{5} = \boxed{.956}$$

| $x_i$ | $x_i - \mu$ | $(x_i - \mu)^2$ |
|---|---|---|
| 3.1 | -.9 | .81 |
| 4.8 | .8 | .64 |
| 5.1 | 1.1 | 1.21 |
| 2.6 | -1.4 | 1.96 |
| 4.4 | + .4 | + .16 |
| | 0 ✓ | 4.78 |

STAND. DEV. $\sigma = \sqrt{\sigma^2} = \sqrt{.956} = \boxed{.9778}$

**Note:** Round all final answers to 4 decimal places (always in this class)

## § 2.4 The Practical Significance of Standard Deviation

---

The larger the stand. dev. is, the higher the variation of the data. (variability)

**Theorem (Chevychev's Theorem):**

Given any number $k \geq 1$, and a set of $n$ pieces of data, the proportion (fraction) of the data that lies within $k$ standard deviations of the mean is at least

$$\left(1 - \frac{1}{k^2}\right).$$

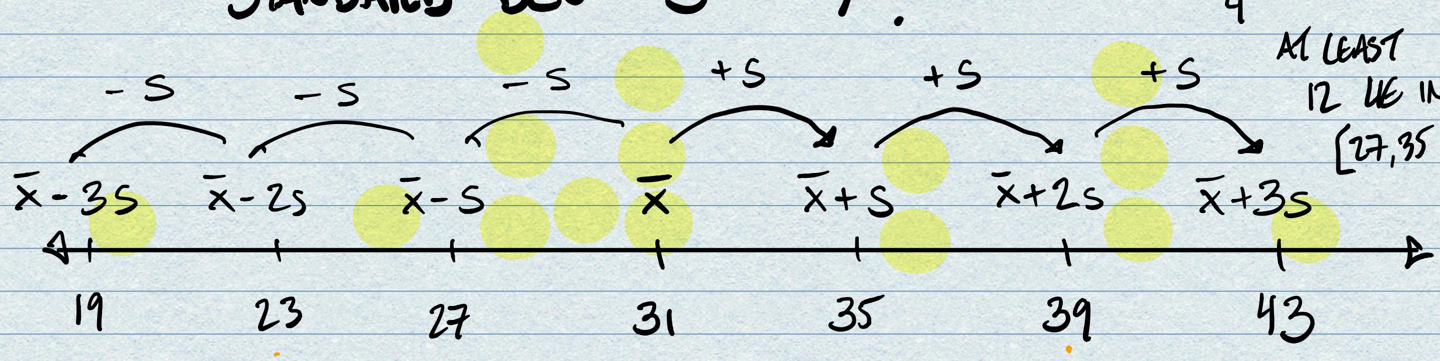
**ex.** Suppose a data set with mean $\bar{x} = 31$

ex. 15 data values, at least $\frac{3}{4}$ lie within $2s$ of $\bar{x}$

STANDARD DEV $s = 4$.

$\frac{3}{4}$ OF 15 = 11.25

AT LEAST 12 LIE IN [27, 35]




(k = 1) #'S FROM 27 TO 35, [27, 35] LIE WITHIN 1s OF $\bar{x}$

(ONE STAND. DEV. OF THE MEAN)

(k = 2) #'S FROM 23 TO 39, [23, 39] LIE WITHIN 2s OF $\bar{x}$

(TWO STAND. DEV. OF THE MEAN)

(k = 3) [19, 43] #'S IN THIS INTERVAL LIE WITHIN 3s OF $\bar{x}$ (THREE STAND. DEV. OF THE MEAN)

THEOREM (CHEBYCHEV'S THEOREM):

GIVEN ANY NUMBER $k \geq 1$, AND A SET OF $n$ PIECES OF DATA, THE PROPORTION (FRACTION) OF THE DATA THAT LIES WITHIN $k$ STANDARD DEVIATIONS OF THE MEAN IS AT LEAST

$$\left(1 - \frac{1}{k^2}\right).$$

- AT LEAST $1 - \frac{1}{1^2} = 0$ (NONE) OF THE DATA LIE WITHIN 1 STAND DEV. OF MEAN [27, 35]

- $k = 2$: AT LEAST $1 - \frac{1}{2^2} = \frac{3}{4}$ (75%) OF THE DATA LIES WITHIN 2 STAND. DEV. OF THE MEAN: [23, 39]

- K = 3 : AT LEAST $1 - \frac{1}{3^2} = \frac{8}{9}$ (88.9%)
  OF THE DATA LIES WITHIN 3 STAND. DEV.
  OF THE MEAN : $[19, 43]$.

NOTE: CHEVYCHEV'S THEOREM HOLDS FOR <u>ALL</u>
DATA SETS, REGARDLESS OF
THE DISTRIBUTION.