Measures of Spread: Standard Deviation

Gaston Sanchez

Summary Measures

- measures of center — *aka* central tendency
- measures of spread — *aka* variation

Measures of Spread

Looking for a value that reflects the **amount of spread**

Equivalent terms

Spread

Scatter

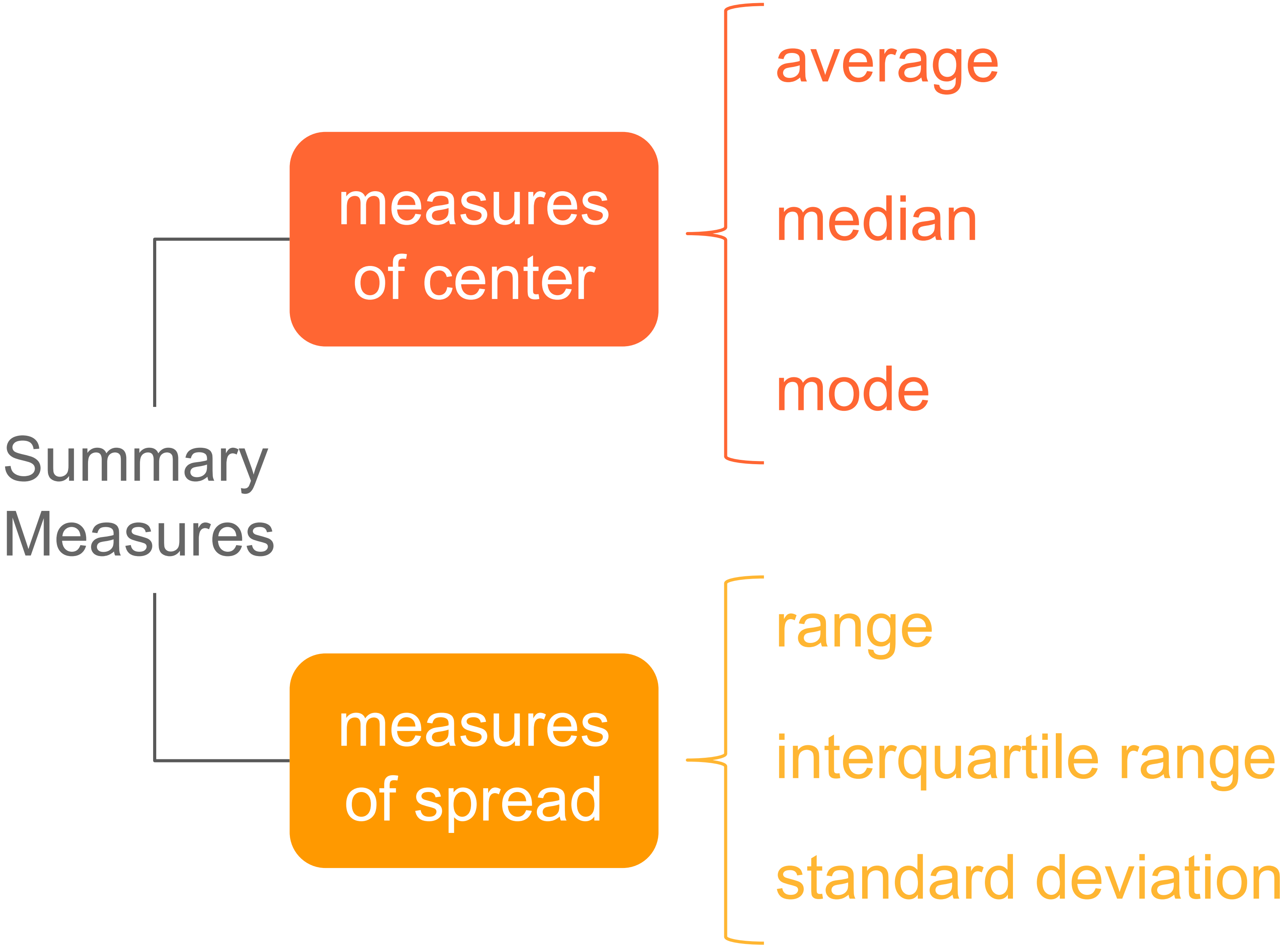
Dispersion

Variation

Measures of spread

Spread Value

Is there a **“representative”** value that tells us how much variation a variable has?


Summary
Measures
measures
of center
average
median
mode
measures
of spread
range
interquartile range
standard deviation

Measures of Spreads

One way to think about measures of spread is in terms of a *typical range of values*.

IQR is a qay to quantify variability relative to the median.

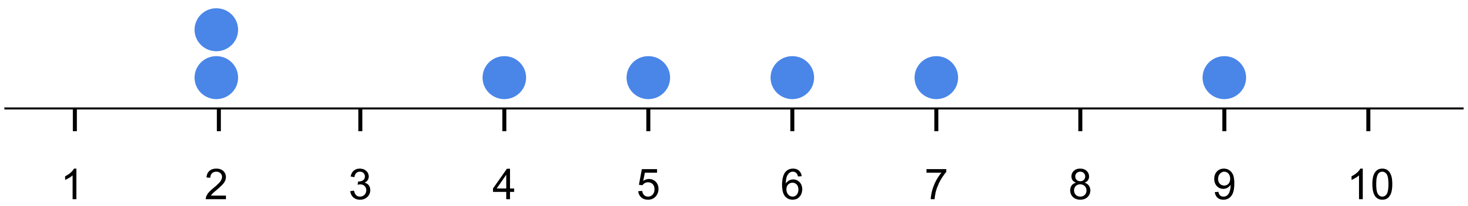
In this slides we develop a numerical measure of spread about the mean.

Average Distance from the Mean (ADM)

Measures of Spreads

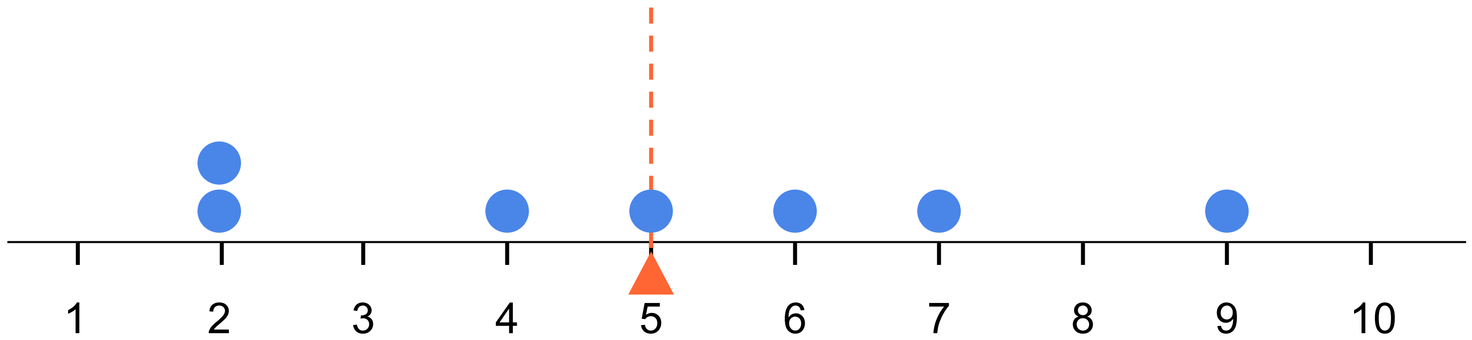
We want to develop a numerical measure of spread that we can use with the mean. In constructing a measure of spread about the mean, we want to compute how far a “typical” number is away from the mean.

Toy data



$$\text{Mean} = \frac{2+2+4+5+6+7+9}{7} = 5$$

Toy data

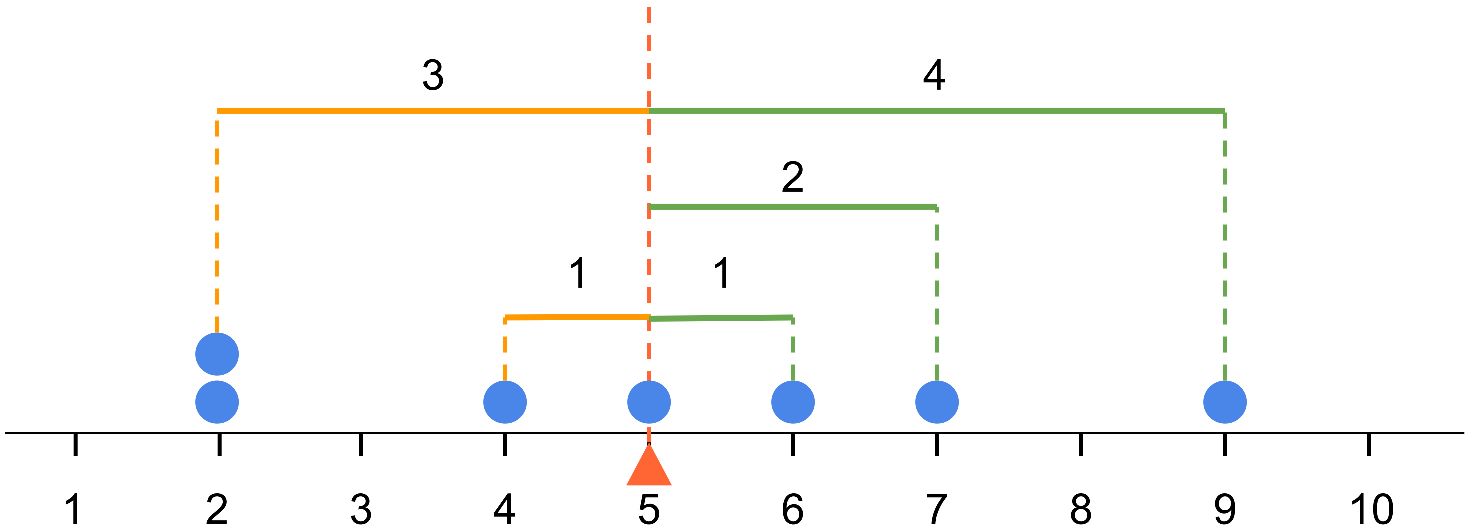


Some data is close to the mean and some data is further from the mean

2 - 5 = -3
2 - 5 = -3
4 - 5 = -1
5 - 5 = 0
6 - 5 = 1
7 - 5 = 2
9 - 5 = 4

deviations from the mean

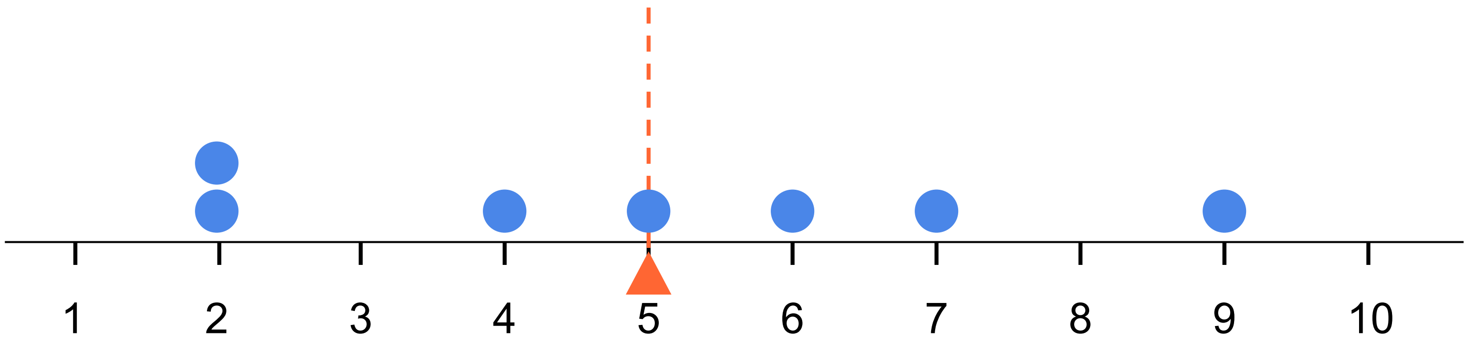
Toy data



Distances between each point and the mean

Our goal is to develop a single measurement that summarizes a typical distance from the mean

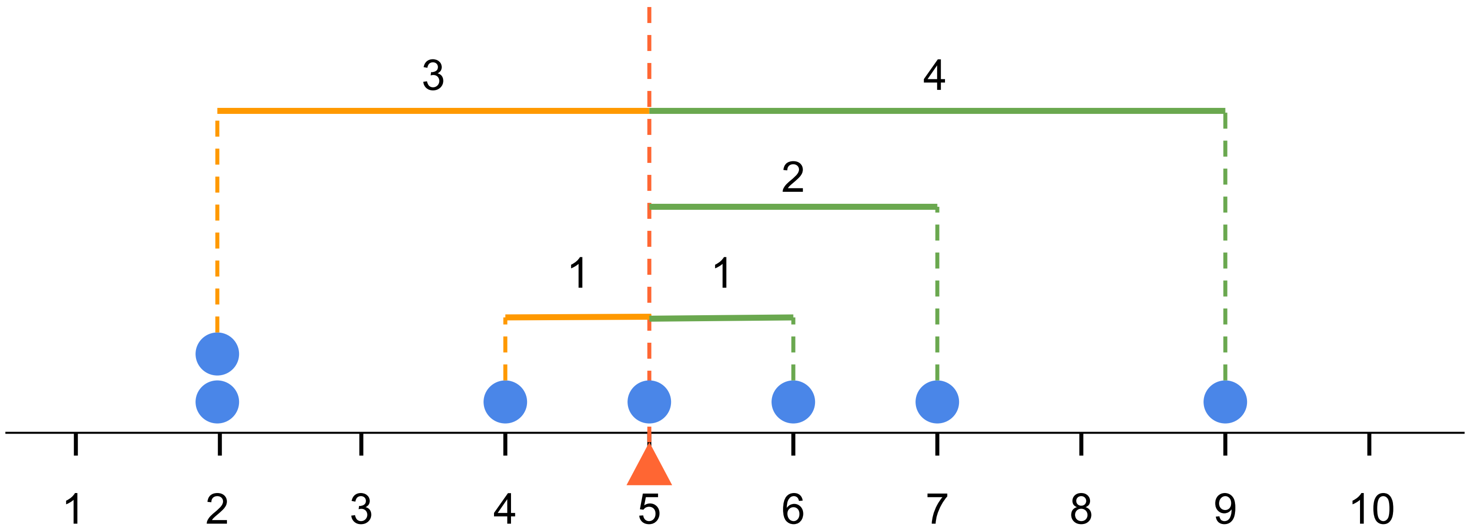
Distance from the Mean



$$|2 - 5| = |-3| = 3$$
$$|2 - 5| = |-3| = 3$$
$$|4 - 5| = |-1| = 1$$
$$|5 - 5| = |0| = 0$$
$$|6 - 5| = |1| = 1$$
$$|7 - 5| = |2| = 2$$
$$|9 - 5| = |4| = 4$$

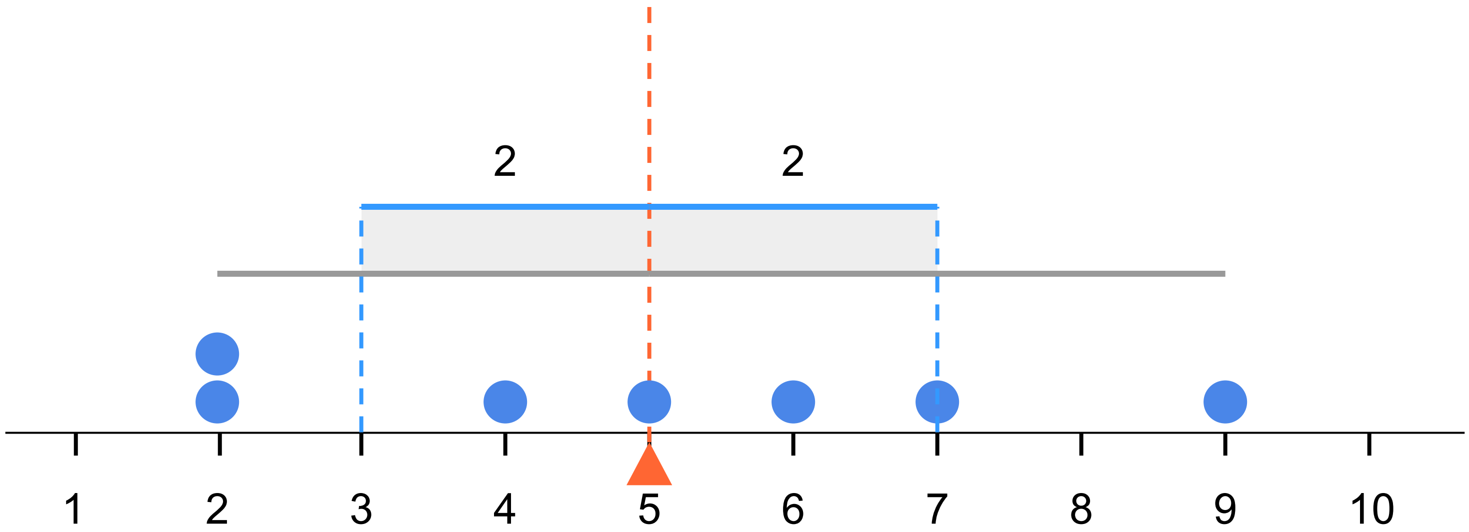
Distance (abs value) of deviations from the mean

Average Distance from the Mean (ADM)



$$\text{ADM} = \frac{3+3+1+0+1+2+4}{7} = 2$$

Average Distance from the Mean (ADM)



We can visualize the ADM using a display similar to a boxplot

About ADM

The **ADM** is a reasonable measure of spread about the mean, but there is another measure that is used much more often: the standard deviation (**SD**).

The standard deviation behaves very much like the average deviation. So all of the work we have done so far is useful in understanding standard deviation. We discuss **Root Mean Square** (RMS) next, which will help us to understand SD.

Root Mean Square (R.M.S.)

Root Mean Square (R.M.S.)

The RMS provides an idea of the **size** of values

RMS

$$\text{r.m.s.} = \sqrt{\text{average of (entries)}^2}$$

How to compute the Root Mean Square

1. **Square** all the entries

2. Take the **mean (average)** of the squares

3. Take the **square root** of the mean

About RMS

How small or big are these values?

0, 5, -8, 7, -3

average = 0.2

range = 7 - (-8) = 15

About RMS

How small or big are these values?

$$0, \quad 5, \quad -8, \quad 7, \quad -3$$

$$0^2 \quad 5^2 \quad (-8)^2 \quad 7^2 \quad (-3)^2$$

About RMS

How small or big are these values?

0, 5, -8, 7, -3

$$\text{r.m.s} = \sqrt{\frac{0^2 + 5^2 + (-8)^2 + 7^2 + (-3)^2}{5}}$$

r.m.s = 5.42

Standard Deviation

Standard Deviation:
How far the data spread out around the average

Deviations

Deviation

deviation from the average = entry - average

Standard Deviation

Root Mean Square size of the deviations

How to find the SD?

What is the SD of the following numbers?

20, 10, 15, 15

step 1

$$\text{average} = \frac{20 + 10 + 15 + 15}{4} = 15$$

How to find the SD?

step 2

Deviations from the avergae

$$20 - 15 = 5$$

$$10 - 15 = -5$$

$$15 - 15 = 0$$

$$15 - 15 = 0$$

How to find the SD?

step 3 R.M.S. size of the deviations

$$SD = \sqrt{\frac{5^2 + (-5)^2 + 0^2 + 0^2}{4}}$$

SD = 3.5355

SD Formula

$$SD = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2}{n}}$$

SD Formula

$$SD = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2}$$

Alternative Formula

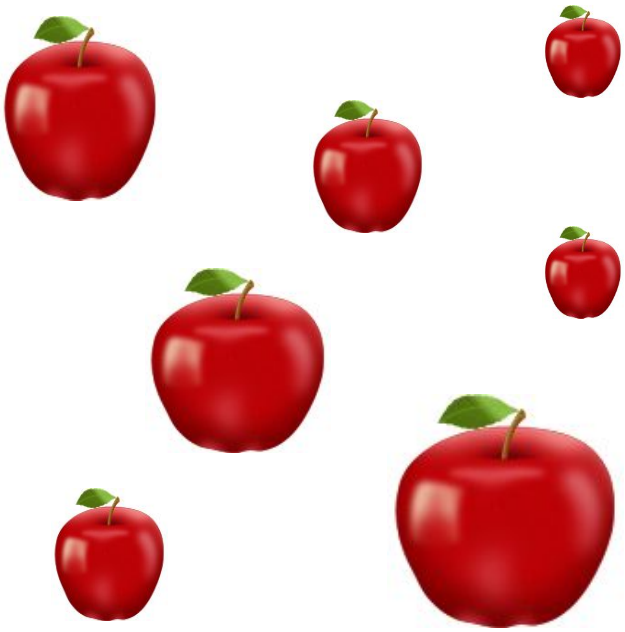
SD alternative formula

$$\text{SD} = \sqrt{\text{average of (entries}^2) - \text{(average of entries)}^2}$$

average of
squared entries

square of the
average of entries

Apple dataset



num	Weight oz	Carbs	Acidity	Shape
1	5	20.0	medium	round
2	6	24.3	high	oval
3	7	25.0	medium	round
4	7	25.5	low	square
5	6	24.7	medium	round
6	8	26.1	low	round
7	6	25.2	high	square
8	9	23.7	high	oval
9	10	21.0	low	round
10	8	27.4	medium	oval

Finding the SD

Apple weight values

5, 6, 7, 7, 6, 8, 6, 9, 10, 8

Mean (average) weight

$\bar{x} = 7.2$

Finding the SD

$$(5 - 7.2)^2 =$$

$$(6 - 7.2)^2 =$$

$$(7 - 7.2)^2 =$$

$$(7 - 7.2)^2 =$$

$$(6 - 7.2)^2 =$$

$$(8 - 7.2)^2 =$$

$$(6 - 7.2)^2 =$$

$$(9 - 7.2)^2 =$$

$$(10 - 7.2)^2 =$$

$$(8 - 7.2)^2 =$$

Finding the SD

$$(5 - 7.2)^2 = (-2.2)^2$$

$$(6 - 7.2)^2 = (-1.2)^2$$

$$(7 - 7.2)^2 = (-0.2)^2$$

$$(7 - 7.2)^2 = (-0.2)^2$$

$$(6 - 7.2)^2 = (-1.2)^2$$

$$(8 - 7.2)^2 = (0.8)^2$$

$$(6 - 7.2)^2 = (-1.2)^2$$

$$(9 - 7.2)^2 = (1.8)^2$$

$$(10 - 7.2)^2 = (2.8)^2$$

$$(8 - 7.2)^2 = (0.8)^2$$

Finding the SD

$$(5 - 7.2)^2 = (-2.2)^2 = 4.84$$
$$(6 - 7.2)^2 = (-1.2)^2 = 1.44$$
$$(7 - 7.2)^2 = (-0.2)^2 = 0.04$$
$$(7 - 7.2)^2 = (-0.2)^2 = 0.04$$
$$(6 - 7.2)^2 = (-1.2)^2 = 1.44$$
$$(8 - 7.2)^2 = (0.8)^2 = 0.64$$
$$(6 - 7.2)^2 = (-1.2)^2 = 1.44$$
$$(9 - 7.2)^2 = (1.8)^2 = 3.24$$
$$(10 - 7.2)^2 = (2.8)^2 = 7.84$$
$$(8 - 7.2)^2 = (0.8)^2 = 0.64$$

Finding the SD

$$
\left.
\begin{array}{lcl}
(5 - 7.2)^2 = (-2.2)^2 & = & 4.84 \\
(6 - 7.2)^2 = (-1.2)^2 & = & 1.44 \\
(7 - 7.2)^2 = (-0.2)^2 & = & 0.04 \\
(7 - 7.2)^2 = (-0.2)^2 & = & 0.04 \\
(6 - 7.2)^2 = (-1.2)^2 & = & 1.44 \\
(8 - 7.2)^2 = (0.8)^2 & = & 0.64 \\
(6 - 7.2)^2 = (-1.2)^2 & = & 1.44 \\
(9 - 7.2)^2 = (1.8)^2 & = & 3.24 \\
(10 - 7.2)^2 = (2.8)^2 & = & 7.84 \\
(8 - 7.2)^2 = (0.8)^2 & = & 0.64
\end{array}
\right\} \sum = 21.6
$$

Finding the SD

$$(5 - 7.2)^2 = (-2.2)^2 = 4.84$$
$$(6 - 7.2)^2 = (-1.2)^2 = 1.44$$
$$(7 - 7.2)^2 = (-0.2)^2 = 0.04$$
$$(7 - 7.2)^2 = (-0.2)^2 = 0.04$$
$$(6 - 7.2)^2 = (-1.2)^2 = 1.44$$
$$(8 - 7.2)^2 = (0.8)^2 = 0.64$$
$$(6 - 7.2)^2 = (-1.2)^2 = 1.44$$
$$(9 - 7.2)^2 = (1.8)^2 = 3.24$$
$$(10 - 7.2)^2 = (2.8)^2 = 7.84$$
$$(8 - 7.2)^2 = (0.8)^2 = 0.64$$

$$SD = \sqrt{\frac{21.6}{10}}$$

$$SD = 1.47$$

SD of apples

Average weight = 7.2 oz

But there are deviations from the average

Some apples are heavier than the average

Some apples are lighter than the average

How big are these deviations? SD is 1.47 oz

Variance

Variance

Measure of the distribution or spread of data around the average value

Variance Formula

$$S^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

Variance

20, 10, 15, 15

average = 15

$$\text{Var} = \frac{(20-15)^2 + (10-15)^2 + (15-15)^2 + (15-15)^2}{4}$$

Var = 12.5

A closer look at the SD formula

Understanding Variance

$$SD = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2}$$

Another way to think about SD

How much variation in data?

How much different values are?

this is kind of an open question?