§2.4 Numerical Measure of Variability

In this section we look at several ways of describing the variability, or spread, of a data set.

Example

Sample 1

Measurement, x	1	1	2	3	4	5	5
Deviation, $x-\bar{x}$	-2	-2	-1	0	1	2	2

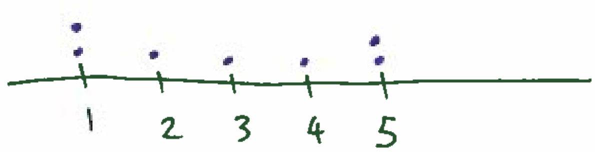
$\bar{x} = \frac{21}{7} = 3$

Sample 2

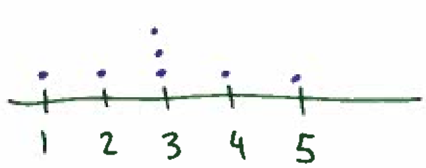
Measurement, x	1	2	3	3	3	4	5
Deviation, $x-\bar{x}$	-2	-1	0	0	0	1	2

$\bar{x} = \frac{21}{7} = 3$

Sample 1



Sample 2



In these two samples, the mean is the same, but the spread of data is quite different.

<u>Definition</u> the <u>range</u> of quantitative data is equal to the largest measurement minus the smallest measurement.

<u>Example</u> The ranges for the two data sets above are both:

$$\text{range} = 5 - 1 = 4$$

Note that range does not differentiate the spread of the two data sets.
Also note that the sum of the Deviation from the mean is zero for both samples.
This will always be true, so ...

<u>Definition</u> the <u>sample variance</u> for n measurements is equal to the sum of the squared deviations from the mean, divided by $(n-1)$. The simbol s^2 is used to represent the sample variance.

$$s^2 = \frac{\sum_{k=1}^{n} (x_k - \bar{x})^2}{n-1} = \frac{\sum_{k=1}^{n} x_k^2 - \frac{1}{n}\left(\sum_{k=1}^{n} x_k\right)^2}{n-1}$$

Example Calculate the sample variance for the two sample data sets above.

In both sample data sets, $n = 7$

Using the first formula on sample 1 we get

$$s^2 = \frac{(-2)^2 + (-2)^2 + (-1)^2 + (0)^2 + (1)^2 + (2)^2 + (2)^2}{(7-1)} = 3$$

Using the second formula on sample 1 we get

$$\sum_{k=1}^{7} x_k = 21 \quad , \quad \sum_{k=1}^{7} x_k^2 = 1^2 + 1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 5^2 = 81$$

so

$$s^2 = \frac{81 - \frac{1}{7}(21)^2}{(7-1)} = 3$$

For sample 2 we get

$$s^2 = \frac{(-2)^2 + (-1)^2 + (0)^2 + (0)^2 + (0)^2 + (1)^2 + (2)^2}{(7-1)} = 1.666\ldots$$

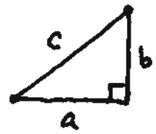
$$= \frac{5}{3}$$

Note that sample variance did differentiate between the two data sets. The first data set has a greater variance than the second one.

Definition the sample standard deviation is the square root of the sample variance.

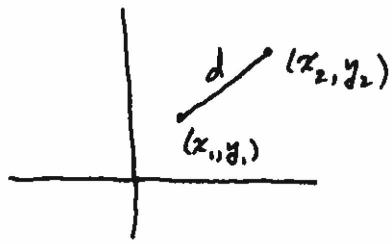
$$s = \sqrt{s^2}$$

This is really an extension to the distance formula in geometry:



$$c = \sqrt{a^2 + b^2}$$

— or —



$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

<u>Example</u> For the 2 sample data sets, the standard deviations are:

For sample 1 : $s = \sqrt{3} = 1.73$

For sample 2 : $s = \sqrt{\frac{5}{3}} = 1.29$