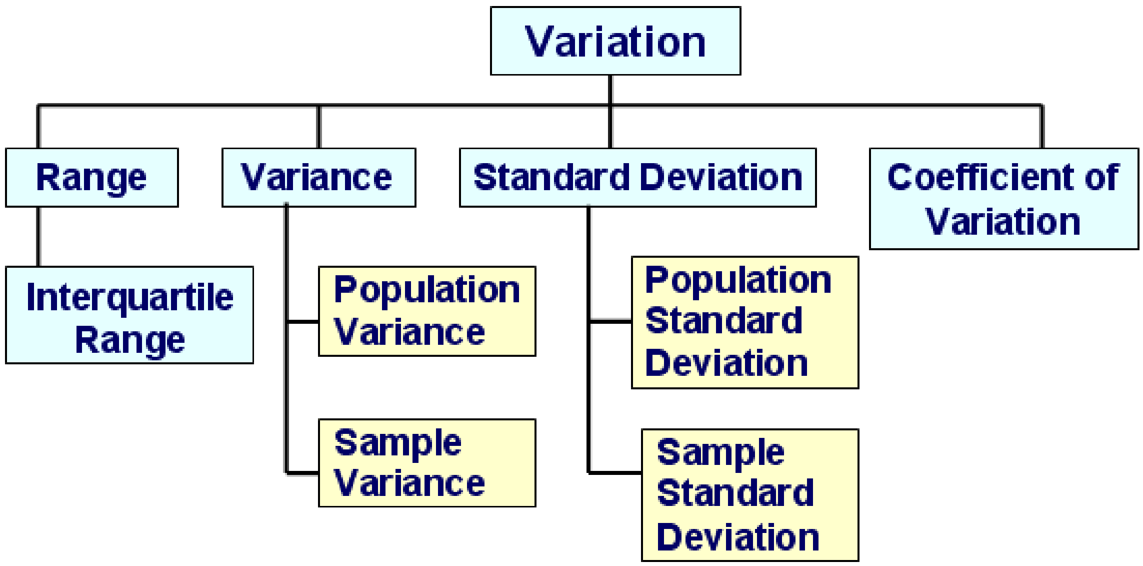
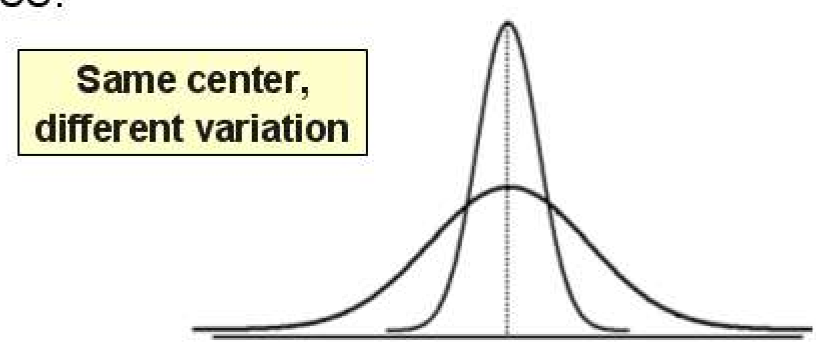
4.2 Notes

Teacher Version

Measures of Variation



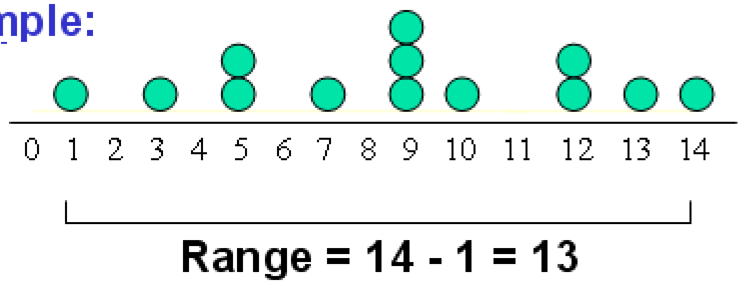
Measures of variation give information on the **spread** or **variability** of the data values.



- **The Range**

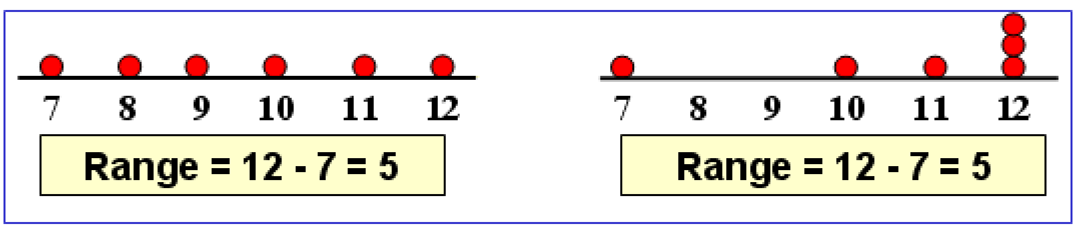
Range = (highest value) – (lowest value)

Example:



<u>Comment</u>: The range is the simplest measure of variation. In certain limited situation it can be very useful. It has obvious disadvantages:

1. It ignores the way in which data are distributed



2. Sensitive to outliers:

1,1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,3,3,3,3,4,5

Range = 5 - 1 = 4

1,1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,3,3,3,3,4,120

Range = 120 - 1 = 119

- **The Variance**

IMPORTANT TO USE CORRECT VARIABLE NOTATIONS
*** small n=sample size while
*** capital N=population size

Population variance:

$$\sigma^2 = \frac{\sum_{i=1}^{N}(X_i - \mu)^2}{N} = \frac{(X_1 - \mu)^2 + (X_2 - \mu)^2 + \cdots + (X_N - \mu)^2}{N}$$

Sample variance

$$s^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1} = \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2}{n-1}$$

which is also called a point estimation of population variance.

Comments:

1. σ^2 is the average squared distance of observations to the population mean.
2. The unit of σ^2 is the square of the unit of the variable.

- **The Standard Deviation**

Population standard deviation: $\sigma = \sqrt{\sigma^2}$, that is,

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(X_i - \mu)^2}{N}} = \sqrt{\frac{(X_1 - \mu)^2 + (X_2 - \mu)^2 + \cdots + (X_N - \mu)^2}{N}}$$

Sample standard deviation: $s = \sqrt{s^2}$ or

$$s = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}} = \sqrt{\frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2}{n-1}}$$

which is called a point estimation of population standard deviation.

Comments

1. σ and σ^2 are always positive.
2. The units of σ are the units of the variable.

An alternative formula for computing s or s^2:

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

Remark: There are several different forms of formulas can be used to calculate the standard deviation of a given data set (sample or population). The tabular computation is recommended when doing manual computation:

An Illustrative Example: suppose we have a data set A={1,4,7}

X	$X-\bar{X}$	$(X-\bar{X})^2$
1	-3	9
4	0	0
7	3	9
Total:12	0	18

FILL IN TABLE

What is the sample mean and sample standard deviation?

Based on the above table, we can see that the mean is 12/3=4, the standard deviation is $\sqrt{18/(3-1)} = 3$.

Another more general example:

Fill in the table to calculate the sample standard deviation.
Use 41.5 as the sample mean (xbar).

x	**Deviation: *x* –** xbar	**Squares: (*x*** - xbar)**²**
41	41 – 41.5 = –0.5	$(-0.5)^2 = 0.25$
38	= –3.5	= 12.25
39	= –2.5	= 6.25
45	= 3.5	= 12.25
47	= 5.5	= 30.25
41	= –0.5	= 0.25
44	= 2.5	= 6.25
41	= –0.5	= 0.25
37	= –4.5	= 20.25
42	= 0.5	= 0.25
Total	Σ = 0	Σ = 88.5

$$s^2 = \frac{\Sigma(x-\bar{x})^2}{n-1} = \frac{88.5}{10-1} \approx 9.8$$

$$s = \sqrt{s^2} = \sqrt{\frac{88.5}{9}} \approx 3.1$$

Examples of datasets that have the same means with different variations