# Measures of Central Tendency

**Central tendency** describes the extent in which values of a numerical variable, group around a central value

**Variation** measures the amount of dispersion or scattering away from the central value that the numerical variable exhibit

The **shape** of the variable describes the pattern of the distribution of values from the lowest value to the highest value

# The Sample Mean

For ungrouped data, the **sample mean** is the sum of all the data points in the sample divided by the total number of values.

To find the sample mean, we use the following formula

$$\textit{Sample Mean} = \frac{\text{Sum of all values in a sample}}{\text{Number of values in a sample}}$$

The mean for sample data or population data are computed the same way…

The sample mean of the score data = 81

66, 71, 73, 77, 77, 88, 90, 93, 94

(assume this data set represents sample data)

$$\bar{x} = \frac{\sum x}{n}$$

$$\bar{x} = \frac{\sum x}{n} = \frac{729}{9} = 81$$

# The Median

The **median** is the midpoint of the values after they have been ordered from the minimum to the maximum

1. There are as many values above the median as there are below it in the data array

2. For an even set of values, the median will be the arithmetic average of the two middle numbers

The median of the score data = 77

66, 71, 73, 77, 77, 88, 90, 93, 94

# The Mode

The **mode** is the value of the observation that appears most frequently in a set of data

66, 71, 73, 77, 77, 88, 90, 93, 94

The mode for the score data = 77

# Measures of Variation and Shape

**Variation** measures the amount of dispersion or scattering away from the central value that the numerical variable exhibit

The **range** is the difference between the maximum and minimum value

The range measures the total spread in the set of data

The range <u>does not</u> indicate whether the values are evenly distributed, clustered near the middle or clustered near one of the extremes

$$Range = x_{maximum} - x_{minimum}$$

66, 71, 73, 77, 77, 88, 90, 93, 94

The range for the score data = 94 - 66 = 28

# The Variance and Standard Deviation

Unlike the range, the **variance** and **standard deviation** does account for how the data is distributed about the mean value

Both the variance and standard deviation are calculated using the squares of the deviations of the data points from the mean value

Since we are dealing with the squares of the deviations, the values of variance and standard deviation will always be positive

# The Sample Variance

The **sample variance** formula is given below

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

$s^2$ = sample variance

x = value of each observation in the sample

$\overline{x}$ = mean of the sample

n = number of observations in the sample

An easier way to compute the sample variance is given by the alternate formula. It will reduce computation time by eliminating multiple subtraction operations.

66, 71, 73, 77, 77, 88, 90, 93, 94

(assume this data set represents sample data)

$$s^2 = \frac{(66-81)^2 + (71-81)^2 + (73-81)^2 \ldots}{9-1} = \frac{864}{8} = 108$$

# The Sample Standard Deviation

The **sample standard deviation** is the square root of the sample variance

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

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

s = sample standard deviation
x = value of each observation in the sample
$\overline{x}$ = mean of the sample
n = number of observations in the sample

66, 71, 73, 77, 77, 88, 90, 93, 94
(assume this data set represents sample data)

$$s = \sqrt{\frac{(66-81)^2 + (71-81)^2 + (73-81)^2 \ldots}{9-1}} = \sqrt{\frac{864}{8}} = \sqrt{108} = 10.39$$

# The Population Mean

For ungrouped data, the **population mean** is the sum of all the data points in the population and divided by the total number of values.

To find the population mean, we use the following formula

$$Population\ Mean = \frac{\text{Sum of all values in a population}}{\text{Number of values in a population}}$$

The mean for sample data or population data are computed the same way…

The population mean of the score data = 78.5

61, 66, 68, 70, 71, 73, 75, 76, 77, 77, 77, 79, 80, 81, 85, 88, 89, 90, 93, 94
(assume this data set represents population data)

$$\mu = \frac{\sum x}{N}$$

$$\mu = \frac{\sum x}{N} = \frac{1570}{20} = 78.5$$

# Population Variance

The **population variance** is the arithmetic mean of the squared deviations from the population mean

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

$\sigma^2$ = symbol for the population variance, “sigma squared”

$\mu$ = the arithmetic mean of the population

x = value of each observation in the population

N = number of observations in the population

61, 66, 68, 70, 71, 73, 75, 76, 77, 77, 77, 79, 80, 81, 85, 88, 89, 90, 93, 94
(assume this data set represents population data)

$$\sigma^2 = \frac{(61-78.55)^2 + (66-78.55)^2 + (68-78.55)^2 \ldots}{20} = \frac{1591}{20} = 79.55$$

# Population Standard Deviation

The **population standard deviation** is the square root of the population variance

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

σ = symbol for the population standard deviation

μ = the arithmetic mean of the population

x = value of each observation in the population

N = number of observations in the population

61, 66, 68, 70, 71, 73, 75, 76, 77, 77, 77, 79, 80, 81, 85, 88, 89, 90, 93, 94
(assume this data set represents population data)

$$\sigma = \sqrt{\frac{(61-78.55)^2 + (66-78.55)^2 + (68-78.55)^2 \ldots}{20}} = \sqrt{\frac{1591}{20}} = \sqrt{79.55} = 8.92$$

# Why the n-1?

Suppose there is a population with **N** items.

Suppose that we take samples of size **n** items from that population. If we could list all possible samples of **n** items that could be selected from the population of **N** items, then we could find the sample variance for each possible sample.

The following should be true:

The average of the sample variances for all possible samples should equal the population variance. It is a logical property and a reasonable thing to happen. This is called "**unbiased sampling**".

If we divide by **(n −1)** when calculating the sample variance, then it turns out that the average of the sample variances for all possible samples will equal the population variance.

So the sample variance is what we call an **unbiased** estimate of the population variance.

Dividing by **n** alone does not give an **unbiased** estimate of the population variance.

$$s=\sqrt{\frac{\sum\left(x-\bar{x}\right)^2}{n-1}} \qquad \sigma=\sqrt{\frac{\sum\left(x-\mu\right)^2}{N}}$$