Section 10.1 - Correlation

A **correlation** exists between two variables when the values of one variable are somehow associated with the values of the other variable.

Scatterplots are often used to identify correlation.

The **linear correlation coefficient** r measures the strength of the linear correlation between paired quantitative x- and y-values in a sample.

$$r = \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{n \sum x^2 - (\sum x)^2}\sqrt{n \sum y^2 - (\sum y)^2}}$$

See the requirements on pages 472-473.

Properties of the Linear Correlation Coefficient

- The value of r is always between -1 and 1 (inclusive).
- The value of r is independent of scale, invariant under transformation, and not affected by the choice of x or y.
- r measures the strength of the *linear* relationship.
- r is sensitive to outliers.

We will use our calculators (or computers) to compute the linear correlation coefficient and an equation for the "best" linear equation that describes the relationship between the variables.

Example

The following data were collected when studying a voltage-controlled amplifier. All values are in volts.

Input Voltage	1.32	1.75	2.23	2.54	3.01	3.37	3.79	4.25	4.62	5.04	5.50
Output Voltage	14	13	12	11	10	9	8	7	6	5	4

In addition to computing the linear correlation coefficient, we can use a P-value to determine whether a linear correlation exists or not (at a certain significance level).

If the computed P-value is less than the significance level, we conclude that there is a linear correlation between the variables. Otherwise, there is not sufficient evidence to support the conclusion of a linear relationship.

The value of r^2 is the proportion of the variation in y that is explained by the linear relationship between x and y.