

### 3.2.2: Correlation Coefficient

The **correlation coefficient** (represented by the letter  $r$ ) measures both the direction and strength of a linear relationship or association between two variables. The value  $r$  will always take on a value between -1 and 1. Values close to zero indicate a very weak correlation. Values close to 1 or -1 indicate a very strong correlation. The correlation coefficient should not be used for non-linear correlation.

It is important to ignore the sign when determining **strength** of correlation. For example,  $r = -0.75$  would indicate a stronger correlation than  $r = 0.62$ , since -0.75 is farther from zero.

We will use technology to calculate the correlation coefficient, but formulas for manually calculating  $r$  are presented at the end of this section.

#### Interpreting the correlation coefficient ( $r$ )

$$-1 \leq r \leq 1$$

$r = 1$  means perfect positive correlation

$r = -1$  means perfect negative correlation

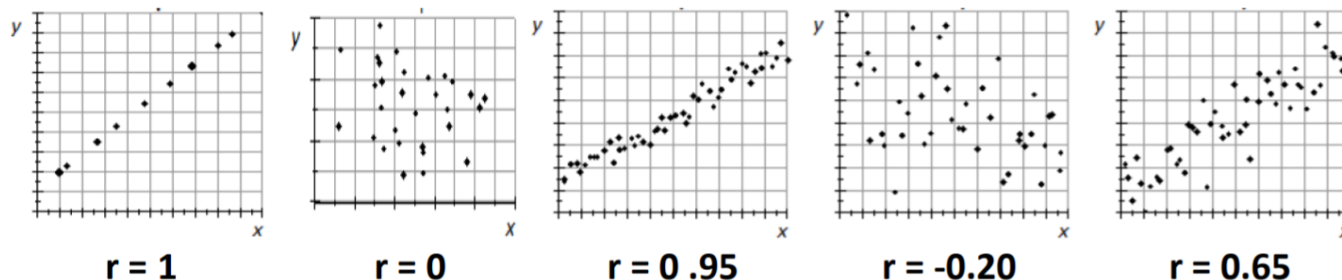
$r = 0$  mean no correlation

The farther  $r$  is from zero, the stronger the correlation

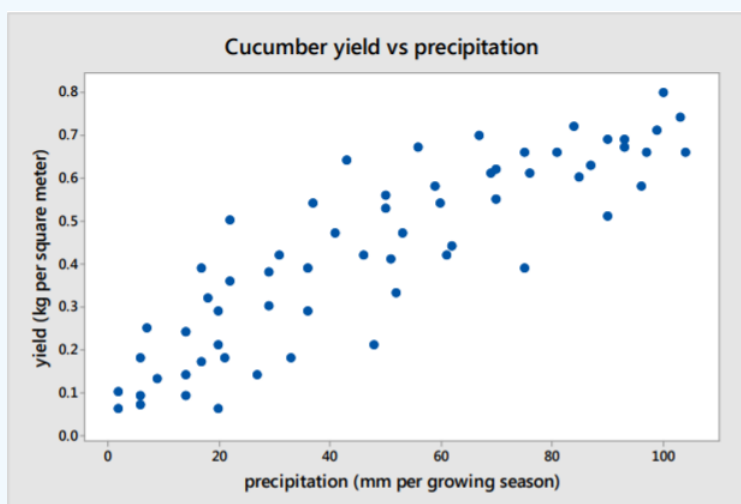
$r > 0$  means positive correlation

$r < 0$  means negative correlation

#### Some Examples



#### Example: Cucumber yield and rainfall

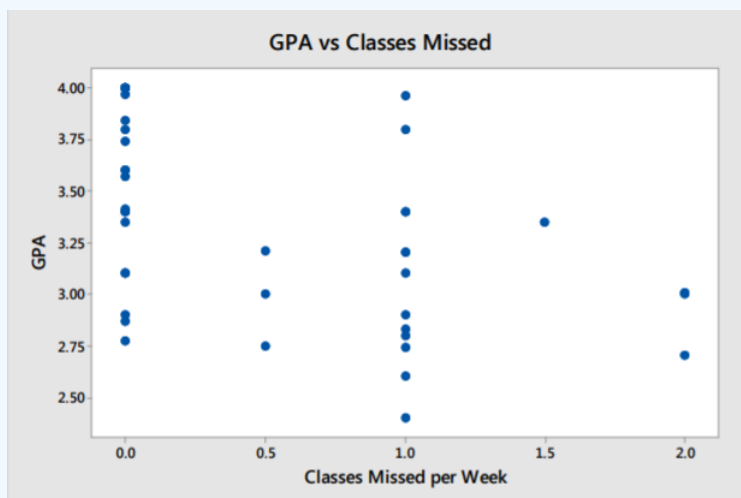


This scatterplot represents randomly collected data on growing season precipitation and cucumber yield.

$r = 0.871$  indicating strong positive correlation.

#### Example: GPA and missing class

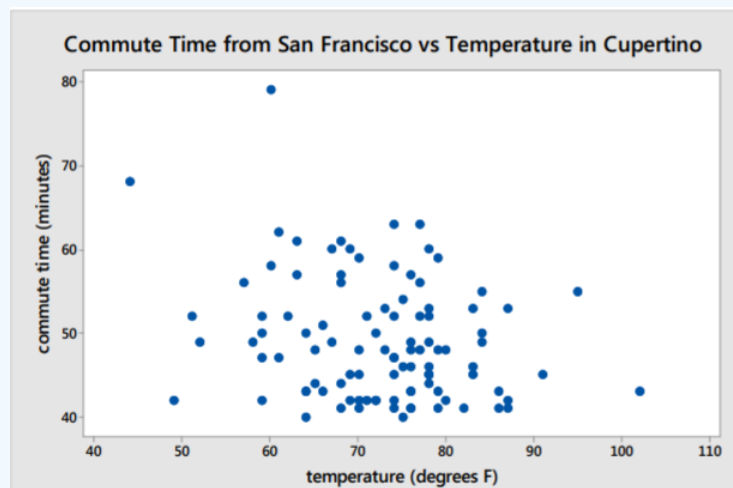
A group of students at Georgia College conducted a survey asking random students various questions about their academic profile. One part of their study was to see if there is any correlation between various students' GPA and classes missed.



$r = -0.236$  indicating weak negative correlation.

#### Example: Commute times and temperature

A mathematics instructor commutes by car from his home in San Francisco to De Anza College in Cupertino, California. For 100 randomly selected days during the year, the instructor recorded the commuting time and the temperature in Cupertino at time of arrival.



$r = -0.02$  indicating no correlation.

### Calculating the correlation coefficient

Manually calculating the correlation coefficient is a tedious process, but the needed formulas and one simple example are presented here:

### Formulas for calculating the correlation coefficient ( $r$ )

$$r = \frac{SSXY}{\sqrt{SSX \cdot SSY}}$$

$$SSX = \sum X^2 - \frac{1}{n}(\sum X)^2$$

$$SSY = \sum Y^2 - \frac{1}{n}(\sum Y)^2$$

$$SSXY = \sum XY - \frac{1}{n}(\sum X \cdot \sum Y)$$

### Example: Sunglasses sales and rainfall

A company selling sunglasses determined the units sold per 1000 people and the annual rainfall in 5 cities.

X = rainfall in inches

Y = sales of sunglasses per 1000 people.

X	Y
10	40
15	35
20	25
30	25
40	15

### Solution

First, find the following sums:

$$\sum X, \sum Y, \sum X^2, \sum Y^2, \sum XY$$

	X	Y	X <sup>2</sup>	Y <sup>2</sup>	XY
	10	40	100	1600	400
	15	35	225	1225	525
	20	25	400	625	500
	30	25	900	625	750
	40	15	1600	225	600
<b>Σ</b>	<b>115</b>	<b>140</b>	<b>3225</b>	<b>4300</b>	<b>2775</b>

Then, find  $SSX$ ,  $SSY$ ,  $SSXY$

$$SSX = 3225 - \frac{115^2}{5} = 580$$

$$SSY = 4300 - \frac{140^2}{5} = 380$$

$$SSXY = 2775 - \frac{(115)(140)}{5} = -445$$

Finally, calculate  $r$

$$r = \frac{SSXY}{\sqrt{SSX \cdot SSY}} = \frac{-445}{\sqrt{580 \cdot 380}} = -0.9479$$

The correlation coefficient is -0.95, indicating a strong, negative correlation between rainfall and sales of sunglasses.

---

3.2.2: Correlation Coefficient is shared under a [CC BY-SA](#) license and was authored, remixed, and/or curated by LibreTexts.

- **3.6.2: Correlation Coefficient** by [Maurice A. Geraghty](#) is licensed [CC BY-SA 4.0](#). Original source: <http://nebula2.deanza.edu/~mo/holisticInference.html>.