

1. DISPLAYING RELATIONSHIP: SCATTERPLOTS

We want to find the relationship between variables in data.

Example 1.1.

Does the grade on Exam1 predict the grade on Exam2 in a class?

Does the temperature influences the number of ice cream sold in the market?

Definition 1.2.

A _____ measures an outcome or result of a study.

An _____ is a variable that we think explains or causes changes in the response variables.

Example 1.3. The data below is the average monthly temperature (in °F) and the average daily kWh of electricity used at a particular home each month.

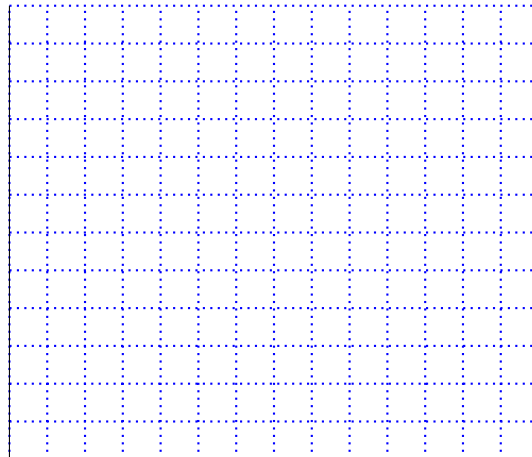
Temp (°F)	86	85	83	77	69	62	55	51	52	61	71	80
Electricity (kWh)	76	70	66	38	26	22	25	33	30	26	38	60

Which is the explanatory variable and which is the response variable?

Definition 1.4. A _____ is a graph of two values with the explanatory variable on the horizontal axis and the response variable on the vertical axis.

Example 1.5. The data below is the average monthly temperature (in °F) and the average daily kWh of electricity used at a particular home each month. Display this information in a scatter plot.

Temp (°F)	86	85	83	77	69	62	55	51	52	61	71	80
Electricity (kWh)	76	70	66	38	26	22	25	33	30	26	38	60



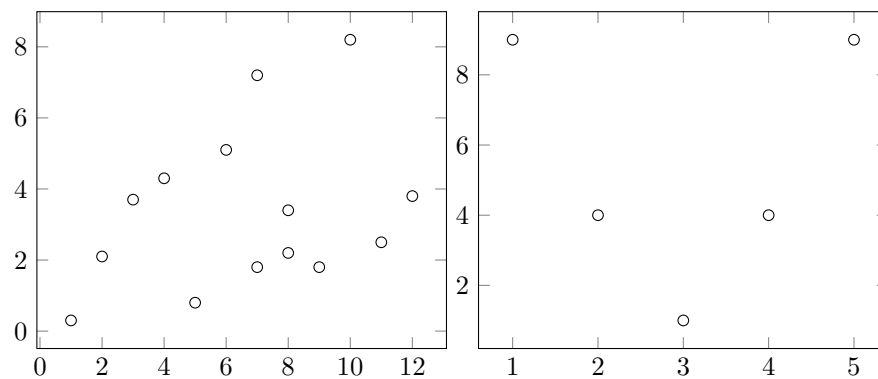
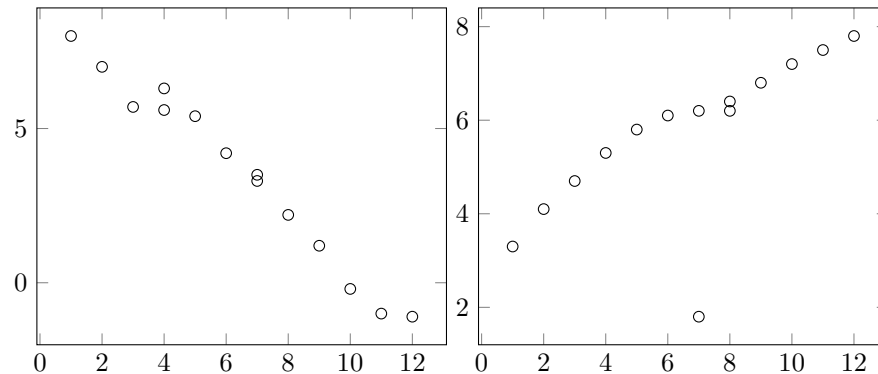
To examine a scatterplot,

- Look at the overall pattern:
 - Is the form linear or not?
 - What is the general direction?
 - How strong is the relationship
- Look for outliers — do any data points fall outside of the general pattern?

Definition 1.6. Two variables are _____ if an increase in one variable tends to accompany an increase in the other variable.

Two variables are _____ if an increase in one variable tends to accompany a decrease in the other variable.

Example 1.7. Examine the scatterplots below and determine if there is an association between the explanatory and response variables. Is it a linear relationship? Is it a strong relationship? Are there outliers?

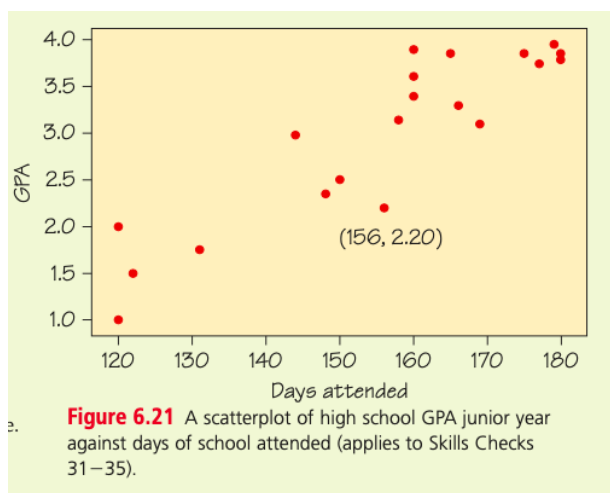


2. MAKING PREDICTIONS: REGRESSION LINES

Definition 2.1. A _____ is a straight line that describes how the response variable changes as the explanatory variable changes. It is a line that is as close as possible to all the points.

Example 2.2.

Draw the estimated regression line for the scatterplots below.



W.H. Freeman (2015), Comap - For All Practical Purposes
Mathematical Literacy in Today's World, Macmillan

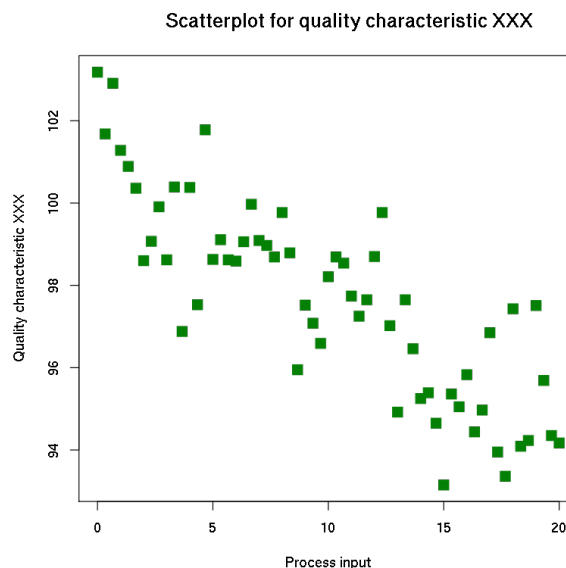


Figure: By Daniel Penfield - Own work,
CC BY-SA 3.0,
<https://commons.wikimedia.org/w/index.php?curid=9402369>

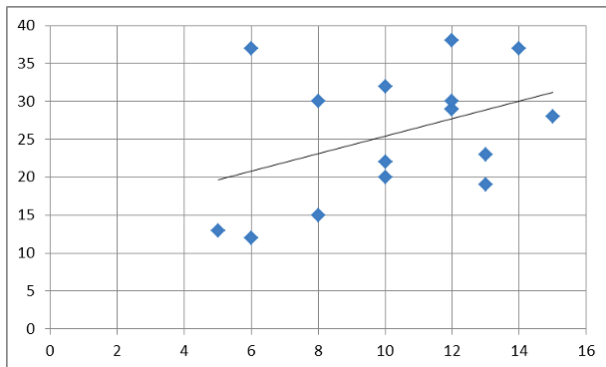
Definition 2.3. Two variables have _____ association if their changes tend to be in the same direction.
Two variables have _____ association if their changes tend to be in opposite direction.

Example 2.4.

Which of the above is Positive association? Negative Association?

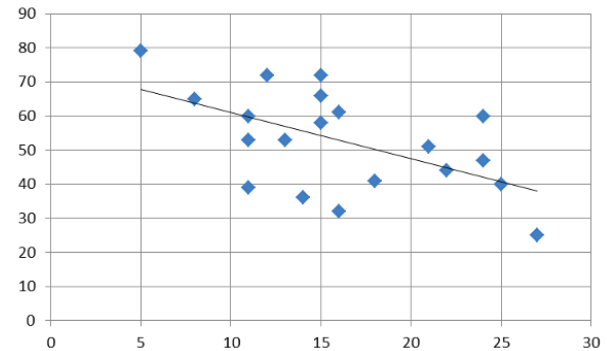
Example 2.5.

Use the given regression line to make predictions about the variable below.



Given above graph,

- If $x = 8$, what is y ?
- If $y = 30$, what is x ?
- If $x = 2$, what is y ?



Given above graph,

- If $x = 15$, what is y ?
- If $y = 50$, what is x ?
- If $x = 3$, what is y ?

Definition 2.6.

A _____ is using the regression line to find values between the minimum and maximum data values.

A _____ is using the regression line to find values that are outside the minimum and maximum values.

Example 2.7.

Which of the above required Interpolation? Extrapolation?

3. CORRELATION

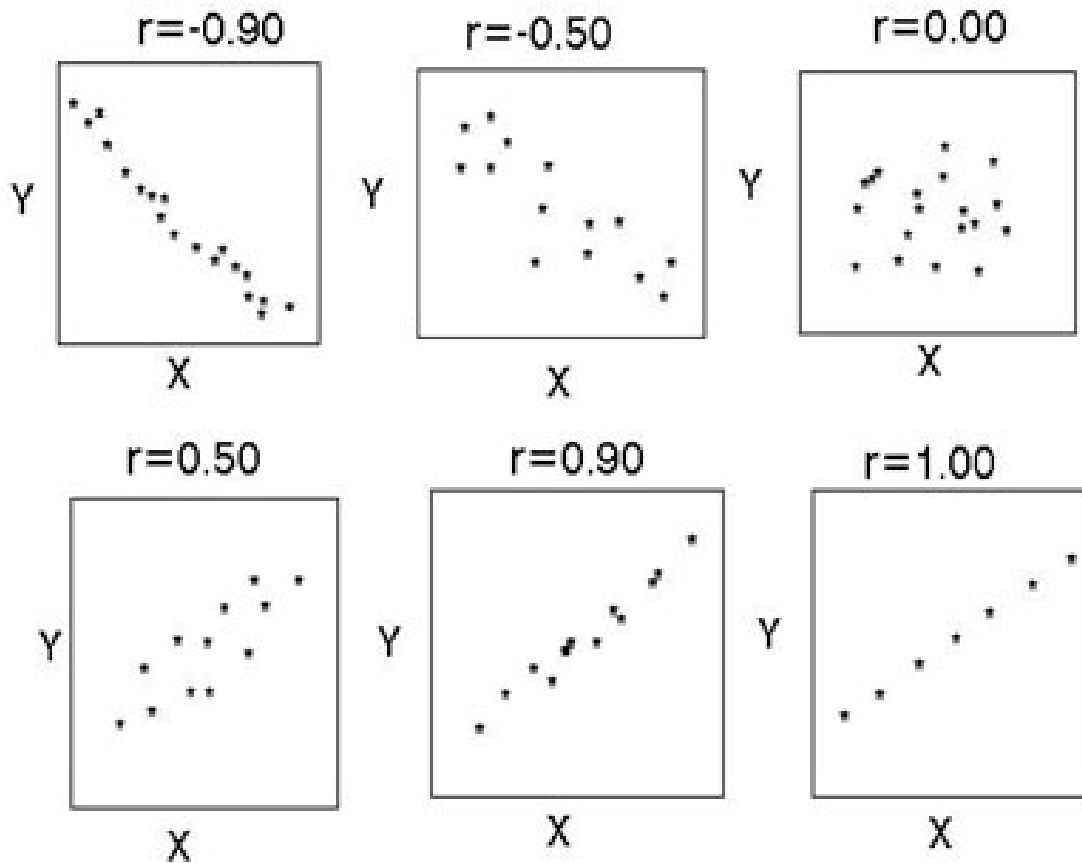
Definition 3.1.

A _____ r measures the direction and strength of the straight-line relationship between two quantitative variables.

The value of the correlation is a number r that is between -1 and 1, inclusive. That is

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

For positive association, $r > 0$. For negative association, $r < 0$. For no straight line association, $r = 0$. The closer $|r|$ is to 1, the stronger the straight line association.



<https://www.simplypsychology.org/correlation-coefficient.jpg>

Example 3.2.

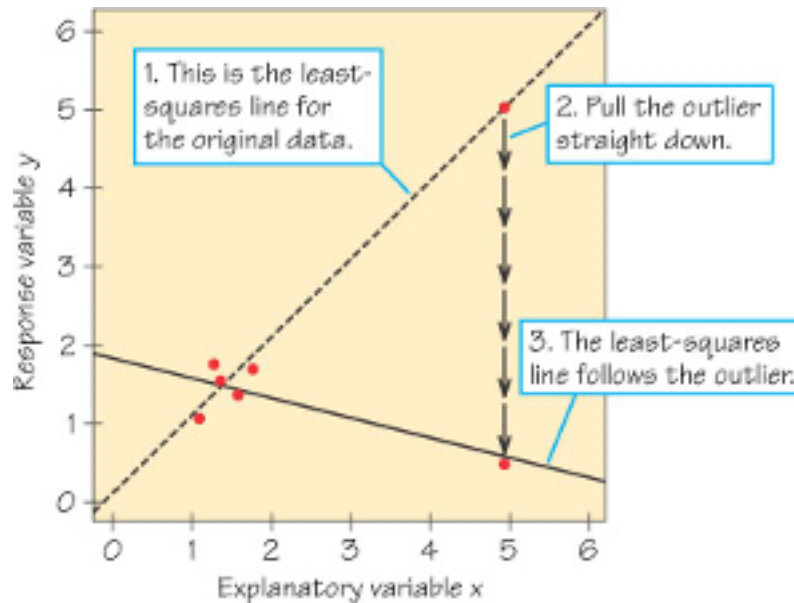
Which represents a stronger linear relationship? $r = 0.7$ or $r = -0.9$?

4. LEAST SQUARES REGRESSION

We will not test or do homework over this section, but I would be happy to answer any questions that you have.

5. INTERPRETING CORRELATION AND REGRESSION

Warning: Look for outliers before you rely on a regression line or a correlation coefficient!!!! Outliers will have a very strong effect on both the regression line and the correlation!



BE CAREFUL if you EXTRAPOLATE beyond your data!

Example 5.1.

A study of children ages 2-12 years old found a growth rate of 2.5 inches per year. If Jane was 48 inches tall at age 12, how tall will she be at age 30?

Example 5.2.

In 1960, the US population was 180,000,000. The annual growth rate at that time was 1.7 percent causing alarm about overpopulation. If that rate continued, the population in the US in 2010 should have been 353,273,143. However, it was 308,745,538 (off by more than 44 million people).

Correlation does not mean causation!

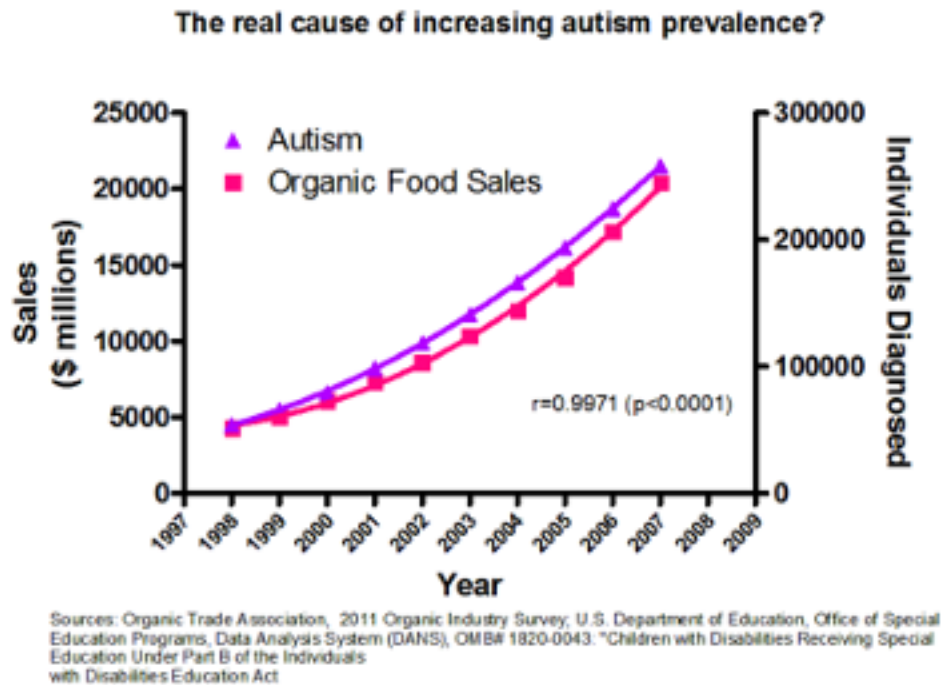
Example 5.3.

A study claimed that “chocolate consumption enhances cognitive function” because of a positive correlation between the country’s chocolate consumption and the number of people from that country who have won the Nobel prize.

<http://epianalysis.wordpress.com/2012/11/19/chocolate/>

Example 5.4.

A study showed a strong correlation between the increase in organic food sales and an increase in the number of people diagnosed with autism. Does this mean that eating organic food causes autism?



<http://boingboing.net/2013/01/01/correlation-between-autism-dia.html>