Unit 3 Glossary Terms

categorical variable

A variable that takes on values which can be divided into groups or categories. For example, color is a categorical variable which can take on the values, red, blue, green, etc.

Two-way table

A way of organizing data from two categorical variables in order to investigate the association between them.

	has a cell phone	does not have a cell phone
10–12 years old	25	35
13–15 years old	38	12
16–18 years old	52	8

Variable (statistics)

A characteristic of individuals in a population that can take on different values

Relative frequency table

A version of a two-way table in which the value in each cell is divided by the total number of responses in the entire table or by the total number of responses in a row or a column.

The table illustrates the first type for the relationship between the condition of a textbook and its price for 120 of the books at a college bookstore.

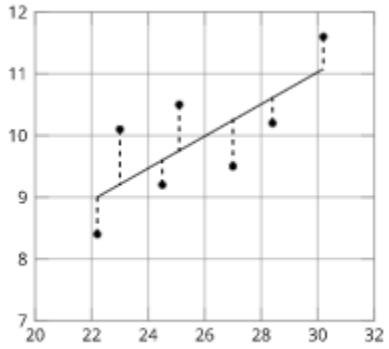
	$10 or less	more than $10 but less than $30	$30 or more
new	0.025	0.075	0.225
used	0.275	0.300	0.100

association

In statistics we say that there is an association between two variables if the two variables are statistically related to each other; if the value of one of the variables can be used to estimate the value of the other.

residual

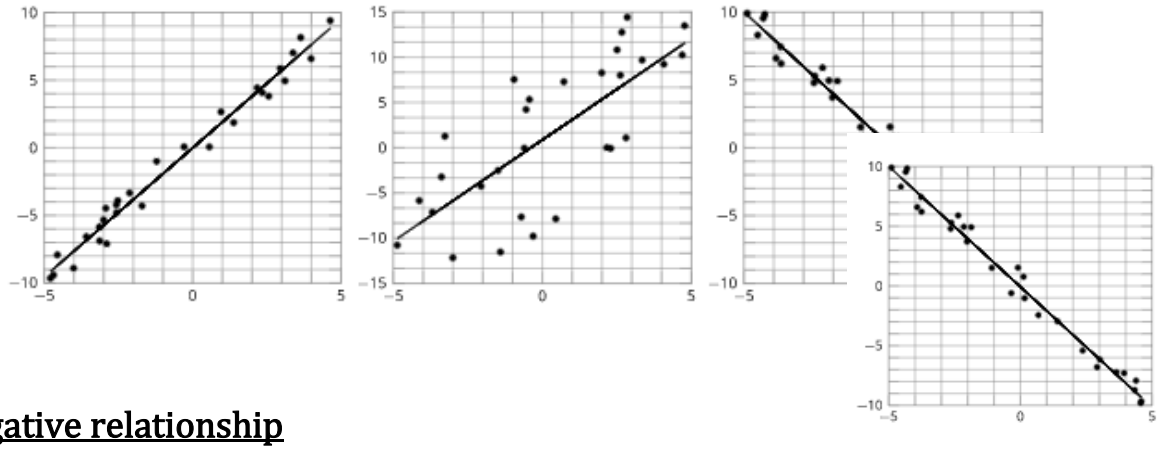
The difference between the -value for a point in a scatter plot and the value predicted by a linear model. The lengths of the dashed lines in the figure are the residuals for each data point.



Correlation coefficient

A number between -1 and 1 that describes the strength and direction of a linear association between two numerical variables. The sign of the correlation coefficient is the same as the sign of the slope of the best fit line. The closer the correlation coefficient is to 0, the weaker the linear relationship. When the correlation coefficient is closer to 1 or -1, the linear model fits the data better.

The first figure shows a correlation coefficient which is close to 1, the second a correlation coefficient which is positive but closer to 0, and the third a correlation coefficient which is close to -1.

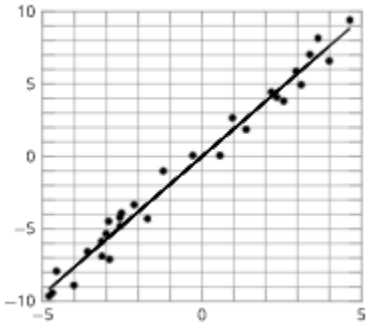


Negative relationship

A relationship between two numerical variables is negative if an increase in the data for one variable tends to be paired with a decrease in the data for the other variable.

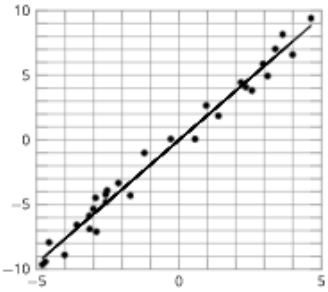
Positive relationship

A relationship between two numerical variables is positive if an increase in the data for one variable tends to be paired with an increase in the data for the other variable.



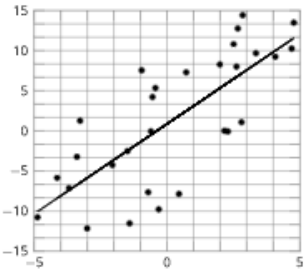
Strong relationship

A relationship between two numerical variables is strong if the data is tightly clustered around the best fit line.



Weak relationship

A relationship between two numerical variables is weak if the data is loosely spread around the best fit line.



Casual relationship

A relationship is one in which a change in one of the variables causes a change in the other variable.