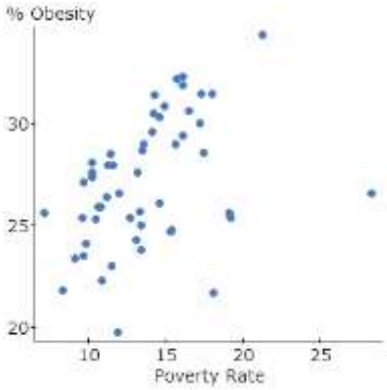
The worksheet data can be found at: http://www.statcrunch.com/app/index.php?dataid=1135370

1. Let's explore this data set using poverty rate and the percentage of the population that is obese. The idea is to explore how poverty associates with obesity. In the 18th century being a bit larger was actually desired as it was a sign of affluence. This altered as food became more readily and cheaply available until today where fast food is one of the easiest to access and cheapest food sources. The question we're trying to answer is whether or not poverty tends to occur with obesity or not.

   a. Using technology, look at a scatter plot with x = poverty rate and y = percentage of the population that is obese in a given state and make a statement about the graph.



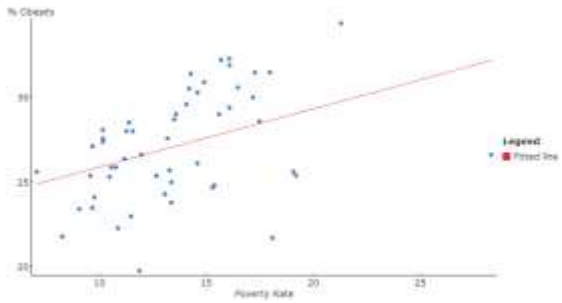
This scatter plot has a weak positive correlation. We would expect r to be between .3 and .5 because there is a correlation but it isn't close to fitting a straight line.

   b. Using technology, find the regression line that uses the poverty rate to predict the percentage of the population that is obese in a given state.

Simple linear regression results:
Dependent Variable: % Obesity
Independent Variable: Poverty Rate
% Obesity = 22.445065 + 0.34355338 Poverty Rate
Sample size: 52
R (correlation coefficient) = 0.40466797
R-sq = 0.16375617
Estimate of error standard deviation: 2.9363454

Parameter estimates:

Parameter	Estimate	Std. Err.	Alternative	DF	T-Stat	P-Value
Intercept	22.445065	1.5601484	≠ 0	50	14.386494	<0.0001
Slope	0.34355338	0.10979356	≠ 0	50	3.1290851	0.0029



   c. What is the explanatory variable?

   The poverty rate is the explanatory variable because we are using it to predict the percentage of the population that is obese.

   d. What is the response variable?

   The percentage of the population that is obese is the response variable because we are trying to predict it.

e. State the regression line.

(% Obesity) = 22.3351 + .3436(Poverty Rate)*

f. Report and interpret the slope.

The slope is .3436 and it indicates that for every unit increase in poverty rate there is a .3436 increase in the percentage of the population that is obese. In other words, for each 1 percent increase in poverty rate we can expect a .3436 increase in the percentage of the population that is obese on average – more impoverished states tend to be more obese.

g. Report and interpret the y-intercept.

The intercept is 22.3351 and it indicates that if a state had 0% poverty we would expect 22.3351% of the population to be obese.

h. Report and interpret the R-sq:

R-sq, r squared, is .1638 and indicates that 16.38 percent of the variation in the percentage of the population that is obese is explained by the poverty *rate.*

i. Predict the percentage of the population that is obese for a state with a poverty rate of 8.3%.

(% Obesity) = 22.3351 + .3436(Poverty Rate)*
(% Obesity) = 22.3351 + .3436(8.3)*
(% Obesity) = 22.3351 + 2.85188
(% Obesity) = 25.18698

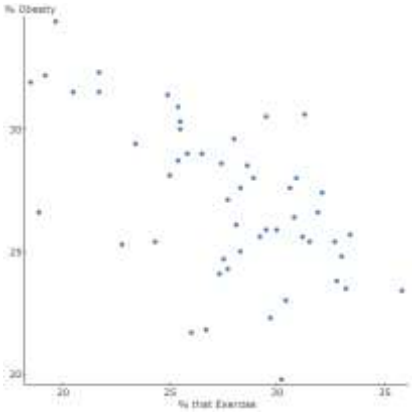
We expect that 25.18698% of the population is obese in the state where the poverty rate is 8.3.

j. If the actual percentage of the population with a poverty rate of 8.3% is 21.8% what is the residual?

Residual = True – Predicted
Residual = 21.8% - 25.18698%
Residual = 3.38698

2. Let's continue exploring this data set using percentage of the population that exercises and the percentage of the population that is obese. The idea is to explore how exercise associates with obesity. It should be obvious that if more of the population exercises that the population should be less obese, but we can look at the statistics.

a. Using technology, look at a scatter plot with x = percentage of the population that exercises and y = percentage of the population that is obese in a given state and make a statement about the graph.



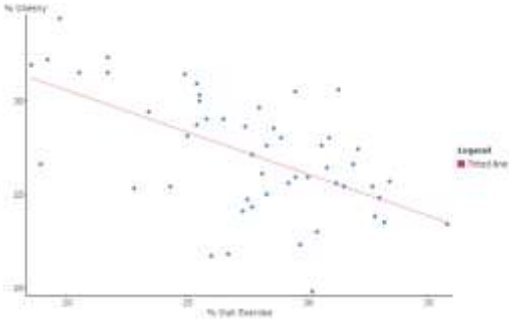
This scatter plot has a moderate negative correlation. We would expect r to be between -.5 and .8 because there is a correlation but it isn't close to fitting a straight line.

b. Using technology, find the regression line that uses the percentage of the population that exercises to predict the percentage of the population that is obese in a given state.

Simple linear regression results:
Dependent Variable: % Obesity
Independent Variable: % that Exercise
% Obesity = 39.493306 - 0.44703596 % that Exercise
Sample size: 52
R (correlation coefficient) = -0.57986723
R-sq = 0.33624601
Estimate of error standard deviation: 2.6160392

Parameter estimates:

Parameter	Estimate	Std. Err.	Alternative	DF	T-Stat	P-Value
Intercept	39.493306	2.4777441	≠ 0	50	15.939219	<0.0001
Slope	-0.44703596	0.088824438	≠ 0	50	-5.0328038	<0.0001



c. What is the explanatory variable?

The percentage of the population that exercises *is the explanatory variable because we are using it to predict the percentage of the population that is obese.*

d. What is the response variable?

The percentage of the population that is obese is the response variable because we are trying to predict it.

e. State the regression line.

(% Obesity) = 39.4933 - .4470(% Exercise)*

f. Report and interpret the slope.

The slope is -.4470 and it indicates that for every unit increase in the percentage of the population that exercises there is a .4470 decrease in the percentage of the population that is obese. In other words, for each 1 percent increase in the percentage of the population that exercises we can expect a .4470 decrease in the percentage of the population that is obese on average – more exercised states tend to be less obese.

g. Report and interpret the y-intercept.
The intercept is 39.4933 and it indicates that if a state had 0% of its population exercising we would expect 39.4933% of the population to be obese.

h. Report and interpret the R-sq:

R-sq, r squared, is .3362 and indicates that 33.62 percent of the variation in the percentage of the population that is obese is explained by the percentage of the population that exercises.
Note: *R-sq is higher here than it was for the poverty rate when R-sq=.1638. This indicates that the percentage of the population that exercises is a better explanatory variable than poverty rate for the percentage of the population that is obese.*

i. Predict the percentage of the population that is obese for a state where 28.6 percent of the population exercises.

(% Obesity) = 39.4933 - .4470(% Exercise)*
(% Obesity) = 39.4933 - .4470(28.6)*
(% Obesity) = 39.4933 - 12.7842
(% Obesity) = 26.7091

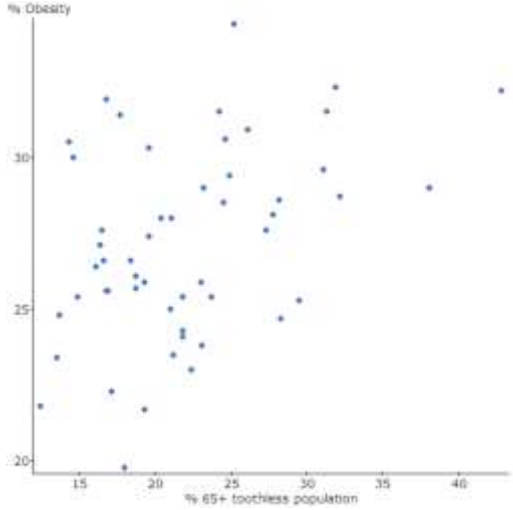
We expect that 26.7091% of the population is obese in the state where 28.6% of the population exercises.

j. If the actual percentage of the population where 28.6% of the population exercises is 28.5% what is the residual?

Residual = True – Predicted
Residual = 28.5% - 26.7091%
Residual = 1.7909

3. Let's continue exploring this data set using percentage of the population over 65 that are toothless and the percentage of the population that is obese. The idea is to explore how the percentage of the population over 65 that are toothless associates with obesity. Do we follow the stereotype of the obese, toothless person or do we say that people without teeth can't eat as much and might be less obese?

a. Using technology, look at a scatter plot with x = percentage of the population over 65 that are toothless and y = percentage of the population that is obese in a given state and make a statement about the graph.



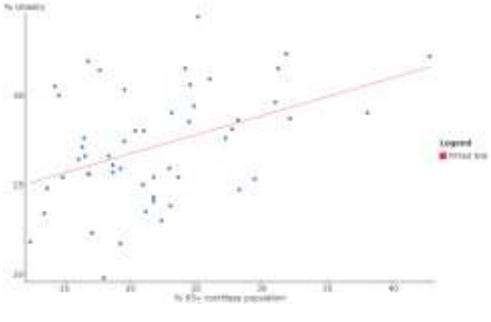
This scatter plot has a very weak positive correlation. We would expect r to be between .1 and .3 because there is a correlation but it isn't close to fitting a straight line.

b. Using technology, find the regression line that uses the percentage of the population over 65 that are toothless to predict the percentage of the population that is obese in a given state.

Simple linear regression results:
Dependent Variable: % Obesity
Independent Variable: % 65+ toothless population
% Obesity = 22.455098 + 0.21293531 % 65+ toothless population
Sample size: 52
R (correlation coefficient) = 0.42429194
R-sq = 0.18002365
Estimate of error standard deviation: 2.9076447

Parameter estimates:

Parameter	Estimate	Std. Err.	Alternative	DF	T-Stat	P-Value
Intercept	22.455098	1.4755097	≠ 0	50	15.218536	<0.0001
Slope	0.21293531	0.064268561	≠ 0	50	3.3132112	0.0017



c. What is the explanatory variable?

The percentage of the population over 65 that are toothless is the explanatory variable because we are using it to predict the percentage of the population that is obese.

d. What is the response variable?

The percentage of the population that is obese is the response variable because we are trying to predict it.

e. State the regression line.

(% Obesity) = 22.4551 + .2129(% toothless over 65)*

f. Report and interpret the slope.

The slope is .2129 and it indicates that for every unit increase in the percentage of the population over 65 that are toothless there is a .2129 increase in the percentage of the population that is obese. In other words, for each 1 percent increase in the percentage of the population over 65 that are toothless we can expect a .2129 increase in the percentage of the population that is obese on average – more toothless states tend to be more obese.

g. Report and interpret the y-intercept.
The intercept is 22.4551 and it indicates that if a state had 0% of its population over 65 being toothless we would expect 22.4551% of the population to be obese.

h. Report and interpret the R-sq:

R-sq, r squared, is .1800 and indicates that 18 percent of the variation in the percentage of the population that is obese is explained by the percentage of the population over 65 that are toothless.
Note: *R-sq is higher here than it was for the poverty rate when R-sq=.1638 and lower than it was for the percentage of the population the exercises, R-sq=.3362. This indicates that the percentage of the population over 65 being toothless is a better predictor than poverty level but a worse predictor than the percent of the population that exercises.*

i. Predict the percentage of the population that is obese for a state where 26.1 percent of the population over 65 is toothless.

(% Obesity) = 22.4551 + .2129(% toothless over 65)*
(% Obesity) = 22.4551 + .2129(26.1)*
(% Obesity) = 22.4551 + .5.55669
(% Obesity) = 28.01179

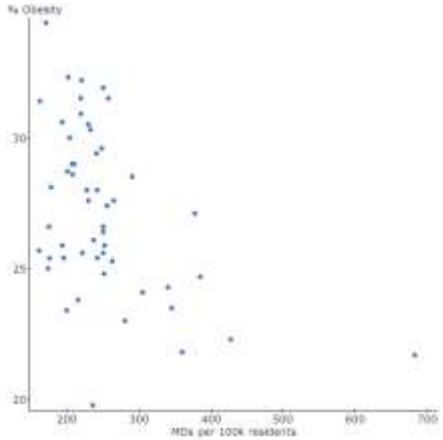
We expect that 28.01179% of the population is obese in the state where 26.1% of the population over 65 is toothless.

j. If the actual percentage of the population where 26.1% of the population over 65 is toothless is 30.9% what is the residual?

Residual = True – Predicted
Residual = 30.9% - 28.01179%
Residual = 2.88821

4. Let's continue exploring this data set using number of doctors per 100k residents and the percentage of the population that is obese. The idea is to explore how the number of doctors per 100k residents associates with obesity. We can think that more doctors would mean a healthier population, but if we think about it like a business we can think that a less healthy population would require more doctors.

a. Using technology, look at a scatter plot with x = number of doctors per 100k residents and y = percentage of the population that is obese in a given state and make a statement about the graph.

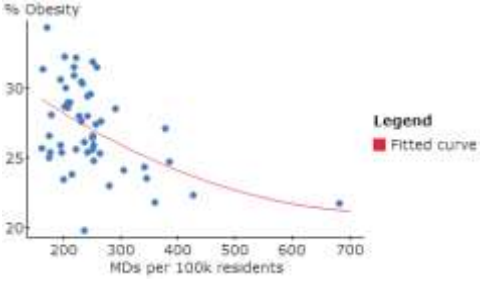


This scatter plot has a very moderate negative correlation. We would expect r to be between -.3 and -.5 because there is a correlation but it isn't close to fitting a straight line. We can, in this case see a slight curve in the data which appears to be better fit by a negative exponential than an ordinary line.

b. Using technology, find the regression line that uses number of doctors per 100k residents to predict the percentage of the population that is obese in a given state.

Polynomial Regression Results:
Dependent Variable: % Obesity
Independent Variable: MDs per 100k residents

Parameter	Estimate	Std. Err.	Alternative	DF	T-Stat	P-Value
Intercept	34.073114	3.2087448	≠ 0	49	10.618829	<0.0001
X	-0.03369492	0.0194088886	≠ 0	49	-1.7360563	0.0888
X^2	0.000021837195	0.000025223896	≠ 0	49	0.86573441	0.3909



c. What is the explanatory variable?

The percentage of the population over 65 that are toothless is the explanatory variable because we are using it to predict the percentage of the population that is obese.

d. What is the response variable?

The percentage of the population that is obese is the response variable because we are trying to predict it.

e. State the regression line.

(% Obesity) = 34.0731 - .0337(# Doctors per 100k) + .0000218*(* # Doctors per 100k)^2

f. Report and interpret the slope.

We won't cover this for the polynomial case, so no worries here.

g. Report and interpret the y-intercept.
The intercept is 34.0731 and it indicates that if a state had 0 doctors per 100k residents we would expect 34.0731% of the population to be obese.

h. Report and interpret the R-sq:

We won't cover this for the polynomial case, so no worries here.

i. Predict the percentage of the population that is obese for a state where there are 242 doctors per 100k residents.

(% Obesity) = 34.0731 - .0337(# Doctors per 100k) + .0000218*(* # Doctors per 100k)^2
(% Obesity) = 34.0731 - .0337(242) + .0000218*(* 242)^2
(% Obesity) = 34.0731 – 8.1554 + 1.2767
(% Obesity) = 27.1944

We expect that 27.1944% of the population is obese in the state that has 242 doctors per 100k residents.

j. If the actual percentage of the population where there are 242 doctors per 100k residents is 25.4% what is the residual?

Residual = True – Predicted
Residual = 25.4% - 27.1944%
Residual = -1.7944

Note there are plenty of other data points and equally exciting questions to answer using the data for this worksheet which can be found at: http://www.statcrunch.com/app/index.php?dataid=1135370

Try to use another variable and follow a similar line of questioning to come up with answers to your own questions!