Correlation between two variables

When comparing two variables (generally through a scatter plot) we see trends. We can describe it through direction and strength. If two variables show correlation it means there is a connection between them.

Direction	
Positive Correlation	Negative Correlation

Strength		
Strong Correlation	Weak Correlation	No Correlation

In order to determine the direction and strength of the model, we use something called a **correlation coefficient**. It is represented by the letter ***r***.

Direction and correlation coefficients

- r is always between -1 and 1
- A positive r is a positive correlation, a negative r is a negative correlation

Strength and correlation coefficients

- A r of 1 is a perfect positive correlation (an r of -1 is a perfect negative correlation)
- The closer the value to 1 (or -1) the stronger the correlation
- Generally 0.8 – 1 represent strong correlations
- 0.7-0.79 represent weak correlation
- Anything below 0.7 generally has no correlation

r = 0.93	r = -0.95	r = -0.75	r = 0. 61	r = 1

Finding correlation coefficient using the graphing calculator:

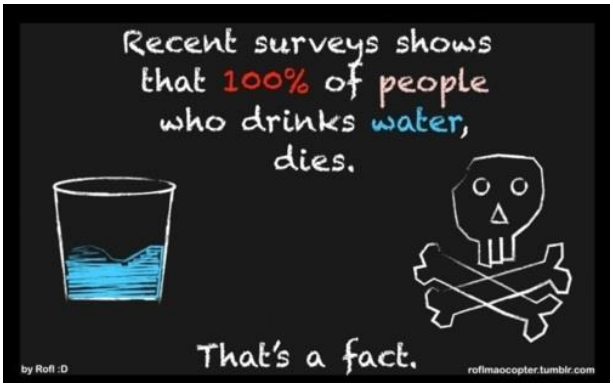
1) Turn on diagnostics (only need to do this once- if r doesn't show up) (2^{nd}) [catalog] (above the 0 key). Scroll down to DiagnosticsOn. Press enter to select it and then enter again to turn it on (it will say Done)
2) Go to [stat] and then [edit] Enter x values into L1 and y values into L2
3) To do anything with the data go to [stat] scroll over to [calc] and select 4:LinReg
4) You will be given the equation and the r (correlation coefficient)

Find the correlation coefficient of the following and determine the strength and direction.

Tables	0	2	6	10	12	13	15
Amount in tips	0	13	42	65	78	92	102

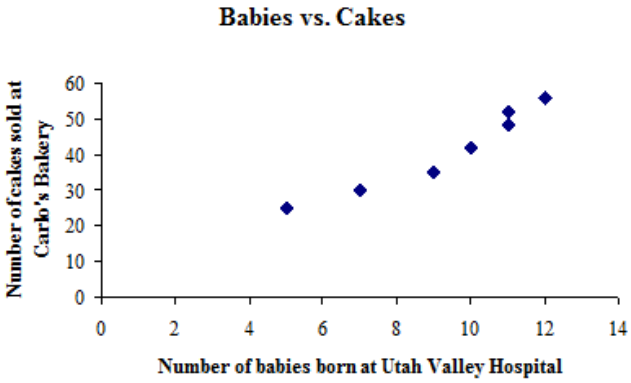
Correlation and Causation

Just because two variables have a strong correlation does not mean that they are causing one another.



Example: During the month of June the number of new babies born at the Utah Valley Hospital were recorded for a week. Over the same time period, the number of cakes sold at Carlo's Bakery in Hoboken, New Jersey was also recorded. What can be said about the correlation? Is there causation? Of course not, the relationship does not make any sense. We have to be careful about what we infer from a statistical analysis.

Number of babies born	Number of cakes sold
5	25
7	30
9	35
10	42
11	48
11	52
12	56



Example: An American medical researcher wants to see if there is a link between a person's socio-economic status (how much money they have) and certain types of cancer. His research seems to indicate that there is a link (rich people seem to suffer from more cancers than poor people do). His Causation Statement: *Being rich will make you more likely to get cancer.* What can be said about the correlation? Is there causation, why or why not?

Not likely. Scientists who use scatter plots to look for correlations between variables watch out for this hidden variable problem. The connection the researcher described may actually exist, but there are certainly other possibilities. People with more money are far more likely to see a doctor than poorer people. The higher rates of cancer in rich people may not be due to their wealth at all, but due instead to the fact that they visit their doctor more often (they can afford it), so more cancers are being diagnosed. Very different lifestyles are most likely the reason.

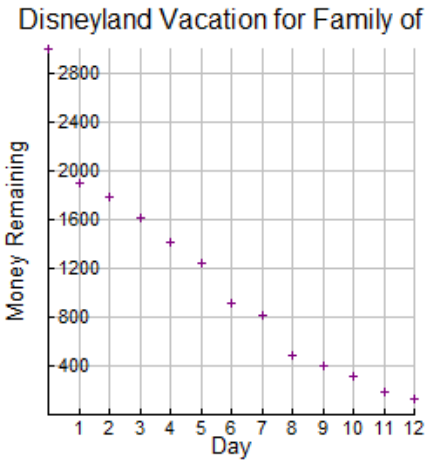
Example: In this present economy, families are trying to find ways to save money Families might be thinking about not eating out to spend less money. Causation Statement: *The more you eat out, the more money you spend at restaurants.*

What can be said about the correlation? Is there causation? Why or why not?

For the given situations below,

a. Is the association positive or negative? (Hint: It may help to sketch a graph of the situation).

b. Is the situation a correlation or a causation?



a.

b.

2. When you are on a diet, the less calories you eat daily vs. the more weight you lose. Causation statement: Therefore, *eating fewer calories makes you lose weight*.

a.

b.

3. The more ice cream consumed on a beach vs. the increased number of people who go in the water. Causation statement: Therefore, *eating more ice cream on the beach makes people go in the water.*

a.

b.

4. The more pets you own vs. the increased amount you spend on pet food. Causation statement: Therefore, *owning more pets makes you spend more money on pet food.*

a.

b.

5. How much you pay for a house vs. how much you pay for a car. Causation statement: Therefore *the more you pay for a house makes you spend more for a car*.

a.

b.

6. When you are trying to improve your grades, the more you study vs. the increase in your GPA. Causation statement: Therefore, *the more you study makes your GPA increase.*

a.

b.

7. The increased of policeman that are visible on a stretch of road vs. the slower speed you travel. Causation Statement: Therefore, *the more policeman that are visible makes you travel slower*

a.

b.

8. A person's height vs. the amount of money that person has. Causation Statement: Therefore, *a person's height determines how much money a person has.*

a.

b.

9. The total amount of money spent on concessions at Phillies games vs. the number of wins the Phillies have. Causation Statement: Therefore, *money spent on concessions determines number of wins the Phillies have.*

a.

b.

10. The increased amount of alcohol people drink vs. the increased incidences of liver disease. Causation Statement: Therefore, *the more alcohol people drink makes the incidences of liver disease increase.*

a.

b.

11. The more people in a family vs. the increased number of cars the family owns. Causation Statement: Therefore, *the more people there are in a family determines how many cars a family owns.*

a.

b.

12. The more problems on a math test vs. the more time it takes students to complete the test. Causation Statement: Therefore, *the more problems there are on a test determines how long students will take to complete the test.*

a.

b.

13. The average speed cars travel from Philadelphia to New York on the turnpike vs. the average amount of times it takes. Causation Statement: Therefore, *the speed cars travel from Philadelphia to New York determines the time it takes to go between them.*

a.

b.

14. The more firemen fighting a fire vs. the bigger the fire is observed to be. Causation statement: Therefore *firemen cause fires.*

a.

b.

Finding a linear regression model

Look back at the example on the first page about money earned in tips and tables that were served by a waitress

Tables	0	2	6	10	12	13	15
Amount in tips	0	13	42	65	78	92	102

What was the correlation coefficient?
What is the shape and strength of this relationship?
Is it causal or just a correlation?

Using your equation, predict how much the waitress will get in tips if she serves 20 tables.

The waitress counted her tips at the end of the night and found out that she had $61. How many tables do you predict she served?

To the right is a table that shows the chirps per seconds of crickets and the temperature outside.

What is the correlation coefficient? Describe the strength and direction.

Is it appropriate to find a linear regression model to predict? Why or why not?

Find the linear regression model

If the crickets chirped 21.1 chirps per second, what could you predict about the temperature outside?

If it is 62° outside how many chirps per second would you predict the crickets would make?

Chirps/Second	Temperature (° F)
20.0	**88.6**
16.0	**71.6**
19.8	**93.3**
18.4	**84.3**
17.1	**80.6**
15.5	**75.2**
14.7	**69.7**
15.7	**71.6**
15.4	**69.4**
16.3	**83.3**
15.0	**79.6**
17.2	**82.6**
16.0	**80.6**
17.0	**83.5**
14.4	**76.3**