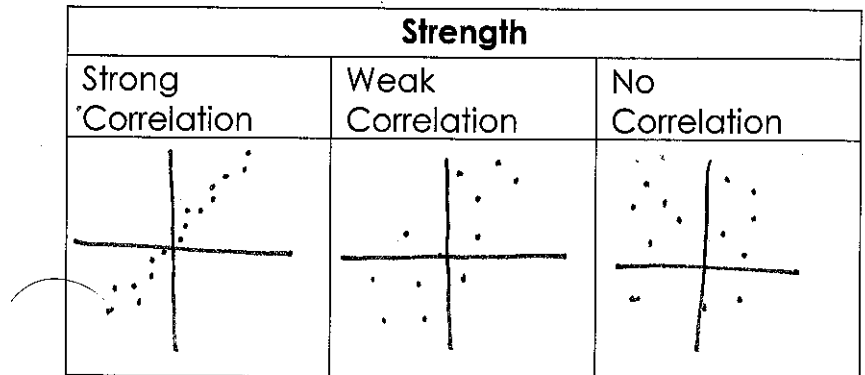
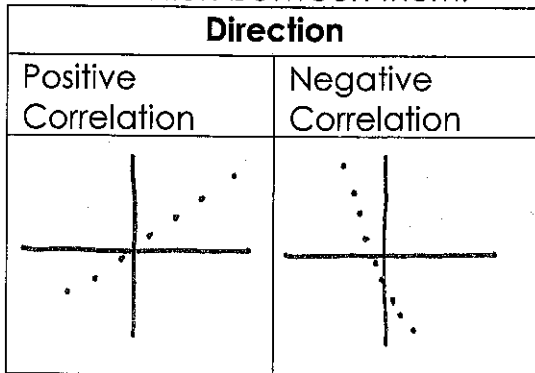


## Correlation between two variables

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.



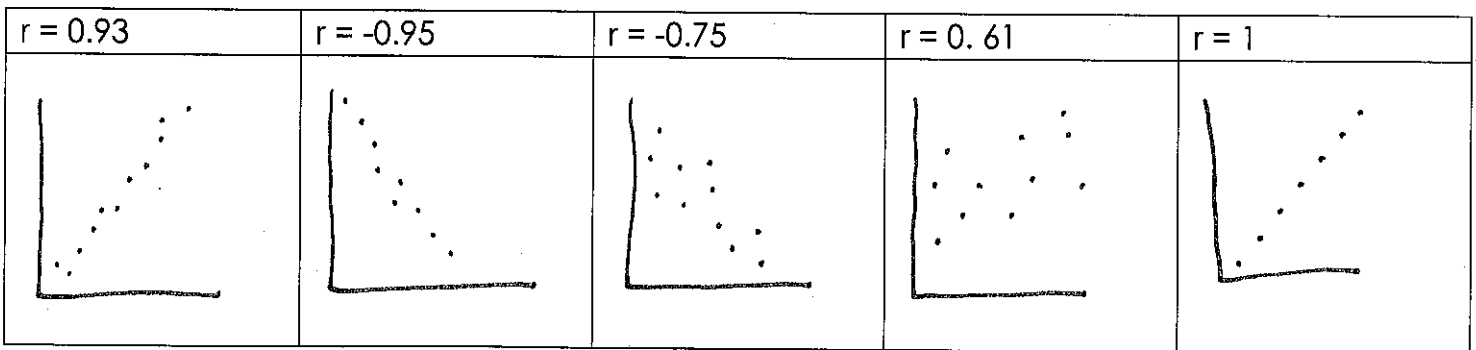
In order to determine the direction and strength of the model, we use something called a **correlation coefficient**. It is represented by the letter  $r$ .

### 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



Finding correlation coefficient using the calculator

- 1) Go to [data] Enter  $x$  values into L1 and  $y$  values into L2
- 2) [2<sup>nd</sup>][data] choose 2-Var Stats (since we have two variables now)
- 3) Scroll down until you find  $r$  ☺

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

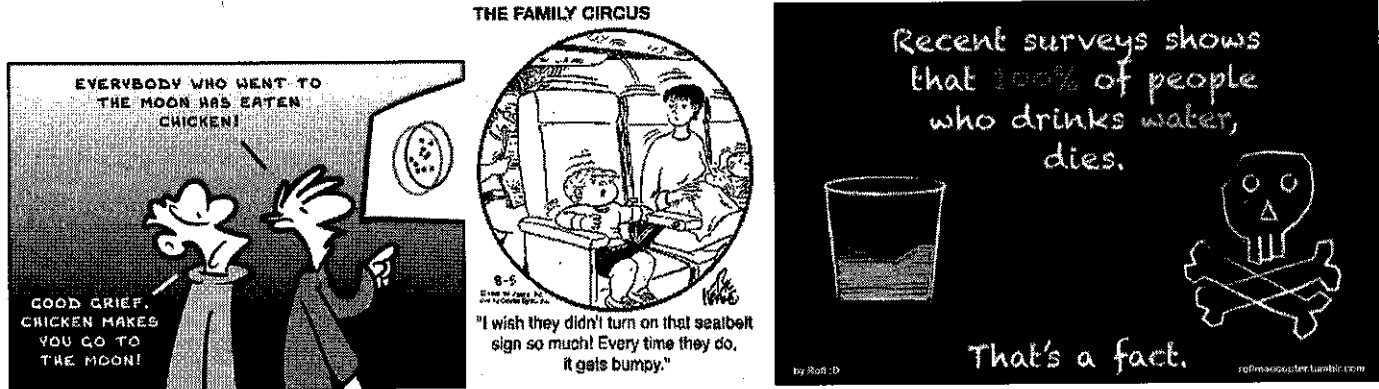
|                |   |    |    |    |    |    |     |
|----------------|---|----|----|----|----|----|-----|
| Tables         | 0 | 2  | 6  | 10 | 12 | 13 | 15  |
| Amount in tips | 0 | 13 | 42 | 65 | 78 | 92 | 102 |

~~0.998~~  
 $r = .998$

STRONG POSITIVE

## Correlation and Causation

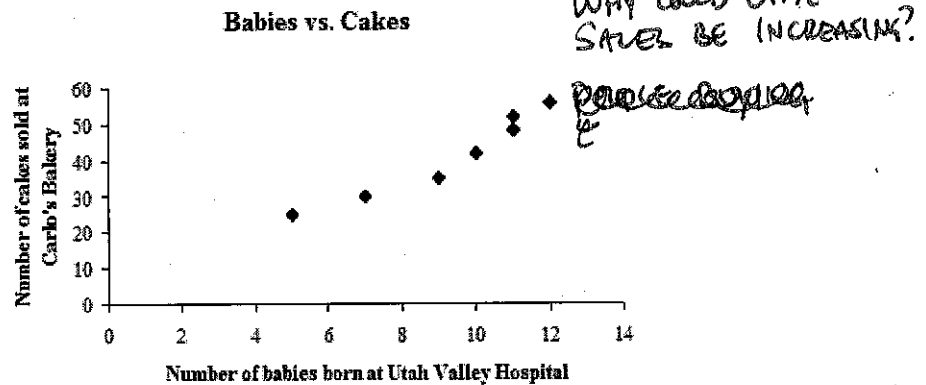
Just because two variables have a strong correlation does not mean that they are causing one another.



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?

CAUSAL YOU HAVE TO SPEND \$ AT RESTAURANTS TO BUY THE FOOD

### 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? ~~0.998~~  $r = .998$

What is the shape and strength of this relationship? **STRONG POSITIVE**

Is it causal or just a correlation? **CAUSAL**

Since we determine that it is a strong relationship (based on the correlation coefficient) we can determine a function (equation) that would model this relationship. This equation can help us predict values that we don't have information for. Remember that lines are always in the form  $y = mx + b$ . Your calculator can determine the  $m$  and the  $b$ . Follow the same steps that you did for finding the  $r$  and look for the variables **a** (which the calculator uses instead of  $m$ ) **and b**. This will be your equation.

~~$y = 3.4x + 22.8$~~   $y = 6.8x - .32$

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

**\$135.68**

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? **ABOUT 9**

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

| 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              |

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

$r = .83$

**STRONG POSITIVE**

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

**YES BECAUSE THERE IS A STRONG CORRELATION**

Find the linear regression model

$y = 3.4x + 22.8$

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

**94.5°**

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

**11.5 CHIRPS**