**Correlation and Causation**

As you have seen in the previous activities, the linear correlation between two variables can be measured by the correlation coefficient, *r*. The correlation coefficient may indicate that the linear relationship is strong or weak, and it may indicate that the trend has a positive or negative slope.

The graphs below, introduced in Activity 5.3.1, show values of *r* for different scatterplots.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Perfect Positive Correlation |  | High  Positive Correlation |  | Low  Positive Correlation |  | No Correlation |  | Low Negative Correlation |  | High  Negative Correlation |  | Perfect Negative  Correlation |
|  | | | | | | | | | | | | | |

In addition to knowing whether two variables are correlated, statisticians often want to know if there is **causation** in the relationship. Causation occurs if changes in one variable *cause* a change in the other variable? It is difficult to prove causation. To prove causation we must perform a carefully designed experiment that measures the effect of changing a variable on another variable. All other variables must be fixed.

1. An ice cream shop keeps track of how much

ice cream they sell versus the temperature

for 12 days.

1. Imagine that a trend line is drawn. Is the

slope positive or negative?

1. Look at the graphs at the top of the page and

describe the strength of the correlation

(strong, weak, or none).

1. Do you think the hotter temperatures caused an increase in ice cream sales? Explain.
2. The Ice Cream shop tracks how many

sunglasses were sold by a big store on

several days and they compare the

number of sunglasses sold each day to

their ice cream sales on that day.

1. Describe the direction of the

correlation.

1. Describe the strength of the correlation.
2. Do you think that the sales of sunglasses could have caused an increase in ice cream sales? Explain.
3. The table shows the 2010 monthly sales (in billions) at Target Stores and the length of sunlight in hours and minutes on the last day of each month in Hartford, CT.

|  |  |  |  |
| --- | --- | --- | --- |
| Target Sales vs. Daylight Hours | | | |
| Month | Sales  (in billions) | Daylight (hours: min) | Daylight Hours |
| July | 4.6 | 14:26 |  |
| August | 5.0 | 13:10 |  |
| September | 5.6 | 11:48 |  |
| October | 4.6 | 10:25 |  |
| November | 6.0 | 9:24 |  |
| December | 9.9 | 9:12 |  |

1. The amount of daylight is given in hours and minutes. Fill in the last column with the daylight hours, rounded to the nearest tenth.
2. Enter the data from the “Sales” column into L1 and the data from the “Daylight Hours” column into L2. Find the regression line and the value of *r*.
3. Describe the strength and direction of the correlation.
4. Why might these variables be correlated?
5. Do you think that the increase in sales at Target could have caused shorter days? Explain.
6. Do you think that shorter days could have caused an increase in sales at Target? Explain.

4. The table shows the length of sunlight in hours and minutes on the last day of each month in Hartford, CT and the average daily temperatures (2010).

|  |  |  |  |
| --- | --- | --- | --- |
| Daylight Hours vs. Average Temperature | | | |
| Month | Daylight (hours: min) | Daylight Hours | Average Temperature |
| July | 14:26 |  | 79 |
| August | 13:10 |  | 75 |
| September | 11:48 |  | 69 |
| October | 10:25 |  | 55 |
| November | 9:24 |  | 44 |
| December | 9:12 |  | 30 |

1. Fill in the column for daylight hours from the table in problem 3.
2. The data from the “Daylight Hours” column should be in L2. Enter the data from the “Average Temperature” column in L3. Using lists L2 and L3, find the regression equation and the value of *r.*
3. Describe the strength and direction of the correlation.
4. Do you think that the decrease in daylight could have caused colder days? Explain.