Lesson 12: Relationships Between Two Numerical Variables

Classwork

A scatter plot is an informative way to display numerical data with two variables. Recall that if the two numerical variables are denoted by $x$ and $y$, the scatter plot of the data is a plot of the $(x, y)$ data pairs.

Example 1: Looking for Patterns in a Scatter Plot

The National Climate Data Center collects data on weather conditions at various locations. They classify each day as clear, partly cloudy, or cloudy. Using data taken over a number of years, they provide data on the following variables:

 $x$ = elevation above sea level (in feet)

 $y $= mean number of clear days per year

The table below shows data for 14 U.S. cities. Here is a scatter plot of the data on elevation and

 mean number of clear days.

|  |  |  |
| --- | --- | --- |
| City | $x$ = Elevation Above Sea Level (ft.) | $y$ = Mean Number of Clear Days per Year |
|  Albany, NY | 275 | 69 |
| Albuquerque, NM | 5,311 | 167 |
| Anchorage, AK | 114 | 40 |
| Boise, ID | 2,838 | 120 |
| Boston, MA | 15 | 98 |
| Helena, MT | 3,828 | 82 |
| Lander, WY | 5,557 | 114 |
| Milwaukee, WI | 672 | 90 |
| New Orleans, LA | 4 | 101 |
| Raleigh, NC | 434 | 111 |
| Rapid City, SD | 3,162 | 111 |
| Salt Lake City, UT | 4,221 | 125 |
| Spokane, WA | 2,356 | 86 |
| Tampa, FL | 19 | 101 |

Data Source: <http://www.ncdc.noaa.gov/oa/climate/online/ccd/cldy.html>

Exercises 1–3

1. Do you see a pattern in the scatter plot, or does it look like the data points are scattered?
2. How would you describe the relationship between elevation and mean number of clear days for these 14 cities? That is, does the mean number of clear days tend to increase as elevation increases, or does the mean number of clear days tend to decrease as elevation increases?
3. Do you think that a straight line would be a good way to describe the relationship between the mean number of clear days and elevation? Why do you think this?

 4. Use the data below to create a scatter plot on the elevation and mean number of partly cloudy days.

|  |  |  |
| --- | --- | --- |
| City | $x$ = Elevation Above Sea Level (ft.) | $w$ = Mean Number of Partly Cloudy Days per Year |
|  Albany, NY | 275 | 111 |
| Albuquerque, NM | 5,311 | 111 |
| Anchorage, AK | 114 | 60 |
| Boise, ID | 2,838 | 90 |
| Boston, MA | 15 | 103 |
| Helena, MT | 3,828 | 104 |
| Lander, WY | 5,557 | 122 |
| Milwaukee, WI | 672 | 100 |
| New Orleans, LA | 4 | 118 |
| Raleigh, NC | 434 | 106 |
| Rapid City, SD | 3,162 | 115 |
| Salt Lake City, UT | 4,221 | 101 |
| Spokane, WA | 2,356 | 88 |
| Tampa, FL | 19 | 143 |

Exercises 5–8: Thinking about Linear Relationships

Below are three scatter plots. Each one represents a data set with eight observations.

The scales on the$ x$ and $y$ axes have been left off these plots on purpose so you will have to think carefully about the relationships.



5. If one of these scatter plots represents the relationship between height and weight for eight adults, which scatter plot do you think it is and why?

1. Ifone of these scatter plots represents the relationship between height and SAT math score for eight high school seniors, which scatter plot do you think it is and why?

1. If one of these scatter plots represents the relationship between the weight of a car and fuel efficiency for eight cars, which scatter plot do you think it is and why?

1. Which of these three scatter plots does not appear to represent a linear relationship? Explain the reasoning behind your choice.

Exercises 9–14: Not Every Relationship is Linear

When a straight line provides a reasonable summary of the relationship between two numerical variables, we say that the two variables are *linearly related* or that there is a *linear relationship* between the two variables.

Take a look at the scatter plots below and answer the questions that follow.

**Scatter Plot 1: Scatter Plot 2:**

****

 Data Source: *Journal of Food Processing and Preservation,* 1995

1. Is there a relationship between number of cell phone calls and age or does it look like the data points are scattered?
2. If there is a relationship between number of cell phone calls and age, does the relationship appear to be linear?
3. Is there a relationship between moisture content and frying time, or do the data points look scattered?
4. If there is a relationship between moisture content and frying time, does the relationship look linear?

**Scatter Plot 3:**

 13. Scatter plot 3 shows data for the prices of bike helmets and

 the quality ratings of the helmets (based on a scale that estimates helmet

 quality). Is there a relationship between quality rating and price, or are

 the data points scattered?

 14. If there is a relationship between quality rating and price for bike

 helmets, does the relationship appear to be linear?

Data Source: www.consumerreports.org/health

Lesson Summary

* A scatter plot can be used to investigate whether or not there is a relationship between two numerical variables.
* A relationship between two numerical variables can be described as a linear or nonlinear relationship.

Problem Set

1. Construct a scatter plot that displays the data for $x$= elevation above sea level (in feet) and $z$ = Mean Number of Cloudy Days per Year

|  |  |  |
| --- | --- | --- |
| City | $x$ **= Elevation Above Sea Level (ft.)** | $z$ = Mean Number of Cloudy Days per Year |
| Albany, NY | 275 | 185 |
| Albuquerque, NM | 5,311 | 87 |
| Anchorage, AK | 114 | 265 |
| Boise, ID | 2,838 | 155 |
| Boston, MA | 15 | 164 |
| Helena, MT | 3,828 | 179 |
| Lander, WY | 5,557 | 129 |
| Milwaukee, WI | 672 | 175 |
| New Orleans, LA | 4 | 146 |
| Raleigh, NC | 434 | 149 |
| Rapid City, SD | 3,162 | 139 |
| Salt Lake City, UT | 4,221 | 139 |
| Spokane, WA | 2,356 | 191 |
| Tampa, FL | 19 | 121 |

1. Based on the scatter plot you constructed in Question 1, is there a relationship between elevation and the mean number of cloudy days per year? If so, how would you describe the relationship? Explain your reasoning.

Consider the following scatter plot for Questions 3 and 4:

 **Scatter Plot 4:** 3. Is there a relationship between finish time and

 age, or are the data points scattered?

 4. Do you think there is a relationship between

 finish time and age? If so, does it look linear?

 Data Source: Sample of 6 women who ran the 2003 NYC marathon

Consider the following scatter plot for Questions 5 and 6:

**Scatter Plot**

 5. A mare is a female horse and a foal is a baby

 horse. Is there a relationship between a foal’s birth

 weight and a mare weight, or are the data points

 scattered?

 Data Source: *Animal Behavior*, 1999

 6. If there is a relationship between baby birth

 weight and mother’s age, does the relationship look

 linear?