

Stat 218 - Day 31 Scatterplots and Association

We will study relationships between two *quantitative* variables for the rest of the course.

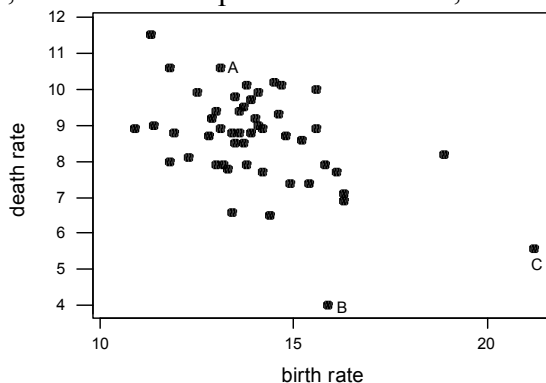
- A **scatterplot** is a graphical display of the relationship between two quantitative variables.
 - The explanatory variable goes on the horizontal (x -) axis, the response on the vertical (y -) axis.
- We examine a scatterplot for evidence of **association** between the variables.

Three aspects of association to look for are:

- Form
 - The form of the association is **linear** if a straight line appears to summarize the relationship between the variables.
- Direction
 - **Positive** association means that larger values of one variable tend to appear with larger values of the other, and smaller values of one variable tend to appear with smaller values of the other.
 - **Negative** association means that larger values of one variable tend to appear with smaller values of the other, and smaller values of one variable tend to appear with larger values of the other.
- Strength
 - Strength refers to the degree to which the data points follow a recognizable form.

Example: Birth rates and death rates

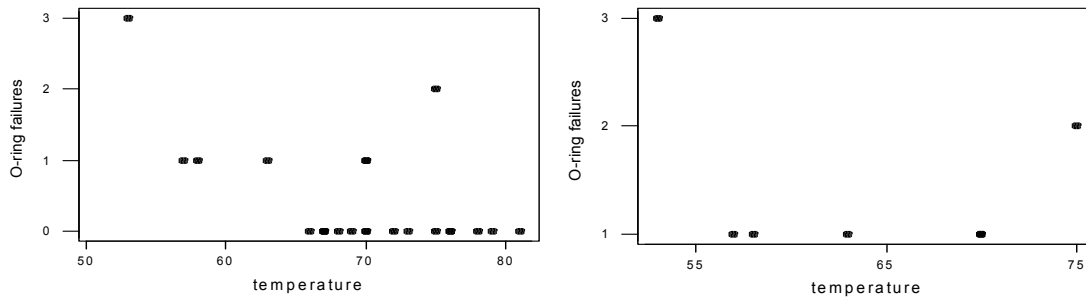
The following scatterplot displays the relationship between the death rate and the birth rate of the 50 states, both measured per 1000 residents, as of 1997:



- Identify the observational units in this study.
- Identify the variables in this study. Also classify each as categorical or quantitative.
- Describe the overall pattern in this scatterplot. Specifically, do states with higher birth rates tend to have lower, higher, or the same death rates as states with lower birth rates?

Example: Space shuttle Challenger disaster

The following scatterplots display the number of O-ring seals showing evidence of thermal distress vs. the air temperature at launch for the 23 space shuttle missions preceding the fatal launch of Challenger in January 1986:



(a) Explain the difference in how the two graphs were constructed.

(b) Would you say that the graph on the left reveals an association between number of O-ring failures and launch temperature? If so, describe its form, direction, and strength.

(c) Repeat (b) for the graph on the right.

(d) Which graph is more informative? Explain.

Example: Televisions and life expectancy

The Minitab worksheet `TVlife.mtw` contains data on the life expectancy and the number of people per television set (as reported in the *World Almanac*) for a sample of 22 countries.

(a) Which country has the fewest people per television? Which has the most? What are those numbers?

(b) Create a scatterplot of life expectancy vs. people per television (`Graph > Scatterplot`, or `MTB > plot c3*c4`). [Note: When you graph y vs. x , the convention is to put y on the vertical axis.] Does the scatterplot reveal an association between a country's life expectancy and its number of people per television set? If so, describe its form (is it linear?), direction, and strength.

(c) Would you conclude that sending televisions to countries with low life expectancy would cause their inhabitants to live longer? If not, suggest a confounding variable that provides an alternative explanation for the association.

Example: Comparing Popular Diets

A recent article in the *Journal of the American Medical Association* (Dansinger, Griffith, Gleason et al., 2005) reported on a randomized, comparative experiment in which 160 subjects were randomly assigned to one of four popular diet plans: Atkins, Ornish, Weight Watchers, and Zone (40 subjects per diet). These subjects were recruited through newspaper and television advertisements in the greater Boston area; all were overweight or obese with body mass index values between 27 and 42. Among the variables measured were:

- which diet the subject was assigned to
- whether or not the subject completed the twelve-month study
- the subject's weight loss after two months, six months, and twelve months (in kilograms, with a negative value indicating weight gain)
- the degree to which the subject adhered to the assigned diet, taken as the average of 12 monthly ratings, each on a 1-10 scale (with 1 indicating complete non-adherence and 10 indicating full adherence)

Data for the 93 subjects who completed the 12-month study are in the Minitab worksheet `ComparingDiets.mtw`.

(a) Examine a scatterplot of subjects' weight loss vs. their initial weight. Comment on whether the plot reveals any association between these two variables.

(b) Examine a scatterplot of subjects' weight loss vs. their adherence levels. Comment on whether the plot reveals any association between these two variables.

- A **labeled scatterplot** includes information from a categorical variable.

(c) Produce a labeled scatterplot (choose `Graph > Scatterplot > With groups`) of weight loss vs. adherence level, using diet as the categorical variable. Comment on whether the relationship differs depending on which diet the person was using.