Scatterplot: Examples

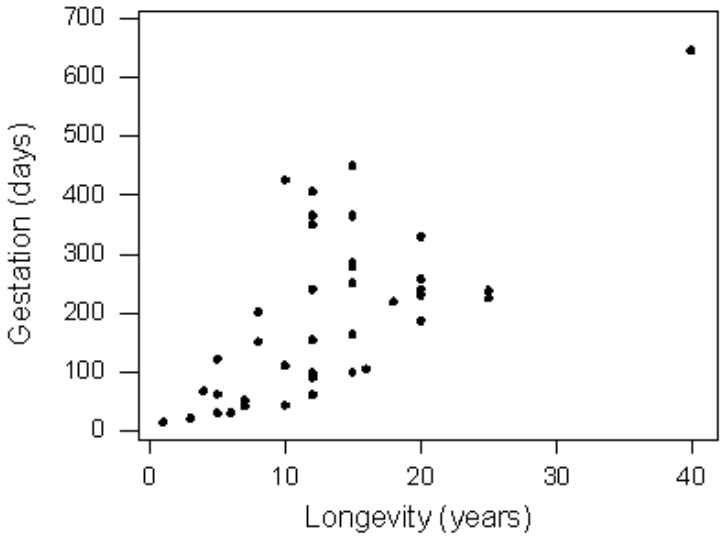
Learning Objective: Graphically display the relationship between two quantitative variables and describe: a) the overall pattern, and b) striking deviations from the pattern.

We will now look at two more examples:

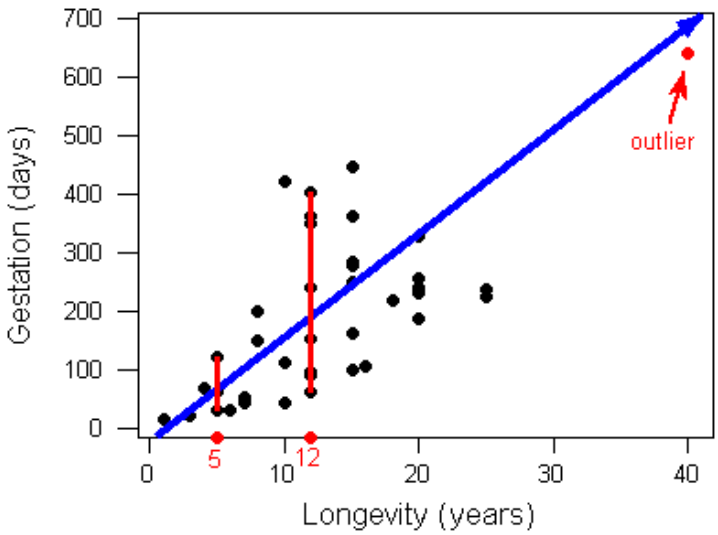
Example: Average Gestation Period

The average gestation period, or time of pregnancy, of an animal is closely related to its longevity (the length of its lifespan.) Data on the average gestation period and longevity (in captivity) of 40 different species of animals have been examined, with the purpose of examining how the gestation period of an animal is related to (or can be predicted from) its longevity. (Source: Rossman and Chance. (2001). Workshop statistics: Discovery with data and Minitab. Original source: The 1993 world almanac and book of facts).

Here is the scatterplot of the data.



What can we learn about the relationship from the scatterplot? The direction of the relationship is **positive**, which means that animals with longer life spans tend to have longer times of pregnancy (this makes intuitive sense). An arrow drawn over the scatterplot below illustrates this:

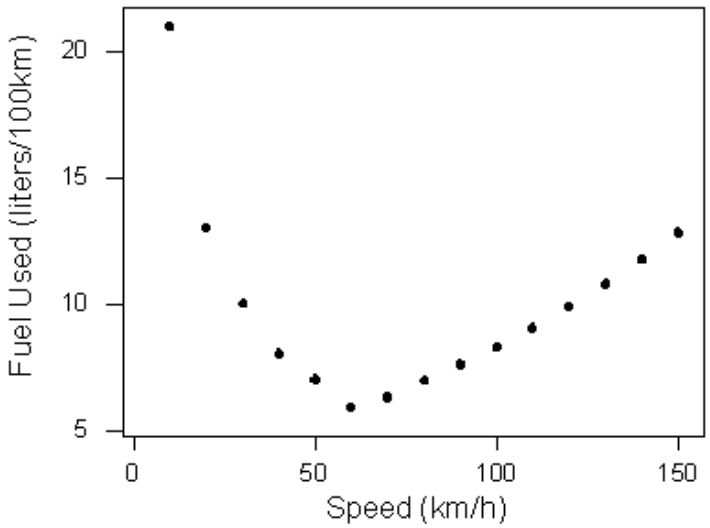


The form of the relationship is again essentially **linear**. There appears to be **one outlier**, indicating an animal with an exceptionally long longevity and gestation period. (This animal happens to be the elephant.) Note that while this outlier definitely deviates from the rest of the data in term of its magnitude, it **does** follow the direction of the data.

Comment: Another feature of the scatterplot that is worth observing is how the variation in gestation increases as longevity increases. This fact is illustrated by the two red vertical lines at the bottom left part of the graph. Note that the gestation periods for animals who live 5 years range from about 30 days up to about 120 days. On the other hand, the gestation period of animals who live 12 years varies much more, and ranges from about 60 days up to more than400 days.

Example: Fuel Usage

As a third example, consider the relationship between the average amount of fuel used (in liters) to drive a fixed distance in a car (100 kilometers), and the speed at which the car is driven (in kilometers per hour). (Source: Moore and McCabe, (2003). Introduction to the practice of statistics. Original source: T.N. Lam. (1985). "Estimating fuel consumption for engine size," Journal of Transportation Engineering, vol. 111)

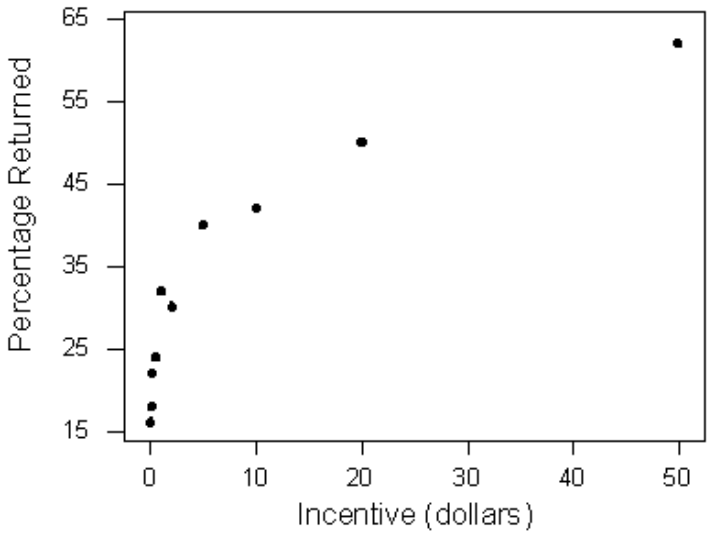


The data describe a relationship that decreases and then increases—the amount of fuel consumed decreases rapidly to a minimum for a car driving 60 kilometers per hour, and then increases gradually for speeds exceeding 60 kilometers per hour. This suggests that the speed at which a car economizes on fuel the most is about 60 km/h. This forms a curvilinear relationship that seems to be very strong, as the observations seem to perfectly fit the curve. Finally, there do not appear to be any outliers.

Learn By Doing

1/1 point (graded)

A study examined how the percentage of participants who completed a survey is related to the monetary incentive that researchers promised to participants. Consider the relationship between these two quantitative variables, displayed in the scatterplot below.



What is the direction of this relationship?

- (●) positive ✔
- ( ) negative
- ( ) neither positive nor negative

Answer
Correct:
As the incentive amount increases the percentage of returned surveys also tends to increase indicating a positive relationship.

Submit

Learn By Doing

1/1 point (graded)
In the context of this example, when researchers promised higher payments, what happened to the percentage of participants who completed the survey?

- (●) increased ✔
- ( ) remained the same

decreased

Answer
Correct:
Since a positive relationship exists, when the values of one variable increase, so do the values of the other variable.

Submit

Learn By Doing

1/1 point (graded)
What is the form of the relationship?

linear

curvilinear ✔

neither linear nor curvilinear

Answer
Correct:
Relationships with a curvilinear form have points dispersed around a curved line, which appears to happen in this scatterplot at the higher incentive amounts.

Submit

Learn By Doing

1/1 point (graded)
Based on the form of the relationship as it is illustrated above, is this a weak relationship or a strong relationship?

strong ✔

weak

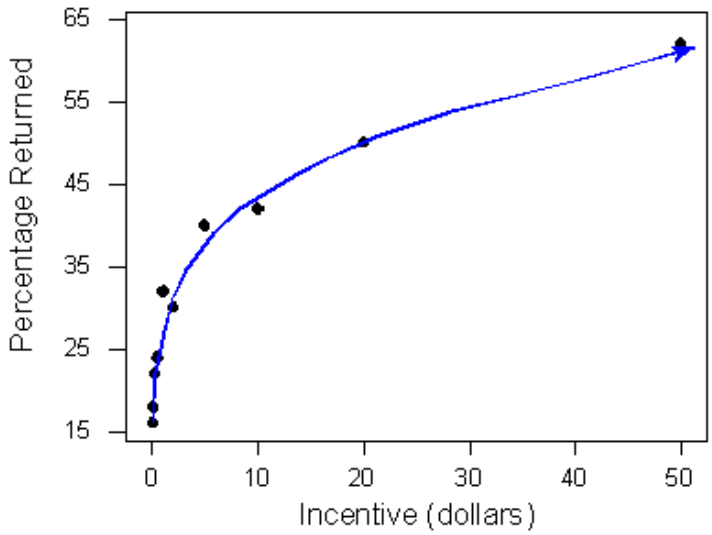
Answer

Correct: The data points fit tightly into a curvilinear form.

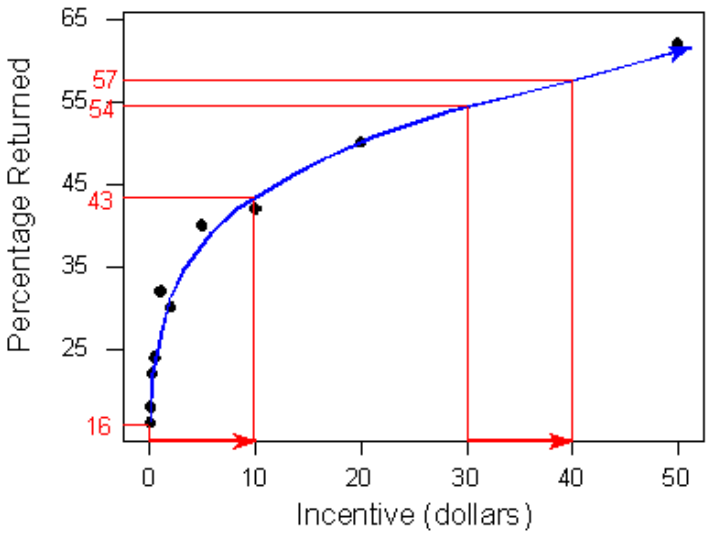
Submit

Comment

The example in the last activity provides a great opportunity for interpretation of the form of the relationship in context. Recall that the example examined how the percentage of participants who completed a survey is affected by the monetary incentive that researchers promised to participants. Here again is the scatterplot that displays the relationship:



The positive relationship definitely makes sense in context, but what is the interpretation of the curvilinear form in the context of the problem? How can we explain (in context) the fact that the relationship seems at first to be increasing very rapidly, but then slows down? The following graph will help us:



Note that when the monetary incentive increases from $0 to $10, the percentage of returned surveys increases sharply—an increase of 27% (from 16% to 43%). However, the same increase of $10 from $30 to $40 doesn't result in the same dramatic increase in the percentage of returned surveys—it results in an increase of only 3% (from 54% to 57%). The form displays the phenomenon of "diminishing returns"—a return rate that after a certain point fails to increase proportionately to additional outlays of investment. $10 is worth more to people relative to $0 than it is relative to $30.
