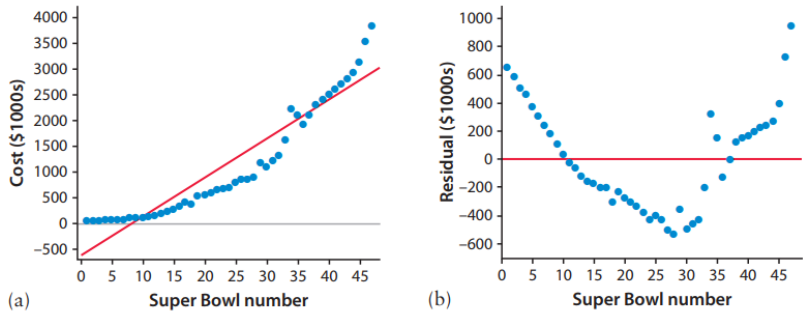
2.7 Assessing a Regression Model (Part 1): Residual Plots

In lesson 2.5, we learned how to calculate and interpret a residual. We can also use residuals to assess whether a regression model is appropriate by making a residual plot.

A **residual plot** is a scatterplot that plots the residuals on the vertical axis and the explanatory variable on the horizontal axis.

The first figure is a scatterplot showing the relationship between Super Bowl number and the cost of a 30-second commercial for the years 1967-2013, along with the least-squares regression line. The resulting residual plot is shown in the other figure.



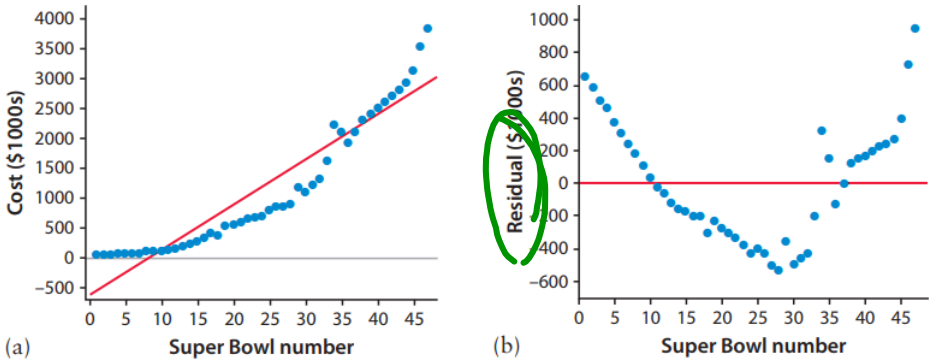
The least-squares regression line clearly doesn't fit this association very well!

In the early years, the actual cost of an ad is always greater than the line predicts, resulting in positive residuals.

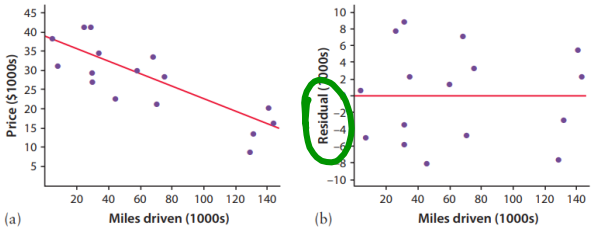
From Super Bowl 11 to Super Bowl 33, the actual cost is always less than the line predicts, resulting in negative residuals.

After Super Bowl 33, the actual cost is almost always greater than the line predicts, again resulting in positive residuals.

The positive-negative-positive pattern in the residual plot indicates that the linear form of our model doesn't match the form of the association. A curved model might be better in this case.



The first figure below gives a scatterplot showing the Ford F-150 data from lesson 2.5 along with the corresponding residual plot.



Looking at the scatterplot, the line seems to be a good fit for the association. You can "see" that the line is appropriate by the lack of a leftover pattern in the residual plot. In fact, the residuals look randomly scattered around the residual = 0 line.

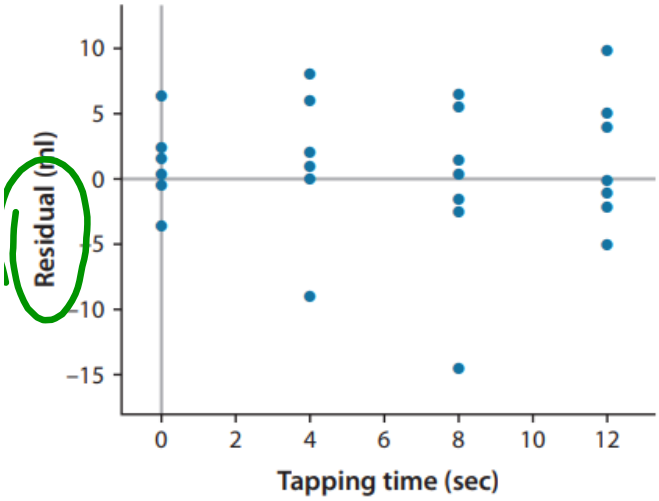
Interpreting a Residual Plot

To determine whether the regression model is appropriate, look at the residual plot.

*If there is no leftover pattern in the residual plot, the regression model is appropriate.

**If there is a leftover pattern in the residual plot, the regression model is not appropriate.

Example: In Lesson 2.5, we used a least-squares regression line to model the relationship between the amount of soda remaining and the tapping time for cans of vigorously shaken soda. Here is the residual plot for that model.



Use the residual plot to determine whether the regression model is appropriate.

Since the plot does not follow a pattern, the model is appropriate.