2.7 Assessing a Regression Model (Part 2): Standard Deviation of Residuals

Once we have all the residuals, we can measure how well the line makes predictions with the standard deviation of the residuals.

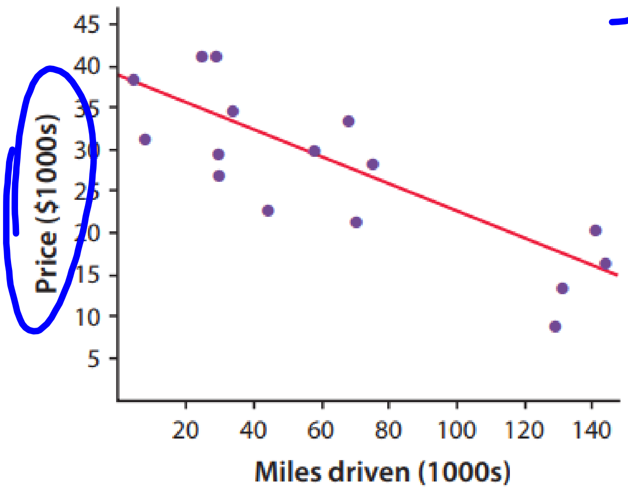
The **standard deviation of the residuals** ***s*** measures the size of a typical residual. That is, *s* measures the typical distance between the actual *y* values and the predicted *y* values.

To calculate the standard deviation of the residuals, we square root each of the residuals, add them, divide the sum by (n - 2), and take the square root.

$$s = \sqrt{\frac{\text{sum of squared residuals}}{n-2}} = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-2}}$$

...Or you can use the "Two Quantitative Variables" function on the website! :)

Example: In Lesson 2.5, we used a least-squares regression line to model the relationship between the price of a Ford F-150 and the number of miles it had driven. The standard deviation of the residuals for this model is *s = $5,740*. Interpret this value.



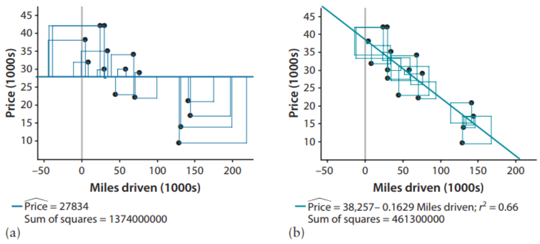
The actual value of the F-150s is typically $5740 away from the predicted value.

2.7 Assessing a Regression Model (Part 3): The Coefficient of Determination r^2

Besides the standard deviation of the residuals *s*, we can also use the coefficient of determination r^2 to measure how well the regression line makes predictions.

The **coefficient of determination r^2** measures the percent reduction in the sum of squared residuals when using the least-squares regression line to make predictions, rather than the mean value of y.
In other words, r^2 measures the $\boxed{\text{percent}}$ of the variability in the response variable that is accounted for by the least-squares regression line.

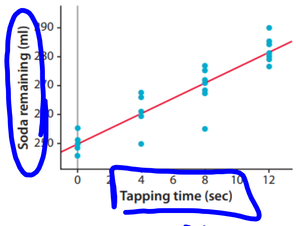
Suppose we wanted to predict the price of a particular Ford F-150 without knowing how many miles it had been driven. The following scatterplots compare the residuals when used with the average price and the residuals when used in the least-squares regression line.



To find r^2, calculate the percent reduction in the sum of squared residuals:

$$r^2 = \frac{1{,}374{,}000{,}000 - 461{,}300{,}000}{1{,}374{,}000{,}000} = \frac{912{,}700{,}000}{1{,}374{,}000{,}000} = 0.66$$

The sum of squared residuals has been reduced by 66%. That is, 66% of the variability in the price of a Ford F-150 is accounted for by the least-squares regression line with x = miles driven.



Example: In Lesson 2.5, we used a least-squares regression line to model the relationship between the amount of soda remaining (in milliliters) and the tapping time (in seconds) for cans of vigorously shaken soda. Interpret the value r^2 = 0.85 for this model.

85%

85% of the variability in the amount of soda remaining is based off the tapping time.