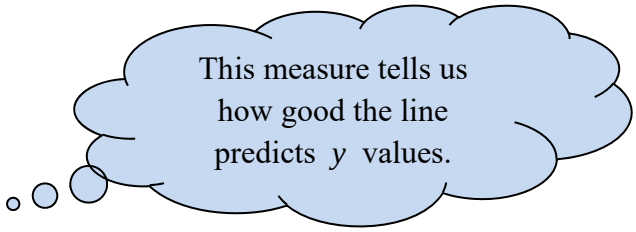
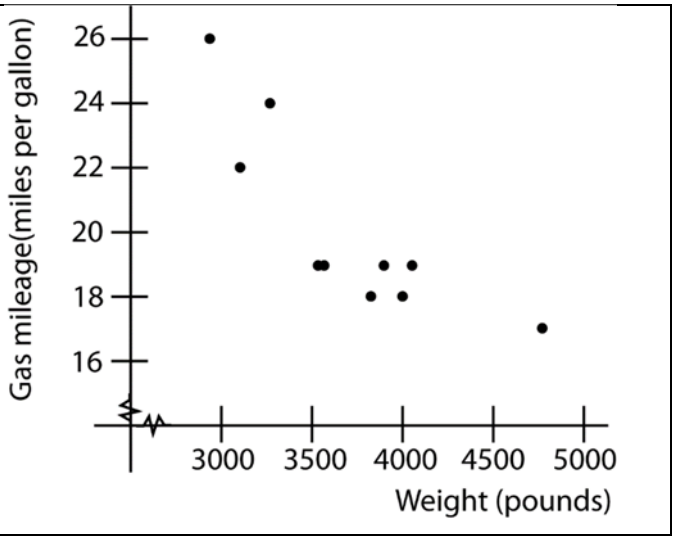


General Education Statistics
Class Notes
The Coefficient of Determination (Section 4.3)

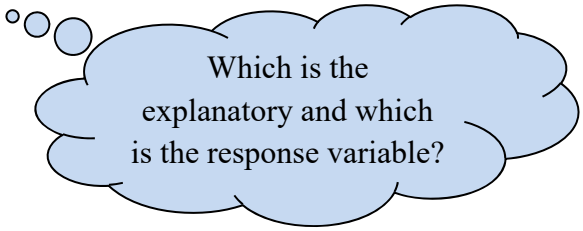
Here, we analyze a measure of the least-squares regression line.

expl 1: Consider the car weight and gas mileage data we have worked with previously.
(source: each manufacturer's website through textbook)

Car	Weight, x (pounds)	Gas Mileage, y (mpg)
Buick LaCrosse	4724	17
Cadillac XTS	4006	18
Chevy Cruze	3097	22
Chevy Impala	3555	19
Chrysler 300	4029	19
Dodge Charger	3934	19
Dodge Dart	3242	24
Ford Focus	2960	26
Ford Mustang	3530	19
Lincoln MKZ	3823	18



You'll recall, in the last section of notes, we found the least-squares regression line to be $\hat{y} = -.00468x + 37.357$ where x represents the weight of the car (in pounds) and $\hat{y}$ represents the gas mileage (in miles per gallon).



We found r, the coefficient of correlation, to be -.84158. This told us the line fit the data well with a negative slope. The calculator also gives us a value for r^2. What is that?

We know that as the weight of the car increases, the gas mileage decreases. This line shows this relationship. But how well does it do?

This value of r^2 measures how well the regression line describes the relationship between the explanatory and response variables.

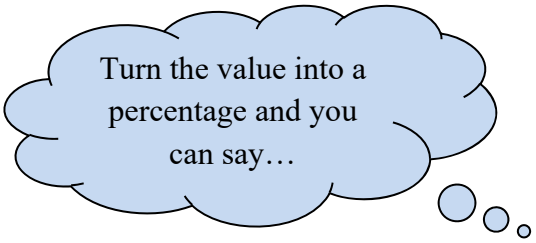
Definition: Coefficient of Determination: The **coefficient of determination**, $\boldsymbol{R^2}$, measures the *proportion of total variation* in the response variable that is explained by the least-squares regression line.

The coefficient of determination is a number between 0 and 1, inclusive. That is, $0 \leq R^2 \leq 1$.

If $R^2 = 0$, then the line has *no* explanatory value.

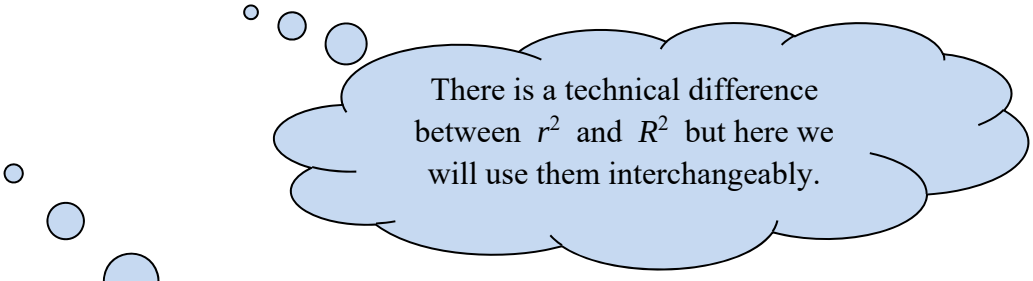
The closer R^2 is to 1, the better the line describes how changes in the explanatory variable affect the value of the response variable.

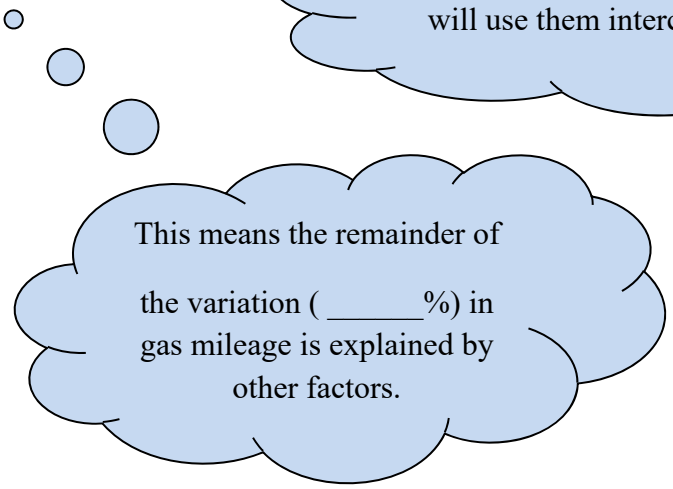
If $R^2 = 1$, then the line explains 100% of the variation in the response variable.



expl 1 continued: The value for r^2 for the car weight/gas mileage example is given, by the calculator, as .708. You could also get that by taking the value for r (-.8416) and squaring it. Turn this into a percent and complete the sentence below. Round to the nearest tenth of a percent.

_____ % of the variation in gas mileage can be explained by the least-squares regression line.

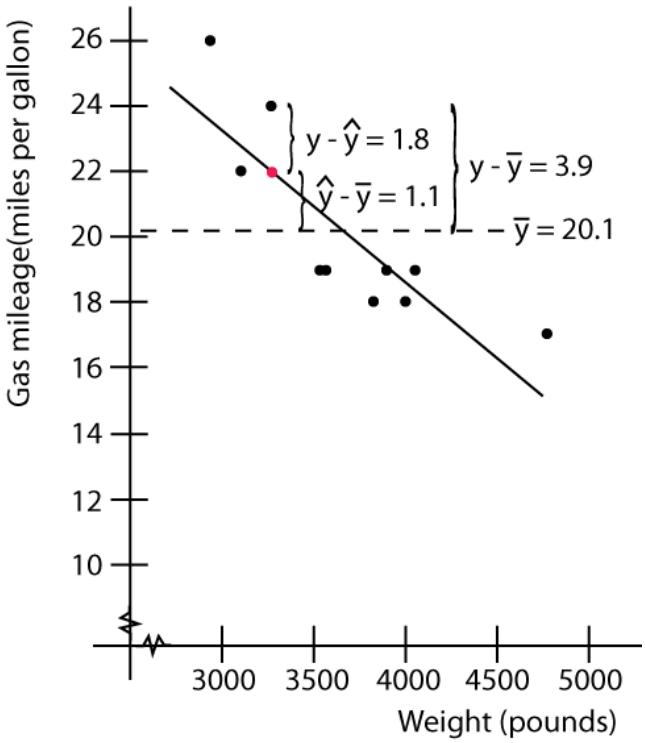




Optional Section: Deviations:

We return to this car weight/gas mileage example. The mean value of the response variable is $\bar{y} = 20.1$ miles per gallon. This is a simple average of the values in the table.

Consider the Dodge Dart (which has an observed gas mileage of 24 miles per gallon and a weight of 3,242 pounds). Find this point on the graph now and circle it. Let's compare our regression prediction against reality and against the mean of the sample.



The difference between the observed value and the mean value is $y - \bar{y} = 24 - 20.1 = 3.9$ miles per gallon. This is called the **total deviation**. Do you see it labeled on the graph?

On the other hand, the least-squares line gives us

$$\begin{aligned}\hat{y} &= -.00468x + 37.357 \\ &= -.00468(3242) + 37.357 \\ &\approx 22.2\end{aligned}$$

The difference between this predicted value (shown as a red dot) and the mean, or $\hat{y} - \bar{y} = 22.2 - 20.1 = 1.1$ miles per gallon, is called the **explained deviation**.

Finally, the difference between the observed value (of 24 miles per gallon) and the predicted value (of 22.2 miles per gallon), or $y - \hat{y} = 24 - 22.2 = 1.8$, is called the **unexplained deviation**.

Now we see why the first is called the total deviation. **Notice how the total deviation equals the explained deviation plus the unexplained deviation.**

It is beyond what we want to discuss here, but it can be shown that the closer the observed y-values are to the regression line (the predicted y-values), the larger R^2 will be. In other words, the value of R^2 will be closer to 1 if the points line up closer to a perfect line. You can use this to estimate the value of R^2 given a graph of points and their regression equation.