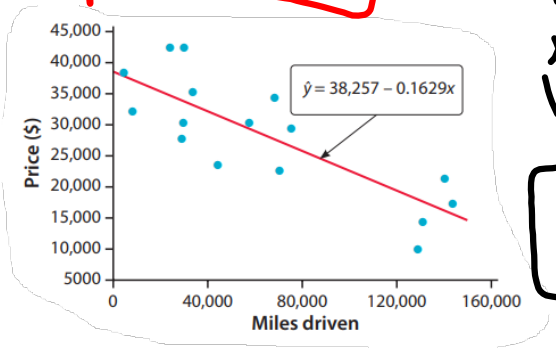
2.5 Regression Lines (Part 1): Making Predictions

When the relationship between two quantitative variables in linear, we can use a **regression line** to model the relationship and make predictions.

A **regression line** is a line that describes how a response variable y changes as an explanatory variable x changes. Regression lines are expressed in the form $\hat{y} = a + bx$, where $\hat{y}$ (pronounced "y-hat") is the predicted value of y for a given value of x.

We could also express the regression line in the form $y = mx + b$ like we do in algebra. However, statisticians prefer this reordered format because it works better when they use several explanatory variables. Just remember, the slope is always the coefficient of the x variable.
The most common use of a regression line is to make predictions.

Example: Everyone knows that cars and trucks lose value the more miles they are driven. Can we predict the price of a Ford F-150 if we know how many miles it has on the odometer? A random sample of 16 Ford F-150 SuperCrew 4x4's was selected from among those listed for sale at autotrader.com. The number of miles driven and price (in dollars) were recorded for each of the trucks. Here is a scatterplot of the data, along with the regression line $\hat{y} = 38{,}257 - 0.1629x$, where x is the miles driven and y is the price. Predict the cost of a Ford F-150 that has been driven 100,000 miles. X



$\hat{y} = 38257 - 0.1629(100000)$

$\hat{y} = 38257 - 16290$

$\hat{y} = \$21967$

Can we predict the price of a Ford F-150 with 300,000 miles driven? We can certainly substitute 300,000 into the equation of the line.

The prediction is:

$$\hat{y} = 38{,}257 - 0.1629(300{,}000) = -\$10{,}613$$

That is, we predict that we would need to pay someone else $10,613 just to take the truck off our hands! This prediction is an **extrapolation**.

Extrapolation is the use of a regression line for prediction far outside the interval of x values used to obtain the line. Such predictions are often not accurate.