Vanderbilt University Political Science - Stats I
Fall 2024 - Prof. Jim Bisbee

15 Lecture 14

- Rewrite sequence of four tasks on board:

  1. DISPLAYING bivariate relationships
  2. SUMMARIZING bivariate relationships NONPARAMETRICALLY
  3. SUMMARIZING bivariate relationships PARAMETRICALLY–that is, the extent to which they approximate a linear relationship
  4. MAKING INFERENCES about the nature of this relationship from a sample to a population

- – Now picking up at section 3:

- recall the correlation coefficient, r :

$$r = \frac{\sum_i \left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sqrt{\sum_i \left(X_i - \overline{X}\right)^2 \sum_i \left(Y_i - \overline{Y}\right)^2}}.$$

  1. – You'll recall from last lecture the strengths and drawbacks of using r as a measure of the association between two variables. Remember that like all statistics, with r we are gaining parsimony while losing details.

- – Because the quantities that appear in r are used often in regression analysis, they have special symbols:

$$\begin{aligned}
S_{xy} &= \textstyle\sum_i \left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right) \\
S_{xx} &= \textstyle\sum_i \left(X_i - \overline{X}\right)^2 = \sum_i \left(X_i - \overline{X}\right)\left(X_i - \overline{X}\right) \\
S_{yy} &= \textstyle\sum_i \left(Y_i - \overline{Y}\right)^2
\end{aligned}$$

So how can we write r?

$$r = \frac{\sum_i \left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sqrt{\sum_i \left(X_i - \overline{X}\right)^2 \sum_i \left(Y_i - \overline{Y}\right)^2}} = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}$$

- now, (finally) linear regression.

- what if we wanted to put some more meat on the extent to which a relationship between x and y is linear? In particular, what if we wanted to say something about the line that best represented the relationship in linear terms? Well, we'd start by specifying a generic formula for that line, which convention has led us to write as follows:

$$y = \beta_0 + \beta_1 x$$

  where β_0 is the intercept and β_1 the slope.

- Draw on board:

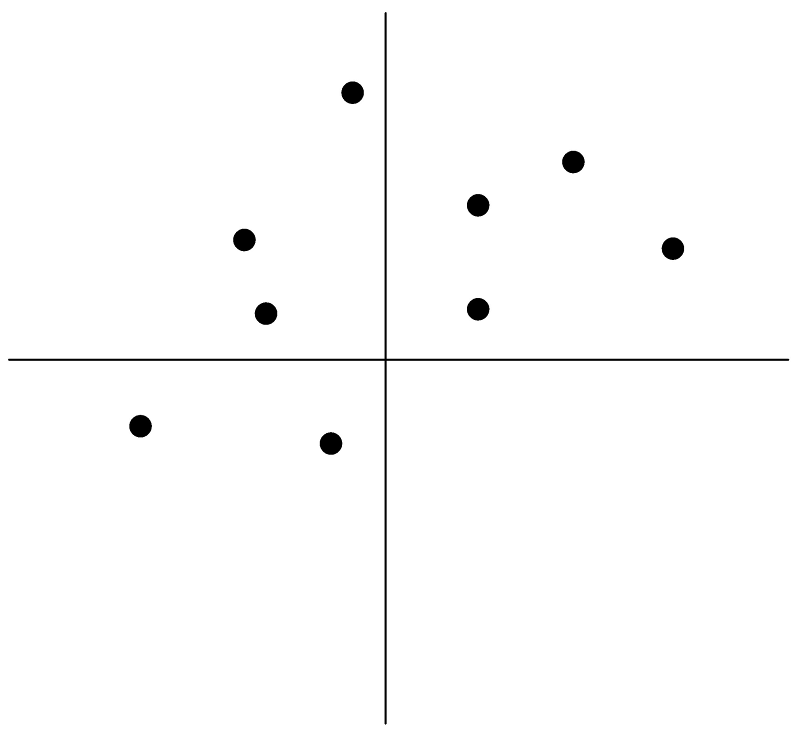

Figure 1: Caption

- There are many such lines we might choose to represent this relationship in linear terms. And even the best line can't hit every point on the nose; it will certainly make mistakes. We define the mistake our line makes for any particular observation i as the observation y_i minus the "fitted value" for that observation, which we'll write as $\widehat{y}_i$:

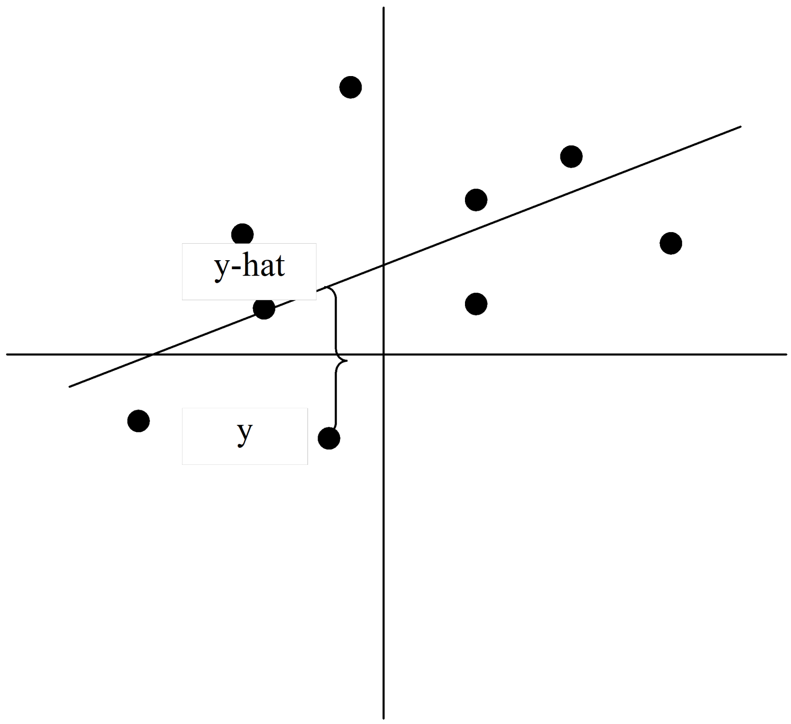

Figure 2: Caption

- Another term for these mistakes is residuals. We for any observation i we write the residual associated with that observation as $\widehat{u}_i = y_i - \widehat{y}_i$.

- An obvious criterion to use for picking the best line would be to minimize the mistakes it makes as it proceeds through the xy plane. That is, we could minimize $|y_i - \widehat{y}_i|$ as much as possible by minimizing $\sum_i |y_i - \widehat{y}_i|$.

- Well (as you will or have already learned), the absolute value function is a lousy one to work with mathematically. It has annoying properties that do not lend itself to easy manipulation. A more useful and easier sum to minimize is

$$SSR = \sum_i (y_i - \widehat{y}_i)^2 ,$$

the **sum of squared residuals (SSR).** Note that the square of the residual has the nice property of becoming bigger as the magnitude of the residual increases. (It has the less felicitous property of counting bigger distances as greater than smaller distances: for example if we double a distance of 3 to 6, the corresponding squares are 9 and 36: a quadrupling. Thus whatever method we pick that minimizes SSR is going to work harder to minimize big deviations from the line than it probably should. But we're getting ahead of ourselves.)

- Now let's go back to our formula for a line. Let's do two things:

  - rewrite it with hats to be clear that we are generating estimates rather than saying anything specific (yet) about population values, and
  - add subscripts for x and y:

$$\widehat{y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_i$$

- Now let's consider the residual again, $\widehat{u}_i$, and substitute:

$$\begin{aligned}
\widehat{u}_i &= y_i - \widehat{y}_i \\
&= y_i - \left(\widehat{\beta}_0 + \widehat{\beta}_1 x_i\right) \\
(\widehat{u}_i)^2 &= \left[y_i - \left(\widehat{\beta}_0 + \widehat{\beta}_1 x_i\right)\right]^2 \\
\textstyle\sum_i (\widehat{u}_i)^2 &= \textstyle\sum_i \left[y_i - \left(\widehat{\beta}_0 + \widehat{\beta}_1 x_i\right)\right]^2
\end{aligned}$$

- So the $\widehat{\beta}_0$ and $\widehat{\beta}_1$ that minimize this quantity are the intercept and the slope of what is known as the **least squares** line–the line that best represents the relationship between x and y. We can use the tools of calculus to find $\widehat{\beta}_0$ and $\widehat{\beta}_1$ by taking the partial derivative of SSR with respect to each of these quantities and setting these derivatives equal to zero:

$$\begin{aligned}
\frac{\partial SSR}{\partial \widehat{\beta}_0} &= \frac{\partial}{\partial \widehat{\beta}_0} \textstyle\sum_i \left[y_i - \left(\widehat{\beta}_0 + \widehat{\beta}_1 x_i\right)\right]^2 \\
&= -2 \textstyle\sum_i y_i - \left(\widehat{\beta}_0 + \widehat{\beta}_1 x_i\right) \\
&= -2 \left(\textstyle\sum_i y_i - n\widehat{\beta}_0 - \widehat{\beta}_1 \textstyle\sum_i x_i\right) = 0
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial SSR}{\partial \widehat{\beta}_1} &= \frac{\partial}{\partial \widehat{\beta}_1} \sum_i \left[y_i - \left(\widehat{\beta}_0 + \widehat{\beta}_1 x_i \right) \right]^2 \\
&= -2 \sum_i \left[y_i - \left(\widehat{\beta}_0 + \widehat{\beta}_1 x_i \right) \right] x_i \\
&= -2 \left(\sum_i x_i y_i - \widehat{\beta}_0 \sum_i x_i - \widehat{\beta}_1 \sum_i x_i^2 \right) = 0
\end{aligned}
$$

- These lead to the **normal equations**

$$
\begin{aligned}
n\widehat{\beta}_0 + \widehat{\beta}_1 \sum_i x_i &= \sum_i y_i \\
\widehat{\beta}_0 \sum_i x_i + \widehat{\beta}_1 \sum_i x_i^2 &= \sum_i x_i y_i
\end{aligned}
$$

- Rewrite in matrix form:

$$
\begin{bmatrix} n & \sum_i x_i \\ \sum_i x_i & \sum_i x_i^2 \end{bmatrix} \begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \end{bmatrix} = \begin{bmatrix} \sum_i y_i \\ \sum_i x_i y_i \end{bmatrix}
$$

- So:

$$
\begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \end{bmatrix} = \begin{bmatrix} n & \sum_i x_i \\ \sum_i x_i & \sum_i x_i^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_i y_i \\ \sum_i x_i y_i \end{bmatrix}
$$

- The inverse of a 2x2 matrix $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is $\mathbf{A}^{-1} = \frac{1}{\det \mathbf{A}} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$

- So

$$
\begin{bmatrix} n & \sum_i x_i \\ \sum_i x_i & \sum_i x_i^2 \end{bmatrix}^{-1} = \frac{1}{n \sum_i x_i^2 - \left(\sum_i x_i\right)^2} \begin{bmatrix} \sum_i x_i^2 & -\sum_i x_i \\ -\sum_i x_i & n \end{bmatrix}
$$

- And thus

$$
\begin{aligned}
\begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \end{bmatrix} &= \frac{1}{n\sum_i x_i^2 - \left(\sum_i x_i\right)^2} \begin{bmatrix} \sum_i x_i^2 & -\sum_i x_i \\ -\sum_i x_i & n \end{bmatrix} \begin{bmatrix} \sum_i y_i \\ \sum_i x_i y_i \end{bmatrix}, \text{ so} \\
\widehat{\beta}_0 &= \frac{\sum_i x_i^2 \sum_i y_i - \sum_i x_i \sum_i x_i y_i}{n\sum_i x_i^2 - \left(\sum_i x_i\right)^2}, \widehat{\beta}_1 = \frac{n\sum_i x_i y_i - \sum_i x_i \sum_i y_i}{n\sum_i x_i^2 - \left(\sum_i x_i\right)^2}.
\end{aligned}
$$

We can simplify these as follows :

$$
\begin{aligned}
\widehat{\beta}_0 &= \frac{n\overline{y}\sum_i x_i^2 - n\overline{x}\sum_i x_i y_i}{n\sum_i x_i^2 - \left(n\overline{x}\right)^2} = \frac{\overline{y}\sum_i x_i^2 - \overline{x}\sum_i x_i y_i}{\sum_i x_i^2 - n\left(\overline{x}\right)^2} \text{ and similarly} \\
\widehat{\beta}_1 &= \frac{n\sum_i x_i y_i - n^2\overline{xy}}{n\sum_i x_i^2 - n^2\left(\overline{x}\right)^2} = \frac{\sum_i x_i y_i - n\overline{xy}}{\sum_i x_i^2 - n\left(\overline{x}\right)^2}.
\end{aligned}
$$

- Simplify further by substituting the symbols we learned earlier with a little manipulation:

$$
\begin{aligned}
S_{xx} &= \sum_i \left(X_i - \overline{X}\right)^2 = \sum_i X_i^2 + \sum_i \overline{X}^2 - \sum_i 2X_i\overline{X} \\
&= \sum_i X_i^2 - n\overline{X}^2; \\
S_{yy} &= \sum_i Y_i^2 - n\overline{Y}^2 \\
S_{xy} &= \sum_i \left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right) = \sum_i X_i Y_i - \sum_i X_i\overline{Y} - \sum_i \overline{X}Y_i + \sum_i \overline{XY} \\
&= \sum_i X_i Y_i - n\sum_i \overline{XY}
\end{aligned}
$$

- So:

$$
\begin{aligned}
\widehat{\beta}_1 &= \frac{S_{xy}}{S_{xx}}. \text{ Note that because } \frac{cov(x,y)}{var(x)} = \frac{\frac{S_{xy}}{n}}{\frac{S_{xx}}{n}}, \\
\widehat{\beta}_1 &= \frac{cov(x,y)}{var(x)}.
\end{aligned}
$$

- We can write $\widehat{\beta}_0$ more simply if note that we can rewrite

$$\begin{aligned}
-2\left(\sum_i y_i - n\widehat{\beta}_0 - \widehat{\beta}_1 \sum_i x_i\right) &= 0 \\
\sum_i y_i - n\widehat{\beta}_0 - \widehat{\beta}_1 \sum_i x_i &= 0 \\
n\overline{y} - n\widehat{\beta}_0 - n\widehat{\beta}_1\overline{x} &= 0 \\
\widehat{\beta}_0 &= \overline{y} - \widehat{\beta}_1\overline{x}, \text{ and so} \\
\widehat{\beta}_0 &= \overline{y} - \frac{S_{xy}}{S_{xx}}\overline{x}.
\end{aligned}$$

- I will spare you a proof that we have satisfied the second-order condition for a minimum!

- Let's do a quick example. ("Handout on the Math of the Least-Squares Line.")

- Note a bit of terminology that gets thrown around:

  - We "regress y on x"
  - x is a "regressor"
  - x is on the "right hand side" of the regression equation

- Note that we NOT talking about making inferences (yet). Everything I've shown you thus far are simply mathematical properties that follow from the desired criterion of fitting a least squares line to data. Although this line tells us how well the data approximates a line, we have so far made no assumptions about the distribution of x and y in the underlying population.

- We will continue in this vein for a little while longer by discussing additional mathematical properties of the least squares line. Here are a few:

  1. $\widehat{\beta}_1 = \frac{\Delta\widehat{y}}{\Delta x}$. If we start with our fitted line,

$$\begin{aligned}
\widehat{y} &= \widehat{\beta}_0 + \widehat{\beta}_1 x \quad \text{and take its derivative with respect to } x, \\
\frac{d\widehat{y}}{dx} &= \widehat{\beta}_1, \text{ we see (obviously) that:} \\
&\quad \text{a one-unit change in } x \text{ is associated with a change of } \widehat{\beta}_1 \text{ units of } \widehat{y}\text{, or} \\
\widehat{\beta}_1 &= \frac{\Delta\widehat{y}}{\Delta x}.
\end{aligned}$$

Note that I'm saying associated with, not "causes." Don't get too consumed by this yet; we'll talk about this a lot more soon.

2. $\sum_i \widehat{u}_i = 0$; $\overline{\widehat{u}} = 0$.

   – The sum of the residuals, $\sum_i \widehat{u}_i$, equals zero. We see this immediately by noting that

$$\begin{aligned} \widehat{u}_i &= y_i - \widehat{y}_i = y_i - \left(\widehat{\beta}_0 + \widehat{\beta}_1 x_i\right), \\ \sum_i \widehat{u}_i &= \sum_i y_i - \left(\widehat{\beta}_0 + \widehat{\beta}_1 x_i\right) \end{aligned}$$

   and noting that the formula for our regression line satisfies the F.O.C. associated with $\widehat{\beta}_0$:

$$\begin{aligned} \frac{\partial SSR}{\partial \widehat{\beta}_0} &= \frac{\partial}{\partial \widehat{\beta}_0} \sum_i \left[y_i - \left(\widehat{\beta}_0 + \widehat{\beta}_1 x_i\right)\right]^2 \\ &= -2 \sum_i y_i - \left(\widehat{\beta}_0 + \widehat{\beta}_1 x_i\right) = 0 \\ &= \sum_i y_i - \left(\widehat{\beta}_0 + \widehat{\beta}_1 x_i\right) = 0, \text{ so} \\ \sum_i \widehat{u}_i &= 0 \end{aligned}$$

   – Note that this is just a property of the mechanics of fitting a line. We say that $\sum_i \widehat{u}_i = 0$ "by construction." This property is always the case, and it tells us nothing about our data or the relationship between x and y.

   – Note that this also means that

$$\begin{aligned} \sum_i \widehat{u}_i &= n\overline{\widehat{u}} = 0, \text{ and so} \\ \overline{\widehat{u}} &= 0. \end{aligned}$$

   – That is, the sample mean of the residuals is zero.

3. $cov(x, \widehat{u}) = 0$.

   – By construction, the sample covariance between the regressors and the residuals is zero. his follows from the F.O.C. associated with $\widehat{\beta}_1$:

$$\begin{aligned}
\frac{\partial SSR}{\partial \widehat{\beta}_1} &= \frac{\partial}{\partial \widehat{\beta}_1} \sum_i \left[y_i - \left(\widehat{\beta}_0 + \widehat{\beta}_1 x_i \right) \right]^2 \\
&= -2 \sum_i \left[y_i - \left(\widehat{\beta}_0 + \widehat{\beta}_1 x_i \right) \right] x_i = 0 \\
\sum_i \left[y_i - \left(\widehat{\beta}_0 + \widehat{\beta}_1 x_i \right) \right] x_i &= 0
\end{aligned}$$

– But since

$$y_i - \left(\widehat{\beta}_0 + \widehat{\beta}_1 x_i \right) = \widehat{u}_i,$$

We can substitute and write

$$\sum_i \widehat{u}_i x_i = 0.$$

– Since the sample covariance of x and $\widehat{u}$ is

$$\begin{aligned}
cov(x, \widehat{u}) &= \frac{\sum_i \left(\widehat{u}_i - \overline{\widehat{u}} \right) (x_i - \overline{x})}{n}, \text{ it follows that} \\
&= \frac{\sum_i (\widehat{u}_i) (x_i - \overline{x})}{n} \\
&= \frac{\sum_i \widehat{u}_i x_i}{n} - \frac{\sum_i \widehat{u}_i \overline{x}}{n} \\
&= 0 - \frac{\overline{x} \sum_i \widehat{u}_i}{n} = 0 - 0 = 0.
\end{aligned}$$

4. $cov\left(\widehat{y}_i, \widehat{u}_i\right) = 0$. [LEAVE AS EXERCISE.]

– The sample covariance between the fitted values and the residuals is zero.

$$\begin{aligned}
cov\left(\widehat{y}_i, \widehat{u}_i\right) &= \frac{\sum_i (y_i - \overline{y}) \left(\widehat{u}_i - \overline{\widehat{u}} \right)}{n} \\
&= \frac{\sum_i (y_i - \overline{y}) (\widehat{u}_i)}{n} \\
&= \frac{\sum_i \widehat{u}_i y_i}{n} - \frac{\sum_i \widehat{u}_i \overline{y}}{n} \\
&= \frac{\sum_i \widehat{u}_i y_i}{n} - \frac{\overline{y} \sum_i \widehat{u}_i}{n}
\end{aligned}$$

– What is $\sum_i \widehat{u}_i y_i$? It's

$$\sum_i \widehat{u}_i y_i = \sum_i (y_i - \widehat{y}_i)\, y_i$$

5. The point $(\overline{x}, \overline{y})$ is always on the regression line.

   – Show this by substituting $\overline{x}$ for x, and $\widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1 \overline{x}$ in the formula for the regression line:

$$\begin{aligned} \widehat{y} &= \widehat{\beta}_0 + \widehat{\beta}_1 x \\ \widehat{y}(\overline{x}) &= \overline{y} - \widehat{\beta}_1 \overline{x} + \widehat{\beta}_1 \overline{x} \\ \widehat{y}(\overline{x}) &= \overline{y}. \end{aligned}$$

6. $\frac{\sum_i \widehat{y}_i}{n} = \frac{\sum_i y_i}{n}$.

   – By construction, the sample average of the fitted values is equal to the sample average of the observed y's, that is $\frac{\sum_i \widehat{y}_i}{n} = \frac{\sum_i y_i}{n}$. To show this, note that

$$\begin{aligned} y_i &= \widehat{y}_i + \widehat{u}_i, \text{ and so} \\ \sum_i y_i &= \sum_i \widehat{y}_i + \sum_i \widehat{u}_i \\ \sum_i y_i &= \sum_i \widehat{y}_i + 0 \\ \frac{\sum_i y_i}{n} &= \frac{\sum_i \widehat{y}_i}{n}. \end{aligned}$$

- We can view the process of fitting a least squares line as decomposing each y_i into two parts: $\widehat{y}_i$ and $\widehat{u}_i$. These values are by construction uncorrelated in the sample.

  – Define the total sum of squares (SST) as

$$SST = \sum_i (y_i - \overline{y})^2.$$

  – Define the explained sum of squares (SSE) as

$$SSE = \sum_i (\widehat{y}_i - \overline{y})^2$$

– Recall that SSR is

$$SSR = \sum_i \widehat{u}_i^2$$

- It can be shown that $SST = SSE + SSR$. (From proof on p. 39 of Wooldridge, but it is complete here.)

- Start with the identity

$$\begin{aligned}
SST &= \sum_i (y_i - \overline{y})^2 = \sum_i (y_i - \widehat{y}_i + \widehat{y}_i - \overline{y})^2 \\
&= \sum_i (\widehat{u}_i + \widehat{y}_i - \overline{y})^2 \\
&= \sum_i (\widehat{u}_i)^2 + \sum_i (\widehat{y}_i - \overline{y})^2 + 2\sum_i \widehat{u}_i (\widehat{y}_i - \overline{y}) \\
&= SSR + SSE + 2\sum_i (\widehat{y}_i - \overline{y})\, \widehat{u}_i
\end{aligned}$$

- Since

$$\begin{aligned}
0 &= cov(\widehat{y}_i, \widehat{u}_i) = \frac{\sum_i (y_i - \overline{y}) \left(\widehat{u}_i - \overline{\widehat{u}}\right)}{n} \\
&= \frac{\sum_i (y_i - \overline{y})(\widehat{u}_i)}{n} = \sum_i (y_i - \overline{y})(\widehat{u}_i),
\end{aligned}$$

we can write

$$SST = SSR + SSE.$$

15.1 More on Sums of Squares

- Last time we defined the three quantities

$$SST = \sum_i (y_i - \overline{y})^2 .$$

–

$$SSE = \sum_i (\widehat{y}_i - \overline{y})^2$$

$$SSR = \sum_i \widehat{u}_i^2$$

- And showed that

$$SST = SSE + SSR.$$

- Let's think about these quantities in a bit more detail:
  - What is SST? It's the sample variance of Y - the extent to which it varies about its mean.
  - If the relationship between x and y could be perfectly described by a line, the fitted values (the y-hats) would would be the same distance from the mean as the observed values every time.
  - In that case $SSE = SST$, and the ratio $\frac{SSE}{SST} = 1$.
  - This is never the case, but we can exploit this fact to construct a measure of the "goodness of fit" of the regression line to the data. Call the ratio $\frac{SSE}{SST} = R^2$.
    * It always ranges between zero and one.
    * When reporting R^2, we typically report it to two decimal places.
    * R^2 is the proportion of the sample variation in y that is explained by x.
    * Also note that $R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST}$.
    * Where did R^2 get its name? Because it is also the case that in the bivariate context, $R^2 = (r)^2$.
      · You'll show that in an exercise on this week's homework.

15.2 Least-Squares Regression is Invariant to Change in Units of Measurement

- What happens when we change the units in which the IV is measured (typically by multiplying these values by some constant, c)?
  - Recall that $\widehat{\beta}_1 =$ the change in $\widehat{y}$ associated with a one-unit change in x.
  - So the change in $\widehat{y}$ associated with a one-unit change in cx should be $\frac{\widehat{\beta}_1}{c}$. And indeed it is:
- So when the IV is multiplied by c,
  - $\widehat{\beta}_1$ is divided by c.

- $\widehat{\beta}_0$ does not change (it is the y-intercept, and $cx = 0$ when $x = 0$)
- R^2 does not change.

- When the DV is multiplied by c,
  - (Again) recall that $\widehat{\beta}_1 =$ the change in $\widehat{y}$ associated with a one-unit change in x.
  - So the change in $c\widehat{y}$ associated with a one-unit change in x is $c\widehat{\beta}_1$.
  - When $x = 0$, the intercept is $c\widehat{y}$.
  - So $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are multiplied by c when the DV is multiplied by c.
  - R^2 does not change.
- All of these will be more exciting exercises on this week's homework, too.