Section 3. Simple Regression – OLS Estimators

1. Last time: Lines, Population Linear Regression
2. Estimators of $\beta_0, \beta_1$
3. Least Squares Estimators
4. OLS Assumptions
1. Review: Lines, Clouds and Pop. Linear Regression

- e.g. Demand for Coffee, Global warming, CA test scores and student-teacher ratios
- Decide which parameters in population we care about ($\beta_0, \beta_1$)
  - just like we did with $\mu$
- Draw a sample and estimate parameters
  - just like we did with $\mu$
- Construct CI for parameters, test hypotheses, make predictions.
  - just like..
## Coffee Example (with a line)

What’s the slope of the line?

About how many cups would you sell at $1?
Global Warming Example

Is the slope statistically different from zero?
The estimated regression line shows a negative relationship between test scores and the student-teacher ratio. If class sizes fall by 1 student, the estimated regression predicts that test scores will increase by 2.28 points.

\[ \text{TestScore} = 698.9 - 2.28 \times STR \]
2. Estimators of $\beta_0, \beta_1$

- Decide which parameters in population we care about ($\beta_0, \beta_1$)
  - we chose Pop. Linear Regression
- Draw a sample and estimate parameters
- Which estimates to use?
- We’d like something with nice properties:
  - unbiased, $E(b_i) = \beta_i$
  - consistent, $b_i \xrightarrow{p} \beta_i$
  - efficient, $V(b_i)$ is smallest
  - with an approximately normal distribution.
3. The Ordinary Least Squares (OLS) Estimators

- Try \( b_0 \) and \( b_1 \) that minimize the sum of \( e_i^2 \)
  \[
  \sum_{i=1}^{n} e_i^2
  \]
  where \( e_i = Y_i - (b_0 + b_1 x) \)

- Why this one? Well, it’s analogous to what we asked for in the population.
- And it seems like a nice property in the sample.
- Deriving formulae for OLS estimators..
  (S&W p. 143)
Minimize \( \sum e_i^2 = \sum (y_i - (b_0 + b_1 x_i))^2 \) by choice of \( b_0, b_1 \)

First order conditions:

\[
\frac{\partial \sum e_i^2}{\partial b_0} = 0, \quad \frac{\partial \sum e_i^2}{\partial b_1} = 0
\]

\[
0 = \frac{\partial \sum e_i^2}{\partial b_0} = \frac{\sum (y_i - (b_0 + b_1 x_i))^2}{\partial b_0} = \frac{\partial (y_i - (b_0 + b_1 x_i))^2}{\partial b_0} + \frac{\partial (y_2 - (b_0 + b_1 x_2))^2}{\partial b_0} + \ldots
\]

\[
= 2e_i \frac{\partial e_i}{\partial b_0} + \ldots
\]

\[
= 2e_i(-1) + 2e_2(-1) + \ldots + 2e_N(-1) = 2(-1) \sum e_i = 0
\]

\( \iff \sum e_i = 0 \neq 1 \)

\[
0 = \frac{\partial \sum e_i^2}{\partial b_1} = \frac{\sum (y_i - (b_0 + b_1 x_i))^2}{\partial b_1} = \frac{\partial (y_i - (b_0 + b_1 x_i))^2}{\partial b_1} + \frac{\partial (y_2 - (b_0 + b_1 x_2))^2}{\partial b_1} + \ldots
\]

\[
= 2e_i \frac{\partial e_i}{\partial b_1} + \ldots
\]

\[
= 2e_i(-x_i) + 2e_2(-x_2) + \ldots + 2e_N(-x_N) = -2 \sum e_i x_i
\]

\( \iff \sum e_i x_i = 0 \neq 2 \)
2. F.O.C. \( \xi c_i = 0 \), \( \xi x_i = 0 \), now solve for \( b_0 \), \( b_1 \).

\[
\xi [y_i - (b_0 + b_1 x_i)] = 0 \Rightarrow \frac{1}{N} \xi y_i - \frac{1}{N} \xi (b_0 + b_1 x_i) = 0
\]

\[
\bar{y} = \frac{1}{N} \sum b_0 + b_1 \bar{x} \Rightarrow b_0 = \bar{y} - b_1 \bar{x}
\]

\[
0 = \xi (y_i - (b_0 + b_1 x_i)) x_i \Rightarrow \xi x_i y_i - b_0 \xi x_i - b_1 \xi x_i^2 = 0
\]

\[
\frac{1}{N} \xi x_i y_i - b_0 \bar{x} - b_1 \frac{\xi x_i^2}{N} = 0
\]

\[
\frac{1}{N} \xi x_i y_i - (\bar{y} - b_1 \bar{x}) \bar{x} - b_0 \frac{\xi x_i^2}{N} = 0
\]

\[
\frac{1}{N} \xi x_i y_i - \bar{y} \bar{x} = b_1 \frac{\xi x_i^2}{N} - b_0 \bar{x}^2
\]

(Use lemma again)

\[
\frac{1}{N} \xi (x_i - \bar{x})(y_i - \bar{y}) = b_1 \left( \frac{\xi x_i^2}{N} - \bar{x}^2 \right) = b_1 \frac{\xi (x_i - \bar{x})^2}{N}
\]

\[
\Rightarrow b_1 = \frac{\xi (x_i - \bar{x})(y_i - \bar{y})}{\xi (x_i - \bar{x})^2}/N
\]
The OLS Estimator, Predicted Values, and Residuals

The OLS estimators of the slope $\beta_1$ and the intercept $\beta_0$ are:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{s_{XY}}{s_X^2} \tag{4.8}$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X}. \tag{4.9}$$

The OLS predicted values $\hat{Y}_i$ and residuals $\hat{u}_i$ are:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1X_i, \ i = 1, \ldots, n \tag{4.10}$$

$$\hat{u}_i = Y_i - \hat{Y}_i, \ i = 1, \ldots, n. \tag{4.11}$$

The estimated intercept ($\hat{\beta}_0$), slope ($\hat{\beta}_1$), and residual ($\hat{u}_i$) are computed from a sample of $n$ observations of $X_i$ and $Y_i, \ i = 1, \ldots, n$. These are estimates of the unknown true population intercept ($\beta_0$), slope ($\beta_1$), and error term ($u_i$).
The Least Squares Assumptions

\[ Y_i = \beta_0 + \beta_1 X_i + u_i, \ i = 1, \ldots, n, \] where:

1. The error term \( u_i \) has conditional mean zero given \( X_i \), that is, \( E(u_i | X_i) = 0 \);
2. \((X_i, Y_i), \ i = 1, \ldots, n\) are independent and identically distributed (i.i.d.) draws from their joint distribution; and
3. \((X_i, u_i)\) have nonzero finite fourth moments.
Next time..

- Confidence intervals for $\beta_0$ and $\beta_1$
- Examples