Linear Regression with One Regressor

Michael Ash

Lecture 8
Linear Regression with One Regressor
a.k.a. Bivariate Regression

An important distinction

1. The Model
   - Effect of a one-unit change in $X$ on the mean of $Y$.
   - Role of randomness and idiosyncratic variation.
   - Mean caveats apply.

2. Estimating the Model
   - Computing sample coefficients with Ordinary Least Squares
   - Hypothesis testing
The Linear Regression Model
Slope, or response

\[ \beta_{\text{ClassSize}} = \frac{\text{Change in TestScore}}{\text{Change in ClassSize}} = \frac{\Delta \text{TestScore}}{\Delta \text{ClassSize}} \]

The Greek capital letter delta \( \Delta \) stands for “change in.” The Greek lower-case letter beta \( \beta \) is the symbol for how \( Y \) (TestScore) responds to a change in \( X \) (ClassSize). In this case, \( \beta \) is measured in test points per student.

Other examples: how murder rates (\( Y \)) respond to poverty (\( X \)); how highway deaths (\( Y \)) respond to drunk-driving penalties (\( X \)); how earnings (\( Y \)) respond to years of schooling (\( X \)). Consider the \( \beta \) in each case.
Using a known $\beta$

Suppose we know that $\beta = -0.6$ test points per student. (Adding one student to the class reduces the class test score by 0.6 points.)

What is the effect of reducing class size by two students?

Rearranging the definition of $\beta$, and then putting in the particular example.

\[
\Delta \text{TestScore} = \beta_{\text{ClassSize}} \times \Delta \text{ClassSize}
\]
\[
= (-0.6) \times (-2) = +1.2
\]

In words:
Reducing class size by two students will raise test scores by 1.2 points.
Building a Model

\[ \text{TestScore} = \beta_0 + \beta_{\text{ClassSize}} \times \text{ClassSize} \]

is a statement about relationship that holds \textbf{on average} across the population of districts.

\[ \text{TestScore} = \beta_0 + \beta_{\text{ClassSize}} \times \text{ClassSize} + \text{other factors} \]

is a statement that is true for any district.

- \( \beta_0 + \beta_{\text{ClassSize}} \times \text{ClassSize} \) is the average effect of class size.
- \textbf{other factors} includes teacher quality, textbook quality, community income or wealth, native English speakers, testing variation, luck.
Linear Regression Model with a Single Regressor

\[ Y_i = \beta_0 + \beta_1 X_i + u_i \]

- \( Y \) is the **dependent variable**, or outcome variable, or left-hand variable. (No one says regressand with a straight face.)
- \( X \) is the **independent variable**, or regressor, or explanatory variable, or right-hand variable.
- \( \beta_0 + \beta_1 X \) is the **population regression line**, or the expected value (mean) of \( Y \) given \( X \), or \( E(Y|X) = \beta_0 + \beta_1 X \)
Linear Regression Model with a Single Regressor

\[ Y_i = \beta_0 + \beta_1 X_i + u_i \]

- \( \beta_1 \) and \( \beta_0 \) are the **coefficients**, or parameters of the regression line.
  - \( \beta_1 \) is the **slope**, the change in \( Y \) associated with a unit change in \( X \).
  - \( \beta_0 \) is the **intercept**, the expected value of \( Y \) when \( X = 0 \).
    (Sometimes \( X = 0 \) doesn’t make any sense.) \( \beta_0 \) raises or lowers the regression line.

- \( u_i \) is the **error term** or **residual**, which includes all of the unique, or idiosyncratic features of observation \( i \), including randomness, measurement error, and luck that affect its outcome \( Y_i \).
Determinism and Randomness

Appreciate determinism

\[ E(Y_i|X_i) = \beta_0 + \beta_1 X_i \]

Appreciate randomness

\[ u_i \]

- Figure 4.1
- Better or worse than predicted
- Determinism and randomness
Draw a best line through a scatterplot.

Figure 4.2

- Choosing $\hat{\beta}_0$ and $\hat{\beta}_1$ defines a line.
- What is the best line?

Recall that the sample mean $\overline{Y}$ minimizes

$$\sum_{i=1}^{n} (Y_i - \overline{m})^2$$
Now instead of getting to choose \( m \), we get to choose \( b_0 \) and \( b_1 \) to minimize

\[
\sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i)^2
\]

Let’s focus on the key term:

\[
(Y_i - b_0 - b_1 X_i) = (Y_i - (b_0 + b_1 X_i)) = (Y_i - E(Y_i|X_i))
\]
Estimating the Coefficients of the Linear Regression Model

\((Y_i - (b_0 + b_1 X_i))\)

The key term expresses how far the actual value of \(Y_i\) is from the expected value of \(Y_i\) given \(X_i\) according to the proposed line. We want to choose \(b_0\) and \(b_1\) to keep these gaps down. The values of \(b_0\) and \(b_1\) that keep

\[
\sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i)^2
\]

as low as possible are called the **ordinary least squares** estimators of \(\beta_0\) and \(\beta_1\). The estimators are named \(\hat{\beta}_0\) and \(\hat{\beta}_1\).
Choose $b_0$ and $b_1$ to minimize $\sum_{i=1}^{n}(Y_i - b_0 - b_1X_i)^2$

1. Could do this by trial and error. (Choose many alternative pairs $b_0$ and $b_1$ and see which gives the smallest sum of squared errors.)

2. Calculus creates simple formulas:

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

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

3. These are an average concept (again!) with all the good properties of sample averages.
Some OLS terminology

- OLS Regression Line
  \[ \hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X \]

- Predicted Value of \( Y_i \) given \( X_i \)
  \[ \hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i \]

- Predicted residual of the \( i^{\text{th}} \) observation
  \[ \hat{u}_i = Y_i - \hat{Y}_i \]
Test Scores and Student-Teacher Ratio

Unit of observation: a California school district \((n = 420)\)
Variables: district average test score, district student-teacher ratio

OLS Results

\[
\begin{align*}
\hat{\beta}_0 &= 698.9 \\
\hat{\beta}_1 &= -2.28
\end{align*}
\]

\[
E(\text{TestScore}) = \beta_0 + \beta_{\text{STR}} \times \text{STR}
\]

\[
\text{TestScore} = 698.9 - 2.28 \times \text{STR}
\]

The slope is \(-2.28\): an increase in the student-teacher ratio (STR) by one student per class is, on average, associated with a decline in the districtwide test score by 2.28 points on the test.
Are the results big?
Consider reducing STR by two students

What would we expect to happen in a district with median student-to-teacher ratio and median test score? (No such district necessarily exists, but it’s a useful reference point.)

Table 4.1

- Reduction of 2 students: from 19.7 (50th percentile) to 17.7 (c. 10th percentile)
- Expected change in test scores: \(-2.28 \times -2 = +4.6\)
- Expected change in test scores: from 654.5 (50th percentile) to 659.1 (c. 60th percentile)?
- Worth it?

- Consider reducing STR by five students
- Beware out-of-sample predictions
Advantages of OLS Estimators

- Widely used method in social sciences, policy, and administration.

Desirable Properties of OLS

1. $\hat{\beta}$ consistent and unbiased estimator of $\beta$
2. $\hat{\beta}$ is approximately normally distributed in large samples
3. With additional assumptions, $\hat{\beta}$ may be the smallest variance estimator of $\beta$. 