Linear Regression with One Regressor

Michael Ash

Lecture 9
Review of Last Time

1. The Linear Regression Model
   - The relationship between independent $X$ and dependent $Y$ is modeled as a straight line (the regression line) with slope ($\beta_1$) and intercept ($\beta_0$)
   - Every datapoint $i$ is above or below the line by an idiosyncratic amount $u_i$

2. Estimating the Model
   - Ordinary Least Squares chooses $\hat{\beta}_0$ and $\hat{\beta}_1$ to best fit the data
The Least Squares Assumptions

The point of the assumptions

- The OLS estimators are unbiased and consistent.
- The OLS estimators are normal in large samples with formulas for standard error to be given.
The Conditional Distribution of $u_i$ given $X_i$ Has a Mean of Zero: $E(u_i | X_i) = 0$

- The regression line needs to be right on average (Figure 4.4). The $u_i$ must be scattered above and below the regression line regardless of the value of $X_i$.
- Failure of this assumption is disastrous. If the regression line is not right, on average, then the estimated slope and intercept will be biased.
- Example: suppose that “other factors” are always bad (or good) in small (or large) classrooms.
- If you knew whether a point would be above the regression line or below the regression line, then you should draw a different regression line.
The Conditional Distribution of $u_i$ given $X_i$ Has a Mean of Zero: $E(u_i|X_i) = 0$

Continued

- $E(u_i|X_i) = 0$ (which is stronger) implies $\text{corr}(X_i, u_i) = 0$ (but not vice versa). So $\text{corr}(X_i, u_i) = 0$ is a necessary but not sufficient condition for $E(u_i|X_i) = 0$.
- The assumption concerns the unknowable value $u_i$, not the computed value $\hat{u}_i$.
  - You cannot test the assumption by checking if $\text{corr}(X_i, \hat{u}_i) = 0$.
  - OLS will draw a line that makes $\text{corr}(X_i, \hat{u}_i) = 0$ look true even if it is not true.
#2 \((X_i, Y_i), \ i = 1, \ldots, n\) Are Independently and Identically Distributed

Usually easy in cross-sectional random samples. The age and earnings of worker \(i\), \((X_i, Y_i)\) in the sample are independent of the age and earnings of worker \(j\), \((X_j, Y_j)\).

Points of concern

- Classical experiments: the experimenter chooses every \(X_i\). Illustrates the importance of randomization in experimental design.
- Stratified sampling draws clustered observations, e.g., workers from the same household.
- Time series: observations that are close together in time likely share common components (or one observation may be a reaction to its predecessor).
We cannot have observations with extremely large values of either $X_i$ or $u_i$.

By **squaring** the gap between each point and the regression line, Ordinary Least **Squares** puts extra weight on outlying data. Note: if the values of $\{u_1, u_2, u_3\} = \{-3, 1, 2\}$, then note that $\{u_1^2, u_2^2, u_3^2\} = \{9, 1, 4\}$, which gives extra weight to the $u_1$ in shaping the regression line.

Regression analysis may not be appropriate for populations that have enormous outliers in either direction.
Sampling Distribution of OLS Estimators

Analogy to the mean

- Population parameters are true, fixed, and unknowable. Estimates of parameters vary because of random sampling, but they are knowable.
- The population coefficients $\beta_0$ and $\beta_1$ are estimated from a sample randomly drawn from the population.
- If we had, by chance, a different sample, we would have slightly different estimates of $\beta_0$ and $\beta_1$. 
Sampling distribution of $\hat{\beta}_0$ and $\hat{\beta}_1$

Because OLS estimation is similar to computing a sample mean, in large samples, Law of Large Numbers and Central Limit Theorem results apply.

- $\hat{\beta}_0$ and $\hat{\beta}_1$ are unbiased estimators of $\beta_0$ and $\beta_1$.

\[
E(\hat{\beta}_0) = \beta_0 \\
E(\hat{\beta}_1) = \beta_1
\]

- $\hat{\beta}_0$ and $\hat{\beta}_1$ are normally distributed around the true values $\beta_0$ and $\beta_1$ with variances $\sigma^2_{\hat{\beta}_0}$ and $\sigma^2_{\hat{\beta}_1}$

\[
\hat{\beta}_1 \sim N(\beta_1, \sigma^2_{\hat{\beta}_1}) \\
\sigma^2_{\hat{\beta}_1} = \frac{1}{n} \frac{\text{var}[(X_i - \mu_X)u_i]}{[\text{var}(X_i)]^2} \\
\hat{\beta}_0 \sim N(\beta_0, \sigma^2_{\hat{\beta}_0}) \\
\sigma^2_{\hat{\beta}_0} = \text{not shown. See Key Concept 4.4}
\]
Summary

- $\beta_0$ and $\beta_1$ are true, fixed population parameters. They do not have distributions.
- $\hat{\beta}_0$ and $\hat{\beta}_1$ are estimators with sampling distributions. The sampling distribution arises because the estimates are computed from sample data.
- $\sigma^2_{\hat{\beta}_0}$ and $\sigma^2_{\hat{\beta}_1}$ describe the spread in the distribution of $\hat{\beta}_0$ and $\hat{\beta}_1$.
- $\sigma^2_{\hat{\beta}_0}$ and $\sigma^2_{\hat{\beta}_1}$ are analogous to the variance of the sample mean (the square of the standard deviation of the sample mean). Bigger values of $\sigma^2_{\hat{\beta}_0}$ and $\sigma^2_{\hat{\beta}_1}$ imply less precision in the estimates of $\hat{\beta}_0$ and $\hat{\beta}_1$.
- We will use $\sigma^2_{\hat{\beta}_0}$ and $\sigma^2_{\hat{\beta}_1}$ to test hypotheses about and make confidence intervals for $\beta_0$ and $\beta_1$. 