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

15 Lecture 15

15.1 Moving from Description to Inference

- As I've said (over and over), all we've done so far is talked about the regression line $\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x$ as a description of how well the relationship between x and y can be approximated by a linear relationship. In this context, $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are simply descriptive statistics, like the empirical mean or empirical variance of the observed values of a variable.

- Now we move to using the formula for the least squares line to make **inferences** about an underlying population from the sample under analysis. Just as we used the statistic $\overline{Y}$ to make inferences about the parameter μ, we will now use the statistics $\widehat{\beta}_1 = \frac{cov(x,y)}{var(x)}$ and $\widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1 \overline{x}$ to make inferences about theparameters β_0 and β_1.

- As usual, we would like to find unbiased, relatively low-variance estimators of these parameters. As it turns out, the formulae for $\widehat{\beta}_0$ and $\widehat{\beta}_1$ that we developed earlier generate unbiased estimates of the parameters β_0 and β_1 as long as certain assumptions hold. We will now develop those assumptions. Later, we will introduce the additional assumptions necessary to show that $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are not only unbiased, but that have smaller variances than any other possible linear unbiased estimator of β_0 and β_1.

- **Assumption 1: The relationship between x and y in the population is linear in its parameters, and it is probabilistic.**

  - We begin with the assumption of a linear relationship between x and y. This not only rules out curvilinear relationships (draw). But (perhaps more importantly), it rules out relationships in which there are diminishing returns to x (draw) or in which the effects of x are smaller at its extreme values (draw sigmoidal function). (Ask for examples of these relationships.)
  - In many cases, these alternate functional forms describe the relationship between x and y with much more verisimilitude than a linear functional form. As we will see, there

are all kinds of ways to account for these nonlinearities; but for now we are stuck in linearities: we assume that the change in y associated with a one-unit change in x is the same across the entire range of x.

- Furthermore, we assume that the linear relationship between x and y is not deterministic – that is, it is not always the case that a particular $y_i = \beta_0 + \beta_1 x_i$. Rather, we assume that the linear relationship is **probabilistic**, and therefore write the complete population model as

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

and we will say that the relationship between any x_i and y_i in the population may be written

$$y_i = \beta_0 + \beta_1 x_i + u_i.$$

Here, u stands for "unexplained," and in this class, we will call u the **error**. For a unit i in the population, u_i is the amount by which the sum $\beta_0 + \beta_1 x_i$ comes in less than y_i. It is thus the variation in y which is unexplained by x.

- We view y, x and u as random variables when stating this model.
- Note the difference between u_i and $\widehat{u}_i$:
  - u_i is the *error* associated with a hypothetical unit in the population. It is never observed.
  - $\widehat{u}_i$ is the *residual* associated with an observed unit in our sample once we've estimated a regression line.
- Note that I have removed the "hats," because we are now talking about β_0 and β_1 as parameters of interest. β_0 and β_1 are unknown and never directly observed.
- **Assumption 2: we have a random sample of (x, y) pairs from the population, making them i.i.d.**
  - Although this assumption can often fail to hold in practice, let's note it for now and move on. Note that this is no stronger of an assumption than we needed in order to say that univariate estimators (such as Y-bar) are unbiased.

- **Assumption 3: The variance of our observed $\mathbf{x}'s$ is nonzero.**
  - Happily, this is a weak assumption and can be easily confirmed by examining our sample.
- **Assumption 4: The error u has the expected value of zero, no matter what the value of x. That is, $E(u|x) = 0$.**
  - This is known as the assumption of "zero conditional mean." It is a statement about the unexplained factors contained in u_i, and it asserts that these other factors are unrelated to x_i in that, given a value of x_i, the mean of the distribution of these other factors equals zero.
  - This is a very strong assumption. It means that these other factors are uncorrelated with x. It practical terms, it means that there are no confounds that could render the relationship between x and y spurious.
    * Example from y = income and x = education.
    * When we write the population model

$$\text{income} = \beta_0 + \beta_1 \text{education} + u,$$

    * we are assuming that

$$\begin{aligned} E(u|\text{education}) &= 0, \text{ or more to he point} \\ cov(\text{education}, u) &= 0. \end{aligned}$$

    * But can we think about a factor that winds up in u (that is, the factor helps to explain income) but is correlated with education? Three prominent examples include: parents' education, ability, motivation. To the extent that these factors explain y and are correlated with x, Assumption 4 does not hold.
    * We'll talk about this in detail when we move to multivariate regression.
    * BTW, can we test this assumption by seeing if $corr(x_i, \widehat{u}_i) = 0$?

- No: the assumption is about errors in the population, not the residuals in our sample. (What conclusion can we draw if we observe $corr(x_i, \widehat{u}_i) = 0$? (Nothing! This is by construction in OLS!)

– The assumption $E(u|x) = 0$ means that in the statistical derivations we are about to do, we will treat x as fixed. Technically, this isn't true as typically we are working with a random sample of (x, y) pairs. But nothing is lost in the derivations by treating the x as nonrandom. This allows us to write that $E(x) = x$.

15.2 The Unbiasedness of the OLS estimator for β_1

- These four assumptions together allow us to say that $\widehat{\beta}_0$ and $\widehat{\beta}_1$ that we developed earlier generate unbiased estimates of the parameters β_0 and β_1. In this context, we call $\widehat{\beta}_0$ and $\widehat{\beta}_1$ the **ordinary least squares (OLS)** estimators of β_0 and β_1.

- Recall the formula for $\widehat{\beta}_1$:

$$\widehat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{\sum (x_i - \overline{x})(y_i - \overline{y})}{\sum (x_i - \overline{x})^2}$$

- In order for $\widehat{\beta}_1$ to be defined, we need Assumption 3: $var(x) > 0$.

- Note that we can rewrite

$$\begin{aligned}
\sum (x_i - \overline{x})(y_i - \overline{y}) &= \sum (x_i - \overline{x}) y_i - \sum (x_i - \overline{x}) \overline{y} \\
&= \sum (x_i - \overline{x}) y_i - \left[\sum x_i \overline{y} - \sum \overline{xy}\right] \\
&= \sum (x_i - \overline{x}) y_i - (n\overline{xy} - n\overline{xy}) \\
&= \sum (x_i - \overline{x}) y_i
\end{aligned}$$

- So

$$
\begin{aligned}
\widehat{\beta}_1 &= \frac{\sum (x_i - \overline{x})\, y_i}{\sum (x_i - \overline{x})^2} \\
&= \frac{\sum (x_i - \overline{x})\,(\beta_0 + \beta_1 x_i + u_i)}{SST_x} \\
&= \frac{\beta_0 \sum (x_i - \overline{x}) + \beta_1 \sum (x_i - \overline{x})\, x_i + \sum (x_i - \overline{x})\, u_i}{SST_x}
\end{aligned}
$$

- Note that $\sum (x_i - \overline{x}) = 0$, and

$$
\begin{aligned}
\sum (x_i - \overline{x})\, x_i &= \sum \left(x_i^2 - \overline{x} x_i\right) \\
&= \sum x_i^2 - \overline{x} \sum x_i \\
&= \sum x_i^2 - n\,(\overline{x})^2 \\
&= \sum x_i^2 - 2n\,(\overline{x})^2 + n\,(\overline{x})^2 \\
&= \sum x_i^2 - 2\overline{x} \sum x_i + \sum (\overline{x})^2 \;\; [\text{since } n\overline{x} = \textstyle\sum x_i, \text{ and } n\,(\overline{x})^2 = \textstyle\sum (\overline{x})^2\,] \\
&= \sum x_i^2 - 2\overline{x} x_i + (\overline{x})^2 \\
&= \sum (x_i - \overline{x})^2 = SST_x.
\end{aligned}
$$

- So:

$$
\begin{aligned}
\widehat{\beta}_1 &= \frac{\beta_1 SST_x + \sum (x_i - \overline{x})\, u_i}{SST_x} \\
&= \beta_1 + \frac{\sum (x_i - \overline{x})\, u_i}{SST_x}
\end{aligned}
$$

- Now let's find the expected value of $\widehat{\beta}_1$:

$$\begin{aligned}
E\left(\widehat{\beta}_1\right) &= E\left[\beta_1 + \frac{\sum (x_i - \overline{x}) u_i}{SST_x}\right] \\
\text{Employing Assumptions 2 and 4, we can proceed} &: \\
&= \beta_1 + \frac{1}{SST_x} E\left[\sum (x_i - \overline{x}) u_i\right] \\
&= \beta_1 + \frac{1}{SST_x} \sum (x_i - \overline{x}) E(u_i) \\
&= \beta_1 + \frac{1}{SST_x} \sum (x_i - \overline{x}) 0 \\
&= \beta_1.
\end{aligned}$$

- And for $\widehat{\beta}_0$:

$$\widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1 \overline{x}$$

$$\begin{aligned}
\text{Since } y_i &= \beta_0 + \beta_1 x_i + u_i, \text{ then } \overline{y} = \beta_0 + \beta_1 \overline{x} + \overline{u}, \text{ substituting:} \\
\widehat{\beta}_0 &= \beta_0 + \beta_1 \overline{x} + \overline{u} - \widehat{\beta}_1 \overline{x} \\
\widehat{\beta}_0 &= \beta_0 + \left(\beta_1 - \widehat{\beta}_1\right) \overline{x} + \overline{u} \\
E\left(\widehat{\beta}_0\right) &= E\left[\beta_0 + \left(\beta_1 - \widehat{\beta}_1\right) \overline{x} + \overline{u}\right] \text{ Again using Assumptions 2 and 4, we proceed:} \\
&= \beta_0 + \left[\beta_1 - E\left(\widehat{\beta}_1\right)\right] \overline{x} + E(\overline{u}) \\
&= \beta_0 + [\beta_1 - \beta_1] \overline{x} + E(\overline{u}) \\
&= \beta_0
\end{aligned}$$

15.3 The variances of the OLS estimators

- Last time, we showed how four assumptions regarding x and y...

  1. Relationship is linear in its parameters
  2. x,y drawn from random sample, making them i.i.d.
  3. variance of x is nonzero
  4. $E(u|x) = 0$

- ..resulted in the least-squares solutions for $\widehat{\beta}_0$ and $\widehat{\beta}_1$ being unbiased estimators for β_0 and β_1.

- $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are statistics, just like (say) $\overline{Y}$. Now, just as we did when making inferences about univariate distributions, we can say something about not only the unbiasedness of our estimates but also how far we can expect them to be away from the true parameter on average. In the univariate context, we were interested in describing the sampling distribution of the statistic X-bar.

- We will do the same thing here and consider the sampling distribution of $\widehat{\beta}_0$ and $\widehat{\beta}_1$. We, of course, already know the means of these sampling distributions: they are β_0 and β_1. Let's now compute the variances of $\widehat{\beta}_0$ and $\widehat{\beta}_1$. To compute these , we make a fifth assumption [add to list under separate heading]:

- **Assumption 5:** $VAR(u|x) = \sigma^2$. That is, the error u has the same variance no matter what the value of x. This is also known as homoskedasticity. When the assumption is violated, we have "heteroskedasticity."

- Draw pictures on board: Wooldridge p. 54 and p.55 – to illustrate homoskedasticity and heteroskedasticity.

  - Note difference between $VAR(u|x) = \sigma^2$ and $E(u|x) = 0$.
  - We did not need Assumption 5 to establish the unbiasedness of the OLS estimators.

- Note that this now allows us to write:

$$\begin{aligned}
y &= \beta_0 + \beta_1 x + u \ \ \text{[the model]} \\
E(y|x) &= E\left[(\beta_0 + \beta_1 x + u)\,|x\right] \ \ \text{[taking expectations conditional on x]} \\
&= E\left(\beta_0|x\right) + E(\beta_1 x|x) + E(u|x) \\
E(y|x) &= \beta_0 + \beta_1 x + 0 \ \ \text{[thanks to assumption 4]}
\end{aligned}$$

- and

$$\begin{aligned} VAR(y|x) &= VAR\left[(\beta_0 + \beta_1 x + u)\,|x\right] \\ &= 0 + 0 + VAR(u|x) \\ &= \sigma^2. \text{ [by assumption 5]} \end{aligned}$$

- What is σ^2?
  - Ask class.
  - Note that it is not $VAR(y)$. It is $VAR(y|x)$.
  - Note that it is a parameter-not an estimate. So it's something in the population, not the sample.
  - σ^2 is a measure of the extent to which unexplained factors are affecting the value of y.
    * Assumption 4 means that we assume that these factors are unrelated to x.
    * Assumption 5 means that these factors are constant regardless of the value of x.
    * When σ^2 is bigger, it is the case that other factors explain a great deal of the variation in y in addition to x.
    * When σ^2 is smaller, it is the case that x is explaining a great deal of the variation in y on its own.
- Later, we'll show how to estimate the variances of the OLS estimators when this assumption is violated.
- Now, we're ready to derive the sampling variances of the OLS estimators. It is the case that:

$$\begin{aligned} VAR\left(\widehat{\beta}_1\right) &= \frac{\sigma^2}{SST_x}; \\ VAR\left(\widehat{\beta}_0\right) &= \frac{\sigma^2 \frac{\sum_i x_i^2}{n}}{SST_x}. \end{aligned}$$

- Proof:

- Recall that in the proof establishing the unbiasedness of $\widehat{\beta}_1$, we wrote

$$\widehat{\beta}_1 = \beta_1 + \frac{\sum (x_i - \overline{x}) u_i}{SST_x}.$$

- We then used the conditional mean assumption in order to say that the value of the last term of this expression is zero.
- Now let's consider

$$\begin{aligned}
VAR\left(\widehat{\beta}_1\right) &= VAR\left[\beta_1 + \frac{\sum (x_i - \overline{x}) u_i}{SST_x}\right]. \\
&= 0 + \frac{1}{(SST_x)^2} VAR\left[\sum (x_i - \overline{x}) u_i\right] \quad \text{[fixed X means we can treat as constant]} \\
&= \frac{1}{(SST_x)^2} \sum (x_i - \overline{x})^2 VAR(u_i) \quad \text{[again]} \\
&= \frac{1}{(SST_x)^2} \sum (x_i - \overline{x})^2 \sigma^2 \quad \text{[by Assumption 5]} \\
&= \frac{SST_x}{(SST_x)^2} \sigma^2 = \frac{\sigma^2}{SST_x}.
\end{aligned}$$

- I'll leave the proof of the variance of $\widehat{\beta}_0$ as an exercise.
- Let's have a look at $VAR\left(\widehat{\beta}_1\right)$. We'd obviously like this to be as (what?) small as possible. What are the two quantities that make it small?
  - σ^2 : as gets smaller, $VAR\left(\widehat{\beta}_1\right)$ gets smaller.
  - SST_x : as gets bigger, $VAR\left(\widehat{\beta}_1\right)$ gets smaller.
    * Now let's take a closer look at SST_x :

$$\begin{aligned}
SST_x &= \sum (x_i - \overline{x})^2 \\
\frac{SST_x}{n} &= \frac{\sum (x_i - \overline{x})^2}{n} \\
\frac{SST_x}{n} &= var(x) \\
SST_x &= n \cdot var(x)
\end{aligned}$$

        * So

$$VAR\left(\widehat{\beta}_1\right) = \frac{\sigma^2}{SST_x} = \frac{\sigma^2}{n \cdot var(x)}.$$

- Of these three properties, what can we really do anything about?
- In most cases, only n:
  - σ^2 is a parameter; it's declines only to the extent that x explains y well.
  - $var(x)$ is the empirical variance of x in our sample. In a random sample, it will of course look like the variance of x in the population. Not a whole lot we can do about that, either.
  - So what does this say about the relationship between sample size and the variance of $\widehat{\beta}_1$?
- Remember $VAR(\overline{Y}) = \frac{\sigma^2}{n}$? Compare to $\frac{\sigma^2}{n \cdot var(x)}$. Again, we have a ratio of something we can't control (the variance of y) over something we can (n).