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

20 Lecture 20

20.1 BLUE

- Go over handout.

- Furthermore, if we add the (recall, troubling) assumption that the population errors u are distributed Normal, then $\widehat{\beta}$ is not only the BLUE of β, it is also the **minimum variance unbiased estimator (MVUE)** of β. That is, no other unbiased estimator of β exists–whether linear or not–that has a lower variance (i.e., is more efficient).

20.2 Interpreting Categorical Dummy Variables

- Go over handout.

20.3 Interaction terms

- When the effect of one variable x1 changes the effect of another x2 on y, we say that an interaction effect exists between x1 and x2.

- We model an interaction effect by creating a new variable that is the product of x1 and x2 and including it in our equation.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 (x_1 \cdot x_2) + u$$

- Note that

$$\frac{\partial y}{\partial x_1} = \beta_1 + \beta_3 x_2$$

  So that the slope of the line describing the relationship between each x and y varies by the value of the other x. Consider the cases where we estimate:

$$y = 1 + 3x_1 + 2x_2 + 4(x_1 \cdot x_2) + u$$

- Here's what y looks like at three different values of x2 (2, 5, and 10):itbpF2.8245in2.0712in0inFigure
- In this case, increasing values of x2 *amplify* the effect of x1 on y, increasing the magnitude of the slope in the signed direction.
- Now consider

$$y = 1 + 3x_1 + 2x_2 - 4\left(x_1 \cdot x_2\right) + u$$

- Here's what y looks like at three different values of x2 (2, 5, and 10):

  itbpF3.5258in2.5849in0inFigure
- In this case, increasing values of x2 *dampen* the effect of x1 on y, increasing the magnitude of the slope in the signed direction.
- To get a sense of interaction effects, you generally need to plot predicted probabilities of y by holding the value of one of the x's constant while varying the other value. You should label each plot accordingly.
- Statistical software programs can't tell the difference between constituitive terms and interaction terms, and so they blindly spit back incorrect standard errors. We have calculated

$$\frac{\partial y}{\partial x_1} = \widehat{\beta}_1 + \widehat{\beta}_3 x_2$$

  and so we are interested in

$$\widehat{\sigma}_{\frac{\partial y}{\partial x_1}} = \sqrt{var\left(\widehat{\beta}_1 + \widehat{\beta}_3 x_2\right)} = \sqrt{var\left(\widehat{\beta}_1\right) + x_2^2 var\left(\widehat{\beta}_3\right) + 2x_2 cov\left(\widehat{\beta}_1, \widehat{\beta}_3\right)}$$

- Why? Recall that $var\left(x + ay\right) = var(x) + a^2 var(y) + 2acov(x, y)$.
- Note how the variance is non constant across values of x_2.
- Not easily calculated, but is definitely available from Stata (go over handout).
- What do we do with $\widehat{\sigma}_{\frac{\partial y}{\partial x_1}}$? We typically use it to see at what values of x_2 the variable x_1 has a non-zero ceteris paribus effect on y. That is, since $T = \frac{\widehat{\beta}_1 + \widehat{\beta}_3 x_2 - 0}{\sqrt{var\left(\widehat{\beta}_1 + \widehat{\beta}_3 x_2\right)}}$ is distributed t with

$N-K-1$ degrees of freedom, we can run the usual significance tests across the entire range of x_2. The values of $x_2$for which T surpasses the significance threshold is where x_1 has a significant effect on y. This may be all, some or only a range of the values of x_2.

20.4 Partialling out

- Here's another way to calculate the OLS estimator of, say, β_1 :

$$\widehat{\beta}_1 = \frac{\sum \widehat{u}_{i1} y_i}{\sum \left(\widehat{u}_{i1}\right)^2} = \frac{cov\left(\widehat{u}_{i1}, y_i\right)}{var\left(\widehat{u}_{i1}\right)},$$

  where $\widehat{u}_{i1}$ are the residuals from a regression of x_1 on all the other x's in the model–that is the variation in x_1 that is not explained by a linear combination of the other x's.

- So one way to think about this estimate is that its numerator is the proportion of this unexplained variation in x_1 that covaries with y.

20.5 Hypotheses about Parameters

- So far, we have focused on hypothesis tests about one parameter (a $\widehat{\beta}_j$) at a time. But there are instances in which you want to test hypotheses involving more than one parameter. Your book has the example where researchers are interested whether the effect on income of an additional year of education at a junior college is as much as the effect of an additional year of education at four-year university. The idea here is that jc's are lower status in the U.S. than universities, so maybe employers value these years of education less. (A complementary hypothesis would be that a jc education may be of lower quality.) The model assumed is

$$\log(wage) = \beta_0 + \beta_1 jc + \beta_2 univ + \beta_3 work + u,$$

  where

$$\begin{aligned} jc &= \text{\# years attending a junior college} \\ univ &= \text{\# years attending a university} \\ work &= \text{\# months in the workforce.} \end{aligned}$$

- If we are interested in whether there is a difference in return to education from junior colleges and universities, what is an appopriate null hypothesis? It's

$$H_0 : \beta_1 = \beta_2.$$

- And an appropriate alternative is

$$H_1 : \beta_1 < \beta_2.$$

- Is a one-sided test appropriate here? Yes: theory justifies this hypothesis.

- So in the case of, say whether two groups have different means, what kind of tests did we run? (Encourage class to come up with them.)

- Rewrite null and alternative as

$$\begin{aligned} H_0 &: \beta_1 - \beta_2 = 0 \\ H_1 &: \beta_1 - \beta_2 < 0 \end{aligned}$$

- We are interested in hypotheses about the quantity $\beta_1 - \beta_2$. The statistic

$$\frac{\widehat{\beta}_1 - \widehat{\beta}_2}{se(\widehat{\beta}_1 - \widehat{\beta}_2)}$$

  is distributed t.

- But how do we get $se(\widehat{\beta}_1 - \widehat{\beta}_2)$? Well,

$$\begin{aligned} se(\widehat{\beta}_1 - \widehat{\beta}_2) &= \sqrt{var(\widehat{\beta}_1 - \widehat{\beta}_2)}, \text{ and} \\ var(\widehat{\beta}_1 - \widehat{\beta}_2) &= var\left(\widehat{\beta}_1\right) + var\left(\widehat{\beta}_2\right) - 2cov\left(\widehat{\beta}_1, \widehat{\beta}_3\right). \text{ So} \\ se(\widehat{\beta}_1 - \widehat{\beta}_2) &= \sqrt{var\left(\widehat{\beta}_1\right) + var\left(\widehat{\beta}_2\right) - 2cov\left(\widehat{\beta}_1, \widehat{\beta}_3\right)} \end{aligned}$$

- Here we reject $H_0 : \beta_1 = \beta_2$ if $\widehat{\beta}_1 - \widehat{\beta}_2 + t_{crit}\left[se(\widehat{\beta}_1 - \widehat{\beta}_2)\right] < 0$.

- We can pull these from the variance-covariance matrix of the estimated betas as we did when estimating the standard errors associated with interaction effects (an example in a minute).
- But there is a much easier way to do this, as described on page 142 of your text. We care about $\beta_1 - \beta_2$, so let's call this a parameter, $\theta = \beta_1 - \beta_2$, and thus $\beta_1 = \theta + \beta_2$. Now

$$\begin{aligned}\log(wage) &= \beta_0 + \beta_1 jc + \beta_2 univ + \beta_3 work + u \text{ becomes} \\ &= \beta_0 + (\theta + \beta_2)\, jc + \beta_2 univ + \beta_3 work + u \\ &= \beta_0 + \theta jc + \beta_2\,(univ + jc) + \beta_3 work + u.\end{aligned}$$

- So if we create a new variable, $univ + jc$, and run the regression

```
reg lnwage jc univplusjc work
```

- the coefficient on jc will be the parameter we care about, and its standard error will be exactly that calculated by Stata.
- Go over example from handout.

20.6 Multiple Linear Restrictions

- The tests we've described so far are about what we call single *restrictions.* That is, we are testing whether the data justify rejecting a hypothesized restriction that $\beta_k = 0$ (in the single parameter case) or β_k is equal to, greater than, or less than some other β_j.
- But there are times when one will wish to conduct tests with multiple linear restrictions, as well. The most common such test is whether a group of variables has no effect on the dependent variable, y.
- Write the *unrestricted* model as

$$y = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k + u \qquad \text{(UR)}$$

- The number of variables we decide to restrict as equal to zero is q, and for convenience we assume that the restricted variables are included in the model after the unrestricted variables, then we can state the null as

$$H_0 : \beta_{k-q+1} = \beta_{k-q+2} = ... = \beta_k = 0$$

- Thus the *restricted* model is

$$y = \beta_0 + \beta_1 x_1 + ... + \beta_{k-q} x_{k-q} + u \qquad \text{(R)}$$

- If we define the F-statistic as

$$F \equiv \frac{(SSR_r - SSR_{ur})/q}{SSR_{ur}/(n-k-1)},$$

- it has an F-distribution with q "numerator d.f." and $n-k-1$ "denominator d.f."

- Note that because $SSR_r \geq SSR_{ur}$ (since ur has more variables than r), F is always non-negative. And what we're really testing here is whether the explanatory power of ur is significantly greater than r. Under the null,

$$F \tilde{} F_{q,n-k-1}$$

- The F-distribution is the ratio of two independent chi-square random variables, divided by their respective degrees of freedom. (Recall that a chi-square is the sum of the squares of independent standard normal RVs, which is what the SSRs are.)

- Like any distribution, we can use the F's density to determine the likelihood that we'd get the F-statistic we see due to chance variation in our data. We reject the null if it's extremely unlikely that we'd obtain the F-statistic by chance.

- If H_0 is rejected, we say that the excluded variables are **jointly significant**. If the null is not rejected, then we say that they are **jointly insignificant**. It's also common to report the p-value associated with an F-test.

- A tidbit: the F statistic obtained by testing the exclusion of a single variable is equal to the square of the the t-statistic obtained on its coefficient via OLS:

$$F_{\text{exclude } \widehat{\beta}_k} = \left(\frac{\widehat{\beta}_k}{\widehat{\sigma}_{\widehat{\beta}_k}} \right)^2$$

- Note that Stata's regression output includes a statistic it calls F, along with "Prob ¿F." These are the (hardly ever used) F-statistic and p-value associated with the test that all coefficients associated with the variables you've included in your regression are zero.

20.7 Quadratics

- When the relationship is curvilinear between a Y and an X, we create the square of X, and include it in our model:
- (see "handout on polynomials")