19 Lecture 19

19.1 The sampling distribution of $\widehat{\beta}$

- Be sure to talk about the interpretation of a t-statistic:
  - Hypothesized mean of zero;
  - two-tailed tests
  - asterisks with p-values.
- There may be times when we want to know something different than whether some β is equal to zero; perhaps $\beta_1 = 1$ in the model [NEXT YEAR, NEW EXAMPLE. THIS ONE DOESN'T QUITE WORK; or elaborate.

$$FT\_othergroup = \beta_0 + \beta_1 \left(FT\_owngroup\right) + \mathbf{Z}\boldsymbol{\beta} + u,$$

where $\mathbf{Z}$ is a matrix of covariates. Here,

$$\begin{aligned} H_0 &: \beta = 1; H_A : \beta \neq 1. \\ \text{test statistic is } t &= \frac{\widehat{\beta} - 1}{\widehat{se\left(\widehat{\beta}\right)}}. \end{aligned}$$

- Note that you can find $\widehat{se\left(\widehat{\beta}\right)}$ by looking at the appropriate diagonal entry of the variance-covariance matrix of the vector $\widehat{\beta}$. It contains $\widehat{var\left(\widehat{\beta}\right)}$, and so $\widehat{se\left(\widehat{\beta}\right)}$ is the square root of this entry.

19.2 Interpreting an OLS regression equation

- Consider the estimated equation

$$\widehat{income} = -17,431 + 2,708(educyears) + 1,050(income16)$$

- Here, the estimates $\widehat{\beta}_1 = 2,708$ and $\widehat{\beta}_2 = 1,050$ have partial effect, or ceteris paribus, interpretations. Note that

$$\frac{\partial \widehat{income}}{\partial educyears} = 2,708 \text{ and } \frac{\partial \widehat{income}}{\partial income16} = 1,050$$

- Thus we can write a statement like, "Holding family income at age 16 constant, each additional year of education is associated with an addition \$2,700 in income." Other ways to say this:

  - Holding family income *fixed*
  - Education has a ceteris paribus association of \$2,708 in income for each additional year of education

- Note that we can also use the equation to generate predictions about y for different values of x.

19.3 Goodness of Fit

- Just as in the simple regression case,

$$R^2 = \frac{\sum (\widehat{y}_i - \overline{y})^2}{\sum (y_i - \overline{y})^2} = \frac{SSE}{SST} = 1 - \frac{SSR}{SST}$$

- R^2 never decreases–and usually increases–whenever we add additional x's to the model by construction. This is because the denominator doesn't change, but the numerator usually increases with the addition of x's.

- For this reason, when comparing the goodness of fit statistics across models, it is better to compare their adjusted R-squared, which is calculated

$$R^2_{adj} = 1 - \left[\frac{\frac{SSR}{(n-k-1)}}{\frac{SST}{(n-1)}} \right] = 1 - \frac{\frac{\widehat{\sigma}^2_u}{(n-k-1)}}{\frac{SST}{(n-1)}}$$

- By construction, R^2_{adj} increases with the introduction of a new regressor into a model if and only if the t-statistic on the new variable's coefficient is greater than one in absolute value.

- Simple algebra gives
$$R^2_{adj} = 1 - \frac{\left(1 - R^2\right)(n-1)}{(n-k-1)}.$$

- When are R^2 and R^2_{adj} likely to be different? When are they likely to be similar?

19.4 Too many variables

- [for this and next subsection, draw familiar diagram of X Y Z on the board; now include W [here] and remove correlation between X and Z [below]

- we call a variable **irrelevant** if it has no partial effect on y in the population. that is, a variable W is irrelevant if:
$$\frac{\partial y}{\partial W} = 0$$

- the inclusion of an irrelevant variable in your model (aka "overspecifying the model") has no effect on the unbiasedness of the estimates of any of the betas, as it does not violate any of the assumptions 1 through 4.

- so if the true model is
$$y = \beta_0 + \beta_1 X + u$$
but you estimate
$$y = \beta_0 + \beta_1 X + \beta_2 W + u,$$
the estimates generated of all the betas (including β_2) will be unbiased: $E(\widehat{\beta}_0) = \beta_0$, etc.

- however, including an irrelvant variable in the model is harmful to the extent that this variable is collinear with other X's in the model. Recall that
$$Var\left(\widehat{\beta}_j\right) = \frac{\sigma^2}{n \cdot var(x_j) \cdot \left(1 - R_j^2\right)},$$
where R_j^2 is the R-squared obtained from regressing x_j on all other independent variables in the model. Well, to the extent that the irrelevant variable W covaries with x's included in

your model, the variances of the estimated coefficients associated with those x's will become inflated.

- The result is that your estimates become less efficient: that is, their statistical power decreases, which increases the likelihood of falsely accepting the null that $\beta_j = 0$.
- To assess the threat of multicollinearity, you can generate the R_j^2 yourself for each variable x_j.
- One way to think about the size of this threat is a quantity known as the variance inflation factor (VIF) associated with each of the x's in your model. It is calculated as

$$VIF\left(\widehat{\beta}_j\right) = \frac{1}{1 - R_j^2},$$

  VIF ranges between unity (when $R_j^2 = 0$) and approaches infinity as R_j^2 approaches 1.
- Now we can re-write $Var\left(\widehat{\beta}_j\right)$ as

$$Var\left(\widehat{\beta}_j\right) = \frac{\sigma^2}{n \cdot var(x_j)} VIF\left(\widehat{\beta}_j\right).$$

- And so VIF_j is the factor by which $Var\left(\widehat{\beta}_j\right)$ is increased due to the fact that x_j is correlated with the other x's in the model.

• When displaying results, it is not always the case that we should take variables that have zero effect on y out of a model. Sometimes we include a variable x_j that we know to have zero effect in order to show our readers that we have controlled for it, and thus we are certain that the estimates on the x's we care about are not confounded by x_j.

19.5 Correlated with y, but not with x

• Recalling that

$$Var\left(\widehat{\beta}_j\right) = \frac{\sigma^2}{n \cdot var(x_j) \cdot \left(1 - R_j^2\right)},$$

note that we will often want to include covariates Z that predict y in a model even if we don't think they are confounds with the x's we care about. That is, it can often be wise to include predictors Z where

$$\frac{\partial x}{\partial Z} = 0 \text{ but } \frac{\partial y}{\partial Z} \neq 0$$

- Why? To the extent that they help explain y, they improve the model's predictive power and thus lower σ^2, making all of our estimates more efficient, even estimates on predictors uncorrelated with these covariates.
  - Country/state/regional fixed effects–which we often include to control for potential confounders–can also be a good example of this.