Econ 103: Multiple Linear Regression II

Manu Navjeevan

UCLA

September 1, 2021

Content Outline

Advanced Inference

- Multiple Hypothesis Testing
- The F-Test

Model Specification

- Indicator Variables
- Omitted Variables Bias

Table of Contents

Advanced Inference: Multiple Hypothesis Testing

So far, we have gone over mainly what we call "single hypothesis testing." That is, our hypotheses have usually involve a single parameter that is being tested.

Advanced Inference: Multiple Hypothesis Testing

So far, we have gone over mainly what we call "single hypothesis testing." That is, our hypotheses have usually involve a single parameter that is being tested.

Examples:

- Testing the null hypothesis $H_0 : \beta_3 = 0$ against an alternative $H_1 : \beta_3 \neq 0$.
- Testing the null hypothesis $H_0 : \lambda \leq 0$ against an alternative $H_1 : \lambda > 0$, where $\lambda = \beta_2 + 3\beta_3$.

Advanced Inference: Multiple Hypothesis Testing

So far, we have gone over mainly what we call "single hypothesis testing." That is, our hypotheses have usually involve a single parameter that is being tested.

Examples:

- Testing the null hypothesis $H_0 : \beta_3 = 0$ against an alternative $H_1 : \beta_3 \neq 0$.
- Testing the null hypothesis $H_0 : \lambda \leq 0$ against an alternative $H_1 : \lambda > 0$, where $\lambda = \beta_2 + 3\beta_3$.
  - Notice that while λ is a linear combination of parameters, we are still only testing the linear combination, not the individual components.

Advanced Inference: Multiple Hypothesis Testing

So far, we have gone over mainly what we call "single hypothesis testing." That is, our hypotheses have usually involve a single parameter that is being tested.

Examples:

- Testing the null hypothesis $H_0 : \beta_3 = 0$ against an alternative $H_1 : \beta_3 \neq 0$.
- Testing the null hypothesis $H_0 : \lambda \leq 0$ against an alternative $H_1 : \lambda > 0$, where $\lambda = \beta_2 + 3\beta_3$.
  - Notice that while λ is a linear combination of parameters, we are still only testing the linear combination, not the individual components.
  - Testing $\lambda = 0$ is different than testing that both $\beta_2 = 0$ and $\beta_3 = 0$.

Advanced Inference: Multiple Hypothesis Testing

However, often in multiple hypothesis testing we would like to test multiple hypotheses at the same time. Consider the multiple linear regression model

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + \epsilon.$$

Now, we will consider testing multiple conjectures about the coefficients.

- Will limit ourselves to "two-sided" alternatives, that is we will only test equality restrictions.

Advanced Inference: Multiple Hypothesis Testing

However, often in multiple hypothesis testing we would like to test multiple hypotheses at the same time. Consider the multiple linear regression model

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + \epsilon.$$

Now, we will consider testing multiple conjectures about the coefficients.

- Will limit ourselves to "two-sided" alternatives, that is we will only test equality restrictions.

Before:

$$H_0 : \beta_1 = 0 \text{ vs. } H_1 : \beta_1 \neq 0.$$

Advanced Inference: Multiple Hypothesis Testing

However, often in multiple hypothesis testing we would like to test multiple hypotheses at the same time. Consider the multiple linear regression model

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + \epsilon.$$

Now, we will consider testing multiple conjectures about the coefficients.

- Will limit ourselves to "two-sided" alternatives, that is we will only test equality restrictions.

Before:

$$H_0 : \beta_1 = 0 \;\; \text{vs.} \;\; H_1 : \beta_1 \neq 0.$$

Now:

$$H_0 : \beta_1 = \beta_2, \beta_3 = 0 \;\; \text{vs.} \;\; H_1 : \text{At least one of these is false.}$$

Advanced Inference: Multiple Hypothesis Testing

However, often in multiple hypothesis testing we would like to test multiple hypotheses at the same time. Consider the multiple linear regression model

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + \epsilon.$$

Now, we will consider testing multiple conjectures about the coefficients.

- Will limit ourselves to "two-sided" alternatives, that is we will only test equality restrictions.

Before:

$$H_0 : \beta_1 = 0 \text{ vs. } H_1 : \beta_1 \neq 0.$$

Now:

$$H_0 : \beta_1 = \beta_2 \text{ and } \beta_3 = 0 \text{ vs. } H_1 : \beta_1 \neq \beta_2 \text{ or } \beta_3 \neq 0.$$

Advanced Inference: Multiple Hypothesis Testing

However, often in multiple hypothesis testing we would like to test multiple hypotheses at the same time. Consider the multiple linear regression model

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + \epsilon.$$

Now, we will consider testing multiple conjectures about the coefficients.

- Will limit ourselves to "two-sided" alternatives, that is we will only test equality restrictions.

Before:

$$H_0 : \beta_1 = 0 \ \text{ vs. } \ H_1 : \beta_1 \neq 0.$$

Now:

$$H_0 : \beta_1 = \beta_2 \text{ and } \beta_3 = 0 \ \text{ vs. } \ H_1 : \beta_1 \neq \beta_2 \text{ or } \beta_3 \neq 0.$$

- Since null hypothesis involves multiple restrictions, this is called a joint hypothesis test

Advanced Inference: Multiple Hypothesis Testing

However, often in multiple hypothesis testing we would like to test multiple hypotheses at the same time. Consider the multiple linear regression model

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + \epsilon.$$

Now, we will consider testing multiple conjectures about the coefficients.

- Will limit ourselves to "two-sided" alternatives, that is we will only test equality restrictions.

Before:

$$H_0 : \beta_1 = 0 \ \text{ vs. } \ H_1 : \beta_1 \neq 0.$$

Now:

$$H_0 : \beta_1 = \beta_2 \text{ and } \beta_3 = 0 \ \text{ vs. } \ H_1 : \beta_1 \neq \beta_2 \text{ or } \beta_3 \neq 0.$$

- Since null hypothesis involves multiple restrictions, this is called a joint hypothesis test
- Alternative is always two sided. Won't consider test something like

$$H_0 : \beta_1 \leq \beta_2 \text{ and } \beta_3 \geq 0 \ \text{ vs. } \ H_1 : \beta_1 > \beta_2 \text{ or } \beta_3 < 0.$$

Example (Demand Estimation)

A hamburger restaurant considers the following model for sales:

$$\text{Sales} = \beta_0 + \beta_1 \text{Price} + \beta_3 \text{Advert} + \beta_4 \text{Advert}^2 + \epsilon.$$

We want to test whether advertising has any effect on sales. In this context, this means testing the joint hypothesis:

$$H_0 : \beta_3 = \beta_4 = 0 \;\; \text{vs.} \;\; H_1 : \beta_3 \neq 0 \text{ or } \beta_3 \neq 0.$$

Example (Demand Estimation)

A hamburger restaurant considers the following model for sales:

$$\text{Sales} = \beta_0 + \beta_1 \text{Price} + \beta_3 \text{Advert} + \beta_4 \text{Advert}^2 + \epsilon.$$

We want to test whether advertising has any effect on sales. In this context, this means testing the joint hypothesis:

$$H_0 : \beta_3 = \beta_4 = 0 \;\; \text{vs.} \;\; H_1 : \beta_3 \neq 0 \text{ or } \beta_3 \neq 0.$$

Notice the difference between running this test and testing something like

$$H_0 : \beta_3 + 2\beta_4 = 0 \;\; \text{vs.} \;\; H_1 : \beta_3 + 2\beta_4 \neq 0.$$

Advanced Inference: Examples

Example (Returns to Education and Experience)

Suppose we estimate the model:

$$\ln(\text{Wage}) = \beta_0 + \beta_1 \text{Edu} + \beta_2 \text{Exper} + \beta_3 \text{Exper}^2 + \beta_4 \text{Exper} \cdot \text{Edu} + \epsilon.$$

We want to test whether experience has any effect on wages, which is equivalent to testing

$$H_0 : \beta_2 = \beta_3 = \beta_4 = 0 \text{ vs. } H_1 : \beta_2 \neq \text{ or } \beta_3 \neq 0 \text{ or } \beta_4 \neq 0.$$

Alternatively, if we wanted to test whether education has any effect on wages we would test:

$$H_0 : \beta_1 = \beta_4 = 0 \text{ vs. } H_1 : \beta_1 \neq 0 \text{ or } \beta_4 \neq 0.$$

Example (Infrastructure)

Suppose LA metro wants to understand whether the number of subway rides is affected by the price of alternative modes of transportation. They estimate the model:

$$\text{No. of Subway Rides} = \beta_0 + \beta_1 \text{Price}_{\text{bus}} + \beta_2 \text{Price}_{\text{gas}} + \beta_3 \text{Price}_{\text{uber}} + \epsilon.$$

The null and alternative hypotheses for whether the prices of substitutes matter are given:

$$H_0 : \beta_2 = \beta_3 = \beta_4 = 0 \quad \text{vs.} \quad H_1 : \beta_2 \neq 0 \text{ or } \beta_3 \neq 0 \text{ or } \beta_4 \neq 0.$$

Example (Model Selection)

Suppose we have estimated the model

$$\text{Anxiety} = \beta_0 + \beta_1 \text{Classes} + \epsilon.$$

We are considering adding information on number of energy drinks and the number of hours of sleep one gets to this model. That is, we are considering estimating the model

$$\text{Anxiety} = \beta_0 + \beta_1 \text{Classes} + \beta_2 \text{Energy Drinks} + \beta_3 \text{Sleep} + \epsilon.$$

We want to know if adding these new covariates adds any explanatory power to our model. This is equivalent to testing

$$H_0 : \beta_2 = \beta_3 = 0 \;\; \text{vs.} \;\; H_1 : \beta_2 \neq 0 \text{ or } \beta_3 \neq 0.$$

Advanced Inference: Testing Procedure

Notice that in all of these we want to test multiple equality restrictions. How do we go about this?

Advanced Inference: Testing Procedure

Notice that in all of these we want to test multiple equality restrictions. How do we go about this?

Our general approach for hypothesis testing has been as follows:

1. Step 1: Formally state the null and the alternative hypothesis
   - Steps that follow depend on what the alternative is
2. Step 2: Look at the data and see whether there is evidence against the null hypothesis
   - Compute the p-value. Does the data look unusual under the assumption that H_0 holds?
3. Step 3: Based on the evidence, decide whether or not to reject H_0.
   - Reject if the p-value is less than α.

Advanced Inference: Testing Procedure

We have now set up our null and alternative hypothesis. We need to now think about what we would expect the distribution of our data to look like under our null hypothesis. To do so, we will define the restricted and the unrestricted models.

Advanced Inference: Testing Procedure

We have now set up our null and alternative hypothesis. We need to now think about what we would expect the distribution of our data to look like under our null hypothesis. To do so, we will define the restricted and the unrestricted models.

- Restricted Model: Incorporates the null hypothesis restrictions on the model.
- Unrestricted Model: More general model specified by the alternative hypothesis.

Advanced Inference: Testing Procedure

We have now set up our null and alternative hypothesis. We need to now think about what we would expect the distribution of our data to look like under our null hypothesis. To do so, we will define the restricted and the unrestricted models.

- Restricted Model: Incorporates the null hypothesis restrictions on the model.
- Unrestricted Model: More general model specified by the alternative hypothesis.

Key Idea: If the null hypothesis is true the unrestricted model shouldn't give a large improvement over the restricted model.

- In any finite sample, the unrestricted model will always give at least a slightly better fit than the restricted model. Under the null hypothesis this improvement shouldn't be very large.

Advanced Inference: Testing Procedure

We have now set up our null and alternative hypothesis. We need to now think about what we would expect the distribution of our data to look like under our null hypothesis. To do so, we will define the restricted and the unrestricted models.

- Restricted Model: Incorporates the null hypothesis restrictions on the model.
- Unrestricted Model: More general model specified by the alternative hypothesis.

Key Idea: If the null hypothesis is true the unrestricted model shouldn't give a large improvement over the restricted model.

- In any finite sample, the unrestricted model will always give at least a slightly better fit than the restricted model. Under the null hypothesis this improvement shouldn't be very large.

Let's go over some examples of restricted and unrestricted models.

Advanced Inference: Testing Procedure

Example (Demand Estimation)

A hamburger restaurant considers the following model to forecast sales:

$$\text{Sales} = \beta_0 + \beta_1 \text{Price} + \beta_2 \text{Advert} + \beta_3 \text{Advert}^2 + \epsilon.$$

As before we want to test the null hypothesis that advertising has no effect on sales ($H_0 : \beta_2 = \beta_3 = 0$). The restricted model is

$$\text{Sales} = \beta_0 + \beta_1 \text{Price} + \epsilon.$$

The unrestricted model is the full model:

$$\text{Sales} = \beta_0 + \beta_1 \text{Price} + \beta_2 \text{Advert} + \beta_3 \text{Advert}^2 + \epsilon.$$

Advanced Inference: Testing Procedure

Example (Returns to Education and Experience)

Suppose we are considering the model:

$$\ln(\text{Wage}) = \beta_0 + \beta_1 \text{Edu} + \beta_2 \text{Exper} + \beta_3 \text{Exper}^2 + \beta_4 \text{Exper} \cdot \text{Edu} + \epsilon.$$

The null hypothesis is that experience has no effect on wages ($H_0 : \beta_2 = \beta_3 = \beta_4 = 0$). The restricted model is:

$$\ln(\text{Wage}) = \beta_0 + \beta_1 \text{Edu} + \epsilon.$$

In constrast, the unrestricted model is the full model:

$$\ln(\text{Wage}) = \beta_0 + \beta_1 \text{Edu} + \beta_3 \text{Exper} + \beta_4 \text{Exper}^2 + \beta_4 \text{Exper} \cdot \text{Edu} + \epsilon.$$

Example (Returns to education and experience)

Suppose we are considering the model:

$$\ln(\text{Wage}) = \beta_0 + \beta_1 \text{Edu} + \beta_2 \text{Exper} + \epsilon.$$

We want to test the null hypothesis that returns to experience are the same as returns to education ($h_0 : \beta_1 = \beta_2$). The restricted model in this case would be

$$\ln(\text{Wage}) = \beta_0 + \beta_1 (\text{Edu} + \text{Exper}) + \epsilon.$$

Whereas the unrestricted model would be

$$\ln(\text{Wage}) = \beta_0 + \beta_1 \text{Edu} + \beta_2 \text{Exper} + \epsilon.$$

We estimate the parameters of the restricted model and the unrestricted model just as before. The restricted model is estimated

$$\hat{\beta}_0^R, \ldots, \hat{\beta}_p^R = \arg \min_{b_0, \ldots, b_p \text{ satisfy } H_0} \frac{1}{n} \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_{1,i} - \cdots - b_p X_{p,i})^2.$$

The unrestricted model is estimated

$$\hat{\beta}_0^R, \ldots, \hat{\beta}_p^R = \arg \min_{b_0, \ldots, b_p} \frac{1}{n} \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_{1,i} - \cdots - b_p X_{p,i})^2.$$

Advanced Inference: Testing Procedure

Example (Returns to Education and Experience)

Suppose we are considering the model:

$$\ln(\mathsf{Wage}) = \beta_0 + \beta_1 \mathsf{Edu} + \beta_2 \mathsf{Exper} + \beta_3 \mathsf{Exper}^2 + \epsilon.$$

The null hypothesis is that experience has no effect on wages ($H_0 : \beta_2 = \beta_3 = 0$). The restricted model is:

$$\ln(\mathsf{Wage}) = \beta_0 + \beta_1 \mathsf{Edu} + \epsilon.$$

This can be estimated by finding

$$\begin{aligned}
\hat{\beta}_0^R, \hat{\beta}_1^R &= \arg \min_{b_0, b_1, b_2 = b_3 = 0} \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - b_0 - b_1 \mathsf{Edu}_i - b_2 \mathsf{Exper}_i - b_3 \mathsf{Exper}_i^2 \right)^2 \\
&= \arg \min_{b_0, b_1} \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - b_0 - b_1 \mathsf{Edu}_i \right)^2 .
\end{aligned}$$

Note: This is the method of estimating the restricted model that we are used to. Nothing has changed. The unrestricted model is estimated as before.

Example (Returns to Education and Experience)

Suppose we are considering the model:

$$\ln(\text{Wage}) = \beta_0 + \beta_1 \text{Edu} + \beta_2 \text{Exper} + \epsilon.$$

We want to test the null hypothesis that returns to experience are the same as returns to education ($H_0 : \beta_1 = \beta_2$). The restricted model is:

$$\ln(\text{Wage}) = \beta_0 + \beta_1 (\text{Edu} + \text{Exper}) + \epsilon.$$

To estimate this model we take

$$\hat{\beta}_0^R, \hat{\beta}_1^R = \arg\min_{b_0, b_1} \frac{1}{n} \sum_{i=1}^{n} \left(\ln(\text{Wage}_i) - b_0 - b_1 (\text{Edu}_i + \text{Exper}_i) \right)^2.$$

Advanced Inference: Testing Procedure

After fitting our restricted model and our unrestricted model, we get two different measures of fit:

- SSE_R: The sum of squared errors from our restricted model

$$\text{SSE}_\text{R} = \sum_{i=1}^{n} (Y_i - \widehat{Y}_i^R)^2.$$

- SSE_U: The sum of squared errors from our unrestricted model

$$\text{SSE}_\text{U} = \sum_{i=1}^{n} (Y_i - \widehat{Y}_i^U)^2.$$

Because the unrestricted model has fewer restrictions on the parameter estimates than the restricted model, we will always have that $\text{SSE}_\text{U} \leq \text{SSE}_\text{R}$, that is the unrestricted model will always have a lower SSE than the restricted model.

Advanced Inference: Testing Procedure

After fitting our restricted model and our unrestricted model, we get two different measures of fit:

- SSE_R: The sum of squared errors from our restricted model

$$\mathsf{SSE}_\mathsf{R} = \sum_{i=1}^{n}(Y_i - \widehat{Y}_i^R)^2.$$

- SSE_U: The sum of squared errors from our unrestricted model

$$\mathsf{SSE}_\mathsf{U} = \sum_{i=1}^{n}(Y_i - \widehat{Y}_i^U)^2.$$

Because the unrestricted model has fewer restrictions on the parameter estimates than the restricted model, we will always have that $\mathsf{SSE}_\mathsf{U} \leq \mathsf{SSE}_\mathsf{R}$, that is the unrestricted model will always have a lower SSE than the restricted model.

- Key Idea: If the null hypothesis restrictions are true SSE_U will not be too much smaller than SSE_R.

Advanced Inference: Testing Procedure

After fitting our restricted model and our unrestricted model, we get two different measures of fit:

- SSE_R: The sum of squared errors from our restricted model

$$\text{SSE}_\text{R} = \sum_{i=1}^{n}(Y_i - \widehat{Y}_i^R)^2.$$

- SSE_U: The sum of squared errors from our unrestricted model

$$\text{SSE}_\text{U} = \sum_{i=1}^{n}(Y_i - \widehat{Y}_i^U)^2.$$

Because the unrestricted model has fewer restrictions on the parameter estimates than the restricted model, we will always have that $\text{SSE}_\text{U} \leq \text{SSE}_\text{R}$, that is the unrestricted model will always have a lower SSE than the restricted model.

- Key Idea: If the null hypothesis restrictions are true SSE_U will not be too much smaller than SSE_R.
- If the null hypothesis is false, than SSE_R should be much larger than SSE_U since we are imposing false restrictions

Advanced Inference: Testing Procedure

Key Idea:

- If the null hypothesis restrictions are true SSE_U will not be too much smaller than SSE_R.

Advanced Inference: Testing Procedure

Key Idea:

- If the null hypothesis restrictions are true SSE_U will not be too much smaller than SSE_R.
- If the null hypothesis is false, than SSE_R should be much larger than SSE_U since we are imposing false restrictions.

Advanced Inference: Testing Procedure

Key Idea:

- If the null hypothesis restrictions are true SSE_U will not be too much smaller than SSE_R.
- If the null hypothesis is false, than SSE_R should be much larger than SSE_U since we are imposing false restrictions.

Testing Procedure: Reject if $\text{SSE}_\text{R} - \text{SSE}_\text{U}$ is "sufficiently" large.

Advanced Inference: The F-Statistic

Formally, we will compare our SSE_R to our SSE_U by constructing the following F-statistic.

$$F^* = \frac{(\text{SSE}_\text{R} - \text{SSE}_\text{U})/J}{\text{SSE}_\text{U}/(n-p-1)}.$$

where

- n is the sample size
- J is the number of restrictions in H_0.
  - Think "count the equality signs"
- $p+1$ is the number of parameters in the unrestricted model (p slope parameters plus an intercept).

Advanced Inference: The F-Statistic

Under the null hypothesis that the restrictions hold, the F-statistic is distributed

$$F^* \sim F(J, n-p-1).$$

The p-value is then computed as probability that a random variable with this distribution would take on a value larger than our observed test statistic F^*.

Advanced Inference: The F-Statistic

Under the null hypothesis that the restrictions hold, the F-statistic is distributed

$$F^* \sim F(J, n - p - 1).$$

The p-value is then computed as probability that a random variable with this distribution would take on a value larger than our observed test statistic F^*.

- We can calculate the probability (under the null hypothesis) that a random variable distributed $F(J, n-p-1)$ takes on a value less than or equal to a constant c using the "pf" command in R:

$$\Pr\left(F(J, n-p-1) \leq c\right) = \mathsf{pf}(c, J, n-p-1).$$

Advanced Inference: The F-Statistic

Under the null hypothesis that the restrictions hold, the F-statistic is distributed

$$F^* \sim F(J, n-p-1).$$

The p-value is then computed as probability that a random variable with this distribution would take on a value larger than our observed test statistic F^*.

The p-value is the probability that we would obtain our observed value of F^* or something even larger (an even larger deviation of SSE_U from SSE_R) under the null. So, the p-value for this test can be computed:

$$p = \Pr\left(F(J, n-p-1) > F^*\right) = 1 - \mathsf{pf}(F^*, J, n-p-1).$$

As before, we reject if this p-value is smaller than some prespecified level $p < \alpha$.

Let's see how this works in practice.

Example (Demand Estimation)

A hamburger restaurant considers the following model for sales:

$$\text{Sales} = \beta_0 + \beta_1 \text{Price} + \beta_2 \text{Advert} + \beta_3 \text{Advert}^2 + \epsilon.$$

We want to test the null hypothesis that advertising has no effect on sales ($H_0 : \beta_2 = \beta_3 = 0$) against the alternative that it does ($H_1 : \beta_2 \neq 0$ or $\beta_3 \neq 0$).

Let's see how this works in practice.

Example (Demand Estimation)

A hamburger restaurant considers the following model for sales:

$$\text{Sales} = \beta_0 + \beta_1 \text{Price} + \beta_2 \text{Advert} + \beta_3 \text{Advert}^2 + \epsilon.$$

We want to test the null hypothesis that advertising has no effect on sales ($H_0 : \beta_2 = \beta_3 = 0$) against the alternative that it does ($H_1 : \beta_2 \neq 0$ or $\beta_3 \neq 0$). After collecting a sample of size $n = 75$ and estimating the restricted model

$$\text{Sales} = \beta_0 + \beta_1 \text{Price} + \epsilon,$$

we find that $\text{SSE}_\text{R} = 1896.391$. Estimating the unrestricted model gives us that $\text{SSE}_\text{U} = 1531.084$. Should we reject our null hypothesis at level $\alpha = 0.05$?

Example (Demand Estimation)

We find that $\text{SSE}_\text{R} = 1896.391$ and $\text{SSE}_\text{U} = 1531.084$. Should we reject our null hypothesis at level $\alpha = 0.05$?

Let's construct our F-Statistic.

- We know that $n = 75$.
- The full model has a total of $p + 1 = 3 + 1 = 4$ parameters.
- Our null hypothesis is $H_0 : \beta_2 = \beta_3 = 0$, for a total of $J = 2$ restrictions

So we can construct our test statistic:

$$F^* = \frac{(\text{SSE}_\text{R} - \text{SSE}_\text{U})/J}{\text{SSE}_\text{U}/(n-p-1)} = \frac{(1892.391 - 1531.084)/2}{1531.084/71} \approx 8.377.$$

Example (Demand Estimation)

We compute the p-value using the $F(J, n-p-1) = F(2,71)$ distribution:

$$\begin{aligned} p = \Pr(F(2,71) > F^*) &= \Pr\left(F(2,71) > 8.3777\right) \\ &= 1 - \Pr\left(F(2,71) \leq 8.3777\right) \\ &= 1 - \text{pf}(8.377, 2, 71) \\ &= 0.0005. \end{aligned}$$

Since $p < \alpha = 0.05$ we reject the null hypothesis and conclude that advertising does have a significant effect on sales.

Example (Model Significance)

Consider the model

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + \epsilon.$$

A classical example of an F-test is testing for the significance of the model.

- This is a test for whether any of our regressors $X_1, \ldots, X_p$ is statistically significant.
- Formally the hypotheses we are interested in are:

$$H_0 : \beta_1 = \ldots \beta_p = 0 \quad \text{vs.} \quad H_1 : \beta_j \neq 0 \text{ for some } 1 \leq j \leq p.$$

- Intuitively, we are just testing whether our model does better at predicting Y than a constant.
- This is the F-statistic that R reports in a regression summary.

Example (Model Significance)

Because the null hypothesis is so restrictive, the formulas simplify considerably. The restricted model sets all slope parameters to zero and so just contains a constant:

$$Y = \beta_0 + \epsilon \implies \hat{\beta}_0^R = \arg\min_{b_0} \frac{1}{n} \sum_{i=1}^{n} (Y_i - b_0)^2 \implies \hat{\beta}_0^R = \bar{Y}.$$

This means that $\text{SSE}_\text{R} = \sum_{i=1}^{n} (Y_i - \bar{Y})^2 = \text{SST}$. The unrestricted model includes all slope parameters estimated normally so $\text{SSE}_\text{U} = \text{SSE}$.

Example (Model Significance)

Recall from our discussion of R^2 that we have the following decomposition:

$$\underbrace{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}_{\text{SST}} = \underbrace{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}_{\text{SSR}} + \underbrace{\sum_{i=1}^{n}\hat{\epsilon}_i^2}_{\text{SSE}}$$

where $\hat{Y}_i$ is the prediction from the unrestricted model and $\hat{\epsilon}_i$ is the estimated residual from the unrestricted model. Using this, and since $R^2 = \text{SSR}/\text{SST}$ we can

simplify the F-statistic:

$$F^* = \frac{(\text{SSE}_\text{R} - \text{SSE}_\text{U})/J}{\text{SSE}_\text{U}/(n-p-1)} = \frac{(\text{SST} - \text{SSE})/p}{\text{SSE}/(n-p-1)} = \frac{R^2/p}{(1-R^2)/(n-p-1)}.$$

Advanced Inference: The F-Statistic

Example (Model Significance)

Recall from our discussion of R^2 that we have the following decomposition:

$$\underbrace{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}_{\text{SST}} = \underbrace{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}_{\text{SSR}} + \underbrace{\sum_{i=1}^{n}\hat{\epsilon}_i^2}_{\text{SSE}}$$

where $\hat{Y}_i$ is the prediction from the unrestricted model and $\hat{\epsilon}_i$ is the estimated residual from the unrestricted model. Using this, and since $R^2 = \text{SSR}/\text{SST}$ we can

simplify the F-statistic:

$$F^* = \frac{(\text{SSE}_\text{R} - \text{SSE}_\text{U})/J}{\text{SSE}_\text{U}/(n-p-1)} = \frac{(\text{SST} - \text{SSE})/p}{\text{SSE}/(n-p-1)} = \frac{R^2/p}{(1-R^2)/(n-p-1)}.$$

Key Idea: The overall significance of the model is determined by the overall fit of the model!

The F-Test: Questions

Questions?

The F-Test: Why Impose Restrictions

Suppose we are just interested in prediction. A natural question here is: why bother imposing restrictions? Why not just estimate all parameters in the unrestricted model?

The F-Test: Why Impose Restrictions

Suppose we are just interested in prediction. A natural question here is: why bother imposing restrictions? Why not just estimate all parameters in the unrestricted model?

- Estimating more parameters increases the variance of each of our estimates.

The F-Test: Why Impose Restrictions

Suppose we are just interested in prediction. A natural question here is: why bother imposing restrictions? Why not just estimate all parameters in the unrestricted model?

- Estimating more parameters increases the variance of each of our estimates.
- Estimating too many parameters can decrease the interpretability of our model and lead to overfitting.

The F-Test: Why Impose Restrictions

Suppose we are just interested in prediction. A natural question here is: why bother imposing restrictions? Why not just estimate all parameters in the unrestricted model?

- Estimating more parameters increases the variance of each of our estimates.
- Estimating too many parameters can decrease the interpretability of our model and lead to overfitting.

However, as we will now see, imposing too many restrictions can lead to problems as well.

Table of Contents

Model Selection and Omitted Variables Bias

For the most part in this lecture, we have taken as a given that we have some model:

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + \epsilon.$$

Model Selection and Omitted Variables Bias

For the most part in this lecture, we have taken as a given that we have some model:

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + \epsilon.$$

However, in practice, we are the ones that must select which variables to include in our model:

- Which of the available data we should use as regressors?
- Should we include transformations of our regressors?
- What are the trade-offs between including and excluding a variable?

Model Selection and Omitted Variables Bias

For the most part in this lecture, we have taken as a given that we have some model:

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + \epsilon.$$

However, in practice, we are the ones that must select which variables to include in our model:

- Which of the available data we should use as regressors?
- Should we include transformations of our regressors?
- What are the trade-offs between including and excluding a variable?

Selecting the right model is a bit of an art, there is no easy rule/recipe to follow. Good model selection combines statistical reasoning as well as knowledge of the problem/setting at hand.

We have covered what happens when we include irrelevant variables. Now let's consider what happens when we exclude a relevant variable.

Recall in the beginning of class we were interested in the relationship between energy drinks consumed and anxiety levels. We looked at a study that (essentially) estimated the following model

$$\text{Anxiety} = \beta_0 + \beta_1 \text{Energy Drinks} + \epsilon$$

and found that $\beta_1 > 0$. We reasoned that this positive association may be due to the fact that people who drink more energy drinks may be taking more classes, and it is the classes that are driving anxiety levels rather than the energy drinks. That is, if we were to instead consider the model

$$\text{Anxiety} = \beta_0^\circ + \beta_1^\circ \text{Energy Drinks} + \beta_2^\circ \text{Classes} + \epsilon^\circ,$$

we would find a value of β_1° that would be much smaller than our β_1 from before. This difference $\beta_1 - \beta_1^\circ$ is called an omitted variables bias.

Omitted Variables Bias

Let's suppose we have access to two possible explanatory variables X_1, X_2 and we consider two models. The first model contains only X_1

$$Y = \beta_0 + \beta_1 X_1 + \epsilon.$$

The second model contains both X_1 and X_2

$$Y = \beta_0^\circ + \beta_1^\circ X_1 + \beta_2^\circ X_2 + \epsilon^\circ.$$

Omitted Variables Bias

Let's suppose we have access to two possible explanatory variables X_1, X_2 and we consider two models. The first model contains only X_1

$$Y = \beta_0 + \beta_1 X_1 + \epsilon.$$

The second model contains both X_1 and X_2

$$Y = \beta_0^\circ + \beta_1^\circ X_1 + \beta_2^\circ X_2 + \epsilon^\circ.$$

Question: What is the relationship between β_1° and β_1?

- In other words, how does the observed relationship between Y and X_1 change when we account for X_2?

Omitted Variables Bias

Let's suppose we have access to two possible explanatory variables X_1, X_2 and we consider two models. The first model contains only X_1

$$Y = \beta_0 + \beta_1 X_1 + \epsilon.$$

The second model contains both X_1 and X_2

$$Y = \beta_0^\circ + \beta_1^\circ X_1 + \beta_2^\circ X_2 + \epsilon^\circ.$$

Question: What is the relationship between β_1° and β_1?

- In other words, how does the observed relationship between Y and X_1 change when we account for X_2?

By performing some algebra, we can find that

$$\beta_1 = \beta_1^\circ + \underbrace{\beta_2^\circ \frac{\mathrm{Cov}(X_1, X_2)}{\mathrm{Var}(X_1)}}_{\text{Omitted Variables Bias}}.$$

Omitted Variables Bias

The omitted variables bias from excluding X_2 in our regression model is given:

$$\beta_2^{\circ}\frac{\mathrm{Cov}(X_1, X_2)}{\mathrm{Var}(X_1)}.$$

Some Intuition:

- If X_2 has a positive relationship with the outcome Y and X_1 and X_2 are positively related, then we will have a postive omitted variables bias, $\beta_1 > \beta_1^{\circ}$.
    - Classes and Anxiety levels have a positive relationship, Classes and Energy Drink consumption have a positive relationship.
    - It will look like energy drink consumption has a stronger positive relationship with anxiety level than it "truly" does since people drinking more energy drinks are likely to be taking more classes.

Omitted Variables Bias

The omitted variables bias from excluding X_2 in our regression model is given:

$$\beta_2^\circ \frac{\mathrm{Cov}(X_1, X_2)}{\mathrm{Var}(X_1)}.$$

Some Intuition:

- If X_2 has a negative relationship with the outcome and X_1 and X_2 are positively related then we will have a negative omitted variables bias, $\beta_1 < \beta_1^\circ$.

  - Suppose we are interested in relationship between anxiety levels, taking Advil PM, and amount of sleep a student is getting. We believe that more sleep helps lower anxiety levels so $\beta_2^\circ < 0$ and that taking Advil PM induces sleep so that $\mathrm{Cov}(X_1, X_2) > 0$.
  - If we were to just regress anxiety levels on whether or not someone is taking Advil PM, we may get a fairly negative value for β_1 and conclude that Advil PM appears to reduce anxiety levels.
  - But, this negative β_1 value is probably due to omitting the sleep variable. Once we include it, it is more likely that we get $\beta_1^\circ \approx 0$.

Omitted Variables Bias

The omitted variables bias from excluding X_2 in our regression model is given:

$$\beta_2^{\circ} \frac{\mathrm{Cov}(X_1, X_2)}{\mathrm{Var}(X_1)}.$$

Some Intuition:

Can keep reasoning through all the different cases, and will do so more in homework. But important to note that it is rare that $\beta_1 = \beta_1^{\circ}$.

- Would need either $\beta_2^{\circ} = 0$, X_2 has no effect on the outcome Y or,
- $\mathrm{Cov}(X_1, X_2) = 0$, X_2 has no (linear) relationship with X_1.

Omitted Variables Bias: Questions?

Questions?

Modeling: Indicator Variables

The final modeling technique we will talk about is using indicator or "Dummy" variables.

Definition (Indicator Variables)

A indicator or "dummy" variable D is a variable D is a variable that only takes on two values. Usually

$$D \in \{0, 1\}.$$

Modeling: Indicator Variables

The final modeling technique we will talk about is using indicator or "Dummy" variables.

Definition (Indicator Variables)

A indicator or "dummy" variable D is a variable D is a variable that only takes on two values. Usually

$$D \in \{0, 1\}.$$

Question: Why would this sort of variable be useful?

- Can turn a categorical variable into a numeric variable
    - Ex. create a dummy variable that is equal to one if a persons favorite color is "blue"
- Helpful for letting parameters of regression be individual to certain subgroups:
    - Can multiply parameters by dummy variables
- Deal with special effects for certain thresholds:
    - College degree vs. no college degree

Modeling: Indicator Variables

The final modeling technique we will talk about is using indicator or "Dummy" variables.

Definition (Indicator Variables)

A indicator or "dummy" variable D is a variable D is a variable that only takes on two values. Usually

$$D \in \{0, 1\}.$$

Question: Why would this sort of variable be useful?

- Can turn a categorical variable into a numeric variable
  - Ex. create a dummy variable that is equal to one if a persons favorite color is "blue"
- Helpful for letting parameters of regression be individual to certain subgroups:
  - Can multiply parameters by dummy variables
- Deal with special effects for certain thresholds:
  - College degree vs. no college degree

Let's see some examples of this

Example (Home Characteristics)

Suppose we are interested in estimating the sales price for a house. In the past we've estimated:

$$\text{Price} = \beta_0 + \beta_1 \text{Sqft} + \epsilon.$$

Problem: There are many qualitative factors that affect the price:

- Is the house close to UCLA?
- Does the house have a pool?

Example (Home Characteristics)

Suppose we are interested in estimating the sales price for a house. In the past we've estimated:

$$\text{Price} = \beta_0 + \beta_1 \text{Sqft} + \epsilon.$$

Problem: There are many qualitative factors that affect the price:

- Is the house close to UCLA?
- Does the house have a pool?

Solution: Model whether the qualitative factor is present by using a dummy variable!

$$D = \begin{cases} 1 & \text{if characteristic is present} \\ 0 & \text{if characteristic is not present} \end{cases}.$$

For example, let's let $D = 1$ if the house is within 5 miles of UCLA and $D = 0$ otherwise.

Indicator Variables: Intercept Changes

Example (Home Characteristics)

Let's let $D = 1$ if the house is within 5 miles of UCLA and $D = 0$ otherwise. We now consider the model

$$\text{Price} = \beta_0 + \delta D + \beta_2 \text{Sqft} + \epsilon.$$

Note: We can now think of δ as the price premium for a house that is close to UCLA.

$$\widehat{\text{Price}} = \begin{cases} (\beta_0 + \delta) + \beta_2 \text{Sqft} & \text{if the house is within 5 miles of UCLA} \\ \beta_0 + \beta_2 \text{Sqft} & \text{otherwise} \end{cases}.$$

This is equivalent to having a different <u>intercept term</u> for houses within 5 miles of UCLA.

Indicator Variables: Intercept Changes

Space to draw what this would look like:

Indicator Variables: Intercept Changes

Question: What if instead of letting $D = 1$ when the house is close to UCLA, we set:

$$LD = \begin{cases} 1 & \text{if house is more than 5 miles from UCLA} \\ 0 & \text{if house is withing 5 miles of UCLA} \end{cases}$$

Note that this is the opposite of what we had before.

Answer: This is perfectly fine, it just changes the interpretation!

Indicator Variables: Intercept Changes

Question: What if instead of letting $D = 1$ when the house is close to UCLA, we set:

$$LD = \begin{cases} 1 & \text{if house is more than 5 miles from UCLA} \\ 0 & \text{if house is withing 5 miles of UCLA} \end{cases}$$

Note that this is the opposite of what we had before.

Answer: This is perfectly fine, it just changes the interpretation!

Example (Home Characteristics)

If instead of using D we modify the previous regression to be

$$\text{Price} = \beta_0 + \delta LD + \beta_2 \text{Sqft} + \epsilon.$$

Then δ is the price discount for not being close to UCLA (expect $\delta < 0$).

Notes:

- The group corresponding to $D = 0$ is sometimes called the reference group.
- Be careful not to include both D and LD and a constant in a regression. Since $D = 1 - LD$, this causes perfect collinearity (rank condition is violated).

Indicator Variables: Slope Changes

Example (Returns to Education)

Suppose we are interested in the relationship between educational attainment and wages. We could suspect that having a college degree has a particular impact above and beyond an additional year of education. Following the example above, we may encode the dummy variable:

$$D = \begin{cases} 1 & \text{if person has a college degree} \\ 0 & \text{otherwise} \end{cases}$$

and estimate the model:

$$\ln(\text{Wages}) = \beta_0 + \delta D + \beta_2 \text{Edu} + \epsilon.$$

Question: But, what if we believe that an additional year of education after completing college has a different effect than an additional year of education before completing college?

Indicator Variables: Slope Changes

Example (Returns to Education)

Question: What if we believe that an additional year of education after completing college has a different effect than an additional year of education before completing college?

Example (Returns to Education)

In this case, we may want the slope parameter to differ for college graduates as well. To model this, we can estimate the model:

$$\ln(\text{Wages}) = \beta_0 + \delta D + \beta_2 \text{Edu} + \gamma D \cdot \text{Edu} + \epsilon.$$

Now we are allowing both the slope and the intercept to change for college graduates:

$$\underbrace{\widehat{Y}(\text{Edu} = 0)}_{\text{Intercept}} = \begin{cases} \beta_0 + \delta & \text{if college graduate} \\ \beta_0 & \text{if not college graduate} \end{cases}$$

and for the slope:

$$\frac{\partial \widehat{Y}}{\partial \text{Edu}} = \begin{cases} \beta_2 + \gamma & \text{if college graduate} \\ \beta_2 & \text{otherwise} \end{cases}.$$

Indicator Variables: Slope Changes

Space to draw what this would look like:

Indicator Variables: Summary

In summary: Indicator variables can allow for a lot of flexibility in our model!

- Allows for the intercepts and slopes to differ by subgroup
- Can allow us to include qualitative data in our models
- Just have to be a bit careful about collinearity

Conclusion

Essentially in this lecture, we have considered model selection. There are two competing risks when doing model selection:

- Including irrelevant variables
  - Increases the variance of each parameter estimate, risk of overfitting, decreases interpretability of our model
  - Can use F-test to check for irrelevant variables

Conclusion

Essentially in this lecture, we have considered model selection. There are two competing risks when doing model selection:

- Including irrelevant variables
  - Increases the variance of each parameter estimate, risk of overfitting, decreases interpretability of our model
  - Can use F-test to check for irrelevant variables
- Excluding relevant variables
  - Leads to omitted variables bias, interpretation of our model can be incorrect,
  - Can think through omitted variables bias formula. This formula is not exact once we consider more variables, but the reasoning is the same. Often useful in causal inference settings.

Conclusion

Essentially in this lecture, we have considered model selection. There are two competing risks when doing model selection:

- Including irrelevant variables
  - Increases the variance of each parameter estimate, risk of overfitting, decreases interpretability of our model
  - Can use F-test to check for irrelevant variables
- Excluding relevant variables
  - Leads to omitted variables bias, interpretation of our model can be incorrect,
  - Can think through omitted variables bias formula. This formula is not exact once we consider more variables, but the reasoning is the same. Often useful in causal inference settings.

There are some statistical procedures that can try to help with model selection. We have gone over one, looking at the adjusted R^2. However, this is a rough selection critetion and there are more sophisticated ones. If you are interested I can send some references.