Lecture 11: Inference in regression models

Pierre Biscaye

Fall 2022

Outline

- Recap of inference for $\hat{\beta}$
- Correlation vs. causality
- One-sided vs. two-sided tests
- p-values
- Economic significance and confidence intervals

Recap: Sampling distribution for $\hat{\beta}$

- $\hat{\beta_1} = \beta_1 + \frac{\sum_i (X_i - \bar{X}) u_i}{\sum_i (X_i - \bar{X})^2}$
    - This is an SLR derivation but intuition for MLR is the same
- Under MLR1-4: $E[\hat{\beta_1}] = \beta_1$
- MLR4-6: $u \sim N(0, \sigma_u^2)$.
- Theorem: this implies

$$\hat{\beta_1} \sim N(\beta_1, var(\hat{\beta_1})) \tag{1}$$

$$\frac{\hat{\beta_1} - \beta_1}{sd(\hat{\beta_1})} \sim N(0, 1) \tag{2}$$

$$\frac{\hat{\beta_1} - \beta_1}{s.e.(\hat{\beta_1})} \sim t_{n-k-1} \tag{3}$$

- Can use for CIs and hypothesis testing with MLR $\hat{\beta}$s, as long as MLR1-6 hold

In small groups: practice R output interpretation

Use what we know to fill in A-D in the below R output estimating

$$gradrate_i = \beta_0 + \beta_1 enroll_i + \beta_2 totcomp_i + \beta_3 staff_i + u_i \qquad (4)$$

Coefficient	Estimate	Std. Errror	t-value	Pr(>|t|)	
(Intercept)	72.6171339	(A)	9.124	<2e-16	***
enroll	-0.0004534	0.00028	-1.618	0.1064	
totcomp	0.0002758	0.00013	2.111	0.0354	(B)
staff	0.0168583	0.05183	(C)	0.7452	

Remember:

- *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, . $p < 0.1$
- $\frac{\hat{\beta_1} - \beta_1}{s.e.(\hat{\beta_1})} \sim t_{n-k-1}$
- R output shows the results for $H_0 : \beta_j = 0$ for each j
- (D) What is t if we want to test $H_0 : \beta_3 = 0.02$?

Practice R output interpretation

$$gradrate_i = \beta_0 + \beta_1 enroll_i + \beta_2 totcomp_i + \beta_3 staff_i + u_i \qquad (5)$$

Coefficient	Estimate	Std. Errror	t-value	Pr(>|t|)	
(Intercept)	72.6171339	(A)	9.124	<2e-16	***
enroll	-0.0004534	0.00028	-1.618	0.1064	
totcomp	0.0002758	0.00013	2.111	0.0354	(B)
staff	0.0168583	0.05183	(C)	0.7452	

- (E) What would we conclude if we wanted to test $H_0 : \beta_1 \geq 0$ at 10% significance level?
- More on this in a bit

Correlation vs. causality

- Our statistical model said that there is a positive statistically significant relationship between teacher compensation and the high school graduation rate.
- But we could find no statistical evidence of a relationship between enrollment and the graduation rate.
- What must be true for these estimates to be causal?
- MLR 1-4 must *all* hold

Last time: MLR1

- Suppose MLR2-4 hold.
- We estimated
  $gradrate_i = \beta_0 + \beta_1 enroll_i + \beta_2 totcomp_i + \beta_3 staff_i + u_i$
- Consider MLR 1: what if in the population $gradrate_i = \beta_0 + \beta_1 log(enroll_i) + \beta_2 log(totcomp_i) + \beta_3 log(staff_i) + u_i$
- Statistical significance may change across specifications: causality will depend on whether we've modeled the population relationship correctly.
- How do we know which specification to use?
  - In some cases, theory may guide appropriate choice.
  - One piece of evidence: higher $\bar{R}^2$ indicates model fits the data better.
  - More on this later.

MLR2: Is our sample randomly drawn from the population?

- Suppose we only sample schools with higher (or lower) enrollment.
- Is the number of students likely to be random?
- What does it imply for our estimates?

To Jupyter!

MLR2: Is our sample randomly drawn from the population?

- Suppose we only sample schools with higher (or lower) enrollment.
- What does it imply for our estimates?
  - Higher enrollment significantly decreases graduation rate in both high- and low-enrollment schools, but much more economically significant in low-enrollment schools
  - Total teacher compensation significantly increases graduation rates in high-enrollment schools, but not in low-enrollment schools
- Can still get interesting results, but no longer about the full population.
- Need to think carefully about how your data were collected and how your sample relates to the full population you care about.

MLR3: What about collinearity?

- Perfect multicollinearity won't be a problem - will drop a variable
- High but not perfect collinearity can affect estimates
- Suppose we estimate
  $gradrate_i = \beta_0 + \beta_1 enroll_i + \beta_2 totcomp_i + \beta_3 staff_i + \beta_4 salary_i + u_i$
- Any potential concerns?

To Jupyter!

MLR3: What about collinearity?

- High but not perfect collinearity can affect estimates
- Suppose we estimate
  $gradrate_i = \beta_0 + \beta_1 enroll_i + \beta_2 totcomp_i + \beta_3 staff_i + \beta_4 salary_i + u_i$
- $\hat{\beta_2}$ no longer significant, and has opposite sign.
- $totcomp_i = salary_i + benefits_i$: What is $\hat{\beta_2}$ estimate now telling us?

MLR4: What about omitted variables?

- We estimated
  $gradrate_i = \beta_0 + \beta_1 enroll_i + \beta_2 totcomp_i + \beta_3 staff_i + u_i$
- What if students' household income levels are associated with these X variables?
- Approximate income with eligibility for school lunch: $gradrate_i = \beta_0 + \beta_1 enroll_i + \beta_2 totcomp_i + \beta_3 staff_i + \beta_4 lnchprg_i + u_i$

To Jupyter!

MLR4: Interpreting results

- Estimating $gradrate_i = \beta_0 + \beta_1 enroll_i + \beta_2 totcomp_i + \beta_3 staff_i + \beta_4 lnchprg_i + u_i$
- What changed with this new regression specification?
- *totcomp* no longer statistically significant, much smaller magnitude
  - Was biased up because higher *totcomp* is associated with lower *lnchprg*, which is associated with higher *gradrate*
- *enroll* now significant at 10% level, about the same magnitude
  - Not strongly correlated with *lnchprg*
  - Adding a relevant variable to model reduced *SSR*, making estimates more precise in general
- In both cases, still not quite convinced we can treat relationships as causal. Future classes will talk about how we can get to this point.

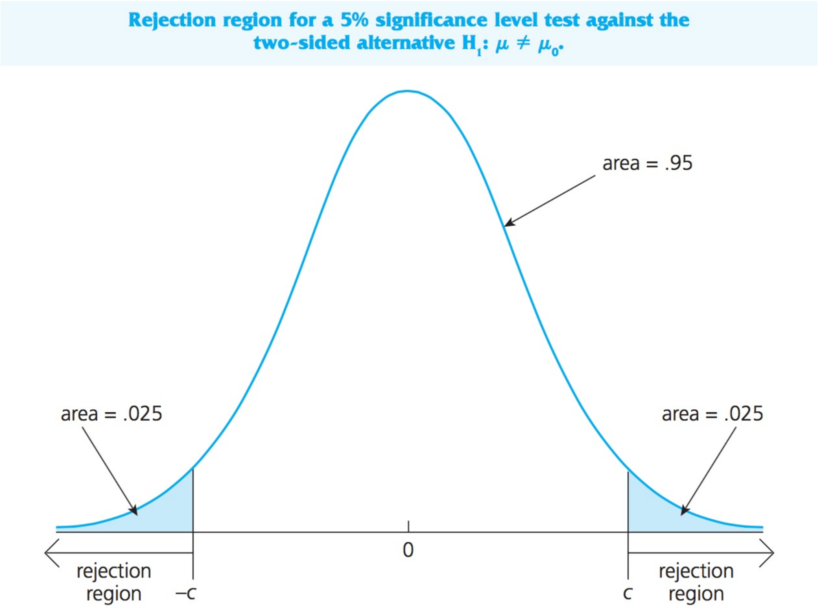
Summary: correlation vs. causality

- MLR1: Can't be sure true of population specification. Can consider theory, test robustness to different functional forms, look at $\bar{R}^2$ to try to find model that most closely fits the actual relationships in the data (but be careful of overfitting!).
- MLR2: Need to think carefully about how your data were collected, and consider what population your sample allows you to make inference about.
- MLR3: Think carefully about relationships among your variables and what adding more controls implies about the interpretation for each coefficient (holding all else constant).
- MLR4: Always be thinking about possible omitted variables and the nature of their relationships with your variables of interest. Later in the course: strategies to help ensure MLR4 holds.

When to use a one-sided or two sided test?

- One-Sided and Two-Sided tests have different critical values
- 95% one-sided positive test would reject H_0 if $t < -1.645$ (for $n > 1000$)
- 95% two-sided test would reject H_0 if $t < -1.96$ or $t > 1.96$
- How can both be correct?

Rejection regions



Interpreting rejection regions

- Under both $H_0 : \beta \geq 0$ and $H_0 : \beta = 0$, the distribution of the test statistic under the null is centered on 0
- When H_0 is true, there is a 5% chance that $t < -1.65$
- When H_0 is true there is also a 5% chance that either $t < -1.97$ or $t > 1.97$
- this means that if $-1.97 < t < -1.65$, you would reject the null if it was a positive one-sided test, but you would not if it was a two-sided test.
- *But* if $t > 1.97$, you would reject the null if it was a two-sided test but you would not if it was a one-sided test.

Interpretation: one-sided test

- In a positive one-sided test
  - $H_0 : \beta \geq 0$
  - $H_1 : \beta < 0$
- Remember, we want to avoid type 1 error (reject the null when the null is true).
- We've essentially expanded the null so that we would fail to reject whenever the sign we hypothesized was true.
- Sacrifice testing equality with 0 for testing the sign
  - In a one-sided test, will fail to reject the null even if you have good evidence that $\beta >> 0$
  - Want to be fairly certain about the hypothesized direction of the effect.
- Choice interpretation: do we think β *might* be greater than zero? Are we interested if so?

Other concern with one-sided tests

- Researchers often face incentives to reject the null
  - We learn more by rejecting the null than by failing to reject the null.
  - For academic publication, or for private profits, often benefits to reporting that you rejected the null.
- Problem: Researchers know what their t statistic ended up being.
- Suppose $t = -1.7$.
  - As a researcher, you know that $Pr(t < -1.7) < 0.05$: reject null under positive one-sided test at 5% significance level.
  - You also know $Pr(t < -1.7 \, or \, t > 1.7) > 0.05$: fail to reject null under two-sided test at 5% significance level.
  - Bad inference to state your hypothesis *after* estimating your t statistic.
- May need a registry to prove intent on using one-sided test: otherwise should be skeptical of research relying heavily on one-sided tests

p-values

- In our analyses, we actually learn more than just whether we could reject H_0 or not at a particular confidence level
- For *lnchprg*, we saw a t-statistic of -6.62. Since |-6.62|>1.97, we rejected $H_0 : \beta_4 = 0$ with 95% confidence
  - But $|-6.62| >> 1.97$
  - We would have rejected H_0 at higher levels of significance as well
- A *p*-value formalizes this

p-values

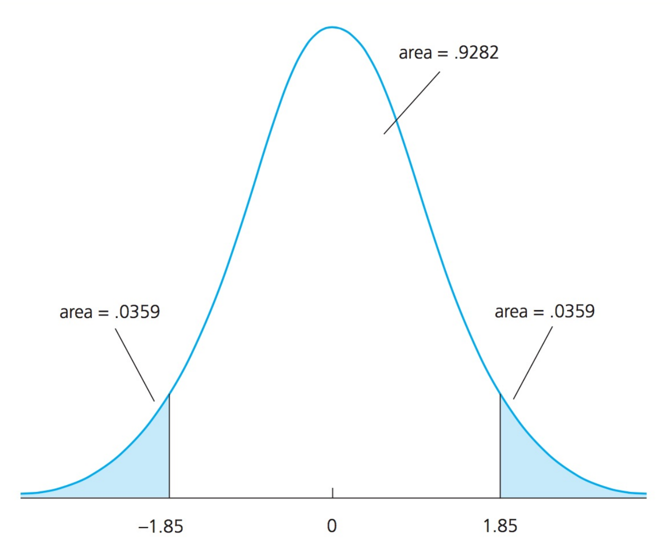
A p-value can be interpreted in 2 ways:

1. What is the highest significance level (smallest α) at which the null hypothesis would be rejected?
2. What is the probability of obtaining a test statistic at least as large as the one sampled if the null hypothesis was true?

- Note: p-values can be one or two-sided. A two-sided p-value will always be 2X as large as a one-sided.

p-values, graphically

Obtaining the p-value against a two-sided alternative, when $t = 1.85$ and $df = 40$.



p-values in R

- We estimated $gradrate_i = \beta_0 + \beta_1 enroll_i + \beta_2 totcomp_i + \beta_3 staff_i + \beta_4 lnchprg_i + u_i$
- What were the p-values?

To Jupyter!

Economic vs. statistical significance

- Economic significance: how large is $\hat{\beta_1}$?
  - Depends in part on the units of Y and X_1
  - Economic significance is often relative, but can be hard to compare across variables with different units
- Statistical significance: is $t = \frac{\hat{\beta_1} - \beta_1}{s.e.(\hat{\beta_1})} > c_\alpha$?
  - So variables with greater economic significance will often be statistically signficant.
  - So will variables with small standard errors.
  - But it is a combination of factors that determines statistical significance. Under MLR1-5:

$$s.e.(\hat{\beta_1}) = \sqrt{\frac{\sum \hat{u}_i{}^2}{(n-k-1)(1-R_j^2)SST_j}} \tag{6}$$

Note typo in SE equation in original notes

Testing other hypotheses

- Economic significance suggests we might want to test some other hypotheses too
  - Is this effect large enough that we should care about it?
- Consider testing $H_0 : \beta_j = b$ against $H_1 : \beta_j \neq b$

Recall our t-statistic

$$t = \frac{\hat{\beta}_j - \beta_j}{s.e.(\hat{\beta}_j)} \tag{7}$$

- Suppose class sizes are concerning if a one percent increase causes a 3 percentage point decrease in graduation rate. We could test:
- $H_0 : \beta_1 = -3$, $H_1 : \beta_1 \neq -3$.
- Based on the R output from the log formulation:

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -23.6120    63.7361  -0.370   0.7112
lenroll      -1.6339     0.9072  -1.801   0.0724 .
ltotcomp     11.6304     5.3074   2.191   0.0290 *
lstaff       -0.6625     5.4829  -0.121   0.9039
```

- What is our test statistic t?

Recall our t-statistic

$$t = \frac{\hat{\beta}_j - \beta_j}{s.e.(\hat{\beta}_j)} \tag{8}$$

- Suppose class sizes are concerning if a one percent increase causes a 3 percentage point decrease in graduation rate. We could test:
- $H_0 : \beta_1 = -3$, $H_1 : \beta_1 \neq -3$.
- Based on the R output from the log formulation:

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -23.6120    63.7361  -0.370   0.7112
lenroll      -1.6339     0.9072  -1.801   0.0724 .
ltotcomp     11.6304     5.3074   2.191   0.0290 *
lstaff       -0.6625     5.4829  -0.121   0.9039
```

$$t = \frac{-1.6339 + 3}{0.9072} = \frac{1.3661}{0.9072} = 1.506 \tag{9}$$

- Can't reject even at 90% confidence level ($c_{.1/2} \approx 1.65$)

Confidence Intervals

- We could construct similar hypothesis tests for *any* b.
  - Not necessarily very useful; would like a range of probable values for true β
- Confidence Intervals show values where for 95% of samples (if using 95% CI), the interval would contain the true parameter.

$$\left[\hat{\beta}_j - c_{\frac{\alpha}{2}} * s.e.(\hat{\beta}_j), \hat{\beta}_j + c_{\frac{\alpha}{2}} * se(\hat{\beta}_j)\right] \tag{10}$$

- A confidence interval for β_j contains all values of β_j where the probability of observing an estimate as large as $\hat{\beta}_j$ is at least α.
  - In other words, the CI contains all b such that we would fail to reject $H_0 : \beta_j = b$ at the α significance level.

Confidence Intervals

- Any H_0 outside of the confidence interval would be rejected with $1-\alpha$ confidence.
- So if $H_0 : \beta_j = 0$ and 0 is outside of the CI, we reject H_0, but if the CI contains 0, we fail to reject H_0 at significance level α.

To Jupyter!