Hypothesis Testing

We've tested whether a coefficient in our population model was equal to zero.

- Gave conditions under which coefficients had normal sampling distribution
- Used t-distribution for hypothesis test
- Used simple t-test to see if each coefficient is different than zero
  - Automatically provided by R in regression output

Specification can be used to test a wide variety of hypotheses.

Testing If Parameters Are Different

- $\log(\text{wage}) = \beta_0 + \beta_1 \text{jc} + \beta_2 \text{univ} + \beta_3 \text{exper} + u$
  - *jc* = years at two-year junior college
  - *univ* = years at four-year university
- Example: modeling wage as function of years spent at a two-year junior college and a four-year university
- Are the two coefficients different?
- Test $H_0 : \beta_1 - \beta_2 = 0$ against $H_0 : \beta_1 - \beta_2 \neq 0$

Setting Up Test Statistic

$$t = \frac{\hat{\beta}_1 - \hat{\beta}_2}{se\left(\hat{\beta}_1 - \hat{\beta}_2\right)}$$

- Estimating standard error, so statistic has *t*-distribution.
- R won't give us this statistic if we run the linear regression.
- We can't compute *t*-statistic we want from output in R.
- What's the standard error of the difference in the denominator?

$$se\left(\hat{\beta}_1 - \hat{\beta}_2\right) = \sqrt{\hat{Var}\left(\hat{\beta}_1 - \hat{\beta}_2\right)} = \sqrt{\hat{Var}\left(\hat{\beta}_1\right) + \hat{Var}\left(\hat{\beta}_2\right) - 2\hat{Cov}\left(\hat{\beta}_1, \hat{\beta}_2\right)}$$

- R tells us variance of $\hat{\beta}_1$ and $\hat{\beta}_2$ but not covariance.

An Alternate Method

- Change the variables:
  - Define $\theta_1 = \beta_1 - \beta_2$
  - Then $H_0 : \beta_1 - \beta_2 = 0$ is equivalent to $H_0 : \theta_1 = 0$
- Rewrite population model:

$$\begin{aligned}\log(wage) &= \beta_0 + (\theta_1 + \beta_2) jc + \beta_2 univ + \beta_3 exper + u \\ &= \beta_0 + \theta_1 jc + \beta_2 (jc + univ) + \beta_3 exper + u\end{aligned}$$

- New variable that represents total years in any college: *jc + univ*
  - Create this variable in our data table, then estimate regression.
  - R estimates θ_1 and tests whether it's statistically significant.
  - If θ_1 is statistically significant, reject original null hypothesis.

OLS Regression Results

$$\underset{}{\hat{\log}(wage)} = \underset{(.021)}{1.472} - \underset{(.0069)}{.0102} jc + \underset{(.0023)}{.0769} totcoll + \underset{(.0002)}{.0049} exper$$

$n = 6{,}763,\ R^2 = .222$

$t = -.0102/.0069 = -1.48$

$p = \Pr(|t| > 1.48) = 0.139$

- Coefficient for *jc* is less than twice its standard error, so not significant.
- *p* = 0.139, so we cannot reject hypothesis that the effects of either type of college are equal.

This method can be generalized to test any linear combination of parameters.