Regression with a Single Regressor: Hypothesis Tests and Confidence Intervals (SW Ch. 5)

Ercan Karadas

Econometrics (ECON 3112)

Belk College of Business, UNCC

October 15, 2018

Outline

Hypothesis Tests Concerning β_1

Confidence Intervals for β_1

Regression when X is Binary

Heteroskedasticity and Homoskedasticity

Using the Student t Distribution when n is Small

A big picture review of where we are going...

- We want to learn about the slope of the population regression line:

$$Y_i = \beta_0 + \beta_1 X_i + u_i, \quad i = 1, 2, \ldots, n$$

- We have data from a sample, so we can compute $\widehat{\beta}_0$ and $\widehat{\beta}_1$ but there is sampling uncertainty that needs to be incorporated into the analysis.
- There are five steps towards this goal:
  - State the population object of interest
  - Provide an estimator of this population object
  - Derive the sampling distribution of the estimator (this requires certain assumptions). In large samples this sampling distribution will be normal by the CLT.
  - The square root of the estimated variance of the sampling distribution is the standard error (SE) of the estimator
  - Use the SE to construct t-statistics (for hypothesis tests) and confidence intervals.

- **Object of interest:** β_1 in

$$Y_i = \beta_0 + \beta_1 X_i + u_i, \quad i = 1, 2, \ldots, n$$

- **Estimator:** $\widehat{\beta}_1$

- **The Sampling Distribution of $\widehat{\beta}_1$:** To derive the large-sample distribution of $\widehat{\beta}_1$, we make the following assumptions:

  A1. For any given value of X, the mean of u is zero: $E[u|X] = 0$

  A2. (X_i, Y_i), $i = 1, \ldots, n$ are i.i.d.

  A3. Large outliers in X and/or Y are rare.

  Under the Least Squares Assumptions, for n large, $\widehat{\beta}_1$ is approximately distributed,

$$\widehat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma_v^2/n}{\left(\sigma_X^2\right)^2}\right), \qquad \left(\sigma_v^2 = V\left[\left(X_i - \mu_X\right) u_i\right]\right)$$

- The last two steps (obtaining the SE and constructing t-statistics) will be our main goal in the rest of this section.

Hypothesis Testing and the Standard Error of $\widehat{\beta}_1$

- The objective is to test a hypothesis, like $\beta_1 = \beta_{1,0}$, using data - to reach a tentative conclusion whether the (null) hypothesis is correct or incorrect.
- **General setup:**
  - Null hypothesis and two-sided alternative:

$$H_0 : \beta_1 = \beta_{1,0} \text{ vs. } H_1 : \beta_1 \neq \beta_{1,0}$$

    where $\beta_{1,0}$ is the hypothesized value under the null.
  - Null hypothesis and one-sided alternative:

$$H_0 : \beta_1 \geq \beta_{1,0} \text{ vs. } H_1 : \beta_1 < \beta_{1,0}$$

$$H_0 : \beta_1 \leq \beta_{1,0} \text{ vs. } H_1 : \beta_1 > \beta_{1,0}$$

- **General approach:** construct t-statistic, and compute p-value (or compare to the N(0,1) critical value)

The t-statistics

- In general
$$t = \frac{\text{Estimator - Hypothesized Value}}{\text{Standard Error of the Estimator}}$$
where the SE of the estimator $= \sqrt{\text{estimator of } var(\text{Estimator})}$
- To test hypotheses regarding μ_Y (the mean of Y) we used
$$t = \frac{\overline{Y} - \mu_{Y,0}}{SE(\overline{Y})}$$
where $\mu_{Y,0}$ is the hypothesized value for μ_Y and $SE(\overline{Y}) = s_Y/\sqrt{n}$
- Similarly, to test hypotheses regarding β_1 we use:
$$t = \frac{\widehat{\beta}_1 - \beta_{1,0}}{SE(\widehat{\beta}_1)}$$
where $\beta_{1,0}$ is the hypothesized value for β_1 and
$SE(\widehat{\beta}_1) = \sqrt{\text{estimator of } var(\widehat{\beta}_1)}$
- So our first task is to obtain a formula for $SE(\widehat{\beta}_1)$.

Formula for $SE(\widehat{\beta}_1)$

- Recall the expression for the variance of $\widehat{\beta}_1$ (when n large)

$$\sigma^2_{\widehat{\beta}_1} = \frac{1}{n} \times \frac{\sigma^2_v}{\left(\sigma^2_X\right)^2}, \qquad \left(\sigma^2_v = var\left[v_i\right], \text{ and } v_i = \left(X_i - \mu_X\right) u_i\right)$$

- The estimator of $\sigma^2_{\widehat{\beta}_1}$ should replace the unknown population values of σ^2_v and σ^2_X by estimators constructed from the data:
  i) estimator of σ^2_v :

$$\widehat{\sigma}^2_v = \frac{1}{n-2} \sum_{i=1}^{n} \widehat{v}_i^2$$

  where $\widehat{v}_i = \left(X_i - \overline{X}\right) \widehat{u}_i$

  ii) estimator of σ^2_X :

$$\widehat{\sigma}^2_X = \frac{1}{n} \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2$$

- Combining these two estimators:

$$\widehat{\sigma}^2_{\widehat{\beta}_1} = \frac{1}{n} \times \frac{\text{estimator of } \sigma_v^2}{\left(\text{estimator of } \sigma_X^2\right)^2} = \frac{1}{n} \times \frac{\frac{1}{n-2}\sum_{i=1}^{n} \widehat{v}_i^2}{\left[\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2\right]^2}$$

- The Standard Error of $\widehat{\beta}_1$: $SE(\widehat{\beta}_1) = \sqrt{\widehat{\sigma}^2_{\widehat{\beta}_1}}$
- This is a bit nasty, but it is less complicated than it seems:
  - The numerator estimates $var(v_i)$, and the denominator estimates $\left[\sigma_X^2\right]^2$.
  - Why the degrees-of-freedom adjustment $n-2$? Because two coefficients have to be estimated ($\widehat{\beta}_0$ and $\widehat{\beta}_1$) to compute $\widehat{u}_i$, which in return used to compute $\widehat{v}_i$.
  - $SE(\widehat{\beta}_i)$ is computed by computer software
  - Your regression software has memorized this formula so you dont need to.

Summary: To Test $H_0 : \beta_1 = \beta_{1,0}$ vs. $H_1 : \beta_1 \neq \beta_{1,0}$

- Construct the t-statistic

$$t^a = \frac{\widehat{\beta}_1^a - \beta_{1,0}}{SE(\widehat{\beta}_1)} = \frac{\widehat{\beta}_1 - \beta_{1,0}}{\sqrt{\widehat{\sigma}^2_{\widehat{\beta}_1}}}$$

- Reject at 5% significance if $|t^a| > 1.96$
- The p-value is $p = P\left[|t| > |t^a|\right] \approx 2\Phi(-|t^a|)$: probability in tails of normal outside $|t^a|$. You reject H_0 at 5% significance level if the p-value is less than 5%
- This procedure relies on the large-n approximation that $\widehat{\beta}_1$ is normally distributed; typically $n = 50$ is large enough for the approximation to be excellent.

Example: Test Score - Class Size data (Cont'd)

Regression Output:

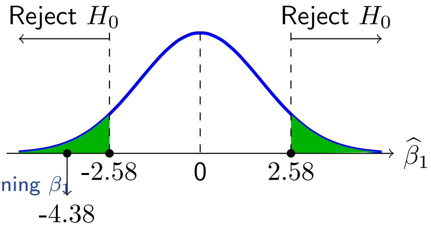
$$\widehat{\text{Test Score}} = \underset{(10.4)}{698.9} - 2.28 \times \underset{(0.52)}{STR}, \quad R^2 = 0.05, \quad SER = 18.6$$

where $SE(\widehat{\beta}_0) = 10.4$ and $SE(\widehat{\beta}_1) = 0.52$

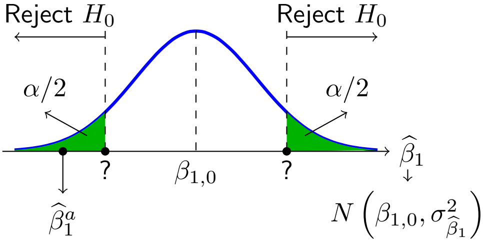
- Hypothesis: $H_0 : \beta_1 = 0$ vs. $H_1 : \beta_1 \neq 0$
- Construct the t-statistic

$$t^a = \frac{\widehat{\beta}_1^a - \beta_{1,0}}{SE(\widehat{\beta}_1)} = \frac{-2.28 - 0}{0.52} = -4.38$$

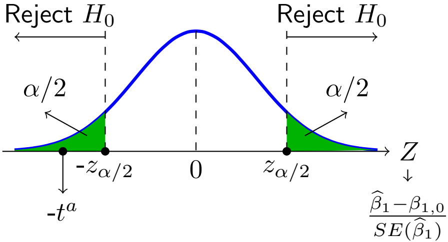
- The 1% 2-sided critical value is 2.58, so we reject the null at the 1% significance level



Pictures ...



Reject H_0
Reject H_0
$\alpha/2$
$\alpha/2$
?
$\beta_{1,0}$
?
$\widehat{\beta}_1$
$\widehat{\beta}_1^a$
$N\left(\beta_{1,0}, \sigma^2_{\widehat{\beta}_1}\right)$



Reject H_0
Reject H_0
$\alpha/2$
$\alpha/2$
-$z_{\alpha/2}$
0
$z_{\alpha/2}$
Z
-t^a
$\frac{\widehat{\beta}_1 - \beta_{1,0}}{SE(\widehat{\beta}_1)}$

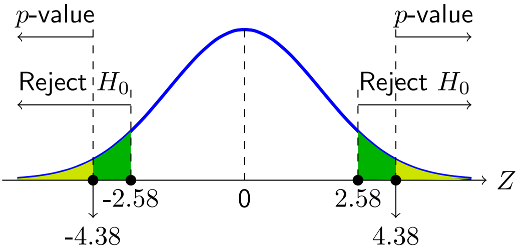
Example: Test Score - Class Size data (Cont'd)

Regression Output:

$$\widehat{\text{Test Score}} = \underset{(10.4)}{698.9} - 2.28 \times \underset{(0.52)}{STR}, \quad R^2 = 0.05, \quad SER = 18.6$$

where $SE(\widehat{\beta}_0) = 10.4$ and $SE(\widehat{\beta}_1) = 0.52$

- Alternatively, we can compute the p-value
- The p-value is $p = P\left[|t| > |t^a|\right] \approx 2\Phi(-|t^a|) = 2\Phi(-4.38) = 0.00001$ which is less than 0.01, so we reject the null.



Confidence Intervals for β_1

- Recall that a 95% C.I., equivalently:
  - The set of points that cannot be rejected at the 5% significance level;
  - A set-valued function of the data (an interval that is a function of the data) that contains the true parameter value 95% of the time in repeated samples.
- Because the t-statistic for β_1 is $N(0,1)$ in large samples, construction of a C.I. for β_1 is just like the case of the sample mean:
  - 90% Confidence Interval for β_1:

$$\widehat{\beta}_1 - 1.64 \times SE(\widehat{\beta}_1) \leq \beta_1 \leq \widehat{\beta}_1 + 1.64 \times SE(\widehat{\beta}_1)$$

  - 95% Confidence Interval for β_1:

$$\widehat{\beta}_1 - 1.96 \times SE(\widehat{\beta}_1) \leq \beta_1 \leq \widehat{\beta}_1 + 1.96 \times SE(\widehat{\beta}_1)$$

  - 99% Confidence Interval for β_1:

$$\widehat{\beta}_1 - 2.58 \times SE(\widehat{\beta}_1) \leq \beta_1 \leq \widehat{\beta}_1 + 2.58 \times SE(\widehat{\beta}_1)$$

- **Interpretation:**(for the second one) 95% of the time the true population slope parameter β_1 will be contained in the set $\{\widehat{\beta}_1 \pm 1.96 \times SE(\widehat{\beta}_1)\}$

Example: Test Score - Class Size data (Cont'd)

$$\widehat{\text{Test Score}} = \underset{(10.4)}{698.9} - 2.28 \times \underset{(0.52)}{STR}, \quad R^2 = 0.05, \quad SER = 18.6$$

- 95% Confidence Interval for β_1:

$$-2.28 - 1.96 \times 0.52 \leq \beta_1 \leq -2.28 + 1.96 \times 0.52$$

$$\implies -3.30 \leq \beta_1 \leq -1.26$$

- The following two statements are equivalent (why?)
  - The 95% confidence interval does not include zero;
  - The hypothesis $H_0 : \beta_1 = 0$ is rejected at the 5% level

Summary of statistical inference about β_0 and β_1

- **Estimation:**
  - OLS estimators $\widehat{\beta}_0$ and $\widehat{\beta}_1$
  - $\widehat{\beta}_0$ and $\widehat{\beta}_1$ have approximately normal sampling distributions in large samples
- **Hypothesis testing:**
  - Hypothesis: $H_0 : \beta_1 = \beta_{1,0}$ vs. $H_1 : \beta_1 \neq \beta_{1,0}$ (two-sided test, there are others too ...)
  - Test Statistic: $t^a = \frac{\widehat{\beta}_1^a - \beta_{1,0}}{SE(\widehat{\beta}_1)}$
  - Compare t^a with the critical values (or use p-value) to conclude the test
- **Confidence intervals:**
  - 95% Confidence Interval for β_1:

$$\widehat{\beta}_1 - 1.96 \times SE(\widehat{\beta}_1) \leq \beta_1 \leq \widehat{\beta}_1 + 1.96 \times SE(\widehat{\beta}_1)$$

  - This is the set of β_1 that is not rejected at the 5% level
  - The 95% C.I. contains the true β_1 in 95% of all samples.

Regression when X is Binary

- Sometimes a regressor is binary:
  - $X = \begin{cases} 1, & \text{if small class} \\ 0, & \text{if not} \end{cases}$
  - $X = \begin{cases} 1, & \text{if female} \\ 0, & \text{if male} \end{cases}$
  - $X = \begin{cases} 1, & \text{if treated (experimental drug)} \\ 0, & \text{if not} \end{cases}$
- Binary regressors are sometimes called dummy variables.
- So far, β_1 has been called a slope, but that doesnt make sense if X is binary.
- How do we interpret regression with a binary regressor?

Interpreting regressions with a binary regressor

$$Y_i = \beta_0 + \beta_1 X_i + u_i, \qquad X = \begin{cases} 1, & \text{if treated} \\ 0, & \text{if not} \end{cases}$$

- When $X_i = 1 \implies Y_i = \beta_0 + \beta_1 + u_i$
  - The mean of Y_i is $\beta_0 + \beta_1$, that is

$$E[Y_i|X_i = 1] = \beta_0 + \beta_1$$

- When $X_i = 0 \implies Y_i = \beta_0 + u_i$
  - The mean of Y_i is β_0, that is

$$E[Y_i|X_i = 0] = \beta_0$$

- Therefore,

$$\begin{aligned} \beta_1 &= E[Y_i|X_i = 1] - E[Y_i|X_i = 0] \\ &= \text{Population Difference in Group Means} \end{aligned}$$

Example

- Tabulation of group means

Class Size	Average Score ($\overline{Y}$)	Std. Dev. (s_Y)	**n**
Small ($STR < 20$)	657.4	19.4	238
Large ($SRT \geq 20$)	650.0	17.9	182

- Difference in means: $\overline{Y}_{\text{small}} - \overline{Y}_{\text{large}} = 657.4 - 650.0 = 7.4$
- Standard Error:

$$SE = \sqrt{\frac{s_s^2}{n_s} + \frac{s_l^2}{n_l}} = \sqrt{\frac{19.4^2}{238} + \frac{17.9^2}{182}} = 1.8$$

Example

- Let $X_i = \begin{cases} 1, & \text{if small class } (STR_i < 20) \\ 0, & \text{if not } (STR_i \geq 20) \end{cases}$
- And specify the population regression model as

$$\text{Test Score}_i = \beta_0 + \beta_1 X_i + u_i$$

- OLS regression gives

$$\widehat{\text{Test Score}}_i = \underset{(1.3)}{650.0} + \underset{(1.8)}{7.4}\, X_i$$

- Interpretation of $\widehat{\beta}_1$: Estimated difference in Test Scores between the districts with small STR and large STR.
- $SE(\widehat{\beta}_1)$ has the usual interpretation
- t-statistics, confidence intervals constructed as usual
- This is another way (an easy way) to do difference-in- means analysis
- The regression formulation is especially useful when we have additional regressors (as we will very soon)

Heteroskedasticity and Homoskedasticity

- What is Heteroskedasticity and Homoskedasticity?
- Consequences of homoskedasticity
- Implication for computing standard errors
- What do these two terms mean?
  - If $var(u|X = x)$ is constant- that is, if the variance of the conditional distribution of u given X does not depend on X - then u is said to be **homoskedastic**. Otherwise, u is **heteroskedastic**.

Example

Example: hetero/homoskedasticity in the case of a binary regressor (that is, the comparison of means)

- Standard error when group variances are **unequal:**

$$SE = \sqrt{\frac{s_s^2}{n_s} + \frac{s_l^2}{n_l}}$$

- Standard error when group variances are **equal:**

$$SE = s_p\sqrt{\frac{s_s^2}{n_s} + \frac{s_l^2}{n_l}}$$

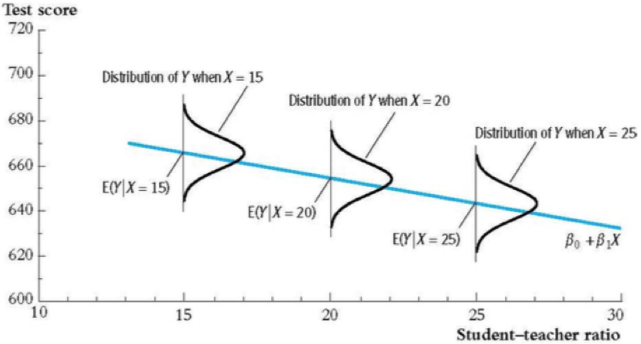
  where

$$s_p^2 = \frac{(n_s - 1)s_s^2 + (n_l - 1)s_l^2}{n_s + n_l - 2}$$

  (s_p = "pooled estimator of σ^2" when $\sigma_s^2 = \sigma_l^2$)

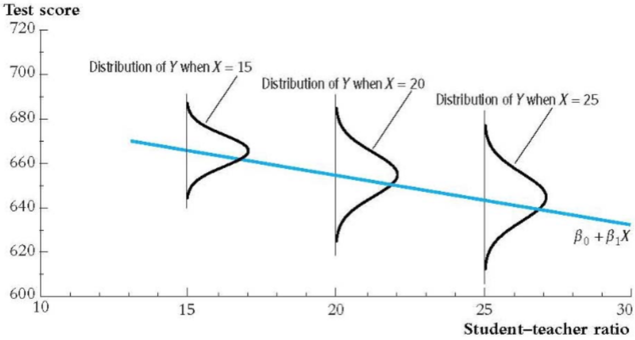
- Equal group variances = homoskedasticity
- Unequal group variances = heteroskedasticity

Homoskedasticity in a Picture



- $E(u|X) = 0$ (u satisfies LSA #1)
- The variance of u **does not** depend on X

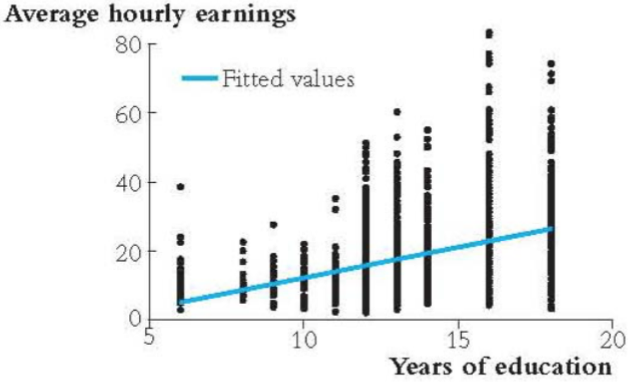
Heteroskedasticity in a Picture



- $E(u|X) = 0$ (u satisfies LSA #1)
- The variance of u **does** depend on X: u is heteroskedastic.

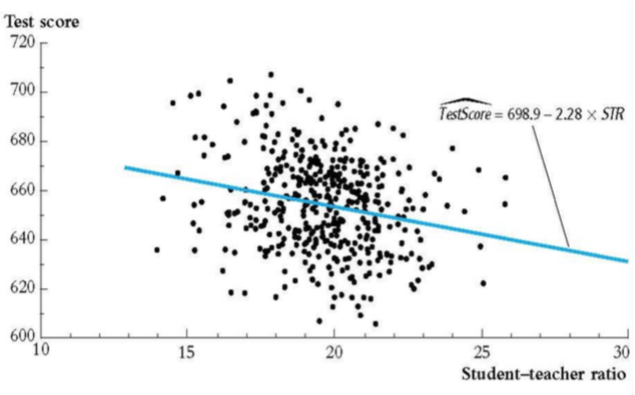
Example: Earnings and Schooling

A real-data example from labor economics: average hourly earnings vs. years of education (data source: Current Population Survey):



- Heteroskedastic or homoskedastic?

Example: The Class Size Data



- Heteroskedastic or homoskedastic?

- So far we have (without saying so) assumed that u might be heteroskedastic.
- Recall the three least squares assumptions:
  A1. For any given value of X, the mean of u is zero: $E[u|X] = 0$

  A2. (X_i, Y_i), $i = 1, \ldots, n$ are i.i.d.

  A3. Large outliers in X and/or Y are rare.
- Heteroskedasticity and homoskedasticity concern $var(u|X = x)$. Because we have not explicitly assumed homoskedastic errors, we have implicitly allowed for heteroskedasticity.

What if the errors are in fact homoskedastic?

- You can prove that OLS has the lowest variance among estimators that are linear in Y
- The formula for the variance of $\widehat{\beta}_1$ and the OLS standard error simplifies:

$$var(\widehat{\beta}_1) = \frac{var\left[(X_i - \mu_X)\, u_i\right]}{\left(n\sigma_X^2\right)^2}, \qquad \text{(general formula)}$$

$$= \frac{\sigma_u^2}{n\sigma_X^2}, \qquad \text{(if } u \text{ is homoscedastic)}$$

- Along with this homoskedasticity-only formula for the variance of $\widehat{\beta}_1$, we have **homoskedasticity-only standard errors:**

$$SE(\widehat{\beta}_1) = \sqrt{\frac{1}{n} \times \frac{\frac{1}{n-2}\sum_{i=1}^{n} \widehat{u}_i^2}{\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}}.$$

Choosing Between two formulas for SE($\widehat{\beta}_1$)

We now have two formulas for standard errors for $\widehat{\beta}_1$.

- **Homoskedasticity-only standard errors** - these are valid only if the errors are homoskedastic.
- **The usual standard errors** - to differentiate the two, it is conventional to call these **heteroskedasticity - robust standard errors**, because they are valid whether or not the errors are heteroskedastic.
- The main advantage of the homoskedasticity-only standard errors is that the formula is simpler. But the disadvantage is that the formula is only correct if the errors are homoskedastic.
- Some people (e.g. Excel programmers) find the homoskedasticity-only formula simpler - but it is wrong unless the errors really are homoskedastic.

Practical implications...

- The homoskedasticity-only formula for the standard error of $\widehat{\beta}_1$ and the "heteroskedasticity-robust" formula differ - so in general, you get different standard errors using the different formulas.
- Homoskedasticity-only standard errors are the default setting in regression software sometimes the only setting (e.g. Excel). To get the general "heteroskedasticity-robust" standard errors you must override the default.
- If you dont override the default and there is in fact heteroskedasticity, your standard errors (and t- statistics and confidence intervals) will be wrong- typically, homoskedasticity-only SEs are too small.

The bottom line:

- If the errors are either homoskedastic or heteroskedastic and you use heteroskedastic-robust standard errors, you are OK
- If the errors are heteroskedastic and you use the homoskedasticity-only formula for standard errors, your standard errors will be wrong (the homoskedasticity-only estimator of the variance of $\widehat{\beta}_1$ is inconsistent if there is heteroskedasticity).
- The two formulas coincide (when n is large) in the special case of homoskedasticity
- So, you should always use heteroskedasticity-robust standard errors.

Inference if u is Homoskedastic and Normally Distributed:

- We are going to derive the sampling distribution of $\widehat{\beta}_1$ under some additional assumptions- precisely two more assumptions
- Extended least squares assumptions:
  A1. For any given value of X, the mean of u is zero: $E[u|X] = 0$
  A2. (X_i, Y_i), $i = 1, \ldots, n$ are i.i.d.
  A3. Large outliers in X and/or Y are rare.
  A4. u is homoskedastic
  A5. u is distributed $N(0, \sigma_u^2)$
- If all five assumptions hold, then:
  - $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are normally distributed **for all n (!)**
  - the t-statistic has a Student t-distribution with $n - 2$ degrees of freedom- this holds exactly for all n (!)

Normality of the Sampling Distribution of $\widehat{\beta}_1$

- We are going to derive the sampling distribution of $\widehat{\beta}_1$ under LSA #1 - #5

$$\widehat{\beta}_1 - \beta_1 = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)u_i}{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}$$

$$= \frac{1}{n}\sum_{i=1}^{n} w_i u_i, \qquad \text{where } w_i = \frac{\left(X_i - \overline{X}\right)}{\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}$$

- What is the distribution of a weighted average of normals?
- Under LSA #1 - #5

$$\widehat{\beta}_1 - \beta_1 \sim N\left(0, \frac{1}{n^2}\left(\sum_{i=1}^{n} w_i^2\right)\sigma_u^2\right)$$

- Substituting w_i into this expression yields the homoskedasticity-only variance formula.

- In addition, under assumptions LSA #1 - #5, under the null hypothesis the t statistic has a Student t distribution with $n-2$ degrees of freedom
- Why $n-2$? Because we estimated 2 parameters, β_0 and β_1
- For $n < 30$, the t critical values can be a fair bit larger than the $N(0,1)$ critical values
- For $n > 50$ or so, the difference in t_{n-2} and $N(0,1)$ distributions is negligible:

d.f.	5% t-distribution critical value
10	2.23
20	2.09
30	2.04
60	2.00
∞	1.96

Practical Implications:

- If $n < 50$ and you really believe that, for your application, u is homoskedastic and normally distributed, then use the t_{n-2} instead of the $N(0,1)$ critical values for hypothesis tests and confidence intervals. Of course, in this case you should also use the homoskedasticity-only SEs to compute t value.
- In most econometric applications, there is no reason to believe that u is homoskedastic and normal- usually, there are good reasons to believe that neither assumption holds.
- Fortunately, in modern applications, $n > 50$, so we can rely on the large-n results presented earlier, based on the CLT, to perform hypothesis tests and construct confidence intervals using the large-n normal approximation.

Summary and Assessment

- The initial policy question:
  - Suppose new teachers are hired so the student-teacher ratio falls by one student per class. What is the effect of this policy intervention (treatment) on test scores?
- Does our regression analysis using the California data set answer this convincingly?
  - Not really- districts with low STR tend to be ones with lots of other resources and higher income families, which provide kids with more learning opportunities outside school...this suggests that $corr(u_i, STR_i) > 0$, so $E(u_i|X_i) \neq 0$.
- It seems that we have omitted some factors, or variables, from our analysis, and this has biased our results...