Economics 6400: Econometrics

Lecture 4: Multiple Regression and Inference

CSU, East Bay

October 13, 2016

Last week...

- We introduced Multiple Regression Analysis:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u$$

- As before, key assumption that is required to obtain unbiased estimates of $\beta_1, \beta_2 \ldots ...., \beta_k$ is:

$$\mathrm{E}(u|x_1, x_2, \ldots, x_k) = 0$$

which implies that *all* of the independent variables are not correlated with the error term

Omitted variable bias

- Consider a population model with two right-hand side variables x_1 and x_2 that are correlated with linear relationship:

$$x_2 = \delta_0 + \delta_1 x_1 + v$$

- Inserting the equation for x_2 into the "true model" yields:

$$\begin{aligned} y =& \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u \\ =& \beta_0 + \beta_1 x_1 + \beta_2(\delta_0 + \delta_1 x_1 + v) + u \\ =& (\beta_0 + \beta_2\delta_0) + (\beta_1 + \beta_2\delta_1)x_1 + (\beta_2 v + u) \end{aligned}$$

- If y is only regressed on x_1 (i.e. x_2 omitted) then:
  - First term is the estimated intercept: $\tilde{\beta}_0 = \hat{\beta}_0 + \hat{\beta}_2\hat{\delta}_0$
  - Second term is the estimated slope coefficient on x_1, and $\hat{\beta}_2\hat{\delta}_1$ is the **omitted variable bias**: $\tilde{\beta}_1 = \hat{\beta}_1 + \hat{\beta}_2\tilde{\delta}_1$
  - Third term is the error term
- All estimated coefficients will be biased!

When are you not in trouble?

- Key relationships:
  - True model: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$
  - Relationship between x_1 and x_2: $x_2 = \delta_0 + \delta_1 x_1 + v$
  - Incorrect and correct estimates of β_1: $\tilde{\beta}_1 = \hat{\beta}_1 + \hat{\beta}_2 \tilde{\delta}_1$

1. If $\beta_2 = 0$ then x_2 is not part of u so $\tilde{\beta}_1 = \hat{\beta}_1$
2. If $\delta_1 = 0$ then x_1 & x_2 are uncorrelated so $\tilde{\beta}_1 = \hat{\beta}_1$

Tracking omitted variable bias

- For simple cases, bias is $\beta_2\delta_1$
- Tracking the direction of bias:

	$\delta_1 > 0$ Corr$(x_1, x_2) > 0$	$\delta_1 < 0$ Corr$(x_1, x_2) < 0$
$\beta_2 > 0$	Positive bias	Negative bias
$\beta_2 < 0$	Negative bias	Positive bias

Omitted variable bias example

- Consider the wage equation

$$wage = \beta_0 + \beta_1 educ + \beta_2 abil + u$$

$$abil = \delta_0 + \delta_1 educ + v$$

- If *abil* omitted from regression then:

$$wage = (\beta_0 + \beta_2\delta_0) + (\beta_1 + \beta_2\delta_1) educ + (\beta_2 v + u)$$

- The return to education β_1 will **upward biased** ($\tilde{\beta}_1 > \hat{\beta}_1$) because $\beta_2\delta_1 > 0$. It will appear that people with many years of education earn very high wages, but this is partly due to the fact that people with more education tend to have greater ability on average ($\delta_1 > 0$)

Omitted variable bias when $k > 2$

- Correlation between a single right-hand side variable and the error generally biases *all* the coefficient estimates.
- Suppose the population model to estimate is:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

  but we omit x_3 and obtain model estimates

$$\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1 + \tilde{\beta}_2 x_2$$

- If $\text{Cov}(x_1, x_3) \neq 0$ but $\text{Cov}(x_2, x_3) = 0$ then
  - $\tilde{\beta}_1$ is biased
  - $\tilde{\beta}_2$ is also biased unless $\text{Cov}(x_1, x_2) = 0$

Roughly approximating direction of the bias when $k > 2$

- Pairwise correlation between right-hand variables makes detecting the direction of the bias difficult
- However, when we are interested in the relationship between a particular explanatory variable then, **as a rough guide only**, we can ignore the other variables and focus on that one variable and the key omitted variable
- Wage equation example:

$$wage = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 abil + u$$

Suppose we omit *abil*. We can focus on the likely bias on $\tilde{\beta}_1$ if we assume that $\text{Cov}(exper, abil) \approx 0$, and $\text{Cov}(educ, exper) \approx 0$

  - Since $\beta_3 > 0$ and *educ* and *abil* are positively correlated, $\tilde{\beta}_1$ will be upward biased

Other regression issues: Perfect collinearity

- Examples:
  1. Relationships between regressors:

     $$voteA = \beta_0 + \beta_1 shareA + \beta_2 shareB + u$$

     There is a perfect linear relationship between the two explanatory variables: $shareA+shareB=1$ ... so you cannot hold *shareB* fixed when increasing *shareA*!

  2. Subtle log relationships:

     $$\log(cons) = \beta_0 + \beta_1 \log(inc) + \beta_2 \log(inc^2) + u$$

     since $\log(inc^2) = 2\log(inc) \Rightarrow x_2 = 2x_1$

- In the case of perfect collinearity, one of the collinear variables will need to be dropped

Perfect collinearity example

```
. reg rd sales_in_millions sales_in_dollars
note: sales_in_millions omitted because of collinearity
```

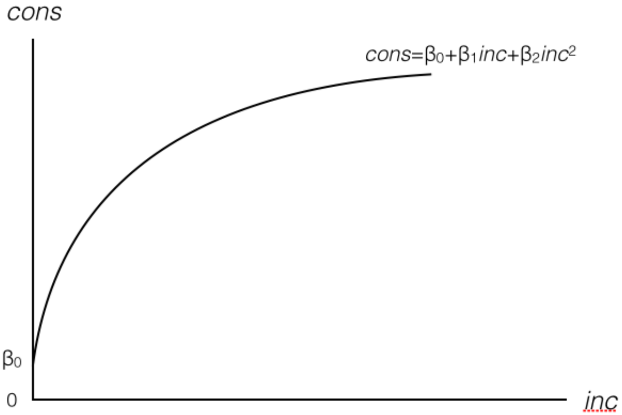
Source	SS	df	MS
Model	2945977.47	1	2945977.47
Residual	321063.086	30	10702.1029
Total	3267040.55	31	105388.405

Number of obs =	32
F(1, 30) =	275.27
Prob > F =	0.0000
R-squared =	0.9017
Adj R-squared =	0.8985
Root MSE =	103.45

rd	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
sales_in_millions	0	(omitted)				
sales_in_dollars	4.06e-08	2.45e-09	16.59	0.000	3.56e-08	4.56e-08
_cons	-.5772171	20.51549	-0.03	0.978	-42.47544	41.321

Family income and family consumption

- Resembles a standard Keynesian consumption curve



Family income and family consumption

- Relationship between consumption (*cons*) and family income (*inc*):

$$cons = \beta_0 + \beta_1 inc + \beta_2 inc^2 + u$$

- Two explanatory factors: income and income squared
- Correct interpretation of the coefficients:

$$\frac{dcons}{dinc} = \beta_1 + 2\beta_2 inc$$

  - Change in consumption if income rises by one unit
  - The change in consumption depends on the current level of family income

Including irrelevant variables in a regression model

- Including irrelevant variables or overspecifying a model will not affect unbiasedness
- Consider the model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

where $\beta_3 = 0$ in the population.

  - Since Assumptions 1-4 hold, $E(\hat{\beta}_3) = \beta_3 = 0$
  - $\hat{\beta}_3$ may not be zero but if enough samples are taken the *average* value will be zero

- Including irrelevant variables may increase sampling variance however

Variance of the OLS estimators

- As before, we add an assumption of homoskedasticity:

$$Var(u_i|x_{i1}, x_{i2}, \ldots, x_{ik}) = \sigma^2$$

- Short-hand notation using vector notation:

$$Var(u_i|\boldsymbol{x}_i) = \sigma^2 \quad \text{with} \quad \boldsymbol{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ik})$$

- Example: wage equation

$$Var(u_i|educ_i, exper_i, tenure_i) = \sigma^2$$

Variance of $\hat{\beta}_j$

- Under the four earlier assumptions plus homoskedasticity:

$$Var(\hat{\beta}_j) = \frac{\sigma^2}{SST_j(1 - R_j^2)}$$

where

- $SST_j = \sum_{i=1}^{n}(x_{ij} - \overline{x}_j)^2$ is the sample variation in right-hand side variable x_j
- R_j^2 is the R^2 from a regression of x_j on the <u>other</u> right-hand side variables (including a constant)

Variance of $\hat{\beta}_j$

- Deriving the variance of $\hat{\beta}_j$:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} \hat{r}_{i1} y_i}{\sum_{i=1}^{n} \hat{r}_{i1}^2} \Rightarrow \hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{n} \hat{r}_{i1} u_i}{\sum_{i=1}^{n} \hat{r}_{i1}^2}$$

where $\hat{r}_{i1}$ are the residuals from a regression of x_1 on $x_2, x_3, \ldots, x_k$

$$\begin{aligned} Var(\hat{\beta}_1|\mathbf{X}) =& \frac{\sum_{i=1}^{n} \hat{r}_{i1}^2 \mathrm{Var}(u_i|\mathbf{X})}{\left(\sum_{i=1}^{n} \hat{r}_{i1}^2\right)^2} \\ =& \frac{\sigma^2}{SSR_1} \\ =& \frac{\sigma^2}{SST_1(1 - R_1^2)} \end{aligned}$$

An aside: Why does partialing out work?

	x1	x2	r1
1	1	1	-.2666667
2	3	2	.8
3	3	3	-.1333333
4	3	4	-1.066667
5	5	5	0
6	7	6	1.066667
7	7	7	.1333333
8	7	8	-.8
9	9	9	.2666667

- r_1 is the residuals from a regression of x_1 on x_2
- r_1 is larger when x_2 does not predict x_1 well
  - This variation in x_1, represented by r_1, can be used to determine the effect of x_1 on y

Components of the variance of $\hat{\beta}_j$

1. The error variance: σ^2
   - A high σ^2 increases the sampling variance because there is more "noise" in the equation
   - A large σ^2 necessarily makes estimates imprecise
   - σ^2 does not decrease with sample size since it is a feature of the population
2. The total sample variation in the explanatory variable: $SST_j = \sum_{i=1}^{n}(x_{ij} - \overline{x}_j)^2$
   - More sample variation leads to more precise estimates
   - Total sample variation automatically increases with the sample size
   - Increasing the sample size is therefore a way to get more precise estimates

Components of the variance of $\hat{\beta}_j$

3. Linear relationships among the independent variables: R_j^2
   - The R^2 of the regression of x_j on the other independent variables will be higher the better x_j can be (linearly) explained by those other independent variables
   - Sampling variance of $\hat{\beta}_j$ will be higher the larger R^2 is
   - The problem of almost linearly dependent variables is called **multicollinearity** (i.e. $R_j^2 \to 1$ for some j).

Multicollinearity example

- Consider a population model of average standardized test scores (*avgscore*) and expenditures across schools:

  $$avgscore = \beta_0 + \beta_1 teachexp + \beta_2 matexp + \beta_3 othexp + \ldots$$

  where

  - *teachexp* is expenditure on teachers
  - *matexp* is expenditures for instructional materials
  - *othexp* is other expenditures
- Each of the expenditures will be strongly correlated
- It will be difficult to estimate the differential effects of different expenditure categories
  - Sampling variance for the coefficients will be large

Multicollinearity "problem"

- In the above example, it would probably make sense to combine all the expenditure categories together since their effects are difficult to disentangle
- In other cases, dropping some independent variables reduces multicollinearity... but this often leads to omitted variable bias!
  - Common practice across social science research =(
- Only the sampling variance of the variables involved in multicollinearity will be inflated; the estimates of the other effects may be very precise
- Multicollinearity is **not** a violation of any of our assumptions
  - With enough data and variation in right-hand side variables, it should not be a problem

Estimating the error variance

- Since we do not observe σ^2, we must obtain an unbiased estimate of it:
$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} \hat{u}_i^2}{n-k-1}$$
- Basis for the degrees of freedom adjustment can be understood by considering the $k+1$ first order conditions for the OLS residuals:
$$\sum_{i=1}^{2} \hat{u}_i = 0, \quad \sum_{i=1}^{n} x_{ij}\hat{u}_i = 0, \quad \text{where } j = 1, 2, \ldots, k.$$
Once you know $n-(k+1)$ of the residuals, you can construct the remaining $k+1$ residuals

Estimating the sampling variance of the OLS estimators

- The true sampling variation of $\hat{\beta}_j$:

$$sd(\hat{\beta}_j) = \sqrt{Var(\hat{\beta}_j)} = \sqrt{\frac{\sigma^2}{SST_j(1 - R_j^2)}}$$

- The **estimated** sampling variation of $\hat{\beta}_j$ or **standard error**:

$$se(\hat{\beta}_j) = \sqrt{\widehat{Var}(\hat{\beta}_j)} = \sqrt{\frac{\hat{\sigma}^2}{SST_j(1 - R_j^2)}}$$

- Formulae are only valid assuming homoskedasticity

Standard errors in Stata output

```
. reg lwage educ exper tenure

      Source |       SS       df       MS              Number of obs =     526
-------------+------------------------------           F(  3,   522) =   80.39
       Model |  46.8741776     3  15.6247259           Prob > F      =  0.0000
    Residual |  101.455574   522  .194359337           R-squared     =  0.3160
-------------+------------------------------           Adj R-squared =  0.3121
       Total |  148.329751   525   .28253286           Root MSE      =  .44086

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |    .092029   .0073299    12.56   0.000     .0776292    .1064288
       exper |   .0041211   .0017233     2.39   0.017     .0007357    .0075065
      tenure |   .0220672   .0030936     7.13   0.000     .0159897    .0281448
       _cons |   .2843595   .1041904     2.73   0.007     .0796756    .4890435
------------------------------------------------------------------------------
```

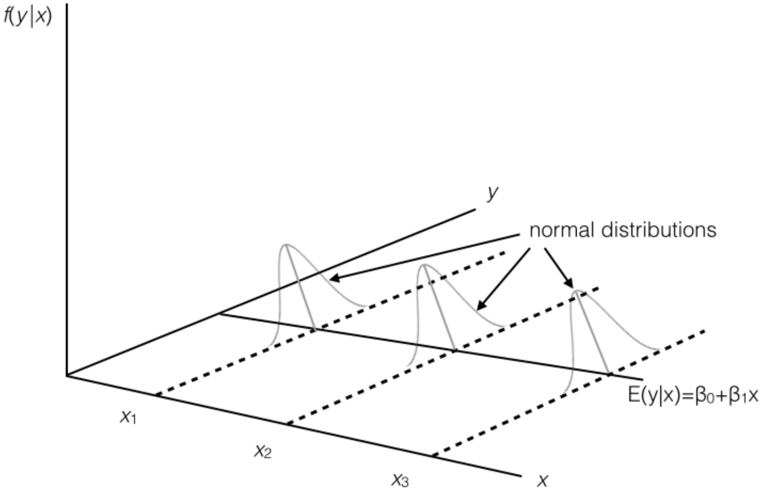
Gauss-Markov Assumptions

- The four assumptions used to establish unbiasedness and the fifth assumption to derive the variance formulas are known as the Gauss-Markov Assumptions:
  1. Linearity: $y = \beta_0 + \beta_1 x + u$
  2. We have a **random** sample of size n, $\{(x_i, y_i) : i = 1, 2, \ldots, n\}$
  3. Sample variation of explanatory variable: The sample outcomes on x are not all the same value
  4. Zero conditional mean: $E(u|x) = 0$
  5. $\text{Var}(u|x_1, x_2 \ldots, x_k) = \sigma^2$
- Under these assumptions, the OLS estimators are the **best linear unbiased estimators (BLUEs)** of the regression coefficients
  - **Best** means smallest variance

Inference

- Having estimated population parameters, we now want to be able to conduct hypothesis tests
  - For example: does gender have no impact on wages (i.e. $\beta_{female} = 0$)?
- To be able to conduct statistical inference we need to make an additional (sixth) assumption:
  6. The population of the error u is normaly distributed with zero mean and variance σ^2: $u \sim \text{Normal}(0, \sigma^2)$.
- These six assumptions are called the **classical linear model (CLM) assumptions**
  - $y|x \sim \mathrm{Normal}(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k, \sigma^2)$

Simple regression model under homoskedasticity and normally distributed errors



Why the normal assumption?

- Basis: u is the sum of many different unobserved factors affecting y so we can invoke the central limit theorem to conclude u has an approximate normal distribution
- Whether normality holds is ultimately an empirical question
  - If the dependent variable cannot be negative (e.g. wage) then normality cannot hold
  - If the dependent variable is *count data* (i.e. 0, 1, 2, 3... with zero being the modal value) then normality cannot hold

Testing hypotheses for a single population parameter

- Note that β_j is an unknown (but real) population parameter, which we try to infer (hypothesize about) using our estimate of this parameter $\hat{\beta}_j$
- Since we do not know σ^2 (and estimate it with $\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} \hat{u}^2}{n-k-1}$) we must use the t distribution instead of the normal distribution:
$$\frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \sim t_{n-k-1} = t_{df}$$
- If n is large, the t distribution is close to the standard normal distribution

t table

Critical Values of the t Distribution

		Significance Level				
	1-Tailed:	**.10**	**.05**	**.025**	**.01**	**.005**
	2-Tailed:	**.20**	**.10**	**.05**	**.02**	**.01**
Degrees of Fre	1	3.078	6.314	12.706	31.821	63.657
	2	1.886	2.920	4.303	6.965	9.925
	3	1.638	2.353	3.182	4.541	5.841
	4	1.533	2.132	2.776	3.747	4.604
	5	1.476	2.015	2.571	3.365	4.032
	6	1.440	1.943	2.447	3.143	3.707
	7	1.415	1.895	2.365	2.998	3.499
	8	1.397	1.860	2.306	2.896	3.355
	9	1.383	1.833	2.262	2.821	3.250
	10	1.372	1.812	2.228	2.764	3.169
	11	1.363	1.796	2.201	2.718	3.106
	12	1.356	1.782	2.179	2.681	3.055
	13	1.350	1.771	2.160	2.650	3.012
	14	1.345	1.761	2.145	2.624	2.977
	15	1.341	1.753	2.131	2.602	2.947
	16	1.337	1.746	2.120	2.583	2.921
	17	1.333	1.740	2.110	2.567	2.898
	18	1.330	1.734	2.101	2.552	2.878
	19	1.328	1.729	2.093	2.539	2.861
	20	1.325	1.725	2.086	2.528	2.845

Testing the null hypothesis $\beta_j = 0$

- In the majority of applications, we will be testing the null hypothesis:

$$\mathrm{H}_0 : \beta_j = 0$$

- This means that **after controlling for the other right-hand side variables**, there is no effect of x_j on y
- Example: $Pr(lung\ cancer) = \beta_0 + \beta_1 cigs + \beta_2 drinks + u$
  - $\mathrm{H}_0 : \beta_2 = 0$
  - After controlling for cigarette consumption (*cigs*), does the number of *drinks* affect the probability of *lung cancer*?

Testing the null hypothesis $\beta_j = 0$

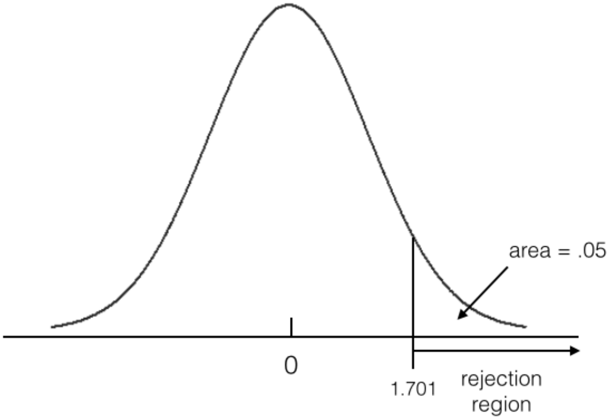
- t-statistic used to test this hypothesis is

$$t_{\hat{\beta}_j} \equiv \frac{\hat{\beta}_j}{se(\hat{\beta}_j)}$$

- The further away from zero $\hat{\beta}_j$ is (i.e. larger numerator), the less likely the null hypothesis will hold
- To determine how far away is "enough," we divide the estimate by its standard deviation
  - Recall that under a normal distribution, 95% of the population lies two standard deviations either side of the mean
- Goal: Define a rejection rule (large enough t-statistic, c) such that if $t_{\hat{\beta}_j}$ exceeds c then H_0 is rejected only with a small probability (or significance level, usually 5%)

Testing against a one-sided alternative

- Test $\mathrm{H}_0 : \beta_j = 0$ vs. $\mathrm{H}_1 : \beta_j > 0$
- If significance level is 5% and $df = 28$ then reject if $t_{\hat{\beta}_j} > c = 1.701$



One-sided alternative wage example

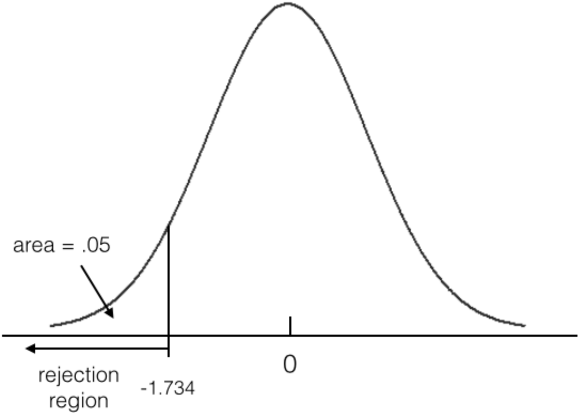
- Each year of experience increases wage by 0.4% (not large)
- Test $\mathrm{H}_0 : \beta_{exper} = 0$ vs. $\mathrm{H}_1 : \beta_{exper} > 0$

lwage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	.092029	.0073299	12.56	0.000	.0776292	.1064288
exper	.0041211	.0017233	2.39	0.017	.0007357	.0075065
tenure	.0220672	.0030936	7.13	0.000	.0159897	.0281448
_cons	.2843595	.1041904	2.73	0.007	.0796756	.4890435

- $t_{exper} = \frac{0.0041}{0.0017} \approx 2.39$
- Critical values: $c_{0.05} = 1.645, c_{0.01} = 2.326\ldots$ so reject H_0 at 5% and 1% levels

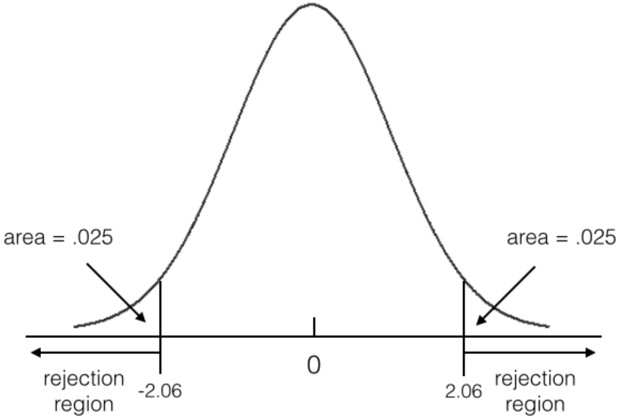
Testing against a one-sided alternative

- Test $\mathrm{H}_0 : \beta_j = 0$ vs. $\mathrm{H}_1 : \beta_j < 0$
- If significance level is 5% and $df = 18$ then reject if $t_{\hat{\beta}_j} < c = -1.734$



Testing against a two-sided alternative

- Test $\mathrm{H}_0 : \beta_j = 0$ vs. $\mathrm{H}_1 : \beta_j \neq 0$
- If significance level is 5% and $df = 25$ then reject if $t_{\hat{\beta}_j} < c = -2.06$ or $t_{\hat{\beta}_j} > c = 2.06$



Determinants of college GPA

- Model to estimate is:

$$colGPA = \beta_0 + \beta_1 hsGPA + \beta_2 ACT + \beta_3 skipped + u$$

where *hsGPA* is school GPA, *skipped* is # of lectures missed

colGPA	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
hsGPA	.4118162	.0936742	4.40	0.000	.2265819	.5970505
ACT	.0147202	.0105649	1.39	0.166	-.0061711	.0356115
skipped	-.0831131	.0259985	-3.20	0.002	-.1345234	-.0317028
_cons	1.389554	.3315535	4.19	0.000	.7339295	2.045178

- $t_{ACT} = \frac{0.0147}{0.0106} \approx 1.39$, $t_{skipped} = \frac{-0.0831}{0.0260} \approx -3.20$
- Critical values: $c_{0.05} = \pm 1.960$, $c_{0.01} = \pm 2.576$
- So reject H_0 for *skipped* but not for *ACT*

Statistical significance

- If a regression coefficient is different from zero in a two-sided test, the corresponding variable is said to be **statistically significant**
  - The vast majority of social science research seeks to establish statistically significant relationships between variables
- If the number of degrees of freedom is large so that the normal distribution is a good approximation for the t-distribution then the following rules of thumb apply
  - $|t_{\hat{\beta}_j}| > 1.645 \Rightarrow$ "statistically significant at 10% level"
  - $|t_{\hat{\beta}_j}| > 1.96 \Rightarrow$ "statistically significant at 5% level"
  - $|t_{\hat{\beta}_j}| > 2.576 \Rightarrow$ "statistically significant at 1% level"

Economic versus statistical significance

- If a variable is statistically significant, consider its magnitude (size) to get an idea if it is large enough to be important
- The mere fact that a coefficient is statistically significant does not mean it is economically significant
- Often coefficients with low t statistics will have the "wrong" sign; these can be ignored

Testing more general hypotheses

- Supposing we wished to test whether β_j is equal to some non-zero value

$$\begin{aligned} \mathrm{H}_0 &: \beta_j = a_j \\ \mathrm{H}_1 &: \beta_j \neq a_j \end{aligned}$$

- Test statistic is now:

$$t_{\hat{\beta}_j} \equiv \frac{\hat{\beta}_j - a_j}{se(\hat{\beta}_j)}$$

where we simply subtract the hypothesized value

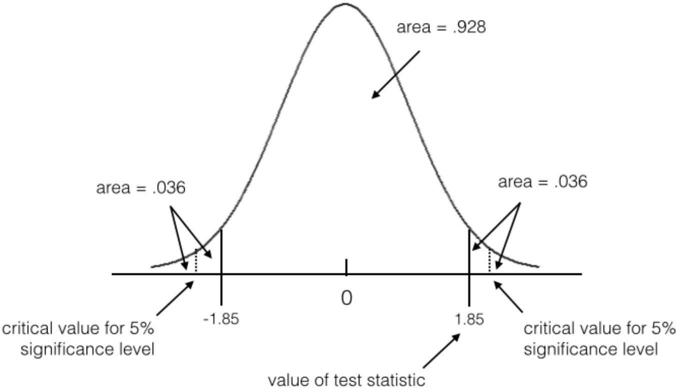
Computing p-value for t-tests

- The smallest significance level at which the null hypothesis is still rejected is called the p-**value** of the hypothesis test
- A small p-value (i.e. 0.05, 0.01, 0.001) is evidence against the null hypothesis because you would not reject the null hypothesis at small significance levels
- A large p-value is evidence in favor of the null hypothesis
- p-values are more informative than tests at fixed significance levels

Testing against a two-sided alternative

- *p*-value is the probability that the *t*-distributed variable takes on a larger absolute value than the realized value of the test statistic:

$$P(|T| > 1.85) = 2(.036) = 0.072$$



Determinants of college GPA

colGPA	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
hsGPA	.4118162	.0936742	4.40	0.000	.2265819	.5970505
ACT	.0147202	.0105649	1.39	0.166	-.0061711	.0356115
skipped	-.0831131	.0259985	-3.20	0.002	-.1345234	-.0317028
_cons	1.389554	.3315535	4.19	0.000	.7339295	2.045178

- p-value is 0.166 for *ACT* and 0.002 for *skipped*
  - Easy to reject H_0 for *ACT* at conventional levels but *skipped* is statistically significant

Confidence intervals

- For a 95 percent confidence interval:

$$P\left(\hat{\beta}_j - c_{0.05} \cdot se(\hat{\beta}_j) \leq \beta_j \leq \hat{\beta}_j + c_{0.05} \cdot se(\hat{\beta}_j)\right) = 0.95$$

- In repeated samples the interval constructed as above will cover the population regression coefficient in 95% of cases
- Relationship between confidence intervals and hypothesis tests:

$$a_j \notin \text{ interval} \Rightarrow \text{reject } \mathrm{H}_0 : \beta_j = a_j$$

Determinants of college GPA

colGPA	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
hsGPA	.4118162	.0936742	4.40	0.000	.2265819	.5970505
ACT	.0147202	.0105649	1.39	0.166	-.0061711	.0356115
skipped	-.0831131	.0259985	-3.20	0.002	-.1345234	-.0317028
_cons	1.389554	.3315535	4.19	0.000	.7339295	2.045178

- 95% confidence interval for *skipped* with $n - k - 1 = 137$ *df* is

$$\hat{\beta}_{skipped} \pm c_{0.05} \cdot se(\hat{\beta}_j)$$
$$-0.0831 \pm 1.98 \cdot 0.0260$$
$$-0.0831 \pm 0.0515$$

Next week

- Inference continued (Chapter 4)