EC2C4: Econometrics II

Multiple Regression: Inference

Ragvir Sabharwal

LSE

Lent Term 2022/23

Introduction

Recap

- So far, what do we know about regression model?
- MLR.1: $y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u$
- MLR.2: random sampling from population
- MLR.3: no perfect collinearity in sample
- MLR.4: $E(u|x_1, \ldots, x_k) = E(u) = 0$ (exogenous regressor)
- MLR.5: $Var(u|x_1, \ldots, x_k) = Var(u) = \sigma^2$ (homoskedasticity)

- (1) Algebraic properties of OLS estimators for any sample, regression anatomy formula, goodness-of-fit R^2, and interpretation of OLS regression line

- (2) Unbiasedness of OLS under MLR.1-4 and omitted variable bias (failure of MLR.4)

- (3) Formula for $Var(\hat{\beta}_j|\mathbf{X})$ under MLR.1-5

Sampling distributions of OLS estimators (Wooldridge, Ch. 4.1)

Testing hypotheses on β_j

- We now want to **test hypotheses about** β_j. Hypothesise that β_j takes certain value, then use data to determine whether to reject the hypothesis or not

- For example, based on ATTEND.dta

$$final = \beta_0 + \beta_1 missed + \beta_2 priGPA + \beta_3 ACT + u$$

where ACT is achievement test score. Null hypothesis that missing lecture has no effect on final exam performance (after controlling for prior GPA and ACT score) is

$$H_0 : \beta_1 = 0$$

What we know about $\hat{\beta}_j$

- To test hypotheses about β_j, we need to know more than just mean and variance of $\hat{\beta}_j$
- Under MLR.1-4, we can compute expected value as

$$E(\hat{\beta}_j) = \beta_j$$

- Under MLR.1-5, we know variance is

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

and $\hat{\sigma}^2 = SSR/(n - k - 1)$ is an unbiased estimator of σ^2

What we want: Sampling distribution of $\hat{\beta}_j$

- Hypothesis testing requires entire **sampling distribution** of $\hat{\beta}_j$. Even under MLR.1-5, sample distributions can be virtually anything
- Write
$$\hat{\beta}_j = \beta_j + \sum_{i=1}^{n} w_{ij} u_i$$
where w_{ij}'s are functions of $\mathbf{X}$
- Conditional on $\mathbf{X}$, the distribution of $\hat{\beta}_j$ is determined by that of $(u_1, \ldots, u_n)$

Assumption MLR.6

- **Assumption MLR.6 (Normality)**

  Error term u is independent of $(x_1, \ldots, x_k)$ and is **normally distributed with mean zero and variance** σ^2

  $$u \sim Normal(0, \sigma^2)$$

- MLR.6 implies MLR.4: $E(u|x_1, \ldots, x_k) = E(u) = 0$
- Also MLR.6 implies MLR.5: $Var(u|x_1, \ldots, x_k) = Var(u) = \sigma^2$
- Now MLR.6 imposes full independence between u and $(x_1, \ldots, x_k)$ (not just mean and variance independence)
- By MLR.6, we now impose very specific distributional assumption for u: the familiar bell-shaped curve

Your turn

- Suppose

$$z \sim Normal(E(z), Var(z))$$

  for $E(z) \neq 0$

- Which is true?
  - A: $\frac{z - E(z)}{Var(z)} \sim Normal(0, 1)$
  - B: $\frac{z - E(z)}{\sqrt{Var(z)}} \sim Normal(0, 1)$
  - C: $\frac{z}{Var(z)} \sim Normal(0, 1)$

Important fact about normal random variables

- Linear combination of normal random variables is also normally distributed
- Because u_i's are independent and identically distributed (called **iid**) as $Normal(0, \sigma^2)$ and $\hat{\beta}_j = \beta_j + \sum_{i=1}^{n} w_{ij} u_i$, we have

$$\hat{\beta}_j | \mathbf{X} \sim Normal\left(\beta_j, Var(\hat{\beta}_j | \mathbf{X})\right)$$

where $\mathbf{X}$ are data for all regressors and we already know the formula for $Var(\hat{\beta}_j | \mathbf{X})$

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

Theorem: Normal sampling distribution

- Under Assumptions MLR.1-6

$$\hat{\beta}_j|\mathbf{X} \sim Normal\left(\beta_j, Var(\hat{\beta}_j|\mathbf{X})\right)$$

  and so

$$\frac{\hat{\beta}_j - \beta_j}{sd(\hat{\beta}_j)} \sim Normal(0,1)$$

- The second result follows from property of normal distribution: if $W \sim Normal$, then $a + bW \sim Normal$ for constants a and b

- Under MLR.1-5, standardized random variable

$$\frac{\hat{\beta}_j - \beta_j}{sd(\hat{\beta}_j)}$$

  always has zero mean and variance one. Under MLR.6, it is also normally distributed

- Notice that second result holds **even when we do not condition on X**

Testing hypotheses about a single population parameter (Wooldridge, Ch. 4.2)

Obtaining a test statistic

- We cannot directly use the result

$$\frac{\hat{\beta}_j - \beta_j}{sd(\hat{\beta}_j)} \sim Normal(0,1)$$

  to test hypotheses about β_j because $sd(\hat{\beta}_j)$ depends on unknown $\sigma = \sqrt{Var(u)}$

- So replace σ with $\hat{\sigma}$ (i.e. replace $sd(\hat{\beta}_j)$ with standard error $se(\hat{\beta}_j)$)

Theorem: t distribution for standardised estimator

- Under Assumptions MLR. 1-6

$$\frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \sim t_{n-k-1} = t_{df}$$

- We will not prove this as the argument is somewhat involved
- Due to replacement of σ with $\hat{\sigma}$, the distribution changes from standard normal to t distribution

t distribution

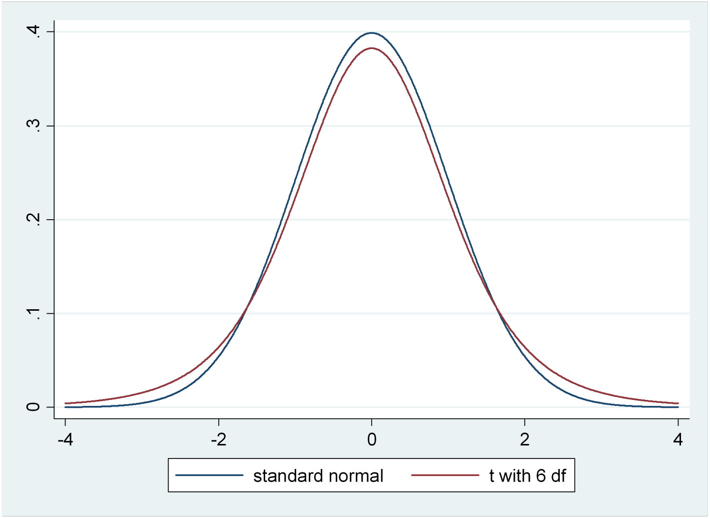
- t distribution also has bell shape but is more spread out than $Normal(0, 1)$

$$
\begin{aligned}
E(t_{df}) &= 0 \text{ if } df > 1 \\
Var(t_{df}) &= \frac{df}{df - 2} > 1 \text{ if } df > 2
\end{aligned}
$$

- If $df = 10$, then $Var(t_{df}) = 1.25$ (25% larger than variance of $Normal(0, 1)$)
- If $df = 120$, then $Var(t_{df}) \approx 1.017$ (only 1.7% larger)
- As $df \to \infty$

$$t_{df} \to Normal(0, 1)$$

Graph of $N(0, 1)$ and t_6



t statistic

- We use result on t distribution to test null hypothesis that x_j has no partial effect on y

$$H_0 : \beta_j = 0$$

- To test $H_0 : \beta_j = 0$, we use t **statistic** (or t **ratio**)

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

- We measure **how far** $\hat{\beta}_j$ **is from zero relative to its standard error**

- Because $se(\hat{\beta}_j) > 0$, $t_{\hat{\beta}_j}$ has same sign as $\hat{\beta}_j$. To test $H_0 : \beta_j = 0$, we need alternative hypothesis

Testing against one-sided alternatives

- First consider the alternative

$$H_1 : \beta_j > 0$$

which means the null is effectively

$$H_0 : \beta_j \leq 0$$

- If we reject $\beta_j = 0$ then reject any $\beta_j < 0$ too
- We often just state $H_0 : \beta_j = 0$ and act like we do not care about negative values

- If $\hat{\beta}_j < 0$, it provides no evidence against H_0 in favor of $H_1 : \beta_j > 0$
- If $\hat{\beta}_j > 0$, the question is: How big does $t_{\hat{\beta}_j} = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)}$ have to be to conclude H_0 is not a credible hypothesis?

Traditional approach to hypothesis testing

- (1) Choose null hypothesis $H_0 : \beta_j = 0$ (or $H_0 : \beta_j \leq 0$)
- (2) Choose alternative hypothesis $H_1 : \beta_j > 0$
- (3) Choose **significance level** for the test. That is, probability of rejecting H_0 when it is in fact true (Type I Error). Suppose we use 5%, so probability of committing Type I error is .05
- (4) Choose critical value c so that **rejection rule**

$$t_{\hat{\beta}_j} > c$$

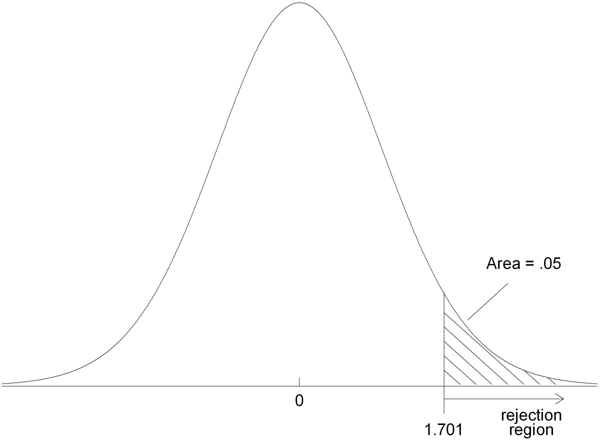
leads to 5% level test

How to get critical value

- Key: **Under the null hypothesis** $H_0 : \beta_j = 0$

$$t_{\hat{\beta}_j} \sim t_{n-k-1} = t_{df}$$

- Use this to obtain critical value c
- Suppose $df = 28$ and 5% significance level. Critical value is $c = 1.701$ (from Table G.2)
- Following picture shows how to find c for **one-tailed test**


Area = .05
0
1.701
rejection region

Rejection rule

- So, with $df = 28$, rejection rule for $H_0 : \beta_j = 0$ against $H_1 : \beta_j > 0$, at 5% level, is
$$t_{\hat{\beta}_j} > 1.701$$
We need t statistic greater than 1.701 to conclude there is enough evidence against H_0
- If $t_{\hat{\beta}_j} \leq 1.701$, we fail to reject H_0 against H_1 at 5% significance level

Different significance level

- Suppose $df = 28$, but we want to carry out test at different significance level (often 10% or 1% level)

$$\begin{aligned} c_{.10} &= 1.313 \\ c_{.05} &= 1.701 \\ c_{.01} &= 2.467 \end{aligned}$$

- If we want to reduce probability of Type I error, we must **increase** critical value (so we reject the null less often)
- If we reject at, say, 1% level, then we must also reject at any **larger** level
- If we fail to reject at, say, 10% level (i.e. $t_{\hat{\beta}_j} \leq 1.313$), then we will fail to reject at any **smaller** level

- With large sample sizes, we can use critical values from standard normal distribution. These are $df = \infty$ entry in Table G.2

$$\begin{aligned} c_{.10} &= 1.282 \\ c_{.05} &= 1.645 \\ c_{.01} &= 2.326 \end{aligned}$$

which we can round to 1.28, 1.65, and 2.33, respectively. The value 1.65 is especially common for one-tailed test

Example: Does ACT score help to predict college GPA (GPA1.dta)

- Model: $colGPA = \beta_0 + \beta_{hsGPA} hsGPA + \beta_{ACT} ACT + u$
  Null hypothesis is $H_0 : \beta_{ACT} = 0$

```
. reg colGPA hsGPA ACT

      Source |       SS       df       MS              Number of obs =     141
-------------+------------------------------           F(  2,   138) =   14.78
       Model |  3.42365506     2  1.71182753           Prob > F      =  0.0000
    Residual |  15.9824444   138  .115814814           R-squared     =  0.1764
-------------+------------------------------           Adj R-squared =  0.1645
       Total |  19.4060994   140  .138614996           Root MSE      =  .34032

------------------------------------------------------------------------------
      colGPA |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       hsGPA |   .4534559   .0958129     4.73   0.000     .2640047    .6429071
         ACT |    .009426   .0107772     0.87   0.383    -.0118838    .0307358
       _cons |   1.286328   .3408221     3.77   0.000      .612419    1.960237
------------------------------------------------------------------------------
```

- From output, $\hat{\beta}_{ACT} = .0094$ and $t_{ACT} = .87$. Even at 10% level ($c = 1.28$), we cannot reject H_0 against $H_1 : \beta_{ACT} > 0$

- Because we fail to reject $H_0 : \beta_{ACT} = 0$, we say that "$\hat{\beta}_{ACT}$ is **statistically insignificant** at 10% level against one-sided alternative"

- Note that estimated effect of ACT is also small. Three more points (slightly more than one standard deviation) only predicts $.0094(3) \approx .028$ increase in *colGPA*

- By contrast, $\hat{\beta}_{hsGPA} = .453$ is large in practical sense and $t_{hsGPA} = 4.73$ is very large. So "$\hat{\beta}_{hsGPA}$ is **statistically significant**" at very small significance levels

Your turn

- Which of the following can cause the usual t test above invalid?
  (a) Heteroskedasticity
  (b) Correlation coefficient of .95 between two regressors
  (c) Omitting an important variable

  - A: All of them can invalidate
  - B: Only (a) can invalidate
  - C: Only (c) can invalidate
  - D: Two of them can invalidate

Again, Your turn

- What is a consequence of using the invalid t test with 5% significance level, say?
  - A: Critical value is too large
  - B: Critical value is too small
  - C: 5% significance level is wrong
  - D: Conclusion (reject or not) is always wrong

Negative one-sided alternative

- For negative one-sided alternative

$$H_1 : \beta_j < 0$$

  we must see significantly negative value for t statistic to reject $H_0 : \beta_j = 0$ in favor of $H_1 : \beta_j < 0$

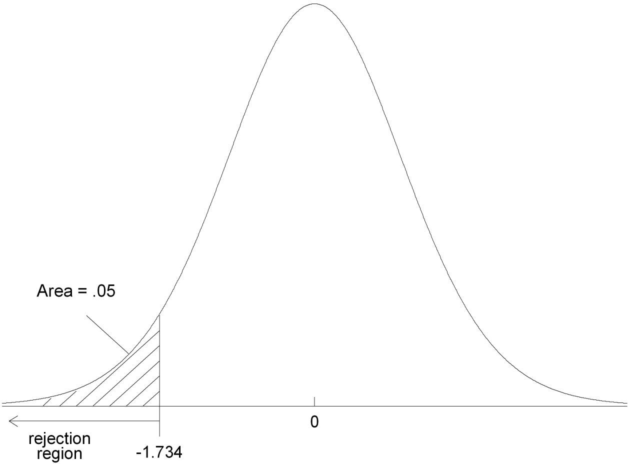
- So the rejection rule is

$$t_{\hat{\beta}_j} < -c$$

  where c is chosen in the same way as in positive case

- With $df = 18$ and 5% level, critical value is $c = -1.734$, so rejection rule is

$$t_{\hat{\beta}_j} < -1.734$$


Area = .05
0
rejection region
-1.734

Testing against two-sided alternatives

- Sometimes we do not know ahead of time whether a variable definitely has positive or negative effect
- Even in the example

$$final = \beta_0 + \beta_1 missed + \beta_2 priGPA + \beta_3 ACT + u$$

  it is conceivable that missing class helps final exam performance (extra time is used for studying, say)
- In this case, null and alternative are

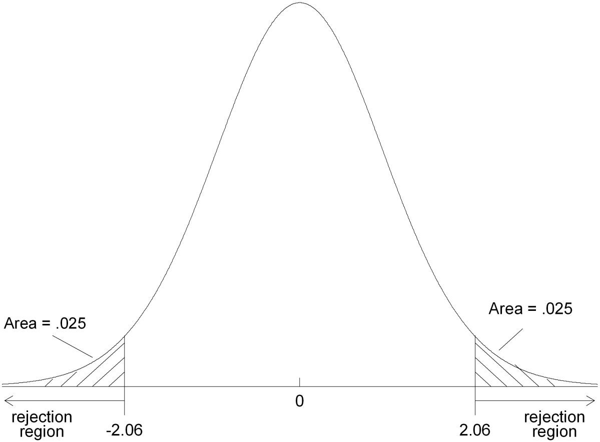
$$\begin{aligned} H_0 &: \beta_j = 0 \\ H_1 &: \beta_j \neq 0 \end{aligned}$$

Rejection rule

- Now we reject if $\hat{\beta}_j$ is sufficiently large in magnitude either positive or negative
- We again use t statistic $t_{\hat{\beta}_j} = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)}$, but now rejection rule is

$$|t_{\hat{\beta}_j}| > c$$

- This results in **two-tailed test** and critical values are given by Table G.2
- For example, if $df = 25$ and 5% level, two-tailed c is 2.06 (97.5-th percentile in t_{25} distribution)
- On the other hand, one-tailed c is 1.71 (95-th percentile in t_{25} distribution)


Area = .025
Area = .025
0
rejection region
-2.06
2.06
rejection region

Example: Factors affecting math pass rates (MEAP93.dta)

- Regress from *math*10 on *lnchprg*, *lsalary*, *enroll*

```
. des math10 lnchprg lsalary enroll

              storage   display      value
variable name   type    format       label      variable label
-----------------------------------------------------------------------------------

math10          float   %9.0g                   perc studs passing MEAP math
lnchprg         float   %9.0g                   perc of studs in sch lnch prog
lsalary         float   %9.0g                   log(salary)
enroll          int     %9.0g                   school enrollment
```

- A priori, we might expect *lnchprg* to has negative effect (it is essentially school-level poverty rate) and *lsalary* to has positive effect. But we can still test against two-sided alternative to avoid specifying alternative ahead of time. *enroll* is clearly ambiguous

- Since $n = 408$, we use standard normal critical values: $c_{.10} = 1.65$, $c_{.05} = 1.96$, and $c_{.01} = 2.58$

```
. reg  math10 lnchprg lsalary enroll

      Source |       SS       df       MS              Number of obs =     408
-------------+------------------------------           F(  3,   404) =   29.60
       Model |  8075.34004     3  2691.78001           Prob > F      =  0.0000
    Residual |  36741.8404   404  90.9451496           R-squared     =  0.1802
-------------+------------------------------           Adj R-squared =  0.1741
       Total |  44817.1805   407  110.115923           Root MSE      =  9.5365

------------------------------------------------------------------------------
      math10 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     lnchprg |  -.2878203   .0380285    -7.57   0.000    -.3625787   -.2130618
     lsalary |   7.969246    3.75663     2.12   0.034     .5842628    15.35423
      enroll |  -.0001741   .0001991    -0.87   0.382    -.0005656    .0002173
       _cons |  -50.69248   39.03804    -1.30   0.195    -127.4355    26.05057
------------------------------------------------------------------------------
```

- Coefficients of *lnchprg* and *lsalary* have anticipated signs. So we easily reject $H_0 : \beta_{lnchprg} = 0$ against $H_1 : \beta_{lnchprg} \neq 0$. Also we reject $H_0 : \beta_{lsalary} = 0$ against $H_1 : \beta_{lsalary} \neq 0$ at 5% level, but not for 1% level.
- *enroll* is different. $|t_{enroll}| = 0.87 < 1.65$, so we fail to reject H_0 at even 10% level

Your turn

- Suppose you do not reject $H_0 : \beta_j = 0$ against two-sided alternative $H_1 : \beta_j \neq 0$ at the 5 % significance level. Based on this and $\hat{\beta}_j > 0$, can we conclude about one-sided test for $\tilde{H}_0 : \beta_j = 0$ against $\tilde{H}_1 : \beta_j > 0$ at 5% level?
  - A: We do not reject $\tilde{H}_0$
  - B: We reject $\tilde{H}_0$
  - C: Not enough information to conclude

Testing other hypotheses about β_j

- Testing the null $H_0 : \beta_j = 0$ is by far most common. That is why Stata automatically reports t statistic for this hypothesis
- It is critical to remember that

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

  is **only for** $H_0 : \beta_j = 0$

- What if we want to test different null value? For example, in constant-elasticity consumption function

$$\log(cons) = \beta_0 + \beta_1 \log(inc) + \beta_2 famsize + \beta_3 pareduc + u$$

we might want to test

$$H_0 : \beta_1 = 1$$

i.e. income elasticity is one (we are pretty sure that $\beta_1 > 0$)

Testing for $H_0 : \beta_j = a_j$

- More generally, suppose the null is

$$H_0 : \beta_j = a_j$$

  where we specify the value a_j (usually zero but in above example $a_j = 1$)

- It is easy to extend t statistic

$$t = \frac{\hat{\beta}_j - a_j}{se(\hat{\beta}_j)}$$

- This t statistic measures how far our estimate $\hat{\beta}_j$ is from the hypothesized value a_j relative to $se(\hat{\beta}_j)$

General expression for t test

- General expression for t test is

$$t = \frac{\textit{estimate} - \textit{hypothesized value}}{\textit{standard error}}$$

- Alternative can be one-sided or two-sided
- We choose critical values in exactly same way as before

Example: Crime and enrollment on college campuses (CAMPUS.dta)

- Bivariate regression

$$\begin{aligned} \log(crime) &= \beta_0 + \beta_1 \log(enroll) + u \\ H_0 &: \beta_1 = 1 \\ H_1 &: \beta_1 > 1 \end{aligned}$$

```
. des crime enroll

              storage   display      value
variable name   type    format       label      variable label
-----------------------------------------------------------------------------------

crime           int     %9.0g                   total campus crimes
enroll          float   %9.0g                   total enrollment
```

```
. reg lcrime lenroll

      Source |       SS       df       MS              Number of obs =      97
-------------+------------------------------           F(  1,    95) =  133.79
       Model |  107.083654     1  107.083654           Prob > F      =  0.0000
    Residual |  76.0358244    95  .800377098           R-squared     =  0.5848
-------------+------------------------------           Adj R-squared =  0.5804
       Total |  183.119479    96  1.90749457           Root MSE      =  .89464

------------------------------------------------------------------------------
      lcrime |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     lenroll |    1.26976    .109776    11.57   0.000     1.051827    1.487693
       _cons |   -6.63137    1.03354    -6.42   0.000    -8.683206   -4.579533
------------------------------------------------------------------------------
```

- We get $\hat{\beta}_1 = 1.27$. So 1% increase in enrollment is estimated to increase crime by 1.27% (so more than 1%). Is this estimate statistically greater than one?

- Although we cannot pull t statistic from output, we can compute it by hand

$$t = \frac{(1.270 - 1)}{.110} \approx 2.45$$

- We have $df = 97 - 2 = 95$. So use $df = 120$ entry in Table G.2. Since $c_{.01} = 2.36$, we reject at 1% level

- Alternatively, we can let Stata do the work using `lincom` ("linear combination" command)

Computing p-values for t tests

- In traditional approach for testing, we choose significance level ahead of time. This can be cumbersome
- Plus, it can hide some information. Even if we reject at 5% level with $c_{.05} = 1.645$, the t statistic of 2 and 4 might convey different information
- Instead of fixing level ahead of time, it is better to answer the following question: Given the observed value of t statistic, **what is the smallest significance level at which we can reject H_0?**
- Such smallest level is known as p-**value**. It allows us to carry out test at any level

- One way to think about p-values is that **it uses the observed statistic as critical value**, and then finds significance level of the test using that critical value

- It is most common to report p-values for two-sided alternatives (this is what Stata does)

- For t testing against two-sided alternative

$$p\text{-value} = P(|T| > |t|)$$

  where t is the value of t statistic and T is a random variable with t_{df} distribution

Interpretation of p-value

- Perhaps the best way to think about p-values: **it is probability of observing the statistic as extreme as we did if H_0 is true**
- So smaller p-values provide more are evidence against the null. For example, if p-value $= .50$, then there is 50% chance of observing t as large as we did (in absolute value). This is not enough evidence against H_0
- If p-value $= .001$, then the chance of seeing t statistic as extreme as we did is .1%. We can conclude that we got very rare sample or that the null hypothesis is highly unlikely

- From

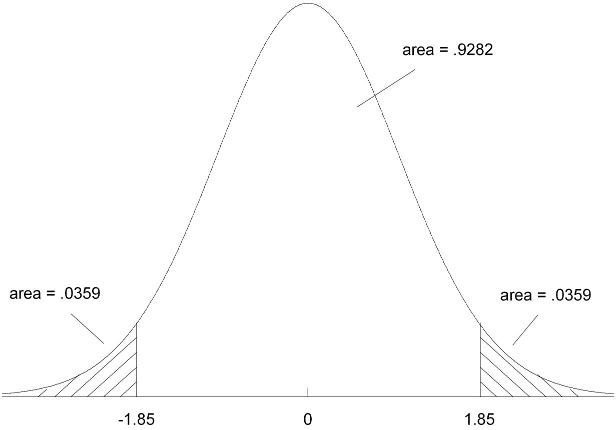
$$p\text{-value} = P(|T| > |t|)$$

we see that as $|t|$ increases p-value decreases. Large absolute t statistics are associated with small p-values

- Suppose $df = 40$ and, from our data, we obtain $t = 1.85$ or $t = -1.85$. Then

$$p\text{-value} = P(|T| > 1.85) = 2P(T > 1.85) = 2(.0359) = .0718$$

where $T \sim t_{40}$. Finding actual number requires Stata


area = .9282
area = .0359
area = .0359
-1.85
0
1.85

Test by *p*-value

- Given *p*-value, we can carry out test at any significance level. If α is chosen level, then

$$\text{Reject } H_0 \text{ if } p\text{-value} < \alpha$$

- For example, in previous example we got p-value $= .0718$. This means we reject H_0 at 10% level but not 5% level. We reject at 8% but not at 7%

- Knowing p-value $= .0718$ is clearly much better than just saying "I fail to reject at 5% level"

Computing p-values for one-sided alternatives

- In Stata, two-sided p-values for $H_0 : \beta_j = 0$ are given in the column labeled "P>|t|"
- With caveat, one sided p-value is given by

$$\text{one-sided } p\text{-value} = \frac{\text{two-sided } p\text{-value}}{2}$$

- We only want the area in one tail, not two tails
- The caveat is: estimated coefficient should be in the direction of the alternative, otherwise one-sided p-value would be above .50

Example: Factors affecting NBA salaries (NBASAL.dta)

```
. des wage games avgmin points rebounds assists

              storage  display     value
variable name   type   format      label      variable label
------------------------------------------------------------------------------------------
wage            float  %9.0g                  annual salary, thousands $
games           byte   %9.0g                  average games per year
avgmin          float  %9.0g                  minutes per game
points          float  %9.0g                  points per game
rebounds        float  %9.0g                  rebounds per game
assists         float  %9.0g                  assists per game

. sum wage games avgmin points rebounds assists

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        wage |       269    1423.828    999.7741        150       5740
       games |       269    65.72491    18.85111          3         82
      avgmin |       269    23.97925    9.731177   2.888889   43.08537
      points |       269    10.21041    5.900667        1.2       29.8
    rebounds |       269    4.401115    2.892573         .5       17.3
-------------+--------------------------------------------------------
     assists |       269    2.408922    2.092986          0       12.6
```

```
. reg lwage games avgmin points rebounds assists
```

Source	SS	df	MS
Model	90.2698185	5	18.0539637
Residual	117.918945	263	.448361006
Total	208.188763	268	.776823743

Number of obs	=	269
F(5, 263)	=	40.27
Prob > F	=	0.0000
R-squared	=	0.4336
Adj R-squared	=	0.4228
Root MSE	=	.6696

lwage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
games	.0004132	.002682	0.15	0.878	-.0048679	.0056942
avgmin	.0302278	.0130868	2.31	0.022	.0044597	.055996
points	.0363734	.0150945	2.41	0.017	.0066519	.0660949
rebounds	.0406795	.0229455	1.77	0.077	-.0045007	.0858597
assists	.0003665	.0314393	0.01	0.991	-.0615382	.0622712
_cons	5.648996	.1559075	36.23	0.000	5.34201	5.955982

- Except for intercept, none of variables is statistically significant at 1% level against two-sided alternative. The closest is *points* with p-value $= .017$ (One-sided p-value is $.017/2 = .0085 < .01$, so it is significant at 1% level against positive one-sided alternative)

- *avgmin* is statistically significant at 5% level because p-value $= .022 < .05$

- *rebounds* is statistically significant at 10% level (against two-sided alternative) because p-value $= .077 < .10$, but not at 5% level. But one-sided p-value is $.077/2 = .0385$

- Both *games* and *assists* have very small t statistics, which lead to p-values close to one (for example, for *assists*, p-value $= .991$). These variables are statistically insignificant

Practical versus statistical significance

- t testing is purely about **statistical significance**. It does not directly speak to the issue of whether a variable has practically or economically large effect
- **Practical or economic significance** depends on the size (and sign) of $\hat{\beta}_j$
- **Statistical significance** depends on $t_{\hat{\beta}_j}$
- It is possible that although the estimate $\hat{\beta}_j$ indicates practically large effect, the estimate is so imprecise that it is statistically insignificant. This is especially an issue with small data sets
- Even more importantly, it is possible to get estimates that are statistically significant but are not practically large. This can happen with very large data sets

Confidence intervals
(Wooldridge, Ch. 4.3)

Confidence interval

- Rather than just testing hypotheses about parameters it is also useful to construct **confidence intervals** (CIs, also known as **interval estimators**)
- Instead of so-called looking at a "point estimator" as we have done so far, we now consider a range of values as our estimator for an unknown population parameter. It takes into account the uncertainty associated with a point estimator
- We will only consider CIs of the form

$$\hat{\beta}_j \pm c \cdot se(\hat{\beta}_j)$$

where $c > 0$ is chosen based on **confidence level**

- We will use 95% confidence level, in which case c comes from 97.5 percentile in t_{df} distribution. In other words, c is 5% critical value against two-sided alternative

- Stata automatically reports at 95% CI for each parameter, based on t distribution using appropriate df

Example (NBASAL.dta)

```
. reg lwage games avgmin points rebounds assists
```

Source	SS	df	MS
Model	90.2698185	5	18.0539637
Residual	117.918945	263	.448361006
Total	208.188763	268	.776823743

Number of obs =	269
F(5, 263) =	40.27
Prob > F =	0.0000
R-squared =	0.4336
Adj R-squared =	0.4228
Root MSE =	.6696

lwage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
games	.0004132	.002682	0.15	0.878	-.0048679	.0056942
avgmin	.0302278	.0130868	2.31	0.022	.0044597	.055996
points	.0363734	.0150945	2.41	0.017	.0066519	.0660949
rebounds	.0406795	.0229455	1.77	0.077	-.0045007	.0858597
assists	.0003665	.0314393	0.01	0.991	-.0615382	.0622712
_cons	5.648996	.1559075	36.23	0.000	5.34201	5.955982

- Notice how three estimates that are not statistically different from zero at 5% level – *games*, *rebounds*, and *assists* – all have 95% CIs that include zero. For example, 95% CI for $\beta_{rebounds}$ is

$$[-.0045, .0859]$$

- By contrast, 95% CI for β_{points} is

$$[.0067, .0661]$$

  which excludes zero

Interpretation of CI

- Properly interpreting CI is a bit tricky. We often see statements like "there is 95% chance that β_{points} is in interval $[.0067, .0661]$." This is incorrect. β_{points} is some fixed value, and it either is or is not in the interval

- Correct way to interpret CI is to remember that the endpoints $\hat{\beta}_j \pm c \cdot se(\hat{\beta}_j)$ change with each sample (i.e. endpoints are random outcomes)

- What 95% CI means is that **in hypothetically repeated random sampling, the interval we compute using the rule $\hat{\beta}_j \pm c \cdot se(\hat{\beta}_j)$ will include the value β_j in 95% of cases**. But for a particular sample we will never know whether β_j is in the interval or not

CI and hypothesis testing

- By 95% CI for β_j, we can test any null value against two-sided alternative at 5% level. Consider

$$\begin{aligned} H_0 &: \beta_j = a_j \\ H_1 &: \beta_j \neq a_j \end{aligned}$$

- (1) If a_j is in 95% CI, then we fail to reject H_0 at 5% level
- (2) If a_j is not in 95% CI then we reject H_0 in favor of H_1 at 5% level

Example (WAGE2.dta)

```
. reg lwage educ IQ exper meduc

      Source |       SS           df       MS      Number of obs   =       857
-------------+----------------------------------   F(4, 852)       =     46.89
       Model |   26.949693         4  6.73742325   Prob > F        =    0.0000
    Residual |  122.411347       852  .14367529    R-squared       =    0.1804
-------------+----------------------------------   Adj R-squared   =    0.1766
       Total |   149.36104       856  .174487197   Root MSE        =    .37905

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0547499   .0077049     7.11   0.000      .039627    .0698727
          IQ |   .0054188   .0010253     5.28   0.000     .0034064    .0074312
       exper |   .0222246   .0034192     6.50   0.000     .0155134    .0289357
       meduc |   .0126417   .0049645     2.55   0.011     .0028977    .0223857
       _cons |   5.108905   .1264354    40.41   0.000     4.860744    5.357066
------------------------------------------------------------------------------
```

- 95% CI for β_{IQ} is about $[.0034, .0074]$. So we can reject $H_0 : \beta_{IQ} = 0$ against two-sided alternative at 5% level. We cannot reject $H_0 : \beta_{IQ} = .005$ (although it is close)

- We can reject return to schooling of 3.5% as being too low, but also 7% is too high

- Just as with hypothesis testing, these CIs are only valid under CLM assumptions. If we have omitted key variables, $\hat{\beta}_j$ is biased. If error variance is not constant, standard errors are improperly computed

Your turn

- What is a consequence of using the invalid confidence interval (say, 95%)?
  - A: CI is too wide
  - B: CI is too narrow
  - C: 95% confidence level is wrong
  - D: Estimator $\hat{\beta}_j$ is biased

Testing a linear restriction involving many parameters (Wooldridge, Ch. 4.4)

Testing linear restriction

- So far, we discussed hypothesis testing only on parameter β_j. But some hypotheses involve **many parameters**
- For example, mother and father's education have same effects on $\log(wage)$? Based on WAGE2.dta, consider

$$\begin{aligned}\log(wage) &= \beta_0 + \beta_1 meduc + \beta_2 feduc \\ &\quad + \beta_3 educ + \beta_4 exper + u\end{aligned}$$

$$\begin{aligned}H_0 &: \beta_1 = \beta_2 \\ H_1 &: \beta_1 \neq \beta_2\end{aligned}$$

Test statistic

- Remember general way to construct t statistic

$$t = \frac{estimate - hypothesized\ value}{standard\ error}$$

- By OLS estimates $\hat{\beta}_1$ and $\hat{\beta}_2$,

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

- Problem: Stata gives us $\hat{\beta}_1$ and $\hat{\beta}_2$ and their standard errors, but that is **not** enough to obtain $se(\hat{\beta}_1 - \hat{\beta}_2)$

Your turn

- Consider two random variables z_1 and z_2. Which is correct expression for $Var(z_1 + z_2)$?
  - A: $Var(z_1 + z_2) = Var(z_1) + Var(z_2)$
  - B: $Var(z_1 + z_2) = Var(z_1) + Var(z_2) + Cov(z_1, z_2)$
  - C: $Var(z_1 + z_2) = Var(z_1) + Var(z_2) - Cov(z_1, z_2)$
  - D: $Var(z_1 + z_2) = Var(z_1) + Var(z_2) + 2Cov(z_1, z_2)$

- Note

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

- Standard error is estimate of its square root

$$se(\hat{\beta}_1 - \hat{\beta}_2) = \{[se(\hat{\beta}_1)]^2 + [se(\hat{\beta}_2)]^2 - 2s_{12}\}^{1/2}$$

  where s_{12} is estimate of $Cov(\hat{\beta}_1, \hat{\beta}_2)$. This is the piece we are missing

- Stata will report s_{12} if we ask, but calculating $se(\hat{\beta}_1 - \hat{\beta}_2)$ is cumbersome. It is easier to use the command `lincom`

- There is also trick of rewriting the model (see Ch. 4.4)

Example (WAGE2.dta)

```
. reg lwage meduc feduc educ exper

      Source |       SS           df       MS      Number of obs   =       722
-------------+----------------------------------   F(4, 717)       =     33.62
       Model |  20.0299189         4  5.00747974   Prob > F        =    0.0000
    Residual |  106.781997       717  .148928866   R-squared       =    0.1579
-------------+----------------------------------   Adj R-squared   =    0.1533
       Total |  126.811916       721  .175883378   Root MSE        =    .38591

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       meduc |   .0116931   .0063125     1.85   0.064    -.0007001    .0240864
       feduc |    .011543   .0055606     2.08   0.038      .000626      .02246
        educ |   .0653767   .0077995     8.38   0.000     .0500641    .0806892
       exper |   .0233539   .0037799     6.18   0.000     .0159329    .0307749
       _cons |   5.397237   .1261321    42.79   0.000     5.149604    5.644869
------------------------------------------------------------------------------
```

- Note that $\hat{\beta}_{meduc} - \hat{\beta}_{feduc} = .0117 - .0115 = .0002$, so difference is very small

```
. lincom meduc - feduc

 ( 1)  meduc - feduc = 0
```

lwage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
(1)	**.0001502**	**.0102829**	**0.01**	**0.988**	**-.020038**	**.0203384**

- Two-sided p-value is .988. Obviously cannot reject $H_0 : \beta_{meduc} = \beta_{feduc}$
- Of course, nothing changes (except sign of estimate) if we use $\beta_{feduc} - \beta_{meduc}$

Testing multiple linear restrictions (Wooldridge, Ch. 4.5)

Testing joint hypotheses

- t test allows us to test single hypothesis, whether it involves one or more than one parameter
- But we sometimes want to test more than one hypothesis
- Generally, it is not valid to look at individual t statistics. We need statistic used to test **joint hypotheses**

Example: Major league baseball salaries (MLB1.dta)

- Consider the model

$$\log(salary) = \beta_0 + \beta_1 years + \beta_2 gamesyr + \beta_3 bavg + \beta_4 hrunsyr + \beta_5 rbisyr + u$$

```
. des salary years gamesyr bavg hrunsyr rbisyr

              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------------------
salary          float  %9.0g                  1993 season salary
years           byte   %9.0g                  years in major leagues
gamesyr         float  %9.0g                  games per year in league
bavg            float  %9.0g                  career batting average
hrunsyr         float  %9.0g                  home runs per year
rbisyr          float  %9.0g                  rbis per year
```

- H_0 : Once we control for experience (*years*) and amount played (*gamesyr*), actual performance has no effect on salary

$$H_0 : \beta_3 = 0,\ \beta_4 = 0,\ \beta_5 = 0$$

- To test H_0, we need **joint (multiple) hypotheses test**
- In this case, we only consider alternative

$$H_1 : H_0 \text{ is not true}$$

i.e. **at least one of** β_3, β_4 and β_5 is different from zero

- One-sided alternatives (where, say, each β is positive) are hard to deal with for multiple restrictions. So, we focus on two-sided alternative

```
. reg lsalary years gamesyr bavg hrunsyr rbisyr

      Source |       SS       df       MS              Number of obs =     353
-------------+------------------------------           F(  5,   347) =  117.06
       Model |  308.989208     5  61.7978416           Prob > F      =  0.0000
    Residual |  183.186327   347  .527914487           R-squared     =  0.6278
-------------+------------------------------           Adj R-squared =  0.6224
       Total |  492.175535   352  1.39822595           Root MSE      =  .72658

------------------------------------------------------------------------------
     lsalary |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       years |   .0688626   .0121145     5.68   0.000     .0450355    .0926898
     gamesyr |   .0125521   .0026468     4.74   0.000     .0073464    .0177578
        bavg |   .0009786   .0011035     0.89   0.376    -.0011918     .003149
     hrunsyr |   .0144295    .016057     0.90   0.369    -.0171518    .0460107
      rbisyr |   .0107657    .007175     1.50   0.134    -.0033462    .0248776
       _cons |   11.19242   .2888229    38.75   0.000     10.62435    11.76048
------------------------------------------------------------------------------
```

- None of three performance variables is statistically significant, even though estimates are all positive
- **Question:** By these insignificant t statistics, should we conclude that none of *bavg*, *hrunsyr*, and *rbisyr* affects baseball player salaries? No. This would be a mistake
- Because of severe multicollinearity (sample correlation between *hrunsyr* and *rbisyr* is about .89), individual coefficients, especially on *hrunsyr* and *rbisyr*, are imprecisely estimated. So we need a joint test

Construct F statistic

- In the general model

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

  we want to test that the last q variables can be excluded

$$H_0 : \beta_{k-q+1} = 0, \ldots, \beta_k = 0$$

- Original model is called **unrestricted model**. When we impose H_0, we get

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_{k-q} x_{k-q} + u$$

  which is called **restricted model**

- Denote SSR from unrestricted and restricted models by SSR_{ur} and SSR_r, respectively. We construct test statistic by comparing SSR_{ur} and SSR_r

- We know that SSR never decreases when regressors are dropped, i.e.

$$SSR_r \geq SSR_{ur}$$

- We ask: **does SSR increase enough** to conclude the restrictions by H_0 are false?

- **F statistic** does degrees of freedom adjustment. In general,

$$F = \frac{(SSR_r - SSR_{ur})/(df_r - df_{ur})}{SSR_{ur}/df_{ur}} = \frac{(SSR_r - SSR_{ur})/q}{SSR_{ur}/(n-k-1)}$$

  where q is number of exclusion restrictions imposed under null and k is number of regressors in unrestricted model (in baseball example, $q = 3$ and $k = 5$)

Rejection rule

- Note that $F \geq 0$ and rejection rule is of the form

$$F > c$$

  where c is appropriately chosen critical value

- We obtain c using (hard to show) fact that, under H_0 (q exclusion restrictions)

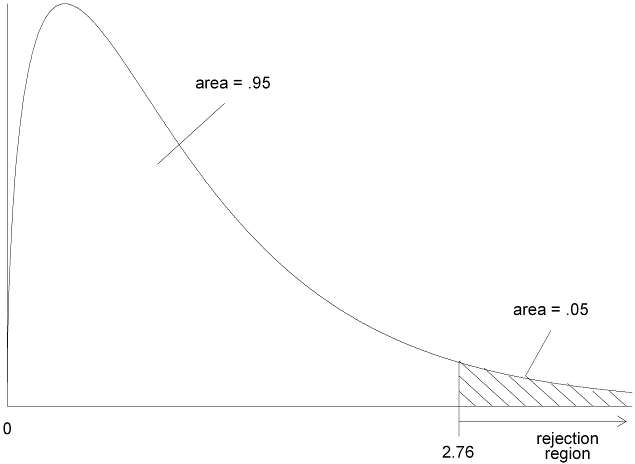
$$F \sim F_{q,n-k-1}$$

  i.e. F distribution with $(q, n-k-1)$ degrees of freedom

- Terminology

$$\begin{aligned} q &= \textbf{numerator df} = df_r - df_{ur} \\ n-k-1 &= \textbf{denominator df} = df_{ur} \end{aligned}$$

- Tables G.3a, G.3b, and G.3c contain critical values for 10%, 5%, and 1% significance levels
- Suppose $q = 3$ and $n - k - 1 = df_{ur} = 60$. Then 5% critical value is 2.76


area = .95
area = .05
0
2.76
rejection region

Example (MB1.dta)

- In MLB example with $n = 353$, $k = 5$, and $q = 3$, we have numerator df $= 3$, denominator df $= 347$. Because the denominator df is above 120, we use "∞" entry. 10% cv is 2.08, 5% cv is 2.60, and 1% cv is 3.78

- As with t testing, it is better to compute p-value, which is reported by Stata after every `test` command

- F statistic for excluding *bavg*, *hrunsyr*, and *rbisyr* from the model is 9.55. This is well above 1% critical value, so we reject at 1% level. In fact, to four decimal places, p-value is zero

- We say that *bavg*, *hrunsyr*, and *rbisyr* are **jointly statistically significant**

- F statistic does not tell us which of the coefficients are different from zero. And t statistics do not help much in this example

```
. reg lsalary years gamesyr bavg hrunsyr rbisyr
```

Source	SS	df	MS
Model	308.989208	5	61.7978416
Residual	183.186327	347	.527914487
Total	492.175535	352	1.39822595

Number of obs =	353
F(5, 347) =	117.06
Prob > F =	0.0000
R-squared =	0.6278
Adj R-squared =	0.6224
Root MSE =	.72658

lsalary	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
years	.0688626	.0121145	5.68	0.000	.0450355	.0926898
gamesyr	.0125521	.0026468	4.74	0.000	.0073464	.0177578
bavg	.0009786	.0011035	0.89	0.376	-.0011918	.003149
hrunsyr	.0144295	.016057	0.90	0.369	-.0171518	.0460107
rbisyr	.0107657	.007175	1.50	0.134	-.0033462	.0248776
_cons	11.19242	.2888229	38.75	0.000	10.62435	11.76048

```
. test bavg hrunsyr rbisyr

 ( 1)  bavg = 0
 ( 2)  hrunsyr = 0
 ( 3)  rbisyr = 0

       F(  3,   347) =    9.55
            Prob > F =    0.0000
```

R^2 form of F statistic

- It is useful to be able to compute F statistic using standard output
- R^2 is reported most of the time, unlike SSR. F statistic for exclusion restrictions can be computed entirely from R^2 for restricted and unrestricted models
- Key: Because same dependent variable is used

$$
\begin{aligned}
SSR_r &= (1 - R_r^2)SST \\
SSR_{ur} &= (1 - R_{ur}^2)SST
\end{aligned}
$$

- Simple algebra shows

$$F = \frac{(R_{ur}^2 - R_r^2)/q}{(1 - R_{ur}^2)/(n-k-1)}$$

- Note: $R_{ur}^2 \geq R_r^2$ which implies $F \geq 0$

F statistic for overall significance of regression

```
. reg ecolbs ecoprc regprc hhsize faminc age educ
```

Source	SS	df	MS
Model	169.099052	6	28.1831753
Residual	4035.03777	653	6.17923089
Total	4204.13682	659	6.3795703

Number of obs =	660
F(6, 653) =	4.56
Prob > F =	0.0002
R-squared =	0.0402
Adj R-squared =	0.0314
Root MSE =	2.4858

ecolbs	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ecoprc	-2.861237	.5919913	-4.83	0.000	-4.023673	-1.6988
regprc	3.006077	.7123078	4.22	0.000	1.607387	4.404767
hhsize	.0630935	.0677804	0.93	0.352	-.0700003	.1961874
faminc	.0021952	.0028653	0.77	0.444	-.0034311	.0078216
age	.0013894	.0067634	0.21	0.837	-.0118914	.0146701
educ	.0343134	.0453141	0.76	0.449	-.0546655	.1232923
_cons	1.05677	.8926501	1.18	0.237	-.6960404	2.809581

- F statistic in upper right corner of output tests very special null hypothesis. In the model

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

  the null is that **all** slope coefficients are zero

$$H_0 : \beta_1 = 0, \ldots, \beta_k = 0$$

- This means none of x_j's helps to explain y
- If we cannot reject this null, we find no factors that explain y

- For this test, $R_r^2 = 0$ (no explanatory variables under H_0), and $R_{ur}^2 = R^2$ from the regression. So, F statistic is

$$F = \frac{R^2/k}{(1-R^2)/(n-k-1)} = \frac{R^2}{(1-R^2)} \cdot \frac{(n-k-1)}{k}$$

  i.e. F is directly related to R^2. As R^2 increases, so does F

- Increasing n increases F. Increasing k decreases F

Dummy variables (Wooldridge, Ch. 7)

Single dummy variable

- For example

$$wage = \beta_0 + \delta_0 female + u$$

- Under SLR.4 $E(u|female) = 0$,

$$E(wage|female) = \beta_0 + \delta_0 female$$

- Average *wage* for men is β_0, average *wage* for women is $\beta_0 + \delta_0$, and δ_0 is difference in average *wage* between women and men

- Inference methods in Ch. 4 directly apply

Interactions among dummy variables

- For two dummy variables, say *female* and *married*, we may consider

$$wage = \beta_0 + \beta_1 female + \beta_2 married + \beta_3 female \cdot married + u$$

  which can estimate average wage for four categories (single male, single female, married male, married female)

- Interaction with quantitative variable provides slope dummy: e.g.

$$wage = \beta_0 + \beta_1 female + \beta_2 exper + \beta_3 female \cdot exper + u$$

- Again inference methods in Ch. 4 directly apply

Testing for differences in regression functions across groups

- We necessarily get same estimated intercepts and slopes if we estimate regressions **separately** for men and women
- Null hypothesis that there is no difference in *wage* between men and women at same levels of experience is

$$H_0 : \beta_1 = 0,\ \beta_3 = 0$$

  which can be tested by F-test
- This is a version of **Chow test** for equality of regression functions across two groups. We test joint significance of dummy variable defining groups as well as interaction terms

Chow test

- In general k variable case, we can define dummy variable w indicating two groups. Then

$$\begin{aligned} y &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k \\ &\quad + \delta_0 w + \delta_1 w \cdot x_1 + \delta_2 w \cdot x_2 + \ldots + \delta_k w \cdot x_k + u \end{aligned}$$

$$H_0 : \delta_0 = 0, \delta_1 = 0, \delta_2 = 0, \ldots, \delta_k = 0$$

  for $k+1$ restrictions

- Do F test for $k+1$ exclusion restrictions

- Chow test statistic often strongly rejects because of $\delta_0 \neq 0$. So it is often of interest to allow $\delta_0 \neq 0$ and just test equality of the slopes (by F test)

$$H_0^S : \delta_1 = 0, \delta_2 = 0, \ldots, \delta_k = 0$$