Motivation

Sampling Distributions of the OLS Estimators

Testing Hypotheses About a Single Population Parameter

Testing Against One-Sided Alternatives

Testing Against Two-Sided Alternatives

Testing Other Hypotheses about the $\beta_j$

Computing $p$-Values for $t$ Tests

Practical (Economic) versus Statistical Significance

Confidence Intervals

Testing Multiple Exclusion

# Multiple Regression Analysis - Inference


Caio Vigo

**The University of Kansas**
Department of Economics


Spring 2019

These slides were based on *Introductory Econometrics* by Jeffrey M. Wooldridge (2015)

# Topics

# Motivation for Inference



**Goal:** We want to test hypothesis about the parameters $\beta_j$ in the population regression model.

We want to know whether the true parameter $\beta_j =$ some value (your hypothesis).

- In order to do that, we will need to add a final assumption **MLR.6**. We will obtain the **Classical Linear Model (CLM)**

# Motivation for Inference



**MLR.1:** $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + u$
**MLR.2:** random sampling from the population
**MLR.3:** no perfect collinearity in the sample
**MLR.4:** $E(u|x_1, ..., x_k) = E(u) = 0$ (exogenous explanatory variables)
**MLR.5:** $Var(u|x_1, ..., x_k) = Var(u) = \sigma^2$ (homoskedasticity)

**MLR.1 - MLR.4**: Needed for unbiasedness of OLS:

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

**MLR.1 - MLR.5**: Needed to compute $Var(\hat{\beta}_j)$:

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

and for efficiency of OLS $\Rightarrow$ **BLUE**.

# Topics

# Sampling Distributions of the OLS Estimators



- Now we need to know the full sampling distribution of the $\hat{\beta}_j$.
- The **Gauss-Markov assumptions** don't tell us anything about these distributions.
- Based on our models, (conditional on $\{(x_{i1}, ..., x_{ik}) : i = 1, ..., n\}$) we need to have $dist(\hat{\beta}_j) = f(dist(u))$, i.e.,

$$\hat{\beta}_j \sim pdf(u)$$

- That's why we need one more assumption.

**MRL.6 (Normality)**

The population error $u$ is independent of the explanatory variables $(x_1, ..., x_k)$ and is normally distributed with mean zero and variance $\sigma^2$:

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

**MLR.1 - MLR.4** $\longrightarrow$ unbiasedness of OLS

**Gauss-Markov assumptions:** MLR.1 - MLR.4 $+$ MLR.5 (homoskedastic errors)

**Classical Linear Model (CLM):** Gauss-Markov $+$ MLR.6 (Normally distributed errors)

# Sampling Distributions of the OLS Estimators



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

- Strongest assumption.
- **MLR.6** implies ⇒ zero conditional mean (**MLR.4**) and homoskedasticity (**MLR.5**)
- Now we have full independence between $u$ and $(x_1, x_2, ..., x_k)$ *(not just mean and variance independence)*
- Reason to call $x_j$ **independent variables**.
- Recall the Normal distribution properties (see slides for **Appendix B**).

Figure: Distribution of $u$: $u \sim N(0, \sigma^2)$

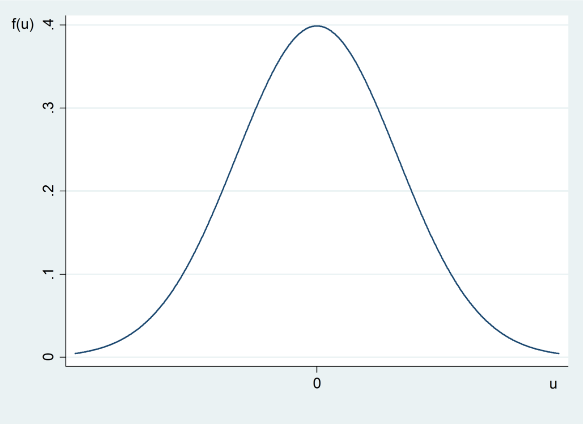
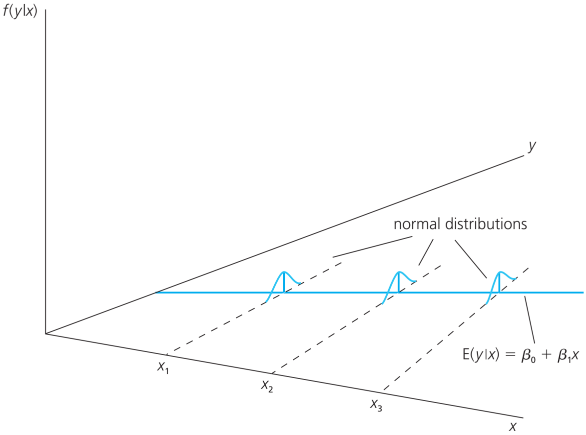


Figure: $f(y|x)$ with homoskedastic normal errors, i.e., $u \sim N(0, \sigma^2)$



# Sampling Distributions of the OLS Estimators



- Property of a **Normal distribution:** if $W \sim Normal$ then $a + bW \sim Normal$ for constants $a$ and $b$.

- What we are saying is that for normal r.v.s, any linear combination of them is also normally distributed.

- Because the $u_i$ are *independent and identically distributed* (*iid*) as $Normal(0, \sigma^2)$

$$\hat{\beta}_j = \beta_j + \sum_{i=1}^{n} w_{ij} u_i \sim Normal\left(\beta_j, Var(\hat{\beta}_j)\right)$$

- Then we can apply the Central Limit Theorem.

# Sampling Distributions of the OLS Estimators

**Theorem: Normal Sampling Distributions**

Under the **CLM** assumptions, conditional on the sample outcomes of the explanatory variables,

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

and so

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

# Topics

**Theorem: $t$ Distribution for Standardized Estimators**

Under the **CLM** assumptions,

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

where $k+1$ is the number of unknown parameter in the population model, and $n-k-1$ is the degrees of freedom (df).

# Testing Hypotheses About a Single Population Parameter: the $t$ Test



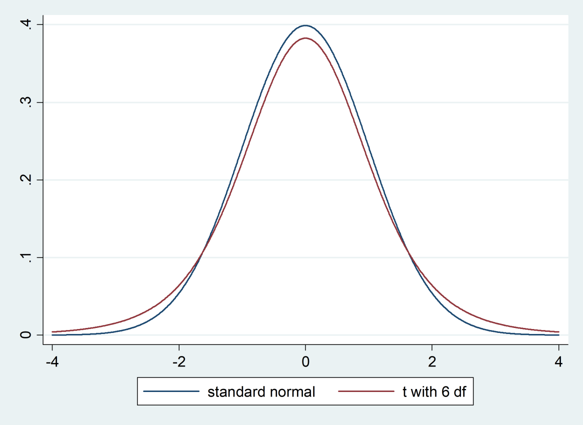
- Compare the ratios of the **previous 2** theorems. What is the difference?
- What is the difference between $sd(\hat{\beta}_j)$ and $se(\hat{\beta}_j)$?
- Recall the $t$ distribution properties (see slides for **Appendix B**).

- The $t$ distribution also has a bell shape, but is more spread out than the $Normal(0, 1)$.
- As $df \to \infty$,

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

- The difference is practically small for $df > 120$.
- See a $t$ table.
- The next graph plots a standard normal pdf against a $t_6$ pdf.

Figure: The pdfs of a standard normal and a $t_6$



• We use the result on the $t$ distribution to test the null hypothesis that $x_j$ has no partial effect on $y$:

$$H_0 : \beta_j = 0$$

$$\begin{aligned} lwage &= \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + u \\ H_0 &: \beta_2 = 0 \end{aligned}$$

• **Interpretation of what we are doing:** Once we control for education and time on the current job ($tenure$), total workforce experience has no affect on $lwage = \log(wage)$.

- To test

$$H_0 : \beta_j = 0$$

we use the **t statistic** (or **t ratio**),

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

- In virtually all cases $\hat{\beta}_j$ is not exactly equal to zero.

- When we use $t_{\hat{\beta}_j}$, we are measuring how far $\hat{\beta}_j$ is from zero *relative to its standard error*.

# Testing Against One-Sided Alternatives



- First consider the alternative

$$H_1 : \beta_j > 0$$

which means the null is effectively

$$H_0 : \beta_j \leq 0$$

- Using a positive one-sided alternative, if we reject $\beta_j = 0$, then we 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.

# Testing Against One-Sided Alternatives



- Because $se(\hat{\beta}_j) > 0$, $t_{\hat{\beta}_j}$ always has the same sign as $\hat{\beta}_j$.

- If the estimated coefficient $\hat{\beta}_j$ is negative, it provides no evidence against $H_0$ in favor of $H_1 : \beta_j > 0$.

- If $\hat{\beta}_j$ is positive, the question is: How big does $t_{\hat{\beta}_j} = \hat{\beta}_j / se(\hat{\beta}_j)$ have to be before we conclude $H_0$ is "unlikely"?

- Let's review the Error Types is Statistics.

# Testing Against One-Sided Alternatives



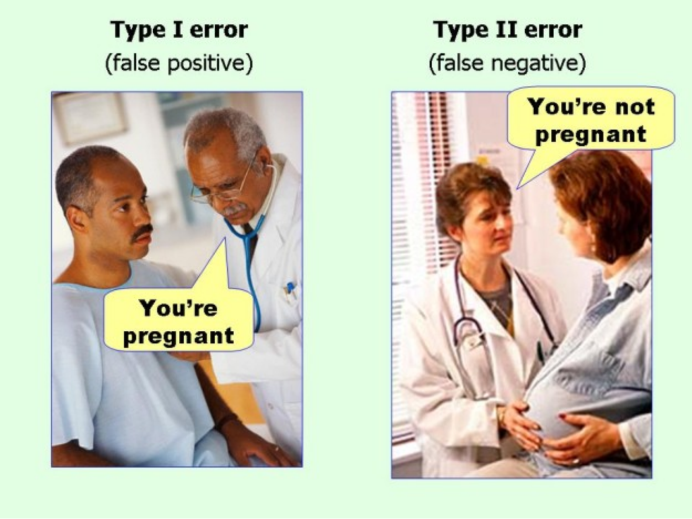
- Consider the following example:

$$H_0 \quad : \quad \text{Not pregnant}$$

$$H_1 \quad : \quad \text{Pregnant}$$

# Testing Against One-Sided Alternatives





# Testing Against One-Sided Alternatives

| | | Reality H0 is actually: | |
|---|---|---|---|
| | | False | True |
| Study Finding | Reject H0 | True Positive (Power) | False Positive Type I Error |
| | Accept H0 | False Negative Type II Error | True Negative |

# Testing Against One-Sided Alternatives



1. Choose a null hypothesis: $H_0 : \beta_j = 0$ (or $H_0 : \beta_j \leq 0$)

2. Choose an alternative hypothesis: $H_1 : \beta_j > 0$

3. Choose a **significance level** $\alpha$ **(**or simply **level**, or **size)** for the test.

That is, the probability of rejecting the null hypothesis when it is in fact true. (Type I Error).

Suppose we use 5%, so the probability of committing a Type I error is .05.

4. Obtain the critical value, $c > 0$, so that the **rejection rule**

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

leads to a 5% level test.

- The key is that, *under the null hypothesis*,

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

and this is what we use to obtain the critical value, $c$.

- Suppose $df = 28$ and we use a 5% test.

- Find the **critical value** in a t-table.
table).

# Testing Against One-Sided Alternatives



Numbers in each row of the table are values on a *t*-distribution with (*df*) degrees of freedom for selected right-tail (greater-than) probabilities (*p*).

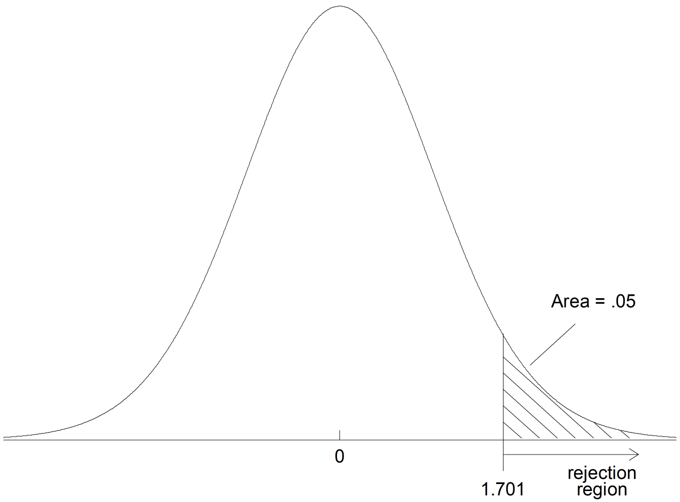

| df/p | 0.40 | 0.25 | 0.10 | 0.05 | 0.025 | 0.01 | 0.005 | 0.0005 |
|---|---|---|---|---|---|---|---|---|
| 1 | 0.324920 | 1.000000 | 3.077684 | 6.313752 | 12.70620 | 31.82052 | 63.65674 | 636.6192 |
| 2 | 0.288675 | 0.816497 | 1.885618 | 2.919986 | 4.30265 | 6.96456 | 9.92484 | 31.5991 |
| 3 | 0.276671 | 0.764892 | 1.637744 | 2.353363 | 3.18245 | 4.54070 | 5.84091 | 12.9240 |
| | | | | | | | | |
| 25 | 0.256060 | 0.684430 | 1.316345 | 1.708141 | 2.05954 | 2.48511 | 2.78744 | 3.7251 |
| 26 | 0.255955 | 0.684043 | 1.314972 | 1.705618 | 2.05553 | 2.47863 | 2.77871 | 3.7066 |
| 27 | 0.255858 | 0.683685 | 1.313703 | 1.703288 | 2.05183 | 2.47266 | 2.77068 | 3.6896 |
| 28 | 0.255768 | 0.683353 | 1.312527 | 1.701131 | 2.04841 | 2.46714 | 2.76326 | 3.6739 |
| 29 | 0.255684 | 0.683044 | 1.311434 | 1.699127 | 2.04523 | 2.46202 | 2.75639 | 3.6594 |
| 30 | 0.255605 | 0.682756 | 1.310415 | 1.697261 | 2.04227 | 2.45726 | 2.75000 | 3.6460 |
| *z* | 0.253347 | 0.674490 | 1.281552 | 1.644854 | 1.95996 | 2.32635 | 2.57583 | 3.2905 |
| CI | ——— | ——— | 80% | 90% | 95% | 98% | 99% | 99.9% |

- The critical value is $c = 1.701$ for 5% significance level (one-sided test).

- The following picture shows that we are conducting a **one-tailed test** (and it is these entries that should be used in the table).

# Testing Against One-Sided Alternatives





# Testing Against One-Sided Alternatives

- So, with $df = 28$, the rejection rule for $H_0 : \beta_j = 0$ against $H_1 : \beta_j > 0$, at the 5% level, is

$$t_{\hat{\beta}_j} > 1.701$$

We need a $t$ statistic greater than $1.701$ to conclude there is enough evidence against $H_0$.

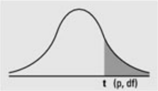
- If $t_{\hat{\beta}_j} \leq 1.701$, we fail to reject $H_0$ against $H_1$ at the 5% significance level.

# Testing Against One-Sided Alternatives



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

Numbers in each row of the table are values on a *t*-distribution with (*df*) degrees of freedom for selected right-tail (greater-than) probabilities (*p*).




| df/p | 0.40 | 0.25 | 0.10 | 0.05 | 0.025 | 0.01 | 0.005 | 0.0005 |
|---|---|---|---|---|---|---|---|---|
| 1 | 0.324920 | 1.000000 | 3.077684 | 6.313752 | 12.70620 | 31.82052 | 63.65674 | 636.6192 |
| 2 | 0.288675 | 0.816497 | 1.885618 | 2.919986 | 4.30265 | 6.96456 | 9.92484 | 31.5991 |
| 3 | 0.276671 | 0.764892 | 1.637744 | 2.353363 | 3.18245 | 4.54070 | 5.84091 | 12.9240 |
| | | | | | | | | |
| 25 | 0.256060 | 0.684430 | 1.316345 | 1.708141 | 2.05954 | 2.48511 | 2.78744 | 3.7251 |
| 26 | 0.255955 | 0.684043 | 1.314972 | 1.705618 | 2.05553 | 2.47863 | 2.77871 | 3.7066 |
| 27 | 0.255858 | 0.683685 | 1.313703 | 1.703288 | 2.05183 | 2.47266 | 2.77068 | 3.6896 |
| 28 | 0.255768 | 0.683353 | 1.312527 | 1.701131 | 2.04841 | 2.46714 | 2.76326 | 3.6739 |
| 29 | 0.255684 | 0.683044 | 1.311434 | 1.699127 | 2.04523 | 2.46202 | 2.75639 | 3.6594 |
| 30 | 0.255605 | 0.682756 | 1.310415 | 1.697261 | 2.04227 | 2.45726 | 2.75000 | 3.6460 |
| *z* | 0.253347 | 0.674490 | 1.281552 | 1.644854 | 1.95996 | 2.32635 | 2.57583 | 3.2905 |
| CI | — | — | 80% | 90% | 95% | 98% | 99% | 99.9% |

- Thus, if $df = 28$, below are the critical values for the following significance levels: 10% level, 5% and 1% level.

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

# Testing Against One-Sided Alternatives



If we want to reduce the probability of Type I error, we must increase the critical value (so we reject the null less often).

- If we reject at, say, the 1% level, then we must also reject at any larger level.

- If we fail to reject at, say, the 10% level – so that $t_{\hat{\beta}_j} \leq 1.313$ – then we will fail to reject at any smaller level.

# Testing Against One-Sided Alternatives



• With large sample sizes – certain when $df > 120$ – we can use critical values from the standard normal distribution.

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

which we can round to $1.28$, $1.65$, and $2.36$, respectively. The value $1.65$ is especially common for a one-tailed test.

# Testing Against One-Sided Alternatives



- Recall our **wage** model example:

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_1 tenure + u$$

- First, let's label the parameters with the variable names: $\beta_{educ}$, $\beta_{exper}$, and $\beta_{\textsf{tenure}}$

- We would like to test:

$$H_0 : \beta_{exper} = 0$$

**Interpretation:** We are testing if workforce experience has no effect on a wage once education and tenure have been accounted for.

# Testing Against One-Sided Alternatives



| | Dependent variable: lwage |
|---|---|
| educ | 0.092*** |
| | (0.007) |
| exper | 0.004** |
| | (0.002) |
| tenure | 0.022*** |
| | (0.003) |
| Constant | 0.284*** |
| | (0.104) |
| Observations | 526 |
| R2 | 0.316 |
| Adjusted R2 | 0.312 |
| Residual Std. Error | 0.441 (df = 522) |
| F Statistic | 80.391*** (df = 3; 522) |
| Note: | *p<0.1; **p<0.05; ***p<0.01 |

# Testing Against One-Sided Alternatives

- What is the $t_{exper}$?

$$t_{exper} = \frac{0.004}{0.002} = 2.00$$

- Now what do you do with this number?
- How many $df$ do we have?
- Which table could I use?
- Using a standard normal table: the one-sided critical value at the 5% level, 1.645.

**Statistical Significance X Economic Importance/Interpretation**

• So "$\hat{\beta}_{exper}$ is **statistically significant**" at 5% level significance level (one-sided test).

• The estimated effect of $exper$, which is its **economic importance** should be interpreted as: another year of experience, holding $educ$ and $tenure$ fixed, is estimated to be worth about $0.4\%$.

# Testing Against One-Sided Alternatives



- For the negative one-sided alternative,

$$
\begin{aligned}
H_0 &: \beta_j \geq 0 \\
H_1 &: \beta_j < 0
\end{aligned}
$$

we use a symmetric rule. But the rejection rule is

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

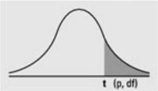
where $c$ is chosen in the same way as in the positive case.

# Testing Against One-Sided Alternatives



- With $df = 28$ and we want to test at a 5% significance level, what is the critical value?

Numbers in each row of the table are values on a *t*-distribution with (*df*) degrees of freedom for selected right-tail (greater-than) probabilities (*p*).




| df/p | 0.40 | 0.25 | 0.10 | 0.05 | 0.025 | 0.01 | 0.005 | 0.0005 |
|---|---|---|---|---|---|---|---|---|
| 1 | 0.324920 | 1.000000 | 3.077684 | 6.313752 | 12.70620 | 31.82052 | 63.65674 | 636.6192 |
| 2 | 0.288675 | 0.816497 | 1.885618 | 2.919986 | 4.30265 | 6.96456 | 9.92484 | 31.5991 |
| 3 | 0.276671 | 0.764892 | 1.637744 | 2.353363 | 3.18245 | 4.54070 | 5.84091 | 12.9240 |
| | | | | | | | | |
| 25 | 0.256060 | 0.684430 | 1.316345 | 1.708141 | 2.05954 | 2.48511 | 2.78744 | 3.7251 |
| 26 | 0.255955 | 0.684043 | 1.314972 | 1.705618 | 2.05553 | 2.47863 | 2.77871 | 3.7066 |
| 27 | 0.255858 | 0.683685 | 1.313703 | 1.703288 | 2.05183 | 2.47266 | 2.77068 | 3.6896 |
| 28 | 0.255768 | 0.683353 | 1.312527 | 1.701131 | 2.04841 | 2.46714 | 2.76326 | 3.6739 |
| 29 | 0.255684 | 0.683044 | 1.311434 | 1.699127 | 2.04523 | 2.46202 | 2.75639 | 3.6594 |
| 30 | 0.255605 | 0.682756 | 1.310415 | 1.697261 | 2.04227 | 2.45726 | 2.75000 | 3.6460 |
| z | 0.253347 | 0.674490 | 1.281552 | 1.644854 | 1.95996 | 2.32635 | 2.57583 | 3.2905 |
| CI | ——— | ——— | 80% | 90% | 95% | 98% | 99% | 99.9% |

# Testing Against One-Sided Alternatives



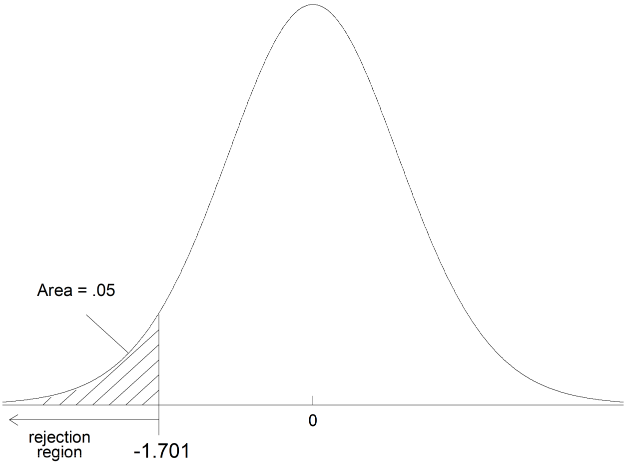
**Intuition:** We must see a significantly negative value for the $t$ statistic to reject the null hypothesis in favor of the alternative hypothesis.

- With $df = 28$ and a 5% test, the critical value is $c = -1.701$, so the rejection rule is

$$t_{\hat{\beta}_j} < -1.701$$

# Testing Against One-Sided Alternatives





**Reminder about Testing**

• Our hypotheses involve the unknown population values, $\beta_j$.

• If in a our set of data we obtain, say, $\hat{\beta}_j = 2.75$, we do not write the null hypothesis as

$$H_0 : 2.75 = 0$$

(which is obviously false).

• Nor do we write

$$H_0 : \hat{\beta}_j = 0$$

(which is also false except in the very rare case that our estimate is exactly zero).

• We do not test hypotheses about the estimate! We know what it is once we collect the sample. We hypothesize about the unknown population value, $\beta_j$.

## Testing Against Two-Sided Alternatives

• Sometimes we do not know ahead of time whether a variable definitely has a positive effect or a negative effect.
• So, in this case the hypothesis should be written as:

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

• Testing against the **two-sided alternative** is usually the default. It prevents us from looking at the regression results and then deciding on the alternative.

# Testing Against Two-Sided Alternatives



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

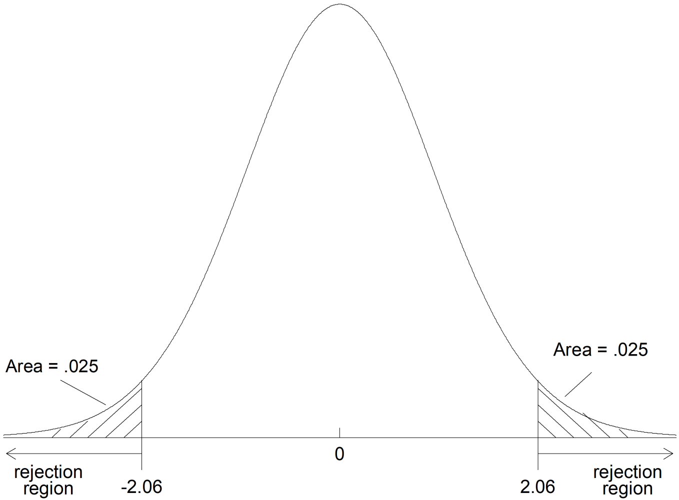
**Two-tailed test**

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

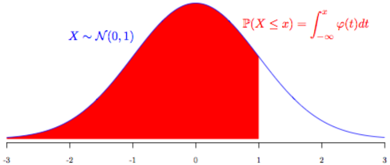
- For example, if we use a 5% level test and $df = 25$, the two-tailed cv is $2.06$. The two-tailed cv is, in this case, the $97.5$ percentile in the $t_{25}$ distribution. (Compare the one-tailed cv, about 1.71, the 95$^{th}$ percentile in the $t_{25}$ distribution).

# Testing Against Two-Sided Alternatives





# Testing Against Two-Sided Alternatives




| | 0.00 | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | 0.5000 | 0.5040 | 0.5080 | 0.5120 | 0.5160 | 0.5199 | 0.5239 | 0.5279 | 0.5319 | 0.5359 |
| 0.1 | 0.5398 | 0.5438 | 0.5478 | 0.5517 | 0.5557 | 0.5596 | 0.5636 | 0.5675 | 0.5714 | 0.5753 |
| 0.2 | 0.5793 | 0.5832 | 0.5871 | 0.5910 | 0.5948 | 0.5987 | 0.6026 | 0.6064 | 0.6103 | 0.6141 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 1.5 | 0.9332 | 0.9345 | 0.9357 | 0.9370 | 0.9382 | 0.9394 | 0.9406 | 0.9418 | 0.9429 | 0.9441 |
| 1.6 | 0.9452 | 0.9463 | 0.9474 | 0.9484 | 0.9495 | 0.9505 | 0.9515 | 0.9525 | 0.9535 | 0.9545 |
| 1.7 | 0.9554 | 0.9564 | 0.9573 | 0.9582 | 0.9591 | 0.9599 | 0.9608 | 0.9616 | 0.9625 | 0.9633 |
| 1.8 | 0.9641 | 0.9649 | 0.9656 | 0.9664 | 0.9671 | 0.9678 | 0.9686 | 0.9693 | 0.9699 | 0.9706 |
| 1.9 | 0.9713 | 0.9719 | 0.9726 | 0.9732 | 0.9738 | 0.9744 | 0.9750 | 0.9756 | 0.9761 | 0.9767 |
| 2.0 | 0.9772 | 0.9778 | 0.9783 | 0.9788 | 0.9793 | 0.9798 | 0.9803 | 0.9808 | 0.9812 | 0.9817 |
| 2.1 | 0.9821 | 0.9826 | 0.9830 | 0.9834 | 0.9838 | 0.9842 | 0.9846 | 0.9850 | 0.9854 | 0.9857 |
| 2.2 | 0.9861 | 0.9864 | 0.9868 | 0.9871 | 0.9875 | 0.9878 | 0.9881 | 0.9884 | 0.9887 | 0.9890 |
| 2.3 | 0.9893 | 0.9896 | 0.9898 | 0.9901 | 0.9904 | 0.9906 | 0.9909 | 0.9911 | 0.9913 | 0.9916 |
| 2.4 | 0.9918 | 0.9920 | 0.9922 | 0.9925 | 0.9927 | 0.9929 | 0.9931 | 0.9932 | 0.9934 | 0.9936 |
| 2.5 | 0.9938 | 0.9940 | 0.9941 | 0.9943 | 0.9945 | 0.9946 | 0.9948 | 0.9949 | 0.9951 | 0.9952 |
| 2.6 | 0.9953 | 0.9955 | 0.9956 | 0.9957 | 0.9959 | 0.9960 | 0.9961 | 0.9962 | 0.9963 | 0.9964 |
| 2.7 | 0.9965 | 0.9966 | 0.9967 | 0.9968 | 0.9969 | 0.9970 | 0.9971 | 0.9972 | 0.9973 | 0.9974 |
| 2.8 | 0.9974 | 0.9975 | 0.9976 | 0.9977 | 0.9977 | 0.9978 | 0.9979 | 0.9979 | 0.9980 | 0.9981 |
| 2.9 | 0.9981 | 0.9982 | 0.9982 | 0.9983 | 0.9984 | 0.9984 | 0.9985 | 0.9985 | 0.9986 | 0.9986 |
| 3.0 | 0.9987 | 0.9987 | 0.9987 | 0.9988 | 0.9988 | 0.9989 | 0.9989 | 0.9989 | 0.9990 | 0.9990 |

# Testing Against One-Sided Alternatives



| | Dependent variable: |
|---|---|
| | lwage |
| educ | 0.092*** |
| | (0.007) |
| exper | 0.004** |
| | (0.002) |
| tenure | 0.022*** |
| | (0.003) |
| Constant | 0.284*** |
| | (0.104) |
| Observations | 526 |
| R2 | 0.316 |
| Adjusted R2 | 0.312 |
| Residual Std. Error | 0.441 (df = 522) |
| F Statistic | 80.391*** (df = 3; 522) |
| Note: | *p<0.1; **p<0.05; ***p<0.01 |

# Testing Against Two-Sided Alternatives



- When we reject $H_0 : \beta_j = 0$ against $H_1 : \beta_j \neq 0$, we often say that $\hat{\beta}_j$ is **statistically different from zero** and usually mention a significance level.

As in the one-sided case, we also say $\hat{\beta}_j$ is **statistically significant** when we can reject $H_0 : \beta_j = 0$.

- Testing the null $H_0 : \beta_j = 0$ is the standard practice.

- **R**, Stata, EViews and all the other regression packages automatically report the $t$ statistic for **this hypothesis** (i.e., two-sided test).

# Testing Other Hypotheses about the $\beta_j$



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

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

we might want to test

$$H_0 : \beta_1 = 1$$

which means an income elasticity equal to one. (We can be pretty sure that $\beta_1 > 0$.)

# Testing Other Hypotheses about the $\beta_j$



**Important observation**

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

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

# Testing Other Hypotheses about the $\beta_j$



- More generally, suppose the null is

$$H_0 : \beta_j = a_j$$

where we specify the value $a_j$

- It is easy to extend the $t$ statistic:

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

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

# Testing Other Hypotheses about the $\beta_j$



**General expression for general $t$ testing**

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

- The alternative can be one-sided or two-sided.
- We choose critical values in exactly the same way as before.

- The language needs to be suitably modified. If, for example,

$$\begin{aligned} H_0 &: \beta_j = 1 \\ H_1 &: \beta_j \neq 1 \end{aligned}$$

is rejected at the 5% level, we say "$\hat{\beta}_j$ is statistically different from one at the 5% level." Otherwise, $\hat{\beta}_j$ is "not statistically different from one." If the alternative is $H_1 : \beta_j > 1$, then "$\hat{\beta}_j$ is statistically greater than one at the 5% level."

**Example:** Crime, police officers and enrollment on college campuses
Let's do the following hypothesis test:

$$\begin{aligned} \log(crime) &= \beta_0 + \beta_1 police + \beta_2 log(enroll) + u \\ H_0 &: \beta_1 = 1 \\ H_1 &: \beta_1 > 1 \end{aligned}$$

# Testing Other Hypotheses about the $\beta_j$



| | Dependent variable: |
|---|---|
| | log(crime) |
| police | 0.0240*** |
| | (0.0073) |
| log(enroll) | 0.9767*** |
| | (0.1373) |
| Constant | -4.3758*** |
| | (1.1990) |
| Observations | 97 |
| R2 | 0.6277 |
| Adjusted R2 | 0.6198 |
| Residual Std. Error | 0.8516 (df = 94) |
| F Statistic | 79.2389*** (df = 2; 94) |
| Note: | *p<0.1; **p<0.05; ***p<0.01 |

# Computing $p$-Values for $t$ Tests



- The traditional approach to testing, where we choose a significance level ahead of time, has a component of **arbitrariness**.
- Different researchers prefer different significance levels (10%, 5%, 1%).
- Committing to a significance level ahead of time can hide useful information about the outcome of hypothesis test.
- **Example:** (On white board)

# Computing $p$-Values for $t$ Tests

• Rather than have to specify a level ahead of time, or discuss different traditional significance levels (10%, 5%, 1%), it is better to answer the following question:

**Intuition:** Given the observed value of the $t$ statistic, what is the *smallest* significance level at which I can reject $H_0$?

• The smallest level at which the null can be rejected is known as the $p$-**value** of a test.

# Computing $p$-Values for $t$ Tests



**$p$-value**

For $t$ testing against a two-sided alternative,

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

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

- The $p$-value is a probability, so it is between zero and one.

# Computing $p$-Values for $t$ Tests



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

- Usually we just report $p$-values for two-sided alternatives.

# Computing $p$-Values for $t$ Tests



**Mnemonic Device**

*Small* $p$-values are evidence *against* the null hypothesis.

*Large* $p$-values provide little evidence *against* the null hypothesis.

**Intuition:** $p$-value is the probability of observing a statistic as extreme as we did if the null hypothesis is true.

# Computing $p$-Values for $t$ Tests



- If $p$-value $= .50$, then there is a 50% chance of observing a $t$ 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 a $t$ statistic as extreme as we did is $.1\%$.

- We can conclude that we got a very rare sample (*unlikely!*) or that the null hypothesis is very likely false.

# Computing $p$-Values for $t$ Tests



- From

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

we see that as $|t|$ increases the $p$-value decreases.

Large absolute $t$ statistics are **associated** with small $p$-values.

**Example:**

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

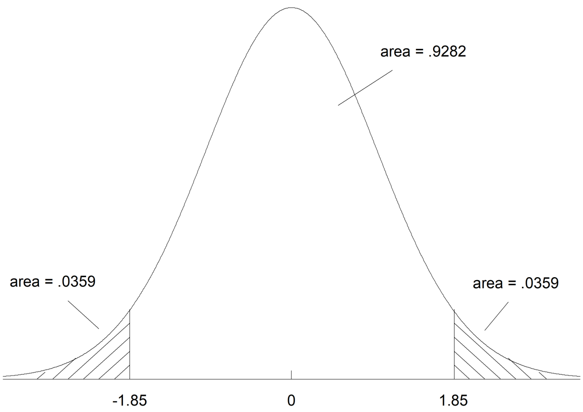
$$p\text{-value} = P(|T| > 1.85) = 2P(T > 1.85) = 2(.0359) = .0718$$

where $T \sim t_{40}$.

# Computing $p$-Values for $t$ Tests



Figure: $t$ distribution with $40$ degrees of freedom



# Computing $p$-Values for $t$ Tests



- Given $p$-value, we can carry out a test at any significance level.

If $\alpha$ is the chosen level, then

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

**Example**

Suppose we obtained $p$-value $= .0718$. This means that we reject $H_0$ at the 10% level but not the 5% level. We reject at 8% but not at 7%.

# Practical versus Statistical Significance

- $t$ testing is purely about *statistical significance*.

- It does not directly speak to the issue of whether a variable has a practically, or economically large effect.

**Practical (Economic) Significance** depends on the size (and sign) of $\hat{\beta}_j$.

**Statistical Significance** depends on $t_{\hat{\beta}_j}$.

# Practical (Economic) versus Statistical Significance



It is possible estimate **practically large effects** but have the estimates so imprecise that they are **statistically insignificant**.

Common with small data sets (but not only small data sets).

**X**

It is possible to get estimates that are **statistically significant** (often with very small $p$-values) but are **not practically large**.

Common with very large data sets.

# Topics

# Confidence Intervals



- **Under the CLM assumptions**, rather than just testing hypotheses about parameters it is also useful to construct **confidence intervals** (also know as **interval estimates**).

**Intuition:** If you could obtain several random samples data, the **confidence interval** tells you that, for a 95% CI, your true $\beta_j$ will lie in this interval $[\beta_j^{lower}, \beta_j^{upper}]$ for 95% of the samples.

- We will construct CIs of the form

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

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

- We will use a 95% confidence level, in which case $c$ comes from the 97.5 percentile in the $t_{df}$ distribution.

- Therefore, $c$ is the 5% critical value against a two-sided alternative.

# Confidence Intervals



**Rule of Thumb**

- For, $df \geq 120$, an approximate 95% CI is:

$$\hat{\beta}_j \pm 2 \cdot se(\hat{\beta}_j) \text{ or } \left[\hat{\beta}_j - 2 \cdot se(\hat{\beta}_j), \hat{\beta}_j + 2 \cdot se(\hat{\beta}_j)\right]$$

- For small $df$, the exact percentiles should be obtained from a $t$ table.

Find the 95% CI for the parameters from the following regression:

| | Dependent variable: |
|---|---|
| | lwage |
| educ | 0.092*** |
| | (0.007) |
| exper | 0.004** |
| | (0.002) |
| tenure | 0.022*** |
| | (0.003) |
| Constant | 0.284*** |
| | (0.104) |
| Observations | 526 |
| R2 | 0.316 |
| Adjusted R2 | 0.312 |
| Residual Std. Error | 0.441 (df = 522) |
| F Statistic | 80.391*** (df = 3; 522) |
| Note: | *p<0.1; **p<0.05; ***p<0.01 |

KU

# Confidence Intervals



- The correct way to interpret a CI is to remember that the endpoints, $\hat{\beta}_j - c \cdot se(\hat{\beta}_j)$ and $\hat{\beta}_j + c \cdot se(\hat{\beta}_j)$, **change** with each sample (or at least can change).

Endpoints are random outcomes that depend on the data we draw.

# Confidence Intervals



A 95% CI means is that for 95% of the random samples that we draw from the population,

the interval we compute using the rule $\hat{\beta}_j \pm c \cdot se(\hat{\beta}_j)$

will include the value $\beta_j$.

**But for a particular sample we do not know whether $\beta_j$ is in the interval.**

- This is similar to the idea that unbiasedness of $\hat{\beta}_j$ does *not* means that $\hat{\beta}_j = \beta_j$. Most of the time $\hat{\beta}_j$ is not $\beta_j$. Unbiasedness means $E(\hat{\beta}_j) = \beta_j$.

## Topics

# Testing Multiple Exclusion Restrictions



- Sometimes want to test **more than one hypothesis**, which then includes **multiple parameters**.

- Generally, it is not valid to look at individual $t$ statistics.

- We need a specific statistic used to test **joint hypotheses**.

# Testing Multiple Exclusion Restrictions

**Example:**

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + u$$

- Let's consider the following null hypothesis:

$$H_0 : \beta_2 = 0,\ \beta_3 = 0$$

- **Exclusion Restrictions**: We want to know if we can exclude some variables jointly.

# Testing Multiple Exclusion Restrictions



- To test $H_0$, we need a **joint (multiple) hypotheses test**.

- A $t$ statistic can be used for a single exclusion restriction; it does not take a stand on the values of the other parameters.

- We are considering the alternative to be:

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

- So, $H_1$ means **at least one** of betas is different from zero.

# Testing Multiple Exclusion Restrictions



- The original model, containing all variables, is the **unrestricted model**:

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + u$$

- When we impose $H_0 : \beta_2 = 0$, $\beta_3 = 0$, we get the **restricted model**:

$$\log(wage) = \beta_0 + \beta_1 educ + u$$

- We want to see how the fit deteriorates as we remove the two variables.

- We use, initially, the **sum of squared residuals** from the two regressions.

- It is an algebraic fact that the SSR must increase (or, at least not fall) when explanatory variables are dropped. So,

$$SSR_r \geq SSR_{ur}$$

# Testing Multiple Exclusion Restrictions

**$F$ test**

Does the $SSR$ increase proportionately by enough to conclude the restrictions under $H_0$ are false?

- In the general model:

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

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

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

- We get $SSR_{ur}$ from estimating the full model.

# Testing Multiple Exclusion Restrictions



- The restricted model we estimate to get $SSR_r$ drops the last $q$ variables ($q$ exclusion restrictions):

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

- The **F statistic** uses a degrees of freedom adjustment. In general, we have

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

where $q$ is the number of exclusion restrictions imposed under the null ($q = 2$ in our example).

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

- The denominator of the $F$ statistic, $SSR_{ur}/df_{ur}$, is the unbiased estimator of $\sigma^2$ from the unrestricted model.
- Note that $F \geq 0$, and $F > 0$ virtually always holds.
- As a computational device, sometimes the formula

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

is useful.

# Testing Multiple Exclusion Restrictions



- Using classical testing, the rejection rule is of the form

$$F > c$$

where $c$ is an appropriately chosen **critical value**.

**Distribution of $F$ statistic**

Under $H_0$ (the $q$ exclusion restrictions)

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

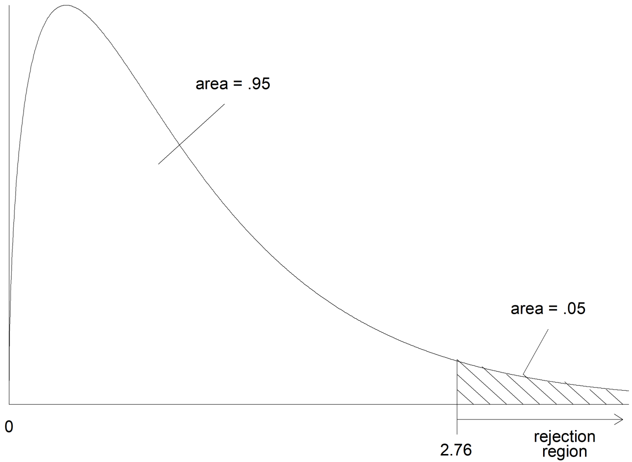
i.e., it has an $F$ distribution with $(q, n-k-1)$ degrees of freedom.

- Recall the $F$ distribution (see slides for **Appendix B**).

# Testing Multiple Exclusion Restrictions



- Suppose $q = 3$ and $n - k - 1 = df_{ur} = 60$. Then the 5% cv is 2.76.



# $R$-Squared Form of the $F$ Statistic



**Question:** Is there a way to compute the $F$ statistic with the information reported in the standard output from any econometric/statistcal package?

- The $R$-squared is always reported.
- The SSR is not reported most of the time.
- It turns out that $F$ tests for exclusion restrictions can be computed entirely from the $R$-squareds for the restricted and unrestricted models.
- Notice that,

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

# $R$-Squared Form of the $F$ Statistic

- Therefore,

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

- Notice how $R_{ur}^2$ comes first in the numerator.
- We know $R_{ur}^2 \geq R_r^2$ so this ensures $F \geq 0$.

# $R$-Squared Form of the $F$ Statistic



### Example

**unrestricted model**: $\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + u$

**restricted model**: $\log(wage) = \beta_0 + \beta_1 educ + u$

# $R$-Squared Form of the $F$ Statistic



| | Dependent variable: lwage |
|---|---|
| educ | 0.092*** |
| | (0.007) |
| exper | 0.004** |
| | (0.002) |
| tenure | 0.022*** |
| | (0.003) |
| Constant | 0.284*** |
| | (0.104) |
| Observations | 526 |
| R2 | 0.316 |
| Adjusted R2 | 0.312 |
| Residual Std. Error | 0.441 (df = 522) |
| F Statistic | 80.391*** (df = 3; 522) |
| Note: | *p<0.1; **p<0.05; ***p<0.01 |

| | Dependent variable: lwage |
|---|---|
| educ | 0.083*** |
| | (0.008) |
| Constant | 0.584*** |
| | (0.097) |
| Observations | 526 |
| R2 | 0.186 |
| Adjusted R2 | 0.184 |
| Residual Std. Error | 0.480 (df = 524) |
| F Statistic | 119.582*** (df = 1; 524) |
| Note: | *p<0.1; **p<0.05; ***p<0.01 |

# $R$-Squared Form of the $F$ Statistic



- We say that $exper$, and $tenure$ are **jointly statistically significant** (or just **jointly significant**), in this case, at any small significance level we want.

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

## The $F$ Statistic for Overall Significance of a Regression

- The $F$ statistic in the **R** output tests a very special null hypothesis.
- In the model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + 0$$

the null is that **all slope coefficients are zero**, i.e,

$$H_0 : \beta_1 = 0, \beta_2 = 0, ..., \beta_k = 0$$

- This means that none of the $x_j$ helps explain $y$.
- If we cannot reject this null, we have found no factors that explain $y$.

# The $F$ Statistic for Overall Significance of a Regression

- For this test,

$R_r^2 = 0$ (no explanatory variables under $H_0$).
$R_{ur}^2 = R^2$ from the regression.

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

# The $F$ Statistic for Overall Significance of a Regression

- As $R^2$ increases, so does $F$.
- A small $R^2$ can lead $F$ to be significant.
- If the $df = n - k - 1$ is large (because of large $n$), $F$ can be large even with a "small" $R^2$.
- Increasing $k$ decreases $F$.

# The $F$ Statistic for Overall Significance of a Regression



| | Dependent variable: lwage |
|---|---|
| educ | 0.092*** |
| | (0.007) |
| exper | 0.004** |
| | (0.002) |
| tenure | 0.022*** |
| | (0.003) |
| Constant | 0.284*** |
| | (0.104) |
| Observations | 526 |
| R2 | 0.316 |
| Adjusted R2 | 0.312 |
| Residual Std. Error | 0.441 (df = 522) |
| F Statistic | 80.391*** (df = 3; 522) |
| Note: | *p<0.1; **p<0.05; ***p<0.01 |

| | Dependent variable: lwage |
|---|---|
| Constant | 1.623*** |
| | (0.023) |
| Observations | 526 |
| R2 | 0.000 |
| Adjusted R2 | 0.000 |
| Residual Std. Error | 0.532 (df = 525) |
| Note: | *p<0.1; **p<0.05; ***p<0.01 |