

nav nav nav nav nav nav

450A Political Methodology I

Lecture 6: Multiple Regression

2. Statistical Inference

Stanford University

Testing Hypotheses about Individual Coefficients

Testing Linear Hypotheses: A Simple Case

Testing Joint Significance

Testing Linear Hypotheses: The General Case

Example: Chilean Referendum on Pinochet

- The 1988 Chilean national plebiscite was a national referendum held to determine whether or not dictator Augusto Pinochet would extend his rule for another eight-year term in office.
- Data: national survey conducted in April and May of 1988 by FLACSO in Chile.
- Outcome: 1 if the respondent intends to vote Yes (for Pinochet), 0 if No. We keep the 1,703 of 2,700 respondents who intend to vote Yes or No and answered every question. With a 0/1 outcome, OLS fits a **linear probability model**, in which each slope is a difference in the probability of a Yes vote. Here that does little harm. The fitted probabilities stay between .24 and .75, and robust SEs (Lecture 6-3) are within 1% of the usual ones.
- Plebiscite was held on October 5, 1988. The No side won with 56% of the vote, with 44% voting Yes.
- We model the intended Pinochet vote as a linear function of gender, education, and age of respondents. Education (educ): 1 = primary, 2 = secondary, 3 = post-secondary.

Hypothesis Testing in R

R Code

```
> fit <- lm(vote1 ~ fem + educ + age, data = d)
> summary(fit)
~~~~~
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.4917751  0.0525717   9.354  < 2e-16 ***
fem          0.1284292  0.0236221   5.437 6.21e-08 ***
educ        -0.1051514  0.0164543  -6.391 2.13e-10 ***
age          0.0033573  0.0008235   4.077 4.77e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4845 on 1699 degrees of freedom
Multiple R-squared:  0.06292,    Adjusted R-squared:  0.06127
F-statistic: 38.03 on 3 and 1699 DF,  p-value: < 2.2e-16
```

R Code

```
> confint(fit)
              2.5 %      97.5 %
(Intercept) 0.388662948 0.594887290
fem          0.082097691 0.174760775
educ        -0.137424277 -0.072878547
age          0.001742153 0.004972387
```

The t-value for Multiple Linear Regression

- Consider testing a hypothesis about a single regression coefficient β_j :

$$H_0 : \beta_j = c$$

- In the simple linear regression we used the **t-value** to test this kind of hypothesis.
- We can consider the same t-value about β_j for the multiple regression:

$$T = \frac{\hat{\beta}_j - c}{\widehat{SE}(\hat{\beta}_j)}$$

- How do we compute $\widehat{SE}(\hat{\beta}_j)$?

$$\widehat{SE}(\hat{\beta}_j) = \sqrt{\widehat{V}(\hat{\beta}_j)} = \sqrt{\widehat{V}(\hat{\beta})_{(j,j)}} = \sqrt{\hat{\sigma}^2(\mathbf{X}'\mathbf{X})_{(j,j)}^{-1}}$$

where $\mathbf{A}_{(j,j)}$ is the (j,j) element of matrix $\mathbf{A}$.

That is, take the variance-covariance matrix of $\hat{\beta}$ and square root the diagonal element corresponding to j .

Hypothesis Testing in R

R Code

```
> fit <- lm(vote1 ~ fem + educ + age, data = d)
> summary(fit)
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.4917751  0.0525717   9.354  < 2e-16 ***
fem          0.1284292  0.0236221   5.437 6.21e-08 ***
educ        -0.1051514  0.0164543  -6.391 2.13e-10 ***
age          0.0033573  0.0008235   4.077 4.77e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We can pull out the variance-covariance matrix $\hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}$ in R from the `lm()` object:

R Code

```
> V <- vcov(fit)
> V
              (Intercept)          fem          educ          age
(Intercept)  2.763788e-03 -3.717198e-04 -6.543766e-04 -3.324904e-05
fem          -3.717198e-04  5.580053e-04  3.781282e-05  9.631845e-07
educ         -6.543766e-04  3.781282e-05  2.707446e-04  3.860967e-06
age          -3.324904e-05  9.631845e-07  3.860967e-06  6.780985e-07

> sqrt(diag(V))
(Intercept)          fem          educ          age
0.0525717389 0.0236221360 0.0164543200 0.0008234674
```

Sampling Distribution of the t-Value

The t-values in multiple regressions essentially have the same statistical properties as the simple regression case. That is,

Theorem (Small-Sample Distribution of the t-Value)

If $H_0 : \beta_j = c$ is true, then under Assumptions I–VI, for any sample size n the t-value has the t distribution with $(n - k - 1)$ degrees of freedom:

$$T = \frac{\hat{\beta}_j - c}{\widehat{SE}[\hat{\beta}_j]} \sim t_{n-k-1}$$

Theorem (Large-Sample Distribution of the t-Value)

If H_0 is true, then under Assumptions I–V, as $n \rightarrow \infty$ the distribution of the t-value approaches the standard normal distribution:

$$T = \frac{\hat{\beta}_j - c}{\widehat{SE}[\hat{\beta}_j]} \stackrel{a.}{\sim} \mathcal{N}(0, 1) \quad \text{as } n \rightarrow \infty$$

- $t_{n-k-1} \rightarrow \mathcal{N}(0, 1)$ as $n \rightarrow \infty$, so the difference disappears when n large.
- In practice people often just use t_{n-k-1} to be on the conservative side.

Using the t -Value as a Test Statistic

The procedure for testing this null hypothesis ($\beta_j = c$) is **identical** to the simple regression case, except that our reference distribution is t_{n-k-1} instead of t_{n-2} .

- 1 Compute the t -value as $T = (\hat{\beta}_j - c) / \widehat{SE}[\hat{\beta}_j]$
- 2 Compare the value to the **critical value** $t_{\alpha/2}$ for the α level test, which under the null hypothesis satisfies

$$P(-t_{\alpha/2} \leq T \leq t_{\alpha/2}) = 1 - \alpha$$

- 3 Decide whether the realized value of T in our data is unusual given the known distribution of the test statistic.
- 4 Finally, either declare that we reject H_0 or not, or report the p -value.

Confidence Intervals

To construct confidence intervals, there is again no difference compared to the case of $k = 1$, except that we need to use t_{n-k-1} instead of t_{n-2}

Since we know the sampling distribution for our t-value:

$$\frac{\hat{\beta}_j - \beta_j}{\widehat{SE}[\hat{\beta}_j]} \sim t_{n-k-1}$$

So we also know the probability that the value of our test statistic falls into a given interval:

$$P\left(-t_{\alpha/2} \leq \frac{\hat{\beta}_j - \beta_j}{\widehat{SE}[\hat{\beta}_j]} \leq t_{\alpha/2}\right) = 1 - \alpha$$

We rearrange:

$$\left[\hat{\beta}_j - t_{\alpha/2} \widehat{SE}[\hat{\beta}_j], \hat{\beta}_j + t_{\alpha/2} \widehat{SE}[\hat{\beta}_j]\right]$$

and thus can construct the confidence intervals as usual using:

$$\hat{\beta}_j \pm t_{\alpha/2} \cdot \widehat{SE}[\hat{\beta}_j]$$

Confidence Intervals in R

R Code

```
> fit <- lm(vote1 ~ fem + educ + age, data = d)
> summary(fit)
~~~~~
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.4917751  0.0525717   9.354  < 2e-16 ***
fem           0.1284292  0.0236221   5.437 6.21e-08 ***
educ        -0.1051514  0.0164543  -6.391 2.13e-10 ***
age           0.0033573  0.0008235   4.077 4.77e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

R Code

```
> confint(fit)
              2.5 %      97.5 %
(Intercept) 0.388662948 0.594887290
fem          0.082097691 0.174760775
educ        -0.137424277 -0.072878547
age          0.001742153 0.004972387
```

Testing Hypotheses about Individual Coefficients

Testing Linear Hypotheses: A Simple Case

Testing Joint Significance

Testing Linear Hypotheses: The General Case

Testing Hypothesis About a Linear Combination of β_j

R Code

```
> fit.dum <- lm(vote1 ~ fem + age + S + PS, data = d)
> summary(fit.dum)
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.4005593  0.0436242   9.182  < 2e-16 ***
fem           0.1280018  0.0236250   5.418  6.89e-08 ***
age           0.0032133  0.0008349   3.849  0.000123 ***
S            -0.1283286  0.0275935  -4.651  3.56e-06 ***
PS           -0.2047191  0.0333376  -6.141  1.02e-09 ***
```

- S and PS are dummies for secondary and post-secondary schooling, with primary as the baseline. Coding education as 1, 2, 3 assumes that each step up in schooling shifts the Yes probability by the same amount. In this model, that assumption is $\beta_{PS} = 2\beta_S$. So we test

$$H_0 : \beta_{PS} - 2\beta_S = 0 \quad \text{against} \quad H_1 : \beta_{PS} - 2\beta_S \neq 0$$

- What would be an appropriate **test statistic** for this hypothesis?

Testing Hypothesis About a Linear Combination of β_j

```
R Code
> fit.dum <- lm(vote1 ~ fem + age + S + PS, data = d)
> summary(fit.dum)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.4005593  0.0436242   9.182  < 2e-16 ***
fem           0.1280018  0.0236250   5.418  6.89e-08 ***
age           0.0032133  0.0008349   3.849  0.000123 ***
S            -0.1283286  0.0275935  -4.651  3.56e-06 ***
PS           -0.2047191  0.0333376  -6.141  1.02e-09 ***
```

- Let's consider a t-value:

$$T = \frac{\hat{\beta}_{PS} - 2\hat{\beta}_S}{\widehat{SE}(\hat{\beta}_{PS} - 2\hat{\beta}_S)}$$

We will reject H_0 if T is sufficiently different from zero.

- Unlike a test of one coefficient against a constant, here both $\hat{\beta}_{PS}$ and $\hat{\beta}_S$ are random, and they are correlated. So the standard error of the combination needs both variances and their covariance.

Worksheet Q1

Return to the Chile model with dummies S (secondary) and PS (post-secondary), with primary schooling as the baseline. The estimated model is

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_{fem} FEM + \hat{\beta}_{age} AGE + \hat{\beta}_S S + \hat{\beta}_{PS} PS.$$

We wish to test whether the 1-2-3 coding of education fits the data:

$$H_0 : \beta_{PS} - 2\beta_S = 0.$$

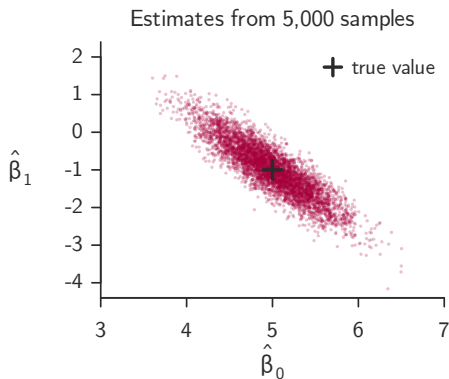
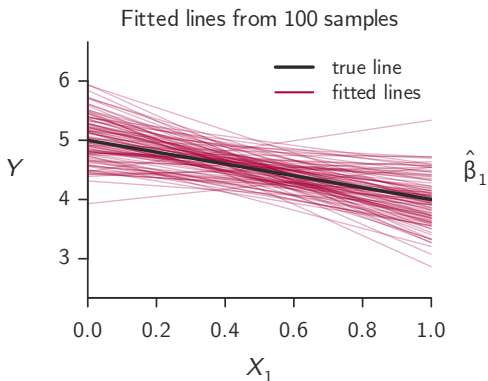
- ① Derive $\text{Var}(\hat{\beta}_{PS} - 2\hat{\beta}_S)$
- ② Let $\mathbf{A} = (\mathbf{X}'\mathbf{X})^{-1} = [a_{rs}]$ and $\widehat{\text{Var}}(\hat{\beta}) = \hat{\sigma}^2 \mathbf{A}$. Explain that

$$\widehat{\text{Var}}(\hat{\beta}_{PS} - 2\hat{\beta}_S) = \hat{\sigma}^2 (a_{PS,PS} + 4a_{S,S} - 4a_{S,PS})$$

is an unbiased estimator of $\text{Var}(\hat{\beta}_{PS} - 2\hat{\beta}_S \mid \mathbf{X})$ under Assumptions I–V.

Joint Normality: Simulation

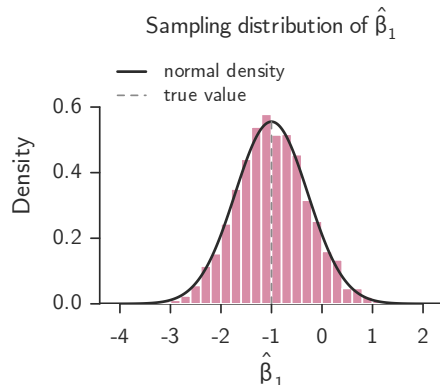
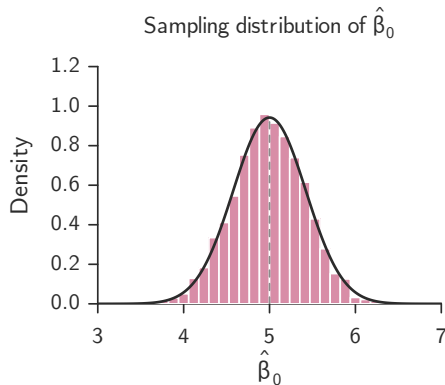
$Y = \beta_0 + \beta_1 X_1 + u$ with $u \sim N(0, \sigma_u^2 = 4)$ and $\beta_0 = 5$, $\beta_1 = -1$, and $n = 100$:



Source: simulation; X_1 drawn once from $U(0, 1)$ and held fixed across 5,000 samples.

Marginal Sampling Distribution

$Y = \beta_0 + \beta_1 X_1 + u$ with $u \sim N(0, \sigma_u^2 = 4)$ and $\beta_0 = 5$, $\beta_1 = -1$, and $n = 100$:

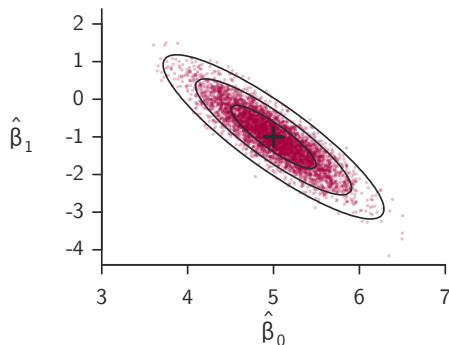


Source: simulation; X_1 drawn once from $U(0, 1)$ and held fixed across 5,000 samples.

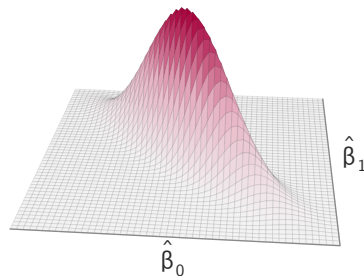
Joint Sampling Distribution

$Y = \beta_0 + \beta_1 X_1 + u$ with $u \sim N(0, \sigma_u^2 = 4)$ and $\beta_0 = 5$, $\beta_1 = -1$, and $n = 100$:

Draws and density contours



Bivariate normal density



The true **variance-covariance matrix** of the estimators, $\sigma_u^2(\mathbf{X}'\mathbf{X})^{-1}$, is:

	$\hat{\beta}_0$	$\hat{\beta}_1$
$\hat{\beta}_0$	.18	-.27
$\hat{\beta}_1$	-.27	.52

Source: simulation; X_1 drawn once from $U(0, 1)$ and held fixed across 5,000 samples. The draws give variances .18 and .50 and covariance $-.26$.

Example: Is the 1-2-3 Coding of Education OK?

R Code

```
> fit.dum <- lm(vote1 ~ fem + age + S + PS, data = d)
> V <- vcov(fit.dum)
> V
```

	(Intercept)	fem	age	S
(Intercept)	1.903067e-03	-3.393292e-04	-3.121865e-05	-6.786308e-04
fem	-3.393292e-04	5.581411e-04	1.019330e-06	4.685980e-05
age	-3.121865e-05	1.019330e-06	6.969858e-07	6.908035e-06
S	-6.786308e-04	4.685980e-05	6.908035e-06	7.613989e-04
PS	-6.961468e-04	7.344137e-05	6.987371e-06	4.232502e-04

	PS
(Intercept)	-6.961468e-04
fem	7.344137e-05
age	6.987371e-06
S	4.232502e-04
PS	1.111396e-03

Example: Is the 1-2-3 Coding of Education OK?

We can then compute the test statistic for the hypothesis of interest:

$$t = \frac{\hat{\beta}_{PS} - 2\hat{\beta}_S}{\widehat{SE}(\hat{\beta}_{PS} - 2\hat{\beta}_S)} \quad \text{where}$$
$$\widehat{SE}(\hat{\beta}_{PS} - 2\hat{\beta}_S) = \sqrt{\widehat{V}(\hat{\beta}_{PS}) + 4\widehat{V}(\hat{\beta}_S) - 4\widehat{\text{Cov}}[\hat{\beta}_S, \hat{\beta}_{PS}]}$$

R Code

```
> se <- sqrt(V["PS","PS"] + 4*V["S","S"] - 4*V["S","PS"])
> se
[1] 0.0496386
>
> tstat <- (coef(fit.dum)["PS"] - 2*coef(fit.dum)["S"])/se
> tstat
      PS
1.046326
```

Dropping the covariance term would inflate the SE from 0.0496 to 0.0645. Plugging in we get $t \simeq 1.05$. So what do we conclude?

We cannot reject the null ($p = 0.30$). The data are consistent with the 1-2-3 coding, under which the Yes-vote gap between post-secondary and primary is twice the gap between secondary and primary.

Testing Hypotheses about Individual Coefficients

Testing Linear Hypotheses: A Simple Case

Testing Joint Significance

Testing Linear Hypotheses: The General Case

F Test for Joint Significance of Coefficients

- In research we often want to test a **joint hypothesis** which involves **multiple linear restrictions** (e.g. $\beta_1 = \beta_2 = \beta_3 = 0$)
- Suppose our regression model is:

$$\text{Vote} = \beta_0 + \gamma_1 \text{FEMALE} + \beta_1 \text{EDUCATION} + \gamma_2 (\text{FEMALE} \cdot \text{EDUCATION}) + \beta_2 \text{AGE} + \gamma_3 (\text{FEMALE} \cdot \text{AGE}) + u$$

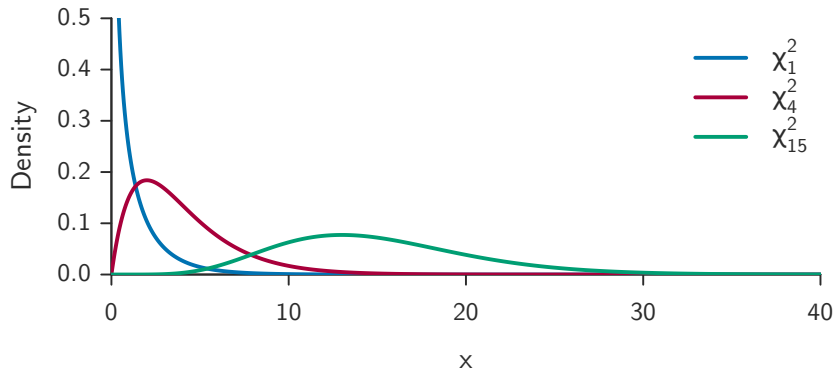
and we want to test

$$H_0 : \gamma_1 = \gamma_2 = \gamma_3 = 0.$$

- Substantively, what question are we asking?
 - Do females and males vote systematically differently from each other?
(Under the null, there is no difference in either the intercept or slopes between females and males).
- This is an example of a joint hypothesis test involving **three restrictions**: $\gamma_1 = 0$, $\gamma_2 = 0$, and $\gamma_3 = 0$.
- If the three estimates are jointly close to zero, relative to their sampling variability, we fail to reject the null of no gender difference.
- **F tests** allow us to test **joint hypotheses**

The χ^2 Distribution

- To test more than one hypothesis jointly we need some additional probability distributions.
- Suppose $Z_1, \dots, Z_n$ are n i.i.d. random variables following $\mathcal{N}(0, 1)$.
- Then, the **sum of their squares**, $X = \sum_{i=1}^n Z_i^2$, is distributed according to the **χ^2 distribution** with n degrees of freedom, $X \sim \chi_n^2$.



Properties: $X > 0$, $E[X] = n$ and $V[X] = 2n$. In R: `dchisq()`, `pchisq()`, `rchisq()`

The F distribution

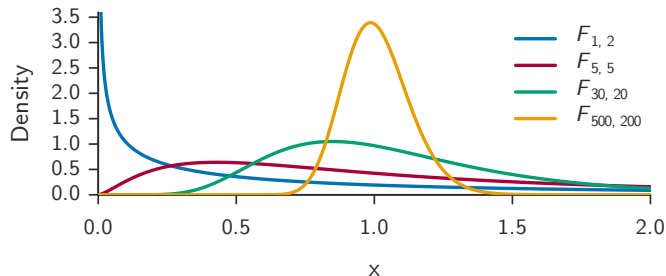
The **F distribution** arises as a ratio of two independent chi-squared distributed random variables:

$$F = \frac{X_1/df_1}{X_2/df_2} \sim \mathcal{F}_{df_1, df_2}$$

where $X_1 \sim \chi_{df_1}^2$, $X_2 \sim \chi_{df_2}^2$, and $X_1 \perp\!\!\!\perp X_2$.

df_1 and df_2 are called the **numerator degrees of freedom** and the **denominator degrees of freedom**.

In R: `df()`, `pf()`, `rf()`



F Test against $H_0 : \gamma_1 = \gamma_2 = \gamma_3 = 0$.

The F statistic can be calculated by the following procedure:

- 1 Fit the **unrestricted model (ur)** which *does not* impose H_0 :

$$\text{Vote} = \beta_0 + \gamma_1 \text{FEM} + \beta_1 \text{EDUC} + \gamma_2 (\text{FEM} * \text{EDUC}) + \beta_2 \text{AGE} + \gamma_3 (\text{FEM} * \text{AGE}) + u$$

- 2 Fit the **restricted model (r)** which *does* impose H_0 :

$$\text{Vote} = \beta_0 + \beta_1 \text{EDUC} + \beta_2 \text{AGE} + u$$

- 3 From the two results, compute the F statistic:

$$F_0 = \frac{(SSR_r - SSR_{ur})/q}{SSR_{ur}/(n - k - 1)}$$

where SSR =sum of squared residuals, q =number of restrictions, k =number of predictors in the unrestricted model, and n = # of observations.

Note that F_0 measures relative increase in (minimized) SSR when going from UR to R.

The F statistics have the following sampling distributions:

- If H_0 is true and Assumptions I–VI hold, $F_0 \sim \mathcal{F}_{q, n-k-1}$ regardless of the sample size.
- If H_0 is true and Assumptions I–V hold, $q F_0 \overset{a.s.}{\sim} \chi_q^2$ as $n \rightarrow \infty$ (see next section).

Unrestricted Model (UR)

R Code

```
> fit.UR <- lm(vote1 ~ fem + educ + age + fem:age + fem:educ, data = Chile)
> summary(fit.UR)
~~~~~
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.358278   0.068998   5.193 2.32e-07 ***
fem           0.421666   0.100858   4.181 3.05e-05 ***
educ        -0.071545   0.022437  -3.189  0.00146 **
age           0.005198   0.001094   4.752 2.19e-06 ***
fem:age      -0.004228   0.001657  -2.551  0.01082 *
fem:educ     -0.073680   0.032931  -2.237  0.02539 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4835 on 1697 degrees of freedom
Multiple R-squared:  0.06785,    Adjusted R-squared:  0.06511
F-statistic: 24.71 on 5 and 1697 DF,  p-value: < 2.2e-16
```

Restricted Model (R)

R Code

```
> fit.R <- lm(vote1 ~ educ + age, data = Chile)
> summary(fit.R)
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.5773293  0.0505810  11.414  < 2e-16 ***
educ        -0.1138543  0.0165133  -6.895  7.58e-12 ***
age           0.0031356  0.0008293   3.781  0.000162 ***
---
Residual standard error: 0.4886 on 1700 degrees of freedom
Multiple R-squared:  0.04662,    Adjusted R-squared:  0.0455
F-statistic: 41.56 on 2 and 1700 DF,  p-value: < 2.2e-16
```

R Code

```
> SSR.UR <- sum(resid(fit.UR)^2) # = 396.7
> SSR.R <- sum(resid(fit.R)^2) # = 405.8

> DFdenom <- df.residual(fit.UR) # = 1697
> DFnum <- 3

> F <- ((SSR.R - SSR.UR)/DFnum) / (SSR.UR/DFdenom)
> F
[1] 12.88682

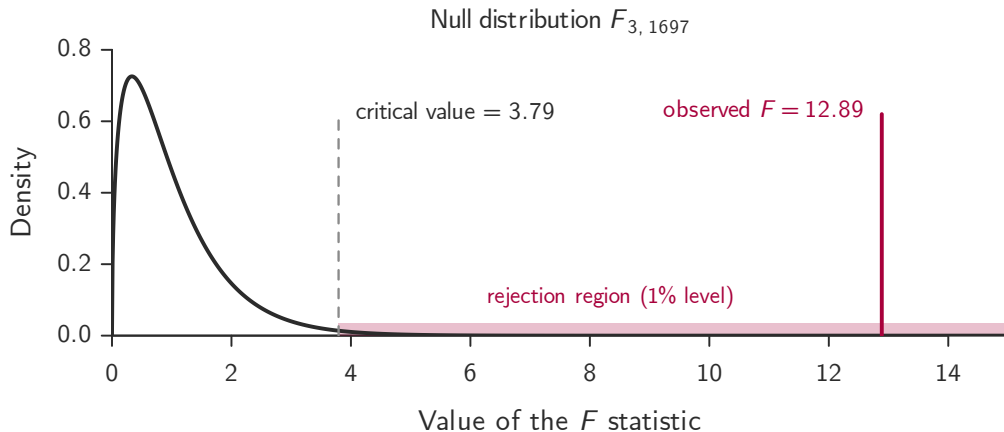
> qf(0.99, DFnum, DFdenom)
[1] 3.793171
```

Given above, what do we conclude?

$F_0 \approx 12.89$ is greater than the **critical value** for a .01 level test. So we *reject* the null hypothesis.

Null Distribution, Critical Value, and Test Statistic

Large values of F mean that imposing H_0 raised the SSR a lot, so we reject only in the right tail of $\mathcal{F}_{q,n-k-1}$ (in large samples, compare qF_0 with χ_q^2).



Worksheet Q2

Consider the multiple regression model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + u,$$

and suppose we wish to test the joint null hypothesis

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0 \quad \text{versus} \quad H_1 : \text{at least one } \beta_j \neq 0.$$

- ① Explain in words what this hypothesis is asking about the explanatory power of the regressors.
- ② Describe step-by-step how to conduct the F test for this hypothesis, including:
 - ① the unrestricted and restricted models,
 - ② how to compute the F statistic from the two sum of squared residuals, and
 - ③ how to determine whether to reject H_0 at significance level α .
- ③ Interpret the meaning of rejecting versus not rejecting H_0 in terms of the overall explanatory power of $X_1, \dots, X_k$.

```
> summary(fit.UR)
~~~~~

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.358278   0.068998   5.193 2.32e-07 ***
fem           0.421666   0.100858   4.181 3.05e-05 ***
educ        -0.071545   0.022437  -3.189  0.00146 **
age           0.005198   0.001094   4.752 2.19e-06 ***
fem:age      -0.004228   0.001657  -2.551  0.01082 *
fem:educ     -0.073680   0.032931  -2.237  0.02539 *
---

Residual standard error: 0.4835 on 1697 degrees of freedom
Multiple R-squared:  0.06785,    Adjusted R-squared:  0.06511
F-statistic: 24.71 on 5 and 1697 DF,  p-value: < 2.2e-16
```

Worksheet Q3

Consider the multiple regression model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_k X_k + u,$$

and suppose we want to test whether the first three regressors have the same effect:

$$H_0 : \beta_1 = \beta_2 = \beta_3 \quad \text{versus} \quad H_1 : \text{at least one differs.}$$

- ① Describe the step-by-step F test procedure, including:
 - ① how to construct the restricted model implied by H_0 ,
 - ② how to compute the F statistic using the two sums of squared residuals, and
 - ③ the reference distribution used for inference.
- ② Provide an intuitive interpretation of what rejecting or failing to reject H_0 implies about the coefficients on X_1 , X_2 , and X_3 in the regression for Y .

Testing Equality of Coefficients in R

R Code

```
> fit.dum <- lm(vote1 ~ fem + age + S + PS, data = d)
> summary(fit.dum)
~~~~~
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.4005593  0.0436242   9.182  < 2e-16 ***
fem           0.1280018  0.0236250   5.418 6.89e-08 ***
age           0.0032133  0.0008349   3.849 0.000123 ***
S            -0.1283286  0.0275935  -4.651 3.56e-06 ***
PS           -0.2047191  0.0333376  -6.141 1.02e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4845 on 1698 degrees of freedom
Multiple R-squared:  0.06353,    Adjusted R-squared:  0.06132
F-statistic: 28.8 on 4 and 1698 DF,  p-value: < 2.2e-16
```

Are the coefficients on S and PS statistically significantly different?

Under $H_0 : \beta_S = \beta_{PS}$, secondary and post-secondary voters differ from primary-school voters by the same amount.

Testing Equality of Coefficients in R

R Code

```
> d$Xstar <- d$S + d$PS
> fit.R2 <- lm(vote1 ~ fem + age + Xstar, data = d)

> SSR.UR2 <- sum(resid(fit.dum)^2)
> SSR.R2 <- sum(resid(fit.R2)^2)

> DFdenom <- df.residual(fit.dum)

> F <- ((SSR.R2 - SSR.UR2)/1) / (SSR.UR2/DFdenom)
> F
[1] 5.685995

> pf(F, 1, DFdenom, lower.tail = FALSE)
[1] 0.01721005
```

So, what do we conclude?

We reject the null that the two coefficients are the same at the .05 level, with a p-value of 0.017.

Consider the hypothesis test of

$$H_0 : \beta_1 = \beta_2 \quad \text{vs.} \quad H_1 : \beta_1 \neq \beta_2$$

What ways have we learned to conduct this test?

- Option 1: Compute $T = (\hat{\beta}_1 - \hat{\beta}_2) / \widehat{SE}(\hat{\beta}_1 - \hat{\beta}_2)$ and do the **t test**.
- Option 2: Create $X^* = X_1 + X_2$, fit the restricted model, compute $F = (SSR_R - SSR_{UR}) / (SSR_{UR} / (n - k - 1))$ and do the **F test**.

It turns out the F test and the two-sided t test give **identical** results. This is because $F = T^2$ in every sample, and

$$X \sim t_{n-k-1} \implies X^2 \sim \mathcal{F}_{1,n-k-1}$$

(The arrow does not reverse: $|X|^2 \sim \mathcal{F}_{1,n-k-1}$ as well, but $|X|$ is never negative, so it is not t distributed.)

- So, for one linear restriction, the F test and the two-sided t test give the same answer.
- Usually, the t test is used for single hypotheses and the F test is used for joint hypotheses.

Some More Notes on F Tests

- When both models have the same outcome variable and observations, the F -value can also be calculated from R^2 :

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

- F tests only work for testing **nested** models, i.e. the restricted model must be a special case of the unrestricted model.

For example F tests cannot be used to test

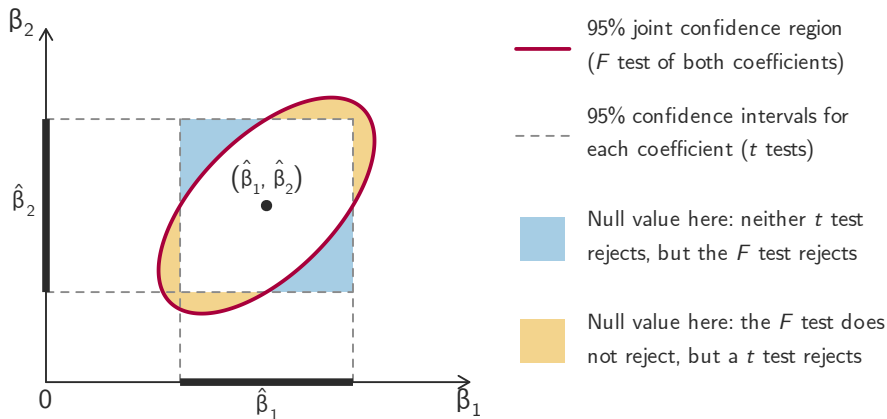
$$Y = \beta_0 + \beta_1 X_1 \quad + \beta_3 X_3 + u$$

against

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \quad + u$$

Some More Notes on F Tests

- Joint significance does not necessarily imply the significance of individual coefficients, or vice versa:



Source: Hayashi, *Econometrics* (2000), Fig. 1.5 ("t- versus F-tests"), redrawn

When t and F Disagree: Age and Age² in Chile

```
R Code
> fit.sq <- lm(vote1 ~ fem + educ + age + I(age^2), data = d)
> summary(fit.sq)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.438e-01  9.811e-02   4.523 6.50e-06 ***
fem          1.274e-01  2.370e-02   5.375 8.74e-08 ***
educ        -1.057e-01  1.648e-02  -6.412 1.86e-10 ***
age           6.105e-03  4.814e-03   1.268   0.205
I(age^2)     -3.301e-05  5.699e-05  -0.579   0.563

> fit.noage <- lm(vote1 ~ fem + educ, data = d)
> anova(fit.noage, fit.sq)
  Res.Df  RSS Df Sum of Sq    F    Pr(>F)
1    1700 402.73
2    1698 398.75  2     3.9807 8.4754 0.0002175 ***
```

Neither age term is significant on its own ($p = 0.20$ and 0.56), yet together they are ($F = 8.48$, $p = 0.0002$). Age and age² move together (correlation 0.99), so the data cannot tell which of the two terms carries the age pattern, and each SE is large. Dropping both terms raises the SSR from 398.75 to 402.73, more than chance would explain. In the figure on the last slide, our null value (0,0) lies in a blue area, inside the box but outside the ellipse.

Testing Hypotheses about Individual Coefficients

Testing Linear Hypotheses: A Simple Case

Testing Joint Significance

Testing Linear Hypotheses: The General Case

Limitation of the F Formula

Consider the following null hypothesis:

$$H_0 : \beta_1 = \beta_2 = \beta_3 = 3$$

or

$$H_0 : \beta_1 = 2\beta_2 = 0.5\beta_3 + 1$$

Can we test them using the F test?

To compute the F value, we need to fit the restricted model. How?

- Some restrictions are difficult to impose when fitting the model.
- Even when we can, the procedure will be ad hoc and require some creativity.
- Is there a general solution?

General Procedure for Testing Linear Hypotheses

- Notice that any set of q linearly independent linear restrictions can be written as

$$\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$$

where

- $\mathbf{R}$ is a $q \times (k + 1)$ matrix of prespecified coefficients on $\boldsymbol{\beta}$ (**hypothesis matrix**); $\mathbf{R}$ has rank q
- $\boldsymbol{\beta} = [\beta_0 \ \beta_1 \ \cdots \ \beta_k]'$
- $\mathbf{r}$ is a $q \times 1$ vector of prespecified constants

Worksheet Q4 Consider the linear regression model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + u,$$

with coefficient vector $\beta = [\beta_0 \ \beta_1 \ \beta_2 \ \beta_3]'$. We wish to test linear hypotheses of the form $R\beta = r$. Construct the corresponding R and r in the following two scenarios.

- 1 Test the hypothesis that all three slope coefficients equal 3:

$$H_0 : \beta_1 = \beta_2 = \beta_3 = 3.$$

Write this null hypothesis in matrix form $R\beta = r$ by explicitly constructing R and r .

- 2 Test the joint hypothesis

$$H_0 : \beta_1 = 2\beta_2 = 0.5\beta_3 + 1.$$

- Let's consider testing $H_0 : \mathbf{R}\beta = \mathbf{r}$, a set of q linearly independent linear restrictions.
- If H_0 is true, $\mathbf{R}\hat{\beta} - \mathbf{r}$ should be zero except for sampling variability.
- To formally evaluate the statistical significance of the deviation from zero, we must transform $\mathbf{R}\hat{\beta} - \mathbf{r}$ to a **statistic** that can be compared to a **reference distribution**.
- It turns out that the following **Wald statistic** can be used:

$$W = (\mathbf{R}\hat{\beta} - \mathbf{r})' \cdot [\hat{\sigma}^2 \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}']^{-1} \cdot (\mathbf{R}\hat{\beta} - \mathbf{r})$$

- Looks complicated? Let's figure out why this makes sense:
 - The first and last components are the deviations $\mathbf{R}\hat{\beta} - \mathbf{r}$; on their own they would give a sum of squares.
 - The middle component is the **inverse** of the estimated variance matrix of $\mathbf{R}\hat{\beta} - \mathbf{r}$. It puts each deviation on a standard scale and removes their correlation, so that under H_0 the statistic is roughly a sum of q squared standard normals.
- We know $\hat{\beta}$ is approximately normal $\Rightarrow \mathbf{R}\hat{\beta} - \mathbf{r}$ should also be normal
 $\Rightarrow W$ should therefore be ... χ^2 distributed!

Sampling Distribution of the Wald Statistic

Theorem (Large-Sample Distribution of the Wald Statistic)

If H_0 is true, then under Assumptions I–V, as $n \rightarrow \infty$ the distribution of the Wald statistic approaches the chi square distribution with q degrees of freedom:

$$W \xrightarrow{d} \chi_q^2 \text{ as } n \rightarrow \infty$$

Theorem (Small-Sample Distribution of the Wald Statistic)

If H_0 is true, then under Assumptions I–VI, for any sample size n the Wald statistic divided by q has the F distribution with $(q, n - k - 1)$ degrees of freedom:

$$W/q \sim \mathcal{F}_{q, n-k-1}$$

- $q\mathcal{F}_{q, n-k-1} \xrightarrow{d} \chi_q^2$ as $n \rightarrow \infty$, so the difference disappears when n large.

```
> pf(3.1, 2, 500, lower.tail=FALSE) [1] 0.04591619
```

```
> pchisq(2*3.1, 2, lower.tail=FALSE) [1] 0.0450492
```

```
> pf(3.1, 2, 50000, lower.tail=FALSE) [1] 0.04505786
```

Testing General Linear Hypotheses in R

In R, the `linearHypothesis()` function in the `car` package does the Wald test for general linear hypotheses.

R Code

```
> fit.dum <- lm(vote1 ~ fem + age + S + PS, data = d)
> R <- matrix(c(0, 0, 0, -2, 1), nrow = 1)
> r <- 0
> linearHypothesis(fit.dum, R, r)
Linear hypothesis test:
- 2 S + PS = 0
```

Model 1: restricted model

Model 2: `vote1 ~ fem + age + S + PS`

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	1699	398.83				
2	1698	398.57	1	0.25698	1.0948	0.2956

This is the test of the 1-2-3 coding again, and $F = 1.09 = 1.046^2 = t^2$. Here the restricted model is also easy to fit.

Under H_0 , $\beta_S S + \beta_{PS} PS = \beta_S (S + 2PS) = \beta_S (\text{educ} - 1)$, so the restricted model is our first model,

`lm(vote1 ~ fem + educ + age)`.