F-test motivation

⇒ We want to test the significance of a group of hypotheses (multiple hypotheses)

$$Grade_{325} = \beta_0 + \beta_1 \#times\_front + \beta_2 \#times\_back + \beta_3 hr\_study + \beta_4 past\_GPA + \beta_5 gender + u$$

H_0: seat position doesn't have impact on GPA

$\beta_1 = 0$ and $\beta_2 = 0$ $\Rightarrow$ $\beta_1 = \beta_2 = 0$

H_a: seat position matters

$\beta_1 \neq 0$ and $\beta_2 \neq 0$

or $\beta_1 \neq 0$ and $\beta_2 = 0$

or $\beta_1 = 0$ and $\beta_2 \neq 0$

} at least one of the $\beta_1, \beta_2 \neq 0$

7 Testing Multiple Linear Restrictions: The F-test

Suppose the model is specified by

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

H_0 : $\beta_2 = 0$ and $\beta_3 = 0$ → want to test if X_1 and X_2 BOTH have no impact on Y.

H_a, H_1 : H_0 is not true

We can use the F-test to test this type of "multiple hypotheses".

1. Our full model is called the "unrestricted" model (ur). Suppose it can be expressed as: ← Big model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$ is true ⇒ Reject H_0

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

2. The model which takes out x (which we think its associated $\beta = 0$) is called the restricted model (r). ← small model

$$y = \beta_0 + \beta_1 x_1 + u$$ is true ⇒ do not reject H_0

Suppose there are "q" number of β that we would like to perform a joint-test of =0
- e.g. in this model q = 2

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_{k-q} x_{k-q} + u$$

H_0: $\beta_{k-q+1} = \beta_{k-q+2} = \ldots = \beta_k = 0$

(the last q βs = 0)

H_a: H_0 is not true.

$$y = \underbrace{\beta_0 + \beta_1 x_1 + \beta_2 x_2 \ldots + \beta_{k-q} x_{k-q}}_{(r)} + \beta_{k-q+1} x_{k-q+1} + \beta_{k-q+2} x_{k-q+2} \ldots + \beta_k x_k + u$$

ur

$$F \equiv \frac{(SSR_r - SSR_{ur})}{q} \Big/ \frac{SSR_{ur}}{(n-k-1)}$$

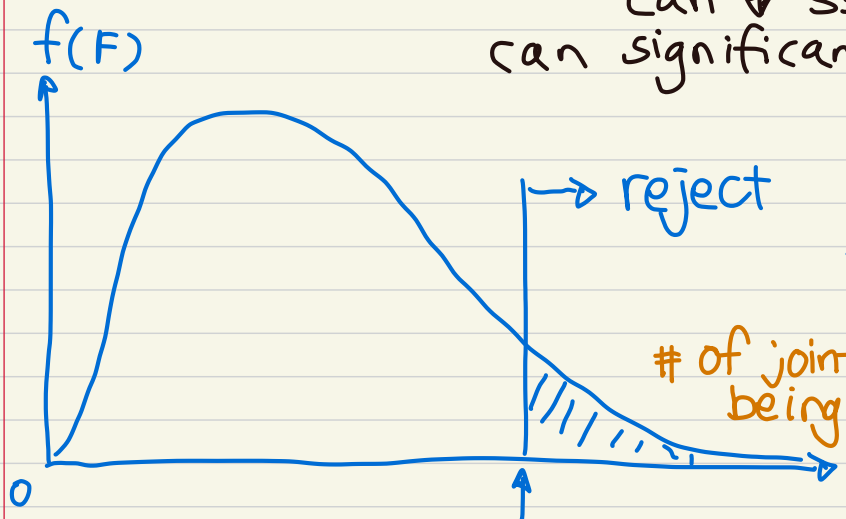
This is always (+) b/c $SSR_{ur} < SSR_r$. Every time you add 1 more x, the model will be better explained.

(n−k−1) ← d.f. of the "ur" model.

- So, if every time you add 1 more X variable, the SSR ↓ and R^2 ↑, why don't we just keep the additional X in the model ??

⇒ Because every time we add 1 more X, $var(\hat{\beta}_s)$ will increase, making the prediction of β less precise. So, we only keep the addition X_s if it/they can improve the model enough

can ↓ SSR (↑ R^2) enough.
can significantly ↓ SSR and ↑ R^2



H_0: $\beta_2 = \beta_3 = \ldots = 0$
H_a: H_0 not true

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

q: # of joint Hypotheses being tested
n-k-1: d.f. of the ur. model

We reject H_0 of jointly no effect if $F > C$

3. Some useful facts

① $R^2_{ur} > R^2_r$ because any additional X would increase R^2 (improve fit).
$\Rightarrow SSR_{ur} < SSR_r$

② By including more X, the model is certainly better explained. However, we would like to reject H_0 if the inclusion of extra variables does not improve the model enough.

4. Other ways to calculate the F-statistics:

$\Rightarrow$ From $R^2 = 1 - \frac{SSR}{SST}$ (SSR → RSS, SST → TSS)

We have $F \equiv \frac{(R^2_{ur} - R^2_r)/q}{(1 - R^2_{ur})/(n-k-1)}$

q: # of β that are set to "0"

k: # of slope β (intercept)

n: # of obs.

$\Rightarrow$ If we want to test the overall significance of the model

$H_0: \beta_1 = \beta_2 = \beta_3 = \dots = \beta_k = 0$, H_a = otherwise

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

R^2 of the model $\approx$ UR

the "r" model has no X at all.

Example: Suppose we are interested in understanding the determinant of a baseball player's salary.

y	*salary*	= season salary
r, ur	*years*	= years in major leagues
r, ur	*gamesyr*	= games per year in the league
r, ur	*bavg*	= career batting average
r, ur	*hrunsyr*	= homeruns per year
r, ur	*rbisyr*	= runs batted in per year

If we want to test whether performance has any impact on salary

$H_0: \beta_{bavg} = \beta_{hrunsyr} = \beta_{rbisyr} = 0$

H_a: otherwise is true

- the unrestricted model (ur) is defined by

ur model

y x

```
. regress log_salary 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

log_salary	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

1 2 3 4 5 intercept

- the restricted model (r) is defined by

When considering each of the performance X one-by-one, none of them has a significant impact at 5%

y x

```
. regress log_salary years gamesyr
```

Source	SS	df	MS
SSE Model	293.864058	2	146.932029
SSR Residual	198.311477	350	.566604221
SST Total	492.175535	352	1.39822595

Number of obs = 353
F(2, 350) = 259.32
Prob > F = 0.0000
R-squared = 0.5971
Adj R-squared = 0.5948
Root MSE = .75273

log_salary	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
years	.071318	.012505	5.70	0.000	.0467236	.0959124
gamesyr	.0201745	.0013429	15.02	0.000	.0175334	.0228156
_cons	11.2238	.108312	103.62	0.000	11.01078	11.43683

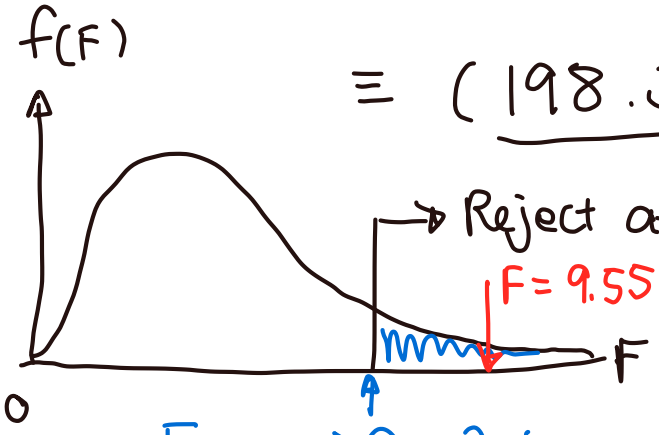
But when performing an F-test, performances have joint impact.

HW:
$F = \frac{(R^2/q)}{(1-R^2)/(n-k-1)} \equiv ??$

Now, our H_0 and H_a becomes

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

$$\equiv \frac{(198.311 - 183.186)/3}{(183.86)/(353-5-1)} \approx 9.55$$



Let's use 5% level of sig.

Since $F = 9.55 > 2.6$, we reject H_0 at 5% level and conclude that performances have joint effects on salary.

8 How the Hypothesis Testing is done in Practice

1. Check the values of $t-statistic$ reported by the statistical software (i.e. STATA, SPSS, SAS)

   $\Rightarrow$ These $t-statistics$ are to test $H_0 : \beta_i = 0$

   $\Rightarrow$ If the d.f. > 30, then when $t > 1.96$, we can reject H_0 with 5% sig. level (z-table)

   $\Rightarrow$ **When $t > 1.96$**, we can say that β_i is **statistically significant** at 5% level. (value of $\beta_i \neq 0$)

   $\Rightarrow$ **When $t < 1.96$** we can say that β_i is **not statistically significant** at 5% level.

   $\Rightarrow$ If $t < 1.96$ we can drop x_i from the model

   $\Rightarrow$ After we drop x_i, we estimate the new regression function and obtain a new set of $\hat{\beta}$.

2. We can also perform other hypothesis testings of interest.

   e.g. $H_0 : \beta_i = \beta_j$

   or $H_0 : \beta_i = 5$ etc.

   or perform an F-test for testing multiple linear restrictions

3. Usually, in economics, the estimation results are reported using this form

Dependent Variable: log(*salary*)			
Independent Variables	**(1)**	**(2)**	**(3)**
log(*sales*)	.224 (.027)	.158 (.040)	.188 (.040)
log(*mktval*)	——	.112 (.050)	.100 (.049)
profmarg	——	−.0023 (.0022)	−.0022 (.0021)
ceoten	——	——	.0171 (.0055)
comten	——	——	−.0092 (.0033)
intercept	4.94 (0.20)	4.62 0.25)	4.57 (0.25)
Observations *R*-squared	177 .281	177 .304	177 .353

sales → other company performance

CEO characteristics

like a simple regression with 1 X.

Multiple Regression Analysis : Further Issues

1 Data scaling on OLS statistics

When we change the unit of measurment of a variable, the value of estimators would chance accordingly. For example

$$\widehat{bweght}_g = \hat{\beta}_0 + \hat{\beta}_1 cigs + \hat{\beta}_2 fa\min c,$$

where
$bwght$ = child birth weight, in grams.
$cigs$ = number of cigarettes smoked by the mother while pregnant, per day.
$fa\min c$ = annual family income, in thousands of dollars.

- What if we use bweght in kilograms??

1 kg. = 1,000 g.

$$\widehat{bweght}_{kg} = \frac{\widehat{bweght}_g}{1000} = \frac{\hat{\beta}_0}{1000} + \frac{\hat{\beta}_1}{1000} cigs + \frac{\hat{\beta}_2}{1000} faminc$$

$$= \hat{\alpha}_0 + \hat{\alpha}_1 cigs + \hat{\alpha}_2 faminc$$

$$\Rightarrow \hat{\alpha}_0 = \frac{\hat{\beta}_0}{1000}, \quad \hat{\alpha}_1 = \frac{\hat{\beta}_1}{1000}, \quad \hat{\alpha}_2 = \frac{\hat{\beta}_2}{1000}$$

- What if we use faminc in USD (instead of 1000 USD)

$$bweght_g = \hat{\beta}_0 + \hat{\beta}_1 cigs + \frac{\hat{\beta}_2}{1000} faminc_{USD}$$

$$= \hat{\beta}_0 + \hat{\beta}_1 cigs + \hat{\theta}_2 faminc_{USD}$$

$$\Rightarrow \hat{\theta}_2 = \frac{\hat{\beta}_2}{1000}$$

The value of this variable is going to be 1000 times larger than faminc in income

in other words $\hat{\theta}_2$ = impact of 1 USD↑ in income

$\hat{\beta}_2$ = ⟶ 1000 USD↑ in income.

- What if we use bweght in kg & income in THB

$$bweght_{kg} = \frac{\hat{\beta}_0}{1000} + \frac{\hat{\beta}_1}{1000} cigs + \frac{\frac{\hat{\beta}_2}{1000}}{30{,}000} faminc_{THB}$$

This value is going to be 30,000 times more than faminc

2 More on functional forms

- Logarithmic Functional Form

$\Delta Y = Y_1 - Y_2$
$\Delta X_1 = X_{11} - X_{12}$

usually means natural log!

$$\log(y) = \beta_0 + \beta_1 \log(x_1) + \beta_2 x_2 + u$$

- $\beta_1 = \frac{d\log(y)}{d\log(x_1)} = \frac{\frac{1}{y}dy}{\frac{1}{x_1}dx_1} = \frac{\frac{1}{y}\Delta y}{\frac{1}{x_1}\Delta x_1} = \frac{100 \times \frac{1}{y}\Delta y}{100 \times \frac{1}{x_1}\Delta x_1} = \frac{\%\Delta y}{\%\Delta x}$

with the log y & log x format, the coefficient is going to be the elasticity! (x_1 elasticity of y)
(price) (demand)

- $\beta_2 = \frac{d\log(y)}{dx_2} = \frac{\frac{1}{y}dy}{dx_2} = \frac{\frac{1}{y}\Delta y}{\Delta x_2}$

⇒ if we want the upper term to be % change, then

$$100\beta_2 = \frac{100\frac{1}{y}\Delta y}{\Delta x_2}$$

$$100\beta_2 = \frac{\%\Delta y}{\Delta x_2}$$

$100\beta_2$ = %Δ in Y given that x_2 increases by 1 unit.

- Models with Quadratics (squares)

⇒ capture increasing/decreasing marginal effects (slope of the relationship between X & Y is not constant.

COVID-19 example

Y (# of cases) ⇒ $y = \beta_0 + \beta_1 x + \beta_2 x^2 + u$

$\frac{dy}{dx} = \beta_1 + 2\beta_2 x$

(+) (+) days

X (days)

Decreasing returns.

profit (π) ⇒ $y = \beta_0 + \beta_1 x + \beta_2 x^2 + u$

$\frac{dy}{dx} = \beta_1 + 2\beta_2 x$

(+) (−)

unit (q)

q_0

Assume

$\pi = (P - mc)q$; $mc = 10$

$\pi = (100 - q - 10)q$ Demand: $P = 100 - q$

F.O.C $\frac{\partial \pi}{\partial q} = 0 = 90 - 2Q$

β_1 is positive

β_2 is (−)

Example : Effects of Pullution on Housing Prices

$$\log(price) = \beta_0 + \beta_1 \log(nox) + \beta_2 \log(dist) + \beta_3 rooms + \beta_4 room^2 + \beta_5 stratio + u$$

Derivatives of exponential and logarithmic functions [edit]

$$\frac{d}{dx}\left(c^{ax}\right) = ac^{ax}\ln c, \qquad c > 0$$

the equation above is true for all c, but the derivative for $c < 0$ yields a complex number.

$$\frac{d}{dx}\left(e^{ax}\right) = ae^{ax}$$

$$\frac{d}{dx}\left(\log_c x\right) = \frac{1}{x\ln c}, \qquad c > 0, c \neq 1$$

the equation above is also true for all c, but yields a complex number if $c < 0$

$$\frac{d}{dx}\left(\ln x\right) = \frac{1}{x}, \qquad x > 0.$$

$\frac{d\ln x}{dx} = \frac{1}{x} \Rightarrow d\ln(x) = \frac{1}{x}dx$

$$\frac{d}{dx}\left(\ln|x|\right) = \frac{1}{x}.$$

$$\frac{d}{dx}\left(x^x\right) = x^x(1+\ln x).$$

$$\frac{d}{dx}\left(f(x)^{g(x)}\right) = g(x)f(x)^{g(x)-1}\frac{df}{dx} + f(x)^{g(x)}\ln\left(f(x)\right)\frac{dg}{dx}, \qquad \text{if } f(x) > 0, \text{ and if } \frac{df}{dx} \text{ and } \frac{dg}{dx} \text{ exist.}$$

$$\frac{d}{dx}\left(f_1(x)^{f_2(x)^{(\ldots)^{f_n(x)}}}\right) = \left[\sum_{k=1}^{n}\frac{\partial}{\partial x_k}\left(f_1(x_1)^{f_2(x_2)^{(\ldots)^{f_n(x_n)}}}\right)\right]\Bigg|_{x_1=x_2=\ldots=x_n=x}, \text{ if } f_{i<n}(x) > 0 \text{ and } \frac{df_i}{dx} \text{ exists.}$$

Logarithmic derivatives [edit]

The logarithmic derivative is another way of stating the rule for differentiating the logarithm of a function (using the chain rule):

$$(\ln f)' = \frac{f'}{f}$$ wherever f is positive.

Logarithmic differentiation is a technique which uses logarithms and its differentiation rules to simplify certain expressions before

where
price = housing price
nox = level of pollution
dist = distance from downtown
rooms = number of rooms
stratio = average student per teacher ratio
The estimation result is given by

→ In the US or many other countries, students can apply to schools in the area without having to take any test. So, the lower stratio, the better the school

```
regress lprice lnox dist rooms rooms_sq stratio
```

Source	SS	df	MS	Number of obs =	506
Model	51.4933152	5	10.298663	F(5, 500) =	155.62
Residual	33.0889098	500	.06617782	Prob > F =	0.0000
				R-squared =	0.6088
Total	84.582225	505	.167489554	Adj R-squared =	0.6049
				Root MSE =	.25725

log(price)

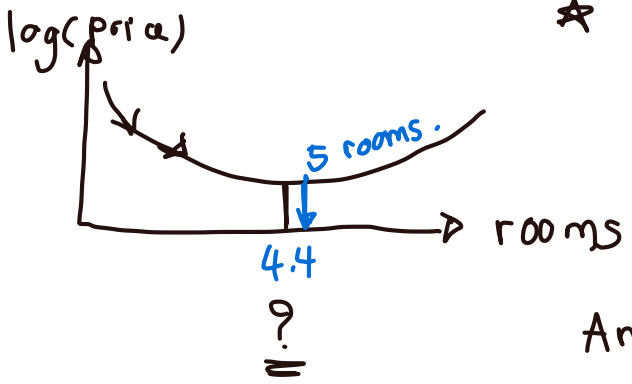
log(nox)

lprice	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lnox (β_1)	-.9767545	.0995938	-9.81	0.000	-1.172429	-.7810806
dist (β_2)	-.0321972	.0094013	-3.42	0.001	-.050668	-.0137264
rooms (β_3)	-.5528032	.1612965	-3.43	0.001	-.8697056	-.2359007
rooms_sq (β_4)	.0624697	.0124867	5.00	0.000	.0379368	.0870025
stratio (β_5)	-.0486667	.0058131	-8.37	0.000	-.0600879	-.0372455
_cons	13.59154	.5650901	24.05	0.000	12.4813	14.70178

$|t| > 1.96$ ↑ ↑ all < 0.05
↝ all variables are significant

Consider the effect of "room"

$$\frac{d\log(price)}{d\,rooms} = \beta_3 + 2\beta_4 rooms = -0.553 + 2(0.062)\cdot rooms$$



★ at how many rooms does 1 additional room has a positive impact on log(price) ??

$$0 = -0.553 + 2(0.062)\cdot rooms$$

$$rooms = 4.4$$

Answer ⇒ at 4.4 rooms or more
at → 5 rooms or more.

What wold be the % change in price when the number of room increases from 5 to 6?

- $\frac{d\log(price)}{d\,rooms} = -0.553 + 2(0.062)\cdot rooms$

total %Δ in price when # rooms ↑ from 5 to 7 is 6.7 + 19.1 % = 25.8%

$$\frac{100 \cdot \frac{1}{price} dprice}{d\,rooms} = 100(-0.553 + 2(0.062)\cdot 5)$$

$$= 100 \times 0.067 = 6.7\% \text{ increase.}$$

→ What about % in price when # rooms increases from 5 to 7 ??

$$\%\Delta price = 100(-0.553 + 2(0.062)\cdot 6) = 19.1\%$$

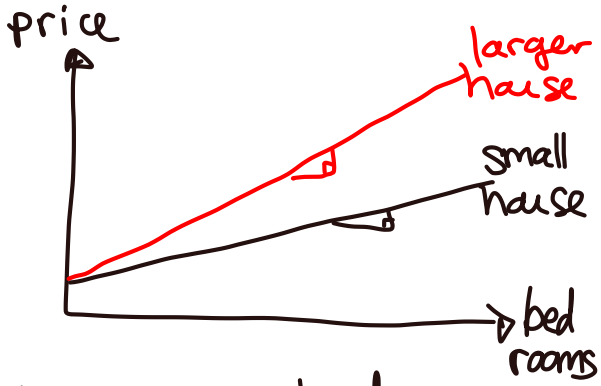
3 Models with Interaction Terms ⇒ Used when the impact of one variable depends on the value (level) of another variable.

Consider

$$price = \beta_0 + \beta_1 sqrft + \beta_2 bdrms + \beta_3 sqrft \times bdrms + \beta_4 bthrms + u$$

X_1 X_2 $X_3 = X_1 \cdot X_2$ X_4

where
- $price$ = housing price
- $sqrft$ = house size (square feet)
- $bdrms$ = number of bedrooms
- $bthrms$ = number of bathrooms



- $\frac{\partial price}{\partial bdrms} = \beta_2 + \beta_3 Sqrft$

⇒ if $\beta_2 > 0$ then, an additional bedroom would increase price more for a larger house!

4 More on the Goodness-of-Fit and Selection of Regressors

- Adding more regressors ALWAYS improve fit → R^2 always ↑

- But we lose the "degree of freedom" (d.f. = free data point used to estimate the parameter)
  → 1 data point is sacrificed every time we estimate a parameter.
- using R^2 would not punish "having too many regressors"
- We use adjusted-R^2 or $\bar{R}^2$ when we want to punish adding too many regressors.

$$R^2 = 1 - \frac{SSR}{SST} = 1 - \frac{SSR/n}{SST/n}$$

$$\text{adj. } R^2 = 1 - \frac{SSR/(n-k-1)}{SST/(n-1)}$$

Using adjusted R-squared to choose between non-nested models (one model is not a subset of another).

Consider Model 1

If we have more k, d.f. = n-k-1 ↓, SSR/(n-k-1) ↑, adj-R^2 ↓

$$\widehat{salary} = \underset{(223.90)}{830.63} + \underset{(0.0089)}{0.0163}sales + \underset{(11.08)}{19.63}roe$$
$$n = 209, \quad R^2 = 0.029, \quad \bar{R}^2 = 0.020$$

Consider Model 2

$$\widehat{\log(salary)} = \underset{(0.29)}{4.36} + \underset{(0.033)}{0.2751}\log(sales) + \underset{(0.004)}{0.0179}roe$$
$$n = 209, \quad R^2 = 0.282, \quad \bar{R}^2 = 0.275$$

27.5% of variation in Y is explained. So, this model is better!