7 Testing Multiple Linear Restrictions: The F-test

Suppose the model is specified by

$$
\begin{aligned}
y &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u \\
H_0 &: \beta_2 = 0 \text{ and } \beta_3 = 0 \\
H_a, H_1 &: H_0 \text{ is not true}
\end{aligned}
$$

→ want to test if x_1 and x_2 Both have no impact

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

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

   $y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + u$ → is true or Reject H_0

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

2. The model which takes out x (which we think its associated $\beta = 0$) is called the restricted model (small model) (r). $y = \beta_0 + \beta_1 X_1 + u$ → is true or 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 $q=2$

  $y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_{K-q} X_{K-q} + u$

  H_0 : $\beta_{K-q+1} = \beta_{K-q+2} = \dots = \beta_K = 0$

  (the last q βs $=0$)

  H_a : H_0 is not true

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

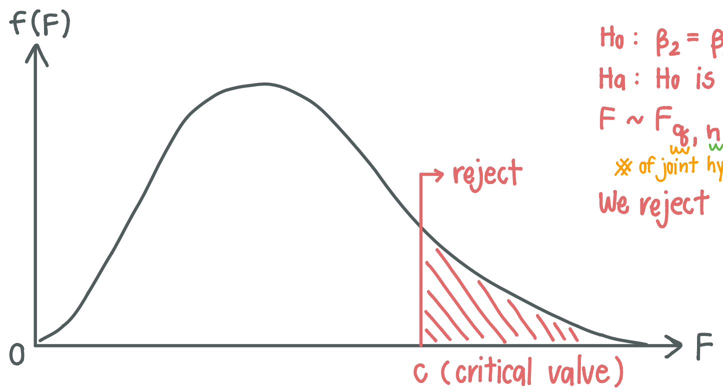
※ more explanatory variable : less residual

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

This is always positive b/c $SSR_{ur} < SSR_r$
Every time you add 1 more x, the model will be better explained. (SSR smaller, r^2 improve)
but variance of each β will be increase, they won't be predicted precisely

$n-k-1$ → degree of the "ur" model

- So, every time you add 1 more x variable, the $SSR \downarrow$ and $R^2 \uparrow$
  Why don't we just keep the addition x in the modle
  → Because every time we add 1 more x, $var(\hat{\beta}_5)$ will increase, making the prediction of β less precise. So, we only keep the addition x_3 if it/they can improve the model enough or can significantly reduce residual and improve R^2
  can $\downarrow SSR$ ($\uparrow R^2$) enough



H_0 : $\beta_2 = \beta_3 = 0$
H_a : H_0 is not true
$F \sim F_{q,\, n-k-1}$
※ of joint hypotheses being tested
d.f. of the ur. model
unrestricted
We reject H_0 of jointly no effect (no impact) if $F > c$

3. Some useful facts

① $R^2_{ur} > R^2_r$ b/c any additional x would increase R^2 (improve fit)

$\Rightarrow SSR_{ur} < SSR_r$

② By increase more x_1, the model is certainly better explained. However, we would like to reject Ho if the inclusion of extra variables do not improve the model enough.

4. Other ways to calculate the F-statistics:

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

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

q: # of β that are set to "0"

n: # of observation; k: # of slop β; 1: intercept

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

Ho: $\beta_1 = \beta_2 = \ldots = \beta_K = 0$

Ha: Otherwise

$$F = \dfrac{\dfrac{R^2}{K}}{\dfrac{(1-R^2)}{n-k-1}}$$

R^2 of the model $\approx$ UR (equivalence to) the "r" model has no X at all.

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

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

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

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

Ha: Otherwise is true

• the unrestricted model (ur) is defined by y x

ur model

```
. regress log_salary 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

------------------------------------------------------------------------------
  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
------------------------------------------------------------------------------
```

same observation

62.78% of the variables y was explained

the more x variables, the more R^2

$t < 1.96$

1 2 3 4 5

intercept $K = 5$

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

- the restricted model (r) is defined by

y x

```
. regress log_salary years gamesyr

      Source |       SS           df       MS      Number of obs   =       353
-------------+----------------------------------   F(2, 350)       =    259.32
       Model |  293.864058         2  146.932029   Prob > F        =    0.0000
    Residual |  198.311477       350  .566604221   R-squared       =    0.5971
-------------+----------------------------------   Adj R-squared   =    0.5948
       Total |  492.175535       352  1.39822595   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
------------------------------------------------------------------------------
```

SSE
SSR
SST

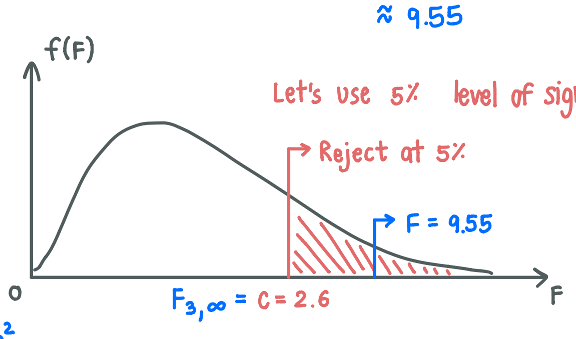
- But when performing an F-test, performance has joint impact

Now, our H_0 and H_a becomes

$$F \equiv \frac{\frac{(SSR_r - SSR_{ur})}{q}}{\frac{SSR_{ur}}{n-k-1}} \equiv \frac{\frac{198.311 - 183.186}{3}}{\frac{183.186}{353-5-1}}$$

$$\approx 9.55$$

Let's use 5% level of significant



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

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

TABLE D.3 UPPER PERCENTAGE POINTS OF THE *F* DISTRIBUTION (*Continued*)

df for denominator N_2	Pr	1	2	3	4	5	6	7	8	9	10	11	12
		df for numerator N_1											
22	.25	1.40	1.48	1.47	1.45	1.44	1.42	1.41	1.40	1.39	1.39	1.38	1.37
	.10	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90	1.88	1.86
	.05	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.26	2.23
	.01	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.18	3.12
24	.25	1.39	1.47	1.46	1.44	1.43	1.41	1.40	1.39	1.38	1.38	1.37	1.36
	.10	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88	1.85	1.83
	.05	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.21	2.18
	.01	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.09	3.03
26	.25	1.38	1.46	1.45	1.44	1.42	1.41	1.39	1.38	1.37	1.37	1.36	1.35
	.10	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86	1.84	1.81
	.05	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.18	2.15
	.01	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	3.02	2.96
28	.25	1.38	1.46	1.45	1.43	1.41	1.40	1.39	1.38	1.37	1.36	1.35	1.34
	.10	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84	1.81	1.79
	.05	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.15	2.12
	.01	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.96	2.90
30	.25	1.38	1.45	1.44	1.42	1.41	1.39	1.38	1.37	1.36	1.35	1.35	1.34
	.10	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82	1.79	1.77
	.05	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.13	2.09
	.01	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.91	2.84
40	.25	1.36	1.44	1.42	1.40	1.39	1.37	1.36	1.35	1.34	1.33	1.32	1.31
	.10	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76	1.73	1.71
	.05	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.04	2.00
	.01	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.73	2.66
60	.25	1.35	1.42	1.41	1.38	1.37	1.35	1.33	1.32	1.31	1.30	1.29	1.29
	.10	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71	1.68	1.66
	.05	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.95	1.92
	.01	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.56	2.50
120	.25	1.34	1.40	1.39	1.37	1.35	1.33	1.31	1.30	1.29	1.28	1.27	1.26
	.10	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65	1.62	1.60
	.05	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.87	1.83
	.01	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.40	2.34
200	.25	1.33	1.39	1.38	1.36	1.34	1.32	1.31	1.29	1.28	1.27	1.26	1.25
	.10	2.73	2.33	2.11	1.97	1.88	1.80	1.75	1.70	1.66	1.63	1.60	1.57
	.05	3.89	3.04	2.65	2.42	2.26	2.14	2.06	1.98	1.93	1.88	1.84	1.80
	.01	6.76	4.71	3.88	3.41	3.11	2.89	2.73	2.60	2.50	2.41	2.34	2.27
∞	.25	1.32	1.39	1.37	1.35	1.33	1.31	1.29	1.28	1.27	1.25	1.24	1.24
	.10	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60	1.57	1.55
	.05	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.79	1.75
	.01	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.25	2.18

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 $H_0 : \beta_i = 0$

   z-table

   $\Rightarrow$ If the d.f. > 30, then when $t \;\; > 1.96$, we can reject H_0 with 5% significant level

   $\Rightarrow$ **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 $\quad H_0 : \beta_i = 5 \quad$ etc.

   or perform an F-test for testing multiple linear restrictions

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

1% increase in sales, will increase salary by 1%

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

impact of sales decrease

the impact get carry by other factors

like a simple regression with 1 x variable

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

$$\underset{(g)}{\widehat{bweght}} = \widehat{\beta}_0 + \widehat{\beta}_1 cigs + \widehat{\beta}_2 fa\min c,$$

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 is we use the wage in kilograms
  1 kg. = 1,000 g.

$$\widehat{bweght}_{kg} = \frac{\widehat{bweght}_g}{1,000} = \frac{\hat{\beta}_0}{1,000} + \frac{\hat{\beta}_1}{1,000} cigs + \frac{\hat{\beta}_2}{1,000} faminc$$
$$= \hat{\alpha}_0 + \hat{\alpha}_1 cigs + \hat{\alpha}_2 faminc$$
$$\Rightarrow \hat{\alpha}_0 = \frac{\hat{\beta}_1}{1,000}, \quad \hat{\alpha}_1 = \frac{\hat{\beta}_1}{1,000}, \quad \hat{\alpha}_2 = \frac{\hat{\beta}_2}{1,000}$$

- What if we use the faminc in USD (instead of 1,000 USD)

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

The value of this variable is going to be 1,000 times larger than faminc

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

$\Rightarrow \hat{\theta}_2 = \frac{\hat{\beta}_2}{1,000}$ in other words $\hat{\theta}_2$ = impact of 1 USD ↑ in income
$\hat{\beta}_2$ = impact of 1,000 USD ↑ in income

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

$$bweght_{kg} = \frac{\hat{\beta}_0}{1,000} + \frac{\hat{\beta}_1}{1,000} cigs + \left[\frac{\frac{\hat{\beta}_2}{1,000}}{3,000}\right] faminc_{THB}$$

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

※ Unit of measurment doesn't change the implication of the estimation result

2 More on functional forms

- Logarithmic Functional Form

usually means natural log

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

- $\beta_1 = \dfrac{d \ln(y)}{d \ln(x)} = \dfrac{\frac{1}{y}dy}{\frac{1}{x_1}dx_1} = \dfrac{\frac{1}{y}\Delta y}{\frac{1}{x_1}\Delta x_1} = \dfrac{100\frac{1}{y}\Delta y}{100\frac{1}{x}\Delta x_1} = \dfrac{\%\Delta Y}{\%\Delta X}$ ($\Delta y = y_1 - y_2$, $\Delta x_1 = x_{11} - x_{12}$)

Note. $\dfrac{d \ln x}{dx} = \dfrac{1}{x}$

$d \ln x = \dfrac{1}{x} dx$

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

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

⇒ If we want the upper term to be % change then $100\beta_2 = \dfrac{100\frac{1}{y}\Delta y}{\Delta x_2}$

$100\beta_2 = \dfrac{\%\Delta Y}{\Delta x_2}$

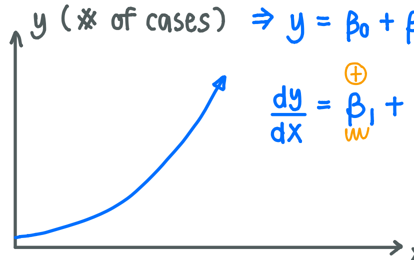
∴ $100\beta_2 = \%\Delta$ in y given that x_2 increases by 1 unit.

- Models with Quadratics

⇒ Capture increasing / decreasing marginal effects (slope of the relationship between x & y is not constant)

ex COVID - 19

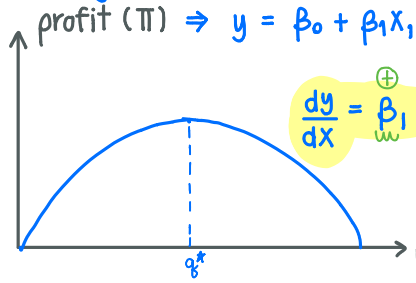
$y = \beta_0 + \beta_1 x + \beta_2 x^2 + u$



$\dfrac{dy}{dx} = \underset{\oplus}{\beta_1} + 2\underset{\oplus}{\beta_2} x$ (days)

Decreasing returns

$y = \beta_0 + \beta_1 x_1 + \beta_2 x^2 + u$



$\dfrac{dy}{dx} = \underset{\oplus}{\beta_1} + 2\underset{\ominus}{\beta_2} x$ (days)

$\Pi = (P - MC)q$; $MC = 10$

Demand : $P = 100 - q$

$\Pi = (100 - q - 10)q$

FOC. $\dfrac{\partial \Pi}{\partial q} = 0 = 90 - 2q$ β_1 is positive β_2 is negative

Example : Effects of Pullution on Housing Prices

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

where

price = housing price
nox = level of pollution
dist = distance from downtown
rooms = number of rooms
stratio = average student per teacher ratio → In the US or many countries, students can apply to schools in the area without having to take any test. So, the lower stratio, the better school

The estimation result is given by

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

Source	SS	df	MS
Model	51.4933152	5	10.298663
Residual	33.0889098	500	.06617782
Total	84.582225	505	.167489554

Number of obs =	506
F(5, 500) =	155.62
Prob > F =	0.0000
R-squared =	0.6088
Adj R-squared =	0.6049
Root MSE =	.25725

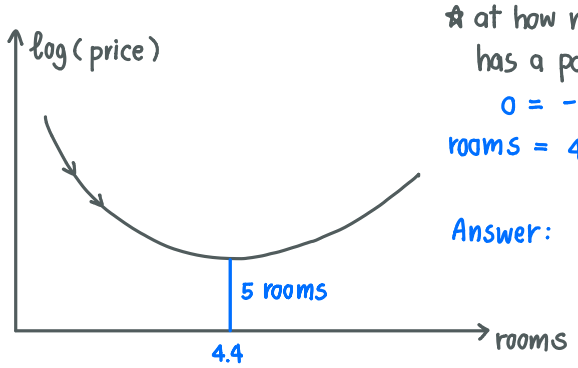
log(price) lprice	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
log(nox) lnox	β_1 -.9767545	.0995938	-9.81	0.000	-1.172429	-.7810806
dist	β_2 -.0321972	.0094013	-3.42	0.001	-.050668	-.0137264
rooms	β_3 -.5528032	.1612965	-3.43	0.001	-.8697056	-.2359007
rooms_sq	β_4 .0624697	.0124867	5.00	0.000	.0379368	.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
5

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

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

$$\frac{100\cdot\frac{1}{price}\,d\,price}{d\,rooms} = 100(-0.553 + 2(0.062)\cdot 5) = 100(0.67) = 6.7\ \%\ \text{increase}$$

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

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

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

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 regression"

- We use adjusted-R^2 or $\bar{R}^2$ when we want to punish adding too many regressors
  $R^2 = 1 - \frac{SSR}{SST} = 1 - \frac{SSR/\cancel{n}}{SST/\cancel{n}}$

Using adjusted R-squared to choose between non-nested models (one model is not a subset of another). adj $R^2 = 1 - \frac{SSR/(n-k-1)}{SST/(n-1)}$

Consider Model 1

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

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

Consider Model 2

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

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