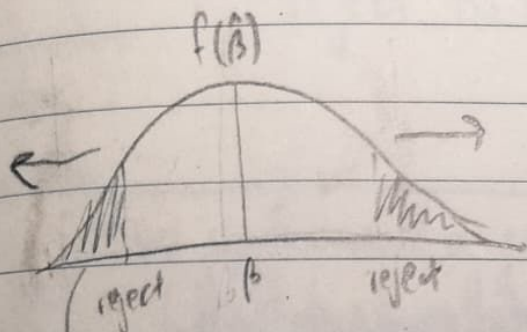


Inference $\rightarrow$ hypothesis testing about " β " the true parameter, but we don't know the true β , so we use $\hat{\beta}$ (estimator) and $s.e.(\hat{\beta})$ to test the hypotheses



1. test if β : some number

$$\frac{\hat{\beta}_j - \beta_j}{s.e.(\hat{\beta}_j)} \sim t_{d.f.}$$

p -value = total shaded area / significant level which we will reject the H_0 .
 $p\text{-value} < \text{significance level} \rightarrow \text{reject } H_0$

Confidence interval

$$CI: \hat{\beta}_j \pm C \times s.e.(\hat{\beta}_j)$$

percentile in the t -distribution with $n-k-1$ d.f.

$\rightarrow$ use one-side on the table (95% $\rightarrow$ 97.5%)

$$95\% \text{ C.I. for } \hat{\beta}_j = [\hat{\beta}_j - 2.06 \times s.e.(\hat{\beta}_j), \hat{\beta}_j + 2.06 \times s.e.(\hat{\beta}_j)]$$

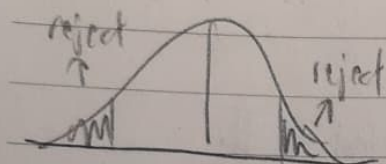


college returns

$$\log(\text{wage}) = \beta_0 + \beta_1 \text{c} + \beta_2 \text{univ} + \beta_3 \text{exper} + u$$

test whether $\beta_1 = \beta_2$

$$\begin{aligned} H_0: \beta_1 = \beta_2 &\rightarrow H_0: \beta_1 - \beta_2 = 0 \\ H_a: \beta_1 \neq \beta_2 &\rightarrow H_a: \beta_1 - \beta_2 \neq 0 \end{aligned} \quad \text{2 tailed test}$$

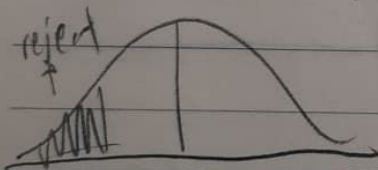


$$t = \frac{(\hat{\beta}_1 - \hat{\beta}_2) - 0}{\text{s.e.}(\hat{\beta}_1 - \hat{\beta}_2)} \quad \text{to compare with critical value}$$

$$\begin{aligned} \text{s.e.}(\hat{\beta}_1 - \hat{\beta}_2) &= \sqrt{\text{var}(\hat{\beta}_1 - \hat{\beta}_2)} \\ &= \sqrt{\text{var}(\hat{\beta}_1) + \text{var}(\hat{\beta}_2) - 2 \text{cov}(\hat{\beta}_1, \hat{\beta}_2)} \end{aligned}$$

OR

$$\begin{aligned} H_0: \beta_1 = \beta_2 &\rightarrow H_a: \beta_1 - \beta_2 = 0 && \text{assume } \beta_1 \text{ never more than } \beta_2 \text{ so} \\ H_a: \beta_1 < \beta_2 &\rightarrow H_a: \beta_1 - \beta_2 < 0 && \text{uni returns} > \text{college returns} \end{aligned}$$



only difference is rejection region

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

$$\text{test the } H_0: \beta_1 - 3\beta_2 = 1$$

1. t-statistic

$$t = \frac{(\hat{\beta}_1 - 3\hat{\beta}_2) - 1}{\text{s.e.}(\hat{\beta}_1 - 3\hat{\beta}_2)}$$

$$2. \text{define } \theta_1 = \hat{\beta}_1 - 3\hat{\beta}_2 \rightarrow H_0: \theta_1 = 1, H_a: \theta_1 \neq 1$$

$$t = \frac{\hat{\theta}_1 - 1}{\text{s.e.}(\hat{\theta}_1)}$$

$$\text{Now } \hat{\beta}_1 = \hat{\theta}_1 + 3\hat{\beta}_2 \text{ or } \beta_1 = \theta_1 + 3\beta_2 \text{ and sub into main regression}$$

$$y = \beta_0 + (\theta_1 + 3\beta_2)x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

$$y = \beta_0 + \theta_1 x_1 + \beta_2 (x_2 + 3x_1) + \beta_3 x_3 + u$$

so x_1 , $x_2 + 3x_1$ and x_3 are explanatory variables

p-value (probability that random t-value is greater (absolute than t in H_0)



reject H_0 if p-value < sig level

F-test (test significance of a group of hypotheses)
(multiple hypotheses)

ex. $H_0: \beta_1 = 0$ and $\beta_2 = 0 \rightarrow \beta_1 = \beta_2 = 0$ (both have to be 0)

$H_0: \beta_1 \neq 0$ and $\beta_2 \neq 0$ } at least one of the

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

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

use F-test to test multiple hypothesis / joint hypothesis (2 models)

1. full/unrestricted model (ur)
 2. restricted/small model (r)
- } test which one is better

ur

* $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \mu$ is true or reject H_0

r

* $y = \beta_0 + \beta_1 x_1 + \dots + \mu$ is true $\rightarrow$ do not reject H_0

suppose there are "q" number of β that we would like to perform a joint-test of $F = 0$

e.g. in this model $q=2$

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

$H_0: \beta_{k-q+1} = \beta_{k-q+2} = \dots = \beta_k = 0 \rightarrow$ the last q $\beta = 0$

$H_a: H_0$ not true

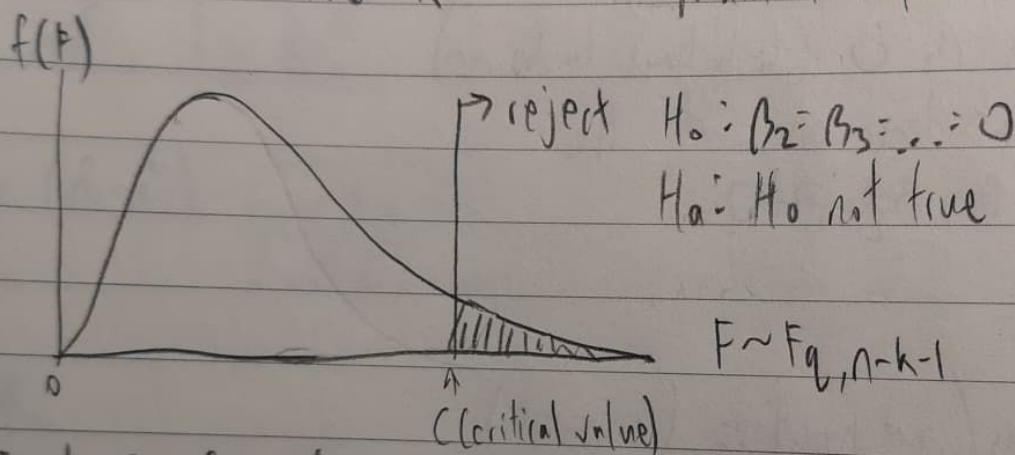


Testing for q # of $\beta = 0$

$\rightarrow$ always smaller than SSR , be better explained with more x variables

$$F = \frac{\frac{SSR_C - SSR_{UR}}{q}}{\frac{SSR_{UR}}{n-k-1}} \leftarrow \text{d.f. of UR model}$$

trade off of adding more x vs. variance of $\hat{\beta}_s$ will increase
 $SSR \downarrow R^2 \uparrow$ prediction of β less precise



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

ways to calculate F -statistic

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

$$F = \frac{R^2_{UR} - R^2_C}{\frac{1 - R^2_{UR}}{n-k-1}}$$

$$F = \frac{\frac{R^2}{K} \rightarrow R^2 \text{ of UR model}}{\frac{1 - R^2}{n-k-1}}$$

Data scaling on OLS statistics

change unit of measurement of a variable $\rightarrow$ change value of estimators
 - e.g. grams to kg

$$\widehat{\text{bweight}} = \hat{\beta}_0 + \hat{\beta}_1 \text{cigs} + \hat{\beta}_2 \text{faminc}$$

$$\begin{aligned} \widehat{\text{bweight}}_{\text{kg}} &= \frac{\widehat{\text{bweight}}}{1000} = \frac{\hat{\beta}_0}{1000} + \frac{\hat{\beta}_1}{1000} \text{cigs} + \frac{\hat{\beta}_2}{1000} \text{faminc} \\ &= \hat{\alpha}_0 + \hat{\alpha}_1 \text{cigs} + \hat{\alpha}_2 \text{faminc} \end{aligned}$$

logarithmic functional form

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

$$\Delta y = y_1 - y_2$$

$$\Delta x_1 = x_{11} - x_{12}$$

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

$\hookrightarrow$ elasticity if log form

$$\beta_2 = \frac{d \log(y)}{dx_2} = \frac{\frac{1}{y} dy}{dx_2} = \frac{\frac{1}{y} \Delta y}{\Delta x_2} \quad \text{multiply with 100 to find \% change}$$

$$100 \beta_2 = \frac{100 \frac{1}{y} \Delta y}{\Delta x_2} = \frac{\% \Delta y}{\Delta x_2} \quad \text{so } 100 \beta_2 = \% \Delta \text{ in } y \text{ given } x_2 \text{ increases by 1 unit}$$

models with quadratics

for increasing/decreasing marginal effects (slope of relationship of x and y is not constant)

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

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

(+) $\uparrow$ rooms

(+ or - depending on situation)

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

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

$\uparrow$ from 5 to 6

what is \% change of price when room $\uparrow$ from 5 to 7

- have to do 5-6 and 6-7 then add

rooms is original e.g. 5 to 6, choose 5



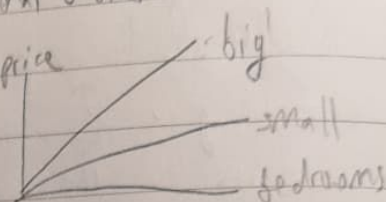
used when the impact of one variable depends on the value of another variable

Models with interaction terms

$$\text{price} = \beta_0 + \beta_1 \text{sqrft} + \beta_2 \text{bdrms} + \beta_3 \text{sqrft} \times \text{bdrms} + \beta_4 \text{bdrms} + u$$

$\xrightarrow{\quad X_3 \quad}$
 $X_1 \quad X_2 \quad X_1 \cdot X_2 \quad X_3$

$\frac{\partial \text{price}}{\partial \text{bdrms}} = \beta_2 + \beta_3 \text{sqrft}$, if $\beta_2 > 0$ then an additional bedroom would increase price more for a larger house



more on goodness-of-fit and selection of regressors

adding more regressors always improves fit $\rightarrow R^2 \uparrow$

but lose d.f. = free data point used to estimate parameter

using R^2 doesn't punish having too many regressors

- instead adjusted R^2 or $\bar{R}^2$ to punish

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

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

dummy variable

married: 1 if married
0 otherwise

$$\begin{aligned} 1. E(\text{wage} | \text{female}, \text{educ}) &= E(\beta_0 + \delta_0 \text{female} + \beta_1 \text{educ} + u | \text{female}, \text{educ}) \\ &= \beta_0 + \delta_0 \text{female} + \beta_1 \text{educ} + E(u | \text{female}, \text{educ}) \\ &= \beta_0 + \delta_0 \text{female} + \beta_1 \text{educ} \quad \rightarrow 0 \text{ (assumption MLR 1-4 holds)} \end{aligned}$$

2. Thus,

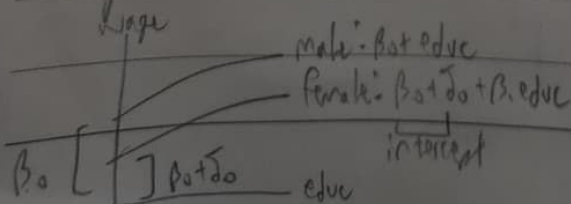
$$\text{♀} : E(\text{wage} | \text{female}=1, \text{educ}) = \beta_0 + \delta_0(1) + \beta_1 \text{educ} = \beta_0 + \delta_0 + \beta_1 \text{educ}$$

$$\text{♂} : E(\text{wage} | \text{female}=0, \text{educ}) = \beta_0 + \delta_0(0) + \beta_1 \text{educ} = \beta_0 + \beta_1 \text{educ}$$

$$\delta_0 = E(\text{wage} | \text{female}=1, \text{educ}) - E(\text{wage} | \text{female}=0, \text{educ})$$

$$\delta_0 = E(\text{wage} | \text{female}, \text{educ}) - E(\text{wage} | \text{male}, \text{educ})$$

given same value of educ, δ_0 is the difference in the expected wage of female v. male



No. _____

Date : _____

not possible to include all dummy alternatives in the same model (if there is intercept) because violates "perfect collinearity"