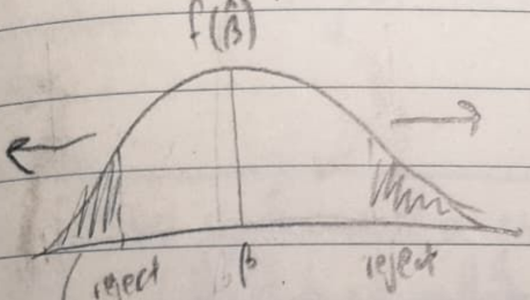
No.__________

Date :__________

Inference → 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-value < significance level → reject H_0

Confidence interval

$CI: \hat{\beta}_j \pm C \times se(\hat{\beta}_j)$

↑

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

↳ use one-side on the table (95% → 97.5%)

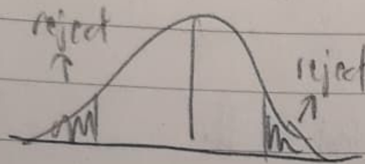
95% C.I. for $\hat{\beta}_j$: $[\hat{\beta}_j - 2.06 \times s.e.(\hat{\beta}_j), \hat{\beta}_j + 2.06 \times se(\hat{\beta}_j)]$

college returns

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

test whether $\beta_1 = \beta_2$

$\left.\begin{array}{l} H_0: \beta_1 = \beta_2 \rightarrow H_0: \beta_1 - \beta_2 = 0 \\ H_a: \beta \neq \beta_2 \rightarrow H_0: \beta_1 - \beta_2 \neq 0 \end{array}\right\}$ 2 tailed test



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

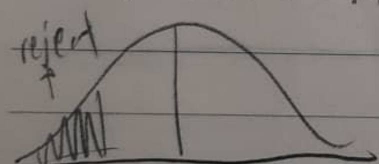
$s.e.(\hat{\beta}_1 - \hat{\beta}_2) = \sqrt{\widehat{var}(\hat{\beta}_1 - \hat{\beta}_2)}$

$= \sqrt{\widehat{var}(\hat{\beta}_1) + \widehat{var}(\hat{\beta}_2) - 2\,\widehat{cov}(\hat{\beta}_1, \hat{\beta}_2)}$

OR

$H_0: \beta_1 = \beta_2 \rightarrow H_a: \beta_1 - \beta_2 = 0$ assume β_1 never more than β_2 so

$H_a: \beta_1 < \beta_2 \rightarrow H_a: \beta_1 - \beta_2 < 0$ uni returns > college returns



only difference is rejection region

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

test the $H_0: \beta - 3\beta_2 = 1$

1. t-statistic

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

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

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

Now $\hat{\beta}_1 = \hat{\theta}_1 + 3\hat{\beta}_2$ or $\beta_1 = \theta_1 + 3\beta_2$ 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

No.

Date :

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)

$H_a: \beta_1 \neq 0$ 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 $\beta_1, \beta_2 \neq 0$

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

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

r

* $y = \beta_0 + \beta_1 x_1 + u$ 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 = 0

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

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

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

H_a: H_0 not true

testing for q# of $\beta = 0$

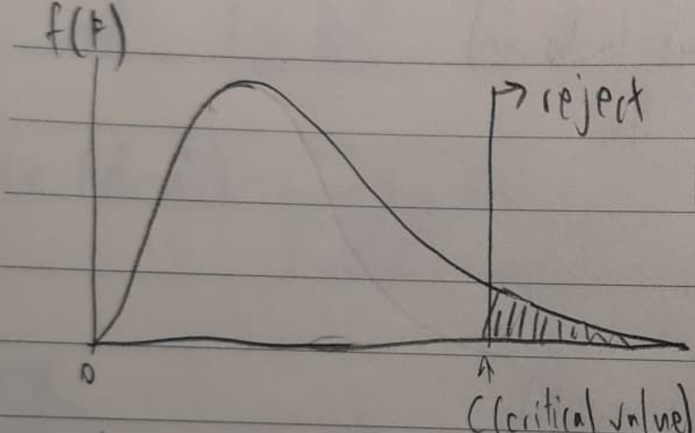
→ always smaller than SSR_r, bc better explained with more x variables

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

$n-k-1$ ← 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 $H_0: \beta_2 = \beta_3 = \dots = 0$

H_a: H_0 not true

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

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

ways to calculate F-statistic

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

$$F = \frac{\frac{R^2_{ur} - R^2_r}{q}}{\frac{1-R^2_{ur}}{n-k-1}} = \frac{\frac{R^2}{k}}{\frac{1-R^2}{n-k-1}}$$

R^2 → R^2 of UR model

No.
Date :

data scaling on OLS statistics

change unit of measurement of a variable → change value of estimators
- e.g. grams to kg

$$\widehat{bwright} = \hat{\beta}_0 + \hat{\beta}_1 \, cigs + \hat{\beta}_2 \, faminc$$

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

logarithmic functional form

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

$$\Delta y = y_1 - y_2$$
$$\Delta x_1 = x_{11} - x_{12}$$

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

↳ elasticity if log form

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

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

models with quadratics

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

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \mu$$
$$\frac{dy}{dx} = \beta_1 + 2\beta_2 x$$

$\hat{\beta}_1$ (+), $\hat{\beta}_2$ days

(+ or - depending on situation)

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

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

for 5 to 6

what is % change of price when room ↑ from 5 to 7
- have to do 5-6 and 6-7 then add

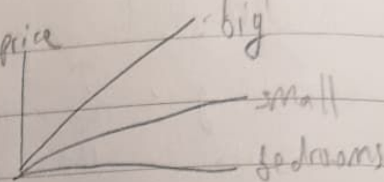
rooms is original e.g. 5 to 6, choose 5

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

<u>Models with interaction terms</u>

$$price = \beta_0 + \beta_1 \underset{x_1}{sqrft} + \beta_2 \underset{x_2}{bdrms} + \beta_3 \overbrace{\underset{x_1 \cdot x_2}{sqft \times bdrms}}^{x_3} + \beta_4 \underset{x_3}{bthms} + u$$

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



<u>more on goodness-of-fit and selection of regressors</u>

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

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

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

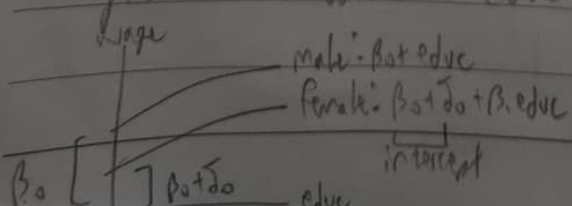
<u>dummy variable</u>

married $=$ 1 if married, 0 otherwise

1. $E(wage \mid female, educ) = E(\beta_0 + \delta_0 female + \beta_1 educ + u \mid female, educ)$
$= \beta_0 + \delta_0 female + \beta_1 educ + E(u \mid female, educ)$
$= \beta_0 + \delta_0 female + \beta_1 educ$ ↳ $= 0$ (assumption MLR 1-4 holds)

2. Thus,
♀: $E(wage \mid female = 1, educ) = \beta_0 + \delta_0(1) + \beta_1 educ = \beta_0 + \delta_0 + \beta_1 educ$
♂: $E(wage \mid female = 0, educ) = \beta_0 + \delta_0(0) + \beta_1 educ = \beta_0 + \beta_1 educ$
$\delta_0 = E(wage \mid female = 1, educ) - E(wage \mid female = 0, educ)$ or
$\delta_0 = E(wage \mid female, educ) - E(wage \mid male, educ)$

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 colinearity"