# Tutorial, Part 5

## Example 1

a) $H_0: \beta_3 = 0$

$H_1: \beta_3 \neq 0$

$$t_{\beta_3} = \frac{b_3 - \beta_{3,0}}{se(b_3)} = \frac{-0.25036 - 0}{\sqrt{0.007921}} = -2.81$$

$t_c\,(m = n-k = 32-3 = 29,\ \alpha/2 = 0.025) = \mp 2.045$

$|t_{\beta_3}| > t_c \Rightarrow$ We reject $H_0$ at $\alpha = 0.05$ and conclude from $H_1$ that the percentage of smokers affects the life expectancy at birth.

(p-value $= 0.009$ $< \alpha = 0.05$)

b) $H_0: 100\beta_2 \leq 0.167$

$H_1: 100\beta_2 > 0.167$

LIFE (in years) $\rightarrow$

2 months $= \frac{2}{12}$ year $= 0.167$ year

$$t = \frac{100b_2 - (100\beta_2)_0}{\sqrt{var(100b_2)}} = \frac{100b_2 - 0.167}{\sqrt{100^2 var(b_2)}} = \frac{100 \cdot 0.00233 - 0.167}{\sqrt{100^2 \cdot 0.0000001376}} = 1.78$$

$t_c\,(m = 29,\ \alpha = 0.05) = 1.699$ (upper tail)

$t > t_c \Rightarrow$ reject $H_0$ at $\alpha = 0.05$ and conclude from $H_1$

($p = 0.0428 < \alpha = 0.05$)

---

Matrix notation:

$H_0: c^T\beta \leq r$

$H_1: c^T\beta > r$

$$c = \begin{bmatrix} 0 \\ 100 \\ 0 \end{bmatrix} \qquad b = \begin{bmatrix} 79.624 \\ 0.00233 \\ -0.25036 \end{bmatrix}$$

$$\text{var-cov}(b) = \begin{bmatrix} 7.82115 & -0.000487 & -0.24151 \\ -0.000487 & 1.376 \cdot 10^{-7} & 9.960 \cdot 10^{-6} \\ -0.24151 & 9.960 \cdot 10^{-6} & 0.007921 \end{bmatrix}$$

$$r = [0.167]$$

$$t = \frac{c^T b - r}{\sqrt{var(c^T b)}} = \frac{c^T b - r}{\sqrt{c^T var\text{-}cov(b) \cdot c}} = \dots = 1.78$$

c) $H_0: 100\beta_2 + \beta_3 = 0$

$H_1: 100\beta_2 + \beta_3 \neq 0$

Like the square of a binomial $(a+b)^2$

$$t = \frac{100 b_2 + b_3 - (100\beta_2 + \beta_3)_0}{\sqrt{var(100 b_2 + b_3)}} = \frac{100 b_2 + b_3}{\sqrt{100^2 var(b_2) + 2 \cdot 100 \cdot cov(b_2, b_3) + var(b_3)}} =$$

$$= \frac{100 \cdot 0.00233 - 0.25036}{\sqrt{0.011289}} = -0.16$$

$t_c(n=29, \alpha/2 = 0.025) = 2.045$

$|t| < t_c \Rightarrow$ We cannot reject $H_0$ (at $\alpha = 0.05$).

Matrix notation:

$H_0: c^T\beta = r$

$H_1: c^T\beta \neq r$

$c = \begin{bmatrix} 0 \\ 100 \\ 1 \end{bmatrix}$

$r = [0]$

$b$ and $var\text{-}cov(b)$ as in part b)

$$t = \frac{c^T b - r}{\sqrt{var(c^T b)}} = \dots = -0.16$$

OR:

$H_0: 100\beta_2 + \beta_3 = 0 \Rightarrow \beta_3 = -100\beta_2$

$H_1: 100\beta_2 + \beta_3 \neq 0$

$LIFE = \beta_1 + \beta_2 EXP + \beta_3 TOBACCO + u_i$ BASE model

$LIFE = \beta_1 + \beta_2 EXP - 100\beta_2 TOBACCO + u_i$

$LIFE = \beta_1 + \beta_2 (EXP - 100\, TOBACCO) + u_i$ RESTRICTED model

$$F = \frac{(RSS_R - RSS_B)/(k_B - k_R)}{RSS_B / (n - k_B)} = \frac{(175.13 - 174.97)/(3-2)}{174.97/(32-3)} =$$

$$= 0.026$$

$F_C(n_1 = k_B - k_R = 1,\ n_2 = n - k_B = 29,\ \underline{\alpha = 0.05}) = 4.18$

$F < F_C \Rightarrow$ cannot reject $H_0$ (at $\alpha = 0.05$)

If $n_1 = 1 \Rightarrow F = t^2 = (-0.16)^2 = 0.026$

# Tutorial, Part 6

## $\boxed{\text{Example 1}}$

a) $\boxed{M_{AR}:}$ $\widehat{WAGE} = b_1 + b_2\, SOB + b_3\, ABR + b_4\, SBI$

$\hookrightarrow R^2_{AR} = 0.95$

$H_0: \beta_j = 0, \quad \forall j = 2,3,4$

$H_1: \beta_j \neq 0, \quad \exists j = 2,3,4$

$$\boxed{F} = \frac{R^2_{AR\,(WAGE)}/(k-2)}{(1-R^2_{AR})/(n-k+1)} = \frac{0.95/(5-2)}{(1-0.95)/(15-5+1)} = 69.7$$

$F_c\,(m_1=3,\ m_2=11,\ \alpha=0.05) = 3.59$

$F > F_c \implies$ reject $H_0$ at $\alpha = 0.05$ and conclude from $H_1$ $\Rightarrow$ variable WAGE causes multicollinearity problems

$$\boxed{VIF}_{WAGE} = \frac{1}{1-R^2_{WAGE}} = \frac{1}{1-0.95} = 20 > 10 \quad \text{(rule of thumb)}$$

$$\boxed{Tol}_{WAGE} = \frac{1}{VIF_{WAGE}} = 0.05 < 0.1$$

Interpretation: Due to $F > F_c$, $VIF > 10$ and $Tol < 0.1$ we can conclude that the variable WAGE causes a serious multicollinearity problem in the model.

**b)** Possible solutions:

- As WAGE is key determinant of consumption we cannot leave it out unless we replace it with a sensible alternative (e.g. total income).
- We can combine all other sources of income:

  $OINC = SOB + ABR + SBI$

  and estimate the model

  $\hat{PC} = b_1 + b_2 \, WAGE + b_3 \, OINC$ .
- We can try to estimate the initial model without variable(s) SOB, ABR and/or SBI.

## Example 2

Checking for possible multicollinearity problem caused by PRE.

MAR: $\hat{PRE}_t = b_1 + b_2 AGE_t + b_3 SD_t + b_4 T3_t$

$\sum PRE_t^2 = 528$ ; $\sum e_{PRE,t}^2 = 14.5$ ; $\overline{PRE} = 2$

$H_0: \beta_j = 0$ ; $\forall j = 2,3,4$
$H_1: \beta_j \neq 0$ ; $\exists j = 2,3,4$

$$R^2_{PRE} = 1 - \frac{RSS_{PRE}}{TSS_{PRE}} = *$$

$$TSS_{PRE} = \sum (PRE_t - \overline{PRE})^2 = \sum (PRE_t^2 - 2 \cdot PRE_t \cdot \overline{PRE} + \overline{PRE}^2) =$$
$$= \sum PRE_t^2 - 2 \cdot \underline{T \cdot \overline{PRE}} \cdot \overline{PRE} + T \cdot \overline{PRE}^2 =$$
$$= \sum PRE_t^2 - T \cdot \overline{PRE}^2 = 528 - 30 \cdot 2^2 = 408$$

$$* = 1 - \frac{14.5}{408} = 0.9645 = R^2_{PRE}$$

$$F = \frac{R^2_{PRE}/(k-2)}{(1-R^2_{PRE})/(n-k+1)} = \frac{0.9645/(5-2)}{(1-0.9645)/(30-5+1)} = 235.2$$

$F_C(m_1=3, m_2=26, \alpha=0.05) = 2.98$

$F > F_C \Rightarrow$ reject $H_0$ at $\alpha = 0.05$

$$VIF_{PRE} = \frac{1}{1-R^2_{PRE}} = \frac{1}{1-0.9645} = 28.14 > 10$$

they found out that there is a serious multicollinearity problem in the base model caused by variable PRE. To overcome the problem there are several possible solutions:

- leave out the problematic variable (if not a key one)
- leave out any other (similar) explanatory variable (e.g. SD or T3)
- transform the model (e.g. take logs).

## Example 3

$M_B$: $PRICE_i = \beta_1 + \beta_2 SIZE_i + \beta_3 AGE_i + \beta_4 ROOMS_i + \beta_5 BEDROOMS_i + u_i$

$R^2 = 0.886$ high

(most) partial regr. coeff. not significant

wrong sign with bedrooms

possible multicollinearity problem

## Correlations between explanatory variables

- $r_{SIZE,ROOMS} = 0.8457 \longrightarrow r^2_{SIZE,ROOMS} = 0.7152$ high
  $(p = 0.0000)$
- $r_{SIZE,BEDROOMS} = 0.7934 \longrightarrow r^2_{SIZE,BEDROOMS} = 0.6295$ high
  $(p = 0.0000)$
- $r_{ROOM,BEDROOMS} = 0.9244 \longrightarrow r^2_{ROOMS,BEDROOMS} = 0.8545 \approx R^2$ serious problem
  $(p = 0.0000)$

## F-tests based on auxiliary regressions for problematic variables

e.g. for ROOMS

$M_{Ae}$ : $\widehat{ROOMS} = b_1 + b_2\,SIZE + b_3\,AGE + b_4\,BEDROOMS$

$\hookrightarrow R^2_{ROOMS} = 0.8887$

$$F = \frac{R^2_{ROOMS}/(k-2)}{(1-R^2_{ROOMS})/(n-k+1)} = \frac{0.8887/(5-2)}{(1-0.8887)/(27-5+1)} = 62.216$$

$F_c(m_1 = 3, m_2 = 23, \alpha = 0.05) = 3.028$

$F > F_c \Longrightarrow$ reject $H_0: \beta_j = 0$ ; $\forall j = 2,3,4$ at $\alpha = 0.05 \Longrightarrow$ ROOMS cause serious multicollinearity

$$VIF_{ROOMS} = \frac{1}{1-0.8887} = 8.98 \approx 10$$

$$Tol_{ROOMS} = \frac{1}{VIF_{ROOMS}} = \frac{1}{8.98} = 0.113 \approx 0.1$$

— 11 —

Possible solutions:

- ROOMS and BEDROOMS are related as bedrooms also count as rooms → no need for BEDROOMS
- ROOMS and SIZE are related as rooms count up for size → no need for ROOMS

↳ exclude variables ROOMS and BEDROOMS from the model

# Testing for normal distribution of disturbances

$$PRICE_i = \beta_1 + \beta_2 SIZE_i + \beta_3 AGE_i + u_i$$

## Jarque-Bera test

$H_0$: Residuals (variable $u$) are normally distributed.
$H_1$: Residuals (variable $u$) are not normally distributed.

$$JB = n \cdot \left( \frac{S^2}{6} + \frac{(K-3)^2}{24} \right) = 27 \left( \frac{(-0.3192)^2}{6} + \frac{(2.3499-3)^2}{24} \right) =$$
$$= 0.934 \sim \chi^2_{(2)}$$

S - coeff. of skewness
K - coeff. of kurtosis

$$\chi^2_{C\,(m=2,\ \alpha=0.05)} = 5.997$$
$JB < \chi^2_C \Rightarrow$ cannot reject $H_0$ (at $\alpha = 0.05$)

(Stata: jb6 $\rightarrow p = 0.6269 > \alpha = 0.05$)