15.5 : Testing for significance

→ Model : $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots \beta_p x_p + \varepsilon$.

$$\left.\begin{aligned} H_0 &: \beta_1 = \beta_2 = \dots = \beta_p = 0 \\ H_1 &: \text{Not all } \beta_j \text{ are zero.} \end{aligned}\right\} \text{F-test.}$$

$n \neq p+1$
ما بزبط يكون عدد المتغيرات أكبر من sample size

مثلاً كان عندي
$n = 6$
$p = 5$
→ $n-p-1 = 0$
وهاد خلط

يعني n لازم تكون كبيرة بالنسبة لل p .

→ ANOVA table :

Source of Variation	df	SS	MS	F	F_α	p-Value
Regression	p	SSR	MSR	F	F_α	p-value
Error	$n-p-1$	SSE	MSE			
Total	$n-1$	SST				

- $MSR = \dfrac{SSR}{p}$ • $MSE = \dfrac{SSE}{n-p-1}$

- $F = \dfrac{MSR}{MSE}$ with $df_1 = p$ and $df_2 = n-p-1$

- Reject H_0 if $F \geq F_\alpha$ or p-value $\leq \alpha$.
  الي فوق

F-test only 1-times

يعني بس بنفحص اذا H_0 reject او لا .

هل Model جابدة من significance او لا .
لو بكاوه عليه By F-test

→ $H_0 : \beta_j = 0$; $H_1 : \beta_j \neq 0$ } t-test

لـ هل variable إلها تأثير significance ولا لا؟ لنجاوب عليه By t-test.

→ **Test statistic :**

$$t = \frac{b_j}{s_{b_j}}$$

→ **Rejection Rule :**

Reject H_0 if $|t| \geq t_{\frac{\alpha}{2}}$ or p-value $\leq \alpha$.

with $df = n-p-1$. By t-table.

→ p- times. ممكن نعمله p-times

→ **Remark:**

$$y = \beta_0 + \sum_{j=1}^{p} \beta_j x_j + \varepsilon .$$

- SSR with p → $MSR = \frac{SSR}{p}$
- SSE with $n-p-1$ → $s^2 = MSE = \frac{SSE}{n-p-1}$
- $F = \frac{MSR}{MSE}$, F_α with $df_1 = p$ and $df_2 = n-p-1$.
- $t^{(j)} = \frac{b_j}{s_{b_j}}$, b_j : estimater for β_j

- $S = \sqrt{S^2} = \sqrt{MSE}$: standard error of the estimate.

- $\delta_{b_j} = \sqrt{Var(b_j)}$: standard deviation of b_j.

- S_{b_j} = estimated standard deviation of b_j

Remark :

Excel : - multiple $R = \sqrt{R^2}$
- standard error $= S = \sqrt{MSE}$.
- significance F = p-value of F-test.

} on Regression statistic.

on Exp : p-value = 0.0003 < $\alpha = 0.01$

$\Rightarrow$ so Reject H_0 ($\beta_1 = \dots \beta_p = 0$) ($\alpha = 0.01$)

$\Rightarrow$ The Model is significance.

on Table of coeffecents.

- standard error : estimated standard deviation of $b_j = S_{b_j}$.
- $t^j = \dfrac{b_j}{S_{b_j}}$ with $df = n-p-1$.
- $(1-\alpha)$ CI $\beta_j = b_j \mp t_{\frac{\alpha}{2}} S_{b_j}$ with $df = n-p-1$.

exp on Excel : on table 3 of coeff. Sheet 3

H_0^1 : $\beta_1 = 0$ → p-value = 0.0005 < α = 0.01

→ Reject H_0^1 (α = 0.01)

→ $\beta_1 \neq 0$ (α = 0.01)

→ x_1 significance variable. (α = 0.01).

H_0^2 : $\beta_2 = 0$ → p-value = 0.004 < α = 0.01

→ Reject H_0^2 (α = 0.01).

→ $\beta_2 \neq 0$ (α = 0.01)

→ x_2 significance variable (α = 0.01)

→ Multicolinearity.

input variable $x_1, x_2, \dots x_p$

Some times some x_i is dependent on the other x_j's, this case is Known as multicolinearity. واحد من المتغيرات input يعتمد على الثانيين.

كيف بدي أعرف إذا هاي الظاهرة موجودة أو لا

→ Variance inflation factor (ViF).

$$ViF = \frac{1}{1 - R_i^2}$$

R_i^2 : Multiple coeffecient of determination for x_i as a function of the other x_j's.

• function means multiple regression.

وقتها لازم أشيل x_i من المودل $ViF(x_i) \geq 10 \Rightarrow x_i$ should be elemeneted.

هذا الحكي حتى لو significance

Model : $Y = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p + \varepsilon$.

* Goodness of fit of Model .
* significance of Model Variable .
* Validity of assumption .

هاي الشغلات إلي درسناها على هاد Model .

→ Multicolinearity $ViF(x_i) = \frac{1}{1 - R_i^2}$

→ R_i^2 : Goodness of fit $(x_i = \gamma_0 + \gamma_1 x_1 + \dots + \gamma_p x_p + \delta)$.

on exp excel (sheet 6) .

$ViF(X_2) = \frac{1}{1 - R_2^2} = \frac{1}{1 - 0.03} = 1.03$.

→ $\hat{x}_2 = 2.32 + 0.007\, X_1$

$ViF(X_1) = \frac{1}{1 - R_1^2} = \frac{1}{1 - 0.03} = 1.03$

من → $\hat{X}_1 = 69.49 + 3.62\, X_2$.

$ViF(X_2) < 10$: there is No colinearity between X_2 and the other variabls (X_1).

$ViF(X_1) < 10$: there is No colinearity between X_1 and the other variable (X_2) .