3.1) Le modèle sous forme matricielle:

Le modèle général suivant...

$$Y_t = \beta_0 + \beta_1 X_{t1} + \dots + \beta_p X_{tp} + \varepsilon_t , \quad t = 1, \dots, n.$$

...représente les n formules suivantes:

$$\begin{aligned} Y_1 &= \beta_0 + \beta_1 X_{1,1} + \dots + \beta_p X_{1p} + \varepsilon_1 \\ &\vdots \\ Y_n &= \beta_0 + \beta_1 X_{n1} + \dots + \beta_p X_{np} + \varepsilon_n \end{aligned}$$

...et peut s'écrire sous forme matricielle comme:

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}_{n\times 1} = \begin{bmatrix} \beta_0 + \beta_1 X_{11} + \dots + \beta_p X_{1p} \\ \vdots \\ \beta_0 + \beta_1 X_{n1} + \dots + \beta_p X_{np} \end{bmatrix}_{n\times 1} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}_{n\times 1}$$

ou encore comme

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}_{n\times 1} = \begin{bmatrix} 1 & X_{11} & \dots & X_{1p} \\ \vdots & \vdots & & \vdots \\ 1 & X_{n1} & & X_{np} \end{bmatrix}_{n\times(p+1)} \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_p \end{bmatrix}_{(p+1)\times 1} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}_{n\times 1}$$

ou...de manière plus compacte, comme:

$$Y = X\beta + \varepsilon$$

, avec

- Y = Vecteur $n \times 1$ des variables réponses
- X = Matrice shéma : $n \times (p+1)$: des variables explicatives
- β = vecteur $(p+1) \times 1$ des coefficients (à estimer)
- ε = Vecteur $n \times 1$ des erreurs. t.q.

$$\rightarrow \mathbb{E}(\varepsilon) = \underset{\sim}{0}_{n \times 1}$$

$$\rightarrow Var(\varepsilon) = \sigma^2 I_{n \times n}$$

matrice identité !

Remarque:

1) On suppose que $(X'X)^{-1}$ existe ... X est de rang complet et que $((X'X)^{-1})' = (X'X)^{-1}$

2) Par un modèle de régression linéaire simple; il ne suffit de définir la matrice shéma telle que:

$$X = \begin{bmatrix} 1 & X_1 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix}_{n \times 2}$$

3) Par un modèle passant l'origine, il n'y a pas de colonne de "1" dans la matrice shéma:

$$X = \begin{bmatrix} X_{11} & \cdots & X_{1p} \\ \vdots & & \vdots \\ X_{n1} & & X_{np} \end{bmatrix}_{n \times p}$$

4) Pour un modèle du type $Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_t^2 + \varepsilon_t$; il ne suffit que de définir la matrice shéma telle que:

$$X = \begin{bmatrix} 1 & X_1 & X_1^2 \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ 1 & X_n & X_n^2 \end{bmatrix} \quad \begin{array}{l} \bullet\ X_{t1} = X_t \\ \bullet\ X_{t2} = X_t^2 . \end{array}$$

3.1.1) Estimateur des moindres carrés (EMC):

On peut démontrer que l'estimateur $\hat{\beta}$ de β qui minimise la somme résiduelle des carrés ...

$$S(\beta) = \sum_{t=1}^{n} (Y_t - \hat{Y}_t)^2$$

$$= \sum_{t=1}^{n} \varepsilon_t^2$$

$$= \underset{\sim}{\varepsilon}' \varepsilon$$

$$= (Y - X\beta)'(Y - X\beta)$$

... est donné par:

$$\boxed{\hat{\beta} = (X'X)^{-1} X'Y}$$

<u>exemple:</u> ...en régression linéaire simple:

$$Y=\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} \; ; \; X=\begin{bmatrix} 1 & X_1 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \Rightarrow X'=\begin{bmatrix} 1 & \cdots & 1 \\ X_1 & \cdots & X_n \end{bmatrix}$$

$$\Rightarrow X'X=\begin{bmatrix} 1 & \cdots & 1 \\ X_1 & \cdots & X_n \end{bmatrix}\begin{bmatrix} 1 & X_1 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix}=\begin{bmatrix} n & \sum_{t=1}^{n} X_t \\ \sum_{t=1}^{n} X_t & \sum_{t=1}^{n} X_t^2 \end{bmatrix}$$

$$\Rightarrow (X'X)^{-1}=\frac{1}{n\sum_{t=1}^{n} X_t^2-(n\bar{X})^2}\begin{bmatrix} \sum_{t=1}^{n} X_t^2 & -n\bar{X} \\ -n\bar{X} & n \end{bmatrix}$$

$$\Rightarrow X'Y=\begin{bmatrix} 1 & \cdots & 1 \\ X_1 & \cdots & X_n \end{bmatrix}\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}=\begin{bmatrix} \sum_{t=1}^{n} Y_t \\ \sum_{t=1}^{n} X_t Y_t \end{bmatrix}=\begin{bmatrix} n\bar{Y} \\ \sum_{t=1}^{n} X_t Y_t \end{bmatrix}$$

<u>Ainsi</u>:

$$\hat{\beta}=\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix}=(X'X)^{-1}X'Y=\begin{bmatrix} \dfrac{n\bar{Y}\sum_{t=1}^{n} X_t^2-n\bar{X}\sum_{t=1}^{n} X_t Y_t}{n\sum_{t=1}^{n} X_t^2-(n\bar{X})^2} \\[2ex] \dfrac{n\sum_{t=1}^{n} X_t Y_t-(n\bar{X})(n\bar{Y})}{n\sum_{t=1}^{n} X_t^2-(n\bar{X})^2} \end{bmatrix}$$

$$= \begin{bmatrix} \dfrac{\bar{Y}\sum_{t=1}^{n} X_t^2 - \bar{X}\sum_{t=1}^{n} X_t Y_t + n\bar{X}^2\bar{Y} - n\bar{X}^2\bar{Y}}{\sum_{t=1}^{n} X_t^2 - n\bar{X}^2} \\[2em] \dfrac{\sum_{t=1}^{n} X_t Y_t - n\bar{X}\bar{Y}}{\sum_{t=1}^{n} X_t^2 - n\bar{X}^2} \end{bmatrix}$$

$$= \begin{bmatrix} \bar{Y} - \hat{\beta}_1 \bar{X} = \hat{\beta}_0 \\[1em] \hat{\beta}_1 \end{bmatrix}$$

les estimateurs de β_0 et β_1 obtenus précédemment !

*Propriétés des estimateurs:

I) Sans biais: $\mathbb{E}(\hat{\beta}) = \mathbb{E}\left((X'X)^{-1}X'Y\right)$

$$\overset{\text{thm.}}{=} (X'X)^{-1}X'\,\mathbb{E}(Y)$$
$$= (X'X)^{-1}X'(X\beta)$$
$$= (X'X)^{-1}(X'X)\beta$$
$$= I\beta$$
$$= \beta.$$

II) Variance-covariance: $Var(\hat{\beta}) = Var\left((X'X)^{-1}X'Y\right)$

$$\overset{\text{thm}}{=} (X'X)^{-1}X'\,Var(Y)\left[(X'X)^{-1}X'\right]'$$
$$= (X'X)^{-1}X'\,\sigma^2 I\left[X\left[(X'X)^{-1}\right]'\right]$$
$$\overset{\text{Remarque 1}}{=} \sigma^2(X'X)^{-1}X'\left[X(X'X)^{-1}\right]$$
$$= \sigma^2(X'X)^{-1}(X'X)(X'X)^{-1}$$
$$= \sigma^2 I(X'X)^{-1}$$
$$= \sigma^2(X'X)^{-1}$$

3.1.2) Résidus et tableau ANOVA:

* Les résidus...

On définit les résidus comme

$$\varepsilon_{n\times 1} = Y - \hat{Y}$$

$$= Y - X\hat{\beta}$$

$$= Y - X\left((X'X)^{-1}X'Y\right)$$

$$= Y - X(X'X)^{-1}X'Y$$

$$= \left(I - X(X'X)^{-1}X'\right)Y$$

$$= (I - H)\,Y$$

↳ Hat matrix = Matrice de projection...

Les sommes des carrés du tableau ANOVA sont données par:

$(n-1)$ d.l. • $SST = \sum_{t=1}^{n} (Y_t - \bar{Y})^2 = \sum_{t=1}^{n} Y_t^2 - n\bar{Y}^2 = Y'Y - n\bar{Y}^2$

$n-(p+1)$ d.l. • $SSE = \sum_{t=1}^{n} (Y_t - \hat{Y}_t)^2 = \sum_{t=1}^{n} \varepsilon_t^2 = \varepsilon'\varepsilon$

Ainsi:

p d.l. • $SSR = \sum_{t=1}^{n} (\hat{Y}_t - \bar{Y})^2 = \sum_{t=1}^{n} \hat{Y}_t^2 - n\bar{Y}^2 = \hat{Y}'\hat{Y} - n\bar{Y}^2$

$= (n-1)$
$-(n-(p+1)) = p$!

Dans le cas de la régression multiple, le tableau ANOVA est :

Source	SS	d.l	MS	F
Régression	SSR	p	SSR/p	MSR/MSE
Erreur	SSE	$n-(p+1)$	SSE $(\cancel{n}-(p+1))$	
Total	SST	$n-1$		

3.1.3) Estimateur de σ^2 :

Dans le cas multiple, on peut démontrer qu'un bon estimateur (sans biais) de σ^2 est S^2 tq. :

$$S^2 = MSE = \frac{SSE}{n-(p+1)}$$

3.2) Intervalles de confiance et tests d'hypothèses :

Essentiellement, on a la même chose qu'au chapitre 2 pour les tests t et F, sauf qu'il faut adapter les degrés de liberté

*Hypothèse de normalité : (Hyp. #4 !)

En supposant que $\varepsilon_t \overset{i.i.d}{\sim} N(0,\sigma^2)$ $\left(\Rightarrow \underset{\sim}{\varepsilon}_{n\times 1} \sim N_n(\underset{\sim}{0}; \sigma^2 I_{n\times n})\right)$

... on a que : • $(Y = X\beta + \varepsilon) \sim N_n(X\beta; \sigma^2 I_{n\times n})$

et que : • $\hat{\beta} \sim N_n(\beta; (X'X)^{-1}\sigma^2)$

3.2.1) Test de Student sur un seul paramètre :

On teste $H_0 : \beta_i = \beta_i^*$ ... (β_i^* = constante (ex: $\beta_i^* = 0$))

$H_1 : \beta_i \neq \beta_i^*$

en utilisant la statistique

$$t = \frac{\hat{\beta}_i - \beta_i^*}{\sqrt{[Var(\hat{\beta})]_{i+1\ i+1}}} \sim N(0,1)$$

ou encore (en remplacant σ^2 par S^2 dans la matrice $Var(\hat{\beta})$) :

$$t = \frac{\hat{\beta}_i - \beta_i^*}{\sqrt{[\widehat{Var}(\hat{\beta})]_{i+1\ i+1}}} \sim t(n-(p+1))$$

et on rejette H_0 au niveau de confiance $100 \times (1-\alpha)\%$ si

$$|t| > t_{\frac{\alpha}{2}}(n-(p+1))$$

Or, on a que : $Var(\hat{\beta}) = (X'X)^{-1}\sigma^2$

Ainsi : $\boxed{\widehat{Var}(\hat{\beta}) = (X'X)^{-1} S^2}$

En "inversant" ce test d'hypothèse, un intervalle de confiance pour β_i (I.C. marginal) est donc:

$$\boxed{\hat{\beta}_i \pm t_{\frac{\alpha}{2}}(n-(p+1)) \times \sqrt{\left[(X'X)^{-1}s^2\right]_{i+1\,i+1}}}$$

3.2.2) Test de Fisher pour la validité globale de la régression:

Dans le cas multiple; on teste

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$$

H_1: Au moins un coefficient parmi $\beta_1, \beta_2, \dots \beta_p$ est $\neq 0$.

avec la statistique

$$F = \frac{MSR}{MSE} \sim F(\underbrace{p}_{\text{d.l. de SSR}}, \underbrace{n-(p+1)}_{\text{d.l. de SSE}})$$

et on rejette H_0 au niveau de confiance $100 \times (1-\alpha)\%$ si:

$$F > F_\alpha(p, n-(p+1))$$

* Remarque IMPORTANTE:

De manière générale (... avec p variables explicatives); on a que:

$$F \neq t^2$$

... l'égalité ne survient que lorsque $p=1$!!!

↑↑↑
Rendu ici
05-10-2010