

Introduction à l'économétrie : Etude des Modèles Linéaires

CQLS : Jean-François Coeurjolly & Rémy Drouilhet
Jean-Francois.Coeurjolly@upmf-grenoble.fr, Remy.Drouilhet@upmf-grenoble.fr

1 Présentation du modèle

1.1 Description générale du modèle

Le modèle linéaire général, appelé aussi modèle de régression multiple, se présente sous la forme suivante

$$Y_t = \beta_0 + \beta_1 x_t^{(1)} + \beta_2 x_t^{(2)} + \dots + \beta_p x_t^{(p)} + U_t, t = 1, 2, \dots, n, \dots.$$

Dans cette étude, on disposera des observations jusqu'au temps n , s'exprimant par le système d'équations linéaires suivant :

$$\begin{cases} Y_1 = \beta_0 + \beta_1 x_1^{(1)} + \beta_2 x_1^{(2)} + \dots + \beta_p x_1^{(p)} + U_1 \\ \vdots \\ Y_t = \beta_0 + \beta_1 x_t^{(1)} + \beta_2 x_t^{(2)} + \dots + \beta_p x_t^{(p)} + U_t \\ \vdots \\ Y_n = \beta_0 + \beta_1 x_n^{(1)} + \beta_2 x_n^{(2)} + \dots + \beta_p x_n^{(p)} + U_n \end{cases}$$

Sous forme matricielle ce système se réécrit comme suit :

$$\underbrace{\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ \vdots \\ Y_n \end{pmatrix}}_{\mathbf{Y}} = \underbrace{\begin{pmatrix} 1 & x_1^{(1)} & x_1^{(2)} & \dots & x_1^{(p)} \\ 1 & x_2^{(1)} & x_2^{(2)} & \dots & x_2^{(p)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_n^{(1)} & x_n^{(2)} & \dots & x_n^{(p)} \end{pmatrix}}_{\mathbf{x}} \underbrace{\begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix}}_{\boldsymbol{\beta}} + \underbrace{\begin{pmatrix} U_1 \\ U_2 \\ \vdots \\ \vdots \\ U_n \end{pmatrix}}_{\mathbf{U}},$$

soit

$$\boxed{\mathbf{Y} = \mathbf{x}\boldsymbol{\beta} + \mathbf{U}}$$

Les vecteurs $\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(p)}$ seront appelés régresseurs et \mathbf{U} le bruit. En particulier le vecteur $\mathbf{x}^{(0)}$ constitué uniquement de 1, noté aussi $\mathbf{1}$, est appelé régresseur constant.

1.2 Modèles linéarisables

Modèle initial	Modèle "linéarisé"
Modèle log-linéaire $Y_t = \kappa \left(x_t^{(1)}\right)^{\beta_1} \left(x_t^{(2)}\right)^{\beta_2} \dots \left(x_t^{(p)}\right)^{\beta_p} \eta_t$	$\log(Y_t) = \beta_0 + \beta_1 \log(x_t^{(1)}) + \dots + \beta_p \log(x_t^{(p)}) + U_t$ avec $\beta_0 = \log(\kappa)$ et $U_t = \log(\eta_t)$
Modèle logarithmique (déjà linéarisé)	$Y_t = \beta_0 + \beta_1 \log(x_t^{(1)}) + \dots + \beta_p \log(x_t^{(p)}) + U_t$
Modèle exponentiel $Y_t = \kappa \exp(\beta_1 x_t^{(1)}) \exp(\beta_2 x_t^{(2)}) \dots \exp(\beta_p x_t^{(p)}) \eta_t$	$\log(Y_t) = \beta_0 + \beta_1 x_t^{(1)} + \dots + \beta_p x_t^{(p)} + U_t$ avec $\beta_0 = \log(\kappa)$ et $U_t = \log(\eta_t)$
Modèle polynômial $Y_t = \beta_0 + \beta_1 x_t + \beta_2 x_t^2 + \dots + \beta_p x_t^p + U_t$	$Y_t = \beta_0 + \beta_1 x_t^{(1)} + \dots + \beta_p x_t^{(p)} + U_t$ avec $x_t^{(1)} = x_t, x_t^{(2)} = x_t^2, \dots, x_t^{(p)} = x_t^p$

1.3 Hypothèses sur le modèle

Nous serons amenés à faire quelques hypothèses sur le modèle dites **hypothèses classiques**. Il convient dans une première lecture de ce document de prendre conscience que tout modèle nécessite des hypothèses particulières sans pour autant les comprendre et leur prêter beaucoup d'attention :

(C1) : \mathbf{Y} et \mathbf{x} sont observés sans erreur.

$$\begin{aligned}
\text{(C2)} : & \left\{ \begin{array}{l} \text{(C2-1)} \left\{ \begin{array}{l} \text{(Homoscédasticité) Pour tout } n > 0, \boldsymbol{\mu}_{\mathbf{U}} \stackrel{\text{Déf.}}{=} \mathbf{E} \left((U_1, \dots, U_n)^T \right) = \mathbf{0} \\ \text{et } \boldsymbol{\Sigma}_{\mathbf{U}} \stackrel{\text{Déf.}}{=} \mathbf{V} \left((U_1, \dots, U_n)^T \right) = \sigma^2 \mathbf{I}_n \text{ avec } \sigma^2 < +\infty. \end{array} \right. \\ \text{(C2-2)} \left\{ \begin{array}{l} \text{(Bruit gaussien)} \\ \text{Pour tout } n > 0, (U_1, \dots, U_n)^T \text{ est un vecteur gaussien.} \end{array} \right. \end{array} \right. \\
\text{(C3)} : & \left\{ \begin{array}{l} \text{(C3-1)} \left\{ \begin{array}{l} \text{(Linéaire indépendance des } \mathbf{x}^{(j)}, j = 1, \dots, p) \\ \text{La matrice } (\mathbf{x}^T \mathbf{x}) \text{ est inversible.} \end{array} \right. \\ \text{(C3-2)} \text{ Lorsque } n \text{ tend vers } +\infty, \frac{1}{n} \mathbf{x}^T \mathbf{x} \text{ tend vers une matrice inversible.} \end{array} \right.
\end{aligned}$$

L'hypothèse (C2-2) exprime entre autres choses que les U_1, \dots, U_n sont des variables aléatoires centrées, mutuellement indépendantes, toutes de loi Normale $\mathcal{N}(0, \sigma)$. Ainsi, le paramètre σ^2 quantifie à lui seul le niveau de bruit. Cette hypothèse (C2-2) constitue un cadre d'étude classique mais parfois restrictif ; il sera possible de s'en abstenir lorsque la taille d'échantillon est suffisamment grande.

2 Estimateurs MCO

L'estimateur des moindres carrés $\hat{\boldsymbol{\beta}}$ est la solution en $\boldsymbol{\beta}$ qui **minimise** le critère MCO (consistant en suivant l'exemple de cours à minimiser la somme des carrés des distances verticales) suivant :

$$\| \mathbf{Y} - \mathbf{x}\boldsymbol{\beta} \|^2 = \sum_{t=1}^n \left(Y_t - \left(\beta_0 + \beta_1 x_t^{(1)} + \beta_2 x_t^{(2)} + \dots + \beta_p x_t^{(p)} \right) \right)^2.$$

La solution à un tel problème peut être obtenue de manière géométrique.

$$\widehat{\boldsymbol{\beta}} \stackrel{\text{Not.}}{=} \widehat{\boldsymbol{\beta}}(\mathbf{Y}|\mathbf{x}) \stackrel{\text{Déf.}}{=} (\mathbf{x}^T \mathbf{x})^{-1} \mathbf{x}^T \mathbf{Y} \iff \|\mathbf{Y} - \mathbf{x}\widehat{\boldsymbol{\beta}}\|^2 = \min_{\boldsymbol{\beta}} \|\mathbf{Y} - \mathbf{x}\boldsymbol{\beta}\|^2$$

Cas particulier d'un seul régresseur ($p = 1$) : le vecteur des estimateurs prend la forme explicite suivante

$$\widehat{\beta}_1 \stackrel{\text{Not.}}{=} \widehat{\beta}_1(\mathbf{Y}|\mathbf{x}) = \frac{\text{cov}(\mathbf{Y}, \mathbf{x}^{(1)})}{\text{var}(\mathbf{x}^{(1)})} \text{ et } \widehat{\beta}_0 \stackrel{\text{Not.}}{=} \widehat{\beta}_0(\mathbf{Y}|\mathbf{x}) = \bar{Y} - \widehat{\beta}_1 \overline{\mathbf{x}^{(1)}},$$

$$\text{avec, } \text{cov}(\mathbf{Y}, \mathbf{x}^{(1)}) = \frac{1}{n} \sum_{t=1}^n (Y_t - \bar{Y}) (x_t^{(1)} - \overline{\mathbf{x}^{(1)}}) = \text{ et } \text{var}(\mathbf{x}^{(1)}) = \frac{1}{n} \sum_{t=1}^n (x_t^{(1)} - \overline{\mathbf{x}^{(1)}})^2.$$

Vecteur des valeurs ajustées et vecteur des résidus

$$\widehat{\mathbf{Y}} = \mathbf{x}\widehat{\boldsymbol{\beta}} \Leftrightarrow \widehat{Y}_t = \widehat{\beta}_0 + \widehat{\beta}_1 x_t^{(1)} + \widehat{\beta}_2 x_t^{(2)} + \dots + \widehat{\beta}_p x_t^{(p)}, \quad t = 1, 2, \dots, n,$$

$$\widehat{\mathbf{U}} = \mathbf{Y} - \widehat{\mathbf{Y}}$$

2.1 Estimateur sans biais

Cette propriété exprime que la moyenne des différentes estimations M.C.O. possibles du vecteur des paramètres de régression correspond au vecteur lui-même.

$$\mathbf{E}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \mathbf{0} \iff \mathbf{E}(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$$

2.2 Estimateur convergent en moyenne quadratique

Puisque l'estimateur $\widehat{\boldsymbol{\beta}}$ est sans biais, ceci se traduit plus simplement en disant que la variance des estimateurs tend vers 0 lorsque $n \rightarrow +\infty$. Cela provient d'une part du fait que la matrice de covariances $\underline{\mathbf{V}}(\widehat{\boldsymbol{\beta}})$ de $\widehat{\boldsymbol{\beta}}$ s'exprime par

$$\underline{\boldsymbol{\Sigma}}_{\widehat{\boldsymbol{\beta}}} \stackrel{\text{Not.}}{=} \underline{\mathbf{V}}(\widehat{\boldsymbol{\beta}}) \stackrel{\text{Déf.}}{=} \mathbf{E}\left(\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)^T\right) = \sigma^2 (\mathbf{x}^T \mathbf{x})^{-1}$$

et d'autre part de l'hypothèse **(C3-2)**. Le $i^{\text{ème}}$ élément diagonal de la matrice $\underline{\boldsymbol{\Sigma}}_{\widehat{\boldsymbol{\beta}}}$ correspond à la variance du $i^{\text{ème}}$ estimateur $\widehat{\beta}_i$ et sera noté $\sigma_{\widehat{\beta}_i}^2$:

$$\sigma_{\widehat{\beta}_i}^2 = \text{Var}(\widehat{\beta}_j(\mathbf{Y}|\mathbf{x})) = \sigma^2 (\mathbf{x}^T \mathbf{x})_{j,j}^{-1}$$

En revanche, il est beaucoup moins connu que :

$$\sigma_{\widehat{\beta}_i}^2 = \text{Var}(\widehat{\beta}_j(\mathbf{Y}|\mathbf{x})) = \frac{\sigma^2}{ns_j^2} \times VIF_j$$

où s_j^2 est la variance empirique du $j^{\text{ème}}$ régresseur et :

$$\boxed{VIF_j = \frac{1}{1 - R_j^2}}$$

est le VIF (Variance Inflation Factor) du $j^{\text{ème}}$ défini à partir de R_j^2 correspondant donc à la part de variance expliquée du $j^{\text{ème}}$ par les autres régresseurs du modèle.

3 Estimation des variances du bruit et des estimateurs M.C.O.

$$\boxed{\hat{\sigma}^2 \stackrel{\text{Not.}}{=} \hat{\sigma}^2(\mathbf{Y}|\mathbf{x}) \stackrel{\text{Déf.}}{=} \frac{1}{n-p-1} \sum_{t=1}^n \hat{U}_t^2.}$$

De cet estimateur on déduit un estimateur de la matrice $\underline{\Sigma}_{\hat{\beta}}$ de covariances des $\hat{\beta}$

$$\boxed{\underline{\hat{\Sigma}}_{\hat{\beta}} \stackrel{\text{Not.}}{=} \underline{\hat{\Sigma}}_{\hat{\beta}}(\mathbf{Y}|\mathbf{x}) \stackrel{\text{Déf.}}{=} \hat{\sigma}^2(\mathbf{Y}|\mathbf{x}) (\mathbf{x}^T \mathbf{x})^{-1}.}$$

Le $i^{\text{ème}}$ élément diagonal de la matrice $\underline{\hat{\Sigma}}_{\hat{\beta}}$ correspond à la variance estimée du $i^{\text{ème}}$ estimateur $\hat{\beta}_i$ et sera noté $\boxed{\hat{\sigma}_{\hat{\beta}_i}^2}$ (et voilà défini l'estimateur du niveau de précision des estimateurs mis en jeu).

4 Analyse de la variance et coefficient de détermination multiple

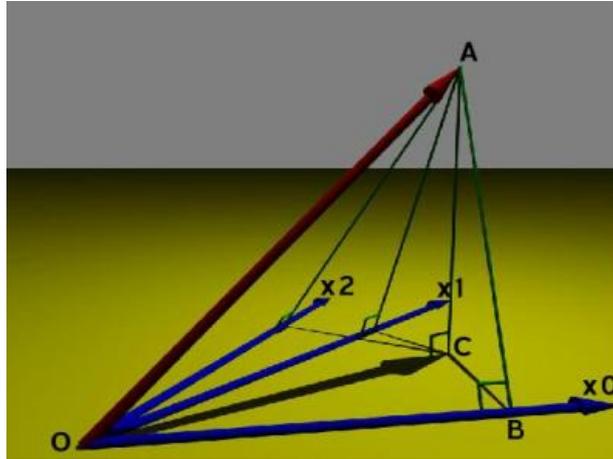


FIGURE 1 – Espace de représentation des variables : il y a correspondance entre les vecteurs \vec{OA} et \mathbf{Y} , \vec{OB} et $\bar{\mathbf{Y}}$, \vec{OC} et $\hat{\mathbf{Y}}$. De plus, comme nous nous sommes placés dans le cadre d’une régression multiple à $p = 2$ régresseurs, les vecteurs $O\vec{X}_0$, $O\vec{X}_1$ et $O\vec{X}_2$ correspondent respectivement aux régresseurs $\mathbf{x}^{(0)} = \mathbf{1}$ (constant), $\mathbf{x}^{(1)}$ et $\mathbf{x}^{(2)}$. Notons que tout vecteur $O\vec{Z}$ (non représenté dans la figure) vivant dans le même plan que les vecteurs $O\vec{X}_0$, $O\vec{X}_1$ et $O\vec{X}_2$ s’écrit comme une combinaison linéaire de ces vecteurs (correspondant alors à $\gamma_0\mathbf{x}^{(0)} + \gamma_1\mathbf{x}^{(1)} + \gamma_2\mathbf{x}^{(2)}$). Le vecteur \vec{OB} étant le vecteur le plus proche dans ce plan du vecteur \vec{OA} , cette figure nous permet donc de comprendre la méthode géométrique conduisant à l’obtention des estimations des paramètres de régression qui ne sont que les coefficients de la combinaison linéaire associée à ce vecteur. En fait, pour faire cette figure avec un logiciel 3D, il suffit de mettre une lumière à l’extrémité du vecteur \vec{OA} dont son ombre sur le plan représente le vecteur \vec{OB} .

Il découle d’une propriété géométrique de l’estimateur MCO l’équation suivante (appelée *équation de l’analyse de la variance*) :

$$\|\mathbf{Y} - \bar{\mathbf{Y}}\|^2 = \underbrace{\|\mathbf{Y} - \hat{\mathbf{Y}}\|^2}_{\|\hat{\mathbf{U}}\|^2} + \underbrace{\|\hat{\mathbf{Y}} - \bar{\mathbf{Y}}\|^2}_{\|\hat{\mathbf{Y}} - \bar{\mathbf{Y}}\|^2}$$

Ceci s’exprime littéralement en divisant tous les termes par n : “variance totale est égale à la somme des variance résiduelle et variance expliquée”. Les données seront d’autant mieux ajustées que la norme $\|\hat{\mathbf{U}}\|^2$ mesurant la dispersion des erreurs est faible. Ceci permet alors d’introduire un indicateur de la qualité d’ajustement appelé *coefficient de détermination multiple* :

$$R^2 = \cos^2(\widehat{ABC}) = \frac{\|\hat{\mathbf{Y}} - \bar{\mathbf{Y}}\|^2}{\|\mathbf{Y} - \bar{\mathbf{Y}}\|^2} = 1 - \frac{\|\hat{\mathbf{U}}\|^2}{\|\mathbf{Y} - \bar{\mathbf{Y}}\|^2} = \text{corr}^2(\mathbf{Y}, \hat{\mathbf{Y}})$$

5 Comparaison des cadres gaussien et asymptotique

Nous pouvons rassembler l'ensemble des résultats des deux sections précédentes dans le tableau suivant :

	Cadre gaussien	Cadre Asymptotique
Paramètre β_i		
Mesure d'écart standardisée	$\Delta_{\widehat{\beta}_i, \beta_i}(\mathbf{Y} \mathbf{x}) \stackrel{\text{Déf.}}{=} \frac{\widehat{\beta}_i(\mathbf{Y} \mathbf{x}) - \beta_i}{\widehat{\sigma}_{\widehat{\beta}_i}(\mathbf{Y} \mathbf{x})}$	
Comportement aléatoire	$\Delta_{\widehat{\beta}_i, \beta_i}(\mathbf{Y} \mathbf{x}) \rightsquigarrow \mathcal{St}(n-p-1)$	$\Delta_{\widehat{\beta}_i, \beta_i}(\mathbf{Y} \mathbf{x}) \stackrel{\text{approx.}}{\rightsquigarrow} \mathcal{N}(0, 1)$
Paramètre β_Q		
Mesure d'écart standardisée	$\Delta_{\widehat{\beta}_Q, \beta_Q}(\mathbf{Y} \mathbf{x}) \stackrel{\text{Déf.}}{=} \frac{1}{q} \left(\widehat{\beta}_Q(\mathbf{Y} \mathbf{x}) - \beta_Q \right)^T \widehat{\Sigma}_{\widehat{\beta}_Q}^{-1}(\mathbf{Y} \mathbf{x}) \left(\widehat{\beta}_Q(\mathbf{Y} \mathbf{x}) - \beta_Q \right)$	
Comportement aléatoire	$\Delta_{\widehat{\beta}_Q, \beta_Q}(\mathbf{Y} \mathbf{x}) \rightsquigarrow \mathcal{F}(q, n-p-1)$	$\Delta_{\widehat{\beta}_Q, \beta_Q}(\mathbf{Y} \mathbf{x}) \stackrel{\text{approx.}}{\rightsquigarrow} q \times \chi^2(q)$
Paramètre σ^2		
Mesure d'écart standardisée	$\Delta_{\widehat{\sigma}^2, \sigma^2}(\mathbf{Y} \mathbf{x}) \stackrel{\text{Déf.}}{=} (n-p-1) \frac{\widehat{\sigma}^2(\mathbf{Y} \mathbf{x})}{\sigma^2}$	
Comportement aléatoire	$\Delta_{\widehat{\sigma}^2, \sigma^2}(\mathbf{Y} \mathbf{x}) \rightsquigarrow \chi^2(n-p-1)$	$\Delta_{\widehat{\sigma}^2, \sigma^2}(\mathbf{Y} \mathbf{x}) \stackrel{\text{approx.}}{\rightsquigarrow} \mathcal{N}(0, 1)$

Remarque : lorsque le paramètre d'intérêt est $\theta = \sigma^2$, le précédent résultat a besoin d'une petite hypothèse supplémentaire que nous noterons **(C2-2*)** :

(C2-2*) Pour tout $t = 1, \dots, n$, $\mathbf{E}(U_t^4) < +\infty$, i.e. que les U_t possèdent un moment d'ordre 4 fini.

Cette hypothèse peut sembler plus technique mais croyez les auteurs elle n'est que très peu restrictive et vous n'auriez vraiment pas de chance si celle-ci n'était pas vérifiée.

6 Test d'hypothèses d'un paramètre de régression

Test d'hypothèses sur β_i	
Hypothèses de test	$\mathbf{H}_0 : \beta_i = \beta_i^0$ contre $\mathbf{H}_1 : \begin{cases} \beta_i > \beta_i^0 & \text{(cas (a))} \\ \beta_i < \beta_i^0 & \text{(cas (b))} \\ \beta_i \neq \beta_i^0 & \text{(cas (c))} \end{cases}$
Statistique de test sous \mathbf{H}_0	$\widehat{\delta}_{\beta_i, \beta_i^0}(\mathbf{Y} \mathbf{x}) \stackrel{\text{Déf.}}{=} \frac{\widehat{\beta}_i(\mathbf{Y} \mathbf{x}) - \beta_i^0}{\widehat{\sigma}_{\widehat{\beta}_i}(\mathbf{Y} \mathbf{x})} \begin{cases} \rightsquigarrow St(n-p-1) & \text{sous l'hypothèse (C2-2)} \\ \overset{\text{approx.}}{\rightsquigarrow} St(n-p-1) & \text{si } n \text{ est grand} \end{cases}$
Règle de décision	<div style="display: flex; align-items: center; justify-content: center;"> <div style="margin-right: 10px;">Rejet de \mathbf{H}_0 si</div> <div style="border: 1px solid black; padding: 5px;"> $\begin{cases} \widehat{\delta}_{\beta_i, \beta_i^0}(\mathbf{y} \mathbf{x}) > \delta_{lim, \alpha}^+ & \text{(cas (a))} \\ \widehat{\delta}_{\beta_i, \beta_i^0}(\mathbf{y} \mathbf{x}) < \delta_{lim, \alpha}^- & \text{(cas (b))} \\ \widehat{\delta}_{\beta_i, \beta_i^0}(\mathbf{y} \mathbf{x}) > \delta_{lim, \alpha/2}^+ & \text{(cas (c))} \end{cases}$ </div> <div style="margin-left: 10px;">ou si</div> </div> <div style="border: 1px solid black; padding: 10px; margin-top: 10px; width: fit-content; margin: 0 auto;"> $p\text{-valeur} = \begin{cases} \mathbb{P}_{\beta_i = \beta_i^0} \left(\widehat{\delta}_{\beta_i, \beta_i^0}(\mathbf{Y} \mathbf{x}) > \widehat{\delta}_{\beta_i, \beta_i^0}(\mathbf{y} \mathbf{x}) \right) & \text{(cas (a))} \\ \mathbb{P}_{\beta_i = \beta_i^0} \left(\widehat{\delta}_{\beta_i, \beta_i^0}(\mathbf{Y} \mathbf{x}) < \widehat{\delta}_{\beta_i, \beta_i^0}(\mathbf{y} \mathbf{x}) \right) & \text{(cas (b))} \\ \mathbb{P}_{\beta_i = \beta_i^0} \left(\widehat{\delta}_{\beta_i, \beta_i^0}(\mathbf{Y} \mathbf{x}) > \widehat{\delta}_{\beta_i, \beta_i^0}(\mathbf{y} \mathbf{x}) \right) & \text{(cas (c))} \end{cases} < \alpha$ </div>

où $\delta_{lim, \alpha}^-$ et $\delta_{lim, \alpha}^+$ sont les quantiles d'ordre α et $1 - \alpha$ de la loi de la statistique de test sous \mathbf{H}_0 . Rappelons qu'en théorie (cf section 5), si on ne suppose pas l'hypothèse (C2-2) et si n est grand la mesure d'écart standardisée de test suit approximativement une $\mathcal{N}(0, 1)$ équivalent à une $St(n-p-1)$ (lorsque n est grand).

Soulignons que dans le cas (c) et $\beta_i^0 = 0$, c'est un **test de significativité locale** du $i^{\text{ème}}$ régresseur.

Test d'hypothèses sur le paramètre de nuisance	
Hypothèses de test	$\mathbf{H}_0 : \sigma^2 = \sigma_0^2$ contre $\mathbf{H}_1 : \begin{cases} \sigma^2 > \sigma_0^2 & \text{(cas (a))} \\ \sigma^2 < \sigma_0^2 & \text{(cas (b))} \\ \sigma^2 \neq \sigma_0^2 & \text{(cas (c))} \end{cases}$
Statistique de test sous \mathbf{H}_0	$\hat{\delta}_{\sigma^2, \sigma_0^2}(\mathbf{Y} \mathbf{x}) \stackrel{\text{Déf.}}{=} \begin{cases} (n-p-1) \frac{\hat{\sigma}^2(\mathbf{Y} \mathbf{x})}{\sigma_0^2} \rightsquigarrow \chi^2(n-p-1) & \text{sous (C2-2)} \\ \frac{\hat{\sigma}^2(\mathbf{Y} \mathbf{x}) - \sigma_0^2}{\sqrt{\frac{\hat{\sigma}^2(\mathbf{Y} \mathbf{x})}{n}}} \underset{\text{approx.}}{\rightsquigarrow} \mathcal{N}(0,1) & \text{si } n \text{ est grand et sous (C2-2*)} \end{cases}$
Règle de décision	Rejet de \mathbf{H}_0 si $\begin{cases} \hat{\delta}_{\sigma^2, \sigma_0^2}(\mathbf{y} \mathbf{x}) > \delta_{lim, \alpha}^+ & \text{(cas (a))} \\ \hat{\delta}_{\sigma^2, \sigma_0^2}(\mathbf{y} \mathbf{x}) < \delta_{lim, \alpha}^- & \text{(cas (b))} \\ \hat{\delta}_{\sigma^2, \sigma_0^2}(\mathbf{y} \mathbf{x}) < \delta_{lim, \alpha/2}^- \text{ ou } \hat{\delta}_{\sigma^2, \sigma_0^2}(\mathbf{y} \mathbf{x}) > \delta_{lim, \alpha/2}^+ & \text{(cas (c))} \end{cases}$ ou si $p\text{-valeur} = \begin{cases} \mathbb{P}_{\sigma^2 = \sigma_0^2} \left(\hat{\delta}_{\sigma^2, \sigma_0^2}(\mathbf{Y} \mathbf{x}) > \hat{\delta}_{\sigma^2, \sigma_0^2}(\mathbf{y} \mathbf{x}) \right) & \text{(cas (a))} \\ \mathbb{P}_{\sigma^2 = \sigma_0^2} \left(\hat{\delta}_{\sigma^2, \sigma_0^2}(\mathbf{Y} \mathbf{x}) < \hat{\delta}_{\sigma^2, \sigma_0^2}(\mathbf{y} \mathbf{x}) \right) & \text{(cas (b))} \\ 2 \times \min \left(\mathbb{P}_{\sigma^2 = \sigma_0^2} \left(\hat{\delta}_{\sigma^2, \sigma_0^2}(\mathbf{Y} \mathbf{x}) < \hat{\delta}_{\sigma^2, \sigma_0^2}(\mathbf{y} \mathbf{x}) \right), \right. \\ \left. \mathbb{P}_{\sigma^2 = \sigma_0^2} \left(\hat{\delta}_{\sigma^2, \sigma_0^2}(\mathbf{Y} \mathbf{x}) > \hat{\delta}_{\sigma^2, \sigma_0^2}(\mathbf{y} \mathbf{x}) \right) \right) & \text{(cas (c))} \end{cases} < \alpha$

où $\delta_{lim, \alpha}^-$ et $\delta_{lim, \alpha}^+$ sont les quantiles d'ordre α et $1 - \alpha$ de la loi de la statistique de test sous \mathbf{H}_0 .

7 Intervalles de confiance

- Paramètre de régression :

$$1 - \alpha = \mathbb{P} \left(\tilde{\beta}_{i, \text{inf}}(\mathbf{Y}|\mathbf{x}) < \beta_i < \tilde{\beta}_{i, \text{sup}}(\mathbf{Y}|\mathbf{x}) \right)$$

avec

$$\begin{aligned} \tilde{\beta}_{i, \text{inf}}(\mathbf{Y}|\mathbf{x}) &= \hat{\beta}_i(\mathbf{Y}|\mathbf{x}) - \delta_{lim, \frac{\alpha}{2}}^+ \times \hat{\sigma}_{\hat{\beta}_i}(\mathbf{Y}|\mathbf{x}) \\ &\text{et} \\ \tilde{\beta}_{i, \text{sup}}(\mathbf{Y}|\mathbf{x}) &= \hat{\beta}_i(\mathbf{Y}|\mathbf{x}) + \delta_{lim, \frac{\alpha}{2}}^+ \times \hat{\sigma}_{\hat{\beta}_i}(\mathbf{Y}|\mathbf{x}) \end{aligned}$$

où $\delta_{lim, \frac{\alpha}{2}}^+$ est le quantile d'ordre $1 - \frac{\alpha}{2}$ d'une loi de $\mathcal{St}(n-p-1)$ sous l'hypothèse (C2-2) et une loi $\mathcal{N}(0,1)$ si n est grand.

- Paramètre de nuisance :

$$1 - \alpha = \mathbb{P} \left(\tilde{\sigma}_{\text{inf}}^2(\mathbf{Y}|\mathbf{x}) < \sigma^2 < \tilde{\sigma}_{\text{sup}}^2(\mathbf{Y}|\mathbf{x}) \right)$$

les bornes étant définies

- sous l'hypothèse **(C2-2)** par :

$$\begin{aligned}\tilde{\sigma}_{\text{inf}}^2(\mathbf{Y}|\underline{\mathbf{x}}) &= \hat{\sigma}^2(\mathbf{Y}|\underline{\mathbf{x}}) \times \frac{n-p-1}{\delta_{\text{lim}, \frac{\alpha}{2}}^-} \\ &\text{et} \\ \tilde{\sigma}_{\text{sup}}^2(\mathbf{Y}|\underline{\mathbf{x}}) &= \hat{\sigma}^2(\mathbf{Y}|\underline{\mathbf{x}}) \times \frac{n-p-1}{\delta_{\text{lim}, \alpha/2}^+}\end{aligned}$$

où $\delta_{\text{lim}, \frac{\alpha}{2}}^-$ et $\delta_{\text{lim}, \frac{\alpha}{2}}^+$ sont les quantiles d'ordre $\frac{\alpha}{2}$ et $1 - \frac{\alpha}{2}$ d'une loi du $\chi^2(n-p-1)$.

- si n est grand et sous l'hypothèse **(C2-2*)** par :

$$\begin{aligned}\tilde{\sigma}_{\text{inf}}^2(\mathbf{Y}|\underline{\mathbf{x}}) &= \hat{\sigma}^2(\mathbf{Y}|\underline{\mathbf{x}}) - \delta_{\text{lim}, \frac{\alpha}{2}}^+ \times \sqrt{\frac{\widehat{\sigma_{\mathbf{U}^2}^2}(\mathbf{Y}|\underline{\mathbf{x}})}{n}} \\ &\text{et} \\ \tilde{\sigma}_{\text{sup}}^2(\mathbf{Y}|\underline{\mathbf{x}}) &= \hat{\sigma}^2(\mathbf{Y}|\underline{\mathbf{x}}) + \delta_{\text{lim}, \frac{\alpha}{2}}^+ \times \sqrt{\frac{\widehat{\sigma_{\mathbf{U}^2}^2}(\mathbf{Y}|\underline{\mathbf{x}})}{n}}\end{aligned}$$

où $\delta_{\text{lim}, \frac{\alpha}{2}}^+$ est le quantile d'ordre $1 - \frac{\alpha}{2}$ d'une loi du $\mathcal{N}(0, 1)$.

8 Prévision

Envisageons maintenant la prévision de la valeur de Y_τ pour un instant τ (pour lequel y_τ est) non observé, en supposant connu le vecteur \mathbf{x}_τ des valeurs des p régresseurs à cet instant τ . La prévision se fait naturellement à partir du modèle ajusté :

$$\hat{Y}_\tau \stackrel{\text{Not.}}{=} \hat{Y}_\tau(\mathbf{Y}|\underline{\mathbf{x}}; \mathbf{x}_\tau) \stackrel{\text{Déf.}}{=} \mathbf{x}_\tau^T \hat{\boldsymbol{\beta}}(\mathbf{Y}|\underline{\mathbf{x}})$$

On peut alors considérer l'erreur de prédiction

$$\hat{U}_\tau \stackrel{\text{Déf.}}{=} Y_\tau - \hat{Y}_\tau(\mathbf{Y}|\underline{\mathbf{x}}; \mathbf{x}_\tau)$$

d'espérance nulle et de variance

$$\sigma_{\hat{U}_\tau}^2 \stackrel{\text{Not.}}{=} V(\hat{U}_\tau) = \sigma^2 \left(1 + \mathbf{x}_\tau^T (\underline{\mathbf{x}}^T \underline{\mathbf{x}})^{-1} \mathbf{x}_\tau \right)$$

naturellement estimée par

$$\hat{\sigma}_{\hat{U}_\tau}^2(\mathbf{Y}|\underline{\mathbf{x}}; \mathbf{x}_\tau) \stackrel{\text{Déf.}}{=} \hat{\sigma}^2(\mathbf{Y}|\underline{\mathbf{x}}) \left(1 + \mathbf{x}_\tau^T (\underline{\mathbf{x}}^T \underline{\mathbf{x}})^{-1} \mathbf{x}_\tau \right)$$

On peut également montrer que :

$$\Delta_{Y_\tau, \hat{Y}_\tau} \stackrel{\text{Déf.}}{=} \Delta_{Y_\tau, \hat{Y}_\tau}(\mathbf{Y}|\underline{\mathbf{x}}; \mathbf{x}_\tau) = \frac{Y_\tau - \hat{Y}_\tau(\mathbf{Y}|\underline{\mathbf{x}}; \mathbf{x}_\tau)}{\hat{\sigma}_{\hat{U}_\tau}(\mathbf{Y}|\underline{\mathbf{x}}; \mathbf{x}_\tau)} \begin{cases} \rightsquigarrow St(n-p-1) & \text{sous l'hypothèse (C2-2)} \\ \underset{\rightsquigarrow}{\text{approx.}} St(n-p-1) & \text{si } n \text{ est grand} \end{cases}$$

Ce résultat permet de construire un intervalle de prévision $\left[\tilde{Y}_{\tau, \text{inf}}(\mathbf{Y}|\underline{\mathbf{x}}; \mathbf{x}_\tau), \tilde{Y}_{\tau, \text{sup}}(\mathbf{Y}|\underline{\mathbf{x}}; \mathbf{x}_\tau) \right]$ de Y_τ au niveau $1 - \alpha$. En effet,

$$1 - \alpha = \mathbb{P} \left(\tilde{Y}_{\tau, \text{inf}}(\mathbf{Y}|\underline{\mathbf{x}}; \mathbf{x}_\tau) < Y_\tau < \tilde{Y}_{\tau, \text{sup}}(\mathbf{Y}|\underline{\mathbf{x}}; \mathbf{x}_\tau) \right)$$

avec

$\tilde{Y}_{\tau, \text{inf}}(\mathbf{Y} \underline{\mathbf{x}}; \mathbf{x}_\tau) = \hat{Y}_\tau(\mathbf{Y} \underline{\mathbf{x}}; \mathbf{x}_\tau) - \delta_{lim, \frac{\alpha}{2}}^+ \times \hat{\sigma}_{\hat{U}_\tau}(\mathbf{Y} \underline{\mathbf{x}}; \mathbf{x}_\tau)$ <p style="text-align: center;">et</p> $\tilde{Y}_{\tau, \text{sup}}(\mathbf{Y} \underline{\mathbf{x}}; \mathbf{x}_\tau) = \hat{Y}_\tau(\mathbf{Y} \underline{\mathbf{x}}; \mathbf{x}_\tau) + \delta_{lim, \frac{\alpha}{2}}^+ \times \hat{\sigma}_{\hat{U}_\tau}(\mathbf{Y} \underline{\mathbf{x}}; \mathbf{x}_\tau)$

où $\delta_{lim, \frac{\alpha}{2}}^+$ est le quantile d'ordre $(1 - \frac{\alpha}{2})$ de la loi de probabilités de $\Delta_{Y_\tau, \hat{Y}_\tau}$. Bien que très ressemblant à un intervalle de confiance, l'intervalle de prévision est fondamentalement différent puisqu'il cherche un encadrement d'une variable aléatoire (en l'occurrence Y_τ) et non pas un paramètre réel. Comme vous en avez l'habitude, nous illustrons cette notion avec l'exemple de cours.