



M1 Miashs

Modèles de régression linéaire et outils du diagnostic

Master 1
2020-2021

I. Régression linéaire multiple

La régression linéaire multiple est une extension de la régression linéaire simple. Elle permet de prendre en compte plusieurs variables explicatives pour expliquer les variations d'une variable aléatoire quantitative notée Y . En d'autres termes, nous recherchons l'effet sur Y de ces variables explicatives.

La régression linéaire multiple peut également être utilisée pour prédire une valeur de Y à partir de valeurs prises par les variables explicatives.

1. Exemple d'introduction : Indice de fécondité en Suisse en 1888

Cet exemple, tiré d'un jeu de données ("swiss") disponible dans le logiciel *R*, concerne 47 provinces francophones de la Suisse autour de l'an 1888. On cherche dans un premier temps à expliquer le niveau de fécondité des populations des provinces par la variable :

- "Agriculture" qui représente le % d'hommes travaillant dans l'agriculture.
- puis par "Education" qui mesure le % de conscrits ayant un niveau d'éducation supérieur à l'école primaire.
- ensuite par "Catholique" qui modélise le % de catholiques (par opposition aux protestants).
- et enfin par "Mortalité infantile" qui mesure le % d'enfants dont l'espérance de vie est inférieure à 1 an.

Pour cela, nous utilisons quatre régressions linéaires simples (une pour chacune des 4 variables explicatives). Rappelons que le modèle de régression linéaire simple peut s'écrire approximativement

$$y_i \approx \beta_0 + \beta_1 x_i$$

où, pour la province $n^{\circ}i$, y_i désigne l'indice de fécondité observé, x_i la valeur observée d'une des 4 variables explicatives, et β_0, β_1 sont les paramètres inconnus à estimer. Les résultats des 4 régressions sont résumés dans le tableau ci-dessous.

Estimation	$\hat{\beta}_0$	$\hat{\beta}_1$	R^2
Agriculture	60.3	0.194	0.1247
Education	79.6	-0.862	0.4406
Catholique	64.4	0.139	0.2150
Mortalité infantile	34.5	1.787	0.1735

$\hat{\beta}_0$ et $\hat{\beta}_1$ sont les estimations de β_0 et β_1 , et R^2 est le coefficient de détermination mesurant la qualité d'ajustement du modèle linéaire.

En réalité, on peut mieux expliquer et prédire l'indice de fécondité par un modèle prenant en compte les 4 variables explicatives en même temps, c'est-à-dire en utilisant une régression linéaire multiple telle que

$$y_i \approx \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4}$$

où, pour la province $n^o i$, y_i désigne l'indice de fécondité observé, et x_{i1} , x_{i2} , x_{i3} , x_{i4} les valeurs observées des variables "Agriculture", "Education", "Catholique" et "Mortalité infantile", respectivement. Le résultats de la régression multiple sont résumés dans le tableau ci-dessous.

Estimation	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	R^2 ajusté
	62.1	-0.155	-0.98	0.125	1.078	0.6707

La valeur du coefficient de détermination ajusté (0.6707), qui tient compte du nombre de régresseurs, montre un meilleur ajustement du modèle de régression linéaire multiple que lorsque l'on considère une régression simple.

2. Modèle de régression linéaire multiple

Nous allons à présent donner l'écriture formelle du modèle de régression linéaire multiple avec p régresseurs (ou variables explicatives).

La variable aléatoire Y est appelée la variable à expliquer (ou variable dépendante). Cette variable aléatoire Y dépend de p régresseurs déterministes x_1, x_2, \dots, x_p (qui sont donc connus et non aléatoires).

Pour l'individu $n^{\circ}i$, elle peut s'écrire au moyen de l'équation ci-dessous :

$$Y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \varepsilon_i \quad (1)$$

où $\beta_0, \beta_1, \dots, \beta_p$ sont des paramètres inconnus à estimer et ε_i désigne l'erreur (aléatoire) que l'on fait lorsque l'on mesure Y_i .

L'erreur ε_i est une variable aléatoire d'espérance nulle et de variance inconnue σ^2 qui est un autre paramètre à estimer.

Comme dans le cas de la régression simple, les paramètres β_j , $j = 0, 1, \dots, p$ peuvent être estimés par la méthode des moindres carrés, c'est-à-dire en minimisant la somme des carrés

$$\sum_{i=1}^n \left(Y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2$$

par rapport à β_0, \dots, β_p (voir plus loin).

Les paramètres estimés sont notés $\hat{\beta}_j$, $j = 0, \dots, p$.

On a alors la valeur estimée (ou ajustée) suivante de Y_i :

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \dots + \hat{\beta}_p x_{ip}$$

et $\hat{\varepsilon}_i = Y_i - \hat{Y}_i$ est le résidu (ou “erreur de reconstitution”) associé à l'individu $n^\circ i$.

3. Ecriture matricielle

Notons $\mathbf{Y} = (Y_1, \dots, Y_n)^T$ le vecteur colonne de taille n des variables aléatoires Y_i associées aux n individus de l'échantillon.

L'équation (1) peut alors s'écrire sous forme matricielle comme ci-dessous :

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

où

- $\mathbf{X} = (1, \mathbf{x}_1, \dots, \mathbf{x}_p)$ est une matrice de dimension $n \times (p + 1)$ vérifiant l'hypothèse (H_1) ci-dessous.

Hypothèse (H_1) : \mathbf{X} est une matrice de plein rang, c'est-à-dire $\text{rang}(\mathbf{X}) = p + 1$.

Chaque régresseur \mathbf{x}_j , $j = 1, \dots, p$, est un vecteur colonne de taille n composé des éléments x_{ij} , $i = 1, \dots, n$, associés aux n individus.

La 1^{re} colonne de la matrice \mathbf{X} contient 1 qui est un vecteur colonne de 1 (de taille n).

- $\beta = (\beta_0, \beta_1, \dots, \beta_p)^T$ est le vecteur colonne des paramètres (de taille $p + 1$). Le paramètre β_0 représente la constante (“intercept” dans les logiciels anglais).
- $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^T$ est le vecteur colonne des erreurs aléatoires associées aux n individus, vérifiant l’hypothèse H_2 ci-dessous.

Hypothèse (H_2) : Les erreurs ε_i sont centrées, de même variance et non corrélées entre elles.

D’après (H_2), on a donc

$$E(\varepsilon) = 0_n \text{ et } \Sigma_\varepsilon = \sigma^2 I_n,$$

où 0_n est le vecteur colonne de zéros (de taille n) et I_n est la matrice identité d’ordre n .

4. Estimateur des moindres carrés

Théorème 1. D'après l'hypothèse (H_1) , la matrice $\mathbf{X}^T \mathbf{X}$ étant inversible, l'estimateur des moindres carrés $\hat{\beta}$ de β a la forme suivante

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$$

Éléments de preuve : la fonction de β à minimiser est

$$\begin{aligned} \|\mathbf{Y} - \mathbf{X}\beta\|^2 &= (\mathbf{Y} - \mathbf{X}\beta)^T (\mathbf{Y} - \mathbf{X}\beta) \\ &= \mathbf{Y}^T \mathbf{Y} - 2\beta^T \mathbf{X}^T \mathbf{Y} + \beta^T \mathbf{X}^T \mathbf{X} \beta \end{aligned}$$

La dérivée par rapport à β est alors égale à :

$$-2\mathbf{X}^T \mathbf{Y} + 2\mathbf{X}^T \mathbf{X} \beta$$

En posant la dérivée égale à zéro, on obtient $\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$.

4.1 Propriétés de l'estimateur $\hat{\beta}$

Sous les hypothèses $(H_1) - (H_2)$, les propriétés suivantes sont vérifiées :

- $\hat{\beta}$ est un estimateur sans biais, c'est-à-dire, $E(\hat{\beta}) = \beta$.
- La matrice de variance-covariance de $\hat{\beta}$ est

$$\sigma_{\hat{\beta}}^2 = \sigma^2(\mathbf{X}^T \mathbf{X})^{-1}. \quad (2)$$

- Parmi les estimateurs sans biais fonctions linéaires des Y_i , $\hat{\beta}$ est de variance minimale (propriété de Gauss-Markov).

4.2 Résidus et estimation de la variance σ^2

Le vecteur des valeurs estimées de \mathbf{Y} est $\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = P_{\mathbf{X}}\mathbf{Y}$ avec $P_{\mathbf{X}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$.

C'est la projection de \mathbf{Y} sur le sous-espace $\text{Vect}(\mathbf{X})$ engendré par les colonnes de \mathbf{X} (voir la figure ci-après). De plus, $P_{\mathbf{X}}$ est la matrice de projection orthogonale sur $\text{Vect}(\mathbf{X})$.

Le vecteur des résidus est défini par $\hat{\boldsymbol{\varepsilon}} = \mathbf{Y} - \hat{\mathbf{Y}} = (\mathbf{I}_n - P_{\mathbf{X}})\mathbf{Y}$

Un estimateur sans biais de la variance σ^2 est alors

$$\hat{\sigma}^2 = \frac{1}{n-p-1} \|\hat{\boldsymbol{\varepsilon}}\|^2 = \frac{1}{n-p-1} \sum_{i=1}^n \hat{\varepsilon}_i^2$$

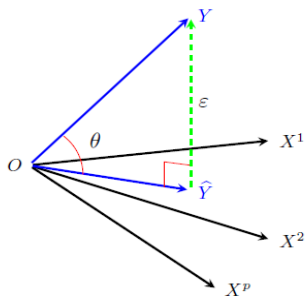


FIGURE 1 – Géométriquement, la régression est la projection \hat{Y} de Y sur l'espace vectoriel $\text{Vect}\{1, X^1, \dots, X^p\}$; de plus $R^2 = \cos^2(\theta)$.

4.3 Estimateur de la variance de $\hat{\beta}$

A partir de (2) et de l'estimateur $\hat{\sigma}^2$ de la variance σ^2 , nous pouvons obtenir un estimateur de la variance de $\hat{\beta}$ en remplaçant σ^2 par son estimateur :

$$\hat{\sigma}_{\hat{\beta}}^2 = \hat{\sigma}^2 (\mathbf{X}^T \mathbf{X})^{-1}$$

Pour chaque $j = 1, \dots, p + 1$, nous avons donc un estimateur de la variance et de l'écart-type de l'estimateur $\hat{\beta}_{j-1}$ de chaque coefficient de la régression β_{j-1} :

$$\hat{\sigma}_{\hat{\beta}_{j-1}}^2 = \hat{\sigma}^2 [(\mathbf{X}^T \mathbf{X})^{-1}]_{jj} \quad \text{et} \quad \hat{\sigma}_{\hat{\beta}_{j-1}} = \sqrt{\hat{\sigma}^2 [(\mathbf{X}^T \mathbf{X})^{-1}]_{jj}} \quad (3)$$

où $[(\mathbf{X}^T \mathbf{X})^{-1}]_{jj}$ est l'élément $n^\circ j$ sur la diagonale de la matrice $(\mathbf{X}^T \mathbf{X})^{-1}$.

4.4 Exemple : Prédiction du sentiment de bien-être

Un des objectifs de l'étude dont est tiré cet exemple était de prédire le sentiment de bien-être (mesuré par un score) sept ans après avoir quitté l'Université, à partir de variables mesurées lorsque l'étudiant se trouvait à l'Université. Les données sont résumées dans le tableau ci-dessous.

Individu	NbEnfU	BEU	NSE	BE
1	2	17	52	21
2	2	20	56	26
3	0	21	27	37
4	0	18	34	40
5	0	31	29	35
6	1	34	38	37
7	0	20	38	35
8	0	17	25	26
9	2	48	53	42
10	0	16	36	38

La variable à expliquer est le sentiment de bien-être (variable notée “BE”), mesuré par un score, et les variables explicatives sont : le nombre d’enfants lors de la scolarité à l’Université (“NbEnfU”), le sentiment de bien-être mesuré au cours de la scolarité à l’Université (“BEU”) et le niveau socio-économique des parents (“NSE”), également mesuré par un score. D’après le tableau des données, on a donc

$$\mathbf{X} = \begin{pmatrix} 1 & 2 & 17 & 52 \\ 1 & 2 & 20 & 56 \\ 1 & 0 & 21 & 27 \\ 1 & 0 & 18 & 34 \\ 1 & 0 & 31 & 29 \\ 1 & 1 & 34 & 38 \\ 1 & 0 & 20 & 38 \\ 1 & 0 & 17 & 25 \\ 1 & 2 & 48 & 53 \\ 1 & 0 & 16 & 36 \end{pmatrix} \quad \text{et} \quad \mathbf{y} = \begin{pmatrix} 21 \\ 26 \\ 37 \\ 40 \\ 35 \\ 37 \\ 35 \\ 26 \\ 42 \\ 38 \end{pmatrix}$$

Nous allons obtenir une estimation des moindres carrés $\hat{\beta}$ du vecteur de paramètres $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)^T$ grâce à l'expression $\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$. En utilisant la transposée et la multiplication de matrices, on a

$$\mathbf{X}^T \mathbf{X} = \begin{pmatrix} 10 & 7 & 242 & 388 \\ 7 & 13 & 204 & 360 \\ 242 & 204 & 6820 & 9679 \\ 388 & 360 & 9679 & 16184 \end{pmatrix}$$

L'inverse de la matrice ci-dessus donne

$$(\mathbf{X}^T \mathbf{X})^{-1} = \begin{pmatrix} 7.9183 & 2.4750 & -0.0493 & -0.2154 \\ 2.4750 & 0.9790 & -0.0132 & -0.0732 \\ -0.0493 & -0.0132 & 0.0013 & 0.0007 \\ -0.2154 & -0.0732 & 0.0007 & 0.0064 \end{pmatrix}$$

D'autre part,

$$\mathbf{X}^T \mathbf{y} = \begin{pmatrix} 337 \\ 215 \\ 8483 \\ 12902 \end{pmatrix}$$

Finalement en multipliant les matrices $(\mathbf{X}^T \mathbf{X})^{-1}$ et $\mathbf{X}^T \mathbf{y}$, on obtient

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} 3.2195 \\ -12.0562 \\ 0.5803 \\ 0.6411 \end{pmatrix}$$

Les valeurs estimées $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ et les résidus $\hat{\boldsymbol{\varepsilon}} = \mathbf{y} - \hat{\mathbf{y}}$ sont

$$\hat{\mathbf{y}} = \begin{pmatrix} 22.3115 \\ 26.6170 \\ 32.7169 \\ 35.4639 \\ 39.8025 \\ 35.2574 \\ 39.1891 \\ 29.1134 \\ 40.9427 \\ 35.5855 \end{pmatrix} \quad et \quad \hat{\boldsymbol{\varepsilon}} = \begin{pmatrix} -1.3115 \\ -0.6170 \\ 4.2831 \\ 4.5361 \\ -4.8025 \\ 1.7426 \\ -4.1891 \\ -3.1134 \\ 1.0573 \\ 2.4145 \end{pmatrix}$$

Une estimation sans biais de la variance est donc

$$\hat{\sigma}^2 = \frac{1}{n - p - 1} \|\hat{\varepsilon}\|^2 = \frac{1}{10 - 4} \sum_{i=1}^{10} \hat{\varepsilon}_i^2 = 16.8852$$

D'après la valeur de $\hat{\beta}$, l'équation de régression est donc :

$$\hat{y} = 3.2195 - 12.0562 * NbEnfU + 0.5803 * BEU + 0.6411 * NSE$$

A partir de cette équation, nous pouvons prédire le score de bien-être. Considérons, par exemple, un ancien étudiant ayant eu, durant sa scolarité à l'Université, 1 enfant, un score de bien-être de 30 et dont le niveau socio-économique des parents est égal à 50.

D'après l'équation, son score de bien-être (7 ans après avoir quitté l'Université) peut être estimé par

$$\hat{y} = 40.6273 \approx 41.$$

5. Coefficient de détermination

Sommes des carrés

On définit SCR comme la somme des carrés des résidus,

$$SCR = ||\mathbf{Y} - \hat{\mathbf{Y}}||^2 = ||\hat{\epsilon}||^2.$$

La somme des carrés expliquée par le modèle est également défini par

$$SCE = ||\hat{\mathbf{Y}} - \bar{y}\mathbf{1}||^2$$

La somme des carrés totale est

$$SCT = ||\mathbf{Y} - \bar{y}\mathbf{1}||^2$$

On a alors

$$SCT = SCE + SCR.$$

Définition du coefficient de détermination

Le coefficient de détermination est le rapport

$$R^2 = \frac{SCE}{SCT}$$

qui est donc la part de variation de Y expliquée par le modèle de régression.

Plus R^2 est proche de 1, meilleure est la qualité d'ajustement du modèle.

Coefficient de détermination ajusté

Le coefficient de détermination ne tient pas compte du nombre de régresseurs. On définit donc un R^2 ajusté

$$\tilde{R}^2 = 1 - \frac{n-1}{n-p-1} \frac{SCR}{SCT}.$$

6. Inférence dans le modèle gaussien

En plus des hypothèses précédentes, nous supposons que les erreurs ε_i vérifient l'hypothèse (H_3) ci-dessous.

Hypothèse (H_3) : Les erreurs ε_i sont gaussiennes, c'est-à-dire que pour tout i ,

$$\varepsilon_i \sim \mathcal{N}(0, \sigma^2).$$

Les hypothèses H_2 et H_3 impliquent que les erreurs ε_i sont indépendantes les unes des autres et on a donc

$$\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^T \sim \mathcal{N}_n(0_n, \sigma^2 I_n).$$

6.1 Estimation par maximum de vraisemblance

Théorème 2. Sous (H_1) - (H_3) , l'estimateur du maximum de vraisemblance $\hat{\beta}_{MV}$ de β est

$$\hat{\beta}_{MV} = \hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}.$$

Éléments de preuve : rappelons tout d'abord que la densité de la variable aléatoire réelle $Y \sim \mathcal{N}(\mu, \sigma^2)$ est donnée par

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$$

Les variables aléatoires $Y_i \sim \mathcal{N}(\beta_0 + \sum_{j=1}^p \beta_j x_{ij}, \sigma^2)$ étant indépendantes, la vraisemblance $\mathcal{L}(\mathbf{y}, \mathbf{X}, \beta, \sigma^2)$ du modèle s'écrit

$$\prod_{i=1}^n f_{Y_i}(y_i) = \left(\frac{1}{2\pi\sigma^2} \right)^{n/2} \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 \right]$$

Donc,

$$\mathcal{L}(\mathbf{y}, \mathbf{X}, \boldsymbol{\beta}, \sigma^2) = \prod_{i=1}^n f_{Y_i}(y_i) = \left(\frac{1}{2\pi\sigma^2} \right)^{n/2} \exp \left[-\frac{1}{2\sigma^2} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2 \right]$$

et la log-vraisemblance est

$$\log \mathcal{L}(\mathbf{y}, \mathbf{X}, \boldsymbol{\beta}, \sigma^2) = -\frac{n}{2} \log \sigma^2 - \frac{n}{2} \log 2\pi - \frac{1}{2\sigma^2} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2.$$

En dérivant par rapport aux paramètres $\boldsymbol{\beta}$ et σ^2 , nous obtenons :

$$\frac{\mathcal{D}}{\mathcal{D}\boldsymbol{\beta}} \log \mathcal{L}(\mathbf{y}, \mathbf{X}, \boldsymbol{\beta}, \sigma^2) = -\frac{1}{2\sigma^2} \frac{\mathcal{D}}{\mathcal{D}\boldsymbol{\beta}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2$$

et

$$\frac{\mathcal{D}}{\mathcal{D}\sigma^2} \log \mathcal{L}(\mathbf{y}, \mathbf{X}, \boldsymbol{\beta}, \sigma^2) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2$$

En posant la première dérivée égale à 0, nous retrouvons bien le même estimateur de β que par les moindres carrés.

Remarque : en posant la seconde dérivée égale à 0, nous obtenons l'estimateur suivant de la variance σ^2 ,

$$\hat{\sigma}_{MV}^2 = \frac{\|\mathbf{Y} - \mathbf{X}\hat{\beta}_{MV}\|^2}{n} = \frac{\|\mathbf{Y} - \hat{\mathbf{Y}}\|^2}{n} = \frac{\|\hat{\epsilon}\|^2}{n},$$

qui n'est pas sans biais. On préférera donc plutôt l'estimateur sans biais $\hat{\sigma}^2$ de la variance σ^2 .

6.2 Propriétés des estimateurs $\hat{\beta}$ et $\hat{\sigma}^2$

Sous les hypothèses (H1) – (H3), nous avons

- i) $\hat{\beta} \sim \mathcal{N}_{p+1}(\beta, \sigma^2(\mathbf{X}^T \mathbf{X})^{-1})$.
- ii) $\frac{(n-p-1)\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-p-1)$.
- iii) $\hat{\beta}$ et $\hat{\sigma}^2$ sont indépendants.
- iv) Pour $j = 1, \dots, p+1$,
$$T_{j-1} = \frac{\hat{\beta}_{j-1} - \beta_{j-1}}{\hat{\sigma} \sqrt{[(\mathbf{X}^T \mathbf{X})^{-1}]_{jj}}} \sim \text{Student}(n-p-1),$$
 où $[(\mathbf{X}^T \mathbf{X})^{-1}]_{jj}$ est l'élément $n^\circ j$ sur la diagonale de la matrice $(\mathbf{X}^T \mathbf{X})^{-1}$.

Remarque : la propriété iv) est une conséquence des propriétés i)-iii) et de (3).

6.3 Intervalle de confiance d'un coefficient β_{j-1}

Pour chaque $j = 1, \dots, p + 1$, d'après la propriété iv), nous pouvons construire un intervalle de confiance à $1 - \alpha$ de β_{j-1} dont les bornes sont

$$\hat{\beta}_{j-1} \mp t_{n-p-1}(1 - \alpha/2) \sqrt{\hat{\sigma}^2 [(\mathbf{X}^T \mathbf{X})^{-1}]_{jj}}$$

où $t_{n-p-1}(1 - \alpha/2)$ est le quantile de probabilité $1 - \alpha/2$ d'une loi de Student à $n - p - 1$ degrés de liberté (ddl).

6.4 Test de Student de signification d'un coefficient β_{j-1}

Nous souhaitons tester $H_0 : \beta_{j-1} = 0$ contre $H_1 : \beta_{j-1} \neq 0$.

Nous rappelons que, sous H_0 , la statistique du test

$$T_{j-1} = \frac{\hat{\beta}_{j-1}}{\hat{\sigma} \sqrt{[(\mathbf{X}^T \mathbf{X})^{-1}]_{jj}}}$$

suit une loi de Student à $(n - p - 1)$ ddl.

La région de rejet du test est :

$$\text{Rejet de } H_0 \text{ si } |T_{j-1}| > t_{n-p-1}(1 - \alpha/2)$$

où $t_{n-p-1}(1 - \alpha/2)$ est le quantile de probabilité $1 - \alpha/2$ d'une loi de Student à $n - p - 1$ degrés de liberté (ddl).

6.5 Exemple : Données économiques

Nous illustrons l'intervalle de confiance et le test de Student de signification d'un coefficient β_{j-1} par un exemple extrait du logiciel *R* ("longley"). Dans cet exemple, on cherche à expliquer le nombre d'actifs ayant un emploi aux Etats-Unis, par deux variables économiques, le produit national brut (PNB) et la population du pays. Le nombre d'actifs et la population sont exprimés en millions d'habitants, et le PNB en milliards de dollars. Les données concernent les années 1955-1962.

Année	PNB	Pop	Actifs
1955	397.469	117.388	66.019
1956	419.180	118.734	67.857
1957	442.769	120.445	68.169
1958	444.546	121.950	66.513
1959	482.704	123.366	68.655
1960	502.601	125.368	69.564
1961	518.173	127.852	69.331
1962	554.894	130.081	70.551

En utilisant les mêmes méthodes que dans l'exemple précédent, nous obtenons

$$\mathbf{X}^T \mathbf{X} = \begin{pmatrix} 8.000 & 3762.336 & 985.184 \\ 3762.336 & 1789380.538 & 464960.671 \\ 985.184 & 464960.671 & 121460.012 \end{pmatrix}$$

et l'inverse de la matrice

$$(\mathbf{X}^T \mathbf{X})^{-1} = \begin{pmatrix} 2813.7174 & 2.6881 & -33.1130 \\ 2.6881 & 0.0027 & -0.0320 \\ -33.1130 & -0.0320 & 0.3912 \end{pmatrix}$$

Pour les 3 estimations de $\hat{\beta}_{j-1}$ ($j = 1, 2, 3$), on a

$$\hat{\beta} = \begin{pmatrix} 120.0801 \\ 0.0882 \\ -0.7570 \end{pmatrix}$$

Une estimation sans biais de la variance est donc

$$\hat{\sigma}^2 = \frac{1}{8-3} \|\mathbf{y} - \mathbf{X}\hat{\beta}\|^2 = 0.2563$$

Par conséquent, les 3 estimations $\hat{\sigma}_{\hat{\beta}_{j-1}} = \sqrt{\hat{\sigma}^2 [(\mathbf{X}^T \mathbf{X})^{-1}]_{jj}}$ des écart-types de $\hat{\beta}_{j-1}$ sont

$$\begin{pmatrix} 26.8533 \\ 0.0262 \\ 0.3167 \end{pmatrix}$$

Ici $n = 8$, $p = 2$ donc pour $\alpha = 5\%$, le quantile de probabilité 97.5% d'une loi de Student à $8 - 2 - 1 = 5$ ddl est

$$t_{n-p-1}(1 - \alpha/2) = t_5(97.5\%) = 2.57058.$$

En appliquant la formule du cours, l'intervalle de confiance de β_1 à 95% ($\alpha = 5\%$) est donc

$$IC(\beta_1) = [0.02089; 0.15547]$$

Concernant maintenant le test de student, pour $j = 1, 2, 3$, les valeurs observées des statistiques T_{j-1} sont respectivement égales à 4.471, 3.369 et -2.391.

En prenant aussi $\alpha = 5\%$, la conclusion est le rejet de l'hypothèse $\beta_{j-1} = 0$ pour $j = 1, 2$ et non rejet de l'hypothèse $\beta_2 = 0$ pour $j = 3$.

6.6 Inférence sur le modèle complet

- Le modèle peut être testé globalement.

Sous l'hypothèse nulle $H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$, la statistique :

$$F = \frac{SCE/p}{SCR/(n-p-1)}$$

avec

$SCE = ||\hat{\mathbf{Y}} - \bar{y}\mathbf{1}||^2 =$ Somme des Carrés Expliquée,

$SCR = ||\mathbf{Y} - \hat{\mathbf{Y}}||^2 =$ Somme des Carrés des Résidus,

suit une loi de Fisher $F_{p,n-p-1}$ à p et $(n-p-1)$ degrés de liberté.

- Les résultats sont habituellement présentés dans un tableau “*d’analyse de la variance*” sous la forme suivante :

Source de variation	d.d.l	Somme des carrés	Variance	F-Statistique
Régression	p	SCE	SCE/p	F
Erreur	$n - p - 1$	SCR	$SCR/(n - p - 1)$	
Total	$n - 1$	SST		

6.7 Inférence sur le modèle réduit

- Le test précédent amène à rejeter H_0 dès que l'une des variables X_j est liée à Y . Il est donc d'un intérêt limité. Il est souvent plus utile de tester un modèle réduit c'est-à-dire dans lequel certains coefficients sont nuls (à l'exception du terme constant) contre le modèle complet avec toutes les variables.
- En ayant éventuellement réordonné les variables, on considère l'hypothèse nulle $H_0 : \beta_1 = \beta_2 = \dots = \beta_q = 0$, $q < p$.

Notons respectivement SCE_q , SCR_q , R_q^2 les sommes de carrés et le coefficient de détermination du modèle réduit à $(p - q)$ variables. Sous H_0 , la statistique

$$F_q = \frac{(SCE - SCE_q)/q}{SCR/(n - p - 1)} = \frac{(R^2 - R_q^2)/q}{(1 - R^2)/(n - p - 1)}$$

suit une loi de Fisher $F_{q, n-p-1}$ à q et $(n - p - 1)$ degrés de liberté.

- Remarque : Dans le cas particulier où $q = 1$ (on teste alors $\beta_1 = 0$), la F -statistique est alors le carrée de la t -statistique (Student) de l'inférence sur un paramètre, et conduit donc au même test.

6.8 Prévisions

- Connaissant les valeurs des variables X_j pour une nouvelle observation : $\mathbf{x}_0^T = [x_{01}, x_{02}, \dots, x_{0p}]$ appartenant au domaine dans lequel l'hypothèse de linéarité reste valide, une prévision, notée \hat{y}_0 de Y est donnée par :

$$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_{01} + \dots + \hat{\beta}_p x_{0p}.$$

- L'intervalle de confiance des prévisions de Y pour *une* valeur $\mathbf{x}_0 \in \mathbb{R}^p$, en posant $\mathbf{v}_0 = (1, \mathbf{x}_0^T)^T \in \mathbb{R}^{p+1}$, est :

$$\left[\hat{y}_0 \pm t_{n-p-1}(1 - \alpha/2) \hat{\sigma} (1 + \mathbf{v}_0^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{v}_0)^{1/2} \right]$$

où $\hat{\sigma}^2$ est un estimateur de la variance σ^2 de l'erreur ε_i .