Table des matières

1 Le modèle linéaire simple **5**
1.1 Introduction à l'économétrie 5
1.2 Le modèle linéaire simple 5
1.3 Présentation du modèle 5
1.3.1 Exemple introductif 5
1.3.2 Spécification 6
1.3.3 Rôle du terme aléatoire 6
1.3.4 Estimation des paramètres 7
1.3.5 Méthode des moindres carrés ordinaires (MCO) 9
1.3.6 Propriétés des estimateurs 9
1.3.7 Construction des tests 11
1.3.8 Équation et tableau d'analyse de la variance 13
1.3.9 La prévision dans le modèle de régression simple 15

2 Le modèle linéaire multiple **17**
2.1 Introduction 17
2.2 La méthode des moindres carrés 17
2.2.1 Les hypothèses de la méthode des MCOs 17
2.2.2 Estimateurs des MCO 18
2.2.3 Propriétés des estimateurs 18
2.3 Équation d'analyse de la variance et qualité d'un ajustement 19
2.4 Construction des tests 22
2.4.1 Construction du tableau d'analyse de la variance et test de signification globale d'une régression 24

Chapitre 1

Le modèle linéaire simple

1.1 Introduction à l'économétrie

L'économétrie est le principal outil d'analyse quantitative utilisé par les économistes et gestionnaires dans divers domaines d'application, comme la macroéconomie, la finance ou le marketing.
Les méthodes de l'économétrie permettent de vérifier l'existence de certaines relations entre des phénomènes économiques, et de mesurer concrètement ces relations, sur la base d'observations de faits réels.
Dans son acception la plus restreinte, l'économétrie est un ensemble de techniques utilisant la statistique mathématique qui vérifient la validité empirique des relations supposées entre les phénomènes économiques et mesurent les paramètres de ces relations. Au sens large, l'économétrie est l'art de construire et d'estimer des modèles empiriques adéquats par rapport aux caractéristiques de la réalité, et intelligibles au regard de la théorie économique.

1.2 Le modèle linéaire simple

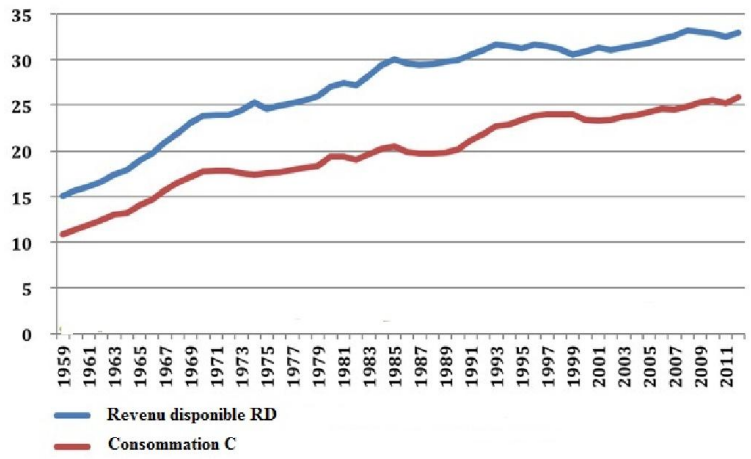
Nous commençons notre étude par le modèle le plus simple : une variable **endogène** est expliquée par une variable **exogène**. Après avoir étudié les conséquences probabilistes de l'erreur d'observation, nous présentons :

- ➔ les formules de base permettant d'estimer les paramètres du modèle
- ➔ les hypothèses stochastiques et leurs conséquences sont étudiées
- ➔ la qualité de l'estimation d'un modèle est examinée à l'aide des premiers tests statistiques (Student, Fisher)
- ➔ le modèle de régression simple est étudié en tant qu'outil de prévision avec le degré de confiance que nous pouvons en attendre.

1.3 Présentation du modèle

1.3.1 Exemple introductif

Admettons que nous constatons le fait économique illustrer dans la figure suivante :
On observe que les 2 courbes évoluent pratiquement dans le même sens : elles augmentent et diminuent simultanément. On peut penser qu'il y a un lien entre ces 2 variables. On peut en effet penser que la



consommation C des ménages est influencée par le revenu disponible RD. Lorsque le revenu augmente, la consommation s'accroît.
Soit la fonction de consommation **keynésienne** :

$$C = a_0 + a_1 Y,$$

où C = consommation, Y = revenu, a_1 = propension marginale à consommer et a_0 = consommation autonome ou incompressible.

1.3.2 Spécification

Nous pouvons distinguer deux types de spécifications :

- **Les modèles en série temporelle**, les variables représentent des phénomènes observés à intervalles de temps réguliers, par exemple la consommation et le revenu annuel sur 20 ans pour un pays donné. Le modèle s'écrit alors :

  $$C_t = a_0 + a_1 Y_t \quad t = 1, \ldots, 20,$$

  où C_t et Y_t sont la consommation et le revenu au temps t.

- **Les modèles en coupe instantanée**, les variables représentent des phénomènes observés au même instant mais concernant plusieurs individus, par exemple la consommation et le revenu observés sur un échantillon de 20 pays. Le modèle s'écrit alors :

  $$C_i = a_0 + a_1 Y_i \quad i = 1, \ldots, 20,$$

  où C_i et Y_i sont la consommation et le revenu du pays i pour une année donnée.

1.3.3 Rôle du terme aléatoire

Il existe une multitude de facteurs susceptibles d'expliquer la consommation. C'est pourquoi nous ajoutons un terme (ε_t) qui synthétise l'ensemble de ces informations non explicitées dans le modèle : $C_t = a_0 + a_1 Y_t + \varepsilon_t$ si le modèle est spécifie en série temporelle ($C_i = a_0 + a_1 Y_i + \varepsilon_i$ si le modèle est spécifie en coupe instantanée).
Le terme ε_t représente l'erreur de specification du modèle, c'est-a-dire l'ensemble des phénomènes explicatifs de la consommation non liés au revenu. Il mesure la différence entre les valeurs réellement observées de C_t et les valeurs qui auraient été observées si la relation spécifiée avait été rigoureusement exacte.

1.3.4 Estimation des paramètres

Soit le modèle suivant :

$$\boxed{y_t = a_0 + a_1 x_t + \varepsilon_t \quad t = 1, \ldots, n.}$$

Afin d'estimer et d'étudier les deux paramètres inconnus a_0 et a_1, nous introduisons les hypothèses suivantes :

- H1 : Le modèle est linéaire en x_t (ou en n'importe quelle transformation de x_t).
- H2 : Les valeurs x_t sont observées sans erreur (x_t non aléatoires).
- H3 : $E(\varepsilon_t) = 0, \quad t = 1, \ldots, n.$
- H4 : $E(\varepsilon_t^2) = \sigma_\varepsilon^2, \quad t = 1, \ldots, n$, la variance de l'erreur est constante.
- H5 : $t \neq t' \Rightarrow E(\varepsilon_t \varepsilon_{t'}) = 0$, les erreurs sont non corrélées.
- H6 : $Cov(x_t, \varepsilon_t) = 0, \quad t = 1, \ldots, n$, l'erreur est indépendante de la variable explicative.

La représentation graphique ne donne qu'une impression de la corrélation entre deux variables sans donner une idée précise de l'intensité de la liaison (voir 1.1), c'est pourquoi nous calculons une statistique appelée coefficient de corrélation linéaire simple, noté $r_{x,y}$. Il est égal à :

$$\boxed{r = r_{x,y} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} = \frac{\sum_{t=1}^n (x_t - \overline{x})(y_t - \overline{y})}{\sqrt{\sum_{t=1}^n (x_t - \overline{x})^2}\sqrt{\sum_{t=1}^n (y_t - \overline{y})^2}}}$$

Ce coefficient n'est calculé que à partir d'un échantillon d'observations et non pas sur l'ensemble des valeurs.

On appelle $\rho_{x,y}$ ce coefficient empirique qui est une estimation du vrai coefficient $r_{x,y}$.
Soit à tester l'hypothèse $H_0 : r_{x,y} = 0$, contre l'hypothèse $H_1 : r_{x,y} \neq 0$.
Nous pouvons démontrer que

$$\boxed{\frac{\rho_{x,y}}{\sqrt{\frac{1-\rho_{x,y}^2}{n-2}}} \rightsquigarrow \mathcal{T}_{n-2}}$$

Nous calculons alors une statistique, appelé le t de Student empirique :

$$\boxed{t^* = \frac{|\rho_{x,y}|}{\sqrt{\frac{1-\rho_{x,y}^2}{n-2}}}}$$

Si $t^* > t_{n-2}^{\alpha/2}$ valeur lue dans une table de Student au seuil $\alpha = 5\%$ à $n-2$ degrés de liberté, nous rejetons l'hypothèse H_0, **le coefficient de correlation est donc significativement different de** 0.
Dans le cas contraire, l'hypothèse d'un coefficient de corrélation nul est acceptée.

Exemple 1.3.1 *On relève* 10 *couples* (x, y) *de données consignés dans le tableau*

x	16	18	23	24	28	29	26	31	32	34
y	20	24	28	22	32	28	32	36	41	41

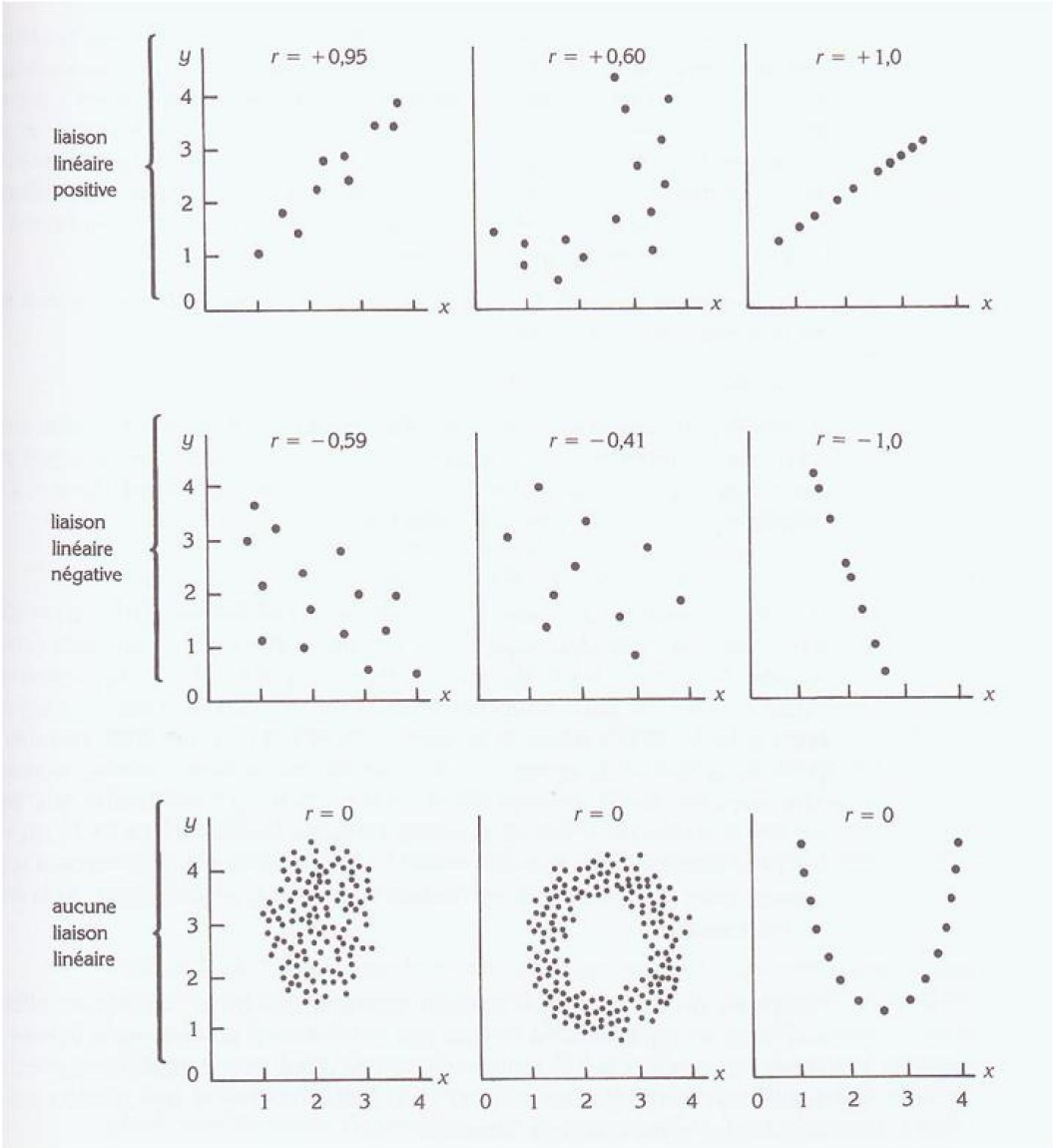
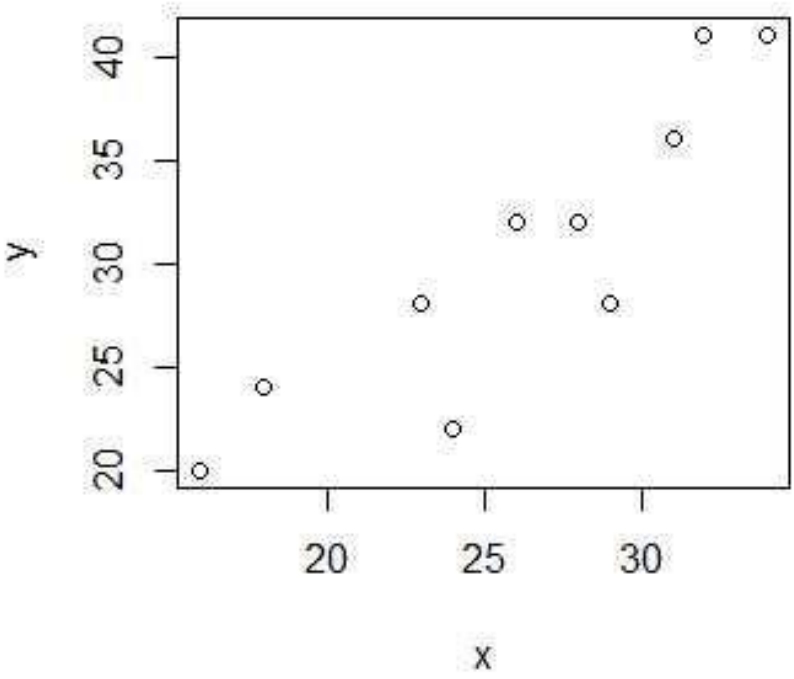


FIGURE 1.1 – Corrélation entre deux variables

1) Tracer le nuage de points et le commenter.
2) Calculer le coefficient de correlation simple et tester sa signification par rapport à 0 *pour un seuil* $\alpha = 0,05$.
Réponses :
1) Le nuage de points indique que les couples de valeurs sont approximativement alignés : les deux variables semblent corrélées positivement.
2) Après calcul, on obtient $\rho_{x,y} = 0,89$.
Le t *de Student empirique est égal à :*

$$t^* = \frac{|\rho_{x,y}|}{\sqrt{\frac{1-\rho_{x,y}^2}{n-2}}} = 5,49 > t_8^{0,025} = 2,306.$$

Donc le coefficient de corrélation entre x *et* y *est significativement différent de* 0.



1.3.5 Méthode des moindres carrés ordinaires (MCO)

Le nuage de points peut s'ajuster à l'aide d'une droite (voir 1.2). L'estimateur des coefficients a_0 et a_1 est obtenu en minimisant la distance au carré entre chaque observation et la droite, d'où le nom d'estimateur **des moindres carrés ordinaires (MCO) :**

$$\boxed{Min \sum_{t=1}^{n} \varepsilon_t^2 = Min \sum_{t=1}^{n} \left(y_t - a_0 - a_1 x_t\right)^2.}$$

La résolution analytique est la suivante :

$$\boxed{\widehat{a}_1 = \frac{\sum_{t=1}^{n}(x_t - \overline{x})(y_t - \overline{y})}{\sum_{t=1}^{n}(x_t - \overline{x})^2} = \frac{\sum_{t=1}^{n}(x_t y_t) - n\bar{x}\bar{y}}{\sum_{t=1}^{n}(x_t)^2 - n\overline{x}^2}}$$

$$\boxed{\widehat{a}_0 = \overline{y} - \widehat{a}_1 \overline{x}}$$

1.3.6 Propriétés des estimateurs

Écrivons

$$y_t = a_0 + a_1 x_t + \varepsilon_t,$$

et

$$\bar{y} = a_0 + a_1 \bar{x} + \bar{\varepsilon},$$

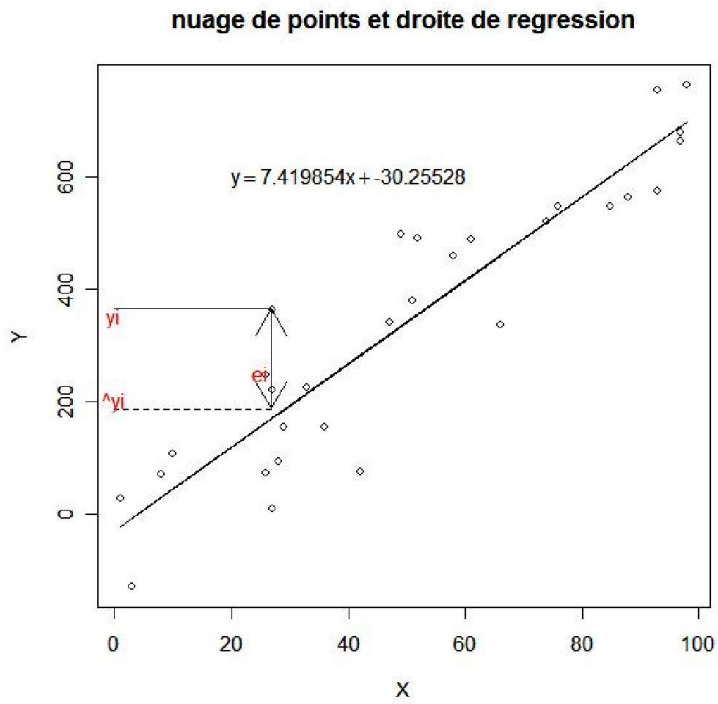


FIGURE 1.2 – Droite de régression

on obtient :

$$\widehat{a}_1 = a_1 + \frac{\sum_{t=1}^{n}(x_t - \overline{x})\varepsilon_t}{\sum_{t=1}^{n}(x_t - \overline{x})^2}.$$

On alors,

$$E(\widehat{a}_1) = E(a_1) + \frac{\sum_{t=1}^{n}(x_t - \overline{x})E(\varepsilon_t)}{\sum_{t=1}^{n}(x_t - \overline{x})^2} = E(a_1)$$

car $E(\varepsilon_t) = 0$.
De même on démontre que $E(\widehat{a}_0) = a_0$. Ce qui signifie que les estimateurs sont **sans biais**.
Puisque les estimateurs sont sans biais, il suffit pour qu'ils soient convergents que :

$$\lim_{n\to+\infty} V(\widehat{a}_1) = \lim_{n\to+\infty} V(\widehat{a}_0) = 0.$$

En effet :

$$V(\widehat{a}_1) = E\left(\widehat{a}_1 - E(\widehat{a}_1)\right)^2 = E\left(\widehat{a}_1 - a_1\right)^2 = E\left[\left(\frac{\sum_{t=1}^{n}(x_t - \overline{x})\varepsilon_t}{\sum_{t=1}^{n}(x_t - \overline{x})^2}\right)^2\right]$$

$$V(\widehat{a}_1) = E\left[\left(\sum_{t=1}^{n}\omega_t\varepsilon_t\right)^2\right] = \sum_{t=1}^{n}\omega_t^2 E(\varepsilon_t^2) + 2\sum_{t<t'}\omega_t\omega_{t'}E(\varepsilon_t\varepsilon_{t'}),$$

où on a posé $\omega_t = \frac{x_t - \overline{x}}{\sum_{t=1}^{n}(x_t - \overline{x})^2}$.
D'après les hypothèses $H4$ et $H5$, on obtient

$$V(\widehat{a}_1) = \sum_{t=1}^{n}(\omega_t^2\sigma_\varepsilon^2) = \frac{\sigma_\varepsilon^2}{\sum_{t=1}^{n}(x_t - \overline{x})^2}.$$

$$\boxed{V(\widehat{a}_1) = \frac{\sigma_\varepsilon^2}{\sum_{t=1}^n (x_t - \overline{x})^2} \to_{n\to+\infty} 0.}$$

Une démonstration analogue pour $\widehat{a}_0$ donne

$$\boxed{V(\widehat{a}_0) = \sigma_\varepsilon^2 \left(\frac{1}{n} + \frac{\overline{x}^2}{\sum_{t=1}^n (x_t - \overline{x})^2} \right) \to_{n\to+\infty} 0.}$$

Théorème de Gauss-Markov : Les estimateurs des MCO ont la plus petite variance parmi les estimateurs linéaires sans biais. On dit que ce sont des estimateurs **BLUE** (Best Linear Unbiased Estimator).

1.3.7 Construction des tests

Nous allons maintenant introduire de nouveau l'hypothèse qui est celle de la normalisé des erreurs. Cette hypothèse n'est pas indispensable afin d'obtenir des estimateurs convergents mais elle va nous permettre de construire des tests statistiques concernant la validité du modèle estimé.
Soit

$$\boxed{\varepsilon_t \rightsquigarrow \mathcal{N}(0, \sigma_\varepsilon^2).}$$

Cette hypothèse permet de définir la loi de probabilité des estimateurs.
L'estimateur de la variance de l'erreur ε noté $\widehat{\sigma}_\varepsilon^2$ est égal à :

$$\boxed{\widehat{\sigma}_\varepsilon^2 = \frac{1}{n-2} \sum_{t=1}^n e_t^2,}$$

où le **résidu** e_t est donné par

$$\boxed{e_t = y_t - \widehat{y}_t, \quad t = 1, \ldots, n.}$$

Ce qui nous permet de définir les estimateurs empiriques de la variance de chacun des coefficients :

$$\boxed{\widehat{\sigma}_{\widehat{a}_1}^2 = \frac{\widehat{\sigma}_\varepsilon^2}{\sum_{t=1}^n (x_t - \overline{x})^2}}$$

$$\boxed{\widehat{\sigma}_{\widehat{a}_0}^2 = \widehat{\sigma}_\varepsilon^2 \left(\frac{1}{n} + \frac{\overline{x}^2}{\sum_{t=1}^n (x_t - \overline{x})^2} \right).}$$

L'hypothèse de normalité des erreurs implique que :

$$\boxed{\frac{\widehat{a}_1 - a_1}{\sigma_{\widehat{a}_1}}, \quad \frac{\widehat{a}_0 - a_0}{\sigma_{\widehat{a}_0}} \rightsquigarrow \mathcal{N}(0,1)}$$

$$\boxed{\frac{\sum_{t=1}^n e_t^2}{\sigma_\varepsilon^2} = (n-2)\frac{\widehat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2} \rightsquigarrow \chi_{n-2}^2}$$

$$\boxed{\Rightarrow \frac{\widehat{a}_1 - a_1}{\widehat{\sigma}_{\widehat{a}_1}}, \quad \frac{\widehat{a}_0 - a_0}{\widehat{\sigma}_{\widehat{a}_0}} \rightsquigarrow \mathcal{T}_{n-2}}$$

Il est donc possible maintenant de mettre en place des tests statistiques afin d'apporter des réponses à des problèmes tels que :

- comparaison d'un coefficient de régression par rapport à une valeur fixée ;

- comparaison de deux coefficients de régression provenant de deux échantillons différents ;
- détermination d'un intervalle de confiance pour un coefficient.

1) Test bilatéral
Soit à tester, à un seuil de 5%, l'hypothèse $H_0 : a_1 = 0$ contre l'hypothèse $H_1 : a_1 \neq 0$.
Sous H_0, on a $\frac{\widehat{a}_1 - a_1}{\widehat{\sigma}_{\widehat{a}_1}} = \frac{\widehat{a}_1 - 0}{\widehat{\sigma}_{\widehat{a}_1}}$ suit une loi de Student à $n-2$ degrés de liberté.
Le test d'hypotheses bilatéral consiste donc à comparer le ratio de Student empirique $t^* = \frac{|\widehat{a}_1|}{\widehat{\sigma}_{\widehat{a}_1}}$ à la valeur du t de Student lue dans la table à $n-2$ degrés de liberté et pour un seuil de probabilité égal à 5%.
Si $n-2 > 30$, on a $t_\infty^{0.05} = 1.96$.
Si $t^* > t_\infty^{0.05}$ nous rejetons l'hypothèse H_0
$\Rightarrow$ le coefficient théorique inconnu a_1 est significativement différent de 0.
2) Test unilatéral
Soit à tester, à un seuil de 5%, l'hypothèse $H_0 : a_1 = 0$ contre l'hypothèse $H_1 : a_1 > 0$ ou $a_1 < 0$
Sous H_0, on a $\frac{\widehat{a}_1 - a_1}{\widehat{\sigma}_{\widehat{a}_1}} = \frac{\widehat{a}_1 - 0}{\widehat{\sigma}_{\widehat{a}_1}}$ suit une loi de Student à $n-2$ degrés de liberté.
Le test d'hypotheses unilatéral consiste donc à comparer le ratio de Student empirique $t^* = \frac{|\widehat{a}_1|}{\widehat{\sigma}_{\widehat{a}_1}}$ à la valeur du t de Student lue dans la table à $n-2$ degrés de liberté et pour un seuil de probabilité égal à 5%.
Si $n-2 > 30$, on a $t_\infty^{0.05} = 1.65$.
Si $t^* > t_\infty^{0.05}$ nous rejetons l'hypothèse H_0
$\Rightarrow$ le coefficient théorique inconnu a_1 est significativement différent de 0.
Remarque : Si nous rejetons l'hypothèse H_0 pour un test bilatéral, alors nous rejetons forcément (pour un même seuil de probabilité) l'hypothèse H_0 pour un test unilatéral.

Exemple 1.3.2 *On s'intéresse à la relation entre les bénéfices réalisés par les entreprises et le budget annuel qu'elles consacrent à la publicité.* 15 *observations ont été réalisées*

Budget	15	8	36	41	16	8	21	21	53	10	32	17	58	6	20
Bénéfices	48	43	77	89	50	40	56	62	100	47	71	58	102	35	60

Répondons aux questions suivantes

1. *Calculer les estimateurs $\widehat{a}_1$, $\widehat{a}_0$ et le coefficient de corrélation r*
2. *Sachant que $\widehat{\sigma}_\varepsilon^2 = 10,155$, procéder à l'estimation des variances de a_1 et a_0.*
3. *Déterminer au seuil de signification de 5%, un intervalle de confiance pour a_1, a_0 et $\widehat{\sigma}_\varepsilon^2$.*
4. *Peut-on affirmer que les coefficients a_1 et a_0 sont significativement différents de 0 pour 5% ?*

1. *Après calculs, on obtient $\widehat{a}_1 = 1,28$, $\widehat{a}_0 = 31,67$ et $r = 0,989$.*
2. *On a $n = 15$, $\bar{x} = 24,13$, $\sum_{t=1}^n (x_t - \overline{x})^2 = 15\sigma_X = 3753,733$ et $\widehat{\sigma}_\varepsilon^2 = 10,155$, donc :*

$$\widehat{\sigma}_{\widehat{a}_1}^2 = \frac{\widehat{\sigma}_\varepsilon^2}{\sum_{t=1}^n (x_t - \overline{x})^2} = 0,0027,$$

$$\widehat{\sigma}_{\widehat{a}_0}^2 = \widehat{\sigma}_\varepsilon^2 \left(\frac{1}{n} + \frac{\overline{x}^2}{\sum_{t=1}^n (x_t - \overline{x})^2} \right) = 2,252.$$

3. *Nous savons que $\frac{\widehat{a}_1 - a_1}{\widehat{\sigma}_{\widehat{a}_1}}$ et $\frac{\widehat{a}_0 - a_0}{\widehat{\sigma}_{\widehat{a}_0}}$ suivent la loi de Student $\mathcal{T}_{n-2}$.*
   L'intervalle de confiance pour a_1 et a_0 nous est donné respectivement par :

$$\boxed{\frac{\widehat{a}_1 - a_1}{\widehat{\sigma}_{\widehat{a}_1}} = \pm t_{n-2}^{\alpha/2} \Rightarrow a_1 = \widehat{a}_1 \pm \widehat{\sigma}_{\widehat{a}_1} \times t_{n-2}^{\alpha/2}}$$

$$\boxed{\frac{\widehat{a}_0 - a_0}{\widehat{\sigma}_{\widehat{a}_0}} = \pm t_{n-2}^{\alpha/2} \Rightarrow a_0 = \widehat{a}_0 \pm \widehat{\sigma}_{\widehat{a}_0} \times t_{n-2}^{\alpha/2}}$$

Avec $n-2 = 13$ degré de liberté et $\alpha/2 = 0,025$, on a $t_{n-2}^{\alpha/2} = 2,16$ lu dans le tableau de Student.
Donc, les intervalles de confiantes pour a_1 et a_0 sont respectivement :

$$\boxed{[\widehat{a}_1 - \widehat{\sigma}_{\widehat{a}_1} \times t_{n-2}^{\alpha/2}; \widehat{a}_1 + \widehat{\sigma}_{\widehat{a}_1} \times t_{n-2}^{\alpha/2}]} = [1,166; 1,391],$$

$$\boxed{[\widehat{a}_0 - \widehat{\sigma}_{\widehat{a}_0} \times t_{n-2}^{\alpha/2}; \widehat{a}_0 + \widehat{\sigma}_{\widehat{a}_0} \times t_{n-2}^{\alpha/2}]} = [28,432; 34,916]$$

Nous savons que $(n-2)\frac{\widehat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2} \rightsquigarrow \chi_{n-2}^2$.
L'intervalle de confiance pour $\widehat{\sigma}_\varepsilon^2$ est

$$\boxed{[(n-2)\frac{\widehat{\sigma}_\varepsilon^2}{\chi_{1-\alpha/2}^2}; (n-2)\frac{\widehat{\sigma}_\varepsilon^2}{\chi_{\alpha/2}^2}]}$$

Avec $n-2 = 13$ degré de liberté et $\alpha/2 = 0,025$, on a $\chi_{\alpha/2}^2 = 5,01$ et $\chi_{1-\alpha/2}^2 = 24,74$ lus dans le tableau de χ^2.
Donc

$$\widehat{\sigma}_\varepsilon^2 \in [5,336; 26,35].$$

4. *On procède à un test d'hypotheses bilatéral qui consiste donc à comparer les ratio de Student empiriques $t^* = \frac{|\widehat{a}_1|}{\widehat{\sigma}_{\widehat{a}_1}} = 24,63$ et*

$t^* = \frac{|\widehat{a}_0|}{\widehat{\sigma}_{\widehat{a}_0}} = 21,10$ à la valeur du $t_{n-2}^{\alpha/2} = 2,16$ de Student lue dans la table à $n-2$ degrés de liberté et pour un seuil de probabilité égal à 5%.
Puisque ces valeurs sont supérieurs en valeur absolu à 2,16, alors les deux paramètres a_1 et a_0 sont significativement différents de 0.
La variable exogène x contribue bien à expliquer la variable endogène y.

1.3.8 Équation et tableau d'analyse de la variance

On peut démontrer que

$$\sum_{t=1}^{n} e_t = 0,$$

et par suite, on trouve l'équation fondamentale d'**analyse de la variance** :

$$\boxed{\underbrace{\sum_{t=1}^{n}(y_t - \bar{y})^2}_{=SCT} = \underbrace{\sum_{t=1}^{n}(\widehat{y}_t - \bar{y})^2}_{=SCE} + \underbrace{\sum_{t=1}^{n} e_t^2}_{=SCR}.}$$

La variabilité totale (**SCT**) est égale à la variabilité expliquée (**SCE**) + la variabilité des résidus (**SCR**).
Cette équation va nous permettre de juger de la qualité de l'ajustement d'un modèle.
En effet, plus la variance expliquée est proche de la variance totale, meilleur est l'ajustement du nuage de points par la droite des moindres carrés.
Il est d'usage de calculer le rapport :

$$\boxed{R^2 = \frac{SCE}{SCT} = 1 - \frac{SCR}{SCT}.}$$

R^2 est appelé le coefficient de détermination, et R le coefficient de corrélation multiple (dans le cas particulier du modèle de régression à une seule variable explicative, il est égal au coefficient de corrélation linéaire simple entre x et y).

Variation	Somme des carrés	Degré de liberté	Carrés moyens
x	$SCE = \sum_{t=1}^{n}(\widehat{y}_t - \bar{y})^2$	1	$\frac{SCE}{1}$
Résidu	$SCR = \sum_{t=1}^{n} e_t^2$	n-2	$\frac{SCR}{n-2}$
Total	$SCT = \sum_{t=1}^{n}(y_t - \bar{y})^2$	$n-1$	

Le test $H_0 : a_1 = 0$ est équivalent au test d'hypothèse $H_0 : SCE = 0$ (la variable explicative x_t ne contribue pas à l'explication du modèle). La statistique de ce test est donnée par :

$$\boxed{F^* = \frac{\frac{SCE}{ddl_{SCE}}}{\frac{SCR}{ddl_{SCR}}} = \frac{\frac{R^2}{1}}{\frac{1-R^2}{n-2}}}$$

F^* suit une statistique de Fisher à 1 et $n-2$ degrés de liberté.
Si $F^* > F^{\alpha}_{1;n-2}$ nous rejetons au seuil α l'hypothèse H_0 et donc la variable x_t est significative.
Dans le cas contraire, nous acceptons l'hypothèse d'égalité des variances, donc la variable x_t n'est pas explicative de la variable y_t.

Remarque 1.3.1 *On peut montrer que*

$$\boxed{(t^*)^2 = F^*.}$$

Exemple 1.3.3 *On s'intéresse à la relation entre les tailles X_i en cm de certains tiges matériaux et leur poids Y_i en Kg. 10 observations ont été réalisées*

taille	150	175	200	225	250	275	300	325	350	375
Poids	18	24	26	23	30	27	34	35	33	40

Donner le tableau d'analyse de la variance associé à cette échantillon. Faire un test de Fisher à un seuil de 5%.
Réponse : *Nous commençons tout d'abord de faire les calculs nécessaires qui sont résumés dans le tableau suivant :*

X_i	Y_i	$(Y_i - \overline{Y})^2$	$\widehat{Y}_i$	$(\widehat{Y}_i - \overline{Y})^2$	$(Y_i - \widehat{Y}_i)^2$
150	18	121	19.84	83.90	3.38
175	24	25	21.87	50.83	4.53
200	26	9	23.91	25.90	4.36
225	23	36	25.95	9.30	8.70
250	30	1	27.98	1.04	4.08
275	27	4	30.02	1.04	9.12
300	34	25	32.05	9.30	3.8
325	35	36	34.09	25.90	0.82
350	33	16	36.13	50.83	9.79
375	40	121	38.16	83.90	3.38
Total		***394***		***341.94***	***51.96***

Traçons maintenant le tableau d'analyse de la variance :

Variation	*Somme des carrés*	*Degré de liberté*	*Carrés moyens*
x	$SCE = 341.94$	*1*	$\frac{SCE}{1} = 341.94$
Résidu	$SCR = 51.96$	*8*	$\frac{SCR}{8} = 6.5$
Total	$SCT = 394$	9	

Soit le test d'hypothèse H_0 : $SCE = 0$ contre H_1 : $SCE \neq 0$.
La statistique de ce test est donnée par :

$$F^* = \frac{\frac{SCE}{ddl_{SCE}}}{\frac{SCR}{ddl_{SCR}}} = \frac{341.94}{6.5} = 52.73.$$

Puisque $F^{\alpha}_{1;n-2} = F^{0.05}_{1;8} = 5.32$, alors $F^* > F^{\alpha}_{1;n-2}$.
Donc, nous rejetons au seuil α l'hypothèse H_0 et donc la variable explicative est significative.

1.3.9 La prévision dans le modèle de régression simple

Lorsque les coefficients du modèle ont été estimés, il est possible de calculer une prévision à un horizon h. Soit le modèle estimé sur la période $t = 1, \ldots, n$;

$$y_t = \widehat{a}_0 + \widehat{a}_1 x_t + e_t$$

Si la valeur de la variable explicative x_t est connue en $n+1$ (x_{n+1}), la prévision est donnée par :

$$\widehat{y}_{n+1} = \widehat{a}_0 + \widehat{a}_1 x_{n+1}$$

L'erreur de prévision est égale à :

$$e_{n+1} = y_{n+1} - \widehat{y}_{n+1} = (a_0 - \widehat{a}_0) + (a_1 - \widehat{a}_1)x_{n+1} + \varepsilon_{n+1}$$

En se référant aux hypothèses du modèle, on a :

$$\boxed{E(e_{n+1}) = 0}$$

$$\boxed{V(e_{n+1}) = \widehat{\sigma}^2_{\varepsilon}\left(\frac{1}{n} + \frac{(x_{n+1} - \overline{x})^2}{\sum_{t=1}^{n}(x_t - \overline{x})^2} + 1\right).}$$

L'hypothèse de normalité de ε_t permet alors de determiner un intervalle à $(1-\alpha)\%$ pour la prévision :

$$\boxed{e_{n+1} \rightsquigarrow \mathcal{N}\left(0, \widehat{\sigma}^2_{\varepsilon}\left(\frac{1}{n} + \frac{(x_{n+1} - \overline{x})^2}{\sum_{t=1}^{n}(x_t - \overline{x})^2} + 1\right)\right)}$$

$$\boxed{\Rightarrow \frac{\widehat{a}_0 + \widehat{a}_1 x_{n+1} - y_{n+1}}{\widehat{\sigma}_{\varepsilon}\sqrt{\left(\frac{1}{n} + \frac{(x_{n+1}-\overline{x})^2}{\sum_{t=1}^{n}(x_t-\overline{x})^2} + 1\right)}} \rightsquigarrow \mathcal{T}_{n-2}.}$$

On obtient alors l'intervalle de prédiction :

$$\boxed{y_{n+1} = \widehat{y}_{n+1} \pm t^{\alpha/2}_{n-2}\widehat{\sigma}_{\varepsilon}\sqrt{\frac{1}{n} + \frac{(x_{n+1} - \overline{x})^2}{\sum_{t=1}^{n}(x_t - \overline{x})^2} + 1.}}$$

Exemple 1.3.4 *A partir de l'exemple précédent, déterminer au seuil* 5%, *un intervalle de confiance pour le poids prévisible relatif à une taille de* $400cm$ *d'un tige.*
Réponse : *On a l'intervalle de prédiction $I_{Y_{40}}$ est donné par :*

$$\boxed{y_{n+1} = \widehat{y}_{n+1} \pm t^{\alpha/2}_{n-2}\widehat{\sigma}_{\varepsilon}\sqrt{\frac{1}{n} + \frac{(x_{n+1} - \overline{x})^2}{\sum_{t=1}^{n}(x_t - \overline{x})^2} + 1}}$$

où $x_{n+1} = 400$, $\overline{x} = 262.5$, $\sum_{t=1}^{n}(x_t - \overline{x})^2 = 51562.5$, $\widehat{y}_{n+1} = 40.02$, $t^{\alpha/2}_{n-2} = t^{0.025}_{8} = 2.306$, $\widehat{\sigma}^2_{\varepsilon} = \frac{SCR}{n-2} = 6.5$.
Donc,

$$I_{Y_{40}} = [21.86; 58.17].$$

Chapitre 2

Le modèle linéaire multiple

2.1 Introduction

Le modèle de régression linéaire multiple n'est qu'une extension du modèle de régression linéaire simple au cas multivarié dans lequel interviennent plusieurs variables exogènes dans l'explication du phénomène étudié.
On parle aussi de modèle de régression linéaire général ou standard pour souligner que ce modèle reste valable quel que soit le nombre d'exogènes qui s'y figurent.
Dans sa forme générale, il s'écrit de la sorte :

$$\boxed{y_t = a_0 + a_1 x_{1t} + \cdots + a_k x_{kt} + \varepsilon_t, \quad t = 1, 2, \ldots, n.}$$

Afin d'en alléger l'écriture et de faciliter l'expression de certains résultats, on a habituellement recours aux notations matricielles.

$$\boxed{\underbrace{Y}_{(n,1)} = \underbrace{X}_{(n,k+1)} \underbrace{a}_{(k+1,1)} + \underbrace{\varepsilon}_{(n,1)},}$$

où

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_t \\ \vdots \\ y_n \end{pmatrix}, X = \begin{pmatrix} 1 & x_{11} & x_{21} & \cdots & x_{k1} \\ 1 & x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_{1t} & x_{2t} & \cdots & x_{kt} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{kn} \end{pmatrix}, a = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_k \end{pmatrix}, \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_t \\ \vdots \\ \varepsilon_n \end{pmatrix}.$$

2.2 La méthode des moindres carrés

La méthode des moindres carrés cherche la meilleure estimation des paramètres a en minimisant la quantité :

$$\boxed{Min \sum_{t=1}^{n} \varepsilon_t^2 = Min \sum_{t=1}^{n} \varepsilon' \varepsilon = Min\,(Y - Xa)'\,(Y - Xa)\,.}$$

2.2.1 Les hypothèses de la méthode des MCOs

Hypothèses probabilistes (hypothèses stochastiques) :

- Les X sont observés sans erreur (non aléatoires)
- $E(\varepsilon) = 0$, en moyenne le modèle est bien spécifié
- $E(\varepsilon^2) = \sigma_\varepsilon^2$ la variance de l'erreur est constante (homoscédasticité)

- $i \neq j \Rightarrow E(\varepsilon_i, \varepsilon_j) = 0$ les erreurs sont non-corrélées (non-autocorrélation des erreurs)
- $Cov(\varepsilon, X) = 0$, l'erreur est indépendante des variables explicatives
- $\varepsilon \rightsquigarrow \mathcal{N}(0, \sigma_\varepsilon^2)$

Hypothèses structurelles :

- $Rang(X'X) = k+1$; $(X'X)^{-1}$ existe, ou encore $det(X'X) \neq 0$
- $(X'X)$ tend vers une matrice finie non singulière quand $n \to +\infty$
- $n > k+1$, le nombre d'observations est supérieur au nombre de paramètres du modèle (variables explicatives + constante)

2.2.2 Estimateurs des MCO

La résolution analytique par MCO est la suivante :

$$\boxed{\widehat{a} = (X'X)^{-1}X'Y}$$

D'une manière développée :

$$\begin{pmatrix} \sum y_t \\ \sum x_{1t}y_t \\ \sum x_{2t}y_t \\ \vdots \\ \vdots \\ \sum x_{kt}y_t \end{pmatrix} = \begin{pmatrix} n & \sum x_{1t} & \sum x_{2t} & \cdots & \sum x_{kt} \\ \sum x_{1t} & \sum x_{1t}^2 & \sum x_{1t}x_{2t} & \cdots & \sum x_{1t}x_{kt} \\ \sum x_{2t} & \sum x_{2t}x_{1t} & \sum x_{2t}^2 & \cdots & \sum x_{2t}x_{kt} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \sum x_{kt} & \sum x_{kt}x_{1t} & \sum x_{kt}x_{2t} & \cdots & \sum x_{kt}^2 \end{pmatrix} \begin{pmatrix} \widehat{a}_0 \\ \widehat{a}_1 \\ \widehat{a}_2 \\ \vdots \\ \widehat{a}_k \end{pmatrix}$$

2.2.3 Propriétés des estimateurs

Écrivons

$$\widehat{a} = (X'X)^{-1}X'Y = (X'X)^{-1}X'(Xa+\varepsilon) = a + (X'X)^{-1}X'\varepsilon,$$

alors,

$$E(\widehat{a}) = E(a) + E((X'X)^{-1}X'\varepsilon) = E(a) + (X'X)^{-1}X'E(\varepsilon) = E(a)$$

car $E(\varepsilon) = 0$.
Ce qui signifie que les estimateurs sont **sans biais**.
La matrice des variances et covariances des coefficients de régression qu'on note $\Omega_{\widehat{a}}$ est donnée par

$$\boxed{\Omega_{\widehat{a}} = \sigma_\varepsilon^2 (X'X)^{-1}}$$

$$\boxed{\Omega_{\widehat{a}} = \begin{pmatrix} Var(\widehat{a}_0) & Cov(\widehat{a}_0, \widehat{a}_1) & \cdots & Cov(\widehat{a}_0, \widehat{a}_k) \\ Cov(\widehat{a}_0, \widehat{a}_1) & Var(\widehat{a}_1) & \cdots & Cov(\widehat{a}_1, \widehat{a}_k) \\ \vdots & \vdots & \cdots & \vdots \\ Cov(\widehat{a}_k, \widehat{a}_0) & Cov(\widehat{a}_k, \widehat{a}_1) & \cdots & Var(\widehat{a}_k) \end{pmatrix}}$$

$$\boxed{\Rightarrow \lim \Omega_{\widehat{a}} = \lim \frac{\sigma_\varepsilon^2}{n} \left(\frac{X'X}{n}\right)^{-1} = 0.}$$

Théorème de Gauss-Markov : Les estimateurs des MCO ont la plus petite variance parmi les estimateurs linéaires sans biais.
Ce sont des estimateurs **BLUE** (Best Linear Unbiased Estimator).

Après un calcul matriciel, il apparaît que nous pouvons estimer sans biais σ_ε^2 par :

$$\boxed{\widehat{\sigma}_\varepsilon^2 = \frac{ee'}{n-k-1}}$$

avec $e_t = y_t - \widehat{y}_t$ est le résidu, c'est-à-dire l'écart entre la valeur observée de la variable à expliquer et sa valeur estimée (ajustée).
En remplaçant la variance de l'erreur par son estimateur, nous obtenons :

$$\boxed{\widehat{\Omega}_{\widehat{a}} = \widehat{\sigma}_\varepsilon^2 (X'X)^{-1}}$$

2.3 Équation d'analyse de la variance et qualité d'un ajustement

L'équation fondamentale d'**analyse de la variance** :

$$\boxed{\underbrace{\sum_{t=1}^{n}(y_t-\bar{y})^2}_{=SCT} = \underbrace{\sum_{t=1}^{n}(\widehat{y}_t-\bar{y})^2}_{=SCE} + \underbrace{\sum_{t=1}^{n} e_t^2}_{=SCR}}$$

va nous permettre de juger de la qualité de l'ajustement d'un modèle ; en effet, plus la variance expliquée est *proche* de la variance totale, meilleur est l'ajustement global du modèle. C'est pourquoi nous calculons le rapport

$$\boxed{R^2 = \frac{SCE}{SCT} = 1 - \frac{SCR}{SCT}}$$

appelé **le coefficient de détermination**, et R le **coefficient de corrélation multiple**. R^2 mesure la proportion de la variance de Y expliquée par la régression de Y sur X.

Exemple 2.3.1 *Soit le modèle $y_t = a_0 + a_1 x_{1t} + a_2 x_{2t} + a_3 x_{3t} + \varepsilon_t$, où :*

t	y	x_1	x_2	x_3
1	12	2	45	121
2	14	1	43	132
3	10	3	43	154
4	16	3	47	145
5	14	7	42	129
6	19	8	41	156
7	21	8	32	132
8	19	5	33	147
9	21	5	41	128
10	16	8	38	163
11	19	4	32	161
12	21	9	31	172
13	25	12	35	174
14	21	7	29	180

1. *Mettre le modèle sous forme matricielle en spécifiant bien les dimensions de chacune des matrices*
2. *Estimer les paramètres du modèle*
3. *Calculer l'estimation de la variance de l'erreur ainsi que les écarts types de chacun des coefficients.*
4. *Calculer le coefficient de détermination et commenter*

Réponses :

1) Forme matricielle : $Y = Xa + \varepsilon$, où

$$Y = \underbrace{\begin{pmatrix} 12 \\ 14 \\ \vdots \\ 10 \end{pmatrix}}_{(14,1)}, X = \underbrace{\begin{pmatrix} 1 & 2 & 45 & 121 \\ 1 & 1 & 43 & 132 \\ 1 & 3 & 43 & 154 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 7 & 29 & 180 \end{pmatrix}}_{(14,4)}, a = \underbrace{\begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{pmatrix}}_{(4,1)}, \varepsilon = \underbrace{\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_{14} \end{pmatrix}}_{(14,1)}.$$

2) Estimation des paramètres : Nous savons que $\widehat{a} = (X'X)^{-1}X'Y$. Donc, on doit calculer $X'X$ puis $(X'X)^{-1}$.

$$X'X = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 2 & 1 & 3 & \cdots & 7 \\ 45 & 43 & 43 & \cdots & 29 \\ 121 & 132 & 154 & \cdots & 180 \end{pmatrix} \begin{pmatrix} 1 & 2 & 45 & 121 \\ 1 & 1 & 43 & 132 \\ 1 & 3 & 43 & 154 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 7 & 29 & 180 \end{pmatrix}$$

$$X'X = \begin{pmatrix} 14 & 85 & 532 & 2094 \\ 85 & 631 & 3126 & 13132 \\ 532 & 3126 & 20666 & 78683 \\ 2094 & 13132 & 78683 & 317950 \end{pmatrix}$$

$$(X'X)^{-1} = \begin{pmatrix} 20,16864 & 0,015065 & -0,23145 & -0,07617 \\ 0,015065 & 0,013204 & 0,001194 & -0,00094 \\ -0,23145 & 0,001194 & 0,003635 & 0,000575 \\ -0,07617 & -0,00094 & 0,000575 & 0,000401 \end{pmatrix}.$$

On a

$$X'Y = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 2 & 1 & 3 & \cdots & 7 \\ 45 & 43 & 43 & \cdots & 29 \\ 121 & 132 & 154 & \cdots & 180 \end{pmatrix} \begin{pmatrix} 12 \\ 14 \\ \vdots \\ 10 \end{pmatrix} = \begin{pmatrix} 248 \\ 1622 \\ 9202 \\ 37592 \end{pmatrix}$$

Alors,

$$\widehat{a} = \begin{pmatrix} 20,16864 & 0,015065 & -0,23145 & -0,07617 \\ 0,015065 & 0,013204 & 0,001194 & -0,00094 \\ -0,23145 & 0,001194 & 0,003635 & 0,000575 \\ -0,07617 & -0,00094 & 0,000575 & 0,000401 \end{pmatrix} \begin{pmatrix} 248 \\ 1622 \\ 9202 \\ 37592 \end{pmatrix}$$

$$\widehat{a} = \begin{pmatrix} \widehat{a}_0 \\ \widehat{a}_1 \\ \widehat{a}_2 \\ \widehat{a}_3 \end{pmatrix} = \begin{pmatrix} 32,89132 \\ 0,801900 \\ -0,38136 \\ -0,03713 \end{pmatrix}$$

3) Calcul de $\widehat{\sigma}_\varepsilon$ et de $\widehat{\sigma}_{\widehat{a}}$. On sait que

$$\widehat{\sigma}_\varepsilon^2 = \frac{ee'}{n-k-1}.$$

Les composantes du résidu $e = Y - \widehat{Y} = Y - X\widehat{a}$ sont :

$$e_t = y_t - (a_0 + a_1 x_{1t} + a_2 x_{2t} + a_3 x_{3t})$$

$$e_t = y_t - 32,89 - 0,80x_{1t} + 0,38x_{2t} + 0,03x_{3t}.$$

Par exemple pour e_1 :

$$e_1 = y_1 - 32,89 - 0,80x_{11} + 0,38x_{21} + 0,03x_{31}$$

$$e_1 = 12 - 32,89 - 0,80 \times 2 + 0,38 \times 45 + 0,03 \times 121 = -0,84$$

Nous résumons dans le tableau suivant l'ensemble des résultats.

t	y_t	$\widehat{y}_t$	e_t	e_t^2
1	12	12,84	−0,84	0,71
2	14	12,39	1,61	2,58
3	10	13,18	−3,18	10,11
4	16	13,39	1,61	2,58
5	14	17,70	−3,70	13,67
6	19	17,88	1,12	1,26
7	21	22,20	−1,20	1,44
8	19	18,86	0,14	0,02
9	21	16,51	4,49	20,14
10	16	18,76	−2,76	7,63
11	19	17,92	1,08	1,17
12	21	21,90	−0,90	0,81
13	25	22,71	2,29	5,27
14	21	20,76	0,24	0,06
Somme			*0*	67,45

On a

$$\widehat{\sigma}_\varepsilon^2 = \frac{ee'}{n-k-1} = \frac{\sum_{t=1}^{14} e_t^2}{14-3-1} = \frac{67,45}{10} = 6,745.$$

$$\Rightarrow \widehat{\Omega}_{\widehat{a}} = \widehat{\sigma}_\varepsilon^2 (X'X)^{-1}$$

$$\widehat{\Omega}_{\widehat{a}} = 6,745 \times \begin{pmatrix} 20,16864 & 0,015065 & -0,23145 & -0,07617 \\ 0,015065 & 0,013204 & 0,001194 & -0,00094 \\ -0,23145 & 0,001194 & 0,003635 & 0,000575 \\ -0,07617 & -0,00094 & 0,000575 & 0,000401 \end{pmatrix}.$$

Les variances des coefficients de régression se trouvent sur la première diagonale :

$$\widehat{\sigma}_{\widehat{a}_0}^2 = 6,745 \times 20,17 = 136,04 \Rightarrow \widehat{\sigma}_{\widehat{a}_0} = 11,66,$$

$$\widehat{\sigma}_{\widehat{a}_1}^2 = 6,745 \times 0,013 = 0,087 \Rightarrow \widehat{\sigma}_{\widehat{a}_1} = 0,29,$$

$$\widehat{\sigma}_{\widehat{a}_2}^2 = 6,745 \times 0,0036 = 0,024 \Rightarrow \widehat{\sigma}_{\widehat{a}_2} = 0,15,$$

$$\widehat{\sigma}_{\widehat{a}_3}^2 = 6,745 \times 0,0004 = 0,0026 \Rightarrow \widehat{\sigma}_{\widehat{a}_3} = 0,05.$$

4) Le coefficient de détermination R^2 est

$$R^2 = \frac{SCE}{SCT} = 1 - \frac{SCR}{SCT} = 1 - \frac{\sum_{t=1}^{14} e_t^2}{\sum_{t=1}^{14} (y_t - \overline{y})^2}.$$

Nous avons $ee' = \sum_{t=1}^{14} e_t^2 = 67,45$ et $\sum_{t=1}^{14} (y_t - \overline{y})^2 = 226,86$, donc

$$R^2 = 1 - \frac{67,45}{226,86} = 0,702.$$

Puisque $R^2 > 0,5$ l'ajustement global du modèle linéaire est bon.

2.4 Construction des tests

Nous pouvons mettre en place un certain nombre de tests statistiques que nous allons expliciter.
1) Comparaison d'un paramètre a_i à une valeur fixée $\overline{a}$

$$H_0 : a_i = \overline{a}, \quad \text{contre} \quad H_1 : a_i \neq \overline{a}.$$

- Si $\boxed{t_{\widehat{a}_i} > t_{n-k-1}^{\alpha/2}}$ nous rejetons l'hypothèse H_0 et alors a_i est significativement different de $\overline{a}$ (au seuil de α).
- Si $\boxed{t_{\widehat{a}_i} \leq t_{n-k-1}^{\alpha/2}}$ nous acceptons l'hypothèse H_0 et alors a_i est n'est pas significativement different de $\overline{a}$ (au seuil de α).

2) Comparaison d'un ensemble de paramètres à un ensemble de valeurs fixées :
Nous cherchons à tester simultanément l'égalité d'un sous-ensemble de coefficients de régression à des valeurs fixées.

$$H_0 : a_q = \overline{a}_q, \quad \text{contre} \quad H_1 : a_q \neq \overline{a}_q,$$

où q étant le nombre de coefficients retenus.
Pour accepter H_0, il suffit que :

$$\boxed{\frac{1}{q}\left(\widehat{a}_q - \overline{a}_q\right)' \widehat{\Omega}_{\widehat{a}_q}^{-1} \left(\widehat{a}_q - \overline{a}_q\right) \leq F^{\alpha}(q, n-k-1).}$$

$F^{\alpha}(q, n-k-1)$ est loi de Fisher au seuil α à q et $n-k-1$ degrés de liberté.
3) Intervalle de confiance de la variance de l'erreur :
L'intervalle de confiance de la variance de l'erreur permet de déterminer une fourchette de variation de l'amplitude de l'erreur.
Pour un intervalle à $(1-\alpha\%)$, il est donné par :

$$\boxed{IC = \left[\frac{(n-k-1)\widehat{\sigma}_{\varepsilon}^2}{\chi_1^2} ; \frac{(n-k-1)\widehat{\sigma}_{\varepsilon}^2}{\chi_2^2}\right]}$$

avec χ_1^2 à $n-k-1$ degrés de liberté et $\alpha/2$ de probabilité d'être dépassée etχ_2^2 à $n-k-1$ degrés de liberté et $1-\alpha/2$ de probabilité d'être dépassée.

Exemple 2.4.1 *En reprenant les données du tableau de l'exemple précédent, on demande de répondre aux questions suivantes :*

1. *Les variables explicatives sont-elles significativement contributives pour expliquer la variable endogène ?*
2. *Le coefficient a_1 est-il significativement inférieur à 1 ?*
3. *Les coefficients a_1 et a_2 sont-ils simultanément et significativement différents de 1 et $-0,5$?*
4. *Quel est l'intervalle de confiance pour la variance de l'erreur ?*

(Les seuils choisis seront de 5%.)
Réponses :
1) Il convient de calculer les trois ratios de Student et de les comparer à la valeur lue dans la table pour un seuil de 5% :

$$t_{\widehat{a}_1}^* = \frac{|\widehat{a}_1|}{\widehat{\sigma}_{\widehat{a}_1}} = \frac{0,80}{0,29} = 2,75 > 2,228 = t_{10}^{0,05} \Rightarrow a_1 \neq 0.$$

Donc, la variable explicative x_1 est contributive à l'explication de y.

$$t_{\widehat{a}_2}^* = \frac{|\widehat{a}_2|}{\widehat{\sigma}_{\widehat{a}_2}} = \frac{|-0,38|}{0,15} = 2,53 > t_{10}^{0,05} \Rightarrow a_2 \neq 0,$$

$$t_{\widehat{a}_3}^* = \frac{|\widehat{a}_3|}{\widehat{\sigma}_{\widehat{a}_3}} = \frac{|-0,03|}{0,05} = 0,60 < t_{10}^{0,05} \Rightarrow a_3 = 0.$$

La variable x_2 est explicative de y alors que la variable x_3 n'est pas contributive à l'explication de y, il convient donc de la retirer de ce modèle et de procéder à une nouvelle estimation.
Nous aurions pu tout aussi bien répondre à cette question en calculant les intervalles de confiance de chacun des coefficients :

$$\boxed{IC_{a_1} = [\widehat{a}_1 - \widehat{\sigma}_{\widehat{a}_1} \times t_{n-k-1}^{\alpha/2}; \widehat{a}_1 + \widehat{\sigma}_{\widehat{a}_1} \times t_{n-k-1}^{\alpha/2}]} = [0,14; 1,45].$$

De même, nous obtenons :

$$IC_{a_2} = [-0,71; -0,04],$$

$$IC_{a_3} = [-0,14; 0,08].$$

La valeur 0 *n'appartient pas à l'intervalle de confiance à* 95% *de a_1 et a_2, donc ces deux coefficients sont significativement différents de* 0 *; en revanche,* 0 *appartient à l'intervalle de confiance de a_3, ce coefficient n'est pas significativement différent de* 0.
2) Nous posons le test d'hypothèses unilatéral suivant :

$$H_0 : a_1 = 1, \quad contre \quad H_1 : a_1 < 1.$$

Sous H_0, nous avons

$$\frac{\widehat{a}_1 - a_1}{\widehat{\sigma}_{\widehat{a}_1}} = \frac{0,80-1}{0,29} = -0,68 > -1,81 = -t_{10}^{0.05}$$

$$\Rightarrow \quad Acceptation\ de \quad H_0.$$

3) Le test d'hypothèses est le suivant :

$$H_0 : \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 1 \\ -0,5 \end{pmatrix}, \quad contre \quad H_1 : \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \neq \begin{pmatrix} 1 \\ -0,5 \end{pmatrix}.$$

Calculons $F^* = \frac{1}{q}\left(\widehat{a}_q - \overline{a}_q\right)' \widehat{\Omega}_{\widehat{a}_q}^{-1} \left(\widehat{a}_q - \overline{a}_q\right)$,
où $q = 2$, $\widehat{a}_q = \begin{pmatrix} 0,80 \\ -0,38 \end{pmatrix}$, $a_q = \begin{pmatrix} 1 \\ -0,5 \end{pmatrix}$ *et*

$$\widehat{\Omega}_{\widehat{a}_q} = 6,745. \begin{pmatrix} 0,013204 & 0,001194 \\ 0,001194 & 0,003635 \end{pmatrix}$$

$$\Rightarrow \widehat{\Omega}_{\widehat{a}_q}^{-1} = \begin{pmatrix} 11,57140 & -3,80213 \\ -3,80213 & 42,03506 \end{pmatrix}.$$

Donc,

$$F^* = \frac{1}{2}(0,8-1; -3,8+0,5) \begin{pmatrix} 11,57140 & -3,80213 \\ -3,80213 & 42,03506 \end{pmatrix} \begin{pmatrix} 0,8-1 \\ -3,8+0,5 \end{pmatrix}$$

$$\Rightarrow F^* = 0,612 < 4,10 = F^{\alpha}(q, n-k-1) = F_{2;10}^{0,05}.$$

Et par suite, on accepte l'hypothèse H_0.
Les données ne sont pas incompatibles avec la possibilité que les coefficients a_1 et a_2 soient simultanément et respectivement égaux à 1 *et* $-0,5$.
4) L'intervalle de confiance de la variance de l'erreur à un seuil $(1-\alpha)\% = 95\%$ *est calcule à partir de la formule*

$$IC = \left[\frac{(n-k-1)\widehat{\sigma}_{\varepsilon}^2}{\chi_1^2}; \frac{(n-k-1)\widehat{\sigma}_{\varepsilon}^2}{\chi_2^2}\right] = \left[\frac{10 \times 6,745}{\chi_{0,025}^2}; \frac{10 \times 6,745}{\chi_{0,975}^2}\right]$$

pour 10 *degrés de liberté.*
Soit $3,30 \leq \sigma_{\varepsilon}^2 \leq 20,75$.
La variance vraie (mais inconnue) σ_{ε}^2 de l'erreur à 95% *de chance de se situer dans cet intervalle.*

2.4.1 Construction du tableau d'analyse de la variance et test de signification globale d'une régression

Dans cette section, nous allons nous interroger sur la signification globale du modèle de régression, c'est-à-dire si l'ensemble des variables explicatives a une influence sur la variable à expliquer.
Ce test peut être formulé de la manière suivante : existe−t−il au moins une variable explicative significative ?
Soit le test d'hypothèses :

$$\begin{aligned} H_0 &: \quad a_1 = a_2 = \cdots = a_k = 0 \\ H_1 &: \quad \text{Il existe au moins } a_i \neq 0 \end{aligned}$$

Le cas où l'hypothèse H_0 est acceptée signifie qu'il n'existe aucune relation linéaire significative entre la variable à expliquer et les variables explicatives (ou encore que la Somme des Carrés Expliqués n'est pas significativement différente de 0).
Nous reprenons l'équation fondamentale d'analyse de la variance :

$$\underbrace{\sum_{t=1}^{n}(y_t-\bar{y})^2}_{=SCT} = \underbrace{\sum_{t=1}^{n}(\widehat{y}_t-\bar{y})^2}_{=SCE} + \underbrace{\sum_{t=1}^{n} e_t^2}_{=SCR}.$$

Nous traçons le tableau d'analyse de la variance permettant d'effectuer le test de Fisher. Soit

$$\boxed{F^* = \frac{\sum_{t=1}^{n}(\widehat{y}_t-\bar{y})^2/k}{\sum_{t=1}^{n} e_t^2/(n-k-1)} = \frac{R^2/k}{(1-R^2)/(n-k-1)}}$$

Variation	SC	DDL	CM
$x_1,\ldots,x_k$	SCE	k	$\frac{SCE}{k}$
Résidu	SCR	$n-k-1$	$\frac{SCR}{n-k-1}$
Total	SCT	$n-1$	

L'hypothese de normalité des erreurs implique que sous H_0, F^* suit une loi de Fisher (rapport de deux chi-deux).
Si $F^* > F^{\alpha}(k, n-k-1)$, nous rejetons H_0 et le modèle est globalement explicatif.

Exemple 2.4.2 *Tester la significativité globale du modèle vu dans l'exemple précédent.*
Réponse :
Le tableau d'analyse de la variance permettant d'effectuer le test de Fisher est :

Variation	*SC*	*DDL*	*CM*
x_1, x_2, x_3	$SCE = 159,41$	*3*	53,13
Résidu	$SCR = 67,45$	10	6,745
Total	$SCT = 226,86$	13	

On a $F^* = \frac{SCE/3}{SCR/10} = 7,87$ *et* $F^{0,95}_{(3;10)} = 3,71$.
Puisque $F^* > F^{0,95}_{(3;10)}$, *nous rejetons* H_0 *et le modèle est globalement explicatif.*

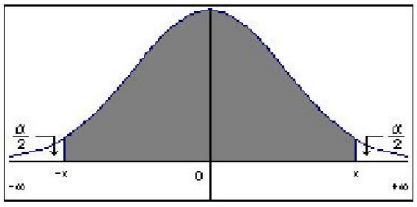
Tables Statistiques usuelles

Loi Normale Centrée Réduite

Exemple : P(Z<1.96)= 0.97500 se trouve en ligne 1.9 et colonne 0.06

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,50000	0,50399	0,50798	0,51197	0,51595	0,51994	0,52392	0,52790	0,53188	0,53586
0,1	0,53983	0,54380	0,54776	0,55172	0,55567	0,55962	0,56356	0,56750	0,57142	0,57535
0,2	0,57926	0,58317	0,58706	0,59095	0,59484	0,59871	0,60257	0,60642	0,61026	0,61409
0,3	0,61791	0,62172	0,62552	0,62930	0,63307	0,63683	0,64058	0,64431	0,64803	0,65173
0,4	0,65542	0,65910	0,66276	0,66640	0,67003	0,67365	0,67724	0,68082	0,68439	0,68793
0,5	0,69146	0,69498	0,69847	0,70194	0,70540	0,70884	0,71226	0,71566	0,71904	0,72241
0,6	0,72575	0,72907	0,73237	0,73565	0,73891	0,74215	0,74537	0,74857	0,75175	0,75490
0,7	0,75804	0,76115	0,76424	0,76731	0,77035	0,77337	0,77637	0,77935	0,78231	0,78524
0,8	0,78815	0,79103	0,79389	0,79673	0,79955	0,80234	0,80511	0,80785	0,81057	0,81327
0,9	0,81594	0,81859	0,82121	0,82382	0,82639	0,82894	0,83147	0,83398	0,83646	0,83891
1,0	0,84135	0,84375	0,84614	0,84850	0,85083	0,85314	0,85543	0,85769	0,85993	0,86214
1,1	0,86433	0,86650	0,86864	0,87076	0,87286	0,87493	0,87698	0,87900	0,88100	0,88298
1,2	0,88493	0,88686	0,88877	0,89065	0,89251	0,89435	0,89617	0,89796	0,89973	0,90148
1,3	0,90320	0,90490	0,90658	0,90824	0,90988	0,91149	0,91309	0,91466	0,91621	0,91774
1,4	0,91924	0,92073	0,92220	0,92364	0,92507	0,92647	0,92786	0,92922	0,93056	0,93189
1,5	0,93319	0,93448	0,93575	0,93699	0,93822	0,93943	0,94062	0,94179	0,94295	0,94408
1,6	0,94520	0,94630	0,94738	0,94845	0,94950	0,95053	0,95154	0,95254	0,95352	0,95449
1,7	0,95544	0,95637	0,95728	0,95819	0,95907	0,95994	0,96080	0,96164	0,96246	0,96327
1,8	0,96407	0,96485	0,96562	0,96638	0,96712	0,96784	0,96856	0,96926	0,96995	0,97062
1,9	0,97128	0,97193	0,97257	0,97320	0,97381	0,97441	0,97500	0,97558	0,97615	0,97670
2,0	0,97725	0,97778	0,97831	0,97882	0,97933	0,97982	0,98030	0,98077	0,98124	0,98169
2,1	0,98214	0,98257	0,98300	0,98341	0,98382	0,98422	0,98461	0,98500	0,98537	0,98574
2,2	0,98610	0,98645	0,98679	0,98713	0,98745	0,98778	0,98809	0,98840	0,98870	0,98899
2,3	0,98928	0,98956	0,98983	0,99010	0,99036	0,99061	0,99086	0,99111	0,99134	0,99158
2,4	0,99180	0,99202	0,99224	0,99245	0,99266	0,99286	0,99305	0,99324	0,99343	0,99361
2,5	0,99379	0,99396	0,99413	0,99430	0,99446	0,99461	0,99477	0,99492	0,99506	0,99520
2,6	0,99534	0,99547	0,99560	0,99573	0,99585	0,99598	0,99609	0,99621	0,99632	0,99643
2,7	0,99653	0,99664	0,99674	0,99683	0,99693	0,99702	0,99711	0,99720	0,99728	0,99736
2,8	0,99744	0,99752	0,99760	0,99767	0,99774	0,99781	0,99788	0,99795	0,99801	0,99807
2,9	0,99813	0,99819	0,99825	0,99831	0,99836	0,99841	0,99846	0,99851	0,99856	0,99861
3,0	0,99865	0,99869	0,99874	0,99878	0,99882	0,99886	0,99889	0,99893	0,99897	0,99900
3,1	0,99903	0,99906	0,99910	0,99913	0,99916	0,99918	0,99921	0,99924	0,99926	0,99929
3,2	0,99931	0,99934	0,99936	0,99938	0,99940	0,99942	0,99944	0,99946	0,99948	0,99950
3,3	0,99952	0,99953	0,99955	0,99957	0,99958	0,99960	0,99961	0,99962	0,99964	0,99965
3,4	0,99966	0,99968	0,99969	0,99970	0,99971	0,99972	0,99973	0,99974	0,99975	0,99976

Loi de Student



α	1	0,8	0,6	0,4	0,2	0,1	0,05	0,02	0,01	0,002	0,001
1 − α	0	0,2	0,4	0,6	0,8	0,9	0,95	0,98	0,99	0,998	0,999
ν = ddl											
1	0,0000	0,3249	0,7265	1,3764	3,0777	6,3137	12,706	31,821	63,656	318,29	636,58
2	0,0000	0,2887	0,6172	1,0607	1,8856	2,9200	4,3027	6,9645	9,9250	22,328	31,600
3	0,0000	0,2767	0,5844	0,9785	1,6377	2,3534	3,1824	4,5407	5,8408	10,214	12,924
4	0,0000	0,2707	0,5686	0,9410	1,5332	2,1318	2,7765	3,7469	4,6041	7,1729	8,6101
5	0,0000	0,2672	0,5594	0,9195	1,4759	2,0150	2,5706	3,3649	4,0321	5,8935	6,8685
6	0,0000	0,2648	0,5534	0,9057	1,4398	1,9432	2,4469	3,1427	3,7074	5,2075	5,9587
7	0,0000	0,2632	0,5491	0,8960	1,4149	1,8946	2,3646	2,9979	3,4995	4,7853	5,4081
8	0,0000	0,2619	0,5459	0,8889	1,3968	1,8595	2,3060	2,8965	3,3554	4,5008	5,0414
9	0,0000	0,2610	0,5435	0,8834	1,3830	1,8331	2,2622	2,8214	3,2498	4,2969	4,7809
10	0,0000	0,2602	0,5415	0,8791	1,3722	1,8125	2,2281	2,7638	3,1693	4,1437	4,5868
11	0,0000	0,2596	0,5399	0,8755	1,3634	1,7959	2,2010	2,7181	3,1058	4,0248	4,4369
12	0,0000	0,2590	0,5386	0,8726	1,3562	1,7823	2,1788	2,6810	3,0545	3,9296	4,3178
13	0,0000	0,2586	0,5375	0,8702	1,3502	1,7709	2,1604	2,6503	3,0123	3,8520	4,2209
14	0,0000	0,2582	0,5366	0,8681	1,3450	1,7613	2,1448	2,6245	2,9768	3,7874	4,1403
15	0,0000	0,2579	0,5357	0,8662	1,3406	1,7531	2,1315	2,6025	2,9467	3,7329	4,0728
16	0,0000	0,2576	0,5350	0,8647	1,3368	1,7459	2,1199	2,5835	2,9208	3,6861	4,0149
17	0,0000	0,2573	0,5344	0,8633	1,3334	1,7396	2,1098	2,5669	2,8982	3,6458	3,9651
18	0,0000	0,2571	0,5338	0,8620	1,3304	1,7341	2,1009	2,5524	2,8784	3,6105	3,9217
19	0,0000	0,2569	0,5333	0,8610	1,3277	1,7291	2,0930	2,5395	2,8609	3,5793	3,8833
20	0,0000	0,2567	0,5329	0,8600	1,3253	1,7247	2,0860	2,5280	2,8453	3,5518	3,8496
21	0,0000	0,2566	0,5325	0,8591	1,3232	1,7207	2,0796	2,5176	2,8314	3,5271	3,8193
22	0,0000	0,2564	0,5321	0,8583	1,3212	1,7171	2,0739	2,5083	2,8188	3,5050	3,7922
23	0,0000	0,2563	0,5317	0,8575	1,3195	1,7139	2,0687	2,4999	2,8073	3,4850	3,7676
24	0,0000	0,2562	0,5314	0,8569	1,3178	1,7109	2,0639	2,4922	2,7970	3,4668	3,7454
25	0,0000	0,2561	0,5312	0,8562	1,3163	1,7081	2,0595	2,4851	2,7874	3,4502	3,7251
26	0,0000	0,2560	0,5309	0,8557	1,3150	1,7056	2,0555	2,4786	2,7787	3,4350	3,7067
27	0,0000	0,2559	0,5306	0,8551	1,3137	1,7033	2,0518	2,4727	2,7707	3,4210	3,6895
28	0,0000	0,2558	0,5304	0,8546	1,3125	1,7011	2,0484	2,4671	2,7633	3,4082	3,6739
29	0,0000	0,2557	0,5302	0,8542	1,3114	1,6991	2,0452	2,4620	2,7564	3,3963	3,6595
30	0,0000	0,2556	0,5300	0,8538	1,3104	1,6973	2,0423	2,4573	2,7500	3,3852	3,6460
40	0,0000	0,2550	0,5286	0,8507	1,3031	1,6839	2,0211	2,4233	2,7045	3,3069	3,5510
50	0,0000	0,2547	0,5278	0,8489	1,2987	1,6759	2,0086	2,4033	2,6778	3,2614	3,4960
60	0,0000	0,2545	0,5272	0,8477	1,2958	1,6706	2,0003	2,3901	2,6603	3,2317	3,4602
70	0,0000	0,2543	0,5268	0,8468	1,2938	1,6669	1,9944	2,3808	2,6479	3,2108	3,4350
80	0,0000	0,2542	0,5265	0,8461	1,2922	1,6641	1,9901	2,3739	2,6387	3,1952	3,4164
90	0,0000	0,2541	0,5263	0,8456	1,2910	1,6620	1,9867	2,3685	2,6316	3,1832	3,4019
100	0,0000	0,2540	0,5261	0,8452	1,2901	1,6602	1,9840	2,3642	2,6259	3,1738	3,3905
200	0,0000	0,2537	0,5252	0,8434	1,2858	1,6525	1,9719	2,3451	2,6006	3,1315	3,3398
∞	0,0000	0,2533	0,5244	0,8416	1,2816	1,6449	1,9600	2,3263	2,5758	3,0903	3,2906

Loi de χ^2

$1-\alpha$	0,001	0,005	0,01	0,025	0,05	0,1	0,5	0,9	0,95	0,975	0,99	0,995	0,999
α	0,999	0,995	0,99	0,975	0,95	0,9	0,5	0,1	0,05	0,025	0,01	0,005	0,001
ν = ddl													
1	0,00	0,00	0,00	0,00	0,00	0,02	0,45	2,71	3,84	5,02	6,63	7,88	10,83
2	0,00	0,01	0,02	0,05	0,10	0,21	1,39	4,61	5,99	7,38	9,21	10,60	13,82
3	0,02	0,07	0,11	0,22	0,35	0,58	2,37	6,25	7,81	9,35	11,34	12,84	16,27
4	0,09	0,21	0,30	0,48	0,71	1,06	3,36	7,78	9,49	11,14	13,28	14,86	18,47
5	0,21	0,41	0,55	0,83	1,15	1,61	4,35	9,24	11,07	12,83	15,09	16,75	20,51
6	0,38	0,68	0,87	1,24	1,64	2,20	5,35	10,64	12,59	14,45	16,81	18,55	22,46
7	0,60	0,99	1,24	1,69	2,17	2,83	6,35	12,02	14,07	16,01	18,48	20,28	24,32
8	0,86	1,34	1,65	2,18	2,73	3,49	7,34	13,36	15,51	17,53	20,09	21,95	26,12
9	1,15	1,73	2,09	2,70	3,33	4,17	8,34	14,68	16,92	19,02	21,67	23,59	27,88
10	1,48	2,16	2,56	3,25	3,94	4,87	9,34	15,99	18,31	20,48	23,21	25,19	29,59
11	1,83	2,60	3,05	3,82	4,57	5,58	10,34	17,28	19,68	21,92	24,73	26,76	31,26
12	2,21	3,07	3,57	4,40	5,23	6,30	11,34	18,55	21,03	23,34	26,22	28,30	32,91
13	2,62	3,57	4,11	5,01	5,89	7,04	12,34	19,81	22,36	24,74	27,69	29,82	34,53
14	3,04	4,07	4,66	5,63	6,57	7,79	13,34	21,06	23,68	26,12	29,14	31,32	36,12
15	3,48	4,60	5,23	6,26	7,26	8,55	14,34	22,31	25,00	27,49	30,58	32,80	37,70
16	3,94	5,14	5,81	6,91	7,96	9,31	15,34	23,54	26,30	28,85	32,00	34,27	39,25
17	4,42	5,70	6,41	7,56	8,67	10,09	16,34	24,77	27,59	30,19	33,41	35,72	40,79
18	4,90	6,26	7,01	8,23	9,39	10,86	17,34	25,99	28,87	31,53	34,81	37,16	42,31
19	5,41	6,84	7,63	8,91	10,12	11,65	18,34	27,20	30,14	32,85	36,19	38,58	43,82
20	5,92	7,43	8,26	9,59	10,85	12,44	19,34	28,41	31,41	34,17	37,57	40,00	45,31
21	6,45	8,03	8,90	10,28	11,59	13,24	20,34	29,62	32,67	35,48	38,93	41,40	46,80
22	6,98	8,64	9,54	10,98	12,34	14,04	21,34	30,81	33,92	36,78	40,29	42,80	48,27
23	7,53	9,26	10,20	11,69	13,09	14,85	22,34	32,01	35,17	38,08	41,64	44,18	49,73
24	8,08	9,89	10,86	12,40	13,85	15,66	23,34	33,20	36,42	39,36	42,98	45,56	51,18
25	8,65	10,52	11,52	13,12	14,61	16,47	24,34	34,38	37,65	40,65	44,31	46,93	52,62
26	9,22	11,16	12,20	13,84	15,38	17,29	25,34	35,56	38,89	41,92	45,64	48,29	54,05
27	9,80	11,81	12,88	14,57	16,15	18,11	26,34	36,74	40,11	43,19	46,96	49,65	55,48
28	10,39	12,46	13,56	15,31	16,93	18,94	27,34	37,92	41,34	44,46	48,28	50,99	56,89
29	10,99	13,12	14,26	16,05	17,71	19,77	28,34	39,09	42,56	45,72	49,59	52,34	58,30
30	11,59	13,79	14,95	16,79	18,49	20,60	29,34	40,26	43,77	46,98	50,89	53,67	59,70

Loi de Fisher

$\alpha = 0{,}975$

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	15	20	30	50	100	200	500	•
1	648	800	864	900	922	937	948	957	963	969	985	993	1001	1008	1013	1016	1017	1018
2	38,5	39,0	39,2	39,3	39,3	39,4	39,4	39,4	39,4	39,4	39,4	39,4	39,5	39,5	39,5	39,5	39,5	39,5
3	17,4	16,0	15,4	15,1	14,9	14,7	14,6	14,5	14,5	14,4	14,3	14,2	14,1	14,0	14,0	13,9	13,9	13,9
4	12,2	10,6	9,98	9,60	9,36	9,20	9,07	8,98	8,90	8,84	8,66	8,56	8,46	8,38	8,32	8,29	8,27	8,26
5	10,0	8,43	7,76	7,39	7,15	6,98	6,85	6,76	6,68	6,62	6,43	6,33	6,23	6,14	6,08	6,05	6,03	6,02
6	8,81	7,26	6,60	6,23	5,99	5,82	5,70	5,60	5,52	5,46	5,27	5,17	5,07	4,98	4,92	4,88	4,86	4,85
7	8,07	6,54	5,89	5,52	5,29	5,12	4,99	4,90	4,82	4,76	4,57	4,47	4,36	4,28	4,21	4,18	4,16	4,14
8	7,57	6,06	5,42	5,05	4,82	4,65	4,53	4,43	4,36	4,30	4,10	4,00	3,89	3,81	3,74	3,70	3,68	3,67
9	7,21	5,71	5,08	4,72	4,48	4,32	4,20	4,10	4,03	3,96	3,77	3,67	3,56	3,47	3,40	3,37	3,35	3,33
10	6,94	5,46	4,83	4,47	4,24	4,07	3,95	3,85	3,78	3,72	3,52	3,42	3,31	3,22	3,15	3,12	3,09	3,08
11	6,72	5,26	4,63	4,28	4,04	3,88	3,76	3,66	3,59	3,53	3,33	3,23	3,12	3,03	2,96	2,92	2,90	2,88
12	6,55	5,10	4,47	4,12	3,89	3,73	3,61	3,51	3,44	3,37	3,18	3,07	2,96	2,87	2,80	2,76	2,74	2,72
13	6,41	4,97	4,35	4,00	3,77	3,60	3,48	3,39	3,31	3,25	3,05	2,95	2,84	2,74	2,67	2,63	2,61	2,60
14	6,30	4,86	4,24	3,89	3,66	3,50	3,38	3,29	3,21	3,15	2,95	2,84	2,73	2,64	2,56	2,53	2,50	2,49
15	6,20	4,76	4,15	3,80	3,58	3,41	3,29	3,20	3,12	3,06	2,86	2,76	2,64	2,55	2,47	2,44	2,41	2,40
16	6,12	4,69	4,08	3,73	3,50	3,34	3,22	3,12	3,05	2,99	2,79	2,68	2,57	2,47	2,40	2,36	2,33	2,32
17	6,04	4,62	4,01	3,66	3,44	3,28	3,16	3,06	2,98	2,92	2,72	2,62	2,50	2,41	2,33	2,29	2,26	2,25
18	5,98	4,56	3,95	3,61	3,38	3,22	3,10	3,01	2,93	2,87	2,67	2,56	2,44	2,35	2,27	2,23	2,20	2,19
19	5,92	4,51	3,90	3,56	3,33	3,17	3,05	2,96	2,88	2,82	2,62	2,51	2,39	2,30	2,22	2,18	2,15	2,13
20	5,87	4,46	3,86	3,51	3,29	3,13	3,01	2,91	2,84	2,77	2,57	2,46	2,35	2,25	2,17	2,13	2,10	2,09
22	5,79	4,38	3,78	3,44	3,22	3,05	2,93	2,84	2,76	2,70	2,50	2,39	2,27	2,17	2,09	2,05	2,02	2,00
24	5,72	4,32	3,72	3,38	3,15	2,99	2,87	2,78	2,70	2,64	2,44	2,33	2,21	2,11	2,02	1,98	1,95	1,94
26	5,66	4,27	3,67	3,33	3,10	2,94	2,82	2,73	2,65	2,59	2,39	2,28	2,16	2,05	1,97	1,92	1,90	1,88
28	5,61	4,22	3,63	3,29	3,06	2,90	2,78	2,69	2,61	2,55	2,34	2,23	2,11	2,01	1,92	1,88	1,85	1,83
30	5,57	4,18	3,59	3,25	3,03	2,87	2,75	2,65	2,57	2,51	2,31	2,20	2,07	1,97	1,88	1,84	1,81	1,79
40	5,42	4,05	3,46	3,13	2,90	2,74	2,62	2,53	2,45	2,39	2,18	2,07	1,94	1,83	1,74	1,69	1,66	1,64
50	5,34	3,98	3,39	3,06	2,83	2,67	2,55	2,46	2,38	2,32	2,11	1,99	1,87	1,75	1,66	1,60	1,57	1,55
60	5,29	3,93	3,34	3,01	2,79	2,63	2,51	2,41	2,33	2,27	2,06	1,94	1,82	1,70	1,60	1,54	1,51	1,48
80	5,22	3,86	3,28	2,95	2,73	2,57	2,45	2,36	2,28	2,21	2,00	1,88	1,75	1,63	1,53	1,47	1,43	1,40
100	5,18	3,83	3,25	2,92	2,70	2,54	2,42	2,32	2,24	2,18	1,97	1,85	1,71	1,59	1,48	1,42	1,38	1,35
200	5,10	3,76	3,18	2,85	2,63	2,47	2,35	2,26	2,18	2,11	1,90	1,78	1,64	1,51	1,39	1,32	1,27	1,23
500	5,05	3,72	3,14	2,81	2,59	2,43	2,31	2,22	2,14	2,07	1,86	1,74	1,60	1,46	1,34	1,25	1,19	1,14
•	5,02	3,69	3,12	2,79	2,57	2,41	2,29	2,19	2,11	2,05	1,83	1,71	1,57	1,43	1,30	1,21	1,13	1,00

Loi de Fisher

$\alpha = 0,95$

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	15	20	30	50	100	200	500	•
1	161	200	216	225	230	234	237	239	241	242	246	248	250	252	253	254	254	254
2	18,5	19,0	19,2	19,2	19,3	19,3	19,4	19,4	19,4	19,4	19,4	19,4	19,5	19,5	19,5	19,5	19,5	19,5
3	10,1	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,70	8,66	8,62	8,58	8,55	8,54	8,53	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,86	5,80	5,75	5,70	5,66	5,65	5,64	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,62	4,56	4,50	4,44	4,41	4,39	4,37	4,37
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	3,94	3,87	3,81	3,75	3,71	3,69	3,68	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,51	3,44	3,38	3,32	3,27	3,25	3,24	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,22	3,15	3,08	3,02	2,97	2,95	2,94	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,01	2,94	2,86	2,80	2,76	2,73	2,72	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,85	2,77	2,70	2,64	2,59	2,56	2,55	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,72	2,65	2,57	2,51	2,46	2,43	2,42	2,40
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,62	2,54	2,47	2,40	2,35	2,32	2,31	2,30
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,53	2,46	2,38	2,31	2,26	2,23	2,22	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,46	2,39	2,31	2,24	2,19	2,16	2,14	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,40	2,33	2,25	2,18	2,12	2,10	2,08	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,35	2,28	2,19	2,12	2,07	2,04	2,02	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,31	2,23	2,15	2,08	2,02	1,99	1,97	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,27	2,19	2,11	2,04	1,98	1,95	1,93	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,23	2,16	2,07	2,00	1,94	1,91	1,89	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,20	2,12	2,04	1,97	1,91	1,88	1,86	1,84
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,15	2,07	1,98	1,91	1,85	1,82	1,80	1,78
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,11	2,03	1,94	1,86	1,80	1,77	1,75	1,73
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,07	1,99	1,90	1,82	1,76	1,73	1,71	1,69
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,04	1,96	1,87	1,79	1,73	1,69	1,67	1,65
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,01	1,93	1,84	1,76	1,70	1,66	1,64	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	1,92	1,84	1,74	1,66	1,59	1,55	1,53	1,51
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,03	1,87	1,78	1,69	1,60	1,52	1,48	1,46	1,44
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,84	1,75	1,65	1,56	1,48	1,44	1,41	1,39
80	3,96	3,11	2,72	2,49	2,33	2,21	2,13	2,06	2,00	1,95	1,79	1,70	1,60	1,51	1,43	1,38	1,35	1,32
100	3,94	3,09	2,70	2,46	2,31	2,19	2,10	2,03	1,97	1,93	1,77	1,68	1,57	1,48	1,39	1,34	1,31	1,28
200	3,89	3,04	2,65	2,42	2,26	2,14	2,06	1,98	1,93	1,88	1,72	1,62	1,52	1,41	1,32	1,26	1,22	1,19
500	3,86	3,01	2,62	2,39	2,23	2,12	2,03	1,96	1,90	1,85	1,69	1,59	1,48	1,38	1,28	1,21	1,16	1,11
•	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,67	1,57	1,46	1,35	1,24	1,17	1,11	1,00