

# Compléments sur la régression linéaire simple

## Anova et inférence sur les paramètres

Frédéric & Myriam Bertrand<sup>1</sup>

<sup>1</sup>IRMA, Université Louis Pasteur  
Strasbourg, France

Master 2ème Année 02-10-2006

Ce cours s'appuie essentiellement sur les deux ouvrages suivants :

- “Analyse de régression appliquée” de Y. Dodge et V. Rousson, Dunod, 2004.
- “Régression non linéaire et applications” de A. Antoniadis, J. Berruyer, R. Carmona, Economica, 1992.

- Il existe plusieurs démarches pour tester la validité de la linéarité d'une régression simple.
- On montre l'équivalence de ces différents tests.
- Conséquence : Cela revient à faire **le test du coefficient de corrélation linéaire**, appelé aussi le coefficient de Bravais-Pearson.

**Remarque :** On peut consulter sur le site un cours sur le coefficient de corrélation linéaire (Cours de L3).

On désire tester l'hypothèse nulle :

$$\mathcal{H}_0 : \rho = 0$$

contre l'hypothèse alternative :

$$\mathcal{H}_1 : \rho \neq 0$$

où

$$\rho = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}[X]\text{Var}[Y]}}$$

avec

$$\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \text{Cov}(Y, X),$$

$$\text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}^2[X] \quad \text{et} \quad \text{Var}[Y] = \mathbb{E}[Y^2] - \mathbb{E}^2[Y].$$

La méthode que nous allons employer ici est :

## la méthode de l'ANOVA

utilisée par les logiciels de statistique.

**Remarque :** ANOVA pour ANalysis Of VAriance ou encore analyse de la variance.

**Remarque :** On peut consulter sur le site un cours sur le test du coefficient de corrélation linéaire (Cours de L3).

On a établi précédemment :

**Somme des Carrés Totale = Somme des Carrés Expliquée  
+ Somme des Carrés Résiduelle**

ce qui s'écrit mathématiquement par :

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2.$$

À chaque somme de carrés est associé son nombre de degrés de liberté (*ddl.*) Ces *ddl* sont présents dans le tableau de l'ANOVA.

Source de variation	SC	ddl	CM
régression SCE	$\sum_{i=1}^n (\hat{y}_i - \bar{y})^2$	1	SCE/1
résiduelle SCR	$\sum_{i=1}^n (y_i - \hat{y}_i)^2$	$n - 2$	SCR/( $n - 2$ )
totale SCT	$\sum_{i=1}^n (y_i - \bar{y})^2$	$n - 1$	

## Remarques :

- Le **coefficient de détermination**

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

mesure le pourcentage d'explication du modèle par la régression linéaire.

- Le rapport

$$s^2 = \frac{SCR}{n-2}$$

est l'**estimation de la variance résiduelle**.

À partir du tableau de l'ANOVA, on effectue **le test de la linéarité de la régression** en calculant **la statistique de Fisher  $F$**  qui suit une loi de Fisher  $F(1, n - 2)$ .

Cette variable aléatoire  $F$  se réalise en :

$$F_{obs} = \frac{SCE/1}{SCR/(n-2)} = (n-2) \frac{SCE}{SCR}.$$

Si

$$F_{obs} \geq F_{\alpha}(1, n - 2),$$

alors on rejette l'hypothèse nulle  $\mathcal{H}_0$  au risque  $\alpha$ , c'est-à-dire qu'il existe une liaison linéaire significative entre  $X$  et  $Y$ .

Si

$$F_{obs} < F_{\alpha}(1, n - 2),$$

alors on accepte l'hypothèse nulle  $\mathcal{H}_0$ , c'est-à-dire qu'il n'existe pas de liaison linéaire entre  $X$  et  $Y$ .

**Remarque :** En effet, si l'hypothèse nulle  $\mathcal{H}_0$  est vérifiée alors cela implique que  $\rho = 0$  c'est-à-dire  $Cov(X, Y) = 0$ . Donc il n'existe aucune liaison linéaire entre  $X$  et  $Y$ .

Le modèle de régression linéaire simple est

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

où les  $\varepsilon_i$  sont des variables aléatoires inobservables, appelées **les erreurs**.

**Conséquence** : Les  $y_i$  sont des variables aléatoires.

**Première hypothèse** :  $\mathbb{E}[\varepsilon_i] = 0$ .

**Conséquence** :  $\mathbb{E}[y_i] = \beta_0 + \beta_1 x_i$ .

D'autre part, on a :

$$\text{Var}[y_i] = \text{Var}[\varepsilon_i].$$

## Les trois hypothèses indispensables pour construire la théorie :

1. La variance des variables aléatoires  $\varepsilon_i$  est égale à  $\sigma^2$  (inconnue) ne dépendant pas de  $x_i$ .  
On a donc pour tout  $i = 1, \dots, n$  :

$$\text{Var}[\varepsilon_i] = \text{Var}[y_i] = \sigma^2.$$

2. Les variables aléatoires  $\varepsilon_i$  sont indépendantes.
3. Les variables aléatoires  $\varepsilon_i$  sont normalement distribuées.

Ces trois hypothèses sont équivalentes à :

**les variables aléatoires  $\varepsilon_i$  sont indépendantes et identiquement distribuées selon une loi normale de moyenne nulle et de variance  $\sigma^2$ .**

On note :

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

## Conséquences importantes :

- La normalité des variables aléatoires  $\varepsilon_i$  implique la normalité des variables aléatoires  $y_i$ .
- L'indépendance des variables aléatoires  $\varepsilon_i$  implique l'indépendance des variables aléatoires  $y_i$ .  
En effet, on montre en calculant que :

$$\begin{aligned} \text{Cov}[y_i, y_j] &= \text{Cov}[\beta_0 + \beta_1 x_i + \varepsilon_i, \beta_0 + \beta_1 x_j + \varepsilon_j] \\ &= \text{Cov}[\varepsilon_i, \varepsilon_j] \\ &= 0. \end{aligned}$$

On a :

$$\hat{\beta}_1 = \frac{\sum (x_i - \bar{x}) y_i}{\sum (x_i - \bar{x})^2},$$

où

$$\bar{x} = \frac{\sum x_i}{n}.$$

Il en résulte que :

- $\hat{\beta}_1$  **est une variable aléatoire** car  $\hat{\beta}_1$  dépend des variables  $y_i$  qui sont des variables aléatoires.
- $\hat{\beta}_1$  **est une fonction linéaire des variables**  $y_i$ .
- Comme les variables  $y_i$  par hypothèse sont normalement distribuées, alors  $\hat{\beta}_1$  **est normalement distribué**.

Il reste donc à calculer ces deux valeurs pour caractériser l'estimateur  $\hat{\beta}_1$  :

- $\mathbb{E} \left[ \hat{\beta}_1 \right]$
- $Var \left[ \hat{\beta}_1 \right]$ .

Par calcul, on montre que :

$$\begin{aligned}
 \mathbb{E} \left[ \hat{\beta}_1 \right] &= \mathbb{E} \left[ \frac{\sum (x_i - \bar{x}) y_i}{\sum (x_i - \bar{x})^2} \right] \\
 &= \frac{\sum (x_i - \bar{x}) \mathbb{E}[y_i]}{\sum (x_i - \bar{x})^2} \\
 &= \frac{\sum (x_i - \bar{x}) (\beta_0 + \beta_1 x_i)}{\sum (x_i - \bar{x})^2} \\
 &= \frac{\beta_0 \sum (x_i - \bar{x}) + \beta_1 \sum (x_i - \bar{x}) x_i}{\sum (x_i - \bar{x})^2} \\
 &= \frac{0 + \beta_1 \sum (x_i - \bar{x}) x_i}{\sum (x_i - \bar{x})^2}.
 \end{aligned}$$

En effet, on montre que :

$$\sum (x_i - \bar{x}) = 0.$$

De plus, comme on a :

$$\sum (x_i - \bar{x})^2 = \sum (x_i - \bar{x})x_i$$

alors on obtient :

$$\mathbb{E} [\hat{\beta}_1] = \beta_1.$$

Donc la variable aléatoire  $\hat{\beta}_1$  est **un estimateur sans biais** du coefficient  $\beta_1$ .

D'autre part, on calcule la variance de  $\hat{\beta}_1$  ainsi :

$$\begin{aligned} \text{Var} [\hat{\beta}_1] &= \text{Var} \left[ \frac{\sum (x_i - \bar{x}) y_i}{\sum (x_i - \bar{x})^2} \right] \\ &= \frac{\sum (x_i - \bar{x})^2 \text{Var}[y_i]}{(\sum (x_i - \bar{x})^2)^2} \\ &= \frac{\sum (x_i - \bar{x})^2 \sigma^2}{(\sum (x_i - \bar{x})^2)^2} \\ &= \frac{\sigma^2}{\sum (x_i - \bar{x})^2}, \end{aligned}$$

ce qui achève la caractérisation de  $\hat{\beta}_1$ .

On a :

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

où

$$\bar{x} = \frac{\sum x_i}{n} \quad \text{et} \quad \bar{y} = \frac{\sum y_i}{n}.$$

- $\hat{\beta}_0$  est une variable aléatoire car  $\hat{\beta}_0$  dépend de  $\hat{\beta}_1$  qui est une variable aléatoire.
- $\hat{\beta}_0$  est une fonction linéaire de  $\hat{\beta}_1$ .
- Comme  $\hat{\beta}_1$  est normalement distribuée, alors  $\hat{\beta}_0$  est normalement distribuée.

Il reste donc à calculer ces deux valeurs pour caractériser l'estimateur  $\hat{\beta}_0$  :

- $\mathbb{E} \left[ \hat{\beta}_0 \right]$
- $Var \left[ \hat{\beta}_0 \right]$ .

Par calcul, on montre que :

$$\begin{aligned}\mathbb{E}[\hat{\beta}_0] &= \mathbb{E}[\bar{y} - \hat{\beta}_1 \bar{x}] \\ &= \mathbb{E}[\bar{y}] - \bar{x} \mathbb{E}[\hat{\beta}_1] \\ &= \mathbb{E}[\bar{y}] - \bar{x} \beta_1,\end{aligned}$$

car on vient de démontrer que  $\hat{\beta}_1$  est un estimateur sans biais du coefficient  $\beta_1$ .

Il reste à calculer la valeur :

$$\mathbb{E}[\bar{y}].$$

Or on a :

$$\begin{aligned}\mathbb{E}[\bar{y}] &= \mathbb{E}\left[\frac{\sum y_i}{n}\right] \\ &= \frac{\sum \mathbb{E}[y_i]}{n} \\ &= \frac{\sum(\beta_0 + \beta_1 x_i)}{n} \\ &= \frac{n\beta_0 + \beta_1 \sum x_i}{n} \\ &= \beta_0 + \bar{x}\beta_1.\end{aligned}$$

On obtient donc :

$$\begin{aligned}\mathbb{E} \left[ \hat{\beta}_0 \right] &= \mathbb{E} [\bar{y}] - \bar{x}\beta_1 \\ &= (\beta_0 + \bar{x}\beta_1) - \bar{x}\beta_1 \\ &= \beta_0.\end{aligned}$$

Donc la variable aléatoire  $\hat{\beta}_0$  est **un estimateur sans biais** du coefficient  $\beta_0$ .

D'autre part, on calcule la variance de  $\hat{\beta}_0$  ainsi :

$$\begin{aligned} \text{Var} [\hat{\beta}_0] &= \text{Var} [\bar{y} - \hat{\beta}_1 \bar{x}] \\ &= \text{Var}[\bar{y}] + \bar{x}^2 \text{Var} [\hat{\beta}_1] - 2 \bar{x} \text{Cov} [\bar{y}, \hat{\beta}_1]. \end{aligned}$$

Il reste donc à calculer la valeur :

$$\text{Cov} [\bar{y}, \hat{\beta}_1].$$

Par les calculs, on montre que :

$$\begin{aligned}
 \text{Cov} \left[ \bar{y}, \hat{\beta}_1 \right] &= \text{Cov} \left[ \frac{\sum y_i}{n}, \frac{\sum (x_j - \bar{x}) y_j}{\sum (x_i - \bar{x})^2} \right] \\
 &= \frac{\sum_i \sum_j (x_j - \bar{x}) \text{Cov}[y_i, y_j]}{n \sum (x_i - \bar{x})^2} \\
 &= \frac{\sum_i (x_i - \bar{x}) \text{Var}[y_i]}{n \sum (x_i - \bar{x})^2} \\
 &= \frac{\sigma^2 \sum_i (x_i - \bar{x})}{n \sum (x_i - \bar{x})^2} \\
 &= 0.
 \end{aligned}$$

Comme

$$\text{Var}[\bar{y}] = \frac{\sigma^2}{n},$$

on obtient :

$$\begin{aligned}\text{Var}[\hat{\beta}_0] &= \text{Var}[\bar{y}] + \bar{x}^2 \text{Var}[\hat{\beta}_1] \\ &= \frac{\sigma^2}{n} + \frac{\bar{x}^2 \sigma^2}{\sum (x_i - \bar{x})^2} \\ &= \frac{\sigma^2 \left( \sum (x_i - \bar{x})^2 + n\bar{x}^2 \right)}{n \sum (x_i - \bar{x})^2}.\end{aligned}$$

En rappelant que :

$$\sum (x_i - \bar{x})^2 = \sum x_i^2 - n\bar{x}^2,$$

on a finalement :

$$\text{Var} [\hat{\beta}_0] = \frac{\sigma^2 \sum x_i^2}{n \sum (x_i - \bar{x})^2}.$$

On rappelle que :

$$\hat{\beta}_1 \sim \mathcal{N}(\beta_1; \sigma^2(\hat{\beta}_1))$$

où

$$\sigma^2(\hat{\beta}_1) = \frac{\sigma^2}{\sum (x_i - \bar{x})^2}.$$

On obtient alors :

$$\frac{\hat{\beta}_1 - \beta_1}{\sigma(\hat{\beta}_1)} \sim \mathcal{N}(0; 1).$$

## Problème :

On ne connaît pas le paramètre  $\sigma^2$ , c'est-à-dire la variance des variables aléatoires  $\varepsilon_j$ .

Que peut-on faire alors pour résoudre ce problème ?

**Solution :** Estimer ce paramètre !

- On estime d'abord  $\sigma^2$  par  $s^2$  l'estimateur sans biais de  $\sigma^2$  :

$$s^2 = \frac{\|\varepsilon\|^2}{n-2} = \frac{\sum (y_i - \hat{y}_i)^2}{n-2}.$$

- On estime ensuite  $\sigma^2(\hat{\beta}_1)$  par :

$$s^2(\hat{\beta}_1) = \frac{s^2}{\sum (x_i - \bar{x})^2}.$$

- On montre alors que :

$$\frac{\hat{\beta}_1 - \beta_1}{s(\hat{\beta}_1)} \sim T_{n-2},$$

où  $T_{n-2}$  désigne une variable aléatoire de Student avec  $(n-2)$  ddl.

On désire tester l'hypothèse nulle :

$$\mathcal{H}_0 : \beta_1 = 0$$

contre l'hypothèse alternative :

$$\mathcal{H}_1 : \beta_1 \neq 0.$$

On utilise la statistique :

$$t_{obs} = \frac{\hat{\beta}_1}{s(\hat{\beta}_1)}$$

pour décider de l'acceptation ou du rejet de l'hypothèse nulle  $\mathcal{H}_0$ .

## Deux conclusions sont possibles :

On rejette l'hypothèse nulle  $\mathcal{H}_0$  au seuil de signification  $\alpha$  si

$$|t_{obs}| > t_{(\alpha/2, n-2)}$$

où la valeur critique  $t_{(\alpha/2, n-2)}$  est le  $(1 - \alpha/2)$ -quantile d'une loi de Student avec  $(n - 2)$  ddl.

Dans ce cas, on dit que la relation linéaire entre  $X$  et  $Y$  est significative au seuil  $\alpha$ .

On accepte l'hypothèse nulle  $\mathcal{H}_0$  au seuil de signification  $\alpha$  si

$$|t_{obs}| < t_{(\alpha/2, n-2)}$$

où la valeur  $t_{(\alpha/2, n-2)}$  est le  $(1 - \alpha/2)$ -quantile d'une loi de Student avec  $(n - 2)$  ddl.

Dans ce cas,  $Y$  ne dépend pas linéairement de  $X$ . Le modèle devient alors :

$$y_i = \beta_0 + \varepsilon_i$$

Le modèle proposé  $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$  est inadéquat. On teste alors un nouveau modèle.

Un intervalle de confiance au niveau  $(1 - \alpha)$  pour le coefficient inconnu  $\beta_1$  est défini par

$$\left[ \hat{\beta}_1 - t_{(\alpha/2, n-2)} \times s(\hat{\beta}_1) ; \hat{\beta}_1 + t_{(\alpha/2, n-2)} \times s(\hat{\beta}_1) \right].$$

Cet intervalle de confiance est construit de telle sorte qu'il contienne le coefficient inconnu  $\beta_1$  avec une probabilité égale à  $(1 - \alpha)$ .

On rappelle que :

$$\hat{\beta}_0 \sim \mathcal{N}(\beta_0; \sigma^2(\hat{\beta}_0))$$

où

$$\sigma^2(\hat{\beta}_0) = \frac{\sigma^2 \sum x_i^2}{n \sum (x_i - \bar{x})^2}.$$

On obtient alors :

$$\frac{\hat{\beta}_0 - \beta_0}{\sigma(\hat{\beta}_0)} \sim \mathcal{N}(0; 1).$$

## Problème :

On ne connaît pas le paramètre  $\sigma^2$ , c'est-à-dire la variance des variables aléatoires  $\varepsilon_j$ .

Que peut-on faire alors pour résoudre ce problème ?

**Solution :** Estimer ce paramètre !

- On estime d'abord  $\sigma^2$  par  $s^2$  l'estimateur sans biais de  $\sigma^2$  et  $s^2$  :

$$s^2 = \frac{\|\varepsilon\|^2}{n-2} = \frac{\sum (y_i - \hat{y}_i)^2}{n-2}.$$

- On estime ensuite  $\sigma^2(\hat{\beta}_0)$  par

$$s^2(\hat{\beta}_0) = \frac{s^2 \sum x_i^2}{n \sum (x_i - \bar{x})^2}.$$

- On montre alors que :

$$\frac{\hat{\beta}_0 - \beta_0}{s(\hat{\beta}_0)} \sim T_{n-2},$$

où  $T_{n-2}$  désigne une variable aléatoire de Student avec  $(n-2)$  ddl.

On désire tester l'hypothèse nulle

$$\mathcal{H}_0 : \beta_0 = 0$$

contre l'hypothèse alternative

$$\mathcal{H}_1 : \beta_0 \neq 0.$$

On utilise la statistique

$$t_{obs} = \frac{\hat{\beta}_0}{s(\hat{\beta}_0)}$$

pour décider de l'acceptation ou du rejet de l'hypothèse nulle  $\mathcal{H}_0$ .

## Deux conclusions sont possibles :

On rejette l'hypothèse nulle  $\mathcal{H}_0$  au seuil de signification  $\alpha$  si :

$$|t_{obs}| > t_{(\alpha/2, n-2)}$$

où la valeur critique  $t_{\alpha/2, n-2}$  est le  $(1 - \alpha/2)$ -quantile d'une loi de Student avec  $(n - 2)$  ddl.

Dans ce cas, le coefficient  $\beta_0$  du modèle est dit significatif au seuil  $\alpha$ .

On accepte l'hypothèse nulle  $\mathcal{H}_0$  au seuil de signification  $\alpha$  si

$$|t_{obs}| < t_{(\alpha/2, n-2)}$$

où la valeur critique  $t_{(\alpha/2, n-2)}$  est le  $(1 - \alpha/2)$ -quantile d'une loi de Student avec  $(n - 2)$  ddl.

Dans ce cas, l'ordonnée de la droite de régression passe par l'origine.

$$y_i = \beta_1 x_i + \varepsilon_i$$

Un intervalle de confiance au niveau  $(1 - \alpha)$  pour le coefficient inconnu  $\beta_0$  est défini par :

$$\left[ \hat{\beta}_0 - t_{(\alpha/2, n-2)} \times s(\hat{\beta}_0) ; \hat{\beta}_0 + t_{(\alpha/2, n-2)} \times s(\hat{\beta}_0) \right].$$

Cet intervalle de confiance est construit de telle sorte qu'il contienne le coefficient inconnu  $\beta_0$  avec une probabilité égale à  $(1 - \alpha)$ .

On va voir comment trouver un intervalle de confiance pour

$$\mu_Y(x) = \beta_0 + \beta_1 x,$$

c'est-à-dire pour l'ordonnée du point d'abscisse  $x$  se trouvant sur la droite de régression.

L'estimateur de  $\beta_0 + \beta_1 x$  est donné par la droite des moindres carrés :

$$\hat{y}(x) = \hat{\beta}_0 + \hat{\beta}_1 x,$$

où

- $\hat{y}(x) \sim \mathcal{N}(\beta_0 + \beta_1 x; \sigma^2(\hat{y}(x)))$

où

$$\sigma^2(\hat{y}(x)) = \sigma^2 \left( \frac{1}{n} + \frac{(x - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right).$$

Ce qui peut s'écrire aussi :

- $\frac{\hat{y}(x) - \mu_Y(x)}{\sigma(\hat{y}(x))} \sim \mathcal{N}(0; 1).$

**Problème :** La variance  $\sigma^2$  est inconnue.

**Solution :**

- On estime d'abord  $\sigma^2$  par  $s^2$ .
- On estime ensuite  $\sigma^2(\hat{y}(x))$  par :

$$s^2(\hat{y}(x)) = s^2 \left( \frac{1}{n} + \frac{(x - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right).$$

- Ainsi on obtient :

$$\frac{\hat{y}(x) - \mu_Y(x)}{s(\hat{y}(x))} \sim T_{n-2}.$$

Un intervalle de confiance au niveau  $(1 - \alpha)$  pour le paramètre inconnu  $\mu_Y(x)$  est défini par :

$$\left[ \hat{y}(x) - t_{(\alpha/2; n-2)} \times s(\hat{y}(x)) ; \hat{y}(x) + t_{(\alpha/2; n-2)} \times s(\hat{y}(x)) \right].$$

Cet intervalle de confiance est construit de telle sorte qu'il contienne le paramètre inconnu  $\mu_Y(x)$  avec une probabilité égale à  $(1 - \alpha)$ .

Observations $i$	Pays	Taux d'urbanisation $x_i$	Taux de natalité $y_i$	Valeurs estimées $\hat{y}_i$	Résidus $e_i$
1	Canada	55,0	16,2	21,05	-4,85
2	Costa Rica	27,3	30,5	32,10	-1,60
3	Cuba	33,3	16,9	29,71	-12,81
4	E.U.	56,5	16,0	20,45	-4,45
5	El Salvador	11,5	40,2	38,40	1,80
6	Guatemala	14,2	38,4	37,33	1,07
7	Haïti	13,9	41,3	37,45	3,83

Observations $i$	Pays	Taux d'urbanisation $x_i$	Taux de natalité $y_i$	Valeurs estimées $\hat{y}_i$	Résidus $e_i$
8	Honduras	19,0	43,9	35,41	8,49
9	Jamaïque	33,1	28,3	29,79	-1,49
10	Mexique	43,2	33,9	25,76	8,14
11	Nicaragua	28,5	44,2	31,62	12,58
12	Trinitade	6,8	24,6	40,28	-15,68
13	Panama	37,7	28,0	27,95	0,05
14	Rép. Dom.	37,1	33,1	28,19	4,91

Le tableur Excel donne successivement :

$$\hat{\beta}_1 = -0,3989,$$

$$\hat{\beta}_0 = 42,991$$

et enfin

$$s^2 = 66,24.$$

## Test sur la pente $\beta_1$ .

$$\mathcal{H}_0 : \beta_1 = 0$$

contre

$$\mathcal{H}_1 : \beta_1 \neq 0.$$

On calcule

$$t_{obs} = \frac{\hat{\beta}_1}{s(\hat{\beta}_1)} = \frac{-0,3989}{\sqrt{0,021}} = -2,75.$$

Or la valeur critique est égale à pour un seuil  $\alpha = 0,050$  :

$$t_{(0,025,12)} = 2,179.$$

Comme

$$|t_{obs}| > t_{(\alpha/2, n-2)},$$

on rejette l'hypothèse nulle  $\mathcal{H}_0$  pour décider l'hypothèse alternative  $\mathcal{H}_1$ .

**En conclusion :** La relation linéaire entre le taux de natalité et le taux d'urbanisation est significative.

Un intervalle de confiance pour le coefficient inconnu  $\beta_1$  au niveau  $(1 - \alpha) = 0,95$  s'obtient en calculant :

$$\hat{\beta}_1 \pm t_{(\alpha/2, n-2)} \times s(\hat{\beta}_1) = -0,3989 \pm 2,179 \times \sqrt{0,021}.$$

On a donc après simplification :

$$[-0,715; -0,083]$$

qui contient la vraie valeur du coefficient inconnu  $\beta_1$  avec une probabilité de 0,95. On remarque que 0 n'est pas compris dans cet intervalle.

## Test sur l'ordonnée $\beta_0$ .

$$\mathcal{H}_0 : \beta_0 = 0$$

contre

$$\mathcal{H}_1 : \beta_0 \neq 0.$$

On calcule

$$t_{obs} = \frac{\hat{\beta}_0}{s(\hat{\beta}_0)} = \frac{42,991}{\sqrt{23,373}} = 8,89.$$

Or la valeur critique est égale à pour un seuil  $\alpha = 0,050$  :

$$t_{0,025,12} = 2,179.$$

Comme

$$|t_{obs}| > t_{\alpha/2, n-2},$$

on rejette l'hypothèse nulle  $\mathcal{H}_0$  pour décider de l'hypothèse alternative  $\mathcal{H}_1$ .

**En conclusion :** La droite de régression ne passe pas par l'origine.

Un intervalle de confiance pour le coefficient inconnu  $\beta_0$  au niveau  $(1 - \alpha) = 0,95$  s'obtient en calculant :

$$\hat{\beta}_0 \pm t_{\alpha/2, n-2} \times s(\hat{\beta}_0) = 42,991 \pm 2,179 \times \sqrt{23,373}.$$

On a donc après simplification :

$$[32,456; 53,526]$$

qui contient la vraie valeur du coefficient inconnu  $\beta_0$  avec une probabilité de 0,95. On remarque que 0 n'est pas compris dans l'intervalle.