Notes de cours
Statistiques
ACT-2000

Gabriel Crépeault-Cauchon

Dernière mise à jour : 11 avril 2018

Table des matières

Chapitre 1

Introduction

1.1 Définitions de propriétés de base

Statistique

Une statistique est une fonction $T(X_1, X_2, X_3, ..., X_n)$ de n variables aléatoire indépendantes et identiquement distribués *(iid)*

Moyenne et variance échantillonale

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$
$$S^2 = \sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{n-1}$$

Remarques

1. $\sum_{i=1}^{n} (X_i - \overline{X})^2 = \sum_{i=1}^{n} X_i^2 - n\overline{X}^2$
2. $\sum_{i=1}^{n} (X_i - \overline{X})^2 = \sum_{i=1}^{n} (X_i - \mu)^2 + n(\mu - \overline{X})^2$

Note : les développement de l'item 1 est disponible à la section 5.2.1.

Espérance et moyenne de la variance échantillonale Par le résultat de la section 1.3.2, on peut déduire la variance de la variance échantillonale :

$$\begin{aligned} E[S^2] &= \frac{\sigma^2}{n-1} E\left[\frac{(n-1)S^2}{\sigma^2}\right] \\ &= \frac{\sigma^2}{n-1}(n-1) \\ &= \sigma^2 \end{aligned}$$

Et la variance,

$$\begin{aligned} Var(S^2) &= \frac{\sigma^4}{(n-1)^2} Var \underbrace{\left(\frac{(n-1)S^2}{\sigma^2}\right)}_{\chi^2_{(n-1)}} \\ &= \frac{2(n-1)\sigma^4}{(n-1)^2} \\ &= \frac{2\sigma^4}{(n-1)} \end{aligned}$$

Étendue échantillonale

Soit $Y_1 = \min(X_1, X_2, ..., X_n)$ et $Y_n = \max(X_1, X_2, ..., X_n)$. La variable aléatoire $R = Y_n - Y_1$ représente l'étendue échantillonale

Moments échantillonnaux

Le k^e moment autour de l'origine :

$$M_k = \sum_{i=1}^{n} \frac{X_i^k}{n}$$

Le k^e moment autour de la moyenne échantillonnale :

$$M'_k = \sum_{i=1}^{n} \frac{(X_i - \overline{X})^k}{n}$$

Remarques

1. $M'_1 = \sum_{i=1}^{n} \frac{(X_i - \overline{X})}{n} = \frac{1}{n}\left(\sum_{i=1}^{n} X_i - n\overline{X}\right) = 0$
2. $M'_2 = \sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{n} = \frac{n-1}{n} \sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{n-1} = \frac{n-1}{n} S^2$

Espérance et variance des moments

$$E[M_k] = E[X^k]$$
$$Var(M_k) = \frac{Var(X^k)}{n}$$

Loi des grands nombres

$$\lim_{n\to\infty} P(|\overline{X} - \mu| < \varepsilon) = 1, \quad \forall \varepsilon > 0$$

1.2 Distribution de $\overline{X}$ pour une population normale

Si on sélectionne un échantillon aléatoire d'une population normale avec moyenne μ et variance σ^2, alors

$$\overline{X} \backsim N(\mu, \frac{\sigma^2}{n})$$

Différence de deux échantillons venant de populations normales : si on a une population $(X_1, X_2, ..., X_n)$ et $(Y_1, Y_2, ..., Y_m)$, alors

$$\overline{X} - \overline{Y} \backsim \left(\mu_X - \mu_Y, \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}\right)$$

1.2.1 Inégalité de Tchebychev

$$P(|X - E[X]| \geq k\sigma) \leq \frac{1}{k^2} \tag{1.1}$$

1.2.2 Théorème central limite

$$F_{W_n}(x) \underset{n\to\infty}{\longrightarrow} F_Z(x)$$

où $W_n = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ et $Z \backsim N(0,1)$. On commence à considérer n grand lorsque $n \geq 30$.

Remarque Pour un n grand, on dira

$$\overline{X} \backsim N(\mu, \sigma^2/n), \quad \sum_{i=1}^{n} X_i \backsim N(n\mu, n\sigma^2)$$

ATTENTION

lorsqu'on utilise une population qui a un *support discret*, il faut faire une correction pour la discontinuité dans notre estimation avec la loi normale, afin de s'assurer d'avoir la borne qu'on mesure.

1.3 Loi du Chi-carré ($\chi^2_{(n)}$)

Il faut se rappeler que la loi du Chi-carré est une Gamma avec $\alpha = \frac{n}{2}$ et $\lambda = \frac{1}{2}$.

$$f(x) = \frac{(1/2)^{n/2} x^{(n/2)-1} e^{-x/2}}{\Gamma(n/2)}, \quad x > 0$$

Propriété de la fonction Gamma		
$\Gamma(n+1) = n\Gamma(n)$	$\Gamma(1/2) = \sqrt{\pi}$	

Éléments à savoir pour la loi du Chi-carré		
$E[X] = n$	$Var(X) = 2n$	$M_X(t) = (1-2t)^{-n/2}$

1.3.1 Propositions importantes

1. Si $X \backsim N(0,1)$, alors

$$X^2 \backsim \chi^2_{(1)}$$

2. Si $X_1, X_2, ..., X_n$ sont n v.a. indépendantes où $X_i \sim N(0,1)$, alors

$$\sum_{i=1}^{n} X_i^2 \backsim \chi^2_{(n)}$$

3. Si $X_1, X_2, ..., X_n$ sont n v.a. indépendantes tel que $X_1 \backsim \chi^2_{(m_1)}$, ..., $X_n \backsim \chi^2_{(m_n)}$, alors

$$\sum_{i=1}^{n} X_i \backsim \chi^2_{(m_1+...+m_n)}$$

4. Si X_1 et X_2 sont des v.a. telle que $X_1 \backsim \chi^2_{(n_1)}$ et $X_1 + X_2 \backsim \chi^2_{(n)}$, alors

$$X_2 \backsim \chi^2_{(n-n_1)}$$

1.3.2 Distribution de $\frac{(n-1)S^2}{\sigma^2}$

Si on a $\overline{X}$ et S^2 qui sont indépendants et qui proviennent d'un échantillon aléatoire de taille n d'une population normale $N(\mu, \sigma^2)$, alors

1. $\overline{X}$ et S^2 sont indépendants (*au niveau des probabilités*)
2. $\frac{(n-1)S^2}{\sigma^2} \backsim \chi^2_{(n-1)}$

1.4 Loi de Student

Définition : Soit X et X des v.a. indépendantes, avec $X \backsim N(0,1)$ et $Y \backsim \chi^2_{(n)}$. Alors, on a la v.a.

$$t_{(n)} = \frac{X}{\sqrt{Y/n}}$$

Avec la fonction de densité de probabilité

$$f_{t_{(n)}}(t) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{\pi n}\Gamma(\frac{n}{2})}(1+\frac{t^2}{n})^{-(\frac{n+1}{2})}$$

Remarque importante : lorsque le nombre d'observations tend vers l'infini,

$$\lim_{n\to\infty} f_{t_{(n)}}(t) = \frac{e^{-t^2/2}}{\sqrt{2\pi}}$$

Voici deux résultats importants à savoir pour faire certaines preuves :

$$\lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n = e^x$$
$$\Gamma(n+1) = n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$$

éléments à savoir pour la loi de Student	
$E[X]=0$	$Var(X)=\frac{n}{n-2}$

1.4.1 Distribution de $\frac{\overline{X}-\mu}{S/\sqrt{n}}$

Si on a un échantillon avec une moyenne μ et une variance échantillonnale S^2, Alors

$$\frac{\overline{X}-\mu}{S/\sqrt{n}} \backsim t_{(n-1)} \tag{1.2}$$

Démonstration. La preuve se trouve à □

1.5 Loi de Fisher

Définition : Soit X et Y des v.a. indépendantes, et on a $X \backsim \chi^2_{(n_1)}$ et $Y \backsim \chi^2_{(n_2)}$. Alors,

$$\frac{X/n_1}{Y/n_2} \backsim F_{n_1,n_2} \tag{1.3}$$

Avec la fonction de densité de probabilité

$$f_{F_{n_1,n_2}}(x) = \frac{\Gamma\left(\frac{n_1+n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)}\left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}} x^{\frac{n_1}{2}-1}\left(1-\frac{n_1}{n_2}x\right)^{-\frac{n_1+n_2}{2}}$$

L'espérance et la variance sont

$$E[F_{n_1,n_2}] = \frac{n_2}{n_2-2}, \quad n_2 > 2$$

$$Var(X) = \frac{2n_2^2(n_1+n_2-2)}{n_1(n_2-2)^2(n_2-4)}, \quad n_2 > 4$$

1.5.1 Propriétés de la Fisher

Si on a une distribution Fisher F_{n_1,n_2}, alors

$$\frac{1}{F_{n_1,n_2}} = F_{n_2,n_1}$$

1.5.2 Distribution de $\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$

Propriété de la Fisher : Soit S_1^2 et S_2^2, les 2 variances échantionnalles provenant de 2 populations normales de taille n_1 et n_2. On sait aussi que les variances de population sont σ_1^2 et σ_2^2. Alors, on a

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \backsim F_{n_1-1,n_2-1}$$

1.6 Statistiques d'ordre

Définition arranger en ordre non-décroissant les éléments d'un échantillon $(X_1,...,X_n)$ dans un ensemble ordonné $(Y_1,...,Y_n)$.

Exemple on a $(X_1,X_2,X_3,X_4) = (10,1,34,2)$. En statistique ordonnée, on aura $(Y_1,Y_2,Y_3,Y_4) = (1,2,10,34)$.

Important : pour une statistique ordonnée,

$$Y_1 = \min(X_1,...,X_n)$$
$$Y_n = \max(X_1,...,X_n)$$

1.6.1 Distribution de Y_k

La fonction de répartition est représentée par

$$F_{Y_k}(y) = \sum_{i=1}^{n} \binom{n}{j} (F(y))^j (1-F(y))^{n-j}$$

La fonction de densité de probabilité :

$$f_{Y_k}(y) = \frac{n!}{(k-1)!(n-k)!} f(y)(F(y))^{k-1}(1-F(y))^{n-k}$$

1.6.2 Distribution du min-max

$$P(Y_1 \leq x, Y_n \leq y) = P(Y_n \leq y) - P(Y_1 > x, Y_n \leq y)$$

Dans le même sens

$$F_{Y_1,Y_n}(x,y) = F(y)^n - (F(y) - F(x))^n$$

La fonction de densité de probabilité est donc

$$f_{Y_1,Y_n}(x,y) = n(n-1)f(x)f(y)(F(y) - F(x))^{n-2}, \quad \text{si } x \leq y$$

1.6.3 Distribution de l'amplitude échantillonnale

Si on a $R = Y_n - Y_1$, qui représente l'amplitude échantillonnale, alors

$$\begin{aligned} P(R \leq r) &= P(Y_n - Y_1 \leq r) \\ &= \int_{-\infty}^{\infty} \int_{x}^{x+r} f_{Y_1,Y_n}(x,y)dydx \end{aligned}$$

Alors,

$$f_R(r) = n(n-1) \int_{-\infty}^{\infty} f(x)f(x+r)[F(x+r) - F(x)]^{n-2}dx \tag{1.4}$$

1.6.4 Distribution de la moyenne entre le min-max (*midrange*)

On a $M = \frac{Y_1+Y_n}{2}$, on peut trouver la fonction de répartition et la fonction de densité de cette distribution :

$$\begin{aligned} P(M \leq m) &= P(Y_n \leq -Y_1 + 2m) \\ &= \int_{-\infty}^{n} \int_{m}^{-x+2m} f_{Y_1,Y_n}(x,y)dydx \end{aligned}$$

1.6.5 Distribution conjointe de $Y_1, \ldots, Y_n$

$$F_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n) = n! \prod_{i=1}^{n} F_X(y_i) \tag{1.5}$$

Les fonctions de densité (continu) et masse (discret) conjointes sont donc

$$f_{Y_1,\ldots,Y_n}(y_1,\ldots,y_n) = n!\prod_{i=1}^{n} f_X(y_i) \tag{1.6}$$

$$P(Y_1 = y_1,\ldots,Y_n = y_n) = n!\prod_{i=1}^{n} P(X = y_i) \tag{1.7}$$

Chapitre 2

Estimation ponctuelle

Disons qu'on prend un échantillon $(X_1, ..., X_n)$ d'une population avec une distribution $F_X(x)$ avec un paramètre θ ou même un vecteur de paramètre $(\theta_1, ..., \theta_n)$. On peut estimer les paramètres avec $\hat{\theta} = T(X_1, ..., X_n)$, une statistique. C'est alors ce qu'on appelle un estimateur ponctuel du paramètre ou du vecteur de paramètre.

Les principaux problèmes de l'estimation ponctuelle sont :

- Estimer la moyenne d'une population
- Estimer la variance d'une population
- probabilité de succès p d'une expérience de Bernouilli
- Estimer la différence entre les moyennes de deux populations
- Estimer la différence de proabibilité de 2 succès Bernouilli
- Estimer le quotient des variances de deux populations

2.1 Méthodes d'estimation ponctuelle

2.1.1 Méthode des moments

On connaît le k^{e} moment centré à zéro,

$$M_k = \sum_{i=1}^{n} \frac{X_i^k}{n} \tag{2.1}$$

Celui centré à la moyenne échantillonale

$$M_k' = \sum_{i=1}^{n} \frac{(X_i - \overline{X})^k}{n} \tag{2.2}$$

Ainsi que le moment conjoint échantillonal

$$M_{k,m} = \sum_{i=1}^{n} \frac{X_i^k Y_i^m}{n} \tag{2.3}$$

La méthode consiste à sortir (selon ce qui nous permettra d'isoler les paramètres dans les moments) les différents moments de la loi de l'échantillon, et isoler les paramètres estimés.

Exemple

On a un échantillon aléatoire de taille n d'une distribution $Gamma(\alpha, \lambda)$. On doit estimer les paramètres α et λ par la méthode des moments :

1. On connaît nos moments :

$$\mu = \frac{\alpha}{\lambda}$$
$$\sigma^2 = \frac{\alpha}{\lambda^2}$$

2. Alors, on est capable d'isoler nos paramètres dans ces équations :

$$\frac{\alpha}{\lambda} = \overline{X}$$
$$\lambda = \frac{\alpha}{\overline{X}}$$

On remplace le λ dans l'équation de la variance :

$$\frac{\alpha}{\lambda^2} = \sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{n}$$
$$\frac{\alpha}{\left(\frac{\alpha^2}{\overline{X}^2}\right)} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n}$$
$$\frac{\overline{X}^2}{\alpha} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n}$$
$$\hat{\alpha} = \frac{n\overline{X}^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

On remplace notre $\hat{\alpha}$ dans l'équation de l'espérance pour trouver notre 2e paramètre :

$$\begin{aligned}\hat{\lambda} &= \frac{\frac{n\overline{X}^2}{\sum_{i=1}^{n}(X_i-\overline{X})^2}}{\overline{X}} \\ &= \frac{n\overline{X}}{\sum_{i=1}^{n}(X_i-\overline{X})^2}\end{aligned}$$

2.1.2 Méthode du maximum de vraisemblance

Nous avons encore un échantillon $(X_1,...,X_n)$ de taille n d'une population donnée avec une fonction de densité de probabilité $f(x;\theta)$ ou $P(X=x,\theta)$ si c'est discret. La fonction de vraisemblance est définie par :

$$L(\theta) = f(x_1,...,x_n;\theta) = \prod_{i=1}^{n} f(x_i,\theta) \tag{2.4}$$

ou (dans le contexte discret)

$$L(\theta) = P(X_1=x_1,...,X_n=x_n;\theta) = \prod_{i=1}^{n} P(X_i=x_i,\theta) \tag{2.5}$$

Notation On peut parfois voir l'estimateur pour le maximum de vraisemblance noté $\hat{\theta}_{MLE}$

Méthode L'estimateur $\hat{\theta}$ de vraisemblance de θ sera la valeur qui maximisera la fonction $L(\theta)$, c'est-à-dire la valeur qui pose la première dérivée égale à zéro. Si nécessaire, on peut aussi poser $\ln L(\theta)$ pour nous aider à isoler notre paramètre.

l'exemple 3, page 15 chapitre 2

Soit $(X_1,...,X_n)$ un échantillon aléatoire d'une distributin $N(\mu,\sigma^2)$. Trouver les estimateurs du maximum de vraisemblance des paramètres μ et σ^2 (Soit $\hat{\mu}_{MLE}$ et $\hat{\sigma^2}_{MLE}$).

On commence par poser $L(\theta) = \prod_{i=1}^{n} f(x;\theta)$:

$$\begin{aligned}
L(\mu,\sigma^2) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2} \\
&= \left(\frac{1}{\sqrt{2\pi}}\right)^n \sigma^{-n} \prod_{i=1}^{n} \exp\left[-\frac{1}{2}\sum_{i=1}\left(\frac{x_i-\mu}{\sigma}\right)^2\right]
\end{aligned}$$

Pour se débarasser du e, on peut poser $\ln(L(\mu,\sigma^2)$:

$$\ln L(\mu,\sigma^2) = -\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2 - \frac{n}{2}\ln(2\pi) - n\ln\sigma$$

Par la suite, on calcule la dérivée (par rapport au paramètre que l'on cherche) pour pouvoir l'égaler à 0 et trouver la valeur qui maximise $L(\theta)$:

$$\begin{aligned}
\frac{\partial \ln L(\mu,\sigma^2)}{\partial \mu} &= -\frac{1}{2}\sum_{i=1}^{n} 2\cdot\left(\frac{x_i-\mu}{\sigma}\right)\cdot\frac{1}{\sigma} \\
0 &= \sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma^2}\right) \\
0 &= \frac{\sum_{i=1}^{n} x_1}{n} \\
\hat{\mu}_{MLE} &= \frac{\sum_{i=1}^{n}}{n} \\
&= \overline{X}
\end{aligned}$$

Même principe pour $\hat{\sigma^2}_{MLE}$:

$$\frac{\partial \ln L(\mu,\sigma^2)}{\partial \sigma^2} =$$

À FINALISER PLUS TARD

Remarque importante l'estimateur a 2 propriétés importante à savoir :

- **propriété d'invariance** : si $\hat{\theta}$ est un estimateur du maximum de vraisemblance, alors $h(\hat{\theta})$ sera un estimateur du maximum de vraisemblance de $h(\theta)$.
- **L'unicité n'est pas** une propriété de l'estimateur du maximum de vraisemblance. Ça veut dire qu'il peut y avoir plusieurs solutions pour

2.1.3 Méthode Bayesienne

On a un échantillon $(X_1,...,X_n)$ d'une population qui a une fonction de répartition $F_X(x,\theta)$ et le paramètre θ suit une distribution $F_\Theta(\theta)$ à priori.

Méthode On pourra trouver notre paramètre estimé $\hat{\theta}$ ainsi :

$$\hat{\theta} = E[\Theta|X_1,...,X_n] \tag{2.6}$$

Rappel de ACT-1002 Les probabilités conditionnelles et le théorème de Bayes :

$$\begin{aligned} P(A|B) &= \frac{P(A\cap B)}{P(B)} \\ &= \frac{P(B|A)P(A)}{P(B)} \end{aligned}$$

2.2 Estimateur non biaisé

Un estimateur $\hat{\theta}$ est non-biaisé si

$$E[\hat{\theta}] = \theta \tag{2.7}$$

2.3 Estimateurs de variance minimale

2.3.1 Efficacité relative

Si on a 2 estimateurs non biaisés $\hat{\theta}_1$ et $\hat{\theta}_2$. On définit l'efficacité relative de $\hat{\theta}_1$ par rapport à $\hat{\theta}_2$ avec

$$e(\hat{\theta}_1,\hat{\theta}_2) = \frac{Var(\hat{\theta}_1)}{Var(\hat{\theta}_2)} \tag{2.8}$$

Interprétation : si $e(\hat{\theta}_1, \hat{\theta}_2) < 1$, alors $\hat{\theta}_1$ est relativement plus efficace que $\hat{\theta}_2$.

2.3.2 Estimateur non biaisé de variance minimale (MVUE)

Si on a un échantillon $(X_1, ..., X_n)$ et Ω représente l'ensemble des estimateurs non biaisés de θ, alors θ^* est un MVUE [1] si

$$Var(\theta^*) \leq Var(\hat{\theta}), \quad \forall \hat{\theta} \in \Omega$$

2.3.3 Borne de Rao-Cramer

Si un échantillon respecte toutes les conditions de régularité suivantes :

- $\frac{\partial \ln f(x;\theta)}{\partial \theta} \quad \exists \forall x$
- La dérivée de l'intégrale est égale à l'intégrale de la dérivée, i.e.

$$\frac{\partial}{\partial \theta} \int ... \int \prod_{i=1}^{n} f(x_1;\theta) dx_1, ..., dx_n = \int ... \int \frac{\partial}{\partial \theta} \prod_{i=1}^{n} f(x_i;\theta) dx_1, ..., dx_n$$

ou en considérant la statistique $\hat{\theta} = T(X_1, ..., X_n)$,

$$\begin{aligned} &\frac{\partial}{\partial \theta} \int ... \int T(x_1, ..., x_n) \prod_{i=1}^{n} f(x_1;\theta) dx_1, ..., dx_n \\ &= \int ... \int T(x_1, ..., x_n) \frac{\partial}{\partial \theta} \prod_{i=1}^{n} f(x_i;\theta) dx_1, ..., dx_n \end{aligned}$$

- $$0 < E\left[\left(\frac{\partial \ln f(X;\theta)}{\partial \theta}\right)^2\right] < \infty \quad , \forall \theta \in \Omega$$

Alors,

$$Var(\hat{\theta}) \geq \frac{1}{nE\left[\left(\frac{\partial \ln f(X;\theta)}{\partial \theta}\right)^2\right]} \tag{2.9}$$

Il faut noter qu'il existe une variante lorsqu'on a une fonction $g(\theta)$:

$$Var(\hat{g}(\theta)) \geq \frac{(g(\theta))^2}{nE\left[\left(\frac{\partial \ln f(X;\theta)}{\partial \theta}\right)^2\right]} \tag{2.10}$$

1. MVUE : Minimum-variance unbiaised estimator

2.4 Efficacité d'un estimateur

L'efficacité d'un estimateur est étroitement liée avec la borne de Rao-Cramer. Un estimateur sera dit efficace si $e(\hat{\theta}) = 1$ dans le calcul suivant :

$$e(\hat{\theta}) = \frac{1}{nVar(\hat{\theta})E\left[\left(\frac{\partial \ln f(X;\theta)}{\partial \theta}\right)^2\right]} \tag{2.11}$$

Remarque

Si $\hat{\theta}$ est efficace, alors cet estimateur est MVUE. Toutefois, si $\hat{\theta}$ est MVUE, il ne sera pas nécessairement efficace.

Identité importante

$$E\left[\left(\frac{\partial \ln f(X;\theta)}{\partial \theta}\right)^2\right] = E\left[\frac{\partial^2}{\partial \theta^2} \ln f(X;\theta)\right] \tag{2.12}$$

2.5 Estimateur consistant

Soit $\hat{\theta}_n$ est un estimateur de θ sur un échantillon n. $\hat{\theta}_n$ est un estimateur consistant seulement (et seulement si) il converge en probabilité vers θ, c'est-à-dire

$$\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| < \varepsilon) = 1 \Leftrightarrow \lim_{n\to\infty} P(|\hat{\theta}_n - \theta| \geq \varepsilon) = 0$$

$\forall \varepsilon > 0$

2.5.1 Conditions suffisantes pour la consistance

On peut dire de $\hat{\theta}_n$ que c'est un estimateur consistant de θ si

1. $E[\hat{\theta}] = \theta$
2. $\lim_{n\to\infty} Var(\hat{\theta}_n) = 0$

2.6 Estimateur exhaustif (suffisant)

Définition $\hat{\theta}$ sera un estimateur exhaustif si la fonction de densité conditionnelle $f_X(x_1,...,x_n|\hat{\Theta}) = \hat{\theta}$ est indépendante de θ.

Vérification Lorsqu'on développe la forme $f_X(x_1,...,x_n|\hat{\Theta}) = \hat{\theta}$, on ne doit pas avoir de θ dans la réponse finale, sans quoi l'estimateur est considéré non-exhaustif.

Interprétation Lorsque l'estimateur choisi $\hat{\theta}$ est exhaustif, on sait que l'échantillon sélectionné ne nous donne pas plus d'information sur le paramètre (du moins, pas plus que le fait déjà $\hat{\theta}$.

2.6.1 Théorème de factorisation de Fisher-Neymann

Soit un échantillon aléatoire $(X_1,...,X_n)$ avec une fonction de densité $f(x;.\because.)$. Alors $\hat{\theta}$ est un estimateur exhaustif de θ si (et seulement si) il existe des fonctions g et h tel que

$$\prod_{i=1}^{n} f(x_i;\theta) = g(\hat{\theta},\theta)h(x_1,...,x_n) \tag{2.13}$$

Où h ne dépend pas de θ.

Chapitre 3

Estimation par intervalles de confiance

Dans ce chapitre, au lieu d'estimer un paramètre de façon *ponctuel* (i.e. trouver une valeur précise à $\hat{\theta}$), on va plutôt bâtir un intervalle de confiance centré à μ pour s'assurer qu'un paramètre est entre 2 bornes (avec un certain % de confiance).

Objectif L'un des objectifs que l'on veut est de bâtir un intervalle de confiance dont l'étendue est minimale (*i.e. la plus petite distance possible entre* $\hat{\theta}_1$ *et* $\hat{\theta}_2$)

3.1 Définitions

Soit $(X_1, ..., X_n)$ un échantillon aléatoire d'une population avec une fonction de probabilité $f(x;\theta)$ et $\hat{\theta}_1 < \hat{\theta}_2$ les bornes inférieure et supérieure de l'estimation de θ. Alors,

Bilatéral $]\hat{\theta}_1\,,\,\hat{\theta}_2[$ est un intervalle de confiance *bilatéral* du paramètre θ avec un niveau de confiance $1-\alpha$, si

$$P(\hat{\theta}_1 < \theta < \hat{\theta}_2) = 1-\alpha$$

Unilatéral à droite $]\hat{\theta}_1\,,\,\infty[$ est un intervalle de confiance *unilatéral à droite* du paramètre θ avec un niveau de confiance $1-\alpha$, si

$$P(\theta > \hat{\theta}_1) = 1-\alpha$$

Unilatéral à gauche $]-\infty, \hat{\theta}_2[$ est un intervalle de confiance *unilatéral à gauche* du paramètre θ avec un niveau de confiance $1-\alpha$, si

$$P(\theta < \hat{\theta}_2) = 1-\alpha$$

3.2 Construction d'un intervalle de confiance

Méthode

1. On cherche une fonction $h(X_1,...,X_n,\theta)$ dont l'expression dépendra de θ et l'échantilon, mais dont la distribution ne dépendra pas de θ. Cette fonction est appelée **quantitée pivot**.
2. On pose l'égalité suivante :

   $$P(a < h\big(X_1,...,X_n,\theta\big) < b) = 1-\alpha)$$

   où a et b sont des constantes.
3. On va trouver notre intervalle de confiance en isolant θ de $h(X_1,...,X_n,\theta)$, tel que

   $$P(\hat{\theta}_1 < \theta < \hat{\theta}_2) = 1-\alpha$$

Certaines constructions d'intervalles de confiance reviennent souvent. La prochaine section discute de ces cas plus *généraux*.

3.3 IC liés à un échantillon de taille n

3.3.1 Moyenne μ d'une distribution $N(\mu,\sigma)$

Si σ^2 est connue

Il y a un seul élément qui provient de l'échantillon, c'est-à-dire $\overline{X}$:

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \backsim N(0,1) \tag{3.1}$$

Construction de l'intervalle de confiance

Alors, on a qu'à isoler la quantité dans l'intervalle :

$$P\left(-z_{\alpha/2} < \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\right) = 1-\alpha$$
$$P\left(\overline{X}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

Par conséquent, l'intervalle de confiance pour la moyenne μ est

$$\left]\overline{X}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{X}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right[$$

Le principe est le même pour tous les autres intervalles de confiance.

Remarque Pour un n grand, on peut utiliser cet intervalle pour d'autres loi que la normale.

Si σ^2 est inconnue

Lorsque la variance n'est pas connue, on va utiliser la loi Student $t_{(n-1)}$ et la quantité pivot suivante :

$$\frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t_{(n-1)} \tag{3.2}$$

3.3.2 Variance d'une distribution $N(\mu, \sigma^2)$

Si μ est connue

Lorsqu'ils sont disponibles, on va toujours prendre les paramètres de la population avant les valeurs échantionnales. Dans le cas ici, on a la moyenne μ :

$$\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi^2_{(n)} \tag{3.3}$$

Si μ est inconnue

Si on n'a pas accès à la moyenne μ, on doit passer par une autre quantité pivot :

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)} \tag{3.4}$$

Pourquoi $(n-1)$?

Lorsqu'on utilise la variance échantillonale, il faut faire une correction à $n-1$.

3.3.3 Paramètre p (probabilité) d'une expérience Bernouilli

On sait que p représente en fait la moyenne d'une Expérience de Bernouilli. La quantité pivot utilisée pour bâtir l'intervalle de confiance est donc identique à celle en 3.1 :

$$\frac{\overline{X}-p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1) \tag{3.5}$$

3.4 IC liés à 2 échantillons de taille n et m

Pour la construction des intervalles de confiance dans cette section, on considère un échantillon $(X_1,\ldots,X_n)$ avec moyenne μ_X et variane σ_X^2 ainsi qu'un 2e échantillon $(Y_1,\ldots,Y_n)$ [1] avec moyenne μ_Y et variance σ_Y^2. Il est aussi essentiel que les 2 échantillons soient indépendants.

3.4.1 Différence de 2 moyennes de distributions normales

Si σ_X^2 et σ_Y^2 sont connus

$(\overline{X}-\overline{Y})$ (la différence de moyenne) suit une loi normale avec moyenne $(\mu_X - \mu_Y)$ et variance $\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}$. Alors,

$$\frac{(\overline{X}-\overline{Y})-(\mu_X-\mu_Y)}{\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}} \sim N(0,1) \tag{3.6}$$

1. attention, on ne fait pas référence ici au statistiques d'ordre !

Si σ_X^2 et σ_Y^2 sont inconnus

On va utiliser sensiblement la même quantité pivot qu'à la section 3.3.1 lorsque les σ sont inconnus, mais cette fois en tenant compte qu'il y a 2 échantillons. On doit donc avoir une mesure qui tient compte des 2 variances échantillonnales, soit S_p^2 :

$$\begin{aligned} S_p^2 &= \frac{S_X^2}{n-1} + \frac{S_Y^2}{m-1} \\ &= \frac{(m-1)S_X^2 + (n-1)S_Y^2}{n+m-2} \end{aligned}$$

Alors, la quantité pivot est

$$\frac{(\overline{X} - \overline{Y}) - (\mu_X - \mu_Y)}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{(n+m-2)} \tag{3.7}$$

3.4.2 Quotient des variances de distributions normales

Si μ_X et μ_Y sont connus

$$\frac{\sum_{i=1}^n \left(\frac{X_i-\mu_X}{\sigma_X}\right)^2 \Big/ n}{\sum_{i=1}^m \left(\frac{Y_i-\mu_Y}{\sigma_Y}\right)^2 \Big/ m} \sim F_{n,m} \tag{3.8}$$

Exemple : on cherche le quotient $\frac{\sigma_X^2}{\sigma_Y^2}$

On sait que

$$P\left(F_{n,m;1-\frac{\alpha}{2}} < \frac{\sum_{i=1}^n \left(\frac{X_i-\mu_X}{\sigma_X}\right)^2 \Big/ n}{\sum_{i=1}^m \left(\frac{Y_i-\mu_Y}{\sigma_Y}\right)^2 \Big/ m} < F_{n,m;\frac{\alpha}{2}}\right) = 1-\alpha$$

$$P\left(\frac{m}{nF_{n,m;\frac{\alpha}{2}}} \frac{\sum_{i=1}^n (X_i-\mu_X)^2}{\sum_{i=1}^m (Y_i-\mu_Y)^2} < \frac{\sigma_X^2}{\sigma_Y^2} < \frac{m}{nF_{n,m;1-\frac{\alpha}{2}}} \frac{\sum_{i=1}^n (X_i-\mu_X)^2}{\sum_{i=1}^m (Y_i-\mu_Y)^2}\right) = 1-\alpha$$

Si μ_X et μ_Y sont inconnus

$$\frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2} \sim F_{n-1,m-1} \tag{3.9}$$

3.4.3 Différence entre 2 probabilité (p_1 et p_2) de succès Bernouilli

La quantité pivot est identique à la quantité 3.6 ,mais appliquée à une expérience ce Bernouilli :

$$\frac{(\overline{X}-\overline{Y})-(p_1-p_2)}{\sqrt{\frac{p_1(1-p_1)}{n}+\frac{p_2(1-p_2)}{m}}} \sim N(0,1) \tag{3.10}$$

3.4.4 Taille d'une échantillon

Évidemment, de la même façon qu'on isole le paramètre pour construire un intervalle de confiance, on peut nous demander la taille n d'échantillon nécessaire pour obtenir un certain niveau de confiance pour un paramètre inconnu.

3.5 IC Bayesien

On cherche à bâtir un intervalle de confiance pour l'estimateur $\hat{\theta}_{\text{BAYES}}$ qui se situera entre $\hat{\theta}_1$ (borne inférieure) et $\hat{\theta}_2$ (borne supérieure)

Méthode

1. On veut trouver la densité de

$$\begin{aligned} f_{\Theta|X_1,\ldots,X_n}(\theta|x_1,\ldots,x_n) &= \frac{f_{X_1,\ldots,X_n|\Theta}(x_1,\ldots,x_n|\theta)}{f_{X_1,\ldots,X_n}(x_1,\ldots,x_n)} \\ &= \frac{f_{X_1,\ldots,X_n|\Theta}(x_1,\ldots,x_n|\theta)}{\int_{-\infty}^{\infty} f_{X_1,\ldots,X_n|\Theta}(x_1,\ldots,x_n|\theta)} \end{aligned}$$

2. On pourra alors trouver une *quantité pivot* à utiliser pour trouver l'IC

3. On isole l'équation suivante pour trouver l'IC de Bayes :

$$P(\hat{\theta}_1 < \text{Quantité pivot} < \hat{\theta}_2) = 1 - \alpha$$

3.6 Autres intervalles de confiance

3.6.1 Moyenne ($\frac{1}{\beta}$) d'une loi Exponentielle

$$\frac{2\sum_{i=1}^{n} X_i}{\beta} \sim \chi^2_{(2n)} \tag{3.11}$$

3.6.2 Moyenne ($\frac{\alpha}{\beta}$) d'une loi Gamma

Sachant que α est connu :

$$2\beta \sum_{i=1}^{n} X_i \sim \chi^2_{(2\alpha n)} \tag{3.12}$$

3.6.3 Paramètre θ d'une fonction $f_X(x;\theta) = \theta x^{\theta-1} 1_{\{0<x<1\}}$

Soit la fonction

$$f_X(x;\theta) = \theta x^{\theta-1} 1_{\{0<x<1\}}$$

La fonction pivot à savoir est

$$\prod_{i=1}^{n} F_X(X_i;\theta) = \prod_{i=1}^{n} X_i^{\theta} \tag{3.13}$$

Avec

$$P\left(a < \prod_{i=1}^{n} X_i^{\theta} < b\right) = 1 - \alpha$$

où a et b sont des bornes (il existe plusieurs possibilités). On recherche un intervalle de confiance d'étendue minimale par analyse numérique.

Chapitre 4

Tests d'hypothèses

Introduction Jusqu'à présent, on a vu de multiples façons de procéder pour estimer les paramètres. Or, une branche importante des statistiques est l'inférence statistique dans laquelle on trouve la production de *tests d'hypothèses*. Essentiellement, l'idée est d'utiliser l'information de *l'échantillon aléatoire* pour vérifier si on doit accepter ou rejeter une hypothèse à propos du comportement d'une v.a. dans la population.

Définition : Test d'hypothèse

procédure qui utilise l'information contenue dans un échantillon aléatoire pour décider si une hypothèse H_0 (hypothèse nulle, par défaut) sur la population doit être acceptée ou rejetée au profit de H_1 (hypothèse de rejet)[a].

a. H_1 est souvent l'évènement complément de H_0.

Idée de la méthode

1. Poser les hypothèses H_0 et H_1.
2. Assumer H_0 vraie et calculer une *mesure de distance* à l'aide de l'échantillon aléatoire.
3. Conclusion :
   - Rejeter H_0 (donc accepter H_1) si la *distance* calculée à l'étape 2 est trop grande.
   - Accepter H_0 (donc rejeter H_1) si la *distance* est acceptable.

4.1 Exemples de tests

4.1.1 Test d'hypothèse à droite

Exemple 1

Une maladie est connue pour tuer 34% des personnes qui en sont atteintes. Un nouveau médicament a été testé sur un échantillon aléatoire de $n = 400$ malades et on a observé un taux de décès de 29% dont 116 décès sur 400 malades. Est-ce que ce médicament est efficace ?

On sait que $X_i \sim Bern(\theta)$ avec $\hat{\theta} = 29\%$.

On pose les hypothèses

H_0 le médicament n'est pas efficace, i.e. $\theta_{POP} = 34\%$

H_1 le médicament est efficace, i.e. $\theta_{POP} < 34\%$

On cherche la *statistique de distance* en supposant que l'hypothèse H_0 est vraie :

Selon les propriétés des estimateurs au maximum de vraisemblance,

$$\begin{aligned}
\overline{X} &= \hat{\theta}_{MLE} \underset{n\to\infty}{\longrightarrow} N(\mu_X, \frac{\sigma_X^2}{n}) \\
&= \hat{\theta}_{MLE} \underset{n\to\infty}{\xrightarrow{H_0}} N(\mu_X = 0,34, \frac{\sigma_X^2}{n} = \frac{0,34(1-0,34)}{400} = 0,000561)
\end{aligned}$$

Afin de tester H_0 contre H_1, on doit déterminer un <u>niveau de confiance</u> $(100 \times (1-\alpha)\%)$ [a] qui déterminera la *zone du possible*.

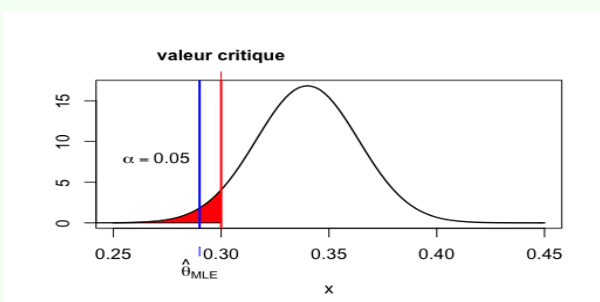
Alors, avec c la valeur critique qui détermine si H_0 est vraie,

$$
\begin{aligned}
\alpha &= P(\hat{\theta}_{MLE} < c | \Theta = \theta_{H_0}) \\
&= P\left(\frac{\hat{\theta}_{MLE} - \mu}{\sqrt{\sigma^2/n}} < \frac{c-\mu}{\sqrt{\sigma^2/n}} \middle| \Theta = 34\%\right) \\
&= \Phi\left(\frac{c - o,34}{\sqrt{0,000561}}\right) \\
\Phi^{-1}(\alpha) &= \Phi^{-1}\left(\Phi\left(\frac{c-o,34}{\sqrt{0,000561}}\right)\right) \\
\Phi^{-1}(\alpha) &= \frac{c-o,34}{\sqrt{0,000561}} \\
c &= 0,34 + \sqrt{0,000561}\Phi^{-1}(\alpha)
\end{aligned}
$$

Si on pose $\alpha = 5\%$,

$$
\begin{aligned}
c &= 0,34 + \Phi^{-1}(0,05)\sqrt{0,000561} \\
&= 0,34 + (-1,645)\sqrt{0,000561} \\
&= 0,30
\end{aligned}
$$

Interprétation : Si H_0 est vraie, alors la plus petite valeur possible pour $\hat{\theta}_{MLE}$ est 30% (c'est un test unilatéral à gauche).



Conclusion Puisque $\hat{\theta}_{MLE} = 29\% < c = 30\%$, il est peu probable au niveau de confiance 95% d'observer $\hat{\theta} = 29\%$ si H_0 est vraie. Donc, on rejette H_0 (on accepte H_1) au niveau de confiance 95% et on conclut que le médicament est efficace.

a. Souvent, α est appelé le *seuil du test* : ça représente la probabilité de rejeter H_0 alors que H_0 est vraie. Il faut choisir une petite valeur pour α, typiquement 1%, 5%, rarement plus que 10%

Exemple 2

Voici les résultats de 2 groupes qui ont passé le même examen :

X	63	49	86	74	56	85	93	32
Y	92	41	62	76	63	66		

On suppose que les notes sont distribués selon des lois normales, i.e. $X_i \sim N(\mu_1, \sigma_1^2)$ et $Y \sim N(\mu_2, \sigma_2^2)$. On veut savoir si les variances sont égales entre les 2 groupes

On pose les hypothèses :

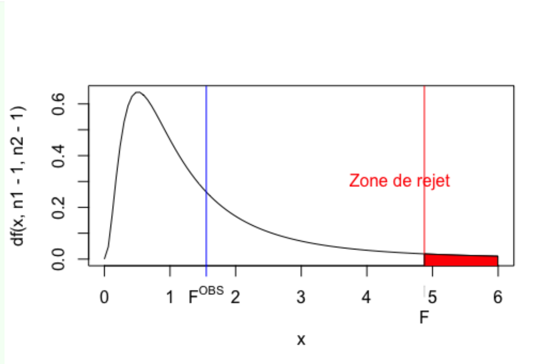
H_0 **:** $\sigma_1^2 = \sigma_2^2 \rightarrow \frac{\sigma_1^2}{\sigma_2^2} = 1$

H_1 **:** $\sigma_1^2 > \sigma_2^2 \rightarrow \frac{\sigma_1^2}{\sigma_2^2} > 1$

On va donc trouver la valeur critique en supposant que H_0 est vraie. Pour se faire, il faut qu'on utilise une quantité pivot perinente pour faire des tests de quotient de variance :

$$\begin{aligned} F &\overset{H_0}{=} \frac{n_1 S_1^2/(n_1-1)}{n_2 S_2^2/(n_2-1)} \sim F_{n_1-1,n_2-1} \\ F^{OBS} &= \ldots \\ &= 1,553 \end{aligned}$$

Alors,



On accepte H_0 au niveau de confiance $1-\alpha = 0.95$ et on conclut que les variances des 2 populations sont les mêmes.

4.1.2 Test unilatéral à gauche

Exemple 3

On reprend les mêmes données que l'exemple 2 précédent :

X	63	49	86	74	56	85	93	32
Y	92	41	62	76	63	66		

On suppose que les notes sont distribués selon des lois normales, i.e. $X_i \sim N(\mu_1, \sigma_1^2)$ et $Y \sim N(\mu_2, \sigma_2^2)$. On veut savoir si la moyenne du groupe 1 (la population) est égale à 60 à partir de l'échantillon qu'on a.

On pose les hypothèses

H_0 **:** $\mu_1 = 60$

H_1 **:** $\mu_1 > 60$

On cherche une valeur critique Si H_0 est vraie, alors

$$X_i \overset{H_0}{\sim} N(\mu_1, \sigma_1^2)$$

$$\overline{X} \underset{n\to\infty}{\overset{H_0}{\sim}} N\left(\mu_2, \frac{\sigma_2^2}{n}\right) \qquad \text{(Lemme de Neymann Pearson)}$$

Ensuite, on connait la quantité pivot de la loi normale,

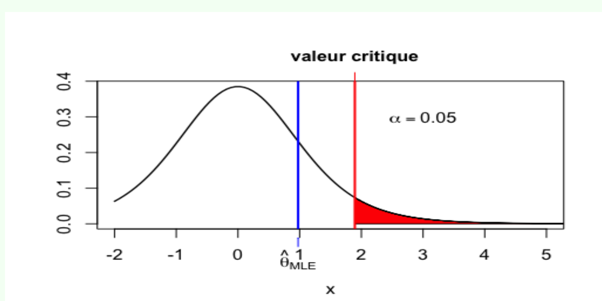
$$Z = \frac{\overline{X} - \mu_1}{\sigma_1/\sqrt{n_1}} \sim N(0,1)$$

Toutefois, puisqu'on ne connaît pas σ_1^2, il faut passer par une autre quantité. Soit la v.a. Y, où

$$Y = \frac{nS_1^2}{\sigma_1^2} \sim \chi^2(n_1 - 1)$$

Et la v.a. T, où

$$\begin{aligned} T &= \frac{Z}{\sqrt{\frac{Y}{n_1-1}}} \\ &= \frac{Z}{\frac{1}{\sqrt{n_1-1}}}\sqrt{\frac{1}{Y}} \\ &= \frac{\frac{\overline{X}-\mu_1}{\sigma_1/\sqrt{n_1}}}{\frac{1}{\sqrt{n_1-1}}}\frac{\sigma_1}{\sqrt{n_1}S_1} \\ &= \frac{\overline{X} - \mu_1(\sigma_1/\sqrt{n_1})}{(\sigma_1/\sqrt{n_1})S_1/\sqrt{n_1-1}} \\ &= \frac{\overline{X} - \mu_1}{S_1/\sqrt{n_1-1}} \sim t_{(n_1-1)} \end{aligned}$$



Alors, Si la valeur observée de T est supérieure à $t_{1-\alpha}(n_1 - 1)$, alors $\overline{X}$ est *trop grand* pour que H_0 soit vraie. Puisqu'il est très peu probable d'observer

T dans la zone de rejet si H_0 est vraie, alor on rejette H_0 au niveau de confiance $(1-\alpha)$ si

$$T = \frac{\overline{X} - \mu_1}{S_1/\sqrt{n_1 - 1}} > t_{1-\alpha}(n_1 - 1)$$

Dans notre cas :

$\overline{X} = \sum_{i=1}^{8} \frac{x_1}{8} = 67,25$	$\mu_1 = 60$
$S_1 = 386,9375$	$t_{1-0,05}(8-1) = 14,067$

Et on peut calculer T :

$$T = \frac{67,25 - 60}{386,9375/\sqrt{8-1}} = 0.975$$

Conclusion Puisque $T \leq t_{1-0.05}(7)$, on accepte $H-0$, soit que $\mu_1 = 60$ avec un niveau de confiance $(1-\alpha) = 0,95$.

4.1.3 Test bilatéral

Exemple 4

On conserve le même contexte qu'aux exemples 2 et 3 (section 4.1.2), sauf qu'on cherche à vérifier que les 2 groupes sont équivalents (i.e. la différence de moyenne est égale à 0).

Hypothèses

H_0 **:** $\mu_1 = \mu_2 \longleftrightarrow \mu_1 - \mu_2 = 0$

H_1 **:** $\mu_1 \neq \mu_2 \longleftrightarrow \mu_1 - \mu_2 \neq 0$

On sait que $\sigma_1^2 = \sigma_2^2$ (prouvé à l'exemple 2 de la section 4.1.1). Sachant cette égalité, on peut utiliser une statistique T basé sur la distribution de différence de moyenne :

$$T = \frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{n_1S_1^2+n_2S_2^2}{n_1+n_2-2}\cdot\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \sim t_{n_1+n_2-2}$$

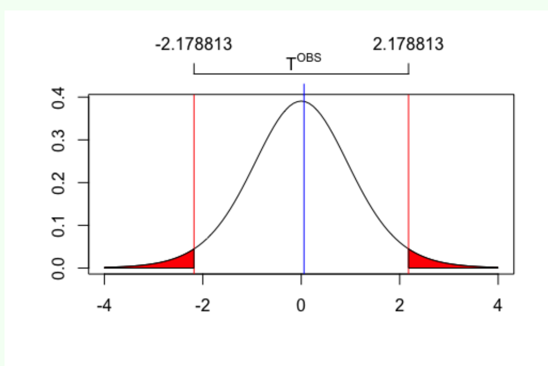
$$T \overset{H_0}{=} \frac{(\bar{X}-\bar{Y})-0}{\sqrt{\frac{n_1S_1^2+n_2S_2^2}{n_1+n_2-2}\cdot\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \sim t_{n_1+n_2-2}$$

$$T^{OBS} = 0,056$$

Ici, puisque c'est un test bilatéral (on veut que notre valeur ne soit pas plus grande ou plus petite qu'un certain niveau), on doit trouver nos 2 valeurs critiques (inférieures et supérieures) avec $\alpha = 5\%$:

$$t_{\alpha/2}(8+6-2) = -2,179$$
$$t_{1-\alpha/2}(8+6-2) = 2,179$$

Alors,



Conclusion On ne peut pas rejeter H_0, car la valeuur T^{OBS} est dans la zone probable. On conclut qu'il est possible d'observer $T = 0.,056$ sous H_0 vraie. En combinant les résultats de cet exemple et de l'exemple 2, on peut conclure que les 2 groupes sont statistiquement équivalents dans la population.

4.2 *p-value*

Note donnée au dépannage En actuariat, on fait surtout des tests d'hypothèses quand on fait de la modélisation. Les logiciels sortent la p-value.

La probabilité d'observer des valeurs aussi ou plus extrêmes que la valeur observée dans l'échantillon, sachant que H_0 est vraie.

Si la valeur p d'une statistique de test est $< \alpha$, alors H_0 peut être rejetée.

Niveau de *p-value* souhaité

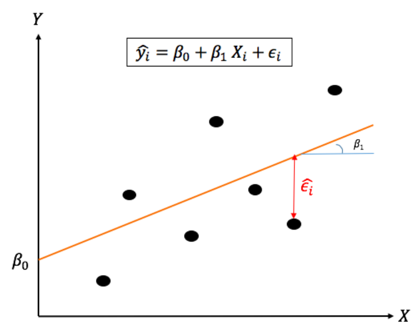
inférieure à 0,001	Excellent (***)
inférieure à 0,05	Suffisant (**)
inférieure à 0,1	peu (*)

En pratique, on calcule la *p-value*, et ensuite on décide si c'est suffisant pour nous.

Chapitre 5

Régression linéaire simple

5.1 Introduction



La régression linéaire sert à bâtir des modèles de prévision probabiliste. De façon générale, on a

Y variable dépendante (réponse ou *output*)

$X_1, \ldots, X_n$ Variable(s) indépendante(s) [1] ou explicative(s) (*input*)

$\beta_0, \beta_1, \ldots, \beta_p$ paramètres du modèle de régression à estimer

Modèle de régression linéaire simple

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \tag{5.1}$$

1. les variables sont indépendantes par rapport à Y, mais ne sont pas nécessairement indépendantes entre-elles.

Mesures de performances du modèle On cherche à trouver une mesure pour calculer la distance globale entre les points (observations) et la droite de régression estimée. 3 mesures sont analysées :

(1) On peut calculer l'erreur totale (ET), soit

$$ET = \sum_{i=1}^{n} \varepsilon_i$$

Cette méthode n'est pas robuste, car les erreur négatives viennent compenser pour les erreurs positives.

(2) L'erreur absolue (EA), soit

$$EA = \sum_{i=1}^{n} |\varepsilon_i|$$

Cette méthode est très robuste. Mais la méthode est très complexe mathématiquement. En effet, pour trouver la valeur des paramètres qui minimisent cette erreur absolue, il faut dériver la valeur absolue...

(3) Erreur quadratique [2] (EQ), soit

$$EQ = \sum_{i=1}^{n} \varepsilon_i^2$$

Cette méthode est très robuste, est simple mathématiquement, et converge généralement vers la meilleure droite. Toutefois, l'erreur quadratique est fortement affectée par des valeurs extrêmes. À moins d'avis contraire, c'est l'erreur quadratique qui sera utilisée pour le reste de ce chapitre.

5.2 Estimation des paramètres

On va estimer β_0 et β_1 par $\hat{\beta}_0$ et $\hat{\beta}_1$, soit les paramètres qui vont minimiser la fonction d'erreur quadratique (i.e. les paramètres qui vont nous donner la meilleure courbe de régression). Un peu comme la méthode du MLE (section 2.1.2), on va dériver la fonction à minimiser par rapport à chacun des paramètres pour obtenir une forme générale.

2. Cette erreur est parfois appelée *Sum of squared errors*.

Quelques formules On peut démontrer [3] que

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n X_i Y_i - n\bar{X}\bar{Y}}{\sum_{i=1}^n X_i^2 - n\bar{X}^2} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} = \frac{S_{XY}}{S_{XX}} \tag{5.2}$$

et

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \tag{5.3}$$

5.2.1 Remarques importantes

(1) On note $\hat{\varepsilon}_i$ le *résidu* généré par le modèle estimé. Par définition,

$$\begin{aligned} \hat{\varepsilon}_i &= Y_i - \hat{Y}_i \\ &= Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i \end{aligned}$$

(2) Le *Centre de gravité* (i.e. le point $(\bar{X}, \bar{Y})$ se trouve toujours dans la droite de régression [4].

Démonstration. On part de l'équation (5.3),

$$\begin{aligned} \hat{\beta}_0 &= \bar{Y} - \hat{\beta}_1 \bar{X} \\ \bar{Y} &= \hat{\beta}_0 + \hat{\beta}_1 \bar{X} \end{aligned}$$

Alors, nécessairement le point $(\bar{X}, \bar{Y})$ est dans la droite de régression. □

(3) Une notation spéciale est parfois utilisée pour représenter le numérateur et le dénominateur de l'estimateur de β_1 :

3. les preuves pour les estimateurs $\hat{\beta}_1$ et $\hat{\beta}_1$ se trouvent en Annexe A.2.
4. Ce résultat est fondamental et est utilisé dans plusieurs autres preuves de régression.

(3.1) Pour le dénominateur[5], on note

$$\begin{aligned}
S_{XX} &= \sum_{i=1}^{n}(X_i-\bar{X})^2 \\
&= \sum_{i=1}^{n}(X_i^2-2X_i\bar{X}+\bar{X}^2) \\
&= \sum_{i=1}^{n}X_i^2-2\bar{X}\frac{\sum_{i=1}^{n}X_i}{n}n+n\bar{X}^2 \\
&= \sum_{i=1}^{n}X_i^2-2n\bar{X}^2+n\bar{X}^2 \\
&= \sum_{i=1}^{n}X_i^2-n\bar{X}^2
\end{aligned}$$

(3.2) Pour le numérateur, on note

$$\begin{aligned}
S_{YX} &= \sum_{i=1}^{n}(Y_i-\bar{Y})(X_i-\bar{X}) \\
&= \sum_{i=1}^{n}(X_iY_i-X_i\bar{Y}-Y_i\bar{X}+\bar{Y}\bar{X}) \\
&= \sum_{i=1}^{n}(X_iY_i-n\bar{X}\bar{Y}-n\bar{Y}\bar{X}+\bar{Y}\bar{X}) \\
&= \sum_{i=1}^{n}X_iY_i-n\bar{X}\bar{Y}\cancel{-n\bar{X}\bar{Y}+n\bar{X}\bar{Y}} \\
&= \sum_{i=1}^{n}X_iY_i-n\bar{X}\bar{Y}
\end{aligned}$$

(4) La somme des résidus de <u>tout</u> modèle de régression linéaire (simple) est **nulle**.

5. On peut aussi avoir $S_{YY}=\sum_{i=1}^{n}(Y_i-\bar{Y})^2=\ldots=\sum_{i=1}^{n}Y_i^2-n\bar{Y}^2$

Démonstration.

$$
\begin{aligned}
\sum_{i=1}^{n} \hat{\varepsilon}_i &= \sum_{i=1}^{n} (Y_i - \hat{Y}_i) \\
&= \sum_{i=1}^{n} (Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i)) \\
&= \sum_{i=1}^{n} (Y_i - ((\bar{Y} - \hat{\beta}_1 \bar{X}) + \hat{\beta}_1 X_i)) \quad \text{, Selon l'équation (5.3)} \\
&= \sum_{i=1}^{n} Y_i - n\bar{Y} + n\bar{X}\hat{\beta}_1 - \hat{\beta}_1 \sum_{i=1}^{n} X_i \\
&= \cancel{\sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} Y_i} + \cancel{\hat{\beta}_1 \sum_{i=1}^{n} X_i - \hat{\beta}_1 \sum_{i=1}^{n} X_i} \\
&= 0
\end{aligned}
$$

□

Exemple numérique de p.12 des notes de cours

On a les données suivantes :

i	X_i	Y_i	X_i^2	X_iY_i
1	2	2	4	4
2	3	5	9	15
3	6	3	36	18
4	9	6	81	56
5	12	5	144	60
Totaux	32	21	274	151

On peut calculer $\hat{\beta}_1$:

$$\begin{aligned}\hat{\beta}_1 &= \frac{\sum_{i=1}^n X_i Y_i - n\overline{XY}}{\sum_{i=1}^n X_i^2 n\overline{X}^2} \\ &= \frac{151 - 5 \times \frac{32}{5} \times \frac{21}{5}}{274 - 5 \times \left(\frac{32}{5}\right)^2} \\ &\approx 0,2399\end{aligned}$$

Vérifier les prévisions

i	X_i	Y_i	$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$	$\hat{\varepsilon} = Y_i - \hat{Y}_i$
1	2	2	3,1145	-1,1445
2	3	5	3,3844	1,6156
3	6	3	4,1041	-1,1041
4	9	6	4,8238	1,1762
5	12	5	5,5435	-0,5435
				$\approx 0,0003$

Le même genre de problème peut se faire dans R aussi :

5.3 Régression passant par l'origine

Dans certaines situations, il est possible que l'on souhaite forcer la droite de régression à passer par le point $(0,0)$. Par exemple,

Y : La consommation d'une voiture en litre d'essence

X : le nombre de kilomètres parcourus par cette voiture.

Dans ce cas, $x = 0 \leftarrow y = 0$, et le modèle sera

$$Y_i = \beta X_i + \varepsilon_i$$

Estimation de $\hat{\beta}$

$$\begin{aligned} EQ(\beta) &= \sum_{i=1}^{n} \varepsilon_i^2 \\ &= \sum_{i=1}^{n} (Y_i - \beta X_i)^2 \end{aligned}$$

Comme pour l'estimation de β_1 et β_2, on va dériver l'erreur quadratique par β :

$$\begin{aligned} \frac{\partial EQ(\hat{\beta})}{\partial \beta} &= 0 \\ \hat{\beta} &= \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2} \end{aligned}$$

5.4 Caractéristiques et postulats sur ε

Caractéristiques du terme d'erreur ε

(1) Capture les effets des autres variables X non-incluses dans le modèle.

(2) Capture aussi le fait que la plupart des phénomènes ne se résument généralement par une simple droite.

4 grands postulats

1. Par définition,

$$E[\varepsilon_i] = 0$$

2. Par hypothèse [6],

$$Var(\varepsilon_i) = \sigma^2$$

3. Par hypothèse,

$$Cov(\varepsilon_i, \varepsilon_j) = 0 \quad , i \neq j$$

4. Par hypothèse,

$$\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$$

6. Hypothèse qu'on fera tomber (dans certains cas) dans le cours de Modèles linéaires.

FIGURE 5.1 – Résumé des 2 propriétés des estimateurs $\hat{\beta}_0$ et $\hat{\beta}_1$

	$\hat{\beta}_0$	$\hat{\beta}_1$
Exactitude	$E[\hat{\beta}_0] = \beta_0$	$E[\hat{\beta}_1] = \beta_1$
Précision	$Var(\hat{\beta}_0) = \frac{\sigma^2}{S_{XX}}$	$Var(\hat{\beta}_1) = \frac{\sigma^2}{n} + \frac{\bar{X}^2\sigma^2}{S_{XX}}$

5.5 Propriétés des estimateurs du modèle

5.5.1 Exactitude

$\hat{\beta}_0$ et $\hat{\beta}_1$ sont des estimateurs exacts (sans biais)

Démonstration. On commence par démontrer pour $\hat{\beta}_1$:

$$\begin{aligned}
E[\hat{\beta}_1] &= E_Y\left[\frac{S_{XY}}{S_{XX}}\right] \\
&= E_Y\left[\frac{\sum_{i=1}^{n}(Y_i-\bar{Y})(X_i-\bar{X})}{S_{XX}}\right] \\
&= \frac{1}{S_{XX}}\sum_{i=1}^{n}(X_i-\bar{X})E[Y_i-\bar{Y}] \\
&= \frac{1}{S_{XX}}\sum_{i=1}^{n}(X_i-\bar{X})E[Y_i]-E[\bar{Y}] \\
&\text{par la remarque (2) de la section 5.2.1,} \\
&= \frac{1}{S_{XX}}\sum_{i=1}^{n}(X_i-\bar{X})(\beta_0+\beta_1X_i-(\beta_0+\beta_1\bar{X})) \\
&= \frac{\sum_{i=1}^{n}(X_i-\bar{X})}{S_{XX}}\beta_1(X_i-\bar{X}) \\
&= \beta_1\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{S_{XX}} \\
&= \beta_1\frac{\cancel{S_{XX}}}{\cancel{S_{XX}}} \\
&= \beta_1
\end{aligned}$$

Sachant que $\hat{\beta}_1$ est exact, la démonstration pour $\hat{\beta}_0$ est simplifiée :

$$\begin{aligned}
E[\hat{\beta}_0] &= E_Y[\bar{Y} - \hat{\beta}_1 \bar{X}] \\
&= E[\bar{Y}] - \bar{X}E[\hat{\beta}_1] \\
&\text{par la remarque (2) de la section 5.2.1,} \\
&= \beta_0 + \cancel{\beta_1 \bar{X}} - \cancel{\bar{X}\beta_1} \\
&= \beta_0
\end{aligned}$$

□

5.5.2 Précision

On peut démontrer [7] que

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{S_{XX}} \tag{5.4}$$

et

$$Var(\hat{\beta}_0) = \frac{\sigma^2}{n} + \frac{\bar{X}^2\sigma^2}{S_{XX}} \tag{5.5}$$

Conclusion pour l'analyse *locale* du modèle Une grande variance de $\hat{\beta}_0$) et/ou $Var(\hat{\beta}_1)$ implique une *mauvaise* précision. Il est difficile de dire en *absolu* si la variance est grande ou petite. On aura besoin de la théorie sur les test d'hypothèses (voir la section 5.6) pour déterminer si ces variances sont grandes ou petites.

5.5.3 Remarque sur la Covariance entre les estimateurs

$$\begin{aligned}
Cov(\hat{\beta}_0, \hat{\beta}_1) &= Cov(\bar{Y} - \hat{\beta}_1 \bar{X}, \hat{\beta}_1) \\
&= \underbrace{Cov(\bar{Y}, \hat{\beta}_1)}_{=0} - \bar{X} \underbrace{Cov(\hat{\beta}_1, \hat{\beta}_1)}_{Var(\hat{\beta}_1)} \\
&= -\bar{X} Var(\hat{\beta}_1)
\end{aligned}$$

7. Les démonstrations se trouvent à l'annexe A.3.

$$Cov(\hat{\beta}_0, \hat{\beta}_1) = \frac{-\bar{X}\sigma^2}{S_{XX}} \tag{5.6}$$

5.6 Intervalles de confiance et tests d'hypothèses sur β_0 et β_1

Contexte Dans ce cas-ci, nous utiliserons les intervalles de confiance et test d'hypothèses pour l'analyse locale de la qualité du modèle de régression obtenu. En d'autres mots, ces outils vont nous servir pour déterminer si $Var(\hat{\beta}_0)$ et/ou $Var(\hat{\beta}_1)$ est *grande* ou *petite* en vue de quantifier le niveau de précision du modèle.

Remarques importantes

- Jusqu'à maintenant, nous n'avons fait aucune hypothèse quant à la distribution des ε_i.
- Par contre, pour obtenir des intervalles de confiance et/ou tests d'hypothèses, il sera nécesaire de greffer une loi à ε_i.

On peut résumer les 4 postulats de ε avec

$$\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$$

Conséquences de l'hypothèse de normalité des ε_i

(1)

$$(Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i) \sim N(\beta_0 + \beta_1 X_i, \sigma^2) \quad i = 1, \ldots, n$$

Attention ! les v.a. sont indépendantes, mais pas identiquement distribués.

(2)

$$(\hat{\beta}_1 = \sum_{i=1}^{n} W_i^* Y_i) \sim N\left(\beta_0, \frac{\sigma^2}{n} + \frac{\bar{X}^2\sigma^2}{S_{XY}}\right)$$

$$(\hat{\beta}_1 = \sum_{i=1}^{n} W_i Y_i) \sim N\left(\beta_1, \frac{\sigma^2}{S_{XX}}\right)$$

(3)

$$\hat{\sigma^2} = \frac{EQ}{n-2} = \frac{\sum_{i=1}^{n} \hat{\varepsilon_i^2}}{n-2}$$
$$= \frac{\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2}{n-2}$$
$$= \frac{\sum_{i=1}^{n} (Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i))^2}{n-2}$$

est un estimateur sans biais de σ^2. On utilise $n-2$ au dénominateur, car on a 2 paramètres à estimer.

(4) On peut aussi montrer que

$$\left(\frac{EQ}{\sigma^2}\right) \sim \chi^2(n-2)$$

5.6.1 Intervalle de confiance pour β_1

Puisque, selon la conséquence (2), on a que

$$\left(\frac{\hat{\beta}_1 - \beta_1}{\sqrt{\sigma^2/S_{XX}}}\right) \sim N(0,1) \tag{5.7}$$

Parce que σ^2 est généralement inconnu, et qu'on doit par conséquent l'estimer par $\hat{\sigma}^2 = \frac{EQ}{n-2}$, on aura que [8]

$$T = \left(\frac{\hat{\beta}_1 - \beta_1}{\sqrt{\hat{\sigma}^2/S_{XX}}}\right) \sim t(n-2) \tag{5.8}$$

Et on a l'intervalle de confiance (à $1-\alpha$). On a donc qu'à isoler β_1 dans l'inter-

8. En pratique, il est maintenant facile d'avoir plus de 30 observations, la loi student et normale se confondent.

valle de confiance :

$$\begin{aligned}
\Pr\left(T \in \left[-t_{\alpha/2}(n-2), t_{\alpha/2}(n-2)\right]\right) &= 1-\alpha \\
\Pr\left(\frac{\hat{\beta}_1 - \beta_1}{\sqrt{\hat{\sigma}^2/S_{XX}}} \in \left[-t_{\alpha/2}(n-2), t_{\alpha/2}(n-2)\right]\right) &= 1-\alpha \\
\Pr\left(\hat{\beta}_1 - \beta_1 \in \left[-\sqrt{\frac{\hat{\sigma^2}}{S_{XX}}} t_{\alpha/2}(n-2), \sqrt{\frac{\hat{\sigma^2}}{S_{XX}}} t_{\alpha/2}(n-2)\right]\right) &= 1-\alpha \\
\Pr\left(\beta_1 \in \left[\hat{\beta}_1 - \sqrt{\frac{\hat{\sigma^2}}{S_{XX}}} t_{\alpha/2}(n-2), \hat{\beta}_1 + \sqrt{\frac{\hat{\sigma^2}}{S_{XX}}} t_{\alpha/2}(n-2)\right]\right) &= 1-\alpha
\end{aligned}$$

Conclusion Un intervalle de confiance au niveau $(1-\alpha)$ pour β_1 est :

$$\left[\hat{\beta}_1 - \sqrt{\frac{\hat{\sigma^2}}{S_{XX}}} t_{\alpha/2}(n-2), \hat{\beta}_1 + \sqrt{\frac{\hat{\sigma^2}}{S_{XX}}} t_{\alpha/2}(n-2)\right]$$

Ou encore (autre façon de l'exprimer) :

$$\hat{\beta}_1 \pm \sqrt{\frac{\hat{\sigma^2}}{S_{XX}}} t_{\alpha/2}(n-2)$$

5.6.2 Intervalle de confiance pour β_0

De manière complètement similaire à l'intervalle de confiance pour β_1, on trouve l'intervalle de confiance pour β_0 à partir de (2) :

$$\left(\frac{\hat{\beta}_0 - \beta_0}{\sqrt{\frac{\sigma^2}{n} + \frac{\bar{X}^2\sigma^2}{S_{XX}}}}\right) \sim N(0,1)$$

Puisqu'on estime $\sigma^2 = \hat{\sigma}^2 = \frac{EQ}{n-2}$, on a

$$\left(\frac{\hat{\beta}_0 - \beta_0}{\sqrt{\frac{\hat{\sigma}^2}{n} + \frac{\bar{X}^2 \hat{sigma}^2}{S_{XX}}}} \right) \sim t(n-2)$$

Alors, on a

$$\Pr\left(\beta_0 \in \left[\hat{\beta}_0 \pm \sqrt{\frac{\hat{\sigma}^2}{n} + \frac{\bar{X}^2 \hat{sigma}^2}{S_{XX}}} t_{\alpha/2}(n-2) \right] \right) = 1 - \alpha$$

Donc, un IC au niveau $(1-\alpha)$ pour β_0 est :

$$\left[\hat{\beta}_0 \pm \sqrt{\frac{\hat{\sigma}^2}{n} + \frac{\bar{X}^2 \hat{sigma}^2}{S_{XX}}} t_{\alpha/2}(n-2) \right]$$

Exemple : (avancé, ne sera pas testé)

On a $n = 14$ observations. On a estimé $\hat{\beta}_0 = 68.494$ et $\hat{\beta}_1 = 0.468$. Alors, on a

$$\begin{aligned} Var(\hat{\beta}) &= \begin{bmatrix} Var(\hat{\beta}_0) & Cov(\hat{\beta}_0, \hat{\beta}_1) \\ Cov(\hat{\beta}_0, \hat{\beta}_1) & Var(\hat{\beta}_1) \end{bmatrix} \\ &= \begin{bmatrix} 66.8511 & -1.2544 \\ -1.2544 & 0.0237 \end{bmatrix} \end{aligned}$$

Avec un intervalle de confiance à 95%,

$$\begin{aligned} \beta_0 &\in [50.67, 86.318] \\ \beta_1 &\in [0.1324, 0.8036] \end{aligned}$$

On peut dire qu'à 95% de niveau de confiance, on peut dire qu'il est pertinent d'utiliser la régression, car l'intervalle de β_1 ne contient pas zéro (i.e. X influe sur le comportement de Y).

5.7 Tests d'hypothèse *Locaux* pour les estimateurs

Souvent, on souhaite utiliser les tests d'hypothèse en régression linéaire simple pour répondre à 2 questions principales :

1. Est-ce que β_0 est significativement différent de 0 ? [9]
2. Est-ce que β_1 est significativement différent de 0 ? [10]

premier test, avec β_0

H_0 **:** $\beta_0 = \beta_0^*$, où β_0^* est une constante

H_1 **:** $\beta_0 \neq \beta_0^*$

On utilise la statistique t, où

$$t = \frac{\hat{\beta}_0 - \beta_0^*}{\sqrt{\frac{\hat{\sigma^2}}{n} + \frac{\bar{X}^2 \hat{\sigma^2}}{S_{XX}}}}$$

Qui est une *distance* standarisée entre $\hat{\beta}_0$ et β_0^*.

On rejète H_0 au niveau de confiance $(1-\alpha)$ si

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

deuxième test, avec β_1

H_0 **:** $\beta_1 = \beta_1^*$, où β_1^* est une constante

H_1 **:** $\beta_1 \neq \beta_1^*$

On utilise la statistique t, où

$$t = \frac{\hat{\beta}_1 - \beta_1^*}{\sqrt{\hat{\sigma}^2 / S_{YX}}}$$

Qui est une *distance* standarisée entre $\hat{\beta}_1$ et β_1^*.

On rejète H_0 au niveau de confiance $(1-\alpha)$ si

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

9. Si le résultat du test est non, on va plutôt utiliser une régression qui passe par l'origine, c'est-à-dire $Y_i = \beta_1 X_i + \varepsilon_i$ (tel que vu à la section 5.3).

10. Si le résultat du test est non, considérer le modèle $Y_i = \beta_0 + \varepsilon_i$,i.e. on utilise l'estimation vu en première partie (chapitre 2). Ça signifie aussi que X n'a pas d'effet sur Y.

Annexe A

Preuves

A.1 Démonstrations des distributions

A.1.1 Distribution de $\frac{\overline{X}-\mu}{S/\sqrt{n}}$

$$\frac{\overline{X}-\mu}{S/\sqrt{n}} \backsim t_{(n-1)}$$

Démonstration. On sait que [1]

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$$

et

$$\frac{(n-1)S_{n-1}^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

Alors,

1. pour la preuve qui suit, S correspond à la variance échantillonale représentée avec $n-1$.

$$
\begin{aligned}
T &= \frac{\bar{X}-\mu}{S/\sqrt{n}} \\
&= \frac{\bar{X}-\mu}{S/\sqrt{n}} \cdot \frac{\sigma}{\sigma} \\
&= \underbrace{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}_{N(0,1)} \frac{\sigma}{S} \\
&= Z\frac{\sigma}{S} \quad \text{, où } Z \text{ suit une loi } N(0,1) \\
&= \frac{Z}{S/\sigma} \\
&= \frac{Z}{\frac{\sqrt{n-1}}{\sqrt{n-1}}\frac{S}{\sigma}} \\
&= \frac{Z}{\frac{\sqrt{n-1}S}{\sigma}\frac{1}{\sqrt{n-1}}} \\
&= \frac{Z}{\sqrt{\underbrace{\frac{(n-1)S^2}{\sigma^2}}_{\chi^2_{(n-1)}}\frac{1}{\sqrt{n-1}}}} \\
&= \frac{Z}{\sqrt{\frac{\chi^2_{(n-1)}}{n-1}}}
\end{aligned}
$$

$$T \sim t_{(n-1)}$$

$\square$

A.2 Estimateur du modèle de Régression linéaire simple

En partant de la fonction d'erreur quadratique

$$EQ = \sum_{i=1}^{n} \varepsilon_i^2$$

Et sachant la formule du modèle de régression linéaire

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

On peut trouver que

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{X})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

Et

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

Démonstration. Avec la méthode des moindres carré, on veut trouver les paramètres β_0 et β_1 qui vont minimiser la fonction d'erreur quadratique, i.e

$$\begin{aligned}
EQ(\beta_0, \beta_1) &= \sum_{i=1}^{n} \varepsilon_i^2 \\
&= \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \\
&= \sum_{i=1}^{n} (Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i))^2
\end{aligned}$$

Pour trouver chacun des 2 paramètres, on va dériver par rapport au paramètre, puis poser la dérivée égale à zéro pour finalement résoudre les 2 équations, 2 inconnus.

Si on commence par $\hat{\beta}_0$,

$$
\begin{aligned}
\frac{\partial}{\partial \beta_0} EQ(\beta_0, \beta_1) &= \frac{\partial}{\partial \beta_0}\left[\sum_{i=1}^{n}(Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i))^2\right] \\
0 &= 2\sum_{i=1}^{n}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)\cdot(-1) \\
&= \sum_{i=1}^{n} Y_i - n\hat{\beta}_0 - \hat{\beta}_1 \sum_{i=1}^{n} X_i \\
n\hat{\beta}_0 &= \sum_{i=1}^{n} Y_i - \hat{\beta}_1 \sum_{i=1}^{n} X_i \\
\hat{\beta}_0 &= \frac{\sum_{i=1}^{n} Y_i - \hat{\beta}_1 \sum_{i=1}^{n} X_i}{n} \\
&= \bar{Y} - \hat{\beta}_1 \bar{X}
\end{aligned}
$$

On peut descendre la même preuve pour $\hat{\beta}_1$ (un peu plus longue) :

$$
\begin{aligned}
\frac{\partial}{\partial \beta_1} EQ(\beta_0, \beta_1) &= \frac{\partial}{\partial \beta_1}\left[\sum_{i=1}^{n}(Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i))^2\right] \\
0 &= 2\sum_{i=1}^{n}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)\cdot(-X_i) \\
&= \sum_{i=1}^{n}(Y_i X_i - \hat{\beta}_0 X_i - \hat{\beta}_1 X_i^2) \\
&= \sum_{i=1}^{n}(Y_i X_i) - \hat{\beta}_0 \sum_{i=1}^{n} X_i - \hat{\beta}_1 \sum_{i=1}^{n} X_i^2 \\
&= \sum_{i=1}^{n}(Y_i X_i) - (\bar{Y} - \hat{\beta}_1 \bar{X}) \sum_{i=1}^{n} X_i - \hat{\beta}_1 \sum_{i=1}^{n} X_i^2 \\
&= \sum_{i=1}^{n} Y_i X_i - \bar{Y}\sum_{i=1}^{n} X_i + \hat{\beta}_1 \bar{X} \sum_{i=1}^{n} X_i - \hat{\beta}_1 \sum_{i=1}^{n} X_i^2 \\
&= \sum_{i=1}^{n} Y_i X_i - n\bar{Y}\bar{X} - \hat{\beta}_1 \left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right) \\
\hat{\beta}_1 &= \frac{\sum_{i=1}^{n} Y_i X_i - n\bar{X}\bar{Y}}{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2}
\end{aligned}
$$

$\square$

A.3 Précision des estimateurs beta

À la section 5.5.2, on précise que

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{S_{XX}}$$

et

$$Var(\hat{\beta}_0) = \frac{\sigma^2}{n} + \frac{\bar{X}\sigma^2}{S_{XX}}$$

Démonstration. On commence par démontrer pour $\hat{\beta}_1$:

$$\begin{aligned}
Var_Y(\hat{\beta}_1) &= Var_Y\left(\sum_{i=1}^{n}\left(\frac{X_i - \bar{X}}{S_{XX}}\right)Y_i\right)\\
&= Var_Y\left(\frac{\sum_{i=1}^{n}(X_i - \bar{X})\cdot(\beta_0 + \beta_1 X_i + \varepsilon_i)}{S_{XX}}\right)\\
&= \frac{1}{S_{XX}^2}Var_Y\left(\sum_{i=1}^{n}(X_i - \bar{X})(\beta_0 + \beta_1 X_i) + \sum_{i=1}^{n}(X_i - \bar{X})\varepsilon_i\right)
\end{aligned}$$

La partie de droite à l'intérieur de la variance est une constante.
On applique les propriétés de la variance :

$$= \frac{1}{S_{XX}^2}Var_Y\left(\sum_{i=1}^{n}(X_i - \bar{X})\varepsilon_i\right)$$

Par l'hypothèse d'indépendance (postulat 3, section 5.4), on a

$$\begin{aligned}
&= \frac{1}{S_{XX}^2}\sum_{i=1}^{n}Var_Y((X_i - \bar{X})\varepsilon_i)\\
&= \frac{1}{\cancel{S_{XX}^2}}\cancel{\sum_{i=1}^{n}(X_i - \bar{X})^2}Var_Y(\varepsilon_i)\\
&= Var(\varepsilon_i)
\end{aligned}$$

par le postulat 2,

$$= \sigma^2$$

Pour ce qui est de la preuve de $\hat{\beta}_0$, la preuve peut être très longue, mais il y a moyen de prendre un raccourci (via la propriété (2) de la section 5.2.1) pour se débarasser facilement de la covariance dans la preuve ci-dessous :

$$
\begin{aligned}
Var_Y(\hat{\beta}_0) &= Var_Y(\bar{Y} - \beta_1 \bar{X}) \\
&= Var_Y(\bar{Y}) + Var_Y(\beta_1 \bar{X}) - 2Cov(\bar{Y}, \beta_1 \bar{X}) \\
&= Var_Y\left(\frac{\sum_{i=1}^n Y_i}{n}\right) + \bar{X}^2 Var_Y(\beta_1) - 2\bar{X}Cov(\bar{Y}, \hat{\beta}_1) \\
&= \frac{1}{n^2} Var_Y(\underbrace{\sum_{i=1}^n \beta_0 + \beta_1 X_i}_{\text{Constantes}} + \varepsilon_i) + \frac{\bar{X}^2 \sigma^2}{S_{XX}} - \underbrace{2\bar{X}Cov(\bar{Y}, \hat{\beta}_1)}_{=0} \\
&= \frac{nVar(\varepsilon_i)}{n^2} + \frac{\bar{X}^2 \sigma^2}{S_{XX}} \\
&= \frac{\sigma^2}{n} + \frac{\bar{X}^2 \sigma^2}{S_{XX}}
\end{aligned}
$$

□