

# Statistique (MATH-F-315, Cours #6)

Thomas Verdebout

Université Libre de Bruxelles

## Plan de la partie Statistique du cours

1. Introduction.
2. Théorie de l'estimation.
3. Tests d'hypothèses et intervalles de confiance.
4. Régression et ANOVA.

## Régression simple

Imaginons que nous possédons des échantillons  $Y_1, \dots, Y_n$  et  $X_1, \dots, X_n$  associés à deux variables aléatoires. Le modèle de régression linéaire simple postule un lien linéaire entre les  $Y_i$  et les  $X_i$  de la forme

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad i = 1, \dots, n,$$

où

- $\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} \in \mathbb{R}^2$  est un paramètre (inconnu) : le *paramètre de régression*, composé d'une *pente*  $\beta_1$  et d'une ordonnée à l'origine  $\beta_0$
- $\varepsilon_1, \dots, \varepsilon_n$  sont des variables aléatoires non observées.

Dans le cas dit "général", les hypothèses traditionnellement faites sur les erreurs sont que  $E[\varepsilon_i] = 0 \forall i$  et

$$E[\varepsilon_i \varepsilon_j] = \begin{cases} \sigma^2 & i = j \\ 0 & i \neq j. \end{cases}$$

Dans le cas dit "gaussien", on renforce cette hypothèse en  $\varepsilon_i$  i.i.d.  $\mathcal{N}(0, \sigma^2)$ ,  $i = 1, \dots, n$ .

## Régression simple

### Remarques:

- ▶ Pour simplifier, nous supposons dans la suite que les variables  $X_1, \dots, X_n$  sont des constantes  $x_1, \dots, x_n$ .
- ▶ Linéarité: Le modèle appartient à la classe des *modèles linéaires* : la moyenne de l'observation  $Y_i$ ,  $E[Y_i] = E[\beta_0 + \beta_1 x_i + \varepsilon_i] = \beta_0 + \beta_1 x_i + E[\varepsilon_i] = \beta_0 + \beta_1 x_i$  est une fonction linéaire du paramètre  $\beta \Rightarrow$  Une augmentation d'une unité de  $x_i$  correspond une augmentation de  $\beta_1$  unités de  $E[Y_i]$ .
- ▶ Homoscédasticité: La variance  $\sigma^2$  est la même pour toutes les observations (et ne dépend donc pas des  $x_i$ )

$$\text{Var}(Y_i) = \text{Var}(\varepsilon_i) = \sigma^2 \quad \text{pour tout } i.$$

La violation de cette hypothèse sera appelée **problème d'hétéroscédasticité**.

- ▶ Non-corrélation: Les observations sont mutuellement non corrélées (surtout important pour des séries temporelles)

$$\text{Cov}(Y_i, Y_j) = \text{Cov}(\varepsilon_i, \varepsilon_j) = 0 \quad \text{pour tout couple } i \neq j.$$

Cette non-corrélation se transforme en indépendance dans le "cas gaussien". La violation de cette hypothèse sera appelée **problème d'autocorrélation**.

## Régression simple

### Estimation des paramètres

La méthode du maximum de vraisemblance s'impose dans le cas gaussien. Puisque les  $Y_i$  sont indépendants, et que  $Y_i \sim \mathcal{N}(\beta_0 + \beta_1 x_i, \sigma^2)$ , la vraisemblance  $L_{\beta_0, \beta_1, \sigma^2}(\mathbf{Y})$  s'écrit ( $\mathbf{Y} := (Y_1, \dots, Y_n)'$ )

$$L_{\beta_0, \beta_1, \sigma^2}(\mathbf{Y}) = (2\pi\sigma^2)^{-n/2} \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 x_i)^2 \right].$$

L'annulation des dérivées (par rapport à  $\beta_0$  et  $\beta_1$ , d'une part, à  $\sigma^2$  d'autre part) de  $\log L_{\beta_0, \beta_1, \sigma^2}(\mathbf{Y})$  conduit (notation usuelles :  $\bar{x} := \frac{1}{n} \sum_{i=1}^n x_i$  et  $\bar{Y} := \frac{1}{n} \sum_{i=1}^n Y_i$ ) aux 3 équations de vraisemblance (exercice):

$$\begin{aligned} \sum_{i=1}^n (x_i - \bar{x})^2 \beta_1 &= \sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y}) \\ \bar{Y} &= \beta_0 + \beta_1 \bar{x} \\ n\sigma^2 &= \sum_i (Y_i - \beta_0 - \beta_1 x_i)^2 \end{aligned}$$

## Régression simple

dont les solutions sont

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{m_{11}}{s_x^2},$$

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

$$\hat{\sigma}^2 = \frac{1}{n} \sum_i (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 = \frac{1}{n} \sum_i \hat{\varepsilon}_i^2$$

avec les notations usuelles:  $s_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$  et  $m_{11} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y})$

- ▶ Les estimateurs  $\hat{\beta}_0$  et  $\hat{\beta}_1$  sont donc les coefficients de la droite de régression (au sens des moindres carrés) de la série empirique  $(x_i, Y_i)$
- ▶ La droite de régression passe par le point  $(\bar{x}, \bar{Y})$  (ce qu'exprime la seconde équation)

## Régression simple

Dans le cas général, la méthode des moindres carrés remplace la méthode du maximum de vraisemblance, et consiste à minimiser (par rapport à  $\beta_0$  et  $\beta_1$ ) la somme des carrés d'écarts

$$\sum_{i=1}^n (Y_i - \beta_0 - \beta_1 x_i)^2 = \sum_{i=1}^n \varepsilon_i^2.$$

Cette minimisation conduit à la même solution (pour  $\beta_0$  et  $\beta_1$ ) que la méthode du maximum de vraisemblance gaussien.

## Régression simple

### Ecriture du problème sous la forme matricielle

Modèle:  $\mathbf{Y} = \mathbf{X}\beta + \varepsilon$

La méthode des moindres carrés consiste à minimiser (par rapport à  $\beta_0$  et  $\beta_1$ ) la somme des carrés d'écarts

$$\begin{aligned} \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 x_i)^2 &= \sum_{i=1}^n \varepsilon_i^2 = \varepsilon' \varepsilon = (\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta) \\ &= \mathbf{Y}'\mathbf{Y} - \beta'\mathbf{X}'\mathbf{Y} - \mathbf{Y}'\mathbf{X}\beta + \beta'\mathbf{X}'\mathbf{X}\beta = \mathbf{Y}'\mathbf{Y} - 2\beta'\mathbf{X}'\mathbf{Y} + \beta'\mathbf{X}'\mathbf{X}\beta \end{aligned}$$

Pour minimiser cette quantité, nous devons la dériver par rapport à  $\beta$ . Or on peut facilement démontrer que:

$$\begin{aligned} \frac{\partial}{\partial \beta} (\beta'\mathbf{X}'\mathbf{Y}) &= \mathbf{X}'\mathbf{Y} \\ \frac{\partial}{\partial \beta} \beta'\mathbf{X}'\mathbf{X}\beta &= 2\mathbf{X}'\mathbf{X}\beta \end{aligned}$$

ce qui implique que

$$\frac{\partial}{\partial \beta} (\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta) = -2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\beta \implies \hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$



## Régression simple

**Inférence relative à  $\beta_0$  et  $\beta_1$ :** Quelques propriétés de  $(\hat{\beta}_0, \hat{\beta}_1)$

**Non-biais:** On calcule aisément que  $E[\hat{\beta}_0] = \beta_0$  et  $E[\hat{\beta}_1] = \beta_1$ .

En effet,

$$\begin{aligned}\hat{\beta} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\beta + \varepsilon) \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon \\ &= \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon,\end{aligned}$$

ainsi  $E[\hat{\beta}] = \beta$  et

$$\hat{\beta} - \beta = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon.$$

## Régression simple

**Variance:** On peut montrer facilement sous forme matricielle que:

$$\begin{aligned}\text{Var}(\hat{\beta}) &= E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'] \\ &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon\varepsilon'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}] \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E[\varepsilon\varepsilon']\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\end{aligned}$$

Or on sait qu' on a les hypothèses suivantes :  $E[\varepsilon_i] = 0 \forall i$  et

$$E[\varepsilon_i\varepsilon_j] = \begin{cases} \sigma^2 & i = j \\ 0 & i \neq j. \end{cases}$$

Donc  $E[\varepsilon\varepsilon'] = \sigma^2\mathbf{I}_{n \times n}$ . Ainsi,

$$\begin{aligned}\text{Var}(\hat{\beta}) &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma^2\mathbf{I}_{n \times n}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{I}_{n \times n}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\end{aligned}$$

Et comme, on a que

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ \dots & \dots \\ 1 & x_n \end{bmatrix} = \begin{bmatrix} n & \sum_i x_i \\ \sum_i x_i & \sum_i x_i^2 \end{bmatrix}$$

## Régression simple

Le déterminant de  $\mathbf{X}'\mathbf{X}$  est donné par:

$$\det(\mathbf{X}'\mathbf{X}) = n \sum_i x_i^2 - \left(\sum_i x_i\right)^2 = n^2 \left[ \frac{1}{n} \sum_i x_i^2 - \left(\frac{1}{n} \sum_i x_i\right)^2 \right] = n^2 s_x^2 \geq 0,$$

où  $s_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$ . Donc  $\mathbf{X}'\mathbf{X}$  est inversible

$$(\mathbf{X}'\mathbf{X})^{-1} = \frac{1}{n^2 s_x^2} \begin{bmatrix} \sum_i x_i^2 & -\sum_i x_i \\ -\sum_i x_i & n \end{bmatrix}$$

Leurs variances (sous forme non matricielle) sont ainsi données par:

$$\text{Var}(\hat{\beta}_0) = \frac{\sigma^2 \sum_i x_i^2}{n^2 s_x^2} = \frac{1}{n} \sigma^2 \left( 1 + \frac{\bar{x}^2}{s_x^2} \right) \quad \text{et} \quad \text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{n s_x^2}$$

## Régression simple

### Lois échantillonnées exactes (cas gaussien)

Dans le cas gaussien,  $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$  étant une transformation linéaire du vecteur normal  $\mathbf{Y} \sim \mathcal{N}(\mathbf{X}\beta, \sigma^2\mathbf{I})$ , est également normal, de moyenne  $E[\hat{\beta}] = \beta$  et de variance  $\text{Var}(\hat{\beta}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ . On a donc  $\hat{\beta} \sim \mathcal{N}(\beta, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$ .

Donc la loi jointe de  $\hat{\beta}_0$  et  $\hat{\beta}_1$  est asymptotiquement normale bivariée :

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} \sim \mathcal{N} \left( \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \frac{\sigma^2}{n^2 s_x^2} \begin{pmatrix} \sum_{i=1}^n x_i^2 & -\sum_{i=1}^n x_i \\ -\sum_{i=1}^n x_i & n \end{pmatrix} \right)$$

On obtient ainsi les résultats suivants :

$$\hat{\beta}_0 \sim \mathcal{N} \left( \beta_0, \frac{\sigma^2}{n} \left( 1 + \frac{\bar{x}^2}{s_x^2} \right) \right)$$

$$\hat{\beta}_1 \sim \mathcal{N} \left( \beta_1, \frac{\sigma^2}{n s_x^2} \right)$$

De surcroît, on peut montrer que

$$\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi_{n-2}^2;$$

et que le couple  $(\hat{\beta}_0, \hat{\beta}_1)$  et l'estimateur de la variance  $\hat{\sigma}^2$  sont indépendants.

## Régression simple

Les résultats sur les lois échantillonnées permettent (cas gaussien) de

- construire des intervalle de confiance pour  $\beta_0$ ,  $\beta_1$  et  $\sigma^2$
- construire des "tests individuels" pour  $\beta_0$  et  $\beta_1$  (et  $\sigma^2$ ).

Exemple : un intervalle de confiance peut être construit pour  $\beta_0$ , et un autre pour  $\beta_1$ , sur base des résultats suivants ( $S^2 = \frac{1}{n-2} \sum_{i=1}^n \hat{\epsilon}_i^2$ ):

$$\frac{\hat{\beta}_0 - \beta_0}{S \sqrt{\frac{\left(1 + \frac{\bar{x}^2}{s_x^2}\right)}{n}}} \sim t_{n-2} \quad \text{et} \quad \frac{\hat{\beta}_1 - \beta_1}{S/s_x \sqrt{n}} \sim t_{n-2}.$$

Une démarche en tous points semblable à celle qui a été suivie dans cas d'une moyenne  $\mu$  fournit, au niveau de confiance  $(1 - \alpha)$ ,

$$\left[ \hat{\beta}_0 \pm t_{n-2; \alpha/2} S \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum_i (x_i - \bar{x})^2}} \right]$$

et

$$\left[ \hat{\beta}_1 \pm t_{n-2; \alpha/2} \frac{S}{\sqrt{\sum_i (x_i - \bar{x})^2}} \right].$$

## Régression multiple

**Le modèle** Le modèle de régression multiple est une extension naturelle du modèle de régression simple au cas de *plusieurs variables explicatives*.

### Exemples

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i$$

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i$$

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i1} x_{i2} + \beta_4 x_{i1}^2 + \beta_5 x_{i2}^2 + \varepsilon_i;$$

La linéarité exigée par le modèle est une linéarité en les paramètres  $\beta_i$ , pas en les covariables.

## Régression multiple

### Estimation des paramètres

Les résultats obtenus sous forme matricielle pour la régression linéaire simple restent valables pour la régression linéaire multiple (estimateur de maximum de vraisemblance dans le cas gaussien, des moindres carrés dans le cas général):

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

avec comme matrice de variance-covariance associée (preuve):

$$\text{Var}(\hat{\beta}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

De surcroît, le maximum de vraisemblance fournit en outre, pour  $\sigma^2$ , l'estimateur

$$\hat{\sigma}^2 = \frac{1}{n}(\mathbf{Y} - \mathbf{X}\hat{\beta})'(\mathbf{Y} - \mathbf{X}\hat{\beta}) = \frac{1}{n}\|\mathbf{Y} - \mathbf{X}\hat{\beta}\|^2.$$

## Régression multiple

Cas particulier de ce modèle: le modèle d'Analyse de la Variance (ANOVA) à un facteur:

$$X_{ij} = \mu_i + \varepsilon_{ij} \quad i = 1 \dots I \quad j = 1 \dots n_i \quad \text{s'écrit sous la forme : } \mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \text{ avec}$$

$$\mathbf{Y}' = (X_{11}, \dots, X_{1n_1}, X_{21}, \dots, X_{I-1, n_{I-1}}, X_{I1}, \dots, X_{I, n_I}),$$

$$\mathbf{X} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 1 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} \quad (n_i \text{ "1" consécutifs dans la } i\text{ème colonne}),$$

$$\boldsymbol{\beta}' = (\mu_1, \dots, \mu_I) \text{ et } \boldsymbol{\varepsilon}' = (\varepsilon_{11}, \dots, \varepsilon_{1n_1}, \dots, \dots, \varepsilon_{in_I}).$$



## Définitions

### Le modèle d'analyse de la variance à un facteur

Le modèle précédent se généralise de façon immédiate au cas de  $J \geq 2$  échantillons en posant

$$X_{ij} = \mu_j + \varepsilon_{ij} \quad \varepsilon_{ij} \text{ i.i.d. } \mathcal{N}(0, \sigma^2) \quad i = 1, \dots, n_j \quad j = 1, \dots, J$$

(première forme du *modèle d'analyse de la variance à un facteur*).

On y distingue  $J$  paramètres d'intérêt,  $\mu_1, \dots, \mu_J \in \mathbb{R}$ , et un paramètre de nuisance,  $\sigma^2 \in \mathbb{R}_0^+$ .

Ceci correspond à la situation suivante :

Echantillon 1 (taille $n_1$ )	...	Echantillon $j$ (taille $n_j$ )	...	Echantillon $J$ (taille $n_J$ )
$X_{11}, \dots, X_{n_1 1}$	...	$X_{1j}, \dots, X_{n_j j}$	...	$X_{1J}, \dots, X_{n_J J}$
iid $\mathcal{N}(\mu_1, \sigma^2)$		iid $\mathcal{N}(\mu_j, \sigma^2)$		iid $\mathcal{N}(\mu_J, \sigma^2)$

Le modèle d'Analyse de la Variance à un facteur n'est donc rien d'autre qu'un modèle de position à  $J$  échantillons.

## Définitions

Posons

$$\mu := \frac{1}{J} \sum_{j=1}^J \mu_j \quad \text{et} \quad \alpha_j := \mu_j - \mu;$$

on a donc, par construction,  $\sum_{j=1}^J \alpha_j = 0$ .

Seconde forme du *modèle d'analyse de la variance à un facteur* :

$$X_{ij} = \mu + \alpha_j + \varepsilon_{ij} \quad \varepsilon_{ij} \text{ i.i.d. } \mathcal{N}(0, \sigma^2) \quad i = 1, \dots, n_j \quad j = 1, \dots, J.$$

Paramètres d'intérêt  $J + 1$  :  $\mu, \alpha_1, \dots, \alpha_J$  liés par une relation linéaire  $\sum_{j=1}^J \alpha_j = 0$   
Paramètre de nuisance,  $\sigma^2 \in \mathbb{R}_0^+$ .

Les  $\alpha_j, j = 1, \dots, J$  sont appelés *effets-traitement* ou *effets-facteur*.

Selon la forme adoptée pour le modèle, les paramètres du modèle sont

$$\mu_1, \dots, \mu_J \quad \text{et} \quad \sigma^2$$

ou

$$\mu, \alpha_1, \dots, \alpha_J \quad (\text{avec } \sum_{j=1}^J \alpha_j = 0) \quad \text{et} \quad \sigma^2.$$

## Définitions

Deux cas peuvent être envisagés (mais ils conduisent à la même solution) :

- (a) le cas gaussien :  $\varepsilon_{ij}$  i.i.d.  $\mathcal{N}(0, \sigma^2)$ —on emploiera la méthode du maximum de vraisemblance gaussien;
- (b) le cas dit *général*  $E(\varepsilon_{ij}) = 0 \forall i, j$ , et

$$E(\varepsilon_{ij}\varepsilon_{kl}) = \begin{cases} \sigma^2 & \text{si } i = k, j = l \\ 0 & \text{sinon} \end{cases}$$

—on emploiera la méthode des moindres carrés.

## Définitions

Méthode du maximum de vraisemblance sous les hypothèses gaussiennes :

$X_{ij} \sim \mathcal{N}(\mu_j, \sigma^2)$   $i = 1 \dots n_j = 1 \dots J$  indépendants.

La densité de  $X_{ij}$  est donc

$$\frac{1}{\sigma\sqrt{2\pi}} \exp \left[ -\frac{1}{2\sigma^2} (X_{ij} - \mu_j)^2 \right],$$

et la vraisemblance s'écrit donc

$$\begin{aligned} L_{\mu_1 \dots \mu_J; \sigma^2}(X_{11} \dots X_{n_J J}) &= \prod_{i,j} \frac{1}{\sigma\sqrt{2\pi}} \exp \left[ -\frac{1}{2\sigma^2} (X_{ij} - \mu_j)^2 \right] \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp \left[ -\frac{1}{2\sigma^2} \sum_{j=1}^J \sum_{i=1}^{n_j} (X_{ij} - \mu_j)^2 \right]. \end{aligned}$$

Maximiser (par rapport à  $\mu_1 \dots \mu_J$ ) la vraisemblance  $L_{\mu_1 \dots \mu_J; \sigma^2}(X_{11} \dots X_{n_J J})$

⇓

Minimiser (par rapport à  $\mu_1 \dots \mu_J$ )  $\sum_{j=1}^J \sum_{i=1}^{n_j} (X_{ij} - \mu_j)^2$ , ce qui, par définition même, n'est rien d'autre que la méthode des moindres carrés.

## Définitions

La solution est

$$\hat{\mu}_j = \bar{X}_{\bullet j} = \frac{1}{n_j} \sum_{i=1}^{n_j} X_{ij} \quad j = 1 \dots J,$$

qui est à la fois l'estimateur *maximum de vraisemblance* (MLE) et l'estimateur *moindres carrés* (OLS) de  $\mu_j$ .

L'estimateur  $\hat{\mu}_j$  est la moyennes observée "dans le jème niveau du traitement". Si la paramétrisation  $(\mu, \alpha_1, \dots, \alpha_J)$  est préférée, on obtient

$$\hat{\mu} = \frac{1}{J} \sum_{j=1}^J \bar{X}_{\bullet j} = \frac{1}{J} \sum_{j=1}^J \underbrace{\frac{1}{n_j} \sum_{i=1}^{n_j} X_{ij}}_{\bar{X}_{\bullet j}} \quad \text{et} \quad \hat{\alpha}_j = \bar{X}_{\bullet j} - \hat{\mu}$$

La méthode du maximum de vraisemblance fournit également un estimateur pour  $\sigma^2$ , ce que ne fait pas la méthode des moindres carrés :

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{j=1}^J \sum_{i=1}^{n_j} (X_{ij} - \bar{X}_{\bullet j})^2 =: \frac{1}{n} \text{SCrés.}$$

=: SCrés

## Définitions

On obtient directement (combinaison linéaire de normales indépendantes) que

$$\hat{\mu}_j = \bar{X}_{\bullet j} \sim \mathcal{N}\left(\mu_j, \frac{\sigma^2}{n_j}\right) \quad j = 1, \dots, J.$$

Donc  $\hat{\mu}_j$  est un estimateur sans biais de  $\mu_j$ .

On peut montrer par ailleurs (sur la base des propriétés des formes quadratiques idempotentes) que

$$\frac{n\hat{\sigma}^2}{\sigma^2} = \frac{\text{SCrés}}{\sigma^2} \sim \chi_{n-J}^2$$

et que  $\hat{\sigma}^2$  et les  $\hat{\mu}_j$  sont mutuellement indépendants. Pour  $J = 1$  et  $J = 2$ , on retrouve ainsi les résultats antérieurs (problèmes à un et deux échantillons).

Il en découle que, ici encore,  $\hat{\sigma}^2$  est un estimateur biaisé de  $\sigma^2$  : on a en effet

$$E[\hat{\sigma}^2] = E[\text{SCrés}/n] = \frac{\sigma^2}{n} E\left[\frac{\text{SCrés}}{\sigma^2}\right] = \frac{\sigma^2}{n}(n - J) = \frac{n - J}{n}\sigma^2 < \sigma^2.$$

Mais, on peut construire un estimateur *sans biais* de  $\sigma^2$  :  $S^2 := \frac{\text{SCrés}}{n-J} = \frac{n}{n-J}\hat{\sigma}^2$ . En effet,  $E[S^2] = \frac{n}{n-J}E[\hat{\sigma}^2] = \sigma^2$ .

## Définitions

La définition des variables de Student permet donc d'affirmer que, pour tout  $j$ ,

$$\frac{\frac{\hat{\mu}_j - \mu_j}{\sigma/\sqrt{n_j}}}{\sqrt{\frac{SCr\acute{e}s}{\sigma^2}/(n - J)}} = \frac{\hat{\mu}_j - \mu_j}{\sqrt{SCr\acute{e}s/n_j(n - J)}} = \frac{\hat{\mu}_j - \mu_j}{\sqrt{S^2/n_j}} \sim t_{n-J}.$$

Ce résultat permet de construire des intervalle de confiance "individuels" pour les  $\mu_j$ , et de construire les test d'hypothèses correspondants.

A titre d'exemple, un intervalle de confiance au niveau  $(1 - \alpha)$  pour  $\mu_1$  est

$$\left[ \bar{X}_{\bullet 1} \pm t_{n-J;\alpha/2} S \sqrt{\frac{1}{n_1}} \right].$$

## Test de l'hypothèse d'absence d'effet-traitement

Le problème de test le plus courant dans ce contexte est celui de l'hypothèse

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_J$$

ou encore

$$H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_J = 0$$

ou, plus intuitivement,

$$H_0 : \text{pas d'effet traitement.}$$

Formellement, le même problème s'écrit :

$$\begin{cases} H_0 : \alpha_j = 0 & \forall j = 1 \dots J \\ H_1 : \exists j : \alpha_j \neq 0 . \end{cases}$$

ou encore:

$$\begin{cases} H_0 : \mu_1 = \dots = \mu_j & \forall j = 1 \dots J \\ H_1 : \text{l'une au moins des moyennes est différente des autres.} \end{cases}$$



## Test de l'hypothèse d'absence d'effet-traitement

### Méthode générale de construction des tests de Fisher

Problème d'absence d'effet-traitement: cas particulier du problème du test de contraintes linéaires sur les paramètres d'un modèle beaucoup plus général.

Dans le cadre gaussien, une solution est fondée sur une méthode générale de construction de tests, dite *méthode du rapport des vraisemblances*. Cette méthode conduit à une classe de tests, dits *tests de Fisher*.

Supposons que l'hypothèse (absence d'effet-facteur)  $H_0 : \mu_1 = \mu_2 = \dots = \mu_J$  impose  $r = (J - 1)$  contraintes linéaires aux  $J$  paramètres du modèle.

La construction de la statistique de test s'effectue en deux étapes.

1. Estimation du modèle "non contraint" par maximum de vraisemblance. La qualité de l'ajustement est mesurée par la somme des carrés résiduels "non-contraints"  $SC_{rés}$ , qui dispose de  $(n - J)$  degrés de liberté.
2. Estimation du modèle "contraint" sous  $H_0$  (= sous les  $r$  contraintes de  $H_0$ ), par maximum de vraisemblance. La qualité de cet ajustement réalisé sous les contraintes est mesurée par la somme des carrés résiduels  $SC_{rés}^0$  ( $(n - J + r)$  degrés de liberté).

## Test de l'hypothèse d'absence d'effet-traitement

Mesure de la détérioration dans la qualité de l'ajustement attribuable à l'introduction des contraintes qui forment l'hypothèse:

$$SC_{rés}^0 - SC_{rés} \geq 0$$

$((n - J + r) - (n - J) = r$  degrés de liberté).

Si cette détérioration est "trop élevée", il est intuitivement raisonnable de l'attribuer à la fausseté de l'hypothèse. Cette détérioration sera mesurée de façon relative, par le rapport  $(SC_{rés}^0 - SC_{rés}) / SC_{rés}$ .

Plus précisément, le test (dit test de Fisher) se présente de la façon suivante.

Statistique de test :  $F := \frac{SC_{rés}^0 - SC_{rés}}{\frac{r}{n-J} SC_{rés}}$

Loi sous  $H_0$  :  $F \sim F_{r;n-J}$

Règle de comportement :  $RH_0$  si  $F > F_{r;n-J;1-\alpha}$

## Test de l'hypothèse d'absence d'effet-traitement

Dans ce cas particulier, le test de Fisher prend une forme particulièrement simple.

1. Estimation dans le modèle non-contraint : elle conduit, on l'a vu, à la somme de carrés résiduelle

$$SC_{rés} = \sum_{j=1}^J \sum_{i=1}^{n_j} (X_{ij} - \bar{X}_{\bullet j})^2 \quad ((n - J) \text{ degrés de liberté}).$$

2. Estimation dans le modèle contraint sous  $H_0$ ,  $\mu_1 = \dots = \mu_J = \mu$  ( $\mu \in \mathbb{R}$  non spécifié).

On dispose donc de  $n := \sum_{j=1}^J n_j$  observations i.i.d.  $\mathcal{N}(\mu, \sigma^2)$ .

L'estimateur maximum de vraisemblance de la moyenne commune  $\mu$  est donc

$$\bar{X} = \frac{1}{n} \sum_{j=1}^J \sum_{i=1}^{n_j} X_{ij}.$$

La somme de carrés résiduelle correspondante est donc

$$SC_{rés}^0 = \sum_{j=1}^J \sum_{i=1}^{n_j} (X_{ij} - \bar{X})^2 \quad ((n - 1) \text{ degrés de liberté}).$$

## Test de l'hypothèse d'absence d'effet-traitement

On obtient donc la statistique de test

$$F = \frac{(\sum_{j=1}^J \sum_{i=1}^{n_j} (X_{ij} - \bar{X})^2 - \sum_{j=1}^J \sum_{i=1}^{n_j} (X_{ij} - \bar{X}_{\bullet j})^2) / (J - 1)}{\sum_{j=1}^J \sum_{i=1}^{n_j} (X_{ij} - \bar{X}_{\bullet j})^2 / (n - J)},$$

et on rejette l'hypothèse  $H_0$  d'absence d'effet-traitement si  $F > F_{J-1, n-J; 1-\alpha}$ .

Cette statistique s'écrit de façon plus simple : décomposons

$$\begin{aligned} SC_{rés}^0 &= \sum_{j=1}^J \sum_{i=1}^{n_j} (X_{ij} - \bar{X})^2 =: SC_{Tot} \quad (\text{somme des carrés totale}) \\ &= \sum_{j=1}^J \sum_{i=1}^{n_j} (X_{ij} - \bar{X}_{\bullet j} + \bar{X}_{\bullet j} - \bar{X})^2 \\ &= \underbrace{\sum_{j=1}^J \sum_{i=1}^{n_j} (X_{ij} - \bar{X}_{\bullet j})^2}_{SC_{rés}} + \sum_{j=1}^J \sum_{i=1}^{n_j} (\bar{X}_{\bullet j} - \bar{X})^2 + 2 \sum_{j=1}^J \sum_{i=1}^{n_j} (X_{ij} - \bar{X}_{\bullet j})(\bar{X}_{\bullet j} - \bar{X}). \end{aligned}$$

## Test de l'hypothèse d'absence d'effet-traitement

Clairement,

$$\sum_{j=1}^J \sum_{i=1}^{n_j} (\bar{X}_{\bullet j} - \bar{X})^2 = \sum_{j=1}^J n_j (\bar{X}_{\bullet j} - \bar{X})^2;$$

$$\sum_{j=1}^J \sum_{i=1}^{n_j} (X_{ij} - \bar{X}_{\bullet j})(\bar{X}_{\bullet j} - \bar{X}) = \sum_{j=1}^J (\bar{X}_{\bullet j} - \bar{X}) \underbrace{\sum_{i=1}^{n_j} (X_{ij} - \bar{X}_{\bullet j})}_0 = 0.$$

La décomposition de la somme des carrés totale s'écrit alors

$$(SC_{\text{rés}}^0 = ) SC_{\text{Tot}} = \sum_{j=1}^J \sum_{i=1}^{n_j} (X_{ij} - \bar{X}_{\bullet j})^2 + \sum_{j=1}^J n_j (\bar{X}_{\bullet j} - \bar{X})^2 = SC_{\text{intra}} + SC_{\text{inter}},$$

- $SC_{\text{inter}} := \sum_{j=1}^J n_j (\bar{X}_{\bullet j} - \bar{X})^2$  : somme des carrés *inter*classe, ou "entre les classes" (en anglais, "between treatments" ou "between sum of squares")
- $SC_{\text{intra}} := \sum_{j=1}^J \sum_{i=1}^{n_j} (X_{ij} - \bar{X}_{\bullet j})^2 = SC_{\text{rés}}$  : somme des carrés *intra*classe, ou "dans les classes" (en anglais, "within treatments" ou "within sum of squares").

et la statistique de test  $F$  prend la forme

$$F = \frac{SC_{\text{inter}}/(J-1)}{SC_{\text{intra}}/(n-J)} \sim F_{J-1, n-J} \quad (\text{loi sous } H_0).$$

## Test de l'hypothèse d'absence d'effet-traitement

On a donc la décomposition (de la somme des carrés totale)

$$(SC_{rés}^0 =) SCTot = SCinter + SCintra,$$

où  $SCinter$  est expliquée par l'action du facteur tandis que  $SCintra$  est liée à la variabilité de l'erreur.

Cette décomposition est généralement présentée dans un tableau, dit *table d'Analyse de la Variance*, de la forme

Source de variation	Somme des carrés	Degrés de liberté	Carré moyen
Traitement (facteur)	$SCinter$	$J - 1$	$\frac{SCinter}{J - 1}$
Erreur	$SCintra$	$n - J$	$\frac{SCintra}{n - J}$
Total	$SCTot$	$n - 1$	

La statistique du test de Fisher s'obtient comme rapport du carré moyen interclasse au carré moyen de l'erreur (le carré moyen intraclasse):

$$F = \frac{SCinter/(J - 1)}{SCintra/(n - J)} \sim F_{J-1, n-J} \quad (\text{loi sous } H_0).$$