

Statistiques et Applications

Analyse de la variance

aurore.lavigne@univ-lille.fr

Les données et le modèle

On considère **1 facteur** à **k niveaux**.

Pour chaque niveau $l \in \{1, 2, \dots, k\}$, on dispose d'un échantillon de taille n_l d'observations de la variable quantitative.

niveau 1	$Y_1^1, Y_2^1, \dots, Y_{n_1}^1$
niveau 2	$Y_1^2, Y_2^2, \dots, Y_{n_2}^2$
\vdots	\vdots
niveau k	$Y_1^k, Y_2^k, \dots, Y_{n_k}^k$

Indépendance

On suppose que les **toutes** les variables sont **indépendantes**.

- Les variables d'un même niveau sont indépendantes :
 $\forall l \in \{1, 2, \dots, k\}, \forall i \neq j, Y_i^l$ et Y_j^l sont indépendantes.
- Les variables de deux niveaux différents sont indépendantes
 $\forall l \neq m, \forall (i, j), Y_i^l$ et Y_j^m sont indépendantes.

Modèle

On suppose de plus que

- toutes les variables suivent une distribution normale
- **l'espérance dépend du niveau k**
- la variance est identique pour toutes les variables

$$Y_i^l \sim \mathcal{N}(\mu_l, \sigma^2), \quad \forall l \in \{1, \dots, k\}, \quad \forall i \in \{1, \dots, n_l\}$$

Modèle

De manière équivalente, on pourra écrire que

$$Y_i^l = \mu_l + \epsilon_i^l \quad \text{avec} \quad \epsilon_i^l \sim \mathcal{N}(0, \sigma^2) \text{ et ind.}$$

- μ_l est l'espérance observée pour le niveau l du facteur.

L'anova est un modèle linéaire

En effet on peut réécrire le modèle ci-dessus de la manière suivante :

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \text{ avec } \boldsymbol{\varepsilon} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_n)$$

et

$$\mathbf{X} = \left(\begin{array}{cccc} \overbrace{1 \ 0 \ \cdots \ 0}^{k \text{ col.}} & & & \\ \vdots & \vdots & \vdots & \vdots \\ 1 \ 0 \ \cdots \ 0 & & & \\ 0 \ 1 \ \cdots \ 0 & & & \\ \vdots & \vdots & \vdots & \vdots \\ 0 \ 1 \ \cdots \ 0 & & & \\ \vdots & \vdots & \vdots & \vdots \\ 0 \ 0 \ \cdots \ 1 & & & \\ \vdots & \vdots & \vdots & \vdots \\ 0 \ 0 \ \cdots \ 1 & & & \end{array} \right) \begin{array}{l} \left. \begin{array}{l} \\ \\ \\ \end{array} \right\} n_1 \text{ li.} \\ \left. \begin{array}{l} \\ \\ \\ \end{array} \right\} n_2 \text{ li.} \\ \left. \begin{array}{l} \\ \\ \\ \end{array} \right\} n_k \text{ li.} \end{array} \quad \text{et } \boldsymbol{\beta} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_k \end{pmatrix}$$

Décomposition de la variance

Dans le cadre d'analyse de la variance, on a :

$$SCT = \sum_{l=1}^k \sum_{i=1}^{n_l} (Y_i^l - \bar{Y})^2 \quad SCM = \sum_{l=1}^k n_l (\bar{Y}_l - \bar{Y})^2 \quad SCR = \sum_{l=1}^k \sum_{i=1}^{n_l} (Y_i^l - \bar{Y}_l)^2$$

avec \bar{Y}_l la moyenne pour le niveau l : $\bar{Y}_l = \frac{1}{n_l} \sum_{i=1}^{n_l} X_i^l$.

REMARQUE : L'espace engendré par les colonnes de \mathbf{X} est de dimension k , le degré de liberté associé à SCM est donc $k - 1$.

Sous l'hypothèse \mathcal{H}_0

$$\frac{SCM/k - 1}{SCR/n - k} \sim \mathcal{F}_{k-1, n-k}$$

Table d'analyse de la variance

Source	Somme des carrés	Degrés de liberté	Som. carrés moyens	Statistique F	Proba. crit.
Facteur	SCM	$k - 1$	$\frac{SCM}{k-1}$	$\frac{SCM/k-1}{SCR/n-k}$	p_c
Résidu	SCR	$n - k$	$\frac{SCR}{n-k}$		
Total	SCT	$n - 1$			

Estimation de l'écart-type σ

$$S = \sqrt{\frac{SCR}{n - k}}$$

est un estimateur de l'écart-type σ .

REMARQUES :

- On peut trouver l'estimation de l'écart-type dans la table d'analyse de la variance.
- La quantité $\frac{SCR}{n-k}$ est souvent nommée *MSE* pour *mean square error* dans les logiciels de statistiques.

Écriture singulière

La plupart des logiciels de statistiques utilisent l'écriture suivante (écriture singulière)

$$Y_i^l = \mu + \alpha_l + \epsilon_i^l \quad \text{avec} \quad \epsilon_i^l \sim \mathcal{N}(0, \sigma^2) \text{ et ind.}$$

- μ est la moyenne générale.
- α_l est l'**effet** du niveau l du facteur

Dans ce cas la matrice \mathbf{X} devient singulière, la première colonne est égale à la somme des k colonnes suivantes.

$$\mathbf{X} = \left(\begin{array}{ccccc} \overbrace{1 \quad 1 \quad 0 \quad \cdots \quad 0}^{k+1 \text{ col.}} & & & & \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & 0 & 0 & \cdots & 1 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & 0 & 0 & \cdots & 1 \end{array} \right) \begin{array}{l} \left. \begin{array}{c} \\ \\ \\ \end{array} \right\} n_1 \text{ li.} \\ \left. \begin{array}{c} \\ \\ \\ \end{array} \right\} n_2 \text{ li.} \\ \left. \begin{array}{c} \\ \\ \\ \end{array} \right\} n_k \text{ li.} \end{array} \quad \text{et } \boldsymbol{\beta} = \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_k \end{pmatrix}$$

Ajout de contraintes d'identifiabilité

Pour rendre le modèle identifiable, on va ajouter une contrainte sur une combinaison linéaire des paramètres $\mu, \alpha_1, \dots, \alpha_l$. Par exemple :

$$\begin{cases} \sum_{l=1}^k n_l \alpha_l = 0 & (1) \\ \alpha_1 = 0 & (2) \end{cases}$$

Avec la contrainte (1), on estime

- μ par \bar{y}
- α_l par $\bar{y}_l - \bar{y}$

Avec la contrainte (2), on estime

- μ par \bar{y}_1
- α_l par $\bar{y}_l - \bar{y}_1$

REMARQUES :

- La contrainte (2) est celle utilisée par **R**.
- Si on ne connaît pas les contraintes, il ne faut pas chercher à interpréter les tests, et les coefficients.

Tests sur les effets

De nombreux logiciels donnent la probabilité critique du test $\mathcal{H}_0 = \{\alpha_l = 0\}$ contre $\mathcal{H}_1 = \{\alpha_l \neq 0\}$.

ATTENTION :

Selon la contrainte utilisée, la signification du test n'est pas la même.

- **Avec (2)** : Le test de $\{\alpha_1 = 0\}$ revient à tester $\{\mu_1 = 0\}$. “La moyenne du groupe 1 est nulle”.
Le test de $\{\alpha_2 = 0\}$ revient à tester $\{\mu_2 = \mu_1\}$. “La moyenne du groupe 2 est égale à la moyenne du groupe 1”.
- **Avec (1)** : Le test de $\{\alpha_1 = 0\}$ revient à tester $\{\mu_1 = \mu\}$. “La moyenne du groupe 1 est égale à la moyenne générale”.

CONCLUSION : Il est fortement déconseillé de se baser sur ces tests pour prendre des décisions sur le modèle.

Contraste

Définition

On appelle contraste L des k moyennes $\mu_1, \mu_2, \dots, \mu_k$ la somme

$$L = \sum_{l=1}^k l_l \mu_l \text{ telle que } \sum_{l=1}^k l_l = 0.$$

EXEMPLES :

- $\mu_1 - \mu_2$: pour comparer μ_1 à μ_2
- $\mu_1 - 2\mu_2 + \mu_3$: pour comparer μ_2 à la moyenne de μ_1 et μ_3 .

INTÉRÊT : On utilise les contrastes pour tester des écarts entre les niveaux d'un même facteur. Comme la somme des coefficients est nulle, les contrastes sont indépendants du choix des contraintes d'identifiabilité.

Estimation

Un estimateur sans biais de L est

$$\hat{L} = \sum_{l=1}^l l_l \hat{\mu}_l = \sum_{l=1}^l l_l \bar{Y}_l$$

Propriétés

On a

- $E(\hat{L}) = L$
- $V(\hat{L}) = \sigma^2 \sum_{l=1}^k \frac{l_l^2}{n_l}$
-

$$\frac{\hat{L} - L}{S \sqrt{\sum_{l=1}^k \frac{l_l^2}{n_l}}} \sim \mathcal{S}_{n-k}$$

Tests sur les contrastes

Tests *a priori*

On sait *a priori* à quelle question doit répondre notre analyse. On définit le contraste en fonction de la problématique et on test $\mathcal{H}_0 = \{L = 0\}$ contre $\mathcal{H}_1 = \{L \neq 0\}$.

- Avantages : on réalise peu de tests
- Inconvénients : il faut à l'avance savoir ce que l'on veut tester

Comparaisons multiples *a posteriori*

On ne sait pas *a priori* ce que l'on cherche, on se trouve dans une démarche exploratoire. On teste tous les contrastes $\mu_l - \mu_{l'}$.

- Avantages : on n'a pas besoin d'avoir une question par avance.
- Inconvénients : tests multiples, on réalise $\frac{k(k-1)}{2}$ tests.

Tests multiples

Soit une famille de m hypothèses de tests \mathcal{H}_{0i} contre \mathcal{H}_{1i} , pour $i \in \{1, 2, \dots, m\}$.

Definition

On appelle *FWER* le *family wise error rate*, la probabilité de rejeter à tort au moins 1 fois une hypothèse \mathcal{H}_{0i} sur les m tests réalisés.

Propriété

Si les m tests sont indépendants et tous de niveau α alors

$$FWER = 1 - (1 - \alpha)^m$$

=> Démonstration

m	1	5	10	20	100
$FWER$	0.05	0.22	0.40	0.64	0.99

CONSÉQUENCE : On ne contrôle plus le risque de première espèce.

Méthode de Bonferroni

On diminue le risque de première espèce α . On prend $\alpha' = \frac{\alpha}{m}$, avec m le nombre de tests à réaliser.

- Avantage : on diminue la probabilité de réaliser au moins une erreur de première espèce sur les m tests.
- Inconvénient : on diminue aussi la puissance du test. On aura des difficultés à repérer les groupes différents.

Les données et le modèle

- On considère **2 facteurs**.
 - le premier facteur est indicé par i et a I niveaux,
 - le deuxième facteur est indicé par j et a J niveaux.
- Y_{ijk} est la k^e observation de la réponse dans le niveau i du facteur 1 et j du facteur 2.
- Il y a n_{ij} observations dans le niveau i du facteur 1 et le niveau j du facteur 2. On note

$$n = \sum_{i=1}^I \sum_{j=1}^J n_{ij} \quad n_{i+} = \sum_{j=1}^J n_{ij} \quad n_{+j} = \sum_{i=1}^I n_{ij}$$

Modèle

On suppose de plus que

- toutes les variables suivent une distribution normale
- **l'espérance dépend du niveau i du facteur 1 et du niveau j du facteur 2**
- la variance est identique pour toutes les variables
- toutes les variables sont **indépendantes**.

$$Y_{ijk} \sim \mathcal{N}(\mu_{ij}, \sigma^2), \quad i.i.d.$$

Modèle

De manière équivalente, on pourra écrire que

$$Y_{ijk} = \mu_{ij} + \epsilon_{ijk} \quad \text{avec} \quad \epsilon_{ijk} \sim \mathcal{N}(0, \sigma^2) \text{ et ind.}$$

- μ_{ij} est l'espérance observée pour le croisement du niveau i du facteur 1 et du niveau j du facteur 2.

Sous forme matricielle

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \text{ avec } \boldsymbol{\varepsilon} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_n)$$

et

$$\mathbf{X} = \left(\begin{array}{cccc} \overbrace{1 \ 0 \ \dots \ 0}^{I \times J \text{ col.}} & & & \\ \vdots & \vdots & \vdots & \vdots \\ 1 \ 0 \ \dots \ 0 & & & \\ 0 \ 1 \ \dots \ 0 & & & \\ \vdots & \vdots & \vdots & \vdots \\ 0 \ 1 \ \dots \ 0 & & & \\ & \vdots & \vdots & \vdots \\ 0 \ 0 \ \dots \ 1 & & & \\ \vdots & \vdots & \vdots & \vdots \\ 0 \ 0 \ \dots \ 1 & & & \end{array} \right) \begin{array}{l} \left. \begin{array}{c} \\ \\ \\ \end{array} \right\} n_{11} \text{ li.} \\ \left. \begin{array}{c} \\ \\ \\ \end{array} \right\} n_{12} \text{ li.} \\ \left. \begin{array}{c} \\ \\ \\ \end{array} \right\} n_{IJ} \text{ li.} \end{array} \quad \text{et } \boldsymbol{\beta} = \begin{pmatrix} \mu_{11} \\ \mu_{12} \\ \vdots \\ \mu_{IJ} \end{pmatrix}$$

Ecriture singulière

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk} \quad \text{avec} \quad \epsilon_{ijk} \sim \mathcal{N}(0, \sigma^2)_{iid}$$

- μ est le terme moyen
- α_i représente l'effet principal du niveau i du facteur 1
- β_j représente l'effet principal du niveau j du facteur 2
- γ_{ij} est le terme d'interaction, il modélise l'écart entre l'effet de la combinaison des niveaux i et j et la somme des effets de chacun des facteurs. En l'absence d'interactions, on suppose que pour tout couple (i, i') $\mu_{ij} - \mu_{i'j} = cte$ pour tout j .

Ajout de contraintes

L'écriture singulière permet de faire apparaître explicitement tous les effets, cependant dans cette écriture le modèle n'est pas identifiable. En effet, dans la première version, la matrice \mathbf{X} est de rang IJ , alors que nous introduisons dans la version singulière $1 + I + J + IJ$ coefficients. Nous allons donc devoir ajouter $1 + I + J$ contraintes sur les coefficients pour le rendre identifiable.

CONSTRAINTES NATURELLES

$$\sum_i n_{i+} \alpha_i = 0 \quad \sum_j n_{+j} \beta_j = 0 \quad \forall i : \sum_j n_{ij} \gamma_{ij} = 0 \quad \forall j : \sum_i n_{ij} \gamma_{ij} = 0$$

Avec ces contraintes les estimateurs sont les moyennes empiriques par sous groupes.

CONSTRAINTES DE R

$$\alpha_1 = 0 \quad \beta_1 = 0 \quad \forall i : \gamma_{i1} = 0 \quad \forall j : \gamma_{1j} = 0$$

Table d'analyse de la variance

On teste

$$\mathcal{H}_0 = \{Y_{ijk} = \mu + \epsilon_{ijk}\} \quad \mathcal{H}_1 = \{Y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk}\}$$

Source	Somme des carrés	Degrés de liberté	Som. carrés moyens	Statistique F	Proba. crit.
Facteur	SCM	$IJ - 1$	$\frac{SCM}{IJ-1}$	$\frac{SCM/IJ-1}{SCR/n-IJ}$	p_c
Résidu	SCR	$n - IJ$	$\frac{SCR}{n-IJ}$		
Total	SCT	$n - 1$			

Considérons les modèles suivant :

ModABi

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk} \quad \text{avec} \quad \epsilon_{ijk} \sim \mathcal{N}(0, \sigma^2) iid$$

ModABa

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk} \quad \text{avec} \quad \epsilon_{ijk} \sim \mathcal{N}(0, \sigma^2) iid$$

ModA

$$Y_{ijk} = \mu + \alpha_i + \epsilon_{ijk} \quad \text{avec} \quad \epsilon_{ijk} \sim \mathcal{N}(0, \sigma^2) iid$$

ModB

$$Y_{ijk} = \mu + \beta_j + \epsilon_{ijk} \quad \text{avec} \quad \epsilon_{ijk} \sim \mathcal{N}(0, \sigma^2) iid$$

La fonction `anova` fournit les résultats des tests de comparaisons de modèles (type I) comme suit :

Source	\mathcal{H}_0	\mathcal{H}_1	Df
Facteur 1	mod0	modA	I
Facteur 2	modA	modABa	J
Interaction	modABa	modABi	$IJ - I - J$

- Sur la ligne "Interaction", le modèle complet avec interaction est comparé au modèle à effets additifs. On teste l'effet de l'interaction sur le modèle.
- Sur la ligne "Facteur 2", on compare le modèle à effets additifs avec le modèle à un facteur : le facteur 1. On teste donc l'utilité du facteur 2, dans un modèle contenant déjà le facteur 1.
- Sur la ligne "Facteur 1", on compare le modèle à un facteur (le facteur 1) avec le modèle avec une constante.
- Remarque : l'ordre dans lequel sont introduits les facteurs ont une importance.

La fonction `Anova` fournit les résultats des tests de comparaisons de modèles (type II) comme suit :

Source	\mathcal{H}_0	\mathcal{H}_1	Df
Facteur 1	modB	modABa	I
Facteur 2	modA	modABa	J
Interaction	modABa	modABi	$IJ - I - J$

- L'ordre dans lequel sont introduits les facteurs ici n'ont pas d'importance. Les deux facteurs ont un rôle symétrique.

Situation du problème

- **En régression simple ou multiple :**

Réponse \Leftarrow Variable(s) explicative(s) **quantitative(s)**

- **En analyse de la variance :**

Réponse \Leftarrow Variable(s) explicative(s) **qualitative(s)**

Nous avons vu que tous ces modèles appartiennent au même cadre, celui du modèle linéaire :

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_n)$$

Analyse de la covariance

Dans une analyse de la covariance, des variables **quantitative(s)** et **qualitative(s)** sont introduites ensembles pour expliquer la variable réponse.

Exemple

On s'intéresse aux dépenses des adolescents aux jeux de hasard, dans les années 1980. On dispose de plusieurs variables :

- **Gamble** la dépense annuelle au jeu de hasard (en livre par an)
- **Income** le revenu (en livre par semaine)
- **Sex** le sexe

On suppose qu'à la fois la fois, le sexe et le revenu peuvent avoir un impact sur la réponse.

Les données et le modèle

On considère

- **1 facteur** (variable qualitative) indicé par i et a I niveaux,
- **1 variable quantitative** x .

On note

- Y_{ik} est la k^e observation de la réponse dans le niveau i du facteur,
- x_{ik} la valeur de la variable x pour la k^e observation dans le niveau i du facteur.

Il y a n_i observations dans le niveau i du facteur. On note

$$n = \sum_{i=1}^I n_i$$

Modèle

On suppose de plus que

- toutes les variables suivent une distribution normale
- **l'espérance dépend du niveau i du facteur 1 et de la variable x**
- la variance est identique pour toutes les variables
- toutes les variables sont **indépendantes**.

$$Y_{ik} \sim \mathcal{N}(\mu_i + \beta_i x_{ik}, \sigma^2), \quad i.i.d.$$

Modèle

De manière équivalente, on pourra écrire que

$$Y_{ik} = \mu_i + \beta_i x_{ik} + \epsilon_{ik} \quad \text{avec} \quad \epsilon_{ik} \sim \mathcal{N}(0, \sigma^2) \text{ et ind.}$$

- μ_i est l'ordonnée à l'origine de la droite de régression pour le niveau i
- β_i est la pente de la droite de régression pour le niveau i

Sous forme matricielle

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \text{ avec } \boldsymbol{\varepsilon} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_n)$$

et

$$\mathbf{X} = \left(\begin{array}{c|c|c} \overbrace{\begin{matrix} 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \end{matrix}}^{I \text{ col.}} & \overbrace{\begin{matrix} x_{11} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ x_{1n_1} & 0 & \cdots & 0 \\ 0 & x_{21} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & x_{2n_2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \end{matrix}}^{I \text{ col.}} & \left. \begin{array}{l} \left. \begin{array}{l} \left. \begin{array}{l} \vdots \\ \vdots \\ \vdots \\ \vdots \end{array} \right\} n_1 \text{ li.} \\ \left. \begin{array}{l} \left. \begin{array}{l} \vdots \\ \vdots \\ \vdots \\ \vdots \end{array} \right\} n_2 \text{ li.} \end{array} \right\} \\ \left. \begin{array}{l} \left. \begin{array}{l} 0 & 0 & \cdots & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 \end{array} \right\} n_I \text{ li.} \end{array} \right\} \end{array} \right) \end{array} \right) \text{ et } \boldsymbol{\beta} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_I \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_I \end{pmatrix}$$

Ecriture singulière

$$Y_{ik} = \mu + \alpha_i + \beta x_{ik} + \gamma_i x_{ik} + \epsilon_{ik} \quad \text{avec} \quad \epsilon_{ik} \sim \mathcal{N}(0, \sigma^2) iid$$

- μ est le terme moyen
- α_i représente l'effet principal du niveau i du facteur 1
- β représente l'effet principal de la variable quantitative
- γ_i est le terme d'interaction, entre le facteur et la variable quantitative. Il signifie que la pente de la droite de régression est différente selon le niveau du facteur.

Ajout de contraintes

Selon l'écriture singulière on doit estimer : $1 + I + 1 + I$ coefficients, alors que le rang de la matrice \mathbf{X} est $2I$. On va donc introduire deux contraintes d'identifiabilités.

CONSTRAINTES DE R

$$\alpha_1 = 0 \quad \beta_1 = 0$$

Table d'analyse de la variance

On teste

$$\mathcal{H}_0 = \{Y_{ik} = \mu + \epsilon_{ik}\} \quad \mathcal{H}_1 = \{Y_{ik} = \mu + \alpha_i + \beta x_{ik} + \gamma_i x_{ik} + \epsilon_{ik}\}$$

Source	Somme des carrés	Degrés de liberté	Som. carrés moyens	Statistique F	Proba. crit.
Facteur	SCM	$2I - 1$	$\frac{SCM}{2I-1}$	$\frac{SCM/2I-1}{SCR/n-2I}$	p_c
Résidu	SCR	$n - 2I$	$\frac{SCR}{n-2I}$		
Total	SCT	$n - 1$			

Considérons les sous-modèles suivant :



ModABi

$$Y_{ik} = \mu + \alpha_i + \beta x_{ik} + \gamma_i x_{ik} + \epsilon_{ik} \quad \text{avec} \quad \epsilon_{ik} \sim \mathcal{N}(0, \sigma^2)_{iid}$$

ModABp (parallèle)

$$Y_{ik} = \mu + \alpha_i + \beta x_{ik} + \epsilon_{ik} \quad \text{avec} \quad \epsilon_{ik} \sim \mathcal{N}(0, \sigma^2)_{iid}$$

ModABo (rdonnée à l'origine)

$$Y_{ik} = \mu + \beta x_{ik} + \gamma_i x_{ik} + \epsilon_{ik} \quad \text{avec} \quad \epsilon_{ik} \sim \mathcal{N}(0, \sigma^2)_{iid}$$

ModB

$$Y_{ik} = \mu + \beta x_{ik} + \epsilon_{ik} \quad \text{avec} \quad \epsilon_{ik} \sim \mathcal{N}(0, \sigma^2)_{iid}$$

ModA

$$Y_{ik} = \mu + \alpha_i + \epsilon_{ik} \quad \text{avec} \quad \epsilon_{ik} \sim \mathcal{N}(0, \sigma^2)_{iid}$$

Mod0

$$Y_{ik} = \mu + \cancel{\alpha_i} + \epsilon_{ik} \quad \text{avec} \quad \epsilon_{ik} \sim \mathcal{N}(0, \sigma^2)_{iid}$$

- Les modèles ModABo et ModABp sont des cas particuliers de ModABi
 - Dans ModABp, les droites de régressions sont parallèles (même pente) mais l'ordonnée à l'origine varie selon le niveau.
 - Dans ModABo, les droites de régressions ne sont pas parallèles mais ont la même ordonnée à l'origine.
 - Dans ModABi, les droites de régressions ne sont pas parallèles et n'ont pas la même ordonnée à l'origine.
- le modèle ModA est un cas particulier du modèle modABp. Dans ce modèle toutes les pentes sont nulles.
- le modèle ModB est un cas particulier du modèle modABo. Dans ce cas, toutes les pentes sont identiques mais non nécessairement nulles.
- le modèle ModO est un cas particulier de tous les autres modèles.

La fonction `anova` fournit les résultats des tests de comparaisons de modèles (effets de type I) comme suit :

Source	\mathcal{H}_0	\mathcal{H}_1	Df
Variable A	mod0	modA	I
Variable B	modA	modABp	$I + 1 - I$
Interaction	modABi	modABp	$2I - (I + 1)$

REMARQUE : les degrés de liberté ne sont valables que si A est le facteur.

Si A est le facteur :

- Sur la ligne "Interaction", le modèle complet avec interaction est comparé au modèle avec droites parallèle. On teste l'effet de l'interaction sur le modèle.
- Sur la ligne "Variable B", on compare le modèle où toutes les droites sont horizontales avec le modèle à droites parallèles.
- Sur la ligne "Variable A", on compare le modèle à plusieurs droites horizontales avec le modèle à une droite horizontale.
- Remarque : l'ordre dans lequel sont introduits les facteurs ont une importance.

La fonction `Anova` fournit les résultats des tests de comparaisons de modèles (effets de type II) comme suit :

Source	\mathcal{H}_0	\mathcal{H}_1	Df
Variable A	modB	modABp	$I + 1 - 1$
Variable B	modA	modABp	$I + 1 - I$
Interaction	modABi	modABp	$2I - (I + 1)$

REMARQUES :

- L'ordre dans lequel sont introduits les facteurs ici n'ont pas d'importance. Les deux facteurs ont un rôle symétrique.
- Les degrés de liberté ne sont valables que si A est le facteur.