ANOVA à deux critères de classification (deux facteurs)

C'est un test d'égalité de la moyenne de plusieurs populations. On étudie l'effet de deux facteurs A et B indépendants l'un de l'autre sur une variable aléatoire quantitative y. Le facteur A à I niveaux et le facteur B à J niveaux. Il existe deux types d'ANOVA2 :

* ANOVA2 avec répétitions (égales ou inégales).
* ANOVA2 sans répétitions.

I. ANOVA2 sans répétitions (sans interaction)

Dans ce cas il n'y a pas de répétitions pour chaque combinaison des niveaux des deux facteurs A et B. On peut pas tester l'interaction entre les deux facteurs.

1. Structure des données :

Les données seront présentées sous forme d'un tableau de contingences de facteur A ayant I modalités et de facteur B ayant J modalités.

	facteur A						
	mod	mod 1	mod 2	...	mod i	mod I	
facteur B	mod 1	y_{11}	y_{12}	...	y_{1i}	y_{1I}	$\bar{y}_{1.}$
	mod 2	y_{21}	y_{22}	...	y_{2i}	y_{2I}	$\bar{y}_{2.}$
	$\vdots$	$\vdots$	$\vdots$	...	$\vdots$	$\vdots$	$\vdots$
	mod j	y_{i1}	y_{i2}	...	y_{ii}	y_{iI}	$\bar{y}_{j.}$
	mod J	y_{J1}	y_{J2}	...	y_{Ji}	y_{JI}	$\bar{y}_{J.}$
		$\bar{y}_{.1}$	$\bar{y}_{.2}$	...	$\bar{y}_{.i}$	$\bar{y}_{.I}$	$\bar{y}$

telle que :

$$\triangleright\ \bar{y}_{.j} = \frac{1}{I}\sum_{i=1}^{I} y_{ij}, \quad j = \overline{1, J}$$

$$\triangleright\ \bar{y}_{i.} = \frac{1}{J}\sum_{j=1}^{J} y_{ij}, \quad i = \overline{1, I}$$

$$\begin{aligned}\triangleright\ \bar{y} &= \frac{1}{n}\sum_{i=1}^{I}\sum_{j=1}^{J} y_{ij} \\ &= \frac{1}{I}\sum_{i=1}^{I} \bar{y_{i.}} \\ &= \frac{1}{J}\sum_{j=1}^{J} \bar{y_{.j}} \qquad \text{(telle que n=I} \times J)\end{aligned}$$

2. Présentation du modèle linéaire

Le modèle d'ANOVA dans ce cas, s'écrit comme suit :

$$y_{ij} = \mu + \alpha_i + \beta_i + \varepsilon_{ij}, \quad \sum_{i=1}^{I} \alpha_i = \sum_{j=1}^{J} \beta_j = 0$$

telle que :

- μ : La moyenne totale de y dans la population.
- α_i : effet du facteur A sur y pour la modalité i.
- β_i : effet du facteur B sur y pour la modalité j.
- ε_{ij} : L'erreur aléatoire, liée aux fluctuations d'échantillonnage pour chaque valeur y_{ij}, $\varepsilon_{ij} \sim N(0, \sigma^2)$.

3. Hypothèses à tester

On teste l'effet du facteur A sur y

► $Test1$
$$\begin{cases} H_0 : \text{Le facteur A n'a pas d'effet sur y, ie :''}\mu_1 = \mu_2 = ... = \mu_i = ... = \mu_I\text{''} \\ H_1 : \text{Le facteur A a un effet sur y, c-à-d : "au moins une des moyennes est différente des autres"} \end{cases}$$

► $Test2$
$$\begin{cases} H_0 : \text{Le facteur B n'a pas d'effet sur y, ie :''}\mu_1 = \mu_2 = ... = \mu_j = ... = \mu_J\text{''} \\ H_1 : \text{Le facteur B a un effet sur y.} \end{cases}$$

4. L'équation fondamentale d'ANOVA2 sans répétitions

$$\underbrace{\sum_{i=1}^{I}\sum_{j=1}^{J}(y_{ij} - \bar{y})^2}_{SCT} = \underbrace{\sum_{i=1}^{I} J(\bar{y_i} - \bar{y})^2}_{SCF_A} + \underbrace{\sum_{j=1}^{J} I(\bar{y_j} - \bar{y})^2}_{SCF_B} + \underbrace{\sum_{i=1}^{I}\sum_{j=1}^{J}(y_{ij} - \bar{y_i} - \bar{y_j} + \bar{y})^2}_{SCR}$$

telle que : $\bar{y_i} = \bar{y_{i.}}$, $\bar{y_j} = \bar{y_{.j}}$

$$SCT = SCF_A + SCF_B + SCR$$

5. Tableau d'ANOVA2 sans répétitions

Source de variation	SC	ddl	MC	Statistique F
Facteur A	SCF_A	I-1	$MCF_A = \frac{SCF_A}{I-1}$	$Fc_A = \frac{MCF_A}{MCR}$
Facteur B	SCF_B	J-1	$MCF_B = \frac{SCF_B}{J-1}$	$Fc_B = \frac{MCF_B}{MCR}$
Résiduelle	SCR	(I-1)(J-1)	$MCR = \frac{SCR}{(I-1)(J-1)}$	/
Totale	SCT	(IJ)-1	$MCT = \frac{SCT}{(IJ)-1}$	/

Lorsque les populations sont normales, les échantillons sont indépendants et les variances sont égales. Chaque hypothèse peut être testée à l'aide d'un test de fisher.

6. Règle de décision

On rejette H_0 au seuil α, si : $Fc_A > F_{\alpha,(I-1),(I-1)(J-1)}$

et

$$Fc_B > F_{\alpha,(J-1),(I-1)(J-1)}$$

II. ANOVA2 avec répétitions (avec interaction)

On étudie deux facteurs A et B ayant respectivement I et J modalités, mais cette fois, on dispose de plusieurs répétitions pour chaque combinaison des des deux facteurs. Le bénéfice ici est qu'on peut tester l'interaction entre les deux facteurs.

1. Structure des données

Les données dans ce cas seront présenter sous forme du tableau de contingence suivant :

	facteur A							
facteur B	mod	mod 1	mod 2	...	mod i	...	mod I	$\bar{y}_{.j.}$
	mod 1	$y_{111}, y_{112}, \ldots, y_{11k}$	$y_{121}, y_{122}\cdot\cdot$	...	$y_{i11}, y_{i12}, \ldots, y_{i1k}$	...	$y_{I11}, y_{I12}, \ldots, y_{I1k}$	$\bar{y}_{.1.}$
	mod 2	$y_{121}, y_{122}, \ldots, y_{12k}$	$y_{221}, y_{222}\cdot\cdot$	...	$y_{i21}, y_{i22}, \ldots, y_{i2k}$	...	$y_{I21}, y_{I22}, \ldots, y_{I1k}$	$\bar{y}_{.2.}$
	$\vdots$	$\vdots$		...		$\vdots$		$\vdots$
	mod j	$y_{j11}, y_{j12}, \ldots$	$y_{j21}, y_{j22}\cdot\cdot$	...	$y_{ji1}, y_{ji2}, \ldots$	...	$y_{jI1}, y_{jI2}, \ldots$	$\bar{y}_{.j.}$
	$\vdots$	$\vdots$		...		$\vdots$		$\vdots$
	mod J	$y_{J11}, y_{J12}, \ldots$	$y_{J21}, y_{J22}\cdot\cdot$	...	$y_{Ji1}, y_{Ji2}, \ldots$	...	$y_{JI1}, y_{JI2}, \ldots$	$\bar{y}_{.J.}$
	$\bar{y}_{i..}$	$\bar{y}_{1..}$	$\bar{y}_{2..}$	...	$\bar{y}_{i..}$	...	$\bar{y}_{I..}$	$\bar{y}$

telle que :

- $\bar{y}_{.j.}$: représente la moyenne pour la j$^{\text{ème}}$ modalité du facteur B.
- $\bar{y}_{i..}$: représente la moyenne pour la i$^{\text{ème}}$ modalité du facteur A.
- $\bar{y}_{ij.}$: la moyenne du i$^{\text{ème}}$ et j$^{\text{ème}}$ modalité.

2. Présentation du modèle linéaire

Le modèle d'ANOVA dans ce cas, s'écrit comme suit :

$$y_{ijk} = \mu + \alpha_i + \beta_i + \gamma_{ij} + \varepsilon_{ijk}$$

avec :

— i=$\overline{1, I}$: le nombre de niveaux du facteur A.

— j=$\overline{1, J}$: le nombre de niveaux du facteur B.

— k=$\overline{1,K}$: le nombre de niveaux des répétitions pour chaque combinaison des niveaux des facteurs.
— μ : La moyenne paramétrique de la population statistique.
— α_i : L'effet du facteur contrôlé A sur y.
— β_i : L'effet du facteur contrôlé B sur y.
— γ_{ij} : L'effet de l'interaction dans le sous groupe d'unités représentant le i$^{\text{ème}}$ groupe de A et le j$^{\text{ème}}$ groupe de B.
— ε_{ij} : L'erreur aléatoire due aux fluctuations d'échantillonnage pour chaque valeur y_{ijk}.

3. Hypothèses à tester

On effectue 3 tests principales

► $Test1 \begin{cases} H_0 : \text{Le facteur A n'a pas d'effet sur y, ie :"}\mu_1 = \mu_2 = ... = \mu_i = ... = \mu_I\text{"} \\ H_1 : \text{Le facteur A a un effet sur y.} \end{cases}$

► $Test2 \begin{cases} H_0 : \text{Le facteur B n'a pas d'effet sur y, ie :"}\mu_1 = \mu_2 = ... = \mu_j = ... = \mu_J\text{"} \\ H_1 : \text{Le facteur B a un effet sur y.} \end{cases}$

► $Test3 \begin{cases} H_0 : \text{Les facteurs A et B n'interagissent pas sur les résultats : "A et B n'ont pas d'effet sur y".} \\ H_1 : \text{Les facteurs A et B interagissent sur les résultats : "L'état du facteur A influence la réponse face au facteur B, et réciproquement".} \end{cases}$

4. L'équation fondamentale d'ANOVA2 avec répétitions

$$\underbrace{\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(y_{ijk}-\bar{y})^2}_{SCT} = \underbrace{\sum_{i=1}^{I} n_{i+}(\bar{y_{i..}}-\bar{y})^2}_{SCF_A} + \underbrace{\sum_{j=1}^{J} n_{+j}(\bar{y_{.j.}}-\bar{y})^2}_{SCF_B} + \underbrace{\sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij}(y_{ij.}-\bar{y_{i..}}-\bar{y_{.j.}}+\bar{y})^2}_{SCF_{AB}}$$
$$+\underbrace{\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(y_{ijk}-i\bar{y}.)^2}_{SCR}$$

telle que : $SCT = SCF_A + SCF_B + SCF_{AB} + SCR.$

où :

○ n_{i+} : représente le nombre de répétitions de niveau i de A.
○ n_{+j} : représente le nombre de répétitions de niveau j de B.
○ $\bar{y}_{i..} = \frac{1}{JK}\sum_{i,k} y_{ijk}$

- $\bar{y}_{.j.} = \frac{1}{IK}\sum_{i,k} y_{ijk}$
- $\bar{y}_{ij.} = \frac{1}{K}\sum_{k}^{K} y_{ijk}$
- $\bar{y} = \frac{1}{n}\sum_{i,j,k} y_{ijk}$

5. Tableau d'ANOVA2 avec répétitions

Source de variation	SC	ddl	MC	Statistique F
Facteur A	SCF_A	I-1	$MCF_A = \frac{SCF_A}{I-1}$	$Fc_A = \frac{MCF_A}{MCR}$
Facteur B	SCF_B	J-1	$MCF_B = \frac{SCF_B}{J-1}$	$Fc_B = \frac{MCF_B}{MCR}$
Interaction AB	SCF_{AB}	(I-1)(J-1)	$MCF_{AB} = \frac{SCF_{AB}}{(I-1)(J-1)}$	$Fc_{AB} = \frac{MCF_{AB}}{MCR}$
Résiduelle	SCR	n-IJ	$MCR = \frac{SCR}{n-IJ}$	/
Totale	SCT	n-1	$MCT = \frac{SCT}{n-1}$	/

6. Règle de décision

$*$ On compare les valeurs observées Fc_A, Fc_B, Fc_{AB} avec respectivement $F_{\alpha,(I-1),IJ(K-1)}$, $F_{\alpha,(J-1),IJ(K-1)}$, $F_{\alpha,(I-1)(J-1),IJ(K-1)}$.

$*$ On rejette H_0 si la valeur observée est plus grande que la valeur seuil.