INTRODUCTION

Analyse de la variance (ANOVA) : Etude de l'influence de un ou plusieurs facteurs sur la réponse observée sur un échantillon

Exemple : niveau d'anxiété de joueurs de tennis suivant le niveau de la compétition

International	National	Régional	Amateur
24,8	45,6	33,4	21,1
26,7	41,1	34,6	35,7
27,5	34,3	36,4	37,3
30,6	37,6	39,1	39,4
32,4	39,5	43,8	40,4
38,2		47,9	44,5
40,5		49,9	45,4
42,9		51,2	49,8
			50,1

ANOVA A UN FACTEUR

Notons A le facteur supposé influent (compétition : 4 niveaux).

Niveaux A	A_1	A_2		A_p
	x_1^1	x_2^1		x_p^1
	x_1^2	x_2^2		x_p^2
			...	
	$x_1^{n_1}$	$x_2^{n_2}$		$x_p^{n_p}$
Moyennes	$\bar{x}_1$	$\bar{x}_2$		$\bar{x}_p$
	X_1	X_2	...	X_p

Effet de A_i

$$x_i^j = \mu + \alpha_i + \varepsilon_i^j$$

$$\Uparrow$$

$$x_i^j = \mu_i + \varepsilon_i^j$$

$$\Uparrow$$

$$\Rightarrow \quad X_i \sim N(\mu_i, \sigma^2)$$

Hypothèse : A influe sur la moyenne et non sur la variance

A est influent $\Leftrightarrow$ Moyennes des sous-échantillons différentes

$$\Longleftrightarrow \begin{cases} H_0 & : \quad \mu_1 = \mu_2 = \ldots = \mu_p = \mu \\ H_1 & : \quad \exists i, j \quad \mu_i \neq \mu_j \end{cases}$$

ANOVA A UN FACTEUR

Décomposition de la variance :

$$S^2 = S_A^2 + S_R^2$$

Variance totale = variance facteur + variance résiduelle

$$S^2 = \sum_i \sum_j (X_i^j - \overline{X})^2 \qquad S_A^2 = \sum_{i=1}^{p} n_i (\overline{X}_i - \overline{X})^2 \qquad S_R^2 = \sum_i \sum_j (X_i^j - \overline{X}_i)^2$$

Si H_0 est vraie alors

$$F = \frac{S_A^2/(p-1)}{S_R^2/(n-p)}$$

suit une loi Fisher-Snedecor $F_\alpha(p\text{-}1,n\text{-}p)$

⇨ Si le rapport est supérieur à la valeur critique de la loi de Fisher-Snedecor, on conclut à une influence significative de A (H_1) avec α chance de se tromper.

ANOVA A UN FACTEUR

Tableau d'analyse de la variance :

Source de variation	Somme des carrées	Degré de liberté	Carré moyen
Expliquée	S_A^2	p-1	$\frac{S_A^2}{p-1}$
Résiduelle	S_R^2	n-p	$\frac{S_R^2}{n-p}$
Totale	S^2	n-1	

$\Rightarrow$ F

Exemple : Compétition

Source de variation	Somme des carrées	Degré de liberté	Carré moyen
Expliquée	**384,18**	**p-1=3**	**128,06**
Résiduelle	**1377,96**	**n-p=26**	**53**
Totale	**1762,14**	**n-1=29**	

$$\Rightarrow \mathbf{F = \frac{128,06}{53} = 2,42}$$

La valeur de F est inférieure au seuil 1% d'une variable F(3;26) qui est 4,64. On conclut que le niveau de compétition n'influe pas sur l'anxiété.

ANOVA A UN FACTEUR

CPI	CPGE	//
470	594	495
792	569	594
544	668	445
792	420	410
643	544	509
594	693	445
643	569	420
643	618	460
445	594	470
841		618
495		742
841		495
519		445
792		445
		594
		504

Exemple : Notes au TOEIC suivant provenance de l'étudiant

Source de variation	Somme des carrées	Degré de liberté	Carré moyen
Expliquée	149847,38	p-1=2	74923,69
Résiduelle	431545,17	n-p=36	11987,37
Totale	581392,56	n-1=38	

$$\mathbf{F} = \frac{74923{,}69}{11987{,}37} = 6{,}25$$

La valeur de F est supérieure au seuil 1% d'une variable F(2;36) qui est 5,25. On conclut que les notes au TOEIC dépendent de la provenance de l'étudiant avec 1% de chance de se tromper.

ANOVA A DEUX FACTEURS

Facteur A à p niveaux / Facteur B à q niveaux / r répétitions

		Niveaux A			
		A_1	A_2	A_i	A_p
Niveaux B	B_1				
	B_2				
	B_j			x_{ij1} x_{ij2} … x_{ijr}	
	B_q				

Moyenne : $\bar{x} = \frac{1}{pqr}\sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{r} x_{ijk}$

Moy. répétitions : $\bar{x}_{ij.} = \frac{1}{r}\sum_{k=1}^{r} x_{ijk}$

Moy. facteur A : $\bar{x}_{i..} = \frac{1}{rq}\sum_{j=1}^{q}\sum_{k=1}^{r} x_{ijk}$

Moy. facteur B : $\bar{x}_{.j.} = \frac{1}{rp}\sum_{i=1}^{p}\sum_{k=1}^{r} x_{ijk}$

Effet moyen

Effet de B_j

$X_{ij} \sim N(\mu_{ij},\sigma^2) \Rightarrow x_{ijk}=\mu_{ij}+\varepsilon_{ijk} \Rightarrow \mathbf{x_{ijk}} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}$

Effet de A_i

Effet interaction

ANOVA A DEUX FACTEURS

Décomposition de la variance :

$$S^2 = S_A^2 + S_B^2 + S_{AB}^2 + S_R^2$$

Var. totale = var. A + var. B + var. interaction + var. résiduelle

Variance totale : $$S^2 = \sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{r}(x_{ijk} - \overline{x})^2$$

Variance facteur A : $$S_A^2 = qr\sum_{i=1}^{p}(\overline{x}_{i..} - \overline{x})^2$$

Variance facteur B : $$S_B^2 = pr\sum_{j=1}^{q}(\overline{x}_{.j.} - \overline{x})^2$$

Variance interaction : $$S_{AB}^2 = r\sum_{i=1}^{p}\sum_{j=1}^{q}(\overline{x}_{ij.} - \overline{x}_{.j.} - \overline{x}_{i..} + \overline{x})^2$$

Variance résiduelle : $$S_R^2 = \sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{r}(x_{ijk} - \overline{x}_{ij.})^2$$

ANOVA A DEUX FACTEURS

Source de variation	Somme des carrées	Degré de liberté	Carré moyen	F
A	S_A^2	p-1	$C_A^2 = \frac{S_A^2}{p-1}$	$F_A = \frac{C_A^2}{C_R^2}$
B	S_B^2	q-1	$C_B^2 = \frac{S_B^2}{q-1}$	$F_B = \frac{C_B^2}{C_R^2}$
AB	S_{AB}^2	(p-1)(q-1)	$C_{AB}^2 = \frac{S_{AB}^2}{(p-1)(q-1)}$	$F_{AB} = \frac{C_{AB}^2}{C_R^2}$
R	S_R^2	pq(r-1)	$C_R^2 = \frac{S_R^2}{pq(r-1)}$	
Totale	S^2	pqr-1		

ANOVA A DEUX FACTEURS

- Si $F_{AB}>F_{\alpha}((p-1)(q-1);pq(r-1))$ alors on considère que l'interaction AB a une influence avec un risque α de se tromper
- Si $F_{A}>F_{\alpha}(p-1;pq(r-1))$ alors on considère que le facteur A a une influence avec un risque α de se tromper
- Si $F_{B}>F_{\alpha}(q-1;pq(r-1))$ alors on considère que le facteur B a une influence avec un risque α de se tromper

Rq. On ne connaît pas le risque (β) associé à la décision inverse