partie 6:ANALYSE DE LA VARIANCE

I - Généralités et définitions

1 : Soit 2 échantillons aléatoires et indépendants extraits de 2 populations qui ne diffèrent seulement que par le traitement qu'on leur a fait subir, et dont on mesure une certaine caractéristique (partie 5 cours 2ème année).
Si le test de comparaison des moyennes correspondantes conduit à rejeter l'hypothèse nulle d'égalité des moyennes, on conclura que la différence des traitements est la cause de la différence des moyennes de la caractéristique étudiée. Une telle démarche constitue le **plan d'expérience** le plus simple possible, où le traitement n'apparaît que sous 2 modalités, et où l'on avait volontairement fait en sorte que " toutes choses soient égales par ailleurs "

2 : Un facteur est une variable, ou un état, qui agit sur une mesure étudiée.

Les facteurs d'une étude sont les paramètres susceptibles d'influer sur une réponse étudiée.

Exemples :
- une dose d'engrais (le facteur) peut augmenter un rendement (la réponse).
- une alimentation enrichie en acides aminés (le facteur) peut favoriser la croissance d'un animal (la réponse).

L'expérimentateur devra établir une liste aussi exhaustive que possible des paramètres pouvant influer la variable mesurée, mais il devra faire la distinction entre :
- les facteurs étudiés introduits volontairement en vue d'en examiner les effets,
- les facteurs contrôlables, facteurs à fixer, mais non étudiés,
- les facteurs non contrôlables mais parfois identifiables, paramètres responsables de fluctuations aléatoires de la mesure étudiée (inhérents au matériel expérimental, à la manière d'opérer de l'expérimentateur).

Un facteur peut être qualitatif (variété, sexe des animaux) ou quantitatif (dose d'un produit).
Un facteur présente plusieurs modalités (variété A, B, …, mâle ou femelle) ou variantes (dose 1, dose 2…). On parle aussi de niveaux de facteurs.

3 : le facteur étudié peut présenter plus de 2 variantes et il est possible que l'on soit amené à étudier plusieurs facteurs simultanément. Dans ce cas on généralisera la méthode par la technique des plans d'expérience qui développe une véritable **stratégie de l'expérimentation** qui tient compte des problèmes spécifiques à étudier et du contexte réel. Elle permet de mettre en évidence l'**influence des facteurs** et leurs **interactions** éventuelles au moyen d'une unique expérimentation préalablement réfléchie. Toute expérimentation réalisée en laboratoire sera plus facile à mettre en œuvre et à analyser que celle réalisée en situation industrielle ou agricole qui présente un certain nombre de facteurs non contrôlés (cours 3ème année).

4 : On appellera traitement sur une unité statistique expérimentale la combinaison de différentes modalités, niveaux ou variantes de facteurs étudiés.

5 : On appellera l'unité statistique l'unité où un traitement et un seul est appliqué. Ce peut être une parcelle, un animal, un groupe d'animaux, une plante, une feuille. Les unités qui reçoivent le même traitement sont désignées comme des répétitions.

6 : Quand on expérimente sur des plantes ou des animaux on constate qu'ils présentent une certaine variabilité.
Dans le domaine végétal, l'hétérogénéité des résultats peut être due à la différence observée entre les caractéristiques physiques et/ou chimiques des parcelles, à la pente, aux irrégularités de pratiques culturales (irrigation, fertilisation…).
En zootechnie, l'hétérogénéité se manifeste dans l'effet « portée » par exemple, ou dans les variations pondérales sur une même journée.
Le dispositif expérimental va permettre :

- de mettre sous contrôle une partie de l'hétérogénéité pour qu'elle ne fausse pas l'analyse des résultats
- d'évaluer l'erreur que l'on qualifiera d'aléatoire ou **variance résiduelle** sur les sources d'hétérogénéité non contrôlables.

!!!Attention aux erreurs systématiques non maîtrisables statistiquement !!!! (des opérateurs différents pour une tâche donnée par exemple)

7 : Une réponse est le résultat de l'expérience. La réponse est quantitative (rendement, coût, nombre d'épis, poids de matière sèche, croissance…). C'est la variable que l'on mesure et qui rentrera dans les calculs nécessaires à l'analyse statistique.

8 : Le principe de l'interprétation statistique des résultats reste analogue à celui des tests d'hypothèses, on teste toujours une hypothèse nulle (on ne peut pas mettre en évidence des effets des facteurs étudiés) en comparant un critère statistique calculé à partir des réponses, à un critère théorique choisi pour un niveau de confiance ($1 - \alpha$), c'est à dire un risque de 1ère espèce égal à α de refuser H_0 alors qu'elle est vraie.

9 : L'analyse de la variance (ANOVA) est à proprement parler, une **méthode statistique d'interprétation des résultats rassemblés au cours d'une expérimentation** . L'analyse de la variance présuppose un modèle probabiliste traduisant la manière d'agir des facteurs, sur l'effet considéré. Ce modèle de type **linéaire** retient l'**additivité des effets moyens** et de leurs interactions ainsi que la **normalité de l'erreur expérimentale**. On suppose aussi que cette erreur suit la même loi pour toutes les observations.
L'analyse de la variance consiste en un **test global** de comparaison des moyennes entre elles et va permettre d'identifier les facteurs dont l'effet serait significatif. Si l'on rejette l'hypothèse nulle d'égalité des moyennes pour un facteur donné, on pourra examiner à titre indicatif les résultats individuels des moyennes en les comparant par 2 à l'aide d'un test de Student, en utilisant l'estimation de l'erreur aléatoire de l'analyse comme estimation de la variance intragroupe.

II - Analyse de la variance à une dimension

On est en présence d'un facteur A étudié à j (j = 1…k) variantes et i (i = 1 …n) répétitions.

$$\begin{array}{cccccc}
x_{11} & x_{12} & \dots & x_{1j} & \dots & x_{1k} \\
x_{21} & x_{22} & \dots & x_{2j} & \dots & x_{2k} \\
\dots & \dots & \dots & \dots & \dots & \dots \\
x_{i1} & x_{i2} & \dots & x_{ij} & \dots & x_{ik} \\
\dots & \dots & \dots & \dots & \dots & \dots \\
x_{n1} & x_{n2} & \dots & x_{nj} & \dots & x_{nk}
\end{array}$$

x_{ij} désigne la répétition de rang **i** de la variante **j**.

Remarque : le nombre des répétitions nj peut être différent, voir exemple en TD.

$$N=\sum_{j=1}^{j=k} n_j \qquad X_j=\sum_{i=1}^{i=n} x_{ij} \qquad \bar{x}_j=\frac{X_j}{n_j}$$

$$X=\sum_{j=1}^{j=k} X_j \qquad \mu=\bar{x}=\frac{X}{N} \qquad C=X\bar{x}$$

SCE désigne la somme des carrés des écarts à la moyenne
ddl désigne le degré de liberté
CM désigne le carré moyen = SCE / ddl

- ♣ le facteur étudié agit seulement sur la moyenne
- ♣ le facteur étudié agit de façon additive
- ♣ la loi de la variable X est une loi Normale de variance σ^2_ε indépendante des variantes du facteur étudié. Il s'agit de la variance aléatoire qui mesure l'erreur expérimentale.
- ♣ les effets des variantes du facteur étudié sont certains, le modèle est donc fixe.

Le modèle: $x_{ij} = \mu + \alpha_j + e_{ij}$ $\qquad \hat{x}_{ij} = \mu + \alpha_j \qquad e_{ij} = x_{ij} - \bar{x}_j$

Comparaison des variances résiduelles:

Après avoir établi la liste des résidus on teste l'hypothèse d'invariance de la variance et la normalité de l'erreur. On doit vérifier l'hypothèse d'invariance des variances intra groupes ($\hat{\sigma}^2{}_j$) au moyen d'un test de comparaison de plusieurs variances.

Tableau de l'analyse de la variance (ANOVA) :
Puis l'hypothèse que l'on va soumettre au test est que tous les effets des variantes du facteur étudié sont nuls ($\alpha_j = 0$). Pour cela on commence par calculer :

- la SCE entre les variantes du facteur notée **SCE a**

$$\text{SCE a} = \sum_{j=1}^{j=k} X_j \overline{x_j} - C$$

- la SCE totale notée **SCE t**

$$\text{SCE t} = \sum_{i,j} x_{ij}^2 - C$$

- la SCE résiduelle notée **SCE e**

$$\text{SCE e} = \sum_{i,j} (\hat{x_{ij}} - x_{ij})^2 = \text{SCE t} - \text{SCE a}$$

H_0: CM a / CM e = 1
H_1 : CM a / CM e > 1

Origine de la variation	SCE	ddl	CM	F calculé
Facteur A	SCE a	k - 1	SCE a / (k – 1) = CM a	**CM a / CM e**
Aléatoire	SCE e	N - k	SCE e / (N - k) = CM e	
total	SCE t	N - 1		

On compare le F calculé à un F théorique lu dans la table F noté $F_{1-\alpha}$ (k-1 ; N - k).

- Si **F calculé < F théorique**, on conserve H_0, ce qui signifie que l'on n'a pas pu mettre en évidence un effet significatif du facteur A sur les moyennes.
- Si l'hypothèse d'homogénéité des $\overline{x}_j$ a été rejetée (**F calculé > F théorique**) avec α% de risque d'erreur, on est naturellement conduit à rechercher, parmi les différentes moyennes celles qui ne sont sans doute pas imputable au hasard, on recherche parmi les k variantes du facteur étudié, celles qui agissent de façon significative.

Comparaison des moyennes :
Pour cela on range les moyennes par ordre croissant et on les compare 2 à 2.
Le critère statistique est celui de la comparaison de 2 moyennes dans le cas des petits échantillons avec comme estimation commune de la variance intra groupe la variance aléatoire estimée (CMe) et ddl = N - k .

Exemple d' ANOVA 1 facteur

A1	A2	A3
5	8	14
7	14	15
6	7	17
9	12	18
13	9	11

Calculs

Groupes	ni	Somme	Moyenne	Produit	SC
A1	5	40	8	320	360
A2	5	50	10	500	534
A3	5	75	15	1125	1155
total	**15**	**165**		**1945**	**2049**

μ = 11
C = 1815

Analyse des résidus

H_0: les k variances résiduelles sont égales
H_1: au moins 1 des k variances est supérieure aux autres

	A1	A2	A3
	-3	-2	-1
	-1	4	0
	-2	-3	2
	1	2	3
	5	-1	-4
SC =	40	34	30
σ^2 (ej) =	10	8,5	7,5

X^2 calculé = 0,084
α = 0,9591 (NS)
On n'a pas pu mettre en évidence qu'au moins une des variances résiduelle était supérieure aux autres, on conserve donc l'hypothèse d'invariance des variances résiduelles.

ANOVA

H_0: CM a / CM e = 1
H_1 : CM a / CM e > 1 (un effet du facteur est mis en évidence)

variations	*SCE*	*ddl*	*CM*	F calculé	*risque α*	*$F_{0,95}$ (2 ; 12)*
facteur A	130	2	65	7,5	0,0077 **	3,89
aléatoire	104	12	8,67			
Total	234	14				

F calculé > F théorique, on rejette H0 avec moins de 5% de risque d'erreur. On a pu mettre en évidence un effet du facteur A avec moins de 5% (0.77%) de chance de se tromper.

Comparaison des moyennes

H_0: $m_j = m_{j+1}$

H_1 : $m_j \neq m_{j+1}$

A1	A2	A3
8	10	15

$\sigma_d(m) = 1,862$

	Δm	t calculé	risque α	
A1 - A2	2	1,07	0,304	NS
A2 - A3	5	2,69	0,0198	*
A1 - A3	7	3,76	0,0027	**

t théorique = $t_{1-\alpha/2}(\nu)$ avec $\nu = 12$

α	$t_{1-\alpha/2}(12)$
0.05	2.179
0.01	3.055
0.001	4.318

Les moyennes m_1 et m_2 ne sont pas significativement différentes entre elles, on a pu mettre en évidence 2 groupes de moyennes, le groupe 1 (A_1 et A_2) et le groupe 2 (A_3).

III - Analyse de la variance à deux dimensions sans répétitions

On a <u>2</u> facteurs étudiés (ou un étudié et un mis sous contrôle = bloc) A et B présentant respectivement i (i = 1.....l) et j (j = 1 ....k) variantes.
Les observations résultent de tirages aléatoires et indépendants dans chacune des sous-populations définies par les combinaisons des variantes des facteurs étudiés.
Pour toute combinaison AiBj des variantes de facteurs, on retient <u>une seule</u> observation.
Les observations sont supposées sans erreur, l'erreur de mesure est d'un ordre de grandeur très inférieur à la variabilité expérimentale.

$$\begin{matrix} x_{11} & x_{12} & \dots & x_{1j} & \dots & x_{1k} \\ x_{21} & x_{22} & \dots & x_{2j} & \dots & x_{2k} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_{i1} & x_{i2} & \dots & x_{ij} & \dots & x_{ik} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_{l1} & x_{l2} & \dots & x_{lj} & \dots & x_{lk} \end{matrix}$$

x_{ij} désigne la valeur mesurée pour la combinaison **AiBj**.

$N=\sum_{j=1}^{j=k} n_{.j}=\sum_{i=1}^{i=l} n_{i.}$	$X_{.j}=\sum_{i=1}^{i=l} x_{ij}$	$X_{i.}=\sum_{j=1}^{j=k} x_{ij}$	$X=\sum_{j=1}^{j=k} X_{.j}=\sum_{i=1}^{i=l} X_{i.}$
$\bar{x}_{i.}=\frac{X_{i.}}{n_{i.}}$	$\bar{x}_{.j}=\frac{X_{.j}}{n_{.j}}$	$\mu=\bar{x}=\frac{X}{N}$	$C=X\bar{x}$

- les facteurs A et B agissent seulement sur la moyenne
- les facteurs étudiés agissent de façon additive
- la loi de la variable X est une loi Normale de variance σ^2_ε indépendante des variances des facteurs étudiés
- les effets de facteurs sont certains
- on supposera négligeable l'interaction du 1er ordre AB

Le modèle :

$x_{ij} = \mu + \alpha_i + \beta_j + e_{ij}$ $\qquad \hat{x}_{ij} = \mu + \alpha_i + \beta_j$ $\qquad e_{ij} = x_{ij} + \mu - \bar{x}_{i.} - \bar{x}_{.j}$

Comparaison des variances résiduelles

On établit la liste des résidus pour vérifier l'hypothèse d'invariance de la variance et la normalité de l'erreur. On doit vérifier l'hypothèse d'invariance des variances résiduelles $\tilde{\sigma}^2_{.j}(\varepsilon)$ et $\tilde{\sigma}^2_{i.}(\varepsilon)$au moyen des tests appropriés.

Tableau de l'analyse de la variance (ANOVA)

Puis l'hypothèse que l'on va soumettre au test est que, pour chacun des facteurs, tous les effets des variantes du facteur étudié sont nuls. Pour cela on commence par calculer :

☞ la **SCE entre les variantes** de chacun des facteur:

$$\text{SCE a} = \sum_{i=1}^{i=l} X_{i.}\bar{x}_{i.} - C \qquad \text{SCE b} = \sum_{j=1}^{j=k} X_{.j}\bar{x}_{.j} - C$$

☞ la SCE totale notée **SCEt**:

$$\text{SCE t} = \sum_{i,j} x_{ij}^2 - C$$

☞ la SCE résiduelle notée **SCEe**:

SCE e = SCE t – SCE a – SCE b

H_0: CM a / CM e = 1 **H'_0: CM b / CM e = 1**
H_1: CM a / CM e > 1 **H'_1: CM b / CM e > 1**

Origine de la variation	**SCE**	**ddl**	**CM**	**F calculé**
Facteur A	SCE a	l - 1	SCE a / (l – 1) = CM a	**CM a / CM e = Fa**
Facteur B	SCE b	k - 1	SCE b / (k – 1) = CM b	**CM b / CM e = Fb**
Aléatoire	SCE e	(k-1)(l-1)	SCE e / (k-1)(l-1) = CM e	
total	SCE t	N - 1		

hypothèses nulles	**critère statistique calculé**	**critère statistique théorique**
pas d'effet du facteur A	F a = CM a / CM e	$F_{1-\alpha}$ [(l-1) ; (k-1)(l-1)]
pas d'effet du facteur B	F b = CM b / CM e	$F_{1-\alpha}$ [(k-1) ; (k-1)(l-1)]

Interprétation des résultats : si F calculé > F théorique alors on a pu mettre en évidence un effet du facteur testé avec un risque de se tromper égal à α.

Comparaison des moyennes

Pour chaque facteur séparément on range les moyennes par ordre croissant et on les compare 2 à 2.

Le critère statistique est celui de la comparaison de 2 moyennes dans le cas des petits échantillons avec comme estimation commune de la variance intra groupe la variance aléatoire estimée (CMe) et ddl = (k-1)(l-1).

Pour chaque facteur significatif on compare et regroupe les moyennes.

Exemple ANOVA 2 facteurs sans répétitions

	B1	B2	B3	B4
A1	44	22	36	34
A2	47	43	41	53
A3	0	9	10	17
A4	36	14	1	34

Calculs

	ni	Somme	Moyenne	Produit
A1	4	136	34	4624,00
A2	4	184	46	8464,00
A3	4	36	9	324,00
A4	4	85	21,25	1806,25
total				**15218,25**
B1	4	127	31,75	4032,25
B2	4	88	22	1936,00
B3	4	88	22	1936,00
B4	4	138	34,5	4761,00
total				**12665,25**

$N = 16$
$X = 441$
$\Sigma x^2_{ij} = 16539$
$\mu = 27,56$
$C = 12155,1$

Analyse des résidus
H_0: les k variances résiduelles selon le facteur A sont égales
H_1: au moins 1 des k variances selon le facteur A est supérieure aux autres
H'_0: les l variances résiduelles selon le facteur B sont égales
H'_1: au moins 1 des l variances selon le facteur B est supérieure aux autres

					σ^2(ei)
	5,81	-6,44	7,56	-6,94	60,18
	-3,19	2,56	0,56	0,06	5,68
	-13,19	5,56	6,56	1,06	83,02
	10,56	-1,69	-14,69	5,81	121,31
σ^2 (ej)	109,81	27,27	105,43	27,68	

Facteur A: **X^2 calculé =5.4** **ddl = 3** **alpha = 0,145**
Facteur B: **X^2 calculé =2.60** **ddl = 3** **alpha = 0,4568**

Pour chaque facteur, on n'a pas pu mettre en évidence qu'au moins une des variances résiduelle était supérieure aux autres, on conserve donc l'hypothèse d'invariance des variances résiduelles.

ANOVA

H_0: CM a / CM e = 1

H_1 : CM a / CM e > 1 (un effet du facteur A est mis en évidence)

H'_0: CM b / CM e = 1

H'_1 : CM b / CM e > 1 (un effet du facteur B est mis en évidence)

variations	SCE	ddl	CM	F calculé	risque α	F $_{0,95}$ (3;9)	
facteur A	3063,19	3	1021,06	11,34	0,0021 (**)	3,86	
facteur B	510,19	3	170,06	1,89	0,2020 (NS)	3,86	
Erreur	810,56	9	90,06				
Total	4383,94	15					

Fa calculé > F théorique, on rejette H0 avec moins de 5% de risque d'erreur. On a pu mettre en évidence un effet du facteur A avec moins de 5% (0.21%) de chance de se tromper.

Fb calculé < F théorique, on conserve H0. On n'a pas pu mettre en évidence un effet du facteur B.

On ne comparera que les moyennes selon le facteur A

Comparaison des moyennes

H_0: $m_j = m_{j+1}$

H_1 : $m_j \neq m_{j+1}$

A3	A4	A1	A2
9	21,25	34	46

σ d(m) = 6.71

	mi - mj	t calculé	alpha	
A3 - A4	12,25	1,83	0,1012	NS
A4 - A1	12,75	1,90	0,0899	NS
A2 - A1	12	1,79	0,1074	NS
A1 - A3	25	3,73	0,0047	**
A4 - A2	24,75	3,69	0,0050	**

t théorique = t $_{1-\alpha/2}$ (ν) avec ν = 9

α	t $_{1-\alpha/2}$ (9)
0.05	2.262
0.01	3.25
0.001	4.781

A3	X		
A4	X	X	
A1		X	X
A2			X

On a pu mettre en évidence 3 groupes de moyennes, le groupe 1 (A_3 et A_4) et le groupe 2 (A_1 et A_4) et le groupe 3 (A_1 et A_2).

IV - Analyse de la variance à 2 dimensions avec répétitions

On a 2 facteurs étudiés A et B présentant respectivement i (i = 1.....l) et j (j = 1 ....k) variantes. Les observations résultent de tirages aléatoires et indépendants dans chacune des sous-populations définies par les combinaisons des variantes des facteurs étudiés. Pour toute combinaison AiBj des variantes de facteurs, on retient r observations (r = 1...n).Les observations sont supposées sans erreur, l'erreur de mesure est d'un ordre de grandeur très inférieur à la variabilité expérimentale.

- les facteurs A et B et leur interaction agissent seulement sur la moyenne μ
- les facteurs étudiés et leur interaction agissent de façon additive
- la loi de la variable X est une loi Normale de variance σ^2_ε indépendante des variances des facteurs étudiés
- les effets principaux et les interactions de facteurs sont certains

le modèle: $x_{ijr} = \mu + \alpha_i + \beta_j + \alpha_i\beta_j + e_{ijr}$

$$\hat{x}_{ijr} = \mu + \alpha_i + \beta_j + \alpha_i\beta_j$$

$$e_{ijr} = x_{ijr} - \overline{x}_{ij}$$

Comparaison des variances résiduelles

On établit la liste des résidus pour vérifier l'hypothèse d'invariance des variances résiduelles de chaque facteur et de celle de l'interaction ainsi que la normalité de l'erreur.

Remarque : pour simplifier les exercices nous ne vérifierons que l'hypothèse d'invariance des variances selon l'interaction. Sauf exception nous supposerons « vérifiée » l'hypothèse d'invariance des variances résiduelles des facteurs.

Tableau de l'analyse de la variance (ANOVA)

Puis l'hypothèse que l'on va soumettre au test est que, pour chacun des facteurs, tous les effets des variantes du facteur étudié sont nuls ainsi que l'effet de l'interaction entre les facteurs.

H_0: CMa / CMe = 1 H'_0: CMb / CMe = 1 H''_0: CMab / CMe = 1

H_1: CMa / CMe > 1 H'_1: CMb / CMe > 1 H''_1: CMab / CMe > 1

origine de la variation	SCE	ddl	CM	F calculé
facteur A	SCE a	l - 1	CM a = SCE a / (l-1)	CM a / CM e
facteur B	SCE b	k - 1	CM b = SCE b / (k-1)	CM b / CM e
interaction A*B	SCE ab	(k-1)(l-1)	CM ab = SCE ab/(k-1)(l-1)	CM ab/ CM e
Aléatoire	SCE e	k1(n-1)	CM e = SCE e / k l (n-1)	
Total	SCE t	N - 1		

En pratique, les éléments du tableau sont déterminés à l'aide des formules suivantes:

SCE a = $\sum_{i=1}^{i=l} X_{i.}\overline{x}_{i.} - C$ SCE b = $\sum_{j=1}^{j=k} X_{.j}\overline{x}_{.j} - C$

SCE ab = $\sum_{ij} X_{ij}\overline{x}_{ij.} - C - SCEa - SCEb$ SCE t = $\sum_{i,j,r} X_{ijr}^2 - C$

SCE e = SCE t – SCE a – SCE b – SCE ab $\quad$ SCE e = $\sum_{i,j}(\hat{x}_{ijr}-x_{ijr})^2$

Interprétation des résultats :

hypothèses nulles	critère statistique calculé	critère statistique théorique
pas d'effet du facteur A	Fa = CM a / CM e	$F_{1-\alpha}$ [(l-1) ; k l (n-1)]
pas d'effet du facteur B	F b = CM b / CM e	$F_{1-\alpha}$ [(k-1) ; k l (n-1)]
pas d'effet d'interaction	Fab =CM ab / CM e	F1- α [(k-1)(l-1) ; k l (n-1)]

si F calculé > F théorique alors on a pu mettre en évidence un effet du facteur et/ou de l'interaction testé avec un risque de se tromper égal à α.

Comparaison des moyennes

- Si l'interaction est non significative on teste et on regroupe les moyennes de chaque facteur influant comme dans les cas précédents. Le critère statistique est celui de la comparaison de 2 moyennes dans le cas des petits échantillons avec comme estimation commune de la variance intra groupe la variance aléatoire estimée (CMe) et ddl = kl(n-1).
- Si l'interaction est significative on classe les mij moyennes dans l'ordre croissant et on applique le test de comparaison de 2 moyennes avec comme estimation commune de la variance intra groupe la variance aléatoire estimée (CMe) et ddl = kl(n-1).

Exemple ANOVA 2 facteurs avec répétitions

	A1	A2
B 1	64	43
B 1	66	58
B 1	62	46
B 1	72	53
B2	70	60
B2	69	76
B2	65	58
B2	68	54

Calculs

B 1	A1	A2	Total
n	4	4	8
Somme	264	200	464
Moyenne	66	50	58
Produit	17424	10000	26912

B2	A1	A2	Total	total B
n	4	4	8	60712
Somme	272	248	520	
Moyenne	68	62	65	total inter
Produit	18496	15376	33800	61296

Total	A1	A2		
n	8	8		
Somme	536	448		
Moyenne	67	56		total A
Produit	35912	25088		61000

N = 16
X = 984
$\Sigma xijr^2$ = 61784
μ = 61.5
C = 60516

Analyse des résidus

H_0: les k variances résiduelles selon le facteur A sont égales
H_1: au moins 1 des k variances selon le facteur A est supérieure aux autres
H'_0: les l variances résiduelles selon le facteur B sont égales
H'_1: au moins 1 des l variances selon le facteur B est supérieure aux autres
H''_0: les kl variances résiduelles sont égales (*)
H''_1: au moins 1 des kl variances est supérieure aux autres

(*) Nous ne testerons ici que le 3ème groupe d'hypothèse pour ne pas alourdir l'exemple.

	A1	A2
B1	-2	-7
	0	8
	-4	-4
	6	3
Variance	18,67	46,00
B2	2	-2
	1	14
	-3	-4
	0	-8
Variance	4,67	93,33

X^2 calc = 5.97
ddl = 3
alpha = 0.1132

On n'a pas pu mettre en évidence qu'au moins une des kl variances résiduelle était supérieure aux autres, on conserve donc l'hypothèse d'invariance des variances résiduelles.

ANOVA
H_0: CM a / CM e = 1
H_1 : CM a / CM e > 1 (un effet du facteur A est mis en évidence)
H'_0: CM b / CM e = 1
H'_1 : CM b / CM e > 1 (un effet du facteur B est mis en évidence)
H''_0: CMab / CM e = 1
H''_1 : CMab / CM e > 1 (un effet d'interaction entre A et B est mis en évidence)

variations	SCE	ddl	CM	F calculé	alpha	$F_{0,95}$ (1;12)
facteur B	196	1	196	4,82	0,0485(*)	4,75
facteur A	484	1	484	11,90	0,0048(**)	4,75
Interaction	100	1	100	2,46	0,1428(NS)	4,75
Aléatoire	488	12	40,67			
Total	1268	15				

Fab calculé < F théorique, on conserve H0. On n'a pas pu mettre en évidence un effet d'interaction.
Fa calculé > F théorique, on rejette H0 avec moins de 5% de risque d'erreur. On a pu mettre en évidence un effet du facteur A avec moins de 5% (0.48%) de chance de se tromper.
Fb calculé > F théorique, on rejette H0 avec moins de 5% de risque d'erreur. On a pu mettre en évidence un effet du facteur B avec moins de 5% (4.85%) de chance de se tromper.
On ne comparera que les moyennes selon le facteur A, puis selon le facteur B.

Comparaison des moyennes

Inutile à faire dans cette situation car chaque facteur ne comporte que 2 modalités! On a donc dans chaque cas 2 moyennes à comparer mais on sait par l'ANOVA qu'elles sont différentes !

V - Analyse de la variance à trois dimensions sans répétition

On a 3 facteurs étudiés A et B et C présentant respectivement **l** et **k** et **s** variantes.
Les observations résultent de tirages aléatoires et indépendants dans chacune des sous-populations définies par les combinaisons des variantes des facteurs étudiés.
Pour toute combinaison **AiBjCt** des variantes de facteurs, on retient une seule observation.
Les observations sont supposées sans erreur, l'erreur de mesure est d'un ordre de grandeur très inférieur à la variabilité expérimentale.

le modèle: $x_{ijt} = \mu + \alpha_i + \beta_j + \gamma_t + \alpha_i\beta_j + \alpha_i\gamma_t + \beta_j\gamma_t + e_{ijt}$

$$e_{ijt} = x_{ijt} - \overline{x}_{ij.} - \overline{x}_{i.t} - \overline{x}_{.jt} + \overline{x}_{i..} + \overline{x}_{.j.} + \overline{x}_{..t} - \mu$$

- ♠ les facteurs A et B et C agissent seulement sur la moyenne
- ♠ les facteurs étudiés agissent de façon additive
- ♠ la loi de la variable X est une loi Normale de variance σ^2_ε indépendante des variances des facteurs étudiés
- ♠ les effets de facteurs sont certains
- ♠ On supposera négligeable l'interaction du 2ème ordre ABC

Comparaison des variances résiduelles

On établit la liste des résidus pour vérifier l'hypothèse d'invariance des variances résiduelles de chaque facteur et de celle de l'interaction ainsi que la normalité de l'erreur.

Remarque : pour simplifier les exercices nous supposerons « vérifiée » l'hypothèse d'invariance des variances résiduelles.

Tableau de l'analyse de la variance (ANOVA)

Les hypothèses nulles soumises aux tests sont l'absence d'effet de chaque facteur et l'absence d'effet de chaque interaction du premier ordre.

H_0: CMa / CMe = 1 H'_0: CMb / CMe = 1 H''_0: CMc / CMe = 1
H_1: CMa / CMe > 1 H'_1: CMb / CMe > 1 H''_1: CMc / CMe > 1

H'''_0: CMab / CMe = 1 H''''_0: CMac / CMe = 1 H'''''_0: CMbc / CMe = 1
H'''_1: CMab / CMe > 1 H''''_1: CMac / CMe > 1 H'''''_1: CMbc / CMe > 1

origine de la variation	SCE	ddl	CM	F calculé
facteur A	SCE a	l - 1	CM a = SCE a / (l-1)	CM a / CM e
facteur B	SCE b	k - 1	CM b = SCE b / (k-1)	CM b / CM e
facteur C	SCE c	s - 1	CM c = SCE c / (s-1)	CM c / CM e
interaction AB	SCE ab	(k-1)(l-1)	CM ab = SCE ab / (k-1)(l-1)	CM ab /CM e
interaction AC	SCE ac	(l-1)(s-1)	CM ac = SCE ac / (l-1)(s-1)	CM ac / CM e
interaction BC	SCE bc	(k-1)(s-1)	CM bc = SCE bc / (k-1)(s-1)	CM bc /CM e
Aléatoire	SCE e	(k-1)(l-1)(s-1)	CM e = SCE e / (k-1)(l-1)(s-1)	
Total	SCE t	N - 1		

En pratique, les éléments du tableau sont déterminés à l'aide des formules suivantes:

$$SCE\ a = \sum_{i=1}^{i=l} X_{i..}\bar{x}_{i..} - C \qquad SCE\ b = \sum_{j=1}^{j=k} X_{.j.}\bar{x}_{.j.} - C \qquad SCE\ c = \sum_{t=1}^{t=s} X_{..t}\bar{x}_{..t} - C$$

$$SCE\ ab = \sum_{i,j} X_{ij.}\bar{x}_{ij.} - C - SCEa - SCEb$$

$$SCE\ bc = \sum_{j,t} X_{.jt}\overline{x_{.jt}} - C - SCEb - SCEc$$

$$SCE\ ac = \sum_{i,t} X_{i.t}\bar{x}_{i.t} - C - SCEa - SCEc$$

$$SCE\ e = \sum_{i,j,t} (\hat{x}_{ijt} - x_{ijt})^2 \qquad SCE\ t = \sum_{i,j,t} x_{ijt}^2 - C$$

SCE e = SCE t – SCE a – SCE b – SCE c – SCE ab – SCE ac – SCE bc

Interprétation des résultats :

hypothèses nulles	Critère statistique calculé	critère statistique théorique
pas d'effet du facteur A	F a = CM a / CM e	$F_{1-\alpha}$ [(l-1) ; (k-1)(l-1)(s-1)]
pas d'effet du facteur B	F b = CM b / CM e	$F_{1-\alpha}$ [(k-1) ; (k-1)(l-1)(s-1)]
pas d'effet du facteur C	F c = CM c / CM e	$F_{1-\alpha}$ [(s-1) ; (k-1)(l-1)(s-1)]
pas d'interaction AB	F ab = CM ab/ CM e	$F_{1-\alpha}$ [(l-1)(k-1) ; (k-1)(l-1)(s-1)]
pas d'interaction AC	F ac = CM ac/ CM e	$F_{1-\alpha}$ [(l-1)(s-1) ; (k-1)(l-1)(s-1)]
pas d'interaction BC	F bc = CM bc/ CM e	$F_{1-\alpha}$ [(s-1)(k-1) ; (k-1)(l-1)(s-1)]

On interprète en premier les effets d'interactions des facteurs entre eux puis ceux des effets principaux des facteurs.
Si F calculé > F théorique alors on a pu mettre en évidence un effet du facteur et/ou de l'interaction testé avec un risque de se tromper égal à α.

Comparaison des moyennes

- Si aucune interaction n'est significative on teste et on regroupe les moyennes de chaque facteur influant comme dans les cas précédents. Le critère statistique est celui de la comparaison de 2 moyennes dans le cas des petits échantillons avec comme estimation commune de la variance intra groupe la variance aléatoire estimée (CMe) et ddl = (k-1).(l.-1) (n-1).
- Si une (ou plusieurs) interaction(s) est (sont) significative(s) on classe les m_{ij} (ou m_{it} ou m_{jt}) moyennes dans l'ordre croissant et on applique le test de comparaison de 2 moyennes avec comme estimation commune de la variance intra groupe la variance aléatoire estimée (CMe) et ddl = (k-1).(l.-1) (n-1).

On pourrait continuer avec 3 facteurs avec répétitions, 4 facteurs , 5 facteurs……..etc mais les calculs seraient un peu longs dans le cadre de ce cours!!! ………Non?