

0-1. Un peu de vocabulaire

Observation, interprétation, expérience, expérimentation, mesure, échelle de mesure, moyenne, variance, hypothèse, test.

0-2. Constituants d'une expérience

0-2-1 Sujets :

Définition : Les unités statistiques sur lesquelles porte l'expérience sont appelées unités expérimentales. L'ensemble des sujets constitue la population (souvent c'est un échantillon). On note S cet ensemble et s un sujet. En psychologie il s'agit souvent d'individus, on les désigne par le terme « sujet ».

0-2-2 Variable dépendante :

Définition : On appelle variable dépendante (VD) la réponse du système aux différentes conditions expérimentales. Cette réponse est mesurée sur les sujets de l'expérience. Elle peut être uni ou multidimensionnelle. On note Y cette variable et $Y(s)$ ou Y_s la valeur qu'elle prend pour le sujet s .

0-2-3 Les facteurs de variation :

0-2-3-1 Les variables indépendantes :

Définition : Les variables indépendantes (VI) sont les variables représentant les causes postulées des variations de la VD. Le but de l'expérimentation étant de mettre en évidence l'effet de ces variables indépendantes.

Remarque : On emploie aussi le terme de facteur de variation ou facteur expérimental ou encore (plus simplement) facteur.

Par usage le terme de « variables indépendantes » est utilisé dans un contexte méthodologique alors que le terme de facteur est utilisé dans un contexte statistique.

Définition : On appelle variables indépendantes provoquées les variables indépendantes pouvant être manipulées (contrôlées) par l'expérimentateur.

Remarque (fondamentale): Dans le cas des variables indépendantes provoquées l'expérimentateur a la possibilité de répartir aléatoirement les sujets dans différents groupes expérimentaux et ceci avant d'agir sur les VI.

Définition : Les variables indépendantes invoquées (ou étiquettes) sont des caractéristiques « naturelles » sur lesquelles l'expérimentateur ne peut agir, mais elles lui servent pour repérer les conditions expérimentales.

Remarque : Certains auteurs réservent le mot « expérimental » aux situations d'observations lorsque toutes les variables indépendantes sont provoquées et « quasi-expérimental » lorsque le plan d'expérience contient des variables indépendantes invoquées.

Définition : On dit que deux variables indépendantes sont confondues si chaque modalité de l'une est associée à une modalité de l'autre.

0-2-3-2 Les variables parasites :

Les variables indépendantes non prises en considération dans l'expérience sont appelées variables parasites. Les variables parasites identifiables sont celles que l'on peut totalement identifier. On cherche alors à les contrôler pour « effacer » leur effet. Les variables parasites non identifiables sont toutes les caractéristiques propres à l'unité statistique. Ces dernières peuvent être contrôlées par randomisation.

0-3. Les effets des facteurs

0-3-1 Les niveaux des effets des facteurs :

Dans de nombreuses études, les phénomènes sont plus complexes que la simple relation de cause à effet entre les VI et les VD. On peut imaginer tout un réseau de causalités. L'effet d'un facteur peut se traduire alors à plusieurs niveaux :

a) effet principal : effet direct sur la VD que l'on aurait mesuré dans une expérience où seul ce facteur serait présent.

b) effet d'interaction : effet du facteur sur les effets des autres facteurs. On distingue alors plusieurs niveaux d'interaction :

- * Interaction d'ordre 1 : interaction d'un facteur sur l'effet d'un autre facteur.
- * Interaction d'ordre 2 : interaction d'un facteur sur l'interaction entre deux autres facteurs
- * etc. ...

0-3-2 Nature des effets des facteurs :

On distingue deux cas :

a) Les facteurs à effets fixes : les modalités du facteur qui intéressent le chercheur sont choisies de façon systématique par l'expérimentateur. On veut mesurer l'effet de chaque modalité prise par elle-même.

b) Les facteurs à effets aléatoires : toutes les modalités du facteur ne sont pas présentes dans l'expérience. Les modalités retenues ne sont pas choisies de façon systématique mais de façon aléatoire. Le chercheur ne s'intéresse donc pas à chaque modalité prise individuellement mais à l'effet global du facteur. Il lui sera alors possible d'étendre des conclusions à d'autres modalités de la variable indépendante non présentes lors de l'expérimentation.

Dans de nombreux modèles on introduit le facteur de variation « sujet » dont les modalités sont les unités statistiques de l'expérience. Le facteur sujet est par construction même de l'expérience un facteur à effets aléatoires.

0-4. Plans d'expérience

0-4-1 Généralités :

L'expérimentateur définit un ensemble de p facteurs élémentaires A_1, A_2, \dots, A_p . Chaque facteur possède un nombre n_i fini de modalités $A_i = \{a_1^i, a_2^i, \dots, a_{n_i}^i\}$.

Chaque observation de la variable dépendante correspond à une combinaison des modalités des facteurs élémentaires $a = \{a_{i1}^1, a_{i2}^2, \dots, a_{ip}^p\}$. $a_{ij}^j \in A_j$

Notons E l'ensemble de toutes les combinaisons des modalités des facteurs élémentaires

$$E = A_1 \times A_2 \times \dots \times A_p.$$

Chaque condition expérimentale envisagée dans l'expérience est décrite par un élément de E

Lors de l'expérience il arrive souvent que toutes les combinaisons ne sont pas retenues.

Cependant chaque modalité de chaque facteur est présente au moins une fois dans l'expérience.

L'ensemble des combinaisons retenues définit le plan d'expérience.

Définition : On appelle plan d'expérience un sous-ensemble E^* de E tel que chaque modalité de chaque facteur soit présente au moins une fois dans les éléments de E^* . Les éléments de E^* sont appelés conditions expérimentales.

0-4-2 Relations entre facteurs :

Un plan d'expérience se conçoit en précisant les facteurs élémentaires pris en compte mais aussi les relations entre ces facteurs. Nous étudierons deux relations fondamentales : la relation de croisement et la relation d'emboîtement. Ces deux relations permettent de décrire un ensemble particulier de plans d'expérience (ensemble des plans quasi-complets).

Définition : Deux facteurs A et B sont croisés si chaque modalité de l'un apparaît avec chaque modalité de l'autre. On note $A \times B$ (ou $B \times A$).

Remarque : La définition précédente se généralise à plusieurs facteurs. p facteurs sont croisés dans leur ensemble si chaque modalité de l'un apparaisse avec toutes les modalités de tous les autres facteurs.

Remarque (importante) : Des facteurs croisés deux à deux ne sont pas nécessairement croisés dans leur ensemble.

Définition : Un facteur A est emboîté dans un facteur B si chaque modalité de A est associée à exactement une modalité de B. A est appelé le facteur emboîté et B le facteur emboîtant. On note $A \langle B \rangle$.

Remarque : un facteur peut être emboîté dans un facteur lui-même emboîté dans un troisième facteur. Dans ce cas le 1^{er} facteur est emboîté dans le 3^{ème}.

Définition : L'emboîtement est dit équilibré (balancé) si et seulement si chaque modalité du facteur emboîtant est associée au même nombre de modalités du facteur emboîté.

Définition : Deux facteurs A et B sont confondus si et seulement si A est emboîté dans B et B est emboîté dans A.

0-4-3 Les plans d'expérience de base:

Définition : Un plan d'expérience est dit complet si et seulement si il est égal au croisement de tous ses facteurs élémentaires ($E^* = E$). Autrement dit il contient toutes les combinaisons des modalités des facteurs élémentaires.

Définition : Un plan d'expérience est dit quasi-complet si et seulement si pour tout couple de facteurs les deux facteurs sont soit croisés soit emboîtés et si tous les facteurs croisés deux à deux sont croisés dans leur ensemble.

Dans toute la suite nous ne nous intéresserons qu'à des plans quasi-complets.

Définition : Un plan d'expérience est dit à un facteur et à mesures indépendantes s'il met en jeu un seul facteur A outre les sujets qui ne sont « utilisés » qu'une seule fois (un sujet ne se retrouve que dans un seul groupe expérimental). On a donc $S \langle A \rangle$. On parle de plan inter.

Définition : Un plan d'expérience est dit à un facteur et à mesures répétées s'il met en jeu un seul facteur A outre les sujets qui sont « utilisés » dans les différentes conditions expérimentales. On a donc $S \times A$. On parle de plan intra.

Définition : Un plan d'expérience est dit à deux facteurs et à mesures indépendantes si on a $S \langle A \times B \rangle$

Définition : Un plan d'expérience est dit à deux facteurs et à mesures complètement répétées si les sujets sont utilisés dans l'ensemble des conditions expérimentales. On a donc $S \times A \times B$.

Définition : Un plan quasi-complet de la forme $S \langle A \rangle \times B$ est dit mixte.

I-1. Modèle statistique

Les données sont constituées d'une famille de N valeurs y_1, y_2, \dots, y_N résultant de l'observation de la variable dépendante dans les différentes conditions expérimentales définies par le plan de l'expérience. On appelle protocole expérimental cette famille et on note $y(e)$ l'observation de la VD dans la condition expérimentale e . L'élaboration du modèle qui décrit le processus de génération des données doit prendre en compte d'une part le fait que les mesures sont faites sur un échantillon aléatoire d'unités statistiques (fluctuations d'échantillonnage) et d'autre part qu'interviennent de façon aléatoire au cours de l'expérimentation des facteurs non contrôlés tels que par exemple les erreurs de mesure ou des facteurs individuels. L'observation $y(e)$ est alors considérée comme la réalisation $Y(e)$ d'une variable aléatoire Y . Le modèle du score décrit les effets des facteurs sur la variable aléatoire par le modèle algébrique : $Y(e) = Z(e) + \varepsilon(e)$
où :

- $Z(e)$, appelé modèle structurel, décrit les effets des facteurs du plan
- $\varepsilon(e)$, appelé résidu, mesure l'écart entre la variable réponse et le modèle structurel dû aux effets de facteurs non pris en compte dans le plan de l'expérience.

Ce modèle linéaire est caractérisé par des hypothèses portant sur le résidu

- $\varepsilon(e)$ est une variable aléatoire de moyenne nulle.
- $\text{Var}(\varepsilon(e)) = \sigma^2$ pour tout e homoscédasticité ou homogénéité des variances des résidus
- $\text{Cor}[(\varepsilon(e), \varepsilon(e')) = 0$ pour tout $e \neq e'$ non corrélation des résidus

Ce modèle linéaire est aussi caractérisé par des hypothèses portant sur le modèle structurel

$$Z(e) = \mu(e) + A(e)$$

- $\mu(e)$ est une fonction déterministe de e qui décrit les effets des modalités des facteurs à effets fixes présentes dans e .
- $A(e)$ est une variable aléatoire non corrélée avec $\varepsilon(e)$, qui décrit les effets des facteurs à effets aléatoires
- $\varepsilon(e)$ suit une loi normale $\mathcal{N}(0, \sigma^2)$

Selon la forme du modèle structurel, on distingue trois classes de modèles d'analyse de la variance :

ANOVA modèle 1 : Le modèle ne contient pas de terme aléatoire $A(e)$: $Y(e) = \mu(e) + \varepsilon(e)$

Ce modèle correspond à des plans où tous les facteurs sont à effets fixes. Leurs effets portent uniquement sur la moyenne de la VD. $Y(e)$ est distribué selon une $\mathcal{N}(\mu(e), \sigma^2)$.

ANOVA modèle 2 : (modèle des composantes de la variance) Le terme déterministe est constant

$$Y(e) = \mu + A(e) + \varepsilon(e)$$

Ce modèle correspond à des plans où tous les facteurs sont à effets aléatoires. Leurs effets portent sur la variance de la VD. $Y(e)$ est distribué selon une $\mathcal{N}(\mu, \sigma^2 + \sigma_{A(e)}^2)$.

ANOVA modèle 3 : (modèle mixte) c'est le modèle qui mélange les deux cas précédents

Ce modèle correspond à des plans où sont présents des facteurs à effets fixes et des facteurs à effets aléatoires. $Y(e)$ est distribué selon une $\mathcal{N}(\mu(e), \sigma^2 + \sigma_{A(e)}^2)$.

I-2. Décomposition de la variance associée à un facteur : (rappel)

Dans toute la suite Y désigne la variable dépendante et G un facteur à r modalités. Les N observations sont réparties en r classes g_1, g_2, \dots, g_r contenant respectivement n_1, n_2, \dots, n_r observations. (On a donc $N = n_1 + n_2 + \dots + n_r$). Notons $y_{s(i)}$ l'observation associée au sujet s dans la classe g_i

Sujets	Facteur G (groupe)						
	g_1	g_2	...	g_i	...	g_r	
1	$y_{1(1)}$	$y_{1(2)}$...	$y_{1(i)}$...	$y_{1(r)}$	
2	$y_{2(1)}$	$y_{2(2)}$...	$y_{2(i)}$...	$y_{2(r)}$	
...	
s	$y_{s(1)}$	$y_{s(2)}$...	$y_{s(i)}$...	$y_{s(r)}$	
...	
...	...	$y_{n_2(2)}$	$y_{n_r(r)}$	
...	$y_{n_1(1)}$	$y_{n_i(i)}$	
Effectif	n_1	n_2	...	n_i	...	n_r	$N = \sum_{i=1}^r n_i$
Total	T_1	T_2	...	T_i	...	T_r	$T = \sum_{i=1}^r T_i$
Moyenne	\bar{y}_1	\bar{y}_2	...	\bar{y}_i	...	\bar{y}_r	$\bar{y} = \frac{1}{N} \sum_{i=1}^r n_i \bar{y}_i$

Les observations de la variable dépendante Y peuvent différer sous l'effet de deux sources de variations :

- La variation à l'intérieur des classes (ou dans les classes) : variation intra
- La variation entre les classes : variation inter.

Le facteur étant *constant* à l'intérieur des classes, seule la variation inter peut être imputée à l'effet du facteur.

La variation intra est une variation *individuelle* correspondant à la variation des observations dans une même classe.

La variation inter est une variation *systématique* correspondant à la variation des classes. Elle exprime l'effet (éventuel) du facteur (VI).

La variation des observations sera mesurée en termes de dispersion autour de la moyenne à l'aide de la variance (ou de la somme des carrés des écarts à la moyenne).

La moyenne générale des observations est donnée par :

$$\bar{y} = \frac{1}{N} \sum_s \sum_i y_{s(i)} \quad \bar{y} = \frac{T}{N} \quad \bar{y} = \frac{1}{N} (n_1 \bar{y}_1 + n_2 \bar{y}_2 + \dots + n_r \bar{y}_r)$$

La variation Totale des observations est mesurée par la Somme des Carrés des écarts entre chacune des N observations et la moyenne générale :

$$SCT = \sum_s \sum_i (y_{s(i)} - \bar{y})^2 = N \times \text{Variance Totale}$$

En notant \bar{y}_i la moyenne des observations de la classe g_i on a : $\bar{y}_i = \frac{1}{n_i} \sum_{s=1}^{n_i} y_{s(i)} \quad \bar{y}_i = \frac{T_i}{n_i}$

La variation à l'Intérieur de la classe g_i est mesurée par la Somme des Carrés des écarts entre chaque observation de la classe et la moyenne de cette classe :

$$\mathbf{SCI}_{\text{Intra}_i} = \sum_{s=1}^{n_i} (y_{s(i)} - \bar{y}_i)^2 = n_i \times \text{Variance}(g_i)$$

La variation à l'intérieur des classes est définie par la somme, pour toutes les classes, des

$$\text{sommes des carrés intra : } \mathbf{SCI}_{\text{Intra}} = \sum_{i=1}^r \mathbf{SCI}_{\text{Intra}_i}$$

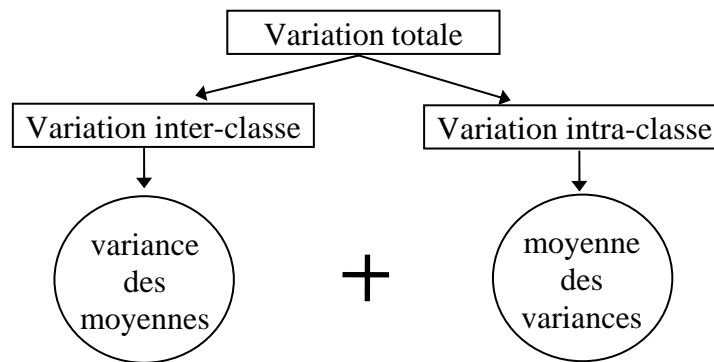
Pour calculer la variation-inter on se place dans le cas de la situation fictive où toutes les observations d'une même classe sont égales à la moyenne \bar{y}_i de cette classe et on calcule la

$$\text{Somme des Carrés Inter : } \mathbf{SCI}_{\text{Inter}} = \sum_{i=1}^r n_i (\bar{y}_i - \bar{y})^2 = N \times \text{Variance}(\bar{y}_i)$$

On démontre la relation fondamentale suivante : $\mathbf{SCT} = \mathbf{SCI}_{\text{Inter}} + \mathbf{SCI}_{\text{Intra}}$

Remarque : En divisant par l'effectif total N on a une relation analogue entre les variances :

$$\text{Variance totale} = \text{Var Inter} + \text{Var Intra}$$



Remarque : On retrouve l'indice de liaison PRE, appelé rapport de corrélation (cf. cours de L1)

en considérant le rapport : $\eta_{Y/G}^2 = \frac{\text{Variation Inter}}{\text{Variation Totale}}$

$\eta_{Y/G}^2$ mesure la liaison entre la variable réponse Y et le facteur G :

- $\eta_{Y/G}^2 = 0$ indique une absence de liaison (la réponse du sujet ne dépend pas du groupe auquel il appartient).
- $\eta_{Y/G}^2 = 1$ indique une liaison parfaite (la réponse du sujet est entièrement déterminée par son appartenance à un groupe).

Ainsi $\eta_{Y/G}^2$ mesure l'intensité de l'effet de G sur la réponse Y .

I-3. Distribution de Fisher-Snedecor

Proposition : Soient $U_1, U_2, \dots, U_m, V_1, V_2, \dots, V_n$ $m+n$ variables aléatoires indépendantes de même loi $\mathcal{N}(0, \sigma^2)$ alors les variables $\frac{1}{\sigma^2} \sum_{i=1}^m U_i^2$ et $\frac{1}{\sigma^2} \sum_{i=1}^n V_i^2$ sont indépendantes et de loi respective χ_m^2 et χ_n^2

Théorème : Soient $U_1, U_2, \dots, U_m, V_1, V_2, \dots, V_n$ $m+n$ variables aléatoires indépendantes de

même loi $\mathcal{N}(0, \sigma^2)$ alors la variable $\frac{\frac{1}{m\sigma^2} \sum_{i=1}^m U_i^2}{\frac{1}{n\sigma^2} \sum_{i=1}^n V_i^2} = \frac{\frac{1}{m} \sum_{i=1}^m U_i^2}{\frac{1}{n} \sum_{i=1}^n V_i^2}$ suit une loi de Fisher-Snedecor

$\mathcal{F}(m, n)$ de (m, n) degrés de liberté.

Remarque : Si \mathcal{F} est une variable aléatoire de loi $\mathcal{F}(m, n)$ alors la variable aléatoire $1/\mathcal{F}$ suit une loi $\mathcal{F}(n, m)$.

Corollaire : Soient U_1, U_2, \dots, U_m m variables aléatoires indépendantes de même loi $\mathcal{N}(\mu_1, \sigma^2)$ et V_1, V_2, \dots, V_n n variables aléatoires indépendantes de même loi $\mathcal{N}(\mu_2, \sigma^2)$ indépendantes des U_i

alors la variable $\frac{\frac{1}{m-1} \sum_{i=1}^m (U_i - \bar{U})^2}{\frac{1}{n-1} \sum_{i=1}^n (V_i - \bar{V})^2}$ suit une loi de Fisher-Snedecor $\mathcal{F}(m-1, n-1)$.

$\bar{U} = \frac{\sum_{i=1}^m U_i}{m}$ et $\bar{V} = \frac{\sum_{i=1}^n V_i}{n}$ Sont les estimateurs respectifs de μ_1 et μ_2

I-2-1 Exemple :

Pour étudier l'influence du facteur « intensité du bruit environnant » sur la capacité d'un sujet à résoudre un problème, l'expérimentateur construit l'expérience suivante : 24 écoliers sont répartis de façon aléatoire dans quatre pièces. Des bruits de la rue ont été enregistrés et sont diffusés dans chaque pièce avec un niveau sonore particulier. Les enfants doivent résoudre une série de problèmes. La variable réponse est la note finale obtenue à la série d'épreuves.

	Niveau sonore				
	1	2	3	4	
	62	56	63	68	
	60	62	67	66	
	63	60	71	71	
	59	61	64	67	
		63	65	68	
		64	66	68	
		63			
		59			
n_i	4	8	6	6	$N = 24$
\bar{y}_i	61	61	66	68	$\bar{y} = 64$
Variance	10/4	48/8	40/6	14/6	$SCI_{Intra} = SCR = 112$
$n_i (\bar{y}_i - \bar{y})^2$	36	72	24	96	$SCI_{Inter} = \sum_{i=1}^r n_i (\bar{y}_i - \bar{y})^2 = 228$

$$SCT = N \times \text{variance totale} = 340 \quad SCI_{Intra} = 228 \quad SCI_{Inter} = 112 \quad \eta_{Y/G}^2 = 0,6706$$

II-1. G est un (seul) facteur à effets fixes

II-1-1 Généralités

G est un facteur à r modalités (groupes) g_1, g_2, \dots, g_r d'effectifs respectifs n_1, n_2, \dots, n_r .
 $N = n_1 + n_2 + \dots + n_r$. Y est la variable dépendante et $y_{s(i)}$ est le score du sujet s dans le groupe g_i
 (c'est l'observation associée au sujet s dans la modalité g_i).

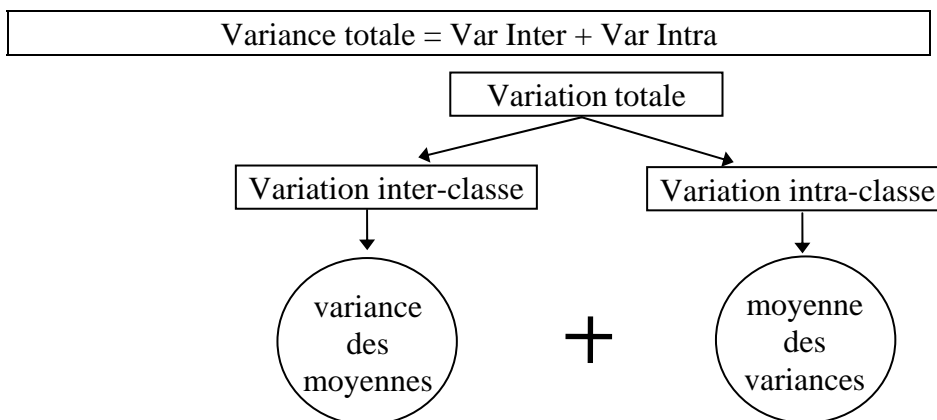
Sujets	Facteur G (groupe)						
	g_1	g_2	...	g_i	...	g_r	
1	$y_{1(1)}$	$y_{1(2)}$...	$y_{1(i)}$...	$y_{1(r)}$	
2	$y_{2(1)}$	$y_{2(2)}$...	$y_{2(i)}$...	$y_{2(r)}$	
...	
s	$y_{s(1)}$	$y_{s(2)}$...	$y_{s(i)}$...	$y_{s(r)}$	
...	
...	...	$y_{n_2(2)}$	$y_{n_r(r)}$	
...	$y_{n_1(1)}$	$y_{n_i(i)}$	
Effectif	n_1	n_2	...	n_i	...	n_r	$N = \sum_{i=1}^r n_i$
Moyenne	\bar{y}_1	\bar{y}_2	...	\bar{y}_i	...	\bar{y}_r	$\bar{y} = \frac{1}{N} \sum_{i=1}^r n_i \bar{y}_i$
Variance	s'_1	s'_2	...	s'_i	...	s'_r	s'

On veut étudier l'effet du facteur G sur la variable réponse Y.

Pour cela on testera l'hypothèse nulle : H_0 : le facteur n'a pas d'effet sur Y

contre l'hypothèse alternative : H_1 : le facteur a de l'effet sur Y.

II-1-2 Décomposition de la variance (Rappel)



II-1-3 Modèle statistique (Modèle du score, modèle linéaire)

On veut étudier l'effet d'un facteur G à r modalités sur une variable réponse Y . On dispose pour chaque modalité i du facteur G de n_i observations. On note $y_{s(i)}$ l'observation du sujet s dans la modalité i .

$y_{s(i)}$ est la réalisation de la variable aléatoire $Y_{s(i)}$ décrite par le modèle suivant : $Y_{s(i)} = \mu + \alpha_i + \varepsilon_{s(i)}$ où :

- μ s'interprète comme un niveau général.
- α_i mesure l'effet de la $i^{\text{ème}}$ modalité du facteur G .
- les variables aléatoires $\varepsilon_{s(i)}$ sont indépendantes et de même loi $\mathcal{N}(0, \sigma^2)$
- Pour tout $i=1, 2, \dots, r$. Les variances des r groupes sont homogènes (homoscédasticité)
- On rajoute une contrainte d'identifiabilité: $\sum_{i=1}^r n_i \alpha_i = 0$ (en moyenne les effets sont nuls)

On peut écrire l'hypothèse nulle et l'hypothèse alternative sous la forme suivante :

hypothèse nulle : $\mathbf{H}_0 : \alpha_i = 0$ pour tout $i = 1, 2, \dots, r$

hypothèse alternative : $\mathbf{H}_1 : \text{l'un au moins des } \alpha_i \text{ est non nul}$

On montre alors les résultats suivants :

Théorème :

- $E(\text{SCR}) = (N - r)\sigma^2$
- $E(\text{SCG}) = (r - 1)\sigma^2 + \sum_i n_i \alpha_i^2$

Remarque : Il est clair que les valeurs des variations (totale, inter et intra) dépendent des effectifs des modalités du facteur G . Aussi pour apprécier plus justement les grandeurs relatives à ces variations (en particulier pour comparer la variation due au facteur à la variation résiduelle), on calcule les carrés moyens des écarts, en divisant chaque variation par son nombre de degrés de

liberté (ddl) : $\text{MC} = \frac{\text{SC}}{\text{ddl}}$

- carrés moyens (total) : MCT

La variation totale concerne N scores (observations) $y_{s(i)}$ liés par une relation $\bar{y} = \frac{1}{N} \sum_{i,j} y_{s(i)}$ donc

présente $N - 1$ ddl et on a : $\text{MCT} = \frac{\text{SCT}}{N - 1}$

- carrés moyens inter : MCG

La variation inter(-groupes) concerne r groupes liés par une relation $\bar{y} = \frac{1}{N} \sum_i n_i \bar{y}_i$ donc présente

$r - 1$ ddl et on a : $\text{MCG} = \frac{\text{SCG}}{r - 1}$

- carrés moyens intra (résiduels) : MCR

La variation intra (résiduelle) concerne N scores (observations) liés par r relations $\bar{y}_i = \frac{1}{n_i} \sum_j y_{ji}$

donc présente $N - r$ ddl et on a : $\text{MCR} = \frac{\text{SCR}}{N - r}$

Remarque : Tout comme les variations, les degrés de liberté sont additifs

$$\text{ddl total} = \text{ddl inter} + \text{ddl intra} \quad N - 1 = r - 1 + N - r$$

Mais les carrés moyens ne sont pas additifs. En général $MCT \neq MCG + MCR$.

Le théorème précédent peut se réécrire : $E(MCR) = \sigma^2$ $E(MCG) = \sigma^2 + \sum_{i=1}^r \frac{n_i \alpha_i^2}{r-1}$

Par conséquent : **MCR** est une estimation sans biais de la variance σ^2 et **MCG** est une estimation de σ^2 augmentée d'un terme positif fonction des effets de groupe ($\sum_{i=1}^r \frac{n_i \alpha_i^2}{r-1}$).

Sous l'hypothèse nulle ($\alpha_i = 0$ pour tout $i = 1, 2, \dots, r$) **MCG** est aussi une estimation de σ^2 .

Théorème :

Dans le cadre du modèle statistique, la statistique $\frac{SCR}{\sigma^2}$ suit une loi du χ^2 à $N - r$ ddl et sous l'hypothèse nulle la statistique $\frac{SCG}{\sigma^2}$ suit une loi du χ^2 à $r - 1$ ddl et ces deux statistiques sont indépendantes.

Corollaire :

Sous l'hypothèse nulle la statistique $\frac{SCG / r - 1}{SCR / N - r} = \frac{MCG}{MCR} = F_G$ suit une loi de Fischer à $r - 1$ et $N - r$ ddl, notée $\mathcal{F}(r - 1, N - r)$.

Conclusion :

Le test d'hypothèse est alors défini au seuil de signification α par la règle de décision suivante :
Si $F_{obs} \geq \lambda_\alpha$ alors on rejette l'hypothèse nulle. λ_α étant donné par l'équation : $\alpha = P(F \geq \lambda_\alpha)$.

II-1-4 Tableau d'analyse de la variance

Les résultats sont souvent présentés sous forme d'un tableau :

Source de variation	Somme des carrés des écarts observés SC_{obs}	ddl	Carrés moyens observés MC_{obs}	F
G, Facteur, Inter-groupes, ...	SCG	$r - 1$	MCG	$F_{obs} = \frac{MCG}{MCR}$
S<G>, Résiduelle, Intra	SCR	$N - r$	MCR	$F_{lu} =$ $\alpha =$
Totale	SCT	$N - 1$		

II-1-5 Exemple :

Pour étudier l'influence du facteur « intensité du bruit environnant » sur la capacité d'un sujet à résoudre un problème, l'expérimentateur construit l'expérience suivante : 24 écoliers sont répartis de façon aléatoire dans quatre pièces. Des bruits de la rue ont été enregistrés et sont diffusés dans chaque pièce avec un niveau sonore particulier. Les enfants doivent résoudre une série de problèmes. La variable réponse est la note finale obtenue à la série d'épreuves.

	Niveau sonore				
	1	2	3	4	
	62	56	63	68	
	60	62	67	66	
	63	60	71	71	
	59	61	64	67	
		63	65	68	
		64	66	68	
		63			
		59			
n_i	4	8	6	6	$N = 24$
\bar{y}_i	61	61	66	68	$\bar{y} = 64$
Variance	10/4	48/8	40/6	14/6	SCIntra= SCR = 112
$n_i (\bar{y}_i - \bar{y})^2$	36	72	24	96	SCInter= $\sum_{i=1}^r n_i (\bar{y}_i - \bar{y})^2 = 228$

$$\text{SCT} = N \times \text{variance totale} = 340 \quad \text{MCO} = \frac{\text{SCO}}{r-1} = \frac{108,9}{2} = 54,45$$

$$\text{MCR} = \frac{\text{SCR}}{(n-1)(r-1)} = \frac{13,8}{6 \times 2} = 1,15.$$

On peut présenter les résultats dans le tableau d'analyse de variance :

Sources de variations	SC_{obs}	ddl	MC_{obs}	Statistique de test = Q
Facteur = Inter = G	228	$r-1 = 3$	76	$F_{obs} = \frac{MCG_{obs}}{MCR_{obs}} = \frac{76}{5,6} = 13,5714^{**}$
$S(G) = \text{Intra}$	112	$N-r = 20$	5,6	-
Total	340	$N-1 = 23$	-	-

En prenant un niveau de signification $\alpha=0,05=5\%$, on peut lire la valeur de λ_α dans la table de Fischer à $r-1 = 3$ et $N-r = 20$ degrés de liberté : $\lambda_\alpha = F_{(3; 20; 0,95)} = 3,10$. Donc, si F_{obs} , la valeur observée de la statistique de test, est supérieure ou égale à 3,10 on rejette H_0 et dans le cas contraire, on conserve l'hypothèse H_0 .

Conclusion : Comme $F_{obs}=13,5714 > \lambda_\alpha=3,10$, on rejette H_0 pour accepter H_1 ; C'est-à-dire qu'il y a effectivement un effet du bruit environnant sur la capacité de résolution des problèmes. On rejette H_0 même avec un $\alpha=0,01=1\%$.

II-2. G est un (seul) facteur à effets aléatoires

II-2-1 Modèle statistique

On choisit un échantillon de r modalités g_1, g_2, \dots, g_r du facteur G . On dispose pour chaque modalité échantillonnée de n_i observations de la variable dépendante ($N=n_1+n_2+\dots+n_r$). On note $y_{s(i)}$ l'observation du sujet s dans le groupe g_i . Chaque observation $y_{s(i)}$ est la réalisation d'une variable aléatoire $Y_{s(i)}$ décrite par le modèle : $Y_{s(i)} = \mu + \Gamma_i + \varepsilon_{s(i)}$
où :

μ est une constante mesurant le niveau général de la réponse

Γ_i est une variable aléatoire qui mesure l'effet aléatoire de G .

$\varepsilon_{s(i)}$ est une variable aléatoire représentant le résidu.

On suppose réalisées les trois hypothèses suivantes :

- les $\varepsilon_{s(i)}$ sont des variables aléatoires indépendantes et de même loi $\mathcal{N}(0, \sigma^2)$

- les Γ_i ($i=1, 2, \dots, r$) sont des variables aléatoires indépendantes de même loi $\mathcal{N}(0, \sigma_G^2)$
- les Γ_i sont indépendantes des $\varepsilon_{s(i)}$.

Le test d'hypothèse s'écrira alors :

$$\mathbf{H}_0 : G \text{ n'a pas d'effet} \Leftrightarrow \sigma_G^2 = 0$$

$$\mathbf{H}_1 : G \text{ a un effet} \Leftrightarrow \sigma_G^2 > 0$$

On montre alors les résultats suivants :

Théorème :

- $E(\mathbf{MCR}) = \sigma^2$ ou encore $E(\mathbf{SCR}) = (N - r)\sigma^2$
- $E(\mathbf{MCG}) = \sigma^2 + K\sigma_G^2$ où K est une constante.

II-2-2 Test statistique

On se ramène donc au cas précédent et on utilisera la même statistique $F = \frac{\mathbf{MCG}}{\mathbf{MCR}}$ pour réaliser le test.

II-2-3 Exemple :

On veut vérifier que l'intensité de traitement perceptif d'un visage dépend du visage examiné (certains visages retiennent plus l'attention que d'autres). Pour mettre à l'épreuve cette hypothèse de recherche, on construit l'expérience suivante :

40 sujets sont choisis au hasard et répartis de façon aléatoire dans 5 groupes de 8 sujets chacun. Chaque groupe examine un visage choisi par l'expérimentateur au hasard dans l'ensemble de visages disponibles.

L'expérimentateur mesure l'intensité du traitement perceptif en observant la dilatation de la pupille lors de l'examen du visage. Il obtient les résultats suivants :

Sujets	Groupe des visages					
	g_1	g_2	g_3	g_4	g_5	
1	58	60	63	64	57	
2	51	61	55	64	59	
3	57	66	57	65	65	
4	59	65	60	61	63	
5	56	59	61	66	62	
6	54	59	62	59	64	
7	53	64	58	67	60	
8	52	63	56	60	63	
n_i	8	8	8	8	8	$N=40$
\bar{y}_i	55	62,125	59	63,25	61,625	$\bar{y} = 60,2$
s'^2_i	7,5	6,6094	7,5	7,4375	6,4844	$s'^2=15,81$

$$\mathbf{SCT} = N s'^2 = 40 \times 15,81 = 632,4$$

$$\mathbf{SCG} = N s'^2_{\bar{y}_i} = N \sum_{i=1}^5 (\bar{y}_i - \bar{y})^2 = 40 \times 8,7038 = 348,15$$

$$\mathbf{SCR} = N \sum_{i=1}^5 s'^2_i = 8 \times [7,5 + 6,6094 + 7,5 + 7,4375 + 6,4844] = 8 \times 35,5313 = 284,25$$

On vérifie que : $\mathbf{SCR} = \mathbf{SCT} - \mathbf{SCG} = 632,4 - 348,15 = 284,25$.

On peut présenter les résultats dans le tableau d'analyse de variance :

Sources de variations	SC _{obs}	ddl	MC _{obs}	Statistique de test = Q
Facteur = Inter = G	348,15	$r-1 = 4$	87,0375	$F_{\text{obs}} = \frac{\text{MCG}_{\text{obs}}}{\text{MCR}_{\text{obs}}} = \frac{87,0375}{8,1214} = 10,717^{**}$
$S(G) = \text{Intra}$	284,25	$N-r = 35$	8,1214	-
Total	632,4	$N-1 = 39$	-	-

En prenant un niveau de signification $\alpha=0,05=5\%$, on peut lire la valeur de λ_α dans la table de Fischer à $r-1 = 4$ et $N-r = 35$ degrés de liberté : $\lambda_\alpha = F_{(4 ; 35 ; 0,95)} = 2,64$

Conclusion : Comme $F_{\text{obs}}=10,717 > \lambda_\alpha=2,64$, on rejette H_0 pour accepter H_1 ; C'est-à-dire qu'il y a effectivement un effet du facteur visage. ($\sigma_G^2 > 0$)

Remarque : En prenant un niveau de signification $\alpha=0,01=1\%$, on peut lire la valeur de λ_α dans la table de Fischer à $r-1 = 4$ et $N-r = 35$ degrés de liberté : $\lambda_\alpha = F_{(4 ; 35 ; 0,99)} = 3,91$ et on tire la même conclusion.

III-1. Plan $S<A \times B>$ avec A et B deux facteurs à effets fixes

III-1-1 Généralités

Un tel plan quasi-complet défini par le croisement de deux facteurs A et B , est aussi appelé plan factoriel complet d'ordre 2. Chaque condition est répliquée sur des sujets différents. Le facteur sujet est emboîté dans le croisement $A \times B$. L'analyse diffère selon que l'emboîtement est équilibré ou non. Dans toute la suite on ne considère que le cas équilibré.

On se propose d'étudier l'effet de la présence simultanée d'un facteur A à r modalités et d'un facteur B à c modalités sur une variable réponse Y . Le plan étant équilibré, on dispose pour chaque croisement des modalités i du facteur A et j du facteur B de n observations. On note $y_{s(i,j)}$ l'observation du sujet s dans le croisement (i,j) .

		Facteur B					
		b_1	b_2	...	b_j	...	B_c
A	a_1						
	a_2						
	...						
	a_i				$y_{1(i,j), y_{2(i,j), \dots, y_{n(i,j)}}$		
	...						
	a_r						

Remarque : dans le cas où $n=1$, l'interaction des facteurs A et B est confondue avec le facteur sujet, on ne peut donc séparer les effets d'interaction des effets résiduels et on ne peut tester leur existence. Un tel plan n'est justifié que lorsque l'on peut modifier a priori la forme de l'interaction entre les facteurs. Dans toute la suite nous développons le cas $n > 1$.

III-1-2 Analyse descriptive des effets (principaux et interaction) :

Considérons, pour chaque croisement (i,j) de la modalité i de A et de la modalité j de B la moyenne obtenue :

		Facteur B						moyennes
		b_1	b_2	...	b_j	...	B_c	
A	a_1	\bar{y}_{11}	\bar{y}_{12}	...	\bar{y}_{1j}	...	\bar{y}_{1c}	\bar{y}_{10}
	a_2	\bar{y}_{21}	\bar{y}_{22}	...	\bar{y}_{2j}	...	\bar{y}_{2c}	\bar{y}_{20}

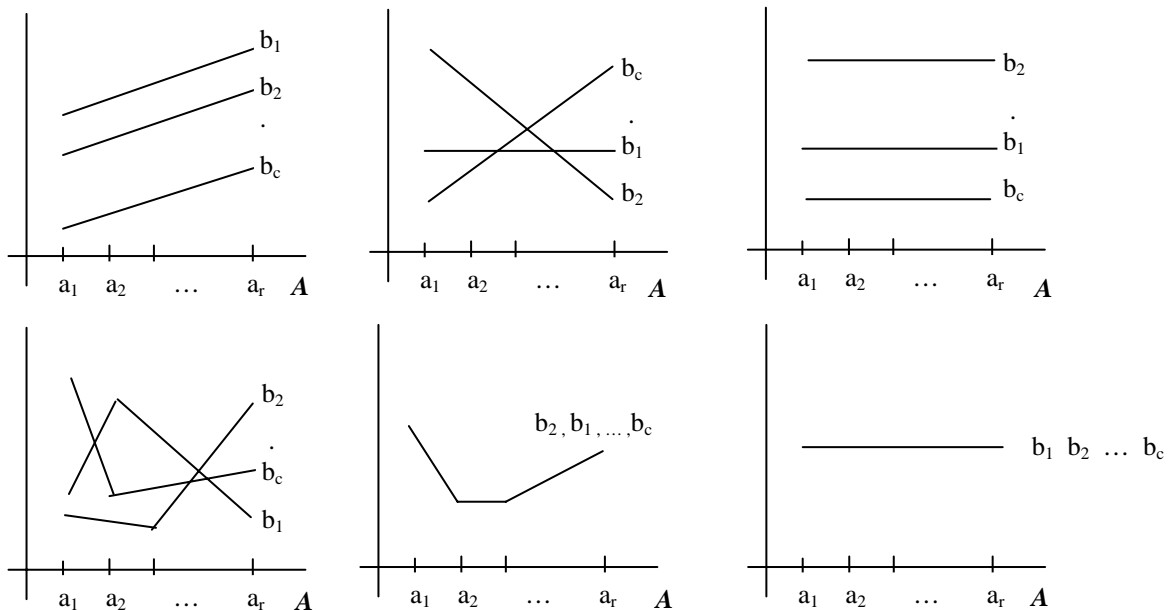
	a_i	\bar{y}_{i1}	\bar{y}_{i2}	...	\bar{y}_{ij}	...	\bar{y}_{ic}	\bar{y}_{i0}

	a_r	\bar{y}_{r1}	\bar{y}_{r2}	...	\bar{y}_{rj}	...	\bar{y}_{rc}	\bar{y}_{r0}
moyennes		\bar{y}_{01}	\bar{y}_{02}	...	\bar{y}_{0j}	...	\bar{y}_{0c}	$\bar{y}_{00} = \bar{y}$

La colonne marginale des moyennes par modalité de A (indépendamment de B) « traduit » l'effet principal du facteur A . De même la ligne marginale des moyennes par modalité de B (indépendamment de A) « traduit » l'effet principal du facteur B . Les différentes moyennes correspondant au croisement des modalités « traduisent » l'interaction des deux facteurs A et B .

III-1-3 Représentation géométrique des différents effets :

Les différentes combinaisons des effets principaux et d'interaction peuvent être illustrés par des graphiques dans lesquels on représente les moyennes en fonction des modalités de l'un des facteurs. Il y a deux graphiques possibles selon le facteur porté en abscisse.



Le parallélisme des lignes traduit le fait que l'effet du facteur A (ou B) sur la variable réponse Y est le même quelque soit la valeur du facteur B (ou A). Les effets des facteurs A et B sont additifs et il n'y a pas d'effet de l'interaction entre les deux facteurs A et B sur Y.

Le non parallélisme des lignes traduit le fait que l'effet du facteur A (ou du facteur B) n'est pas le même en fonction de la modalité du facteur B (ou du facteur A). Les effets des facteurs A et B sur la variable Y ne sont pas additifs. Aux effets simples des facteurs s'ajoutent l'effet de l'interaction entre les deux facteurs.

Le parallélisme avec l'axe des abscisses traduit l'absence d'effet du facteur A (ou B). De plus si les différentes lignes correspondant aux modalités de B (ou A) sont confondues c'est l'absence totale d'effets.

III-1-4 Le modèle :

Rappelons que l'on veut étudier l'effet de la présence simultanée d'un facteur A à r modalités et d'un facteur B à c modalités sur une variable réponse Y. Le plan étant équilibré, on dispose pour chaque modalité i du facteur A et pour chaque modalité j du facteur B de n observations (n > 1). On note $y_{s(i,j)}$ l'observation du sujet s dans le croisement (i,j). $y_{s(i,j)}$ est la réalisation de la variable aléatoire $Y_{s(i,j)}$ décrite par le modèle suivant :

$$Y_{s(i,j)} = \mu_{ij} + \varepsilon_{s(i,j)}$$

où les variables aléatoires $\varepsilon_{s(i,j)}$ sont indépendantes et de même loi $\mathcal{N}(0, \sigma^2)$ et

$$\mu_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} \quad \text{où}$$

- μ mesure la moyenne de la population
- les α_i mesurent les effets principaux du facteur A
- les β_j mesurent les effets principaux du facteur B
- les $(\alpha\beta)_{ij}$ mesurent les effets d'interaction.

On rajoute (comme hypothèses) les contraintes d'identifiabilité suivantes :

$$\sum_i \alpha_i = 0 \quad \sum_j \beta_j = 0 \quad \sum_j (\alpha\beta)_{ij} = 0 \quad \sum_i (\alpha\beta)_{ij} = 0$$

III-1-5 Décomposition de la variation :

A partir du tableau des données on calcule les Sommes des Carrés associées aux différentes sources de variation. Le croisement des facteurs A et B permet de décomposer la variation totale en somme de deux variations :

$$\mathbf{SCT}_{\text{obs}} = \mathbf{SCI}_{\text{intercase}} + \mathbf{SCI}_{\text{intracase}} = \mathbf{SC}(A \times B)_{\text{obs}} + \mathbf{SCR}_{\text{obs}}$$

avec :

$$\mathbf{SCT}_{\text{obs}} = \sum_i \sum_j \sum_s (y_{s(i,i)} - \bar{y})^2 \quad \mathbf{SC}(A \times B)_{\text{obs}} = n \sum_i \sum_j (\bar{y}_{ij} - \bar{y})^2 \quad \mathbf{SCR}_{\text{obs}} = \sum_i \sum_j \sum_s (y_{s(i,i)} - \bar{y}_{ij})^2$$

D'autre part le facteur élémentaire A (resp B) définit une partition de l'ensemble des observations en classes à laquelle correspond une somme de carrés observée mesurant la variation associée à l'effet principal de A (resp de B) :

$$\mathbf{SCA}_{\text{obs}} = nc \sum_i (\bar{y}_{i0} - \bar{y})^2 \quad \mathbf{SCB}_{\text{obs}} = nr \sum_j (\bar{y}_{0j} - \bar{y})^2$$

Définition : On appelle somme des carrés associée à l'interaction la quantité :

$$\mathbf{SCAB}_{\text{obs}} = \mathbf{SC}(A \times B)_{\text{obs}} - \mathbf{SCA}_{\text{obs}} - \mathbf{SCB}_{\text{obs}}$$

$$\text{On montre que : } \mathbf{SCAB}_{\text{obs}} = \sum_i \sum_j n(\bar{y}_{ij} - \bar{y}_{i0} - \bar{y}_{0j} + \bar{y})^2$$

Dans le cadre du modèle statistique ces sommes de carrés sont des réalisations de variables aléatoires dont on calcule les espérances et plus généralement les distributions des probabilités.

On montre le théorème fondamental suivant :

Théorème : Sous les hypothèses du modèle, les statistiques **SCA**, **SCB**, **SCAB** et **SCR** sont indépendantes et d'espérances respectives :

$$E(\mathbf{SCA}) = (r-1)\sigma^2 + \sum_i nc\alpha_i^2$$

$$E(\mathbf{SCB}) = (c-1)\sigma^2 + \sum_j nr\beta_j^2$$

$$E(\mathbf{SCAB}) = (r-1)(c-1)\sigma^2 + \sum_i \sum_j n(\alpha\beta)_{ij}^2$$

$$E(\mathbf{SCR}) = (N - rc)\sigma^2 = rc(n-1)\sigma^2$$

On ramène toutes les sommes de carrés à des moyennes de carrés en divisant par les degrés de liberté correspondant. $\mathbf{MCA} = \frac{\mathbf{SCA}}{r-1}$ $\mathbf{MCB} = \frac{\mathbf{SCB}}{c-1}$ $\mathbf{MCAB} = \frac{\mathbf{SCAB}}{(r-1)(c-1)}$ $\mathbf{MCR} = \frac{\mathbf{SCR}}{rc(n-1)}$

Ceci permet de réécrire le théorème précédent pour les moyennes des carrés :

Théorème : Sous les hypothèses du modèle, les statistiques **MCA**, **MCB**, **MCAB** et **MCR** sont indépendantes et d'espérances respectives :

$$E(\mathbf{MCA}) = \sigma^2 + \frac{nc}{(r-1)} \sum_i \alpha_i^2$$

$$E(\mathbf{MCB}) = \sigma^2 + \frac{nr}{(c-1)} \sum_j \beta_j^2$$

$$E(\mathbf{MCAB}) = \sigma^2 + \frac{n}{(r-1)(c-1)} \sum_i \sum_j (\alpha\beta)_{ij}^2$$

$$E(\mathbf{MCR}) = \sigma^2$$

III-1-6 Tests et décisions statistiques :

Le théorème précédent nous permet de tester l'existence des différents effets des facteurs. On peut construire des tests indépendants sur chacune des sources de variations :

Test 1 :

hypothèse nulle \mathbf{H}_{0A} : pas d'effet principal du facteur A $\Leftrightarrow (\forall i) \alpha_i = 0$ contre
l'hypothèse alternative : \mathbf{H}_{1A} : il existe un effet principal du facteur A $\Leftrightarrow (\exists i) \alpha_i \neq 0$

Test 2 :

hypothèse nulle \mathbf{H}_{0B} : pas d'effet principal du facteur B $\Leftrightarrow (\forall j) \beta_j = 0$ contre
l'hypothèse alternative : \mathbf{H}_{1B} : il existe un effet principal du facteur B $\Leftrightarrow (\exists j) \beta_j \neq 0$

Test 3 :

hypothèse nulle \mathbf{H}_{0AB} : pas d'effet d'interaction $\Leftrightarrow (\forall i,j) (\alpha\beta)_{ij} = 0$ contre
l'hypothèse alternative : \mathbf{H}_{1AB} : il existe une interaction $\Leftrightarrow (\exists i,j) (\alpha\beta)_{ij} \neq 0$

La construction de ces tests résulte des trois corollaires du théorème suivant :

Théorème :

Dans le cadre du modèle statistique, la statistique \mathbf{SCR}/σ^2 suit une loi de χ^2 à $rc(n-1)$ degrés de liberté, sous l'hypothèse nulle \mathbf{H}_{0A} la statistique \mathbf{SCA}/σ^2 suit une loi de χ^2 à $r-1$ degrés de liberté, sous l'hypothèse nulle \mathbf{H}_{0B} la statistique \mathbf{SCB}/σ^2 suit une loi de χ^2 à $c-1$ degrés de liberté, sous l'hypothèse nulle \mathbf{H}_{0AB} la statistique \mathbf{SCAB}/σ^2 suit une loi de χ^2 à $(r-1)(c-1)$ degrés de liberté et ces statistiques sont indépendantes.

Corollaire : sous l'hypothèse \mathbf{H}_{0A} , la statistique $\mathbf{F}_A = \frac{\mathbf{MCA}}{\mathbf{MCR}}$ suit une loi de Fischer à $r-1$ et $rc(n-1)$ degrés de liberté.

Le test1 est alors défini au seuil de signification α par la règle de décision suivante :

si $\mathbf{F}_{A \text{ obs}} \geq \lambda_\alpha$ alors on rejette l'hypothèse nulle

où λ_α est donné par l'équation : $\alpha = P(\mathbf{F}_A \geq \lambda_\alpha)$.

Corollaire : sous l'hypothèse \mathbf{H}_{0B} , la statistique $\mathbf{F}_B = \frac{\mathbf{MCB}}{\mathbf{MCR}}$ suit une loi de Fischer à $c-1$ et $rc(n-1)$ degrés de liberté.

Le test2 est alors défini au seuil de signification α par la règle de décision suivante :

si $\mathbf{F}_{B \text{ obs}} \geq \lambda_\alpha$ alors on rejette l'hypothèse nulle

où λ_α est donné par l'équation : $\alpha = P(\mathbf{F}_B \geq \lambda_\alpha)$.

Corollaire : sous l'hypothèse \mathbf{H}_{0AB} , la statistique $\mathbf{F}_{AB} = \frac{\mathbf{MCAB}}{\mathbf{MCR}}$ suit une loi de Fischer à $(r-1)(c-1)$ et $rc(n-1)$ degrés de liberté.

Le test3 est alors défini au seuil de signification α par la règle de décision suivante :

si $\mathbf{F}_{AB \text{ obs}} \geq \lambda_\alpha$ alors on rejette l'hypothèse nulle

où λ_α est donné par l'équation : $\alpha = P(\mathbf{F}_{AB} \geq \lambda_\alpha)$.

On présente l'étude dans le :

tableau d'analyse de la variance :

Source	SC_{obs}	ddl	MC_{obs}	F_{obs}
A	SCA_{obs}	$r - 1$	MCA_{obs}	$F_{A\ obs}$
B	SCB_{obs}	$c - 1$	MCB_{obs}	$F_{B\ obs}$
AB	$SCAB_{obs}$	$(r - 1)(c - 1)$	$MCAB_{obs}$	$F_{AB\ obs}$
R	SCR_{obs}	$rc(n - 1)$	MCR_{obs}	
Total	SCT_{obs}	$rcn - 1$		

Pratique des calculs :

Calcul de SCA :

$$SCA_{obs} = nc \sum_i (\bar{y}_{i0} - \bar{y})^2 = ncr \times \text{variance des moyennes de A}$$

Calcul de SCB :

$$SCB_{obs} = nr \sum_{j=1}^c (\bar{y}_{0j} - \bar{y})^2 = ncr \times \text{variance des moyennes de B}$$

Calcul de SC(A×B) :

$$SC(A \times B)_{obs} = n \sum_{i=1}^r \sum_{j=1}^c (\bar{y}_{ij} - \bar{y})^2 = ncr \times \text{variance des moyennes des carrés croisés}$$

Calcul de SC(AB) :

$$SC(AB)_{obs} = SC(A \times B)_{obs} - SCA_{obs} - SCB_{obs}$$

Calcul de SCT :

$$SCT_{obs} = ncr \times \text{variance de toutes les données}$$

Calcul de SCR :

$$SCR_{obs} = SCT_{obs} - SC(A \times B)_{obs} = SCT_{obs} - SC(AB)_{obs} - SCA_{obs} - SCB_{obs}$$

III-1-7 Exemple :

Un expérimentateur veut mettre à l'épreuve l'hypothèse que le niveau d'excitation physiologique d'une personne frustrée dépend du type d'agression autorisée. Il construit alors l'expérience suivante : 30 sujets choisis au hasard sont répartis de façon aléatoire en 6 groupes de 5 sujets, correspondant aux 6 conditions expérimentales définies par le croisement de deux facteurs

A : Facteur frustration :

2 modalités : A₁ : frustré A₂ : non frustré

B : Facteur agression :

3 modalités : B₁ : non agressif B₂ : moyennement agressif
B₃ : très agressif

Les résultats sont donnés par le tableau ci-contre.

	A ₁					A ₂				
B ₁	1	6	4	3	0	8	8	10	8	11
B ₂	3	5	2	4	4	6	8	4	4	3
B ₃	6	7	4	5	3	5	7	3	-1	1

Calcul de SCA :

$$SCA_{obs} = nc \sum_i (\bar{y}_{i0} - \bar{y})^2 = ncr \times \text{variance}$$

des moyennes de A

$$SCA_{obs} = 30 \times 1,0 = 30,0$$

Calcul de SCB :

$$SCB_{obs} = nr \sum_j (\bar{y}_{00j} - \bar{y})^2 = ncr \times \text{variance}$$

des moyennes de B $SCB_{obs} = 30 \times 0,6067 = 18,2$

Calcul de SC(A×B) :

$$SC(A \times B)_{obs} = n \sum_{i=1}^r \sum_j (\bar{y}_{ij} - \bar{y})^2 = ncr \times \text{variance des moyennes des carrés croisés}$$

$$SC(A \times B)_{obs} = 30 \times 4,1867 = 125,6$$

Calcul de SC(AB) :

$$SC(AB)_{obs} = SC(A \times B)_{obs} - SCA_{obs} - SCB_{obs} = 125,6 - 30,0 - 18,2 = 77,4$$

Calcul de SCT :

$$SCT_{obs} = ncr \times \text{variance de toutes les données} \quad SCT_{obs} = 30 \times 7,16 = 214,8$$

Calcul de SCR :

$$SCR_{obs} = SCT_{obs} - SC(A \times B)_{obs} = 214,8 - 125,6 = 89,2$$

On présente l'étude dans le :

tableau d'analyse de la variance :

Source	SC _{obs}	ddl	MC _{obs}	F _{obs}
A	SCA _{obs} = 30,0	r - 1 = 1	MCA _{obs} = 30,0	F _{A obs} = 8,0717
B	SCB _{obs} = 18,2	c - 1 = 2	MCB _{obs} = 9,1	F _{B obs} = 2,4484
AB	SCAB _{obs} = 77,4	(r - 1)(c - 1) = 2	MCAB _{obs} = 38,7	F _{AB obs} = 10,4126
R	SCR _{obs} = 89,2	rc(n - 1) = 24	MCR _{obs} = 3,7167	
Total	SCT _{obs} = 214,8	rcn - 1 = 29		

IV-1. O est un (seul) facteur à effets fixes

IV-1-1 Généralités

O est un facteur à r modalités o_1, o_2, \dots, o_r (Occasions). S est le facteur sujet (à effets aléatoires) dont les modalités sont les n sujets s_1, s_2, \dots, s_n . Pour chaque modalité du facteur O on « utilise » les n sujets. On a donc des mesures y_{si} répétées pour chaque sujet s dans l'occasion (modalité) O_i . En psychologie et en méthodologie on parle de plan intra, en statistique on parle de plan à mesures répétées.

On dispose donc de $N=n \times r$ mesures y_{si} que l'on présente sous la forme d'un tableau :

		Facteur O (Occasions)				
Sujets	o_1	o_2	...	o_i	...	o_r
s_1	y_{11}	y_{12}	...	y_{1i}	...	y_{1r}
s_2	y_{21}	y_{22}	...	y_{2i}	...	y_{2r}
...
s_s	y_{s1}	y_{s2}	...	y_{si}	...	y_{sr}
...
s_n	y_{n1}	y_{n2}		y_{ni}		y_{nr}

Remarque : Un premier avantage d'un tel plan par rapport au plan $S < O >$ est une « économie » du nombre de sujets. D'autre part avec un tel plan on peut diminuer l'erreur expérimentale (résidu). En effet le contrôle du facteur sujet permet de séparer son effet des effets résiduels. Mais le résidu sera confondu avec l'interaction entre S et O .

IV-1-2 Modèle univarié :

Les données peuvent être d'une part, considérées comme les $N=n \times r$ observations de la variable réponse Y dans les $N=n \times r$ conditions expérimentales décrites par le croisement du facteur sujet et du facteur occasion. Elles peuvent d'autre part, être regardées comme les n observations d'un vecteur de r variables réponses. Ces deux points de vue conduisent à l'élaboration de deux modèles, le modèle mixte univarié dans le premier cas et le modèle multivarié dans le second (voir condition de validation IV-1-5).

Nous étudierons le modèle mixte univarié dans lequel on considère que les modalités du facteur sujet ont été obtenues par échantillonnage, le facteur sujet est donc un facteur aléatoire. Le facteur occasion est lui un facteur à effets fixes.

Chaque donnée y_{si} correspond alors à l'observation d'une variable aléatoire réelle Y_{si} décrite par le modèle suivant : $Y_{si} = \mu_i + \pi_s + \varepsilon_{si}$

Où :

- les μ_i sont des constantes mesurant les effets fixes des modalités i ($i=1, 2, \dots, r$) du facteur O
- les π_s sont des variables aléatoires indépendantes de loi $\mathcal{N}(0, \sigma^2_{\pi})$ mesurant les effets aléatoires des modalités s ($s=1, 2, \dots, n$) du facteur S
- les résidus ε_{si} sont des variables aléatoires indépendantes de loi $\mathcal{N}(0, \sigma^2)$. On suppose en plus que les π_s et les ε_{si} sont indépendantes.

On peut réécrire l'effet de la modalité i sous la forme : $\mu_i = \mu + \alpha_i$ on a donc : $Y_{si} = \mu + \alpha_i + \pi_s + \varepsilon_{si}$

Où :

- μ constante s'interprétant comme un niveau général de réponse

- les α_i sont des constantes mesurant les effets fixes des modalités i et vérifiant :

$$\sum_i \alpha_i = 0 \text{ (contrainte d'identifiabilité)}$$

Conséquence : D'après ce modèle les Y_{si} sont des variables aléatoires de loi $\mathcal{N}(\mu + \alpha_i, \sigma_\pi^2 + \sigma^2)$.

L'effet du facteur fixe se traduit sur la moyenne de la variable réponse tandis que l'effet du facteur aléatoire se traduit sur la variance de la variable réponse.

Pour ce modèle, l'interaction entre le facteur sujet et le facteur occasion est confondue avec le résidu car pour chaque modalité de l'un et de l'autre on ne dispose que d'une seule observation.

IV-1-3 Décomposition de la variation :

On décompose la variation totale en variation inter-sujets (due au facteur sujet) et variation intra-sujets (due à l'effet du facteur O). $\mathbf{SCT}_{\text{obs}} = \mathbf{SC}_{\text{inter-sujets}} + \mathbf{SC}_{\text{intra-sujets}}$

La variation inter-sujets est due au facteur sujet et la variation intra-sujets est due d'une part à l'effet du facteur O et d'autre part à l'effet des autres facteurs non contrôlés

$$\mathbf{SC}_{\text{intra-sujets}} = \mathbf{SCO}_{\text{obs}} + \mathbf{SCR}_{\text{obs}}$$

Conclusion : Dans le cas du plan complet $S \times O$, la variation totale se décompose de façon additive en : $\mathbf{SCT}_{\text{obs}} = \mathbf{SCS}_{\text{obs}} + \mathbf{SCO}_{\text{obs}} + \mathbf{SCR}_{\text{obs}}$

De même on a la décomposition des degrés de liberté : $nr-1 = (n-1) + (r-1) + (n-1)(r-1)$

Les différentes sommes de carrés se calculent par :

$$\mathbf{SCT}_{\text{obs}} = \sum_{s=1}^n \sum_{i=1}^r (y_{si} - \bar{y})^2 = nr \times \text{variance totale}$$

$$\mathbf{SCS}_{\text{inter-sujet}} = \mathbf{SCS}_{\text{obs}} = r \sum_{s=1}^n (\bar{y}_{s0} - \bar{y})^2 = nr \times \text{variance des moyennes des sujets}$$

$$\mathbf{SCO}_{\text{obs}} = n \sum_{i=1}^r (\bar{y}_{0i} - \bar{y})^2 = nr \times \text{variance des moyennes par occasion}$$

$$\mathbf{SCR} \text{ est calculé par : } \mathbf{SCR}_{\text{obs}} = \mathbf{SCT}_{\text{obs}} - \mathbf{SCS}_{\text{obs}} - \mathbf{SCO}_{\text{obs}}$$

En pratique on peut effectuer les calculs en utilisant le tableau suivant :

Sujets	Facteur O (Occasions)						\bar{y}_{s0}	$(\bar{y}_{s0} - \bar{y})^2$
	o_1	o_2	...	o_i	...	o_r		
s_1	y_{11}	y_{12}	...	y_{1i}	...	y_{1r}	\bar{y}_{10}	$(\bar{y}_{10} - \bar{y})^2$
s_2	y_{21}	y_{22}	...	y_{2i}	...	y_{2r}	\bar{y}_{20}	$(\bar{y}_{20} - \bar{y})^2$
...
s_s	y_{s1}	y_{s2}	...	y_{si}	...	y_{sr}	\bar{y}_{s0}	$(\bar{y}_{s0} - \bar{y})^2$
...
s_n	y_{n1}	y_{n2}	...	y_{ni}	...	y_{nr}	\bar{y}_{n0}	$(\bar{y}_{n0} - \bar{y})^2$
\bar{y}_{0i}	\bar{y}_{01}	\bar{y}_{02}	...	\bar{y}_{0i}	...	\bar{y}_{0r}	$\bar{y}_{00} = \bar{y}$	$r \sum_{s=1}^n (\bar{y}_{s0} - \bar{y})^2$
$(\bar{y}_{0i} - \bar{y})^2$	$(\bar{y}_{01} - \bar{y})^2$	$(\bar{y}_{02} - \bar{y})^2$		$(\bar{y}_{0i} - \bar{y})^2$		$(\bar{y}_{0r} - \bar{y})^2$	$n \sum_{i=1}^r (\bar{y}_{0i} - \bar{y})^2$	

IV-1-4 Statistiques des tests :

On montre le théorème suivant :

Théorème : sous les hypothèses du modèle on a

- $E(\mathbf{SCO}) = (r - 1)\sigma^2 + \sum n\alpha_i^2$
- $E(\mathbf{SCS}) = (n - 1)(\sigma^2 + r\sigma_\pi^2)$
- $E(\mathbf{SCR}) = (n - 1)(r - 1)\sigma^2$

En considérant les moyennes des carrés : $\mathbf{MCO} = \frac{\mathbf{SCO}}{r - 1}$ $\mathbf{MCR} = \frac{\mathbf{SCR}}{(r - 1)(n - 1)}$

On peut réécrire le théorème précédent sous la forme :

Théorème : sous les hypothèses du modèle on a

- $E(\mathbf{MCO}) = \sigma^2 + \frac{n}{r - 1} \sum \alpha_i^2$
- $E(\mathbf{MCS}) = \sigma^2 + \frac{r}{n - 1} \sigma_\pi^2$
- $E(\mathbf{MCR}) = \sigma^2$

De nouveau on constate que **MCR** est un estimateur sans biais de la variance σ^2 et que **MCO** est un également un estimateur de la variance σ^2 augmentée d'une valeur positive traduisant l'effet du facteur. On considérera donc la statistique $\mathbf{F}_O = \frac{\mathbf{MCO}}{\mathbf{MCR}}$ pour effectuer le test suivant :

Test de l'effet du facteur O sur la variable Y :

hypothèse nulle $\mathbf{H}_0 : O$ n'a pas d'effet $\Leftrightarrow (\forall i) \alpha_i = 0$ (tous les α_i sont nuls) contre l'hypothèse alternative : $\mathbf{H}_1 : O$ a un effet $\Leftrightarrow (\exists i) \alpha_i \neq 0$ (l'un au moins des α_i est non nul).

Sous l'hypothèse nulle **MCO** et **MCR** sont deux estimateurs indépendants de σ^2 , leur rapport \mathbf{F}_O est distribué selon une loi de Fischer à $v_1 = r - 1$ et $v_2 = (n - 1)(r - 1)$ ddl.

Le test est alors défini au seuil de signification α par la règle de décision suivante :

si $\mathbf{F}_{O \text{ obs}} \geq \lambda_\alpha$ alors on rejette l'hypothèse nulle

où λ_α est donné par l'équation : $\alpha = P(\mathbf{F}_O \geq \lambda_\alpha)$.

Les résultats sont présentés dans le tableau d'analyse de la variance suivant :

Source de variation	SC	ddl	MC	\mathbf{F}_O
Inter S	$\mathbf{SCS}_{\text{obs}}$	$n - 1$	$\mathbf{MCS}_{\text{obs}}$	$\mathbf{F}_{O \text{ obs}}$
Intra S	$\mathbf{SC}_{\text{intra-sujets}}$	$n(r - 1)$		
O	$\mathbf{SCO}_{\text{obs}}$	$r - 1$	$\mathbf{MCO}_{\text{obs}}$	
R	$\mathbf{SCR}_{\text{obs}}$	$(n - 1)(r - 1)$	$\mathbf{MCR}_{\text{obs}}$	
Total	$\mathbf{SCT}_{\text{obs}}$	$rn - 1 = N - 1$		

IV-1-5 Condition de validation :

Les données peuvent être d'une part, considérées comme les $N=n \times r$ observations de la variable réponse Y dans les $N=n \times r$ conditions expérimentales décrites par le croisement du facteur sujet et du facteur occasion. Elles peuvent d'autre part, être regardées comme les n observations d'un vecteur de r variables réponses. Ces deux points de vue conduisent à l'élaboration de deux modèles, le modèle mixte univarié dans le premier cas et le modèle multivarié dans le second.

Le fait de mesurer plusieurs fois la variable réponse sur le même sujet introduit des corrélations entre les observations faites sur ce même sujet. Dans le cas du modèle mixte univarié on montre

que :

$$\text{cov}(Y_{si}, Y_{s'i'}) = \begin{cases} 0 & \text{si } s \neq s' \\ \sigma_{\pi}^2 & \text{si } s = s' \text{ et } i \neq i' \\ \sigma_{\pi}^2 + \sigma^2 & \text{si } s = s' \text{ et } i = i' \end{cases}$$

$$\text{et } \text{cor}(Y_{si}, Y_{s'i'}) = \frac{\sigma_{\pi}^2}{\sigma_{\pi}^2 + \sigma^2}$$

Dans le cas du modèle multivarié on considère que les données sont les réalisations de r vecteurs aléatoires de dimension n . $Y_s = (Y_{s1}, Y_{s2}, \dots, Y_{sr})$ indépendants et de même loi normale caractérisés par : $E(Y_{si}) = \mu_i$ $\text{Var}(Y_{si}) = \sigma_i^2$ $\text{cov}(Y_{si}, Y_{s'i'}) = \text{cov}_{ii'}$

Le modèle mixte univarié est donc un cas particulier du modèle multivarié se caractérisant par une matrice de variance-covariance Σ de la forme :

$$\Sigma = \begin{matrix} & O_1 & O_2 & \dots & O_i & \dots & O_r \\ \begin{matrix} O_1 \\ O_2 \\ \dots \\ O_i \\ \dots \\ O_r \end{matrix} & \begin{bmatrix} \sigma_{\pi}^2 + \sigma^2 & \sigma_{\pi}^2 & \dots & \sigma_{\pi}^2 & \dots & \sigma_{\pi}^2 \\ \sigma_{\pi}^2 & \sigma_{\pi}^2 + \sigma^2 & \dots & \sigma_{\pi}^2 & \dots & \sigma_{\pi}^2 \\ \vdots & \vdots & \ddots & \vdots & \dots & \vdots \\ \sigma_{\pi}^2 & \sigma_{\pi}^2 & \dots & \sigma_{\pi}^2 + \sigma^2 & \dots & \sigma_{\pi}^2 \\ \vdots & \vdots & \dots & \vdots & \ddots & \vdots \\ \sigma_{\pi}^2 & \sigma_{\pi}^2 & \dots & \sigma_{\pi}^2 & \dots & \sigma_{\pi}^2 + \sigma^2 \end{bmatrix} \end{matrix}$$

Une telle matrice où tous les termes sont positifs et où tous les termes diagonaux sont égaux et tous les autres sont égaux est dite matrice circulaire, et cette propriété est appelée la symétrie composée de la matrice.

Sur la diagonale principale de la matrice des variances-covariances Σ se trouvent les variances et les éléments hors diagonale sont des covariances.

L'hypothèse d'homogénéité des variances (*homoscédasticité*) se traduit par l'égalité des éléments de la diagonale de la matrice Σ , Les éléments en dehors de la diagonale principale de la matrice Σ doivent être de même ordre.

De même on peut considérer la matrice des corrélations ρ :

$$\rho = \begin{matrix} & O_1 & O_2 & \dots & O_i & \dots & O_r \\ \begin{matrix} O_1 \\ O_2 \\ \dots \\ O_i \\ \dots \\ O_r \end{matrix} & \begin{bmatrix} 1 & \frac{\sigma_{\pi}^2}{\sigma_{\pi}^2 + \sigma^2} & \dots & \frac{\sigma_{\pi}^2}{\sigma_{\pi}^2 + \sigma^2} & \dots & \frac{\sigma_{\pi}^2}{\sigma_{\pi}^2 + \sigma^2} \\ \frac{\sigma_{\pi}^2}{\sigma_{\pi}^2 + \sigma^2} & 1 & \dots & \frac{\sigma_{\pi}^2}{\sigma_{\pi}^2 + \sigma^2} & \dots & \frac{\sigma_{\pi}^2}{\sigma_{\pi}^2 + \sigma^2} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \frac{\sigma_{\pi}^2}{\sigma_{\pi}^2 + \sigma^2} & \frac{\sigma_{\pi}^2}{\sigma_{\pi}^2 + \sigma^2} & \dots & 1 & \dots & \frac{\sigma_{\pi}^2}{\sigma_{\pi}^2 + \sigma^2} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \frac{\sigma_{\pi}^2}{\sigma_{\pi}^2 + \sigma^2} & \frac{\sigma_{\pi}^2}{\sigma_{\pi}^2 + \sigma^2} & \dots & \frac{\sigma_{\pi}^2}{\sigma_{\pi}^2 + \sigma^2} & \dots & 1 \end{bmatrix} \end{matrix}$$

De ce fait, dans le cas du choix de modèle mixte univarié, en calculant les variances-covariances nous obtenons $\hat{\Sigma}$ l'estimation de la matrice des variances-covariances (et aussi l'estimation de la matrice des corrélations).

Si les éléments de la diagonale de la matrice des variances-covariances $\hat{\Sigma}$ sont de même ordre, alors la condition de l'homoscédasticité (homogénéité des variances) est vérifiée.

Si les éléments hors diagonale de la matrice des variances-covariance (ou des corrélations) sont de même ordre que les variances, alors le choix du modèle mixte univarié est justifié. C'est-à-dire que la condition de la circularité (de sphéricité) est acquise.

IV-1-6 Exemple :

Pendant ¼ d'heure on compte le nombre d'actions exercées par chacun des 7 rats sur un levier. Et ceci dans trois conditions de renforcement :

1^{ère} condition : on présente au rat des aliments très appréciés.

2^{ème} condition : on présente au rat des aliments moyennement appréciés.

3^{ème} condition : on présente au rat des aliments peu appréciés.

On obtient les résultats donnés par le tableau ci-contre.

Sujets	Facteur O			moyenne	$(\bar{y}_{s0} - \bar{y})^2$
	A	B	C		
s_1	8	6	2	$\bar{y}_{10}=5,333$	1,416
s_2	6	5	1	$\bar{y}_{20}=4$	0,020
s_3	7	5	0	$\bar{y}_{30}=4$	0,020
s_4	9	3	3	$\bar{y}_{40}=5$	0,734
s_5	5	4	1	$\bar{y}_{50}=3,333$	0,656
s_6	7	5	2	$\bar{y}_{60}=4,667$	0,274
s_7	6	2	0	$\bar{y}_{70}=2,667$	2,178
moyenne	$\bar{y}_{01}=6,857$	$\bar{y}_{02}=4,286$	$\bar{y}_{03}=1,286$	$\bar{y} = T/N = 4,143$	$5,298 \times 3 = 15,90$
$(\bar{y}_{0i} - \bar{y})^2$	7,365	0,020	8,162	$15,547 \times 7 = 108,857$	

$$SCT=138,5714$$

$$SCO=108,857$$

$$SCS=15,90$$

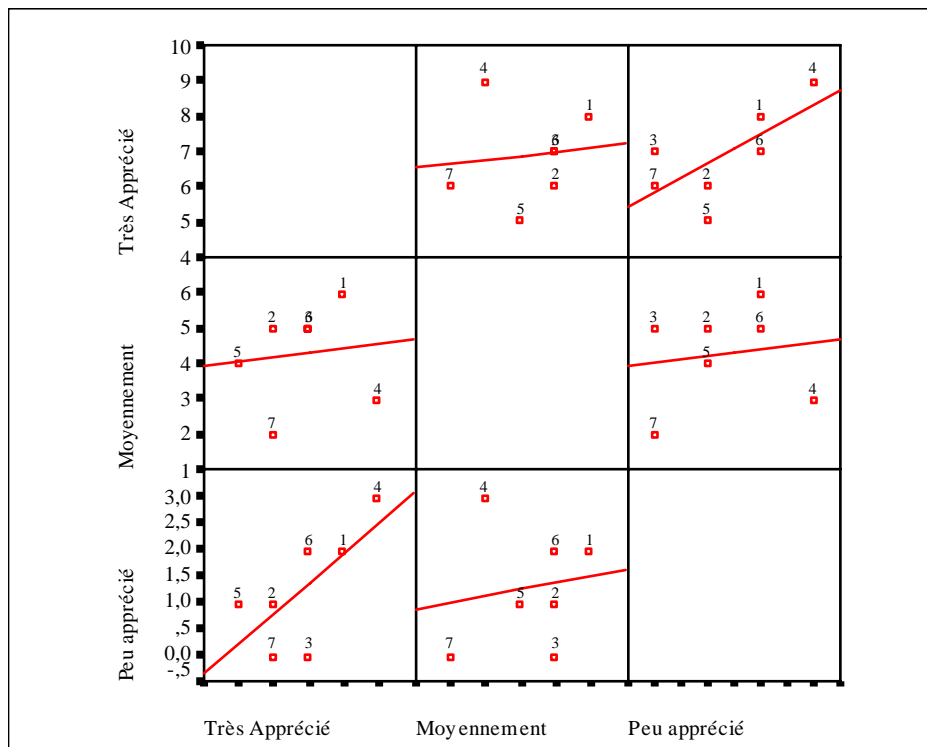
$$SCR=SCT- SCO- SCS=13,8$$

$$MCO = \frac{SCO}{r-1} = \frac{108,9}{2} = 54,45 \quad MCR = \frac{SCR}{(n-1)(r-1)} = \frac{13,8}{6 \times 2} = 1,15 \quad F_{obs} = \frac{MCO}{MCR} = \frac{54,45}{1,15} = 47,34$$

Les résultats sont présentés dans le tableau d'analyse de la variance suivant :

Source de variation	SC	ddl	MC	F_o
Inter S	$SCS_{obs}=15,90$	$n-1=6$	$MCS_{obs}=2,65$	$F_{O obs}=47,34$
Intra S	$SC_{intra-sujets}$	$n(r-1)=14$		
O	$SCO_{obs}=108,857$	$r-1=2$	$MCO_{obs}=54,45$	
R	$SCR_{obs}=13,8$	$(n-1)(r-1)=12$	$MCR_{obs}=1,15$	
Total	$SCT_{obs}=138,5714$	$rn-1=N-1=20$		

Pour vérifier la condition de la symétrie composée, nous allons établir la matrice des variances-covariances et aussi la matrice des corrélations pour les colonnes du tableau des observations.



Matrice des variances-covariances

$$\begin{matrix}
 & O_1 & O_2 & O_3 \\
 O_1 & \begin{bmatrix} 1,551 \\ 0,184 \\ 0,898 \end{bmatrix} \\
 O_2 & & \begin{bmatrix} 1,633 \\ 0,204 \end{bmatrix} \\
 O_3 & & & \begin{bmatrix} 1,061 \end{bmatrix}
 \end{matrix}$$

Matrice des corrélations

$$\begin{matrix}
 & O_1 & O_2 & O_3 \\
 O_1 & \begin{bmatrix} 1 \\ 0,115 \\ 0,700 \end{bmatrix} \\
 O_2 & & \begin{bmatrix} 1 \\ 0,155 \end{bmatrix} \\
 O_3 & & & \begin{bmatrix} 1 \end{bmatrix}
 \end{matrix}$$

On constate que les estimations des variances ne sont pas très différentes. L'hypothèse d'homogénéité des variances (homoscédasticité) peut être considérée comme acquise.

En ce qui concerne les covariances ou les corrélations, on constate une corrélation relativement forte entre les résultats des conditions 1 et 3 ($r_{13} = 0,7$).

Dans la condition 3 (aliments peu appréciés), les rats n'ont pas trop d'actions. Mais, les rats très actifs dans la condition 1 (aliments très appréciés) sont aussi actifs dans la condition 3, alors que les résultats de la condition 2 ne sont pas corrélés ni avec la condition 1 ni avec la condition 2.

Ces résultats suggèrent une réflexion supplémentaire sur ces données. Peut être faut-il examiner un modèle multivarié et effectuer une *MANOVA*.

V-1. A et B sont deux facteurs à effets fixes

V-1-1 Généralités

On considère le plan complet $S \times A \times B$ défini par le croisement de trois facteurs S , A et B . S est le facteur sujet (aléatoire) à n modalités. A et B sont deux facteurs à effets fixes, A possède r modalités et B c modalités. Chacun des n sujets est observé dans chacun des $r \times c$ croisements des modalités de A et B .

On dispose donc de $N = n \times r \times c$ observations y_{sij} que l'on présente sous la forme d'un tableau :

Sujets	a_1					...	a_i					...	a_r				
	b_1	...	b_j	...	b_c	...	b_1	...	b_j	...	b_c	...	b_1	...	b_j	...	b_c
1	y_{111}	...	y_{11j}	...	y_{11c}	...	y_{1i1}	...	y_{1ij}	...	y_{1ic}	...	y_{1r1}	...	y_{1rj}	...	y_{1rc}
2	y_{211}	...	y_{21j}	...	y_{21c}	...	y_{2i1}	...	y_{2ij}	...	y_{2ic}	...	y_{2r1}	...	y_{2rj}	...	y_{2rc}
...
s	y_{s11}	...	y_{s1j}	...	y_{s1c}	...	y_{si1}	...	y_{sij}	...	y_{sic}	...	y_{sr1}	...	y_{srj}	...	y_{src}
...
n	y_{n11}	...	y_{n1j}	...	y_{n1c}	...	y_{ni1}	...	y_{nij}	...	y_{nic}	...	y_{nr1}	...	y_{nrj}	...	y_{nrc}

V-1-2 Modèle univarié :

Les données peuvent être d'une part, considérées comme les $N = n \times r \times c$ observations d'une seule variable aléatoire Y dans les $N = n \times r \times c$ conditions expérimentales décrites par le croisement des trois facteurs S , A et B . Elles peuvent d'autre part, être regardées comme les n observations d'un vecteur de $r \times c$ variables aléatoires. Ces deux points de vue conduisent à l'élaboration de deux modèles, le modèle mixte univarié dans le premier cas, et le modèle multivarié dans le second.

Comme pour le plan $S \times O$, nous allons étudier le modèle mixte univarié dans lequel nous considérons que les modalités du facteur sujet ont été obtenues par échantillonnage, le facteur sujet est donc un facteur aléatoire. Les facteurs A et B sont des facteurs à effets fixes. Chaque donnée y_{sij} correspond alors à l'observation d'une variable aléatoire Y_{sij} et on pose le modèle mixte suivant :

$$Y_{sij} = \mu_{ij} + \pi_s + (\alpha\pi)_{is} + (\beta\pi)_{js} + (\alpha\beta\pi)_{sij} + e_{sij}$$

Pour chaque croisement des modalités s , i et j nous ne disposons que d'une seule observation, l'interaction $(\alpha\beta\pi)_{sij}$ sera confondue avec le résidu et on pose : $\varepsilon_{sij} = (\alpha\beta\pi)_{sij} + e_{sij}$

$$\text{On a donc : } Y_{sij} = \mu_{ij} + \pi_s + (\alpha\pi)_{is} + (\beta\pi)_{js} + \varepsilon_{sij}$$

Où

Les μ_{ij} ($i=1, 2, \dots, r, j=1, 2, \dots, c$) sont des constantes qui mesurent les effets fixes des modalités (i,j) du croisement $A \times B$

Les π_s ($s = 1, 2, \dots, n$) sont des variables aléatoires indépendantes et identiquement distribuées (*iid*) de loi $\mathcal{N}(0; \sigma^2_\pi)$ qui mesurent les effets aléatoires des modalités s du facteur S

Les $(\alpha\pi)_{si}$ ($s = 1, 2, \dots, n, i = 1, 2, \dots, r$) sont des variables aléatoires indépendantes et identiquement distribuées (*iid*) de loi $\mathcal{N}(0; \sigma^2_{\alpha\pi})$ qui mesurent les effets aléatoires d'interactions entre le facteur S et le facteur A

Les $(\beta\pi)_{sj}$ ($s = 1, 2, \dots, n, j = 1, 2, \dots, c$) sont des variables aléatoires indépendantes et identiquement distribuées (*iid*) de loi $\mathcal{N}(0; \sigma^2_{\beta\pi})$ qui mesurent les effets aléatoires d'interactions entre le facteur S et le facteur B

Les résidus ε_{sij} sont des variables aléatoires indépendantes et identiquement distribuées (*iid*) de loi $\mathcal{N}(0; \sigma^2)$

De plus, on suppose que les résidus ε_{sij} sont indépendantes des π_s , des $(\alpha\pi)_{si}$ et des $(\beta\pi)_{sj}$.

Nous pouvons réécrire l'effet fixe de la modalité (i,j) sous la forme suivante :

$$\mu_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}$$

Où :

Le paramètre μ s'interprète comme un niveau général de réponse commun pour l'ensemble des observations,

Le paramètre $\alpha_i = \mu_i - \mu$ s'interprète comme l'effet de la modalité i du facteur A

Le paramètre $\beta_j = \mu_j - \mu$ s'interprète comme l'effet de la modalité j du facteur B

Le paramètre $(\alpha\beta)_{ij} = \mu_{ij} - \mu_i - \mu_j + \mu$ s'interprète comme l'effet de la modalité (i,j) du croisement $A \times B$ (l'effet d'interaction de A et B).

Le modèle mixte univarié peut finalement s'écrire :

$$Y_{sij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \pi_s + (\alpha\pi)_{is} + (\beta\pi)_{js} + \varepsilon_{sij}$$

Où : les Y_{sij} sont des variables aléatoires de loi : $\mathcal{N}(\mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}, \sigma_\pi^2 + \sigma_{\alpha\pi}^2 + \sigma_{\beta\pi}^2 + \sigma^2)$

Une telle paramétrisation nécessite de rajouter les contraintes d'identifiabilité suivantes :

$$\sum_{i=1}^r \alpha_i = 0 \quad \sum_{j=1}^c \beta_j = 0 \quad \sum_{i=1}^r (\alpha\beta)_{ij} = 0 \quad \sum_{j=1}^c (\alpha\beta)_{ij} = 0$$

V-1-3 Décomposition de la variation :

On commence par décomposer la variation totale en variation inter-sujets et variation intra-sujets :

$$\mathbf{SCT}_{\text{obs}} = \mathbf{SC}_{\text{inter-sujets}} + \mathbf{SC}_{\text{intra-sujets}}$$

$$\mathbf{SC}_{\text{inter-sujet}} = rc \sum_{s=1}^n (\bar{y}_{s00} - \bar{y})^2 = N \times \text{Variance des moyennes des sujets}$$

$$\mathbf{SC}_{\text{intra-sujet}} = \sum_{s,i,j} (y_{sij} - \bar{y}_{s00})^2$$

La variation inter-sujets est due au facteur sujet et la variation intra-sujets est due aux effets du croisement $A \times B$ et à l'effet des autres facteurs non contrôlés. L'effet du facteur $A \times B$ se fait à deux niveaux : effet principal et effet d'interaction.

$$\mathbf{SC}_{\text{intra-sujets}} = \mathbf{SC}(A \times B)_{\text{obs}} + \mathbf{SC}(A \times B)S_{\text{obs}} + \mathbf{SCR}_{\text{obs}}$$

$$\text{Avec : } \mathbf{SC}(A \times B)_{\text{obs}} = \mathbf{SC}(A)_{\text{obs}} + \mathbf{SC}(B)_{\text{obs}} + \mathbf{SC}(AB)_{\text{obs}}$$

$$\text{et : } \mathbf{SC}(A \times B)S_{\text{obs}} = \mathbf{SC}(AS)_{\text{obs}} + \mathbf{SC}(BS)_{\text{obs}} + \mathbf{SC}(ABS)_{\text{obs}}$$

Ne disposant que d'une seule observation pour chaque croisement (i,j,s) on ne peut donc séparer l'interaction SAB du résidu on a donc : $\mathbf{SC}(ABS)_{\text{obs}} = \mathbf{SCR}$

Conclusion : Dans le cas du plan complet $S \times A \times B$, la variation totale se décompose de façon additive en :

$$\mathbf{SCT}_{\text{obs}} = \mathbf{SCS}_{\text{obs}} + \mathbf{SCA}_{\text{obs}} + \mathbf{SCB}_{\text{obs}} + \mathbf{SCAB}_{\text{obs}} + \mathbf{SCAS}_{\text{obs}} + \mathbf{SCBS}_{\text{obs}} + \mathbf{SCR}_{\text{obs}}$$

De même on a la décomposition des degrés de liberté :

$$N-1 = ncr-1 = (n-1) + n(rc-1) = (n-1) + n(rc-1) - (c-1)(r-1)(n-1) + (c-1)(r-1)(n-1)$$

$$N-1 = ncr-1 = (n-1) + (r-1) + (c-1) + (c-1)(r-1) + (r-1)(n-1) + (c-1)(n-1)$$

Les différentes sommes de carrés se calculent de façon habituelles par :

$$\mathbf{SCT}_{\text{obs}} = \sum_{s=1}^n \sum_{j=1}^c \sum_{i=1}^r (y_{sij} - \bar{y})^2 = N \times \text{variance de toutes les observations}$$

$$\mathbf{SCS}_{\text{obs}} = rc \sum_{s=1}^n (\bar{y}_{s00} - \bar{y})^2 = N \times \text{variance des } n \text{ moyennes des sujets}$$

$$\mathbf{SCA}_{\text{obs}} = nc \sum_{i=1}^r (\bar{y}_{0i0} - \bar{y})^2 = N \times \text{variance des } r \text{ moyennes par modalités de } A$$

$$\mathbf{SCB}_{\text{obs}} = nr \sum_{j=1}^c (\bar{y}_{00j} - \bar{y})^2 = N \times \text{variance des } c \text{ moyennes par modalités de } B$$

$$\mathbf{SC}(A \times B)_{\text{obs}} = n \sum_{i=1}^r \sum_{j=1}^c (\bar{y}_{0ij} - \bar{y})^2 = N \times \text{variance des } r \times c \text{ moyennes par croisement de } A \text{ et } B$$

$$\mathbf{SC}(AB)_{\text{obs}} = \mathbf{SC}(A \times B)_{\text{obs}} - \mathbf{SCA}_{\text{obs}} - \mathbf{SCB}_{\text{obs}}$$

$$\mathbf{SC}(A \times S)_{\text{obs}} = c \sum_{s=1}^n \sum_{i=1}^r (\bar{y}_{si0} - \bar{y})^2 = N \times \text{variance des } r \times n \text{ moyennes par croisement de } A \text{ et } S$$

$$\mathbf{SC}(AS)_{\text{obs}} = \mathbf{SC}(A \times S)_{\text{obs}} - \mathbf{SCA}_{\text{obs}} - \mathbf{SCS}_{\text{obs}}$$

$$\mathbf{SC}(B \times S)_{\text{obs}} = r \sum_{s=1}^n \sum_{j=1}^c (\bar{y}_{s0j} - \bar{y})^2 = N \times \text{variance des } n \times c \text{ moyennes par croisement de } B \text{ et } S$$

$$\mathbf{SC}(BS)_{\text{obs}} = \mathbf{SC}(B \times S)_{\text{obs}} - \mathbf{SCB}_{\text{obs}} - \mathbf{SCS}_{\text{obs}}$$

$$\mathbf{SCR}_{\text{obs}} = \mathbf{SCT}_{\text{obs}} - \mathbf{SCS}_{\text{obs}} - \mathbf{SCA}_{\text{obs}} - \mathbf{SCB}_{\text{obs}} - \mathbf{SC}(AB)_{\text{obs}} - \mathbf{SC}(AS)_{\text{obs}} - \mathbf{SC}(BS)_{\text{obs}}$$

Dans le cadre du modèle statistique ces sommes de carrés sont des réalisations de variables aléatoires dont on calcule les espérances et plus généralement les distributions des probabilités. On montre le théorème fondamental suivant :

Théorème : Sous les hypothèses du modèle, les statistiques **SCA**, **SCB**, **SCS**, **SCAB**, **SCAS**, **SCBS** et **SCR** sont indépendantes et d'espérances respectives :

$$E(\mathbf{SCS}) = (n-1)\sigma^2 + rc\sigma_{\pi}^2$$

$$E(\mathbf{SCA}) = (r-1)(\sigma^2 + c\sigma_{\alpha\pi}^2) + \sum_i nc\alpha_i^2$$

$$E(\mathbf{SCB}) = (c-1)(\sigma^2 + r\sigma_{\beta\pi}^2) + \sum_j nr\beta_j^2$$

$$E(\mathbf{SC}(AB)) = (r-1)(c-1)\sigma^2 + \sum_i \sum_j n(\alpha\beta)_{ij}^2$$

$$E(\mathbf{SC}(AS)) = (n-1)(r-1)(\sigma^2 + c\sigma_{\alpha\pi}^2)$$

$$E(\mathbf{SC}(BS)) = (n-1)(c-1)(\sigma^2 + r\sigma_{\beta\pi}^2)$$

$$E(\mathbf{SCR}) = (r-1)(c-1)(n-1)\sigma^2$$

On ramène toutes les sommes de carrés à des moyennes de carrés en divisant par les degrés de liberté correspondants.

$$\mathbf{MCA} = \frac{\mathbf{SCA}}{r-1} \quad \mathbf{MCB} = \frac{\mathbf{SCB}}{c-1} \quad \mathbf{MC}(AS) = \frac{\mathbf{SC}(AS)}{(r-1)(n-1)} \quad \mathbf{MCAB} = \frac{\mathbf{SC}(AB)}{(r-1)(c-1)}$$

$$\mathbf{MC}(BS) = \frac{\mathbf{SC}(BS)}{(c-1)(n-1)} \quad \mathbf{MCR} = \frac{\mathbf{SCR}}{(r-1)(c-1)(n-1)}$$

Ceci nous permet de réécrire le théorème précédent sous la forme suivante :

Théorème : Sous les hypothèses du modèle, les statistiques **MCA**, **MCB**, **MCS**, **MCAB**, **MCAS**, **MCBS** et **MCR** sont indépendantes et d'espérances respectives :

$$E(\text{MCS}) = \sigma^2 + \frac{rc}{(n-1)} \sigma_\pi^2$$

$$E(\text{MCA}) = \sigma^2 + c\sigma_{\alpha\pi}^2 + \frac{1}{r-1} \sum_i n c \alpha_i^2$$

$$E(\text{MCB}) = \sigma^2 + r\sigma_{\beta\pi}^2 + \frac{1}{c-1} \sum_j n r \beta_j^2$$

$$E(\text{MC(AB)}) = \sigma^2 + \frac{1}{(r-1)(c-1)} \sum_i \sum_j n (\alpha\beta)_{ij}^2$$

$$E(\text{MC(AS)}) = \sigma^2 + c\sigma_{\alpha\pi}^2$$

$$E(\text{MC(BS)}) = \sigma^2 + r\sigma_{\beta\pi}^2$$

$$E(\text{SCR}) = \sigma^2$$

Ceci nous permet de tester l'existence des différents effets des facteurs. On peut construire des tests indépendants sur chacune des sources de variations :

Test 1 :

hypothèse nulle \mathbf{H}_{0A} : pas d'effet principal du facteur A $\Leftrightarrow (\forall i) \alpha_i = 0$ contre
l'hypothèse alternative : \mathbf{H}_{1A} : il existe un effet principal du facteur A $\Leftrightarrow (\exists i) \alpha_i \neq 0$

Théorème : sous l'hypothèse \mathbf{H}_{0A} , la statistique $\mathbf{F}_A = \frac{\text{MCA}}{\text{MC(AS)}}$ suit une loi de Fischer à $r - 1$ et $(n - 1)(r - 1)$ degrés de liberté.

Le test1 est alors défini au seuil de signification α par la règle de décision suivante :

si $\mathbf{F}_{A \text{ obs}} \geq \lambda_\alpha$ alors on rejette l'hypothèse nulle

où λ_α est donné par l'équation : $\alpha = P(\mathbf{F}_A \geq \lambda_\alpha)$.

Test 2 :

hypothèse nulle \mathbf{H}_{0B} : pas d'effet principal du facteur B $\Leftrightarrow (\forall j) \beta_j = 0$ contre
l'hypothèse alternative : \mathbf{H}_{1B} : il existe un effet principal du facteur B $\Leftrightarrow (\exists j) \beta_j \neq 0$

Théorème : sous l'hypothèse \mathbf{H}_{0B} , la statistique $\mathbf{F}_B = \frac{\text{MCB}}{\text{MC(BS)}}$ suit une loi de Fischer à $c - 1$ et $(n - 1)(c - 1)$ degrés de liberté.

Le test2 est alors défini au seuil de signification α par la règle de décision suivante :

si $\mathbf{F}_{B \text{ obs}} \geq \lambda_\alpha$ alors on rejette l'hypothèse nulle

où λ_α est donné par l'équation : $\alpha = P(\mathbf{F}_B \geq \lambda_\alpha)$.

Test 3 :

hypothèse nulle \mathbf{H}_{0AB} : pas d'effet d'interaction $\Leftrightarrow (\forall i,j) (\alpha\beta)_{ij} = 0$ contre
l'hypothèse alternative : \mathbf{H}_{1AB} : il existe une interaction $\Leftrightarrow (\exists i,j) (\alpha\beta)_{ij} \neq 0$

Théorème : sous l'hypothèse \mathbf{H}_{0AB} , la statistique $\mathbf{F}_{AB} = \frac{\text{MC(AB)}}{\text{MCR}}$ suit une loi de Fischer à $(r - 1)(c - 1)$ et $(n - 1)(c - 1)(r - 1)$ degrés de liberté.

Le test3 est alors défini au seuil de signification α par la règle de décision suivante :

si $F_{AB\text{ obs}} \geq \lambda_\alpha$ alors on rejette l'hypothèse nulle

où λ_α est donné par l'équation : $\alpha = P(F_{AB} \geq \lambda_\alpha)$.

On présente l'étude dans le tableau d'analyse de la variance :

Source	SC_{obs}	ddl	MC_{obs}	F_{obs}
Inter-sujets : S	SCS_{obs}	$n - 1$		
Intra-sujets :				
A	SCA_{obs}	$r - 1$	MCA_{obs}	$F_{A\text{ obs}}$
AS	$SC(AS)_{\text{obs}}$	$(r - 1)(n - 1)$	$MC(AS)_{\text{obs}}$	
B	SCB_{obs}	$c - 1$	MCB_{obs}	$F_{B\text{ obs}}$
BS	$SC(BS)_{\text{obs}}$	$(c - 1)(n - 1)$	$MC(BS)_{\text{obs}}$	
AB	$SC(AB)_{\text{obs}}$	$(r - 1)(c - 1)$	$MC(AB)_{\text{obs}}$	$F_{AB\text{ obs}}$
R	SCR_{obs}	$(r - 1)(c - 1)(n - 1)$	MCR_{obs}	
Total	SCT_{obs}	$N - 1$		

V-1-5 Condition de validation :

Le fait de mesurer plusieurs fois la variable réponse sur le même sujet introduit des corrélations entre les observations faites sur ce même sujet. Dans le cas du modèle mixte univarié on montre

$$\text{cov}(Y_{sij}, Y_{s'i'j'}) = \begin{cases} 0 & \text{si } s \neq s' \\ \sigma_\pi^2 & \text{si } s = s' \text{ et } i \neq i' \text{ et } j \neq j' \\ \sigma_\pi^2 + \sigma_{\alpha\pi}^2 & \text{si } s = s' \text{ et } i = i' \text{ et } j \neq j' \\ \sigma_\pi^2 + \sigma_{\beta\pi}^2 & \text{si } s = s' \text{ et } i \neq i' \text{ et } j = j' \\ \sigma_\pi^2 + \sigma_{\alpha\pi}^2 + \sigma_{\beta\pi}^2 & \text{si } s = s' \text{ et } i = i' \text{ et } j = j' \end{cases}$$

que :

$$\text{cor}(Y_{sij}, Y_{s'i'j'}) = \frac{\sigma_\pi^2}{\sigma_\pi^2 + \sigma_{\alpha\pi}^2 + \sigma_{\beta\pi}^2 + \sigma^2}$$

$$\text{cor}(Y_{sij}, Y_{s'i'j}) = \frac{\sigma_\pi^2 + \sigma_{\alpha\pi}^2}{\sigma_\pi^2 + \sigma_{\alpha\pi}^2 + \sigma_{\beta\pi}^2 + \sigma^2}$$

$$\text{cor}(Y_{sij}, Y_{s'ij'}) = \frac{\sigma_\pi^2 + \sigma_{\beta\pi}^2}{\sigma_\pi^2 + \sigma_{\alpha\pi}^2 + \sigma_{\beta\pi}^2 + \sigma^2}$$

Dans le cas du modèle multivarié on considère que les données sont les réalisations de n vecteurs aléatoires de dimension rc . $Y_s = (Y_{s1}, Y_{s2}, \dots, Y_{src})$ indépendants et de même loi normale caractérisés par : $E(Y_{sk}) = \mu_{ij}$ $\text{Var}(Y_{sk}) = \sigma_{ij}^2$ $\text{cov}(Y_{sk}, Y_{sk'}) = \text{cov}_{kk'}$

Le modèle mixte univarié est donc un cas particulier du modèle multivarié correspondant aux hypothèses de circularité de la matrice de variance-covariance Σ de la forme :

		A ₁						A _i						A _r		
		B ₁	...	B _c	...	B ₁	...	B _j	...	B _c	...	B ₁	...	B _c		
A ₁	B ₁	$\sigma^2_{\pi} + \sigma^2_{\pi\alpha} + \sigma^2_{\pi\beta} + \sigma^2$														
	B _c	$\sigma^2_{\pi} + \sigma^2_{\pi\alpha}$		$\sigma^2_{\pi} + \sigma^2_{\pi\alpha} + \sigma^2_{\pi\beta} + \sigma^2$												
A _i	B ₁	$\sigma^2_{\pi} + \sigma^2_{\pi\beta}$		σ^2_{π}		$\sigma^2_{\pi} + \sigma^2_{\pi\alpha} + \sigma^2_{\pi\beta} + \sigma^2$										
	B _j	σ^2_{π}		σ^2_{π}		$\sigma^2_{\pi} + \sigma^2_{\pi\alpha}$		$\sigma^2_{\pi} + \sigma^2_{\pi\alpha} + \sigma^2_{\pi\beta} + \sigma^2$								
	B _c	σ^2_{π}		$\sigma^2_{\pi} + \sigma^2_{\pi\beta}$		$\sigma^2_{\pi} + \sigma^2_{\pi\alpha}$		$\sigma^2_{\pi} + \sigma^2_{\pi\alpha}$		$\sigma^2_{\pi} + \sigma^2_{\pi\alpha} + \sigma^2_{\pi\beta} + \sigma^2$						
	B _c	σ^2_{π}		$\sigma^2_{\pi} + \sigma^2_{\pi\beta}$		σ^2_{π}		σ^2_{π}		σ^2_{π}		$\sigma^2_{\pi} + \sigma^2_{\pi\alpha} + \sigma^2_{\pi\beta} + \sigma^2$				
A _r	B ₁	$\sigma^2_{\pi} + \sigma^2_{\pi\beta}$		σ^2_{π}		$\sigma^2_{\pi} + \sigma^2_{\pi\beta}$		σ^2_{π}		σ^2_{π}		$\sigma^2_{\pi} + \sigma^2_{\pi\alpha} + \sigma^2_{\pi\beta} + \sigma^2$				
	B _c	σ^2_{π}		$\sigma^2_{\pi} + \sigma^2_{\pi\beta}$		σ^2_{π}		σ^2_{π}		$\sigma^2_{\pi} + \sigma^2_{\pi\beta}$		$\sigma^2_{\pi} + \sigma^2_{\pi\alpha}$		$\sigma^2_{\pi} + \sigma^2_{\pi\alpha} + \sigma^2_{\pi\beta} + \sigma^2$		

		B ₁						B _j						B _c		
		A ₁	...	A _r	...	A ₁	...	A _i	...	A _r	...	A ₁	...	A _r		
B ₁	A ₁	$\sigma^2_{\pi} + \sigma^2_{\pi\alpha} + \sigma^2_{\pi\beta} + \sigma^2$														
	A _r	$\sigma^2_{\pi} + \sigma^2_{\pi\beta}$		$\sigma^2_{\pi} + \sigma^2_{\pi\alpha} + \sigma^2_{\pi\beta} + \sigma^2$												
B _j	A ₁	$\sigma^2_{\pi} + \sigma^2_{\pi\alpha}$		σ^2_{π}		$\sigma^2_{\pi} + \sigma^2_{\pi\alpha} + \sigma^2_{\pi\beta} + \sigma^2$										
	A _i	σ^2_{π}		σ^2_{π}		$\sigma^2_{\pi} + \sigma^2_{\pi\beta}$		$\sigma^2_{\pi} + \sigma^2_{\pi\alpha} + \sigma^2_{\pi\beta} + \sigma^2$								
	A _r	σ^2_{π}		$\sigma^2_{\pi} + \sigma^2_{\pi\alpha}$		$\sigma^2_{\pi} + \sigma^2_{\pi\beta}$		$\sigma^2_{\pi} + \sigma^2_{\pi\beta}$		$\sigma^2_{\pi} + \sigma^2_{\pi\alpha} + \sigma^2_{\pi\beta} + \sigma^2$						
	A _r	$\sigma^2_{\pi} + \sigma^2_{\pi\alpha}$		σ^2_{π}		$\sigma^2_{\pi} + \sigma^2_{\pi\alpha}$		σ^2_{π}		σ^2_{π}		$\sigma^2_{\pi} + \sigma^2_{\pi\alpha} + \sigma^2_{\pi\beta} + \sigma^2$				
B _c	A ₁	σ^2_{π}		$\sigma^2_{\pi} + \sigma^2_{\pi\alpha}$		σ^2_{π}		σ^2_{π}		σ^2_{π}		$\sigma^2_{\pi} + \sigma^2_{\pi\beta}$		$\sigma^2_{\pi} + \sigma^2_{\pi\alpha} + \sigma^2_{\pi\beta} + \sigma^2$		
	A _r	σ^2_{π}		$\sigma^2_{\pi} + \sigma^2_{\pi\alpha}$		σ^2_{π}		σ^2_{π}		$\sigma^2_{\pi} + \sigma^2_{\pi\alpha}$		$\sigma^2_{\pi} + \sigma^2_{\pi\beta}$		$\sigma^2_{\pi} + \sigma^2_{\pi\alpha} + \sigma^2_{\pi\beta} + \sigma^2$		

V-1-6 Exemple :

Un expérimentateur veut étudier l'effet de la consommation de lécithine sur les troubles de mémoire. Il choisit 4 sujets auxquels il administre un traitement quotidien. Au bout d'un mois, de deux mois et de six mois de traitements, il fait passer à chaque sujet deux tests. Le premier test (test 1) est le même chaque mois, le deuxième test (test 2) est une forme parallèle chaque mois.

Sujet	Test 1			Test 2		
	M ₁	M ₂	M ₆	M ₁	M ₂	M ₆
1	10	11	9	3	6	3
2	18	20	17	16	20	14
3	6	8	8	5	6	3
4	4	9	9	10	10	6

Pratique des calculs :

Sujet	Test 1				Test 2				
	M ₁	M ₂	M ₆		M ₁	M ₂	M ₆		
1	10	11	9	$\bar{y}_{110} = 10$	3	6	3	$\bar{y}_{120} = 4$	$\bar{y}_{100} = 7$
2	18	20	17	$\bar{y}_{210} = 18,33$	16	20	14	$\bar{y}_{220} = 16,66$	$\bar{y}_{200} = 17,5$
3	6	8	8	$\bar{y}_{310} = 7,33$	5	6	3	$\bar{y}_{320} = 4,66$	$\bar{y}_{300} = 6$
4	4	9	9	$\bar{y}_{410} = 7,33$	10	10	6	$\bar{y}_{420} = 8,66$	$\bar{y}_{400} = 8$
$\bar{y}_{011} = 9,5$			$\bar{y}_{012} = 12$	$\bar{y}_{013} = 10,75$	$\bar{y}_{021} = 8,5$			$\bar{y}_{022} = 10,5$	$\bar{y}_{023} = 6,5$
$\bar{y}_{010} = 10,75$				$\bar{y}_{020} = 8,5$					
				$\bar{y}_{001} = 9$					
				$\bar{y}_{002} = 11,25$					
				$\bar{y}_{003} = 8,625$					
$\bar{y} = 8$									

Pratique des calculs : autre méthode

A \ B	M ₁	M ₂	M ₆	moyennes
T ₁	9,50	12,00	10,75	10,75
T ₂	8,50	10,50	6,50	8,50
moyennes	9,00	11,25	8,625	9,625

A \ S	1	2	3	4
T ₁	10,00	18,33	7,33	7,33
T ₂	4,00	16,67	4,67	8,67
moyennes	7,00	17,50	6,00	8,00

B \ S	1	2	3	4
M ₁	6,50	17,00	5,50	7,00
M ₂	8,50	20,00	7,00	9,50
M ₆	6,00	15,50	5,50	7,50
moyennes	7,00	17,50	6,00	8,00

Calcul des sommes des carrées :

SCT=645,625 SCS=508,125 SCA=30,375 SCB=32,25
 SC(A×B)=74,875 SC(AB)=12,25 SC(A×S)=579,5652 SC(AS)=41,133
 SC(B×S)=546,125 SC(BS)=5,75 SCR=15,742

Tableau d'analyse de la variance :

Source	SC _{obs}	ddl	MC _{obs}	F _{obs}
Inter S	SCS _{obs} =508,125	$n - 1=3$		
Intra S				
A	SCA _{obs} =30,375	$r - 1=1$	MCA _{obs} =30,375	F _{A obs} =2,215
AS	SC(AS) _{obs} =41,133	$(r - 1)(n - 1)=3$	MC(AS) _{obs} =13,711	
B	SCB _{obs} =32,25	$c - 1=2$	MCB _{obs} =16,125	F _{B obs} =16,827
BS	SC(BS) _{obs} =5,75	$(c - 1)(n - 1)=6$	MC(BS) _{obs} =0,9583	
AB	SC(AB) _{obs} =12,25	$(r - 1)(c - 1)=2$	MC(AB) _{obs} =6,125	F _{AB obs} =2,334
R	SCR _{obs} =15,742	$(r - 1)(c - 1)(n - 1)=6$	MCR _{obs} =2,624	
Total	SCT _{obs} =645,625	$N - 1=23$		

Test de l'existence des différents effets :

Test 1 : Effet principal du facteur A

hypothèse nulle \mathbf{H}_{0A} : pas d'effet principal du facteur A $\Leftrightarrow (\forall i) \alpha_i = 0$ contre
 l'hypothèse alternative : \mathbf{H}_{1A} : il existe un effet principal du facteur A $\Leftrightarrow (\exists i) \alpha_i \neq 0$

Théorème : sous l'hypothèse \mathbf{H}_{0A} , la statistique $F_A = \frac{MCA}{MC(AS)}$ suit une loi de Fischer à $r - 1=1$ et $(n - 1)(r - 1)=3$ degrés de liberté.

Le test1 est alors défini au seuil de signification $\alpha=5\%$ par la règle de décision suivante :

si $F_{A\text{ obs}} \geq \lambda_\alpha=10,1$ alors on rejette l'hypothèse nulle

où λ_α est donné par l'équation : $\alpha = P(F_A \geq \lambda_\alpha)$.

Or $F_{A\text{ obs}} = 2,215 < \lambda_\alpha=10,1$ on en conclue qu'il n'y a pas d'effet principal du facteur A

Test 2 : Effet principal du facteur B

hypothèse nulle \mathbf{H}_{0B} : pas d'effet principal du facteur B $\Leftrightarrow (\forall j) \beta_j = 0$ contre
 l'hypothèse alternative : \mathbf{H}_{1B} : il existe un effet principal du facteur B $\Leftrightarrow (\exists j) \beta_j \neq 0$

Théorème : sous l'hypothèse \mathbf{H}_{0B} , la statistique $F_B = \frac{MCB}{MC(BS)}$ suit une loi de Fischer à $c - 1=2$ et $(n - 1)(c - 1)=6$ degrés de liberté.

Le test2 est alors défini au seuil de signification α par la règle de décision suivante :

si $F_{B\text{ obs}} \geq \lambda_\alpha = 5,14$ alors on rejette l'hypothèse nulle

où λ_α est donné par l'équation : $\alpha = P(F_B \geq \lambda_\alpha)$.

Or $F_{B\text{ obs}} = 16,827 > \lambda_\alpha=5,14$ on en conclue qu'il y a un effet principal du facteur B

Test 3 : Test de l'effet d'interaction des facteurs A et B

hypothèse nulle \mathbf{H}_{0AB} : pas d'effet d'interaction $\Leftrightarrow (\forall i,j) (\alpha\beta)_{ij} = 0$ contre
 l'hypothèse alternative : \mathbf{H}_{1AB} : il existe une interaction $\Leftrightarrow (\exists i,j) (\alpha\beta)_{ij} \neq 0$

Théorème : sous l'hypothèse H_{0AB} , la statistique $F_{AB} = \frac{MC(AB)}{MCR}$ suit une loi de Fischer à $(r - 1)(c - 1) = 2$ et $(n - 1)(c - 1) = 6$ degrés de liberté.

Le test3 est alors défini au seuil de signification α par la règle de décision suivante :

si $F_{AB\text{ obs}} \geq \lambda_\alpha = 5,14$ alors on rejette l'hypothèse nulle

où λ_α est donné par l'équation : $\alpha = P(F_{AB} \geq \lambda_\alpha)$.

Or $F_{AB\text{ obs}} = 2,334 < \lambda_\alpha = 5,14$ on en conclue qu'il n'y a pas d'effet d'interaction des facteurs A et B

Condition de validation du modèle :

Calculons la matrice de variances covariances empirique (et aussi la matrice des corrélations) pour examiner la condition de validité du modèle mixte univarié. Les résultats sont présentés sous forme de blocs homogènes (qui devraient l'être).

Matrice de Variances Covariances

	A	Test1			Test2		
A	B	M ₁	M ₂	M ₆	M ₁	M ₂	M ₆
Test1	M ₁	28,750					
	M ₂	24,500	22,500				
	M ₆	17,875	17,000	13,188			
Test2	M ₁	16,250	18,750	15,875	25,250		
	M ₂	24,250	18,750	20,125	27,750	32,750	
	M ₆	19,250	19,750	15,875	21,750	25,750	20,250

Matrice des corrélations.

	A	Test 1			Test 2		
A	B	M ₁	M ₂	M ₆	M ₁	M ₂	M ₆
Test 1	M ₁	1					
	M ₂	0,9633	1				
	M ₆	0,9180	0,9869	1			
Test 2	M ₁	0,6031	0,7866	0,8700	1		
	M ₂	0,7903	0,7276	0,9684	0,9650	1	
	M ₆	0,7978	0,9253	0,9714	0,9619	0,9999	1

Le premier bloc (en haut et à gauche) est le bloc des variables issues des modalités de B alors que A prend la modalité « test1 ». Le deuxième bloc (en bas et à droite) est le bloc des variables issues des modalités de B alors que A prend la modalité « test2 ». Et le troisième bloc (en bas et à gauche) est le bloc des variables issues des modalités de B et aussi des modalités de A (test1 et test2) à la fois. Dans ce dernier bloc, on distingue les éléments de la diagonale, dont les modalités de B sont les mêmes des éléments hors diagonale.

Nous constatons qu'à l'intérieur des blocs nous avons, (plus ou moins) une homogénéité des covariances (et aussi des corrélations). En ce qui concerne la condition d'homoscédasticité (homogénéité des variances), nous ne constatons pas une différence importante sur la diagonale principale de la matrice, donc nous pouvons penser que cette condition est également acquise.

VI-1. A et B sont deux facteurs à effets fixes

VI-1-1 Généralités

On considère le plan quasi-complet $S \times A \times B$ où les sujets sont répartis dans des groupes définis par les r modalités du facteur A . Chaque sujet est observé dans les différentes occasions définies par les c modalités du facteur B . Il y a n sujets par groupe. On a donc échantillonné $n \times r$ sujets et on dispose donc de $N = n \times r \times c$ observations $y_{s(i)j}$ que l'on présente sous la forme d'un tableau :

Sujets	a_1					...	a_i					...	a_r				
	b_1	...	b_j	...	b_c	...	b_1	...	b_j	...	b_c	...	b_1	...	b_j	...	b_c
1	$y_{1(1)1}$...	$y_{1(1)j}$...	$y_{1(1)c}$...	$y_{1(i)1}$...	$y_{1(i)j}$...	$y_{1(i)c}$...	$y_{1(r)1}$...	$y_{1(r)j}$...	$y_{1(r)c}$
2	$y_{2(1)1}$...	$y_{2(1)j}$...	$y_{2(1)c}$...	$y_{2(i)1}$...	$y_{2(i)j}$...	$y_{2(i)c}$...	$y_{2(r)1}$...	$y_{2(r)j}$...	$y_{2(r)c}$
...
s	$y_{s(1)1}$...	$y_{s(1)j}$...	$y_{s(1)c}$...	$y_{s(i)1}$...	$y_{s(i)j}$...	$y_{s(i)c}$...	$y_{s(r)1}$...	$y_{s(r)j}$...	$y_{s(r)c}$
...
n	$y_{n(1)1}$...	$y_{n(1)j}$...	$y_{n(1)c}$...	$y_{n(i)1}$...	$y_{n(i)j}$...	$y_{n(i)c}$...	$y_{n(r)1}$...	$y_{n(r)j}$...	$y_{n(r)c}$

VI-1-2 Modèle univarié :

Les données peuvent être d'une part, considérées comme les $N = n \times r \times c$ observations d'une seule variable aléatoire Y dans les $N = n \times r \times c$ conditions expérimentales décrites par le croisement des trois facteurs S emboîté dans A et B . Elles peuvent d'autre part, être regardées comme les $n \times r$ observations d'un vecteur de c variables aléatoires. Ces deux points de vue conduisent à l'élaboration de deux modèles, le modèle mixte univarié dans le premier cas, et le modèle multivarié dans le second.

Comme pour les plans $S \times O$ et $S \times A \times B$ nous allons étudier le modèle mixte univarié dans lequel nous considérons que les modalités du facteur sujet ont été obtenues par échantillonnage, le facteur sujet est donc un facteur aléatoire. Les facteurs A et B sont des facteurs à effets fixes. Chaque donnée $y_{s(i)j}$ correspond alors à l'observation d'une variable aléatoire $Y_{s(i)j}$ et on pose le modèle mixte suivant :

$$Y_{s(i)j} = \mu_{ij} + \pi_{s(i)} + (\pi\beta)_{sj} + e_{s(i)j}$$

Pour chaque croisement des modalités s et j nous ne disposons que d'une seule observation, l'interaction $(\pi\beta)_{sj}$ sera confondue avec le résidu et on pose : $\varepsilon_{s(i)j} = (\pi\beta)_{sj} + e_{s(i)j}$

$$\text{On a donc : } Y_{s(i)j} = \mu_{ij} + \pi_{s(i)} + \varepsilon_{s(i)j}$$

Où

Les μ_{ij} ($i=1, 2, \dots, r, j=1, 2, \dots, c$) sont des constantes qui mesurent les effets fixes des modalités (i,j) du croisement $A \times B$

Les $\pi_{s(i)}$ ($s=1, 2, \dots, n, i=1, 2, \dots, r$) sont des variables aléatoires indépendantes et identiquement distribuées (*iid*) de loi $\mathcal{N}(0; \sigma^2_\pi)$ qui mesurent les effets aléatoires des modalités s du facteur S

Les résidus $\varepsilon_{s(i)j}$ sont des variables aléatoires indépendantes et identiquement distribuées (*iid*) de loi $\mathcal{N}(0; \sigma^2)$

De plus, on suppose que les résidus $\varepsilon_{s(i)j}$ sont indépendantes des $\pi_{s(i)}$.

Nous pouvons réécrire l'effet fixe de la modalité (i,j) sous la forme suivante :

$$\mu_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}$$

Où :

Le paramètre μ s'interprète comme un niveau général de réponse commun pour l'ensemble des observations,

Le paramètre $\alpha_i = \mu_i - \mu$ s'interprète comme l'effet de la modalité i du facteur A

Le paramètre $\beta_j = \mu_j - \mu$ s'interprète comme l'effet de la modalité j du facteur B

Le paramètre $(\alpha\beta)_{ij} = \mu_{ij} - \mu_i - \mu_j + \mu$ s'interprète comme l'effet d'interaction des modalités i et j .

Le modèle mixte univarié peut finalement s'écrire :

$$Y_{s(i)j} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \pi_{s(i)} + \varepsilon_{s(i)j}$$

Où : les $Y_{s(i)j}$ sont des variables aléatoires de loi : $\mathcal{N}(\mu + \alpha_i + \beta_j, \sigma_\pi^2 + \sigma^2)$

Une telle paramétrisation nécessite de rajouter les contraintes d'identifiabilité suivantes :

$$\sum_{i=1}^r \alpha_i = 0 \quad \sum_{j=1}^c \beta_j = 0 \quad \sum_{i=1}^r (\alpha\beta)_{ij} = 0 \quad \sum_{j=1}^c (\alpha\beta)_{ij} = 0$$

VI-1-3 Décomposition de la variation :

On note : $\bar{y}_{s(i)0}$ la moyenne des c observations de la modalité $s(i)$ de l'emboîtement $S \langle A \rangle$

$\bar{y}_{0(i)0}$ la moyenne des nc observations de la modalité i de A

$\bar{y}_{0(0)j}$ la moyenne des nr observations de la modalité j de B

$\bar{y}_{0(i)j}$ la moyenne des n observations de la modalité (i,j) du croisement $A \times B$ et \bar{y} la moyenne générale

On commence par décomposer la variation totale en variation inter-sujets et variation intra-sujets :

$$\mathbf{SCT}_{\text{obs}} = \mathbf{SC}_{\text{inter-sujets}} + \mathbf{SC}_{\text{intra-sujets}}$$

$$\mathbf{SC}_{\text{inter-sujet}} = \mathbf{SCS}_{\text{obs}} = c \sum_i \sum_{s(i)} (\bar{y}_{s(i)0} - \bar{y})^2 = N \times \text{Variance des moyennes des sujets}$$

$$\mathbf{SC}_{\text{intra-sujet}} = \sum_i \sum_{s(i)} \sum_j (y_{s(i)j} - \bar{y}_{s(i)0})^2$$

On décompose la variation inter-sujets en variation inter A et variation intra A . On obtient :

$$\mathbf{SC}_{\text{inter-sujets}} = \mathbf{SCS}_{\text{obs}} = \mathbf{SC}_{\text{inter } A} + \mathbf{SC}_{\text{intra } A} \quad \text{où :}$$

$$\mathbf{SC}_{\text{inter } A} = \mathbf{SCA}_{\text{obs}} = nc \sum_{i=1}^r (\bar{y}_{0(i)0} - \bar{y})^2$$

$$\mathbf{SC}_{\text{intra } A} = \mathbf{SCS(A)}_{\text{obs}} = c \sum_i \sum_{s(i)} (\bar{y}_{s(i)0} - \bar{y}_{0(i)0})^2$$

La variation intra-sujets est due aux effets du facteur B et à l'effet des autres facteurs non contrôlés.

L'effet du facteur B se fait à deux niveaux : effet principal et effet d'interaction avec A . On a :

$$\mathbf{SC}_{\text{intra-sujets}} = \mathbf{SCB}_{\text{obs}} + \mathbf{SC(AB)}_{\text{obs}} + \mathbf{SCR}_{\text{obs}}$$

Avec : $\mathbf{SC(AB)}_{\text{obs}} = \mathbf{SC(A \times B)}_{\text{obs}} - \mathbf{SC(A)}_{\text{obs}} - \mathbf{SC(B)}_{\text{obs}}$

Conclusion : Dans le cas du plan complet $S \langle A \rangle \times B$, la variation totale se décompose de façon additive en :

$$\mathbf{SCT}_{\text{obs}} = \mathbf{SCA}_{\text{obs}} + \mathbf{SCS(A)}_{\text{obs}} + \mathbf{SCB}_{\text{obs}} + \mathbf{SCAB}_{\text{obs}} + \mathbf{SCR}_{\text{obs}}$$

De même on a la décomposition des degrés de liberté :

$$N-1 = ncr-1 = (r-1) + r(n-1) + (c-1) + (r-1)(c-1) + r(c-1)(n-1)$$

Les différentes sommes de carrés se calculent de façon habituelles par :

$$\mathbf{SCT}_{\text{obs}} = \sum_{s=1}^n \sum_{j=1}^c \sum_{i=1}^r (y_{s(i)j} - \bar{y})^2 = N \times \text{variance de toutes les observations}$$

$$\mathbf{SCA}_{\text{obs}} = nc \sum_{i=1}^r (\bar{y}_{0(i)0} - \bar{y})^2 = N \times \text{variance des } r \text{ moyennes par modalités de } A$$

$$\mathbf{SCS}_{\text{obs}} = c \sum_{s=1}^n (\bar{y}_{s(i)0} - \bar{y})^2 = N \times \text{variance des } r \times n \text{ moyennes des sujets}$$

$$\mathbf{SCS}(A)_{\text{obs}} = \mathbf{SCS}_{\text{obs}} - \mathbf{SCA}_{\text{obs}}$$

$$\mathbf{SCB}_{\text{obs}} = nr \sum_{j=1}^c (\bar{y}_{0(0)j} - \bar{y})^2 = N \times \text{variance des } c \text{ moyennes par modalités de } B$$

$$\mathbf{SC}(A \times B)_{\text{obs}} = n \sum_{i=1}^r \sum_{j=1}^c (\bar{y}_{0(i)j} - \bar{y})^2 = N \times \text{variance des } r \times c \text{ moyennes par croisement de } A \text{ et } B$$

$$\mathbf{SC}(AB)_{\text{obs}} = \mathbf{SC}(A \times B)_{\text{obs}} - \mathbf{SCA}_{\text{obs}} - \mathbf{SCB}_{\text{obs}}$$

$$\mathbf{SCR}_{\text{obs}} = \mathbf{SCT}_{\text{obs}} - \mathbf{SCA}_{\text{obs}} - \mathbf{SCB}_{\text{obs}} - \mathbf{SC}(AB)_{\text{obs}} - \mathbf{SCS}(A)_{\text{obs}}$$

Dans le cadre du modèle statistique ces sommes de carrés sont des réalisations de variables aléatoires dont on calcule les espérances et plus généralement les distributions des probabilités.

On montre le théorème fondamental suivant :

Théorème : Sous les hypothèses du modèle, les statistiques **SCA**, **SCB**, **SCAB**, **SCS(A)**, et **SCR** sont indépendantes et d'espérance respective :

$$E(\mathbf{SCA}) = (r-1)(\sigma^2 + c\sigma_{\pi}^2) + \sum_i n c \alpha_i^2$$

$$E(\mathbf{SCS}(A)) = r(n-1)(\sigma^2 + c\sigma_{\pi}^2)$$

$$E(\mathbf{SCB}) = (c-1)(\sigma^2 + \sigma_{\beta\pi}^2) + \sum_j n r \beta_j^2$$

$$E(\mathbf{SC}(AB)) = (r-1)(c-1)(\sigma^2 + \sigma_{\beta\pi}^2) + \sum_i \sum_j n (\alpha\beta)_{ij}^2$$

$$E(\mathbf{SCR}) = r(c-1)(n-1)(\sigma^2 + \sigma_{\beta\pi}^2)$$

On ramène toutes les sommes de carrés à des moyennes de carrés en divisant par les degrés de liberté correspondants.

$$\mathbf{MCA} = \frac{\mathbf{SCA}}{r-1} \quad \mathbf{MCB} = \frac{\mathbf{SCB}}{c-1} \quad \mathbf{MCS}(A) = \frac{\mathbf{SCS}(A)}{r(n-1)} \quad \mathbf{MCAB} = \frac{\mathbf{SC}(AB)}{(r-1)(c-1)} \quad \mathbf{MCR} = \frac{\mathbf{SCR}}{r(n-1)(c-1)}$$

Ceci nous permet de réécrire le théorème précédent sous la forme suivante :

Théorème : Sous les hypothèses du modèle, les statistiques **MCA**, **MCB**, **MCS(A)**, **MCAB**, et **MCR** sont indépendantes et d'espérances respectives :

$$E(\mathbf{MCA}) = \sigma^2 + c\sigma_{\pi}^2 + \frac{1}{r-1} \sum_i n c \alpha_i^2 \quad E(\mathbf{MCB}) = \sigma^2 + \sigma_{\beta\pi}^2 + \frac{1}{c-1} \sum_j n r \beta_j^2$$

$$E(\mathbf{MC}(AB)) = \sigma^2 + \sigma_{\beta\pi}^2 + \frac{1}{(r-1)(c-1)} \sum_i \sum_j n (\alpha\beta)_{ij}^2$$

$$E(\mathbf{MC}(S(A))) = \sigma^2 + c\sigma_{\pi}^2 \quad E(\mathbf{MCR}) = \sigma^2 + \sigma_{\beta\pi}^2$$

Ceci nous permet de tester l'existence des différents effets des facteurs. On peut construire des tests indépendants sur chacune des sources de variations :

Test 1 :

hypothèse nulle \mathbf{H}_{0A} : pas d'effet principal du facteur A $\Leftrightarrow (\forall i) \alpha_i = 0$ contre
 l'hypothèse alternative : \mathbf{H}_{1A} : il existe un effet principal du facteur A $\Leftrightarrow (\exists i) \alpha_i \neq 0$

Théorème : sous l'hypothèse \mathbf{H}_{0A} , la statistique $F_A = \frac{MCA}{MCS(A)}$ suit une loi de Fischer à $r - 1$ et $r(n - 1)$ degrés de liberté.

Le test1 est alors défini au seuil de signification α par la règle de décision suivante :

si $F_{A \text{ obs}} \geq \lambda_\alpha$ alors on rejette l'hypothèse nulle

où λ_α est donné par l'équation : $\alpha = P(F_A \geq \lambda_\alpha)$.

Test 2 :

hypothèse nulle \mathbf{H}_{0B} : pas d'effet principal du facteur B $\Leftrightarrow (\forall j) \beta_j = 0$ contre
 l'hypothèse alternative : \mathbf{H}_{1B} : il existe un effet principal du facteur B $\Leftrightarrow (\exists j) \beta_j \neq 0$

Théorème : sous l'hypothèse \mathbf{H}_{0B} , la statistique $F_B = \frac{MCB}{MCR}$ suit une loi de Fischer à $c - 1$ et $r(n - 1)(c - 1)$ degrés de liberté.

Le test2 est alors défini au seuil de signification α par la règle de décision suivante :

si $F_{B \text{ obs}} \geq \lambda_\alpha$ alors on rejette l'hypothèse nulle

où λ_α est donné par l'équation : $\alpha = P(F_B \geq \lambda_\alpha)$.

Test 3 :

hypothèse nulle \mathbf{H}_{0AB} : pas d'effet d'interaction $\Leftrightarrow (\forall i,j) (\alpha\beta)_{ij} = 0$ contre
 l'hypothèse alternative : \mathbf{H}_{1AB} : il existe une interaction $\Leftrightarrow (\exists i,j) (\alpha\beta)_{ij} \neq 0$

Théorème : sous l'hypothèse \mathbf{H}_{0AB} , la statistique $F_{AB} = \frac{MC(AB)}{MCR}$ suit une loi de Fischer à $(r - 1)(c - 1)$ et $r(n - 1)(c - 1)$ degrés de liberté.

Le test3 est alors défini au seuil de signification α par la règle de décision suivante :

si $F_{AB \text{ obs}} \geq \lambda_\alpha$ alors on rejette l'hypothèse nulle

où λ_α est donné par l'équation : $\alpha = P(F_{AB} \geq \lambda_\alpha)$.

On présente l'étude dans le tableau d'analyse de la variance :

Source	SC_{obs}	ddl	MC_{obs}	F_{obs}
Intra-S :				
A	SCA_{obs}	$r - 1$	MCA_{obs}	$F_{A \text{ obs}}$
S(A)	$SCS(A)_{\text{obs}}$	$r(n - 1)$	$MCS(A)_{\text{obs}}$	
Inter-S :				
B	SCB_{obs}	$c - 1$	MCB_{obs}	$F_{B \text{ obs}}$
AB	$SC(AB)_{\text{obs}}$	$(r - 1)(c - 1)$	$MC(AB)_{\text{obs}}$	$F_{AB \text{ obs}}$
R	SCR_{obs}	$r(c - 1)(n - 1)$	MCR_{obs}	
Total	SCT_{obs}	$N - 1$		

VI-1-5 Condition de validation :

Le fait de mesurer plusieurs fois la variable réponse sur le même sujet introduit des corrélations entre les observations faites sur ce même sujet. Dans le cas du modèle mixte univarié on montre

que :

$$\text{cov}(Y_{s(i)j}, Y_{s'(i')j'}) = \begin{cases} 0 & \text{si } s(i) \neq s'(i') \\ \sigma_{\pi}^2 & \text{si } j \neq j' \\ \sigma_{\pi}^2 + \sigma^2 & \text{si } j = j' \end{cases}$$

$$\text{cor}(Y_{s(i)j}, Y_{s'(i')j'}) = \frac{\sigma_{\pi}^2}{\sigma_{\pi}^2 + \sigma^2}$$

Dans le cas du modèle multivarié on considère que les données sont les réalisations de n vecteurs aléatoires de dimension rc . $Y_s = (Y_{s(i)1}, Y_{s(i)2}, \dots, Y_{s(i)c})$ indépendants et de même loi normale

caractérisés par : $E(Y_{s(i)j}) = \mu_{ij}$ $\text{Var}(Y_{s(i)j}) = \sigma_{ij}^2$ $\text{cov}(Y_{s(i)j}, Y_{s(i)j'}) = \text{cov}_{(ij)j'}$

Le modèle mixte univarié est donc un cas particulier du modèle multivarié correspondant aux hypothèses de circularité et d'homogénéité des matrices des variances-covariances Σ_i ($i = 1, 2, \dots, r$) de la forme :

$$\Sigma_i = \begin{bmatrix} \sigma_{\pi}^2 + \sigma^2 & \sigma_{\pi}^2 & \dots & \sigma_{\pi}^2 & \dots & \sigma_{\pi}^2 \\ \sigma_{\pi}^2 & \sigma_{\pi}^2 + \sigma^2 & \dots & \sigma_{\pi}^2 & \dots & \sigma_{\pi}^2 \\ \vdots & \vdots & \ddots & \vdots & \dots & \vdots \\ \sigma_{\pi}^2 & \sigma_{\pi}^2 & \dots & \sigma_{\pi}^2 + \sigma^2 & \dots & \sigma_{\pi}^2 \\ \vdots & \vdots & \dots & \vdots & \ddots & \vdots \\ \sigma_{\pi}^2 & \sigma_{\pi}^2 & \dots & \sigma_{\pi}^2 & \dots & \sigma_{\pi}^2 + \sigma^2 \end{bmatrix}$$

De même on peut considérer la matrice des corrélations ρ :

$$\rho = \begin{bmatrix} 1 & \frac{\sigma_{\pi}^2}{\sigma_{\pi}^2 + \sigma^2} & \dots & \frac{\sigma_{\pi}^2}{\sigma_{\pi}^2 + \sigma^2} & \dots & \frac{\sigma_{\pi}^2}{\sigma_{\pi}^2 + \sigma^2} \\ \frac{\sigma_{\pi}^2}{\sigma_{\pi}^2 + \sigma^2} & 1 & \dots & \frac{\sigma_{\pi}^2}{\sigma_{\pi}^2 + \sigma^2} & \dots & \frac{\sigma_{\pi}^2}{\sigma_{\pi}^2 + \sigma^2} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \frac{\sigma_{\pi}^2}{\sigma_{\pi}^2 + \sigma^2} & \frac{\sigma_{\pi}^2}{\sigma_{\pi}^2 + \sigma^2} & \dots & 1 & \dots & \frac{\sigma_{\pi}^2}{\sigma_{\pi}^2 + \sigma^2} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \frac{\sigma_{\pi}^2}{\sigma_{\pi}^2 + \sigma^2} & \frac{\sigma_{\pi}^2}{\sigma_{\pi}^2 + \sigma^2} & \dots & \frac{\sigma_{\pi}^2}{\sigma_{\pi}^2 + \sigma^2} & \dots & 1 \end{bmatrix}$$

V-1-6 Exemple :

Dans une expérience les sujets doivent estimer la longueur d'une barre métallique. Les barres présentées ont trois longueurs différentes. On forme deux groupes de 4 sujets distincts, dans chaque groupe on présente à chaque sujet trois barres de longueurs différentes ...

Sujet	G 1			G 2		
	L 1	L 2	L 3	L 1	L 2	L 3
1	10	11	9	3	6	3
2	18	20	17	16	20	14
3	6	8	8	5	6	3
4	4	9	9	10	10	6

Pratique des calculs :

Sujet	G 1				G 2			
	L 1	L 2	L 3		L 1	L 2	L 3	
1	10	11	9	$\bar{y}_{1(1)0} = 10$	3	6	3	$\bar{y}_{1(2)0} = 4$
2	18	20	17	$\bar{y}_{2(1)0} = 18,33$	16	20	14	$\bar{y}_{2(2)0} = 16,67$
3	6	8	8	$\bar{y}_{3(1)0} = 7,33$	5	6	3	$\bar{y}_{3(2)0} = 4,67$
4	4	9	9	$\bar{y}_{4(1)0} = 7,33$	10	10	6	$\bar{y}_{4(2)0} = 8,67$
			$\bar{y}_{0(1)1} = 9,5$		$\bar{y}_{0(2)1} = 8,5$	$\bar{y}_{0(2)2} = 10,5$	$\bar{y}_{0(2)3} = 6,5$	
			$\bar{y}_{0(1)2} = 12$		$\bar{y}_{0(2)0} = 8,5$			
			$\bar{y}_{0(1)3} = 10,75$					
			$\bar{y}_{0(1)0} = 10,75$					
			$\bar{y}_{0(0)1} = 9$					
			$\bar{y}_{0(0)2} = 11,25$					
			$\bar{y}_{0(0)3} = 8,625$					
								$\bar{y} = 9,625$

Pratique des calculs : autre méthode

A \ B	L ₁	L ₂	L ₃	moyennes
G ₁	9,50	12,00	10,75	10,75
G ₂	8,50	10,50	6,50	8,50
moyennes	9,00	11,25	8,625	9,625

A \ S	1	2	3	4
G ₁	10,00	18,33	7,33	7,33
G ₂	4,00	16,67	4,67	8,67

$SCT=645,625$

$SCS=579,853$

$SCA=30,375$

$SCB=32,25$

$SC(A \times B)=74,875$

$SC(AB)=12,25$

$SCS(A)=549,478$

$SCR=21,272$

Tableau d'analyse de la variance :

Source	SC _{obs}	ddl	MC _{obs}	F _{obs}
Intra-S :				
A	$SCA_{obs}=30,375$	$r - 1=1$	$MCA_{obs}=30,375$	$F_{A obs}=0,331 !$
S(A)	$SCS(A)_{obs}=549,478$	$r(n - 1)=6$	$MC(AS)_{obs}=91,579$	
Inter-S :				
B	$SCB_{obs}=32,25$	$c - 1=2$	$MCB_{obs}=16,125$	$F_{B obs}=9,099$
AB	$SC(AB)_{obs}=12,25$	$(r - 1)(c - 1)=2$	$MC(AB)_{obs}=6,125$	$F_{AB obs}=3,4565$
R	$SCR_{obs}=21,272$	$r(c - 1)(n - 1)=12$	$MCR_{obs}=1,772$	
Total	$SCT_{obs}=645,625$	$N - 1=23$		

Test de l'existence des différents effets :

Test 1 : Effet principal du facteur A

hypothèse nulle \mathbf{H}_{0A} : pas d'effet principal du facteur A $\Leftrightarrow (\forall i) \alpha_i = 0$ contre
l'hypothèse alternative : \mathbf{H}_{1A} : il existe un effet principal du facteur A $\Leftrightarrow (\exists i) \alpha_i \neq 0$

Théorème : sous l'hypothèse \mathbf{H}_{0A} , la statistique $F_A = \frac{MCA}{MCS(A)}$ suit une loi de Fischer à $r - 1$ et $r(n - 1)$ degrés de liberté.

Le test1 est alors défini au seuil de signification $\alpha=5\%$ par la règle de décision suivante :

si $F_{A \text{ obs}} \geq \lambda_\alpha = 10,1$ alors on rejette l'hypothèse nulle

où λ_α est donné par l'équation : $\alpha = P(F_A \geq \lambda_\alpha)$.

Or $F_{A \text{ obs}} = 0,331 < 1$ Dans ce cas particulier on considère $1/F_{A \text{ obs}} = 1/0,331 = 3,015$ et la comparaison se fait par rapport à la valeur lue dans la table de Fisher en inversant les degrés de liberté. $3,015 < \lambda'_{\alpha} = 234$ on en conclue qu'il n'y a pas d'effet principal du facteur A

Test 2 : Effet principal du facteur B

hypothèse nulle \mathbf{H}_{0B} : pas d'effet principal du facteur B $\Leftrightarrow (\forall j) \beta_j = 0$ contre
l'hypothèse alternative : \mathbf{H}_{1B} : il existe un effet principal du facteur B $\Leftrightarrow (\exists j) \beta_j \neq 0$

Théorème : sous l'hypothèse \mathbf{H}_{0B} , la statistique $F_B = \frac{MCB}{MCR}$ suit une loi de Fischer à $c - 1$ et $r(n - 1)(c - 1)$ degrés de liberté.

Le test2 est alors défini au seuil de signification α par la règle de décision suivante :

si $F_{B \text{ obs}} \geq \lambda_\alpha = 3,89$ alors on rejette l'hypothèse nulle

où λ_α est donné par l'équation : $\alpha = P(F_B \geq \lambda_\alpha)$.

Or $F_{B \text{ obs}} = 9,099 > \lambda_\alpha = 3,89$ on en conclue qu'il y a un effet principal du facteur B

Test 3 : Test de l'effet d'interaction des facteurs A et B

hypothèse nulle \mathbf{H}_{0AB} : pas d'effet d'interaction $\Leftrightarrow (\forall i,j) (\alpha\beta)_{ij} = 0$ contre
l'hypothèse alternative : \mathbf{H}_{1AB} : il existe une interaction $\Leftrightarrow (\exists i,j) (\alpha\beta)_{ij} \neq 0$

Théorème : sous l'hypothèse \mathbf{H}_{0AB} , la statistique $F_{AB} = \frac{MC(AB)}{MCR}$ suit une loi de Fischer à $(r - 1)(c - 1)$ et $r(n - 1)(c - 1)$ degrés de liberté.

Le test3 est alors défini au seuil de signification α par la règle de décision suivante :

si $F_{AB \text{ obs}} \geq \lambda_\alpha = 3,89$ alors on rejette l'hypothèse nulle

où λ_α est donné par l'équation : $\alpha = P(F_{AB} \geq \lambda_\alpha)$.

Or $F_{AB \text{ obs}} = 3,4565 < \lambda_\alpha = 3,89$ on en conclue qu'il n'y a pas d'effet d'interaction des facteurs A et B

Condition de validation du modèle :

Calculons la matrice de variances-covariances empirique (et aussi la matrice des corrélations) pour examiner la condition de validité du modèle mixte univarié.

Matrice des variances-covariances

Groupe 1 (A=G ₁)				Groupe 2 (A=G ₂)			
	L ₁	L ₂	L ₃		L ₁	L ₂	L ₃
L ₁	$\left[\begin{array}{ccc} 28,752 & & \\ 24,500 & 22,502 & \\ 17,875 & 17,000 & 13,188 \end{array} \right]$			L ₁	$\left[\begin{array}{ccc} 22,500 & & \\ 27,750 & 32,750 & \\ 21,750 & 25,750 & 20,250 \end{array} \right]$		
L ₂							
L ₃							

On considère la matrice des variances-covariances « moyenne »

$$\begin{array}{c} L_1 \quad L_2 \quad L_3 \\ L_1 \left[\begin{array}{ccc} 27,000 & & \\ 26,125 & 27,625 & \\ 19,813 & 21,375 & 16,719 \end{array} \right] \\ L_2 \\ L_3 \end{array}$$

et la matrice des corrélations associée :

$$\begin{array}{c} L_1 \quad L_2 \quad L_3 \\ L_1 \left[\begin{array}{ccc} 1 & & \\ 0,957 & 1 & \\ 0,933 & 0,995 & 1 \end{array} \right] \\ L_2 \\ L_3 \end{array}$$

La validité du modèle mixte univarié est considérée comme acquise.

VII-1. Généralités

VII-1-1 Notion de comparaison

Contrairement à l'analyse de la variance canonique nous nous intéressons aux situations où les hypothèses portent sur des comparaisons entre les effets associés à certaines modalités des facteurs. De ce fait on ne considère que des facteurs à effets fixes. Dans le cas d'un facteur à effets fixes, le modèle pose que les effets du facteur portent sur la moyenne de la VD. Comparer les effets de certaines modalités revient donc à comparer les moyennes qui leur sont associées dans le modèle.

Définition 1 : On appelle contraste entre k moyennes $\mu_1, \mu_2, \dots, \mu_k$ toute combinaison linéaire de la forme : $C = \sum_{i=1}^k c_i \mu_i$ où $\sum_{i=1}^k c_i = 0$

On appelle comparaison tout ensemble de contrastes.

VII-1-2 Construction d'un contraste

Une hypothèse portant sur une comparaison entre des moyennes est traduite par la donnée d'un contraste entre les moyennes. Ce contraste est caractérisé par la donnée d'un ensemble de coefficients de somme nulle. Nous donnons un procédé particulier permettant de choisir ces coefficients. Ce procédé repose sur une propriété classique de la moyenne arithmétique pondérée.

Proposition 1 : Etant donnée un ensemble de valeurs et une pondération quelconque sur cet ensemble de valeurs, la somme des écarts pondérés à la moyenne pondérée est nulle.

Comparer des moyennes revient à

- Ordonner les moyennes ou les écarts entre les moyennes.
- Associer à chaque moyenne un rang selon l'hypothèse à tester.
- Calculer la moyenne des rangs, puis centrer les rangs par rapport à leur moyenne.
- Les coefficients sont alors calculés à partir des rangs centrés en utilisant la proposition précédente.

Remarque 1 : L'usage conduit à considérer des pondérations proportionnelles aux poids des modalités. Ainsi quand les groupes sont équilibrés, on calcule une simple moyenne arithmétique. Par contre, dans le cas d'un plan non équilibré, les effectifs des groupes constituent les poids des modalités.

Certaines moyennes peuvent ne pas être concernées par des hypothèses. Les coefficients relatifs à ces moyennes seront nuls.

VII-1-3 Tests d'hypothèses sur les comparaisons

Nous avons vu que poser une hypothèse sur certains effets du facteur revient à construire un contraste C . Pour définir le test d'hypothèses, nous pouvons écrire les hypothèses de test sur le contraste C : $H_0 : C = 0$ contre $H_1 : C \neq 0$ (respectivement $C > 0$ ou $C < 0$)

Dans le cas d'une comparaison simple sur deux modalités particulières, le contraste correspondant est écrit sur les moyennes théoriques μ_1 et μ_2 : $C = \mu_1 - \mu_2$

et la statistique de test est définie à partir des observations : $\hat{C} = \bar{Y}_1 - \bar{Y}_2$

La distribution d'échantillonnage de la statistique de test $\hat{C} = \bar{Y}_1 - \bar{Y}_2$ dépend des hypothèses faites sur le modèle, (taille, échantillons, ...) et peut être soit la loi normale soit la loi de Student.

Dans le cas général, la construction des tests de comparaisons est fondée sur l'évaluation des contrastes à partir des observations. Pour effectuer un test d'hypothèses sur le contraste

$$C = \sum_{i=1}^k c_i \mu_i \quad \text{où} \quad \sum_{i=1}^k c_i = 0 \quad \text{la statistique de test est définie par :} \quad \hat{C} = \sum_{i=1}^k c_i \bar{Y}_i$$

On généralise ainsi le test de comparaison entre deux moyennes au test de comparaison entre deux groupes de moyennes.

VII-1-4 Les erreurs de première espèce

La question principale qui se pose dans tout examen des procédures de comparaisons multiples est liée à la probabilité de commettre des erreurs de première espèce. La plupart des différences entre les méthodes proposées sont dues au fait qu'on y adopte des approches différentes dans la manière de contrôler les erreurs. Le problème est en partie technique, mais il relève bien plus encore d'une question subjective : comment veut-on définir le taux d'erreur et quelle marge se donne-t-on pour le taux maximal d'erreur possible ?

Étant donné un ensemble de comparaisons, on est amené à distinguer deux types d'erreurs de première espèce :

- l'erreur par comparaison notée $\alpha(\text{EC})$:

$\alpha(\text{EC}) = P(\text{rejeter } H_0 \text{ à tort dans une comparaison donnée.})$

Nous avons déjà utilisé l'erreur par comparaison (EC). Il s'agit de la probabilité de commettre une erreur de première espèce dans une comparaison donnée. Si, par exemple, nous procédons à une comparaison en effectuant un test t entre deux groupes et rejetons l'hypothèse nulle parce que notre t excède $t_{0,05}$, nous travaillons avec une erreur par comparaison de $\alpha=0,05$.

- l'erreur de l'ensemble notée $\alpha(\text{EE})$:

$\alpha(\text{EE}) = P(\text{rejeter au moins une fois } H_0 \text{ à tort dans l'ensemble des comparaisons.})$

Lorsque nous aurons effectué un ensemble de comparaisons entre nos moyennes de groupes, nous arrivons à un ensemble (souvent appelé famille) de conclusions. Par exemple, la famille pourrait se composer des affirmations : $\mu_1 < \mu_2$ et $\mu_3 < \mu_4$ et $\mu_1 < (\mu_3 + \mu_4)/2$.

La probabilité de voir cette famille de conclusions contenir au moins une erreur de première espèce est appelée erreur de l'ensemble $\alpha(\text{EE})$.

De nombreuses procédures que nous examinons visent de façon spécifique à contrôler l'erreur d'ensemble $\alpha(\text{EE})$.

Dans une expérience qui n'implique qu'une seule comparaison on a $\alpha(\text{EC}) = \alpha(\text{EE})$. Par contre, à mesure que le nombre de comparaisons augmente, ces deux probabilités se différencient. On peut donner un calcul approché de la relation entre ces deux probabilités $\alpha(\text{EC})$ et $\alpha(\text{EE})$.

Considérons une famille de r comparaisons C^1, C^2, \dots, C^r et soit α le seuil de signification pour chaque comparaison simple : $\alpha(\text{EC}) = \alpha$.

Pour une comparaison donnée, $P(\text{ne pas rejeter } H_0) = 1 - \alpha$.

Pour un ensemble de comparaisons :

$P(\text{aucune erreur de première espèce}) = P(\text{pas d'erreur de première espèce pour } C^1 \cap \text{pas d'erreur de première espèce pour } C^2 \cap \dots \cap \text{pas d'erreur de première espèce pour } C^r)$

Si les comparaisons sont indépendantes alors : $P(\text{aucune erreur de première espèce}) = P(\text{pas d'erreur de première espèce pour } C^1) \times P(\text{pas d'erreur de première espèce pour } C^2) \times \dots \times P(\text{pas d'erreur de première espèce pour } C^r) = (1 - \alpha) \times (1 - \alpha) \times \dots \times (1 - \alpha) = (1 - \alpha)^r$

Et donc : $\alpha(\text{EE}) = 1 - (1 - \alpha(\text{EC}))^r$.

Par exemple, pour $c = 5$ et $\alpha(\text{EC}) = 0,05$, on obtient $\alpha(\text{EE}) = 1 - (1 - 0,05)^5 = 0,226$. On constate ici l'importance de contrôler l'erreur de l'ensemble.

Si on fixe la valeur de $\alpha(\text{EE})$ dans le cas de comparaisons indépendantes, on peut déduire la valeur de $\alpha(\text{EC})$ pour chacune des comparaisons : $\alpha(\text{EC}) = 1 - (1 - \alpha(\text{EE}))^{\frac{1}{r}}$

Par exemple pour cinq comparaisons indépendantes : $\alpha(\text{EE}) = 10\% \Rightarrow \alpha(\text{EC}) = 2,09\%$
 $\alpha(\text{EE}) = 5\% \Rightarrow \alpha(\text{EC}) = 1,02\% \quad \alpha(\text{EE}) = 1\% \Rightarrow \alpha(\text{EC}) = 0,20\%$

VII-1-5 Comparaisons a priori et a posteriori

Une comparaison traduit une hypothèse de l'expérimentateur. On peut distinguer alors deux types de comparaisons :

- les comparaisons a priori définies avant la collecte des données et qui traduisent une hypothèse théorique de l'expérimentateur,
- les comparaisons a posteriori planifiées lorsque l'expérimentateur a rassemblé les données, examiné les moyennes et noté quelles moyennes sont éloignées et quelles moyennes sont proches l'une de l'autre.

La distinction entre deux types d'erreur apparaît bien lorsque l'on compare les procédures de comparaisons a priori et les procédures de comparaisons a posteriori. Considérons le cas où le facteur possède 5 modalités et nous avons 5 moyennes. Dans ce cas, il y a 10 comparaisons possibles impliquant des paires de moyennes (ex. \bar{Y}_1 par rapport à \bar{Y}_2 , \bar{Y}_1 par rapport à \bar{Y}_3 , etc.).

Supposons que l'hypothèse nulle (absence d'effet) soit vérifiée.

Pour effectuer ce test nous sommes amenés à comparer les moyennes empiriques. Supposons que \bar{Y}_3 et \bar{Y}_4 soient suffisamment éloignées pour nous amener à rejeter l'hypothèse $\mu_3 = \mu_4$, et aucun autre test ne révélant une différence significative entre les moyennes. Si nous devons planifier à l'avance une comparaison unique parmi les 10 comparaisons par paires possibles, nous avons une chance sur dix de faire la comparaison qui implique une erreur de première espèce. Par contre, si nous examinons les données avant de faire le test, nous sommes certains de faire une erreur de première espèce. En effet, au vu des données, même si on veut n'effectuer qu'une seule comparaison, on procède implicitement à l'examen de toutes les comparaisons par paires avant de

choisir la plus significative. Cela équivaut alors à la situation où l'on effectue une famille de comparaisons.

Les procédures de comparaisons a posteriori ajustent α (EC) comme s'il s'agissait d'une famille de comparaisons.

Cet exemple montre que si les comparaisons sont planifiées a priori et constituent un sous-ensemble strict de l'ensemble de toutes les comparaisons possibles, alors la probabilité de commettre une erreur de première espèce est plus petite que si les comparaisons interviennent a posteriori.

VII-2. Analyse des comparaisons a priori dans un plan à un facteur

On étudie des tests portant sur des hypothèses a priori dans le cas d'un plan (inter $S \times A$ ou intra $S \times A$) où A est un facteur à effets fixes. Le facteur A à k modalités et le plan peut être soit équilibré et dans ce cas on dispose de n observations de la variable dépendante Y dans chacune des k conditions expérimentales ($N=kn$), soit un plan non équilibré et on dispose de n_i observations de la variable dépendante Y dans la modalité i du facteur A ($N = \sum_{i=1}^k n_i$).

Rappelons que les données sont alors décrites par le modèle : l'observation du sujet s dans la condition i est la réalisation d'une variable aléatoire Y_{si} de loi $\mathcal{N}(\mu_i, \sigma^2_Y)$, ($s=1, \dots, n_i$; $i=1, \dots, k$)

avec : $\sigma_Y^2 = \begin{cases} \sigma^2 & \text{dans le cas d'un plan inter} \\ \sigma^2 + \sigma_\pi^2 & \text{dans le cas d'un plan intra} \end{cases}$

VII-2-1 Analyse d'un contraste

Soit C un contraste entre les moyennes : $C = \sum_{i=1}^k c_i \mu_i$ où $\sum_{i=1}^k c_i = 0$

On veut tester $H_0 : C=0$ contre $H_1 : C \neq 0$ (respectivement $C > 0$ ou $C < 0$) au vu des N observations de la variable dépendante Y .

Proposition 2 : En estimant les moyennes μ_i par les \bar{Y}_i on peut construire un estimateur de C par :

$$\hat{C} = \sum_{i=1}^k c_i \bar{Y}_i .$$

Sous les hypothèses du modèle, \hat{C} est une variable aléatoire normalement distribuée d'espérance :

$$E(\hat{C}) = C \text{ et de variance : } \text{Var}(\hat{C}) = \sum_{i=1}^k c_i^2 \frac{\sigma^2}{n_i}$$

L'estimation $\hat{c} = \sum_{i=1}^k c_i \bar{y}_i$ est une réalisation de cette variable aléatoire \hat{C} .

Sous l'hypothèse nulle H_0 la moyenne de \hat{C} est nulle et on est donc ramené au test de comparaison à 0 de la moyenne d'une loi normale de variance inconnue. On peut procéder de deux manières.

Test de Student :

Pour un plan équilibré on prendra la statistique de test : $T = \frac{\hat{C}}{\frac{S}{\sqrt{n}} \sqrt{\sum_{i=1}^k c_i^2}} = \frac{\hat{C} \sqrt{n}}{S \sqrt{\sum_{i=1}^k c_i^2}}$

Pour un plan non équilibré on prendra la statistique de test : $T = \frac{\hat{C}}{S \sqrt{\sum_{i=1}^k \frac{c_i^2}{n_i}}}$

Où S^2 est un estimateur de σ^2 . Or on sait que MCR est aussi un estimateur de σ^2 , on a donc :

Pour un plan équilibré on prendra la statistique de test :
$$T = \frac{\hat{C}\sqrt{n}}{\sqrt{\text{MCR} \sum_{i=1}^k c_i^2}}$$

Pour un plan non équilibré on prendra la statistique de test :
$$T = \frac{\hat{C}}{\sqrt{\text{MCR} \sum_{i=1}^k \frac{c_i^2}{n_i}}}$$

Proposition 3 : Sous l'hypothèse nulle H_0 la statistique T suit une loi de Student à $v=\text{ddl}(\text{SCR})$ degrés de liberté.

Décision statistique :

Le test $H_0 : C=0$ contre $H_1 : C \neq 0$ de niveau α est défini par la règle de décision : si $|t|=|T_{obs}| \geq \lambda_\alpha$ on rejette H_0 avec λ_α donné par la table de Student à v degrés de liberté : $P(\lambda_\alpha)=1-\alpha/2$

Le test $H_0 : C=0$ contre $H_1 : C > 0$ de niveau α est défini par la règle de décision : si $t=T_{obs} \geq \lambda_\alpha$ on rejette H_0 avec λ_α donné par la table de Student à v degrés de liberté : $P(\lambda_\alpha)=1-\alpha$

Le test $H_0 : C=0$ contre $H_1 : C < 0$ de niveau α est défini par la règle de décision : si $t=T_{obs} \leq -\lambda_\alpha$ on rejette H_0 avec λ_α donné par la table de Student à v degrés de liberté : $P(\lambda_\alpha)=1-\alpha$

Test de Fisher :

On a vu qu'un contraste exprime la différence entre deux groupes de moyennes. En regroupant les moyennes correspondantes à chacun des deux groupes, on se ramène à l'analyse de la variance à un facteur à deux modalités. MCR est calculé par l'analyse de variance canonique. On montre que :

Pour un plan équilibré :
$$SC_{\text{contraste}} = SC_{\text{inter}} = \frac{n(\hat{C})^2}{\sum_{i=1}^k c_i^2}$$

Pour un plan non équilibré :
$$SC_{\text{contraste}} = SC_{\text{inter}} = \frac{(\hat{C})^2}{\sum_{i=1}^k \frac{c_i^2}{n_i}}$$

Cette somme des carrés est appelée somme des carrés du contraste et notée SC_C . On peut présenter les calculs sous la forme d'une table d'analyse de la variance :

Source	SC	ddl	MC	F
Contraste C	SC_C	1	$MC_C=SC_C$	$F_C = \frac{MC_C}{MRC}$
Résidu R	SCR	v	MCR	

Remarque 2 : Le degré de liberté d'un contraste est toujours égal à 1 (deux groupes) on a donc $MC_c= SC_c$

Pour un plan équilibré on prendra la statistique de test :
$$F_C = \frac{MC_C}{MRC} = \frac{n(\hat{C})^2}{MRC \sum_{i=1}^k c_i^2}$$

Pour un plan non équilibré on prendra la statistique de test :
$$F_C = \frac{MC_C}{MRC} = \frac{(\hat{C})^2}{MRC \sum_{i=1}^k \frac{c_i^2}{n_i}}$$

Proposition 4 : Sous l'hypothèse nulle H_0 la statistique F_C suit une loi de Fisher à $v_1=1$ et $v_2=v=ddl(SCR)$ degrés de liberté.

Décision statistique :

Le test $H_0 : C=0$ contre $H_1 : C \neq 0$ de niveau α est défini par la règle de décision : si $f_C = F_{C\text{ obs}} \geq \lambda_\alpha$ on rejette H_0 avec λ_α donné par la table de Fisher à 1 et v degrés de liberté : $P(\lambda_\alpha)=1-\alpha$

Remarque 3 : Les deux méthodes sont équivalentes car $F_C=T^2$

VII-2-2 Contrastes indépendants Contrastes orthogonaux

Définition 2 : étant donnés r contrastes C^1, C^2, \dots, C^r et r nombres $\lambda_1, \lambda_2, \dots, \lambda_r$ l'expression $\lambda_1 C^1 + \lambda_2 C^2 + \dots + \lambda_r C^r$ est une combinaison linéaire des r contrastes.

Définition 3 : un contraste est (linéairement) indépendant d'un ensemble de contrastes s'il ne peut s'écrire comme combinaison linéaire de ces contrastes.

Proposition 5 : Tout ensemble de contrastes portant sur k moyennes contient au plus $k-1$ contrastes linéairement indépendants.

Définition 4 : Considérons deux contrastes

$$C^1 = \sum_{i=1}^k c_i^1 \mu_i \quad \text{avec} \quad \sum_{i=1}^k c_i^1 = 0 \quad \text{et} \quad C^2 = \sum_{i=1}^k c_i^2 \mu_i \quad \text{avec} \quad \sum_{i=1}^k c_i^2 = 0$$

Dans le cas d'un plan équilibré les deux contrastes C^1 et C^2 sont orthogonaux si et seulement

$$\sum_{i=1}^k c_i^1 c_i^2 = 0$$

Dans le cas d'un plan non équilibré les deux contrastes C^1 et C^2 sont orthogonaux si et seulement

$$\sum_{i=1}^k \frac{c_i^1 c_i^2}{n_i} = 0$$

Proposition 6 : Dans le cas d'un plan inter, si deux contrastes sont orthogonaux alors leurs estimateurs sont non corrélés. (Ce qui équivaut, dans le cas de variables gaussiennes, à l'indépendance).

Proposition 7 : Soit une comparaison multiple définie par r contrastes C^1, C^2, \dots, C^r alors la somme des carrés associée à la comparaison est donnée par : $SC_{\text{comparaison}} = SC_{C^1} + SC_{C^2} + \dots + SC_{C^r}$ et son degré de liberté est égal à r .

Remarque 4 : Dans le cas de contrastes orthogonaux, l'usage veut de fixer l'erreur de première espèce $\alpha(ER)$.

VII-2-3 Analyse de tendance et contrastes polynomiaux

Lorsque les modalités du facteur sont ordonnées le long d'un continuum, on peut modéliser l'effet du facteur sur la variable dépendante par la donnée d'une fonction f telle que : $E(VD)=f(\text{facteur})$

La fonction f est caractérisée par une expression : $y=f(x)$. Les fonctions élémentaires sont les fonctions polynomiales : $f(x)=a_0+a_1x+a_2x^2+ \dots +a_px^p$ p est le degré du polynôme $a_0+a_1x+ \dots +a_px^p$.

Ce polynôme est une somme de monômes de degré $q \leq p$.

$q=1$: composante linéaire a_1x $q=2$: composante quadratique a_2x^2

$q=3$: composante cubique a_3x^3 $q=4$: composante d'ordre 4 a_4x^4

Les composantes sont appelées tendance.

Tester une forme particulière de liaison entre le facteur et la variable dépendante revient à tester la présence d'une composante d'un certain degré de la fonction f . Ceci revient à tester certains contrastes.

Il existe des procédures très générales permettant de calculer les coefficients des contrastes associés aux différentes tendances. Dans le cas particulier où le plan est équilibré et où les modalités du facteur peuvent être considérées comme régulièrement espacées sur un continuum les coefficients du contraste se déduisent de l'écart entre les rangs des moyennes.

Proposition 8 : Etant donné k points il existe un et un seul polynôme de degré $k-1$ passant par ces k points. Il suffit, pour un facteur à k modalités, de considérer les $k-1$ polynômes orthogonaux de degré $\leq k-1$.

Tester une tendance revient à tester un contraste. Pour connaître la contribution des tendances dans l'effet globale on mesure son intensité par : $\frac{SC_{\text{tendance}}}{SC_{\text{facteur}}}$

VII-2-4 Comparaisons non orthogonales

Dans le cas de comparaisons multiples définies a priori par un ensemble de contrastes non orthogonaux on peut contrôler l'erreur de première espèce de l'ensemble $\alpha(\text{EE})$ et ceci par différentes procédures :

Proposition 9 : Etant donnée une comparaison multiple définie à partir de r contrastes C^1, C^2, \dots, C^r On a les inégalités :

$$\text{Inégalité de Bonferroni : } \alpha(\text{EE}) \leq \sum_{i=1}^r \alpha(\text{EC}^i)$$

$$\text{Inégalité de Sidak : } \alpha(\text{EE}) \leq 1 - \prod_{i=1}^r (1 - \alpha(\text{EC}^i))$$

Dans le cas où les erreurs de première espèce par contraste sont toutes égales à $\alpha(\text{EC})$, les inégalités précédentes deviennent :

$$\text{Inégalité de Bonferroni : } \alpha(\text{EE}) \leq r\alpha(\text{EC})$$

$$\text{Inégalité de Sidak : } \alpha(\text{EE}) \leq 1 - (1 - \alpha(\text{EC}))^r$$

$$\text{Et de plus : } \alpha(\text{EE}) \leq 1 - (1 - \alpha(\text{EC}))^r \leq r\alpha(\text{EC})$$

Remarque 5 : L'inégalité de Sidak est plus précise mais toutes les deux donnent une borne supérieure à $\alpha(\text{EE})$ et permettent de répartir $\alpha(\text{EE})$ sur chacun des contrastes de façon à contrôler $\alpha(\text{EE})$ afin de ne pas dépasser le seuil de signification α .

VII-2-4-1 Test de Bonferroni et test de Sidak

A partir de l'inégalité de Bonferroni $\alpha(\text{EE}) \leq r\alpha(\text{EC})$ il suffit de prendre $\alpha(\text{EC}) \leq \frac{\alpha}{r}$ pour assurer

$\alpha(\text{EE}) \leq \alpha$. Donc il suffit de tester chaque contraste au niveau de signification : $\alpha' = \frac{\alpha}{r}$

Le test de Bonferroni est un test portant sur un contraste. La statistique de test est $T = \frac{\hat{C}\sqrt{n}}{\sqrt{\text{MCR} \sum_{i=1}^k c_i^2}}$

pour un plan équilibré et $T = \frac{\hat{C}}{\sqrt{\text{MCR} \sum_{i=1}^k \frac{c_i^2}{n_i}}}$ pour un plan non équilibré.

Sous l'hypothèse H_0 cette statistique est distribuée selon une loi de student à v ddl(MCR) mais les valeurs seront lues dans la table de Bonferroni appropriée.

A partir de l'inégalité de Sidak $\alpha(\text{EE}) \leq 1 - (1 - \alpha(\text{EC}))^r$ il suffit de prendre $\alpha(\text{EC}) \leq 1 - (1 - \alpha)^{\frac{1}{r}}$ pour assurer $\alpha(\text{EE}) \leq \alpha$. Donc il suffit de tester chaque contraste au niveau de signification :

$$\alpha' = 1 - (1 - \alpha)^{\frac{1}{r}}$$

VII-2-4-2 Le test d'Holm

Holm a proposé une alternative au test de Bonferroni pour augmenter la puissance de test tout en garantissant que $\alpha(\text{EE}) \leq \alpha$.

La statistique de test étant la même que pour le test de Bonferroni on calcule les r statistiques $t^{(i)}$ que l'on range par ordre croissant de leur valeur absolue : $|t^{(1)}| \leq |t^{(2)}| \leq \dots \leq |t^{(r)}|$

Le premier test de signification porte alors sur la significativité de $|t^{(r)}|$ pour un seuil $\alpha' = \frac{\alpha}{r}$. Si cette valeur est significative, on teste la significativité de la deuxième plus grande valeur $|t^{(r-1)}|$ pour un seuil $\alpha' = \frac{\alpha}{r-1}$. On réitère la procédure jusqu'à ce que le test donne un résultat non significatif.

VII-2-4-3 Le test de Dunnett

Dunnett a proposé une alternative au test de Bonferroni dans le cas d'une comparaison de toutes les modalités du facteur à une modalité de contrôle ou témoin. Dans un plan équilibré cette variante est plus puissante que le test de Bonferroni.

On connaît la relation exacte qui lie $\alpha(\text{EE})$ et $\alpha(\text{EC})$. Dunnett a établi la table donnant les valeurs critiques $\lambda_{\alpha'}$ pour la statistique T en fonction du nombre k de moyennes, du degré de liberté $v = \text{ddl}(\text{MCR})$ et du seuil α choisi pour $\alpha(\text{EE})$.

Notons \bar{y}_0 la moyenne du groupe contrôle et $\bar{y}_1, \bar{y}_2, \dots, \bar{y}_{k-1}$ les $k-1$ moyennes des groupes expérimentaux. Les contrastes correspondants aux différentes comparaisons sont donnés sous la forme : $C^i : \mu_i - \mu_0$ et les tests de Dunnett sont de la forme : $H_0^i : C^i = 0$ contre $H_1^i : C^i \neq 0$

La statistique de test est donnée par :
$$T = \frac{|\bar{y}_i - \bar{y}_0|}{\sqrt{\frac{2\text{MCR}}{n}}}$$

Si $t_{\text{obs}} \geq \lambda_{\alpha'}$ lu dans la table de Dunnett pour le degré de liberté $v = \text{ddl}(\text{MCR})$ et k alors H_0^i est rejetée

Une façon équivalente de réaliser ce test consiste à calculer la valeur critique $\lambda_c = \lambda_{\alpha'} \sqrt{\frac{2\text{MCR}}{n}}$ et comparer chaque différence $|\bar{y}_i - \bar{y}_0|$ à cette valeur.

VII-3. Exemple 1 : (tolérance à la morphine)

Cet exemple est tiré de Siegel (1975) « Evidence from rats that morphine tolerance is a learned response ». J. of Comparative and Physiological Psychology 80, pp 498-506.

On le retrouve dans D. C. Howell avec des données fictives mais les moyennes (et le degré de signification des différences entre les moyennes) sont les mêmes que celles de l'article de Siegel. La morphine est un médicament souvent utilisé pour atténuer la douleur. Cependant, des administrations répétées de morphine provoquent un phénomène de tolérance : la morphine a de moins en moins d'effet au fil de temps (la réduction de la douleur est de moins en moins forte). Pour mettre en évidence cette tolérance, on a souvent recours à une expérience qui consiste à placer un rat sur un support que l'on chauffe. Lorsque la chaleur devient insupportable, le rat va se mettre à se lécher les pattes. Le temps de latence qui précède le moment où le rat commence à se lécher les pattes est utilisé comme mesure de sa sensibilité à la douleur. Un rat qui vient de recevoir pour la première fois une injection de morphine montre en général un temps de latence plus long, ce qui montre que sa sensibilité à la douleur est réduite. Le développement de la tolérance à la morphine est indiqué par le fait que les temps de latence se raccourcissent progressivement (signe d'une sensibilité accrue) sous l'effet des injections répétées de morphine.

Siegel a constaté que dans plusieurs situations impliquant des médicaments autres que la morphine, les réponses conditionnées (appries) au médicament vont en sens inverse des effets inconditionnés (naturels) du médicament. Par exemple, un animal à qui l'on a injecté de l'atropine a tendance à manifester une diminution prononcée de la salivation. Par contre, si, après des injections répétées d'atropine, on administre soudain, dans le même environnement physique, une solution saline (qui ne devrait avoir absolument aucun effet), la salivation de l'animal va augmenter. C'est comme si celui-ci compensait l'effet anticipé de l'atropine. Dans ce type d'étude, il semble qu'un mécanisme compensatoire appris se développe au fil des essais pour contrebalancer l'effet du médicament. Siegel a formulé la théorie selon laquelle ce processus pourrait contribuer à expliquer la tolérance à la morphine. D'après son raisonnement, si, pendant une série d'essais préliminaires, on injecte de la morphine à l'animal placé sur un support chauffé, une certaine tolérance à la morphine va se développer. Donc, si le sujet reçoit une nouvelle injection de morphine lors d'un test ultérieur, il sera aussi sensible à la douleur qu'un animal qui n'a jamais reçu de morphine. Par contre, si lors du test ultérieur, on remplace la morphine par une solution saline, l'animale montrera une hypersensibilité et donc un temps de latence très court. En outre, selon Siegel, si on administre cette solution dans un environnement différent de celui des premières injections, le nouvel environnement ne suscitera pas d'hypersensibilité et donc l'animal réagira comme un animale qui reçoit sa première injection.

L'expérience implique cinq groupes de huit rats. Chaque groupe participe à quatre essais et les données d'analyse sont uniquement prélevées lors du dernier essai dit essai test.

On désigne les groupes en indiquant le traitement appliqué lors des trois premiers essais puis du quatrième.

- Le groupe M-M a reçu des injections de morphine lors des trois premiers essais puis de nouveau lors du quatrième toujours dans le même environnement, appelé environnement standard de test.
- Le groupe M-S a reçu des injections de morphine lors des trois premiers essais puis une solution saline lors du quatrième toujours dans l'environnement standard de test.

- Le groupe Mc-M a reçu des injections de morphine lors des trois premiers essais qui avaient lieu dans leur cage habituelle, puis une nouvelle injection de morphine lors du quatrième essai, mais dans l'environnement standard de test qui ne le connaissait pas. Dans ce groupe, les indices initialement associés à l'injection de morphine n'étaient pas présents lors du test ; on ne devait donc pas s'attendre à constater, chez ces animaux, une tolérance à la morphine lors du test.
- Le groupe S-M a reçu des injections de solution saline durant les trois premiers essais puis une injection de morphine lors du quatrième toujours dans l'environnement standard de test. On s'attend à ce que ces animaux manifestent la sensibilité la plus réduite à la douleur puisqu'ils n'ont eu aucune occasion de développer une certaine tolérance à la morphine.
- Enfin, le groupe S-S a reçu une injection de solution saline lors des quatre essais toujours dans l'environnement standard de test.

Si Siegel a raison, c'est le groupe S-M qui devrait présenter les temps de latence les plus longs (indiquant une sensibilité minimale) et le groupe M-S les temps de latence les plus courts (sensibilité maximale). Le groupe Mc-M devrait se rapprocher du groupe S-M puisque les indices associés aux trois premiers essais du groupe Mc-M ne sont pas présents lors du test. Les groupes M-M et S-S devraient se situer à un niveau intermédiaire. L'égalité ou non des groupes M-M et S-S dépendra de la vitesse à laquelle se développe la tolérance à la morphine. Le schéma des résultats ainsi anticipés est le suivant :

$$S-M = Mc-M > M-M ? S-S > M-S$$

Le point d'interrogation indique l'absence de prédiction. La variable dépendante (VD) est le temps de latence (en secondes) qui s'écoule avant que l'animal ne commence à se lécher les pattes.

On obtient les données suivantes :

	M-S	M-M	S-S	S-M	Mc-M
	3	2	14	29	24
	5	12	6	20	26
	1	13	12	36	40
	8	6	4	21	32
	1	10	19	25	20
	1	7	3	18	33
	4	11	9	26	27
	9	19	21	17	30
Moyenne \bar{y}_i	4,00	10,00	11,00	24,00	29,00
Ecart-types s'_i	3,16	5,13	6,72	6,37	6,16

On construit alors le tableau d'analyse de variance

Source des variations	SC	ddl	MC	F
Traitement	3497,6	4	874,4	27,33***
Résidu (Erreur)	1120,0	35	32,0	
Total	4617,6	39		

L'analyse globale de la variance est nettement significative, ce qui indique la présence de différences entre les cinq groupes de traitement. Mais cette analyse n'apporte pas de réponse aux questions posées.

Supposons que nous ayons prévu (a priori) de comparer les deux groupes recevant une solution saline lors de la quatrième injection avec les autres groupes. Nous avons également prévu de comparer le groupe Mc-M au groupe M-M, ainsi que le groupe M-S au groupe S-S pour des raisons

exposées lors de l'examen des tests t multiples. Enfin, nous avons prévu de comparer le groupe M-M au groupe S-S pour voir si une tolérance à la morphine s'est développée au point que les animaux ayant toujours reçu des injections de morphine ne se différencient pas, après quatre essais, des animaux ayant toujours reçu des injections de solution saline.

Nous pouvons écrire les contrastes correspondants suivants :

	M-S	M-M	S-S	S-M	Mc-M	$\sum c_i^2$	\hat{C}_{obs}^i
C^1	-3	2	-3	2	2	30	81
C^2	0	-1	0	0	1	2	19
C^3	-1	0	1	0	0	2	7
C^4	0	1	-1	0	0	2	-1

Nous allons tester l'hypothèse $H_0 : C^i = 0$ contre $H_1 : C^i \neq 0$ pour chacun des contrastes (Remarquer que H_1 est toujours bilatérale). Dans ce cas, nous pouvons utiliser aussi bien la statistique $Q_c = T$ que la statistique $Q_c = F_{obs}$. Bien sûr, nous pouvons tester l'hypothèse unilatérale pour certains contrastes et dans ce cas nous devons utiliser la statistique $Q_c = T$ en se servant de la table de Student.

Le plan est équilibré. On estime chacun des contrastes C^i par \hat{C}_{obs}^i

$$\hat{C}_{obs}^1 = (-3) \times 4 + 2 \times 10 + (-3) \times 11 + 2 \times 24 + 2 \times 29 = 81$$

$$\hat{C}_{obs}^2 = 0 \times 4 + (-1) \times 10 + 0 \times 11 + 0 \times 24 + 1 \times 29 = 19$$

$$\hat{C}_{obs}^3 = (-1) \times 4 + 0 \times 10 + 1 \times 11 + 0 \times 24 + 0 \times 29 = 7$$

$$\hat{C}_{obs}^4 = 0 \times 4 + 1 \times 10 + (-1) \times 11 + 0 \times 24 + 0 \times 29 = -1$$

Calculons la statistique $F_{C^i_{obs}}$

$$SC_{C^1} = \frac{8 \times 81^2}{30} = 1749,60 \Rightarrow F_{C^1} = \frac{1749,60}{32} = 54,675^{**}$$

$$SC_{C^2} = \frac{8 \times 19^2}{2} = 1444,00 \Rightarrow F_{C^2} = \frac{1444,00}{32} = 45,125^{**}$$

$$SC_{C^3} = \frac{8 \times 7^2}{2} = 196,00 \Rightarrow F_{C^3} = \frac{196,00}{32} = 6,125^*$$

$$SC_{C^4} = \frac{8 \times (-1)^2}{2} = 4,00 \Rightarrow F_{C^4} = \frac{4,00}{32} = 0,125 \text{ NS}$$

Chacune de ces valeurs peut être évaluée par rapport à la valeur $\lambda_\alpha = F_{(1; 35; 0,95)} = 4,12$.

Les trois premiers contrastes comme attendus sont significatifs. Le dernier contraste correspondant à "?" indique qu'une tolérance à la morphine se développe après seulement quatre injections de morphine.

Concernant l'erreur de l'ensemble des comparaisons effectuées, si nous nous limitons aux trois premiers contrastes et avec une erreur par comparaison $\alpha(EP) = 5\%$, nous obtenons une erreur d'ensemble $\alpha(EE)$ égale à : $\alpha(EE) = 1 - (1 - \alpha(EP))^r = 1 - (1 - 0,05)^3 = 1 - (0,95)^3 = 14,26\%$

Bien sûr, en rajoutant la quatrième comparaison cette erreur d'ensemble sera encore plus grande (à peu près 18%). Cette probabilité d'erreur peut paraître trop importante. Il existe plusieurs possibilités pour la réduire :

- Effectuer chaque comparaison avec un $\alpha(EP)$ plus petit, par exemple $\alpha(EP) = 1\%$, donne $\alpha(EE) = 4\%$,
- Utiliser une procédure permettant de contrôler $\alpha(EE)$.
- Réduire le nombre de contrastes en prenant des contrastes indépendants.

VII-4. Exemple 2 :

Dans une expérience concernant le traitement de l'information visuelle, la tâche des sujets est de détecter le plus rapidement possible une cible parmi des distracteurs. Les stimuli varient selon cinq niveaux de complexité, de 1, le plus simple, à 5, le plus complexe (un nombre identique de stimuli est présenté pour chaque niveau de complexité). La complexité est évaluée en terme de quantité d'information présente dans le stimulus. Plus un stimulus est complexe, plus il fournit d'information et plus le décodage est long. En revanche, pour les stimuli les plus simples, le décodage est rapide, mais l'information est pauvre et ralentit donc le temps de décision. Les auteurs font donc l'hypothèse que la performance du sujet va passer par un optimum correspondant à un compromis entre simplicité et quantité d'information. Les temps de réaction (en centième de secondes) pour des réponses correctes sont présentés dans le tableau suivant :

Sujets	Niveau de complexité				
	1	2	3	4	5
1	121	58	47	72	135
2	97	103	58	76	183
3	87	105	72	91	129
4	90	87	62	102	159
5	85	102	75	76	138
6	87	79	74	128	132
7	119	96	81	99	132
Moyenne \bar{y}_i	98	90	67	92	144
Ecart-types s'_i					

Dans la suite nous analyserons ces données en supposant que les cinq groupes sont indépendants puis en supposant qu'un même sujet effectue les cinq tâches.

Le Facteur complexité (A) comprend cinq modalités, on définit alors les quatre contrastes polynomiaux en cherchant les coefficients du contraste dans la table appropriée.

	1	2	3	4	5		
Moyenne \bar{y}_i	98	90	67	92	144	$\sum c_i^2$	\hat{C}_{obs}^i
Linéaire	-2	-1	0	1	2	10	94
Quadratique	2	-1	-2	-1	2	14	168
Cubique	-1	2	0	-2	1	10	42
Degré 4	1	-4	6	-4	1	70	-84

Le plan est équilibré. On estime chacun des contrastesⁱ par \hat{C}_{obs}^i

$$\hat{C}_{obs}^1 = (-2) \times 98 + (-1) \times 90 + 0 \times 67 + 1 \times 92 + 2 \times 144 = 94$$

$$\hat{C}_{obs}^2 = 2 \times 98 + (-1) \times 90 + (-2) \times 67 + (-1) \times 92 + 2 \times 144 = 168$$

$$\hat{C}_{obs}^3 = (-1) \times 98 + 2 \times 90 + 0 \times 67 + (-2) \times 92 + 1 \times 144 = 42$$

$$\hat{C}_{obs}^4 = 1 \times 98 + (-4) \times 90 + 6 \times 67 + (-4) \times 92 + 1 \times 144 = -84$$

Calculons la somme des carrés des contrastes : $SC_{\text{contraste}} = \frac{n(\hat{C})^2}{\sum_{i=1}^k c_i^2}$

$$SC_{C^1} = \frac{7 \times 94^2}{10} = 6185,2 \quad SC_{C^2} = \frac{7 \times 168^2}{14} = 14112,0$$

$$SC_{C^3} = \frac{7 \times 42^2}{10} = 1234,8 \quad SC_{C^4} = \frac{7 \times (-84)^2}{70} = 705,6$$

Si nous analysons ces données en supposant que les cinq groupes sont indépendants (plan inter $S \times A$), on calcule de façon habituelle les sommes des carrés : $SCT=30985,6$ et $SCA=22237,6$. On en déduit $SCR=8748$.

Comme les quatre contrastes correspondant aux différentes tendances sont orthogonaux on peut vérifier que $SCA=SC_{linéaire}+SC_{quadratique}+SC_{cubique}+SC_{degré\ 4}$

Le calcul des F_{obs} se fait aussi de manière habituelle par le rapport de la moyenne des carrés du contraste à la moyenne des carrés du résidu et on présente les résultats dans

le tableau d'analyse de la variance :

Sources des variations	SC	ddl	MC	F
Complexité A	22237,6	4	5559,4	$F_A = 19,065^{***}$
Linéaire	6185,2	1	6185,2	$F_{lin} = 21,211^{***}$
Quadratique	14112,0	1	14112,0	$F_{quad} = 48,395^{***}$
Cubique	1234,8	1	1234,8	$F_{cub} = 4,235^*$
Degré 4	705,6	1	705,6	$F_{deg\ 4} = 2,420$ (ns)
Résidu R	8748	30	291,6	
Total	30985,6	34		

$$\lambda_\alpha = F_{(1; 30; 0,95)}=4,17 \quad \lambda_\alpha = F_{(1; 30; 0,99)}=7,56$$

Si nous analysons ces données en supposant qu'un même sujet effectue les cinq tâches (plan intra $S \times A$), on calcule de façon habituelle les sommes des carrés : $SCT=30985,6$; $SCA=22237,6$ et $SCS=1154,4$. On en déduit $SCR=7593,6$.

Comme les quatre contrastes correspondant aux différentes tendances sont orthogonaux on peut vérifier que $SCA=SC_{linéaire}+SC_{quadratique}+SC_{cubique}+SC_{degré\ 4}$

Le calcul des F_{obs} se fait aussi de manière habituelle par le rapport de la moyenne des carrés du contraste à la moyenne des carrés du résidu et on présente les résultats dans

le tableau d'analyse de la variance :

Sources des variations	SC	ddl	MC	F
Sujet S	1154,4	6		
Complexité A	22237,6	4	5559,4	$F_A = 17,571^{***}$
Linéaire	6185,2	1	6185,2	$F_{lin} = 19,549^{***}$
Quadratique	14112,0	1	14112,0	$F_{quad} = 44,602^{***}$
Cubique	1234,8	1	1234,8	$F_{cub} = 3,906$ (ns)
Degré 4	705,6	1	705,6	$F_{deg\ 4} = 2,230$ (ns)
Résidu R	7593,6	24	316,4	
Total	30985,6	34		

$$\lambda_\alpha = F_{(1; 24; 0,95)}=4,26 \quad \lambda_\alpha = F_{(1; 24; 0,99)}=7,82$$

Coefficient d'intensité :

On mesure l'intensité de la contribution de chaque tendance par le rapport : $\frac{SC_{tendance}}{SC_{facteur}}$

$$\text{Tendance linéaire : } \frac{SC_{linéaire}}{SC_{facteur}} = \frac{6185,2}{22237,6} = 0,2781 = 27,81\%$$

$$\text{Tendance quadratique : } \frac{SC_{quadratique}}{SC_{facteur}} = \frac{14112,0}{22237,6} = 0,6346 = 63,46\%$$

$$\text{Tendance cubique : } \frac{SC_{cubique}}{SC_{facteur}} = \frac{1234,8}{22237,6} = 0,0555 = 5,55\%$$

$$\text{Tendance degré 4 : } \frac{SC_{degré\ 4}}{SC_{facteur}} = \frac{705,6}{22237,6} = 0,0317 = 3,17\%$$

Table des matières

Eléments de Méthodologie	1
0-1. Un peu de vocabulaire	1
0-2. Constituants d'une expérience.....	1
0-3. Les effets des facteurs	2
0-4. Plans d'expérience.....	2
Modélisation statistique.....	4
I-1. Modèle statistique	4
I-2. Décomposition de la variance associée à un facteur : (rappel)	5
I-3. Distribution de Fisher-Snedecor	7
Analyse du plan $S<G>$	9
II-1. G est un (seul) facteur à effets fixes	9
II-2. G est un (seul) facteur à effets aléatoires.....	12
Analyse du plan $S<A \times B>$	15
III-1. Plan $S<A \times B>$ avec A et B deux facteurs à effets fixes	15
Analyse du plan $S \times 0$	21
IV-1. O est un (seul) facteur à effets fixes.....	21
Analyse du plan $S \times A \times B$.....	27
V-1. A et B sont deux facteurs à effets fixes	27
Analyse du plan $S<A> \times B$	36
VI-1. A et B sont deux facteurs à effets fixes	36
Analyse des comparaisons.....	44
VII-1. Généralités	44
VII-2. Analyse des comparaisons a priori dans un plan à un facteur	47
VII-3. Exemple 1 : (tolérance à la morphine).....	52
VII-4. Exemple 2 :	55