

---



---

## Cours 9 : Plans à plusieurs facteurs

---



---

### Table des matières

Section 1. Diviser pour regner, rassembler pour saisir .....	3
Section 2. Définitions et notations.....	3
2.1. Définitions .....	3
2.2. Notations.....	4
Section 3. Effets principaux, interaction, effets simples.....	5
3.1. Type d'effets principaux.....	5
3.2. Types d'effets simples.....	6
Section 4. Répartition de la somme des carrés et des degrés de liberté .....	7
Section 5. Exemples.....	9
5.1. Plan factoriel $2 \times 2$ : Thérapie $\times$ Séances sans interaction .....	9
5.2. Plan factoriel $2 \times 2$ revisité : Thérapie $\times$ Séances avec interaction.....	11
5.3. Plan $3 \times 3$ : Niveau d'association et nombre de mots sur la mémoire .....	13
a. Moyenne par.....	13
Section 6. Comparaison de moyennes.....	15
6.1. Test de Tukey .....	16
a. Plan à un seul facteur .....	16
b. Plan à deux facteurs.....	17
6.2. Tableau récapitulatif .....	18
Section 7. Conclusion .....	18
Exercices .....	19

### Lectures

Suggérée : Howell, chapitre 13.1 à 13.4 et pour les comparaisons de moyennes, 12.1 à 12.5 (seulement la section « la statistique d'écart studentisée ») et 12.10, 12.11.

### Objectifs

Réaliser une analyse de variances sur  $p \times q$  moyennes provenant d'une expérience factorielle où deux facteurs A et B ont été manipulés. Plus précisément, pouvoir identifier si

une interaction est présente, et si oui, faire les effets simples, sinon, faire les effets principaux. Pouvoir faire un test de Tukey sur les moyennes intermédiaires quand le nombre de niveau est supérieur à 2.

---

## Section 1. Diviser pour regner, rassembler pour saisir

---

Le mathématicien Henri Poincaré faisait remarquer au début du XX<sup>ème</sup> siècle que la science progresse de façon très particulière: une théorie représente une avancée significative si elle remplace plusieurs théories distinctes ou si elle englobe plusieurs phénomènes d'allure différents. Ce fut le cas pour la théorie de Newton qui permettait d'englober la mécanique céleste et la théorie de la chute des graves (la pesanteur). Or, au fur et à mesure que la science fait des théories de plus en plus générale expliquant de plus en plus de facteurs ayant de plus en plus de niveaux, de nouveaux phénomènes inexpliqués apparaissent à la frange du compris. Comme le dit Poincaré:

Dans l'histoire de la [science], on distingue deux tendances inverses. D'une part, on découvre à chaque instant des liens nouveaux entre des objets qui semblaient devoir rester étrangers; ils tendent à s'ordonner en une imposante synthèse. La science marche vers l'unité et la simplicité. D'autre part, l'observation nous révèle tous les jours des faits nouveaux; ce que nous croyions simple redevient complexe et la science paraît marcher vers la variété et la complication.

De ces deux tendances inverses, laquelle l'emportera? Si c'est la première, la science est possible; mais rien ne le prouve à priori. (La science et l'hypothèse, 1902, p. 183)

---

## Section 2. Définitions et notations

---

### 2.1. Définitions

---

Un plan factoriel est un plan d'expérience où l'on fait, dans une même expérience, l'étude simultanée de deux ou plusieurs variables indépendantes (facteurs), afin de connaître le rôle propre de chaque variable indépendante, leur importance relative et leur interaction.

Le terme facteur est utilisé dans le même sens que les termes traitement ou variable indépendante. Le traitement réfère toujours à un traitement expérimental qui peut être ou ne pas être un traitement thérapeutique.

Un plan factoriel est un plan d'expérience dans lequel toutes les combinaisons possibles des niveaux de tous les facteurs sont représentées. Supposons par exemple que l'on veuille évaluer, dans une même expérience, l'effet de trois dosages différents d'un médicament donné sur le comportement de deux catégories de patients (dépressifs et schizophrènes). Dans ce cas, on utilisera un plan factoriel à deux facteurs: Dosages (dose 1, dose 2, dose 3) par Catégories (dépressifs, schizophrènes). Le plan complet comprendra six cellules, tel qu'illustré ici :

	Dosage (A)		
Catégorie (B)	1	2	3
Dépressifs	grp 1	grp 2	grp 3
Schizophrène	grp 4	grp 5	grp 6

Chaque cellule du plan représente une combinaison des deux facteurs : Dosage  $\times$  Catégorie. On peut utiliser (comme ici) des groupes de sujets différents pour chacun des dosages, et dans ce cas, on parle d'un plan factoriel à groupes indépendants. Si on utilise, pour chaque catégorie de patient, le même groupe de sujets sous chacun des dosages, on parle d'un plan factoriel à mesures répétées sur le facteur Dosage. Dans certains cas, un groupe de sujets peut être utilisé sous tous les niveaux de tous les facteurs, nommé un plan factoriel à mesures répétées sur tous les facteurs. Dans le cours 9, nous nous concentrons sur les plans factoriels à groupes indépendants. Le cours suivant est consacré aux plans avec des mesures répétées.

## 2.2. Notations

---

Les traitements d'un plan factoriel seront identifiés dans la suite par des lettres majuscules du début de l'alphabet, tel  $A, B, C$ , etc. Les niveaux des traitements sont désignés par les lettres minuscules correspondantes et indicées. Par exemple,  $a_1$  indique le premier niveau du facteur  $A$ . Dans l'exemple précédent,  $a_1$  désigne le dosage 1 et  $b_2$  la catégorie des schizophrènes. Le nombre total de niveaux pour chaque traitement sera désigné par les lettres  $p, q, r$ , etc. (au cours précédent, nous utilisions  $p$  pour le nombre de groupes puisqu'il n'y avait d'un seul traitement administré à des groupes indépendants). Pour identifier le nombre de sujets à l'intérieur de chaque cellule, nous utiliserons la lettre  $n$ . Si les cellules ont un nombre inégal de sujets (si et seulement si ce sont des groupes indépendants), nous utiliserons  $n_{ij}$  pour indiquer le nombre de sujets dans la cellule correspondant au niveau  $a_i b_j$  des facteurs  $A$  et  $B$ .

En bref, le traitement  $A$  possède  $p$  niveaux notés  $a_1, a_2, \dots, a_p$ , le traitement  $B$  possède  $q$  niveaux notés  $b_1, b_2, \dots, b_q$ , etc. La donnée brute obtenue par le  $k$ ème sujet au niveau  $a_i b_j$  des facteurs  $A$  et  $B$  est notée  $X_{ijk}$ . La moyenne des  $k$  sujets de cette cellule est notée  $\bar{X}_{ij}$ . Finalement, la moyenne d'une colonne (à un niveau  $a_i$  donné) est noté  $\bar{X}_{i\cdot}$  et la moyenne d'une ligne (à un niveau  $b_j$  donné) est noté  $\bar{X}_{\cdot j}$ .

La notation symbolique d'un plan factoriel sera indiquée par le nombre de facteurs et le nombre de niveaux de chaque facteur. Ainsi, l'appellation « plan  $p \times q$  » désigne un plan factoriel à deux facteurs comportant  $p$  niveaux du facteur  $A$  et  $q$  du facteur  $B$ . Par exemple, un «  $3 \times 2 \times 5$  » désigne un plan à trois facteurs, avec un total de 30 cellules correspondantes aux croisements des trois niveaux du facteur  $A$  par les deux niveaux du facteur  $B$  par les cinq niveaux du facteur  $C$ .

L'utilisation de parenthèses dans la notation symbolique d'un plan factoriel identifie les facteurs à mesures répétées. Ainsi, la notation  $p \times (q)$  désigne un plan factoriel comprenant  $p$  groupes indépendants, un pour chaque niveau de  $A$ , chaque groupe étant mesuré  $q$  fois sur chacun des niveaux du facteur  $B$ . De même, un plan  $(p \times q)$  désigne un plan factoriel avec un seul groupe, mais mesuré sur toutes les combinaisons des  $p \times q$  combinaisons des facteurs  $A$  et  $B$ . On étend facilement cette notation aux plans à trois facteurs, tel un  $p \times (q \times r)$ .

**Section 3. Effets principaux, interaction, effets simples**

La force des plans factoriels est de permettre d'analyser plusieurs V. I. simultanément. Cependant, les patrons de résultats possibles se complexifient rapidement. En particulier, on peut obtenir des interactions, concept qu'on avait rapidement abordé lors des tableaux de contingences (cours 6). Afin de simplifier l'exposé, imaginons des résultats obtenus dans une expérience de type  $2 \times 2$  avec les facteurs A ( $a_1, a_2$ ) et B ( $b_1, b_2$ ). Par exemple, ce pourrait être l'impact de la fatigue (24 heures sans sommeil, 36 heures sans sommeil) et de l'alcool (6 doses, 10 doses) sur le nombre moyen d'erreurs dans une tâche de mémoire. Pour illustrer nos résultats hypothétiques, nous utilisons un graphique des moyennes. Les barres d'erreurs sont absentes, mais supposons que des moyennes proches ne diffèrent pas significativement.

**3.1. Type d'effets principaux**

*Absence d'effet autant du traitement A que du traitement B.* Dans l'exemple, on conclurait que la fatigue et l'alcool n'ont aucune influence sur le nombre d'erreurs dans la tâche de mémoire (variable dépendante, V. D.). Visuellement, on voit que les résultats sont plats et non distinguables entre les niveaux de B. Le premier panneau de la Figure 1 en donne un exemple.

*Effet principal du traitement A seulement.* Ici, seul la fatigue a un effet. Un plus haut niveau de fatigue amène le sujet à commettre plus d'erreurs. Le facteur alcool, pour sa part, n'a aucun impact sur les moyennes. Le second panneau de la Figure 1 en donne un exemple.

*Effet principal du traitement B seulement.* L'inverse du cas précédent. Seulement le niveau d'alcool affecte le nombre d'erreurs. Pour sa part, le niveau de fatigue n'influence pas le nombre moyen d'erreurs. Le troisième panneau de la Figure 1 en donne un exemple.

*Effets principaux des traitements A et B.* Les deux patrons précédents jumelés dans le même graphe. Autant la fatigue que l'alcool augmentent le nombre d'erreurs dans la tâche de mémoire. De plus, l'effet des deux facteurs se conjuguent, de telle façon que si  $a_i$  ou  $b_j$  réduisent la performance, le décretement conjoint causé par  $a_i$  et  $b_j$  simultanément est au maximum. Autrement dit, les effets s'accroissent. Le dernier panneau de la Figure 1 en donne un exemple.

*Effet principal (déf.) :* Un effet principal se dit de l'effet d'un facteur qui est présent peu importe le niveau des autres facteurs. Dans notre exemple, on dira que la fatigue a un effet principal si l'effet de la fatigue est le même peu importe la dose d'alcool. Les effets principaux sont les plus simples à analyser car ils peuvent être décrits chacun à leur tour. Dans le dernier graphe de la Figure 1, on pourra consacrer un paragraphe à interpréter l'effet de l'alcool, puis

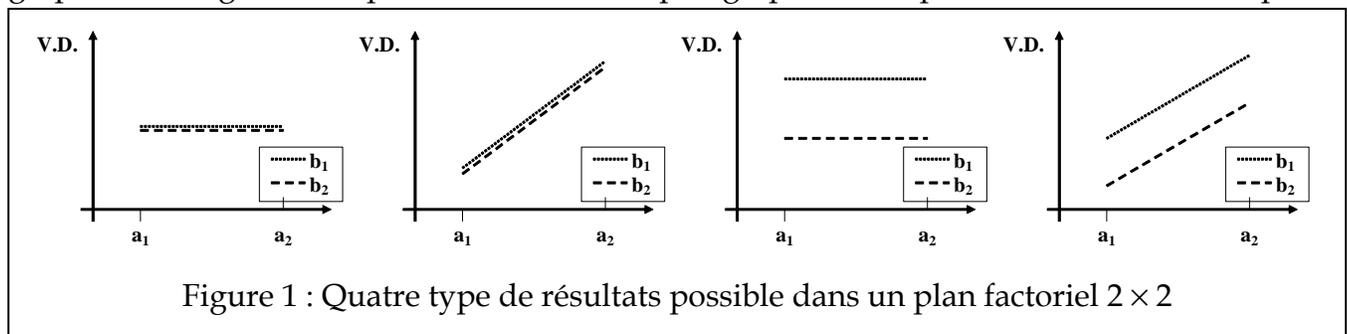


Figure 1 : Quatre type de résultats possible dans un plan factoriel  $2 \times 2$

un second pour interpréter l'effet de la fatigue (ou vice versa).

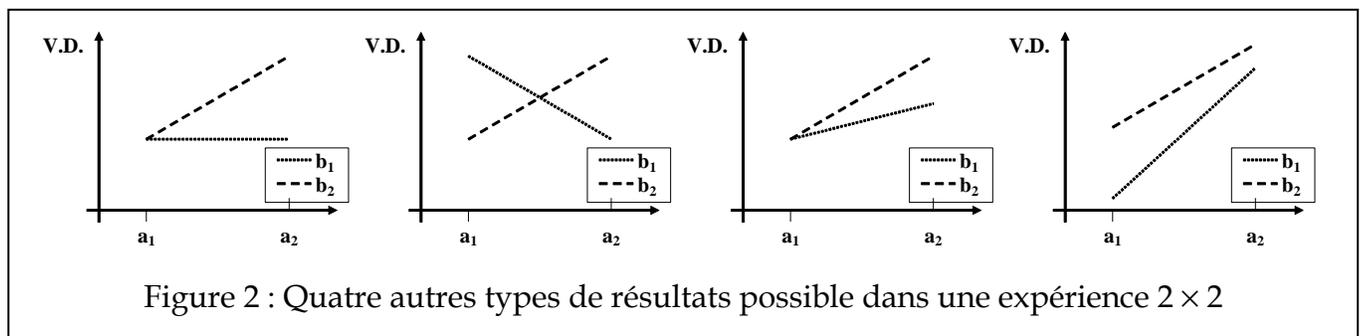
### 3.2. Types d'effets simples

*Interaction conditionnelle* : Effet simple du traitement B au niveau de  $a_2$  seulement. Dans l'exemple du premier panneau de la Figure 2, avec 24 heures sans sommeil, la dose d'alcool n'influence pas la performance dans la tâche de mémoire. Cependant, après 36 heures sans sommeil, la dose d'alcool est déterminante pour savoir si les performances vont se détériorer. 8 doses réduisent les performances significativement plus que 6 doses seulement.

*Interaction croisée*: Effet simple du traitement B au niveau de  $a_1$ , et aussi un effet simple du traitement B au niveau de  $a_2$ , mais inverse. Cet étrange patron de résultat indique que lorsque la privation de sommeil dure 24 heures, 6 doses d'alcool sont plus nuisibles à la performance que 8. Par contre, après 36 heures de privation, 8 doses sont plus nuisibles que 6. Le second panneau de la Figure 2 en donne un exemple.

*Interaction sur-additive* difficile à interpréter car si on regarde aux niveaux du traitement A, on voit que le traitement B est sans effet au niveau  $a_1$ , et a un effet au niveau  $a_2$ . Par contre, si l'on porte le regard opposé, c'est à dire aux niveaux du traitement B, on voit que A a un effet au niveau de  $b_1$  et que A a aussi un effet au niveau de  $b_2$ . L'interaction vient du fait que l'effet du traitement A n'est pas égal aux différents niveaux de B. Le troisième panneau de la Figure 2 en donne un exemple.

*Interaction sous-additive* aussi difficile : Au niveau de  $a_1$ , l'effet de B est significatif. Au niveau de  $a_2$ , l'effet de B est aussi significatif, mais moins important qu'au niveau  $a_1$ . Si le chercheur opte pour le point de vue complémentaire, il conclue que l'effet de A est significatif au niveau  $b_1$  et qu'il l'est aussi au niveau  $b_2$ , mais moins important dans ce second cas. Le dernier panneau de la Figure 2 en donne un exemple.



*Interaction* (déf.) : Une interaction dans les données est présente lorsque les effets visualisés sur un graphe ne sont pas parallèles. Lorsqu'une interaction est présente (et significative), la notion d'effet principale n'a plus de sens. Le chercheur, après avoir rapporté la présence d'une interaction, ne doit pas rapporter de statistiques aux sujets des effets principaux. Il doit plutôt se concentrer immédiatement sur les effets simples. Celles-ci forcent le chercheur à interpréter ses résultats en procédant à une *décomposition* des effets.

Une façon de quantifier une interaction dans un plan  $2 \times 2$  est de rapporter l'interaction moyenne (très rarement fait, cependant). L'interaction moyenne se calcule en additionnant les

moyennes obtenues aux niveaux externes des deux variables moins la somme des moyennes obtenues aux niveaux internes. Par exemple:  $(\bar{X}_{22} + \bar{X}_{11}) - (\bar{X}_{12} + \bar{X}_{21})$ . Si le résultat est zéro, il n'y a pas d'interaction présente dans les données. Si l'interaction moyenne est supérieure à zéro, on parle d'une interaction sur-additive. Dans le cas opposé, nous avons une interaction sous-additive. Certaines théories font des prédictions précises sur la valeur numérique de l'interaction moyenne.

*Décomposition* : Décomposer une interaction signifie procéder à l'analyse des effets d'un facteur (disons A) pour chaque niveau de l'autre facteur (soit  $b_1, b_2, \dots$ ). Une décomposition nécessite donc plus de travail. De plus, l'interprétation doit être plus nuancée car le chercheur va rapporter des effets qui dépendent du niveau de l'autre facteur. Le choix d'interpréter A aux différents niveaux de B, ou d'interpréter B aux différents niveaux de A est libre (et dépend des hypothèses et des V. I. présentes). Cependant, le chercheur n'a pas le droit de procéder aux deux décompositions puisqu'elles s'excluent mutuellement.

*Effet simple* : nom donné à l'effet d'un facteur (disons A) lorsqu'il est conditionnel au niveau de l'autre facteur (disons  $b_1$ ).

#### Section 4. Répartition de la somme des carrés et des degrés de liberté

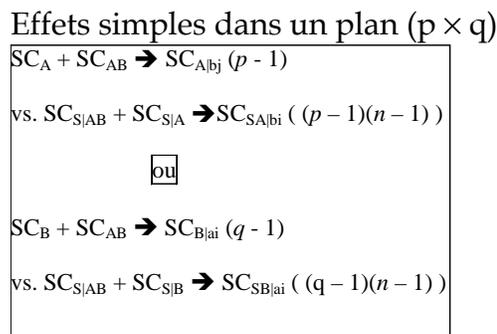
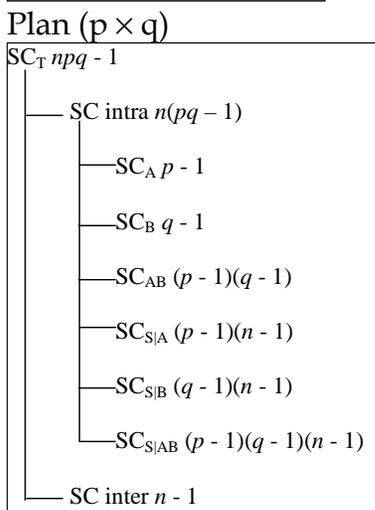
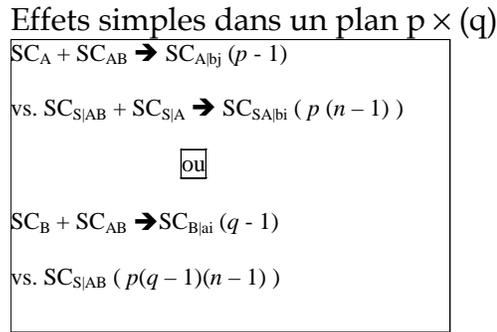
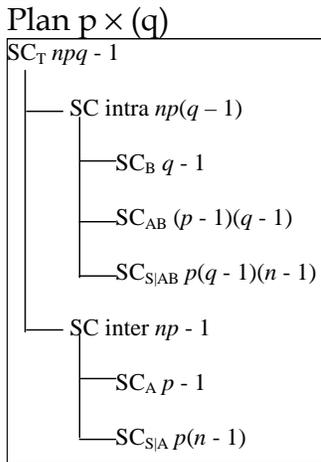
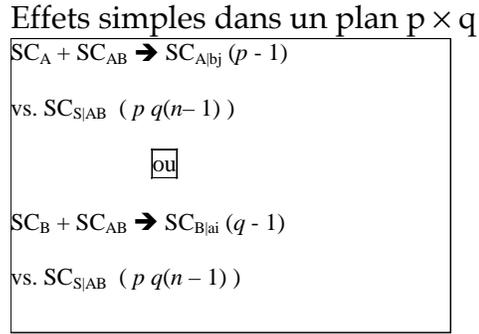
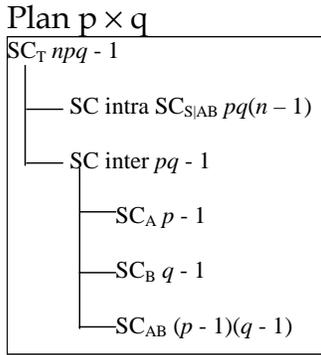
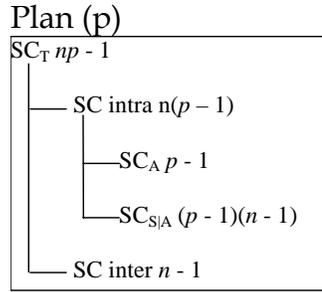
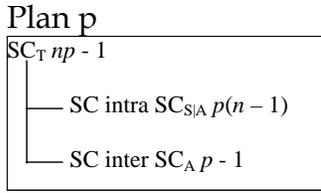
Pour démontrer la répartition de la somme des carrés, il faut procéder à des manipulations algébriques semblables à celle du cours précédent. On utilise comme point de départ la relation :

$$\mathbf{X}_{ijk} - \bar{\mathbf{X}} = (\mathbf{X}_{ijk} - \bar{\mathbf{X}}_{ij}) + (\bar{\mathbf{X}}_{ij} - \bar{\mathbf{X}}_{i\cdot} - \bar{\mathbf{X}}_{\cdot j} + \bar{\mathbf{X}}) + (\bar{\mathbf{X}}_{i\cdot} - \bar{\mathbf{X}}) + (\bar{\mathbf{X}}_{\cdot j} - \bar{\mathbf{X}})$$

En élevant chaque côté au carré, en faisant la triple somme sur tous les niveaux de  $a_i, b_j$  et tous les sujets  $n$ , on obtient la répartition de la somme des carrés. En substance, le premier terme de droite représente la somme des carrés des sujets par rapport à leurs traitements respectifs  $SC_{S|AB}$ , le troisième et quatrième terme sont la somme des carrés d'un traitement par rapport à la moyenne globale, soit  $SC_A$  et  $SC_B$ . Le second terme représente l'effet d'interaction,  $SC_{AB}$ . Comme toujours, la variance se répartit en variance intra groupe et intergroupe. Dans la variance intergroupe se trouve la variance des traitements et de leur interaction puisqu'il s'agit d'un plan à groupes indépendants. La variance intra groupe est un estimé de l'erreur expérimentale, et constitue donc le terme d'erreur lorsque nous appliquerons le test  $F$ .

La page suivante illustre la répartition de la somme des carrés pour les plans à un et à deux facteurs, avec et sans mesures répétées. Le plan  $p$  a été couvert au cours 8, le plan  $p \times q$  l'est dans ce cours, et les plans  $(p)$ ,  $p \times (q)$ , et  $(p \times q)$  le seront au cours 10.

9.3 (suite) Répartition de la somme des carrés et des degrés de liberté pour les plans à un et à deux facteurs



**Section 5. Exemples**

**5.1. Plan factoriel 2 × 2 : Thérapie × Séances sans interaction**

Soit une expérience dans laquelle un chercheur veut évaluer l'effet de deux types de thérapie (immersion, désensibilisation) et du nombre de séances (5, 10) sur le comportement de 20 sujets divisés en quatre groupes indépendants de 5 sujets sur le traitement de la peur de parler en public. La variable dépendante étant le nombre de minutes pendant lesquelles le patient peut parler devant un public de 20 personnes avant de manifester des signes d'anxiétés. Le seuil de signification  $\alpha$  est de 5%.

Les données brutes sont données dans ce tableau :

Séances (B)		
Thérapie (A)	5 ( $b_1$ )	10 ( $b_2$ )
immersion ( $a_1$ )	2.40, 2.95, 2.05, 2.75, 2.10	4.20, 4.90, 4.10, 4.00, 4.20
désensibilisation ( $a_2$ )	3.80, 4.75, 4.55, 4.95, 4.30	6.75, 6.95, 6.60, 6.90, 6.55

On synthétise facilement les résultats dans ce tableau des moyennes :

Séances (B)			Moyenne par thérapie
Thérapie (A)	5 ( $b_1$ )	10 ( $b_2$ )	
immersion ( $a_1$ )	2.45	4.28	3.37
désensibilisation ( $a_2$ )	4.47	6.75	5.61
Moyenne par séance	3.46	5.52	<b>4.49</b>

Dans un premier regard, il semble qu'il n'y ait pas d'interaction (l'interaction moyenne est de 0.45, comparable à l'écart type à l'intérieur d'un groupe). Si vous faites le graphique, on voit que les résultats semblent être parallèles, un autre signe d'absence d'interaction. En utilisant un logiciel, on obtient cette table d'ANOVA synthétisant les résultats :

Analyse de variance  
Expérience Thérapie par Séances

source de variation	SC	dl	CM	F
Intergroupe	46.57	3		
Thérapie (A)	25.20	1	25.20	196.59
Séances (B)	21.12	1	21.12	164.72
Interaction (AB)	0.25	1	0.25	1.97
Intra groupe (S   AB) (erreur)	2.05	16	0.12	
<b>Total</b>	<b>48.62</b>	<b>19</b>		

Vérifiez premièrement que la somme des carrés est bien additive. Vérifiez aussi que les degrés de libertés sont bien répartis. Vérifiez aussi que les carrés moyens sont calculés correctement. Finalement, vérifiez les ratios  $F$ . Remarquez que deux ratios semblent nettement supérieurs à 1 (dans un test  $F$ , on veut savoir si le ratio est  $> 1$ ). Une fois la validité du tableau vérifiée, il nous reste à le comprendre.

La première hypothèse que nous devons vérifier dans le tableau concerne l'interaction. En effet, si une interaction est présente, il faut immédiatement passer aux effets simples (et procéder à des analyses complémentaires). Si l'interaction n'est pas significative, on peut analyser les effets principaux directement.

L'hypothèse implicite dans l'interaction est que l'interaction n'est pas différente de zéro. Formellement, dans notre plan  $2 \times 2$ , nous notons :

$H_0$  : Interaction  $A \times B$  absente

On rejette  $H_0$  si  $\frac{CM_{AB}}{CM_{S|AB}} > s_{AB}(\alpha)$  dans laquelle la valeur  $\frac{CM_{AB}}{CM_{S|AB}}$  se distribue comme une

distribution de  $F$  avec les degrés de liberté  $(p-1)(q-1)$ ,  $pq(n-1)$ , soit dans notre exemple,  $(1, 16)$ . La valeur critique  $s_{AB}(\alpha)$  est obtenue par la table  $F$  et est de 4.494. Comme on le voit, le ratio  $\frac{CM_{AB}}{CM_{S|AB}}$  dans le tableau à la ligne Interaction (AB) est de 1.97, pas supérieur à 4.494. On ne rejette donc pas  $H_0$ , et concluons que l'interaction n'est pas significative.

Étant donné l'absence d'interaction, on peut examiner les effets principaux. Les hypothèses sont:

$H_{0A}$  Absence d'effet du facteur A

$H_{0B}$  : Absence d'effet du facteur B

En substance, on teste si les moyennes marginales sont les mêmes ou différents.

On rejette les hypothèses  $H_{0A}$  si  $\frac{CM_A}{CM_{S|AB}} > s_A(\alpha)$  et  $H_{0B}$  si  $\frac{CM_B}{CM_{S|AB}} > s_B(\alpha)$ . Il apparaît qu'ici,

les valeurs critiques  $s_A(\alpha)$  et  $s_B(\alpha)$  sont identiques puisque les degrés de libertés sont les mêmes  $(1, 16)$ . Ce n'est pas forcément le cas, puisque les degrés de libertés dépendent du nombre de niveaux pour chaque facteur.

Dans notre cas, les deux effets principaux excèdent la valeur critique. Ne reste plus qu'à conclure. Dans l'expérience, l'effet de la variable Thérapie est significative ( $F(1,16) = 196.6$ ,  $p < .05$ ). La thérapie par désensibilisation permet aux participants de parler en public plus longtemps avant de manifester des signes d'anxiétés (5.6 min) qu'avec la thérapie par immersion (3.7 min). Le nombre de séances a aussi un effet significatif ( $F(1,16) = 164.7$ ,  $p < .05$ ). Les participants ayant reçus 10 sessions parlent en public en moyenne 5.5 min avant de manifester des signes d'anxiétés, contre 3.5 min avec seulement 5 séances. L'interaction Séance par Thérapie n'est pas significative ( $F(1,16) = 1.97$ ,  $p > .05$ ).

5.2. Plan factoriel 2 × 2 revisité : Thérapie × Séances avec interaction

Supposons plutôt que les résultats obtenus dans l'expérience décrite ci-haut aient été :

		Séances (B)	
Thérapie (A)		5 (b <sub>1</sub> )	10 (b <sub>2</sub> )
immersion (a <sub>1</sub> )		3.40, 3.95, 4.05, 3.75, 4.35	2.80, 2.90, 3.10, 3.00, 2.35
désensibilisation (a <sub>2</sub> )		2.80, 3.75, 2.55, 2.70, 3.30	7.75, 7.95, 8.60, 8.90, 7.80

On synthétise facilement les résultats dans ce tableau des moyennes :

		Séances (B)		Moyenne par thérapie
Thérapie (A)		5 (b <sub>1</sub> )	10 (b <sub>2</sub> )	
immersion (a <sub>1</sub> )		3.90	2.83	3.37
désensibilisation (a <sub>2</sub> )		3.02	8.20	5.61
Moyenne par séance		3.46	5.52	<b>4.49</b>

Remarquez que les moyennes marginales sont absolument identiques à ce que nous avons obtenues précédemment. Est-ce à dire que les résultats sont comparables ? Si on jette un coup d'œil informel à l'interaction moyenne, on trouve une valeur de -6.25, une valeur nettement supérieure à l'écart type à l'intérieur d'un groupe, ce qui suggère une très forte interaction sous-additive. Si vous faites le graphique, on voit que les lignes sont loin d'être parallèles, appuyant le résultat de l'interaction moyenne. En utilisant un logiciel, on obtient cette table d'ANOVA synthétisant les résultats :

Analyse de variance  
Expérience Thérapie par Séances (II)

source de variation	SC	dl	CM	F
Intergroupe	95.15	3		
Thérapie (A)	25.20	1	25.20	139.23
Séances (B)	21.12	1	21.12	116.66
Interaction (AB)	48.83	1	48.83	269.77
Intra groupe (S   AB) (erreur)	2.90	16	0.18	
<b>Total</b>	<b>98.05</b>	<b>19</b>		

Commencez par vérifier la validité du tableau en vérifiant les nombres de degrés de liberté, les carrés moyens et les ratios.

La première hypothèse que nous devons vérifier dans le tableau concerne l'interaction. En effet, si une interaction est présente, il faut immédiatement passer aux effets simples (et procéder à des analyses complémentaires). Inutile de regarder les effets principaux (même s'ils sont automatiquement fourni par le listing du logiciel).

Autrement dit:

$H_0$ : Absence d'interaction  $A \times B$

$H_0$ : Interaction  $A \times B$

On peut rejeter  $H_0$  si  $\frac{CM_{AB}}{CM_{S|AB}} > s_{AB}(\alpha)$  dans laquelle la valeur  $\frac{CM_{AB}}{CM_{S|AB}}$  se distribue

comme une distribution de F avec les degrés de liberté  $(p - 1)(q - 1), pq(n - 1)$ , soit dans notre exemple, 1,16. La valeur critique  $s_{AB}(\alpha)$  est obtenue par la table F et est de 4.494.

Comme on le voit, le ratio  $\frac{CM_{AB}}{CM_{S|AB}}$  dans le tableau à la ligne Interaction (AB) est de 269.77, très

nettement supérieur à  $s_{AB}(\alpha)$ . L'interaction est donc significative. Autrement dit, les moyennes marginales ne reflètent absolument pas les moyennes des différentes cellules.

Nous sommes donc tenu de procéder à une décomposition. Nous choisissons (arbitrairement dans cet exemple) de décomposer suivant les niveaux de B. Nous aurons donc une analyse pour l'effet de A quand le nombre de séances est 5, et un effet de A quand le nombre de séances est 10. Le logiciel fourni le tableau suivant de la nouvelle analyse :

Analyse de variance  
Expérience Thérapie par Séances (II) : Décomposition suivant Séances

source de variation	SC	dl	CM	F
Thérapie à 5 séances (A   b <sub>1</sub> )	1.93	1	1.93	10.70
Thérapie à 10 séances (A   b <sub>2</sub> )	72.10	1	72.10	398.30
Intra groupe (S   AB) (erreur)	2.90	16	0.18	

Pour conclure, on dira qu'après 5 séances, l'effet de la thérapie par immersion produit des résultats significativement supérieurs à ceux obtenus avec la thérapie par désensibilisation ( $F(1,16) = 10.70, p < .05$ ). La performance moyenne est de 3.90 min contre 3.02.min. Toutefois, après 10 séances, l'effet s'inverse, et la thérapie par désensibilisation est significativement supérieure à la thérapie par immersion ( $F(1,16) = 398.30, p < .05$ ). L'interaction entre l'effet de la thérapie et l'effet du nombre de séance est significative ( $F(1,16) = 269.77, p < .05$ ).

On aurait pu faire l'analyse des effets simples de séances à chacun des niveaux de thérapie (B | a<sub>1</sub> et B | a<sub>2</sub>). Les résultats seront évidemment les mêmes. Cependant, dans certains cas, l'interprétation des résultats est plus facile suivant une décomposition plutôt que l'autre. Il est très important de noter que ces deux décompositions ne sont pas orthogonales. C'est pourquoi le chercheur devra choisir entre une ou l'autre analyse des effets simples.

**5.3. Plan 3 × 3 : Niveau d'association et nombre de mots sur la mémoire**

Supposons une expérience sur la mémoire verbale où un chercheur étudie les effets du niveau d'association entre les mots (bas, moyen, élevé), tel Docteur-infirmière, et la longueur de la liste de mots à mémoriser (8, 12, 16) sur le nombre de mots mémorisés par les participants. Neuf groupes indépendants sont soumis à l'expérience. Les données brutes sont:

	Longueur des listes (B)		
Association (A)	8 (b <sub>1</sub> )	12 (b <sub>2</sub> )	16 (b <sub>3</sub> )
Élevée (a <sub>1</sub> )	8, 7, 8, 8	12, 12, 11, 12	10, 10, 13, 11
Moyenne (a <sub>2</sub> )	6, 5, 6, 6	11, 9, 8, 10	12, 14, 13, 10
Basse (a <sub>3</sub> )	6, 3, 3, 5	5, 8, 6, 5	10, 12, 10, 11

suit du tableau des moyennes:

	Longueur des listes (B)			a. Moyenne par
Association (A)	8 (b <sub>1</sub> )	12 (b <sub>2</sub> )	16 (b <sub>3</sub> )	association
Élevée (a <sub>1</sub> )	7.75	11.75	11.00	10.17
Moyenne (a <sub>2</sub> )	5.75	9.50	12.25	9.17
Basse (a <sub>3</sub> )	4.25	6.00	10.75	7.00
Moyenne par longueur	5.92	9.08	11.33	8.78

Quand le plan est supérieur à 2 × 2, la formule de l'interaction moyenne devient très complexe à calculer. Il est préférable de faire un graphe. Dans ce cas-ci, on observe des effets non parallèles, suggérant une interaction. Cette table d'ANOVA synthétisant les résultats :

Analyse de variance  
Expérience Thérapie par Séances (II)

source de variation	SC	dl	CM	F
Intergroupe	274.22	8		
Association (A)	62.89	2	31.44	22.64
Longueur (B)	177.22	2	88.86	63.98
Interaction (AB)	34.11	4	8.53	6.14
Intra groupe (S   AB) (erreur)	37.50	27	1.39	
<b>Total</b>	<b>311.72</b>	<b>35</b>		

Commencez par vérifier la validité du tableau en vérifiant les nombres de degrés de liberté, les carrés moyens et les ratios.

La première hypothèse que nous devons vérifier dans le tableau concerne l'interaction (l'hypothèse implicite dans l'interaction est que l'interaction moyenne n'est pas différente de zéro). Formellement, dans l'analyse de ce plan  $3 \times 3$ , nous notons :

$$H_0 : \text{Interaction } A \times B \text{ absente}$$

$$H_1 : \text{Interaction } A \times B \text{ présente}$$

On peut rejeter  $H_0$  si  $\frac{CM_{AB}}{CM_{S|AB}} > s_{AB}(\alpha)$  dans laquelle la valeur  $\frac{CM_{AB}}{CM_{S|AB}}$  se distribue

comme une distribution de F avec les degrés de liberté  $(p - 1)(q - 1)$ ,  $p q (n - 1)$ , soit dans notre exemple, (4, 27). La valeur critique  $s_{AB}(\alpha)$  est obtenue par la table F et est de 2.728.

Comme on le voit, le ratio  $\frac{CM_{AB}}{CM_{S|AB}}$  dans le tableau à la ligne Interaction (AB) est de 6.14,

supérieur à  $s_{AB}(\alpha)$ . L'interaction est donc significative. Autrement dit, les moyennes marginales ne reflètent absolument pas les moyennes des différentes cellules. Nous choisissons (arbitrairement dans cet exemple) de décomposer suivant les niveaux de B. Nous aurons donc une analyse pour l'effet de A quand la longueur de la liste est 8, un effet de A quand la longueur de la liste est 12, et un effet de A quand la longueur de la liste est 16. Le logiciel fourni le tableau suivant de la nouvelle analyse :

Analyse de variance  
Expérience Association par Longueur de listes (II) : Décomposition suivant Longueur

source de variation	SC	dl	CM	F
Association pour 8 mots (A   b <sub>1</sub> )	24.67	2	12.33	8.88
Association pour 12 mots (A   b <sub>2</sub> )	67.17	2	33.58	24.18
Association pour 16 mots (A   b <sub>3</sub> )	5.17	2	2.58	1.86
Intra groupe (S   AB) (erreur)	37.50	27	1.39	

Dans le cas de ces décomposition, la valeur critique  $s(\alpha)$  se trouve dans la table avec (2, 27) degrés de liberté. On trouve 3.354. Pour conclure, on dira qu'après une liste de 8 mots, la force d'association a un effet significatif sur le rappel ( $F(2, 27) = 8.88$ ,  $p < .05$ ). Les mots avec une force d'association élevées sont mieux rappelés que les mots avec une faible force d'association (7.75 vs. 4.25 mots, respectivement). Il en va de même pour des listes de 12 mots ( $F(2,27) = 24.18$ ,  $p < .05$ ), où le nombre de succès est de 11.75 et 6.00 pour les mots à force d'association élevée et faible, respectivement. Pour les listes de 16 mots cependant, il n'existe plus de différence entre les forces d'association ( $F(2,27) = 1.86$ ,  $p > .05$ ). Le nombre moyen de mots rappelés pour les listes de 16 mots est de 11.33. L'interaction entre la longueur des listes et la force d'association est significative ( $F(4,27) = 6.14$ ,  $p < .05$ ).

Il faut noter qu'une ANOVA n'informe pas à savoir si, pour les listes de 8 et 12 mots, la force d'association moyenne est équivalente à la force d'association élevée, à la force d'association faible ou entre les deux. Pour avoir des informations plus fines, il faut procéder à une comparaison de moyenne.

---

## Section 6. Comparaison de moyennes

---

Une ANOVA ne permet de conclure que sur la plus grande différence. Cependant, si votre plan possède trois niveaux, on ne sait pas si le niveau intermédiaire est significativement différent des deux autres ou s'il ressemble à l'une ou l'autre des moyennes extrêmes. Souvent, notre recherche est plus intéressée dans les valeurs extrêmes, mais il peut arriver que l'on veuille en savoir plus sur la valeur intermédiaire. Si tel est le cas, il faut procéder à une comparaison de moyennes.

Le problème de la comparaison des moyennes est un problème complexe qui a suscité de nombreuses études. Le problème vient du fait que l'on ne peut pas faire un test  $t$  par couple de variables. Par exemple, pour un plan avec un seul facteur ayant 4 niveaux, il faudrait comparer  $\bar{X}_1$  et  $\bar{X}_2$ ,  $\bar{X}_1$  et  $\bar{X}_3$ ,  $\bar{X}_1$  et  $\bar{X}_4$ ,  $\bar{X}_2$  et  $\bar{X}_3$ ,  $\bar{X}_2$  et  $\bar{X}_4$ ,  $\bar{X}_3$  et  $\bar{X}_4$ , soit 6 comparaisons. Chaque test  $t$  a un risque  $\alpha$  de retourner la mauvaise décision. Si on réalise 6 comparaisons, on a beaucoup plus de chance de faire une mauvaise décision (exactement  $1 - (1 - \alpha)^6$  soit 26%), ce qui signifie que l'on court vers de gros problèmes. Pour éviter ce problème de comparaisons multiples, il faut utiliser un test subalterne à l'ANOVA qui va corriger pour ce risque accru d'erreur  $\alpha$ .

On peut distinguer plusieurs situations :

- Si l'hypothèse de recherche prévoit les tests d'au plus  $p - 1$  comparaisons (ou  $p \times q - 1$  dans le cas d'un plan factoriel à deux facteurs), il est possible d'évaluer les « contrastes » désirés au moyen de la technique dite des comparaisons a priori orthogonales.
- Si l'hypothèse de recherche prévoit plus de  $p - 1$  comparaisons ( $p \times q - 1$  dans le cas d'un plan factoriel à deux facteurs), les contrastes peuvent être évalués selon la méthode de Bonferroni, soit une méthode de comparaison a priori).
- Si les comparaisons ne sont pas prédéterminées par les hypothèses, il faut recourir aux techniques de comparaisons multiples, appelées comparaisons a posteriori. Parmi celles-ci, deux sont plus utiles, a) le test de Tukey pour comparer toutes les moyennes entre elles, et b) le test de Dunnett pour comparer des moyennes avec la moyenne d'un groupe contrôle.

Avant de faire des comparaisons a posteriori, il faut s'assurer que l'effet principal ou l'effet simple (ayant plus de 2 niveaux) est significatif. En effet, si les moyennes extrêmes ne diffèrent pas, il n'est pas utile de regarder les moyennes intermédiaires. Lors de comparaison a priori, il ne faut pas regarder les résultats de l'ANOVA. Dans ce cours, nous n'expliquons que la méthode de comparaisons de moyennes à posteriori développée par Tukey. Il en existe d'autres, et le lecteur devrait être au courant des forces et faiblesses de chacune.

**6.1. Test de Tukey**

Un test de Tukey compare des moyennes à deux niveaux, par exemple  $\bar{X}_r$  et  $\bar{X}_s$ . Il s'agit toujours d'un test bicaudal. Il est de la forme

$$\text{rejet de } H_0 \text{ si } \frac{|\bar{X}_r - \bar{X}_s|}{\sqrt{\frac{CM_e}{N}}} > s(\alpha)$$

dans lequel la valeur  $\frac{|\bar{X}_r - \bar{X}_s|}{\sqrt{\frac{CM_e}{N}}}$  est distribuée comme une Studentized Range distribution. Le

nombre de degrés de liberté de cette statistique est donné par deux valeurs, (V, dl<sub>e</sub>) (voir Table 6). La seule difficulté dans le test de Tukey est d'identifier correctement le terme d'erreur CM<sub>e</sub>, la valeur N, et le premier degré de liberté V. Ici, V représente toujours le nombre de groupes qui seront comparés (soit le nombre de niveaux de la variable).

**a. Plan à un seul facteur**

L'utilisation d'un test de Tukey avec un plan à un seul facteur est très facile. Premièrement, il n'y a qu'un seul terme d'erreur, CM<sub>S|A</sub> (contrairement aux plans factoriels à mesures répétées que nous verrons au prochain cours). La valeur V représente le nombre de niveaux de l'unique facteur A. N représente le nombre de sujets par groupe.

*a.1. Exemple*

Soit un plan à 6 groupes indépendants de 4 sujets chacun. Nous obtenons une ANOVA significative, comme le montre le tableau suivant :

Analyse de variance  
Test de Tukey à un seul facteur sur un plan *p*

source de variation	SC	dl	CM	F
Inter groupe (A)	472.00	5	94.40	5.90
Intra groupe (S AB) (erreur)	288.00	18	16.00	
<b>Total</b>	<b>760.00</b>	<b>23</b>		

Les moyennes extrêmes diffèrent donc. Les moyennes sont :

$$\bar{X}_1 = 10, \bar{X}_2 = 14, \bar{X}_3 = 16, \bar{X}_4 = 18, \bar{X}_5 = 20, \bar{X}_6 = 24$$

Le nombre de groupes V est de 6. La valeur critique au seuil de 5% pour (6, 18) degrés de libertés est de 4.49. Le dénominateur est  $\sqrt{\frac{CM_e}{N}} = \sqrt{\frac{16.00}{4}} = 2.0$  Si on manipule l'équation précédente, on trouve l'intervalle de confiance :

$$\text{rejet de } H_0 \text{ si } |\bar{X}_r - \bar{X}_s| > s(\alpha) \times \sqrt{\frac{CM_e}{N}} = 4.49 \times 2 = 8.98$$

Autrement dit, si la distance entre deux moyennes excède 8.98, elle sera significative au seuil de 5%. On présente généralement les résultats dans une matrice triangulaire semblable à celle-ci, où l'on voit la différence entre deux moyennes :

différences	$\bar{X}_2$	$\bar{X}_3$	$\bar{X}_4$	$\bar{X}_5$	$\bar{X}_6$
$\bar{X}_1$	4	6	8	10 *	14 *
$\bar{X}_2$		2	4	6	10 *
$\bar{X}_3$			2	4	8
$\bar{X}_4$				2	6
$\bar{X}_5$					4

\* :  $p < .05$

Ce que le tableau nous indique, c'est que les groupes 1 et 6 diffèrent significativement, de même que les groupes 1 et 5, et 2 et 6. Pour vous aider, faites le graphique des moyennes, et essayer de voir les résultats obtenus par le test de Tukey. Notez que nous savions que la différence entre le groupe 1 et 6 était significative puisque l'ANOVA était significative.

**b. Plan à deux facteurs**

Ici, la procédure à suivre dépend de l'interaction. Si l'interaction n'est pas significative, alors les moyennes marginales représentent les effets des deux facteurs, A et B. Il est alors permis de faire un test de Tukey soit sur les moyennes des colonnes, soit sur les moyennes des lignes ou sur les deux. Lorsqu'on analyse les moyennes du facteur A ( $\bar{X}_{1.}, \bar{X}_{2.}, \dots, \bar{X}_{p.}$ ), le nombre de moyennes comparées correspond au nombre de niveaux de la variable A, soit  $p$ . Le nombre de sujets est  $nq$ , puisque l'on fusionne les  $q$  niveaux de l'autre facteur ensemble. Finalement, le carré moyen d'erreur est obtenu par l'ANOVA directement avec ses degrés de liberté.

Lorsque l'interaction est significative, les effets simples ont été analysés (disons  $A | b_j$ ). Pour chaque effet simple significatif (par exemple,  $A | b_1$ ), un test de Tukey peut être appliqué, comparant les moyennes (c'est à dire la moyenne  $a_1 | b_1, a_2 | b_1, \dots, a_p | b_1$ , que l'on note ( $\bar{X}_{11}, \bar{X}_{21}, \dots, \bar{X}_{p1}$ )). Dans ce cas, le nombre de sujets est  $n$  (nombre de sujets par groupe). Ici, tout comme il n'est pas permis de décomposer  $A | b_j$  et  $B | a_i$ , il n'est pas permis de faire un Tukey sur les moyennes obtenues en  $A | b_j$  puis sur les moyennes obtenues en  $B | a_i$ .

6.2. Tableau récapitulatif

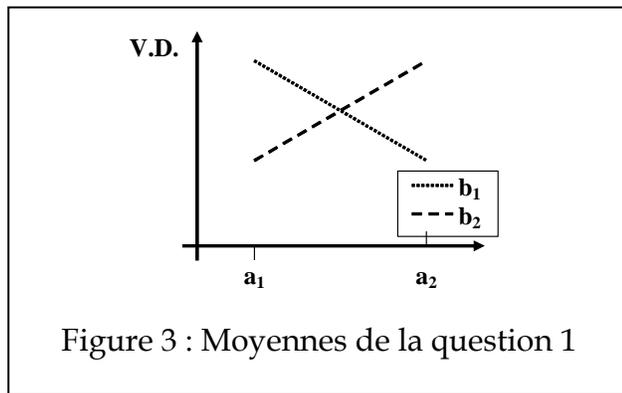
plan	interaction	sur les niveaux	$N$	$V$	$dl_e$	$CM_e$
$p / (p)$	n/a	A	$n$	$p$	$p(n-1) / (p-1)(n-1)$	$CM_{S A}$
$p \times q$	n.s.	A	$n q$	$p$	$pq(n-1)$	$CM_{S AB}$
	n.s.	B	$n p$	$q$	$pq(n-1)$	$CM_{S AB}$
	significative	A   $b_j$	$n$	$p$	$pq(n-1)$	$CM_{S AB}$
	significative	B   $a_i$	$n$	$q$	$pq(n-1)$	$CM_{S AB}$
$p \times (q)$	n.s.	A	$n q$	$p$	$p(n-1)$	$CM_{S A}$
	n.s.	B	$n p$	$q$	$p(q-1)(n-1)$	$CM_{S AB}$
	significative	A   $b_j$	$n$	$p$	$p(q-1)(n-1)$	$CM_{SA bj}$
	significative	B   $a_i$	$n$	$q$	$p(q-1)(n-1)$	$CM_{SB ai}$
$(p \times q)$	n.s.	A	$n q$	$p$	$(p-1)(n-1)$	$CM_{S A}$
	n.s.	B	$n p$	$q$	$(q-1)(n-1)$	$CM_{S B}$
	significative	A   $b_j$	$n$	$p$	$(p-1)(q-1)(n-1)$	$CM_{SA bj}$
	significative	B   $a_i$	$n$	$q$	$(p-1)(q-1)(n-1)$	$CM_{SB ai}$

Note :  $n$  représente le nombre de sujets par groupe ou la moyenne harmonique du nombre de sujets dans le cas de groupes inégaux.

Section 7. Conclusion

*Exercices*

1. Soit ce graphique des moyennes de la Figure 3 obtenues dans une expérience  $2 \times 2$  à 4 groupes indépendants dans laquelle l'axe des Y représente une mesure du Q. I. et l'axe des X le sexe des participants ( $a_1$  = homme et  $a_2$  = femme). Les chercheurs ont étudié l'effet de la peur ( $b_1$  dans une pièce calme,  $b_2$  des terroristes entrent avec des mitraillettes). Supposant que les différences sont toutes significatives, quelle sorte d'interprétation pourriez-vous écrire?



2. Vrai ou faux : Il faut toujours faire un graphique des moyennes lors d'une ANOVA?
3. Peut-on faire un Tukey si les groupes ont un nombre inégal de sujets?
4. Puisque l'ANOVA est basée sur l'analyse de variance, est-il préférable d'avoir des groupes où les sujets varient beaucoup entre eux (d'où une variabilité élevée)?