

# Analyse de la variance pour plans à mesures répétées

Frédéric Bertrand<sup>1</sup> & Myriam Maumy<sup>1</sup>

<sup>1</sup>IRMA, Université de Strasbourg  
Strasbourg, France

DUS2  
20-06-2011

## Référence

Ce cours s'appuie essentiellement sur

- 1 le livre de David C. Howell **Méthodes statistiques** en sciences humaines, traduit de la sixième édition américaine, aux éditions de Boeck, 2008.
- 2 le livre de Kutner, Nachtsheim, Neter, Li **Applied Linear Statistical Models**, Fifth Edition, aux éditions McGraw-Hill Irwin, 2004

# Sommaire

- 1 Introduction
  - Exemple
  - Définition
  - Équation de l'ANOVA
- 2 Répétitions suivant le facteur de l'ANOVA
- 3 Répétitions suivant un des deux facteurs de l'ANOVA
- 4 Répétitions suivant les deux facteurs de l'ANOVA

# Sommaire

- 1 Introduction
  - Exemple
  - Définition
  - Équation de l'ANOVA
- 2 Répétitions suivant le facteur de l'ANOVA
- 3 Répétitions suivant un des deux facteurs de l'ANOVA
- 4 Répétitions suivant les deux facteurs de l'ANOVA

## Introduction

Lorsque nous mesurons plusieurs fois de suite, par exemple à des instants différents, la même grandeur sur les mêmes sujets, nous sommes confrontés à certains problèmes. En effet, nous ne pouvons plus utiliser l'analyse de la variance classique, car

- 1 la condition d'indépendance n'est plus vérifiée,
- 2 il faut prendre en compte la variabilité « Inter-sujets ».

Une solution est alors d'utiliser l'ANOVA pour plans à mesures répétées.

## Exemple

Nous mesurons le QI dans trois situations distinctes, les sujets ont passé le test dans chacune des trois situations. Comment pouvons-nous modéliser une telle expérience ?

- Nous pourrions dire qu'il y a trois variables « QI » perdant ainsi le lien entre les trois situations.
- Nous pourrions dire qu'il y a une seule variable « QI » et une variable « Situation », mais il faudra tenir compte que chacun des sujets ont passé le test dans les trois situations.

Cette expérience va se modéliser par un certain type de plan que nous appelons un **plan à mesures répétées**.

## ANOVA pour plans à mesures répétées

Nous utilisons souvent l'ANOVA pour plans à mesures répétées, lorsque nous mesurons plusieurs fois une même grandeur pour en percevoir l'évolution au cours du temps - ou dans diverses situations, pour chaque sujet.

## Nouvelle modélisation

Nous allons utiliser une description incorrecte mathématiquement, mais facile à comprendre en parlant de variable « Intra-sujets » et de variable « Inter-sujets ».

## Une nouvelle décomposition des sommes de carrés

Variation totale ( $ddl = IS - 1$ ) = Variation « Inter-sujets »  
 ( $ddl = S - 1$ ) + Variation « Intra-sujets » ( $ddl = S(I - 1)$ )  
 et

Variation « Intra-sujets » ( $ddl = S(I - 1)$ ) = Variation « Facteur »  
 ( $ddl = (I - 1)$ ) + Variation « Erreur » ( $ddl = (S - 1)(I - 1)$ ).



# Sommaire

- 1 Introduction
  - Exemple
  - Définition
  - Équation de l'ANOVA
- 2 Répétitions suivant le facteur de l'ANOVA
- 3 Répétitions suivant un des deux facteurs de l'ANOVA
- 4 Répétitions suivant les deux facteurs de l'ANOVA

## Exemple issu du livre de Howell

Nous étudions l'efficacité des techniques de relaxation pour contrôler les maux de tête.

Dans le cadre de cette expérience, nous avons recruté 9 personnes souffrant de maux de tête et leur avons demandé d'enregistrer la fréquence et la durée de leurs migraines.

Aux 4 semaines de relevés de base, sans aucun exercice, a succédé une période de 6 semaines consacrée à la relaxothérapie.

Pour notre exemple, nous analyserons les données correspondant aux 2 dernières semaines de relevés de base et aux 3 dernières semaines d'exercices.

La variable dépendante est la durée (heures/semaines) des migraines durant chacune de ces 5 semaines.

## Tableau des données

Sujet	Sem. 1	Sem. 2	Sem. 3	Sem. 4	Sem. 5
1	21	22	8	6	6
2	20	19	10	4	4
3	17	15	5	4	5
4	25	30	13	12	17
5	30	27	13	8	6
6	19	27	8	7	4
7	26	16	5	2	5
8	17	18	8	1	5
9	26	24	14	8	9

## Le modèle

Le modèle statistique s'écrit de la façon suivante :

$$Y_{is} = \mu + \alpha_i + \tau_s + \varepsilon_{is}$$

où  $i = 1, \dots, I$ ,  $s = 1, \dots, S$ ,

avec la contrainte supplémentaire :

$$\sum_{i=1}^I \alpha_i = 0$$

où  $Y_{is}$  est la valeur prise par la réponse  $Y$  dans les conditions  $(\alpha_i, \tau_s)$ .

Nous notons  $n = I \times S$  le nombre total de mesures ayant été effectuées.

## Contexte

- Un facteur contrôlé  $\alpha$  se présente sous  $I$  modalités, chacune d'entre elles étant notée  $\alpha_j$ .
- Les termes  $\tau_S$  représentent un échantillon de taille  $S$  prélevé dans une population importante. Nous admettrons que les effets des  $\tau_S$  sont distribués suivant une loi normale centrée de variance  $\sigma_\tau^2$ .

## Conditions liées à ce type d'analyse

Nous supposons que

- les effets aléatoires  $\tau_S$  sont indépendants
- les effets aléatoires  $\tau_S$  ont même variance  $\sigma_\tau^2$  inconnue
- les effets aléatoires  $\tau_S$  sont de loi gaussienne.

## Conditions classiques de l'ANOVA

Nous postulons les hypothèses classiques de l'ANOVA pour les variables erreurs  $\mathcal{E}_{is}$  :

- les erreurs sont indépendantes
- les erreurs ont même variance  $\sigma^2$  inconnue
- les erreurs sont de loi gaussienne.

## Ajout d'une condition

Nous ajoutons l'hypothèse d'indépendance entre les effets aléatoires et les erreurs :

- les effets aléatoires  $\tau_s$  et les erreurs  $\mathcal{E}_{is}$  sont indépendants, pour tout  $i$  variant de 1 à  $I$  et pour tout  $s$  variant de 1 à  $S$ .

## Relation fondamentale de l'ANOVA

Nous supposons que les conditions d'utilisation de ce modèle sont bien remplies.

Nous utilisons les quantités  $SC_{\alpha}$ ,  $SC_{\tau}$ ,  $SC_R$ ,  $SC_{TOT}$  déjà introduites au chapitre précédent.

Nous rappelons la relation fondamentale de l'ANOVA :

$$SC_{TOT} = SC_{\alpha} + SC_{\tau} + SC_R$$

ou encore qui peut s'écrire dans ce contexte :

$$SC_{TOT} = SC_{Inter-sujets} + SC_{Intra-sujets}.$$



## Tableau de l'ANOVA

Variation	$SC$	$ddl$	$CM$	$F_{obs}$	$F_c$
Due au facteur $\alpha$	$sc_\alpha$	$l - 1$	$cm_\alpha$	$\frac{cm_\alpha}{cm_R}$	$c$
Due au facteur $\tau$	$sc_\tau$	$S - 1$	$cm_\tau$	$\frac{cm_\tau}{cm_R}$	$c$
Résiduelle	$sc_R$	$(l - 1)(S - 1)$	$cm_R$		
Totale	$sc_{TOT}$	$IS - 1$			

## Tests d'hypothèses

L'analyse de la variance pour plans à mesures répétées permet deux tests de Fisher.

## Premier test

Nous testons l'hypothèse nulle

$$(\mathcal{H}_0) : \alpha_1 = \alpha_2 = \dots = \alpha_I = 0$$

contre l'hypothèse alternative

$$(\mathcal{H}_1) : \text{Il existe } i_0 \in \{1, 2, \dots, I\} \text{ tel que } \alpha_{i_0} \neq 0.$$

Sous l'hypothèse nulle ( $\mathcal{H}_0$ ) précédente d'absence d'effet du facteur  $\alpha$  et lorsque les conditions de validité du modèle sont respectées,  $F_{\alpha, obs}$  est la réalisation d'une variable aléatoire qui suit une loi de Fisher à  $I - 1$  et  $(I - 1)(S - 1)$  degrés de liberté.

## Premier test - Suite

Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  du test, ou à l'aide d'une table, rejet si la valeur  $F_{\alpha,obs}$  est supérieure ou égale à la valeur critique issue de la table de Fisher.

Lorsque l'hypothèse nulle ( $\mathcal{H}_0$ ) est rejetée, nous pouvons procéder à des tests de comparaisons multiples des différents effets des niveaux du facteur  $\alpha$ .

## Deuxième test

Nous testons l'hypothèse nulle

$$(\mathcal{H}_0) : \sigma_\tau^2 = 0$$

contre l'hypothèse alternative

$$(\mathcal{H}_1) : \sigma_\tau^2 \neq 0.$$

Sous l'hypothèse nulle ( $\mathcal{H}_0$ ) précédente d'absence d'effet du facteur  $\tau$  et lorsque les conditions de validité du modèle sont respectées,  $F_{\tau,obs}$  est la réalisation d'une variable aléatoire qui suit une loi de Fisher à  $S - 1$  et  $(I - 1)(S - 1)$  degrés de liberté.

## Deuxième test - Suite

Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  du test, ou à l'aide d'une table, rejet si la valeur  $F_{\tau,obs}$  est supérieure ou égale à la valeur critique issue de la table de Fisher.

## Remarque

En fonction du contexte et de la problématique posée, vous pourrez être amené à vous en servir.

## Retour à l'exemple - Sortie avec MINITAB

Analyse de la variance pour Relevés, avec utilisation de la somme des carrés ajustée pour les tests

Source	DL	SomCar séq	CM ajust	F	P
Sujet	8	486,71	60,84	8,45	0,000
Semaine	4	2449,20	612,30	85,04	0,000
Erreur	32	230,40	7,20		
Total	44	3166,31			

S = 2,68328

R carré = 92,72 % R carré (ajust) = 89,99 %

## Procédure de vérification des conditions d'application

- 1 **L'additivité du modèle.** Il est conseillé de tracer le nuage de points de la réponse en fonction du facteur, et ce pour chaque sujet. Ce graphique permet de détecter si un terme d'interaction doit être ajouté ou pas.
- 2 **L'indépendance du facteur  $\tau$ .** Il n'existe pas de test. Il faut simplement s'assurer que tous les sujets sont tous indépendants.
- 3 **La normalité du facteur  $\tau$ .** Nous pouvons la vérifier en calculant les quantités correspondantes. Pour cela, il faut calculer et stocker les coefficients du modèle.
- 4 **L'homogénéité des variances du facteur  $\tau$ .** Nous ne pouvons pas évaluer cette hypothèse en l'absence de répétitions.



## Vérification des conditions

- 5 **L'indépendance des erreurs.** Il n'existe toujours pas de test.
- 6 **La normalité des erreurs.** Nous utiliserons la même procédure et donc le même test que dans les chapitres précédents.
- 7 **L'homogénéité des variances des erreurs.** Nous pouvons la tester par rapport au facteur à effets fixes et nous nous contentons de l'évaluer graphiquement pour le facteur à effets aléatoires.

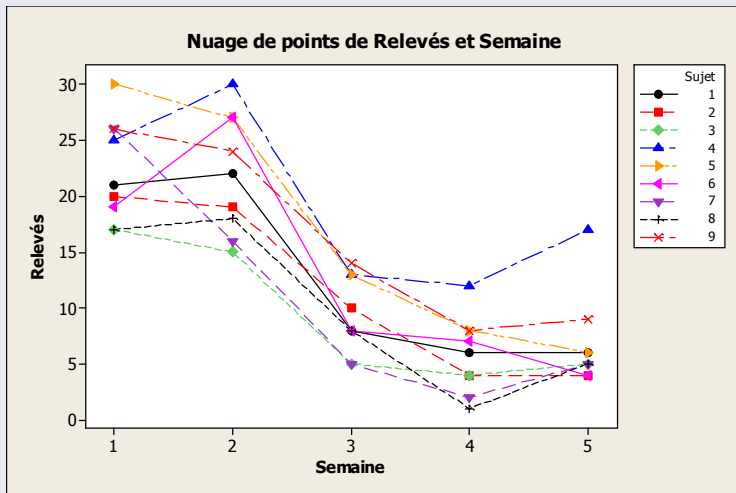
## Retour à l'exemple

- 1 **L'additivité du modèle.** Voir le graphique 1.
- 2 **L'indépendance du facteur  $\tau$ .** D'après Howell, les 9 sujets sont indépendants.
- 3 **La normalité du facteur  $\tau$ .** Dans le cas présent, nous ne le ferons pas, car le nombre de sujets est insuffisant.
- 4 **L'homogénéité des variances du facteur  $\tau$ .** Nous ne pouvons pas évaluer cette hypothèse en l'absence de répétitions au sein de chacun des patients.

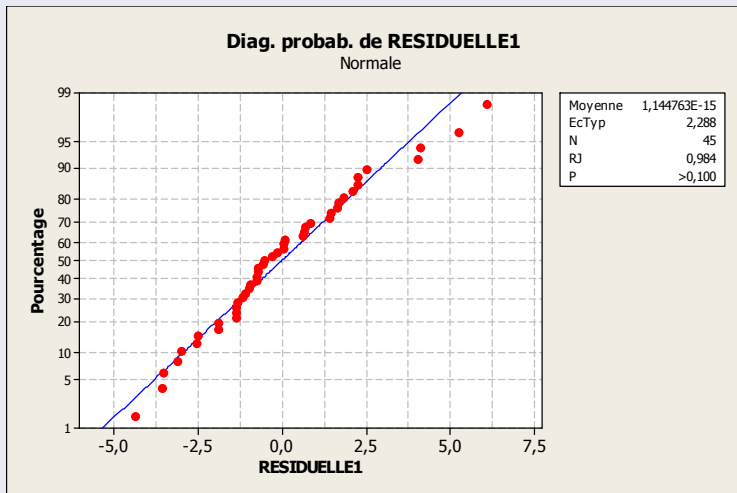
## Retour à l'exemple - Suite et fin

- 5 **L'indépendance des erreurs.**
- 6 **La normalité des erreurs.** Voir le graphique 2.
- 7 **L'homogénéité des variances des erreurs.** Voir les graphiques 3 et 4.

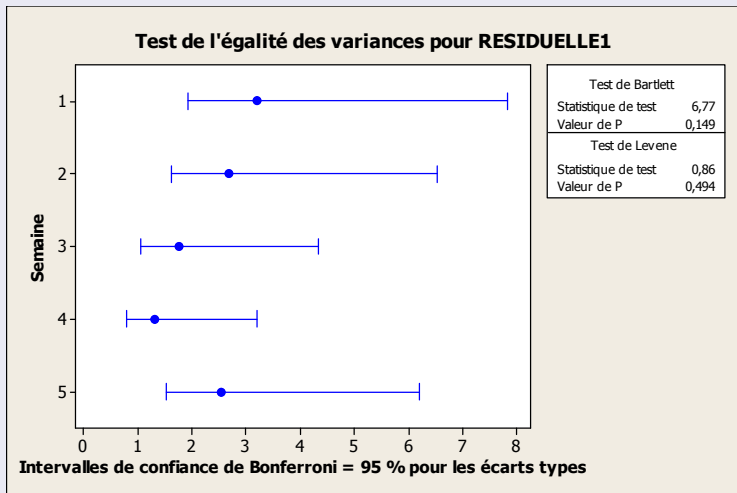
## Graphique 1



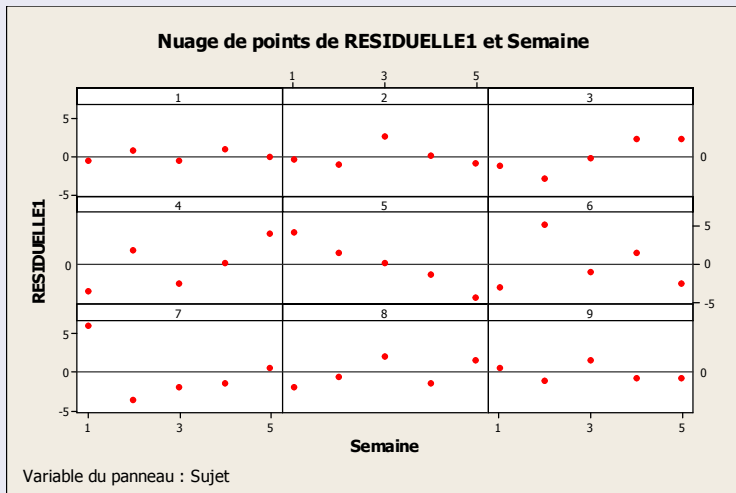
## Graphique 2



## Graphique 3



## Graphique 4



## Analyse des résultats

Afin d'étudier les effets de la relaxation sur la durée des migraines, 9 participants ont évalué la durée de leur migraine pendant deux semaines consécutives avant de suivre une thérapie de relaxation et puis pendant trois semaines au cours de leur thérapie.

Une analyse de la variance globale avec mesures répétées révèle la présence d'une différence significative d'une semaine à l'autre ( $F(4,32)=85,04$ ,  $P=0,000$ ).

La durée moyenne pendant les semaines de relevés de base est de 22,166 et descend à une moyenne de 7,296 pendant les exercices, ce qui correspond à une différence de 14,870. Nous voudrions maintenant savoir où résident ces différences statistiquement significatives.



## Contrastes linéaires

Pour répondre à la dernière question, nous allons utiliser la méthode des contrastes linéaires. Nous allons contraster les moyennes de l'ensemble des relevés de base avec celles de l'ensemble des semaines d'exercices. Les coefficients qui permettent de faire cela sont donnés dans le tableau ci-dessous :

	Sem. 1	Sem. 2	Sem. 3	Sem. 4	Sem. 5
Coefficient	1/2	1/2	-1/3	-1/3	-1/3
Moyenne	22,333	22,000	9,333	5,778	6,778

## Réalisation pratique du test des contrastes linéaires

$$\begin{aligned}
 L_{obs} &= \left(\frac{1}{2} \times 22,333\right) + \left(\frac{1}{2} \times 22,000\right) - \left(\frac{1}{3} \times 9,333\right) \\
 &\quad - \left(\frac{1}{3} \times 5,778\right) - \left(\frac{1}{3} \times 6,778\right) \\
 &= 14,870.
 \end{aligned}$$

Calculons la statistique  $t$  du test :

$$t_{obs} = \frac{L_{obs}}{s_{L_{obs}}} = \frac{14,870}{\sqrt{7,20 \left(\frac{0,833}{9}\right)}} = 18,21 > 2,0369.$$

Le contraste est significatif, au seuil  $\alpha = 5\%$ .

# Sommaire

- 1 Introduction
  - Exemple
  - Définition
  - Équation de l'ANOVA
- 2 Répétitions suivant le facteur de l'ANOVA
- 3 Répétitions suivant un des deux facteurs de l'ANOVA
- 4 Répétitions suivant les deux facteurs de l'ANOVA

## Exemple issu du livre de Kutner, Nachtsheim, Neter et Li

Une grande chaîne de distribution nationale souhaite étudier l'effet de deux campagnes publicitaires (facteur  $A$ ) sur l'évolution au cours du temps (facteur  $B$ ) du volume des ventes de paires de chaussures d'athlétisme.

10 supermarchés tests (sujet  $S$ ) ayant des caractéristiques semblables ont été choisis au hasard pour être inclus dans cette étude.

Les deux campagnes publicitaires ( $A_1$  et  $A_2$ ) sont semblables à tout point de vue si ce n'est l'image du sportif associé à la campagne. Les volumes de vente ont été relevés à 3 reprises pendant des périodes de 2 semaines ( $B_1 = 2$  semaines avant le début de la campagne,  $B_2 = 2$  semaines pendant la campagne,  $B_3 = 2$  semaines après la fin de la campagne.)

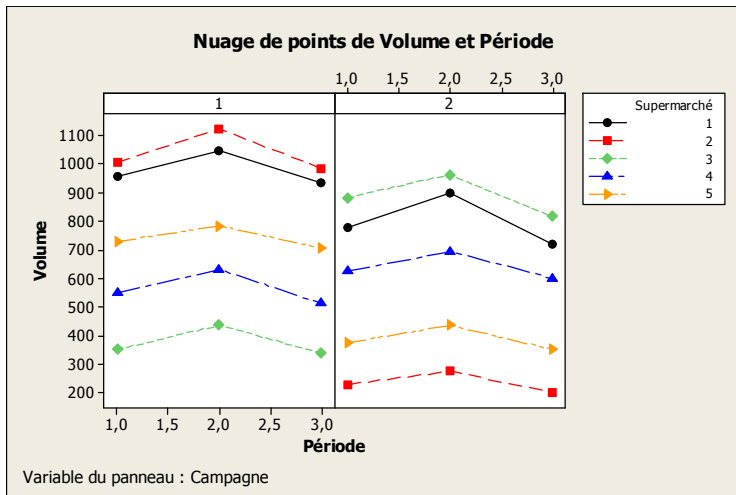
## Exemple issu du livre de Kutner, Nachtsheim, Neter et Li - Suite et fin

L'expérience a été menée pendant une période de 6 semaines lors de laquelle le volume des ventes des paires de chaussures d'athlétisme est considéré comme stable.

Les données sont reproduites dans le tableau qui va suivre. Elles sont représentées dans la figure ci-dessous. Il ne semble pas y avoir besoin d'inclure d'interaction entre les supermarchés tests et les campagnes publicitaires.

## Tableau des données

Campagne	Supermarché	Avant	Pendant	Après
1	1	958	1 047	933
1	2	1 005	1 122	986
1	3	351	436	339
1	4	549	632	512
1	5	730	784	707
2	1	780	897	718
2	2	229	275	202
2	3	883	964	817
2	4	624	695	599
2	5	375	436	351



## Le modèle

Le modèle statistique s'écrit de la façon suivante :

$$Y_{isj} = \mu + \alpha_i + \tau_{s(i)} + \beta_j + (\alpha\beta)_{ij} + \mathcal{E}_{isj}$$

avec  $i = 1, \dots, I, s = 1, \dots, S, j = 1, \dots, J$

avec les contraintes supplémentaires :

$$\sum_{i=1}^I \alpha_i = 0, \sum_{j=1}^J \beta_j = 0, \sum_{i=1}^I (\alpha\beta)_{ij} = 0, \sum_{j=1}^J (\alpha\beta)_{ij} = 0 \text{ pour tout } i, j$$

et où  $Y_{isj}$  est la valeur prise par la réponse  $Y$  dans les conditions  $(\alpha_i, \beta_j)$  pour le  $s$ -ème sujet.

Nous notons  $n = I \times S \times J$  le nombre total de mesures ayant été effectuées.



## Contexte

- Un facteur contrôlé  $\alpha$  se présente sous  $I$  modalités, chacune d'entre elles étant notée  $\alpha_j$ .
- Les  $\tau_{S(i)}$  représentent un échantillon de taille  $S$  prélevé dans une population importante. Nous admettrons que les effets des  $\tau_{S(i)}$  sont distribués suivant une loi normale centrée de variance  $\sigma_{\tau|\alpha}^2$ .
- Un facteur contrôlé  $\beta$  se présente sous  $J$  modalités, chacune d'entre elles étant notée  $\beta_j$ .

## Conditions liées à ce type d'analyse

Nous supposons que

- les effets aléatoires  $\tau_{S(i)}$  sont indépendants
- $\mathcal{L}(\tau_{S(i)}) = \mathcal{N}(0, \sigma_{\tau|\alpha}^2)$ , pour tout  $s$  variant de 1 à  $S$ .

## Conditions classiques de l'ANOVA

Nous postulons les hypothèses classiques de l'ANOVA pour les variables erreurs  $\mathcal{E}_{isj}$  :

- les erreurs sont indépendantes
- les erreurs ont même variance  $\sigma^2$  inconnue
- les erreurs sont de loi gaussienne.

## Ajout d'une condition

Nous ajoutons l'indépendance des effets aléatoires  $\tau_{S(i)}$  et des erreurs  $\mathcal{E}_{isj}$ .

## Relation fondamentale de l'ANOVA

Nous supposons que les conditions d'utilisation de ce modèle sont bien remplies.

Nous utilisons les quantités  $SC_{\alpha}$ ,  $SC_{\tau|\alpha}$ ,  $SC_{\beta}$ ,  $SC_{\alpha\beta}$ ,  $SC_R$ ,  $SC_{TOT}$  déjà introduites au chapitre précédent.

Nous rappelons la relation fondamentale de l'ANOVA :

$$SC_{TOT} = SC_{\alpha} + SC_{\tau|\alpha} + SC_{\beta} + SC_{\alpha\beta} + SC_R.$$

## Tableau de l'ANOVA

Variation	$SC$	$ddl$	$CM$	$F_{obs}$	$F_c$
Due au fact. $\alpha$	$sc_\alpha$	$I - 1$	$cm_\alpha$	$\frac{cm_\alpha}{cm_{\tau \alpha}}$	$c$
Due au fact. $\tau$ dans $\alpha$	$sc_{\tau \alpha}$	$I(S - 1)$	$cm_{\tau \alpha}$	$\frac{cm_{\tau \alpha}}{cm_R}$	$c$
Due au fact. $\beta$	$sc_\beta$	$J - 1$	$cm_\beta$	$\frac{cm_\beta}{cm_R}$	$c$
Due à l'inter.	$sc_{\alpha\beta}$	$(I - 1)(J - 1)$	$cm_{\alpha\beta}$	$\frac{cm_{\alpha\beta}}{cm_R}$	$c$
Résiduelle	$sc_R$	$I(S - 1)(J - 1)$	$cm_R$		
Totale	$sc_{TOT}$	$n - 1$			

## Tests d'hypothèses

L'analyse de la variance à 2 facteurs avec mesures répétées sur l'un des deux facteurs permet quatre tests de Fisher.

## Premier test

Nous testons l'hypothèse nulle

$$(\mathcal{H}_0) : \alpha_1 = \alpha_2 = \dots = \alpha_I = 0$$

contre l'hypothèse alternative

$$(\mathcal{H}_1) : \text{Il existe } i_0 \in \{1, 2, \dots, I\} \text{ tel que } \alpha_{i_0} \neq 0.$$

Sous l'hypothèse nulle ( $\mathcal{H}_0$ ) précédente d'absence d'effet du facteur  $\alpha$  et lorsque les conditions de validité du modèle sont respectées,  $F_{\alpha, obs}$  est la réalisation d'une variable aléatoire qui suit une loi de Fisher à  $I - 1$  et  $I(S - 1)$  degrés de liberté.

## Premier test - Suite

Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  du test, ou à l'aide d'une table, rejet si la valeur  $F_{\alpha, obs}$  est supérieure ou égale à la valeur critique issue de la table de Fisher.

Lorsque l'hypothèse nulle ( $\mathcal{H}_0$ ) est rejetée, nous pouvons procéder à des tests de comparaisons multiples des différents effets des niveaux du facteur  $\alpha$ .



## Deuxième test

Nous testons l'hypothèse nulle

$$(\mathcal{H}_0) : \sigma_{\tau|\alpha}^2 = 0$$

contre l'hypothèse alternative

$$(\mathcal{H}_1) : \sigma_{\tau|\alpha}^2 \neq 0.$$

Sous l'hypothèse nulle ( $\mathcal{H}_0$ ) précédente d'absence d'effet du facteur  $\tau$  dans le facteur  $\alpha$  et lorsque les conditions de validité du modèle sont respectées,  $F_{\tau|\alpha,obs}$  est la réalisation d'une variable aléatoire qui suit une loi de Fisher à  $I(S - 1)$  et  $I(S - 1)(J - 1)$  degrés de liberté.

## Remarque

En fonction du contexte et de la problématique posée, vous pourrez être amené à vous en servir.

## Troisième test

Nous testons l'hypothèse nulle

$$(\mathcal{H}_0) : \beta_1 = \beta_2 = \dots = \beta_J = 0$$

contre l'hypothèse alternative

$$(\mathcal{H}_1) : \text{Il existe } j_0 \in \{1, 2, \dots, J\} \text{ tel que } \beta_{j_0} \neq 0.$$

Sous l'hypothèse nulle ( $\mathcal{H}_0$ ) précédente d'absence d'effet du facteur  $\beta$  et lorsque les conditions de validité du modèle sont respectées,  $F_{\beta, obs}$  est la réalisation d'une variable aléatoire qui suit une loi de Fisher à  $J - 1$  et  $I(S - 1)(J - 1)$  degrés de liberté.

## Troisième test - Suite

Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  du test, ou à l'aide d'une table, rejet si la valeur  $F_{\beta,obs}$  est supérieure ou égale à la valeur critique issue de la table de Fisher.

Lorsque l'hypothèse nulle ( $\mathcal{H}_0$ ) est rejetée, nous pouvons procéder à des tests de comparaisons multiples des différents effets des niveaux du facteur  $\beta$ .

## Quatrième test

Nous testons l'hypothèse nulle

$$(\mathcal{H}_0) : (\alpha\beta)_{11} = \dots = (\alpha\beta)_{IJ} = 0$$

contre l'hypothèse alternative

$$(\mathcal{H}_1) : \text{Il existe } (i_0, j_0) \in \{1, \dots, I\} \times \{1, \dots, J\} \text{ tq } (\alpha\beta)_{i_0 j_0} \neq 0.$$

Sous l'hypothèse nulle ( $\mathcal{H}_0$ ) précédente d'absence d'effet du facteur  $\alpha\beta$  et lorsque les conditions de validité du modèle sont respectées,  $F_{\alpha\beta, obs}$  est la réalisation d'une variable aléatoire qui suit une loi de Fisher à  $(I - 1)(J - 1)$  et  $I(S - 1)(J - 1)$  degrés de liberté.

## Retour à l'exemple - Sortie avec MINITAB

Analyse de la variance pour Volume, avec utilisation de la somme des carrés ajustée pour les tests

Sour.	DL	SC séq	CM ajust	F	P
Cam.	1	168151	168151	0,73	0,417
Sup.					
(Cam.)	8	1833681	229210	640,31	0,000
Pér.	2	67073	33537	93,69	0,000
C.*P.	2	391	196	0,55	0,589
Err.	16	5727	358		
Tot.	29	2075023			

S = 18,9200

R carré = 99,72% R carré (ajust) = 99,50 %

## Procédure de vérification des conditions d'application - Retour à l'exemple

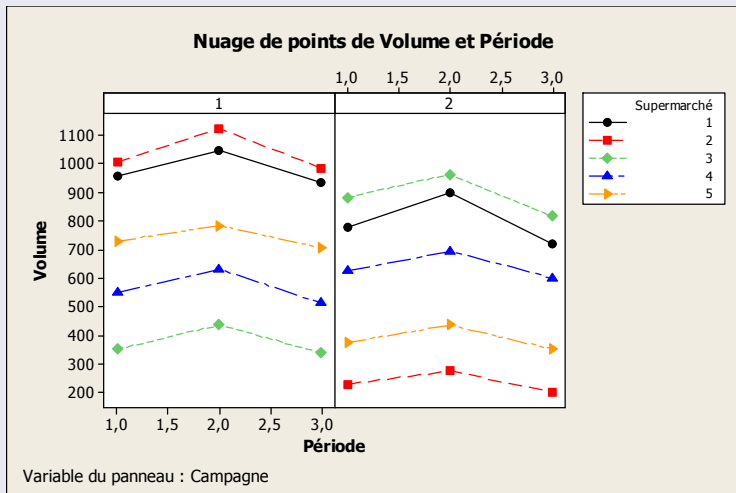
- 1 **L'additivité du modèle.** Voir le graphique 1.
- 2 **L'indépendance du facteur « Supermarché ».** D'après Kutner et al., les 10 supermarchés sont indépendants.
- 3 **La normalité du facteur « Supermarché ».** Dans le cas présent, nous ne le ferons pas, car le nombre de supermarchés est insuffisant.
- 4 **L'homogénéité des variances du facteur « Supermarché ».**

## Retour à l'exemple - Suite et fin

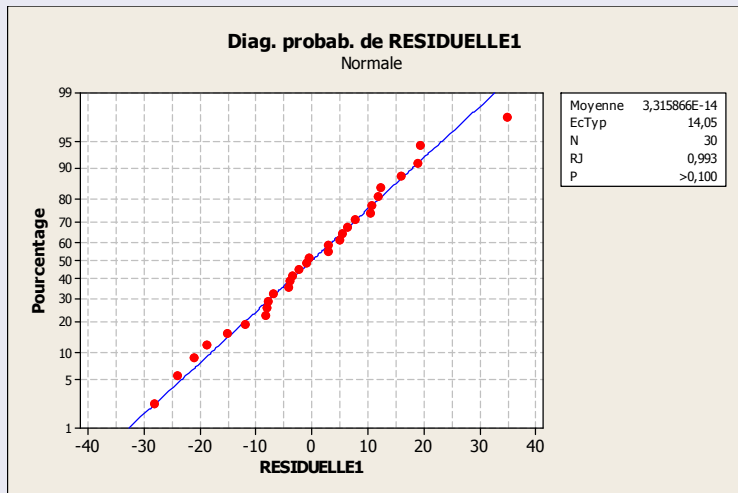
- 5 **L'indépendance des erreurs.**
- 6 **La normalité des erreurs.** Voir le graphique 2.
- 7 **L'homogénéité des variances des erreurs.** Voir les graphiques 3 et 4.



## Graphique 1



## Graphique 2



## Graphique 3

## Test de l'égalité des variances pour RESIDUELLE1

Campagne

Période

1

1

2

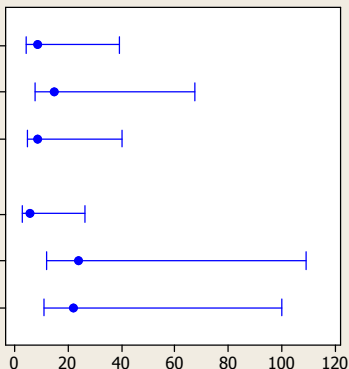
3

2

1

2

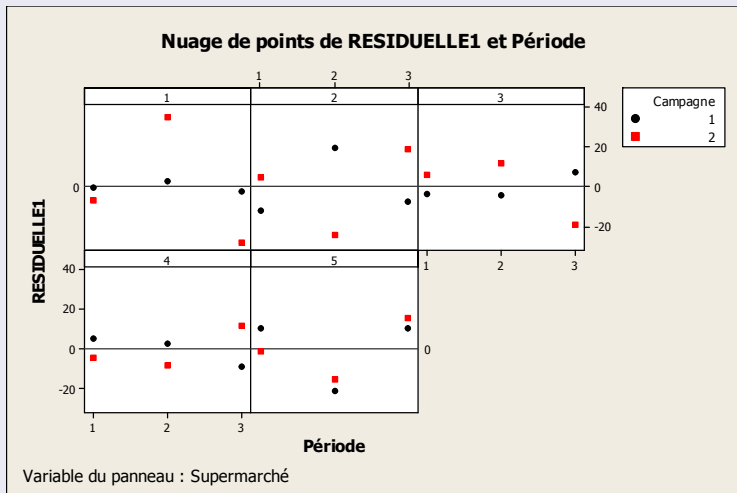
3



Test de Bartlett	
Statistique de test	10,45
Valeur de P	0,063
Test de Levene	
Statistique de test	1,17
Valeur de P	0,353

Intervalles de confiance de Bonferroni = 95 % pour les écarts types

## Graphique 4



## Analyse des résultats

- Pour le premier test, P-value = 0,417, nous décidons de ne pas refuser l'hypothèse nulle ( $\mathcal{H}_0$ ). Par conséquent, nous n'avons pas réussi à mettre en évidence d'effet du facteur fixe « Campagne ». Le risque associé à cette décision est un risque de deuxième espèce. Pour l'évaluer, il resterait à calculer la puissance de ce test.
- Pour le troisième test, P-value = 0,000, nous décidons de refuser l'hypothèse nulle ( $\mathcal{H}_0$ ). Par conséquent, nous pouvons dire, au seuil  $\alpha = 5\%$ , qu'il y a un effet significatif du facteur fixe « Période ».

## Analyse des résultats

- Pour le quatrième test, P-value = 0,589, nous décidons de ne pas refuser l'hypothèse nulle ( $\mathcal{H}_0$ ). Par conséquent, nous n'avons pas réussi à mettre en évidence d'effet du facteur fixe « Interaction ». Le risque associé à cette décision est un risque de deuxième espèce. Pour l'évaluer, il resterait à calculer la puissance de ce test.

## Remarque

Nous pourrions envisager de faire des tests de comparaisons multiples sur le facteur « Période ». Malheureusement, avec MINITAB, ce ne sera pas possible, à cause de la présence du facteur aléatoire. Par contre, nous pourrions envisager de faire les tests de comparaisons multiples « à la main » ou d'utiliser un autre logiciel.

## Test de comparaisons multiples

La procédure de Tukey est utilisée pour réaliser ce test.

$$c = \frac{1}{\sqrt{2}} q_{\overline{T}}(0,99; 3, 16) = \frac{1}{\sqrt{2}}(4,78) = 3,38.$$

et

$$\frac{s_R^2}{2} \left( \frac{1}{n'_i} + \frac{1}{n_i} \right) = \frac{358}{2} \left( \frac{1}{5} + \frac{1}{5} \right) = 71,60.$$

D'où  $c \times \sqrt{\frac{s_R^2}{2} \left( \frac{1}{n'_i} + \frac{1}{n_i} \right)} = 3,38 \sqrt{71,60} = 28,60$ , en se

souvenant que le carré moyen résiduel observé,  $cm_R$ , est également noté  $s_R^2$  car c'est une estimation sans biais de la variance des erreurs  $\sigma^2$ .

## Calculs des différences entre les modalités du facteur période

$$\bar{y}_{..2} - \bar{y}_{..1} = 728,80 - 648,40 = 80,40$$

$$\bar{y}_{..3} - \bar{y}_{..1} = 616,40 - 648,40 = -32,00$$

et

$$\bar{y}_{..3} - \bar{y}_{..2} = 616,40 - 728,80 = -112,40.$$



## Calculs des intervalles de confiance

$$52 \leq \mu_{..2} - \mu_{..1} \leq 109$$

$$-61 \leq \mu_{..3} - \mu_{..1} \leq -3$$

et

$$-141 \leq \mu_{..3} - \mu_{..2} \leq -84.$$

## Conclusion

En utilisant des intervalles de confiance ayant un taux de confiance simultané de 99%, nous concluons que :

- 1 pendant la campagne publicitaire, il y a une augmentation immédiate des moyennes des ventes comprise entre 52 et 109 (8% à 17%)
- 2 après la campagne publicitaire, il y a une diminution des moyennes des ventes comprise entre 3 et 61 (0,5% et 9%).

# Sommaire

- 1 Introduction
  - Exemple
  - Définition
  - Équation de l'ANOVA
- 2 Répétitions suivant le facteur de l'ANOVA
- 3 Répétitions suivant un des deux facteurs de l'ANOVA
- 4 Répétitions suivant les deux facteurs de l'ANOVA

## Exemple issu du livre de Kutner, Nachtsheim, Neter et Li

Un clinicien étudie les effets de deux drogues utilisées indépendamment ou simultanément sur la circulation sanguine chez des sujets humains. 12 sujets masculins en bonne santé et à l'âge adulte ont été choisis au hasard pour participer à cette étude. Les 4 traitements utilisés dans cette étude sont définis de la façon suivante :

- $A_1 B_1$  un placebo (aucune drogue)
- $A_1 B_2$  uniquement la drogue  $B$
- $A_2 B_1$  uniquement la drogue  $A$
- $A_2 B_2$  les 2 drogues simultanément.

## Exemple issu du livre de Kutner, Nachtsheim, Neter et Li - Suite et fin

Les 12 sujets reçoivent chacun les 4 traitements dans des ordres indépendamment aléatoires. La variable réponse est l'augmentation dans la circulation sanguine entre avant et juste après l'administration du traitement. Les traitements sont administrés jours après jours.

## Tableau des données

Sujet	$A_1B_1$	$A_1B_2$	$A_2B_1$	$A_2B_2$
1	2	10	9	25
2	-1	8	6	21
3	0	11	8	24
...	...	...	...	...
10	-2	10	10	28
11	2	8	10	25
12	-1	8	6	23

## Le modèle

Le modèle statistique s'écrit de la façon suivante :

$$Y_{ijs} = \mu + \alpha_i + \beta_j + \tau_s + (\alpha\beta)_{ij} + (\alpha\tau)_{is} + (\beta\tau)_{js} + \varepsilon_{ijs}$$

avec  $i = 1, \dots, I, j = 1, \dots, J, s = 1, \dots, S$

avec les contraintes supplémentaires :

$$\sum_{i=1}^I \alpha_i = 0, \sum_{j=1}^J \beta_j = 0, \sum_{i=1}^I (\alpha\beta)_{ij} = 0, \sum_{j=1}^J (\alpha\beta)_{ij} = 0 \text{ pour tout } i, j$$

et où  $Y_{ijs}$  est la valeur prise par la réponse  $Y$  dans les conditions  $(\alpha_i, \beta_j)$  pour le  $s$ -ème sujet.

Nous notons  $n = I \times J \times S$  le nombre total de mesures ayant été effectuées.

## Contexte

- Un facteur contrôlé  $\alpha$  se présente sous  $I$  modalités, chacune d'entre elles étant notée  $\alpha_i$ .
- Un facteur contrôlé  $\beta$  se présente sous  $J$  modalités, chacune d'entre elles étant notée  $\beta_j$ .
- Les  $\tau_S$  représentent un échantillon de taille  $S$  prélevé dans une population importante. Nous admettrons que les effets des  $\tau_S$  sont distribués suivant une loi normale centrée de variance  $\sigma_T^2$ .



## Conditions liées à ce type d'analyse

Nous supposons que

$$\mathcal{L}(\tau_s) = \mathcal{N}(0, \sigma_\tau^2), \text{ pour tout } s, \quad 1 \leq s \leq S,$$

$$\mathcal{L}((\alpha\tau)_{is}) = \mathcal{N}\left(0, \frac{I-1}{I} \sigma_{\alpha\tau}^2\right), \forall (i, s), 1 \leq i \leq I, 1 \leq s \leq S$$

$$\mathcal{L}((\beta\tau)_{js}) = \mathcal{N}\left(0, \frac{J-1}{J} \sigma_{\beta\tau}^2\right), \forall (j, s), 1 \leq j \leq J, 1 \leq s \leq S.$$

## Conditions liées à ce type d'analyse

Nous supposons aussi l'indépendance des effets aléatoires :

- les effets aléatoires  $\tau_S$  sont indépendants
- les effets aléatoires  $(\alpha\tau)_{iS}$  sont indépendants
- les effets aléatoires  $(\beta\tau)_{jS}$  sont indépendants
- les effets aléatoires  $\tau_S$  et  $(\alpha\tau)_{iS}$  sont indépendants
- les effets aléatoires  $\tau_S$  et  $(\beta\tau)_{jS}$  sont indépendants
- les effets aléatoires  $(\alpha\tau)_{iS}$  et  $(\beta\tau)_{jS}$  sont indépendants.

## Conditions classiques de l'ANOVA

Nous postulons les hypothèses classiques de l'ANOVA pour les variables erreurs  $\mathcal{E}_{ijs}$  :

- les erreurs sont indépendantes
- les erreurs ont même variance  $\sigma^2$  inconnue
- les erreurs sont de loi gaussienne.

## Ajout de conditions

Nous ajoutons :

- l'indépendance des effets aléatoires  $\tau_s$  et des erreurs  $\mathcal{E}_{ijs}$
- l'indépendance des effets aléatoires  $(\alpha\tau)_{is}$  et des erreurs  $\mathcal{E}_{ijs}$
- l'indépendance des effets aléatoires  $(\beta\tau)_{js}$  et des erreurs  $\mathcal{E}_{ijs}$

## Relation fondamentale de l'ANOVA

Nous supposons que les conditions d'utilisation de ce modèle sont bien remplies.

Nous utilisons les quantités  $SC_\alpha$ ,  $SC_\beta$ ,  $SC_\tau$ ,  $SC_{\alpha\beta}$ ,  $SC_{\alpha\tau}$ ,  $SC_{\beta\tau}$ ,  $SC_R$ ,  $SC_{TOT}$  déjà introduites au chapitre précédent.

Nous rappelons la relation fondamentale de l'ANOVA :

$$SC_{TOT} = SC_\alpha + SC_\beta + SC_\tau + SC_{\alpha\beta} + SC_{\alpha\tau} + SC_{\beta\tau} + SC_R.$$

## Tableau de l'ANOVA

Variation	$SC$	$ddl$	$CM$	$F_{obs}$	$F_c$
Due au fact. $\alpha$	$sc_{\alpha}$	$I - 1$	$cm_{\alpha}$	$\frac{cm_{\alpha}}{cm_{\alpha\tau}}$	$c$
Due au fact. $\beta$	$sc_{\beta}$	$J - 1$	$cm_{\beta}$	$\frac{cm_{\beta\tau}}{cm_{\alpha\beta}}$	$c$
Due au fact. $\tau$	$sc_{\tau}$	$S - 1$	$cm_{\tau}$	$\frac{cm_{\tau}}{cm_{\alpha\beta}}$	$c$
Due à l'inter.	$sc_{\alpha\beta}$	$(I - 1)(J - 1)$	$cm_{\alpha\beta}$	$\frac{cm_{\alpha\beta}}{cm_R}$	$c$
Due à l'inter.	$sc_{\alpha\tau}$	$(I - 1)(S - 1)$	$cm_{\alpha\tau}$	$\frac{cm_{\alpha\tau}}{cm_R}$	$c$

## Tableau de l'ANOVA

Variation	$SC$	$ddl$	$CM$	$F_{obs}$	$F_c$
Due à l'inter.	$sc_{\beta\tau}$	$(J - 1)(S - 1)$	$cm_{\alpha\beta}$	$\frac{cm_{\alpha\beta}}{cm_R}$	$c$
Résiduelle	$sc_R$	$(I - 1)(J - 1)(S - 1)$	$cm_R$		
Totale	$sc_{TOT}$	$n - 1$			

## Tests d'hypothèses

L'analyse de la variance à 2 facteurs avec mesures répétées sur les deux facteurs permet six tests de Fisher.

## Premier test

Nous testons l'hypothèse nulle

$$(\mathcal{H}_0) : \alpha_1 = \alpha_2 = \dots = \alpha_I = 0$$

contre l'hypothèse alternative

$$(\mathcal{H}_1) : \text{Il existe } i_0 \in \{1, 2, \dots, I\} \text{ tel que } \alpha_{i_0} \neq 0.$$

Sous l'hypothèse nulle ( $\mathcal{H}_0$ ) précédente d'absence d'effet du facteur  $\alpha$  et lorsque les conditions de validité du modèle sont respectées,  $F_{\alpha, obs}$  est la réalisation d'une variable aléatoire qui suit une loi de Fisher à  $I - 1$  et  $(I - 1)(S - 1)$  degrés de liberté.



## Premier test - Suite

Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  du test, ou à l'aide d'une table, rejet si la valeur  $F_{\alpha,obs}$  est supérieure ou égale à la valeur critique issue de la table de Fisher. Lorsque l'hypothèse nulle ( $\mathcal{H}_0$ ) est rejetée, nous pouvons procéder à des tests de comparaisons multiples des différents effets des niveaux du facteur  $\alpha$ .

## Deuxième test

Nous testons l'hypothèse nulle

$$(\mathcal{H}_0) : \beta_1 = \beta_2 = \dots = \beta_I = 0$$

contre l'hypothèse alternative

$$(\mathcal{H}_1) : \text{Il existe } j_0 \in \{1, 2, \dots, J\} \text{ tel que } \beta_{j_0} \neq 0.$$

Sous l'hypothèse nulle ( $\mathcal{H}_0$ ) précédente d'absence d'effet du facteur  $\beta$  et lorsque les conditions de validité du modèle sont respectées,  $F_{\beta,obs}$  est la réalisation d'une variable aléatoire qui suit une loi de Fisher à  $J - 1$  et  $(J - 1)(S - 1)$  degrés de liberté.

## Deuxième test - Suite

Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  du test, ou à l'aide d'une table, rejet si la valeur  $F_{\beta,obs}$  est supérieure ou égale à la valeur critique issue de la table de Fisher. Lorsque l'hypothèse nulle ( $\mathcal{H}_0$ ) est rejetée, nous pouvons procéder à des tests de comparaisons multiples des différents effets des niveaux du facteur  $\beta$ .

## Troisième test

Nous testons l'hypothèse nulle

$$(\mathcal{H}_0) : \sigma_\tau^2 = 0$$

contre l'hypothèse alternative

$$(\mathcal{H}_1) : \sigma_\tau^2 \neq 0.$$

Sous l'hypothèse nulle ( $\mathcal{H}_0$ ) précédente d'absence d'effet du facteur  $\tau$  et lorsque les conditions de validité du modèle sont respectées,  $F_{\tau,obs}$  est la réalisation d'une variable aléatoire qui suit une loi de Fisher à  $S - 1$  et  $(I - 1)(J - 1)(S - 1)$  degrés de liberté.

## Remarque

En fonction du contexte et de la problématique posée, vous pourrez être amené à vous en servir.

## Quatrième test

Nous testons l'hypothèse nulle

$$(\mathcal{H}_0) : (\alpha\beta)_{11} = (\alpha\beta)_{12} = \dots = (\alpha\beta)_{IJ} = 0$$

contre l'hypothèse alternative

$$(\mathcal{H}_1) : \text{Il existe } (i_0, j_0) \in \{1, \dots, I\} \times \{1, \dots, J\} \text{ tq } (\alpha\beta)_{i_0 j_0} \neq 0.$$

Sous l'hypothèse nulle ( $\mathcal{H}_0$ ) précédente d'absence d'effet du facteur  $\alpha\beta$  et lorsque les conditions de validité du modèle sont respectées,  $F_{\alpha\beta, obs}$  est la réalisation d'une variable aléatoire qui suit une loi de Fisher à  $(I - 1)(J - 1)$  et  $(I - 1)(J - 1)(S - 1)$  degrés de liberté.

## Quatrième test - Suite

Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  du test, ou à l'aide d'une table, rejet si la valeur  $F_{\alpha\beta,obs}$  est supérieure ou égale à la valeur critique issue de la table de Fisher. Lorsque l'hypothèse nulle ( $\mathcal{H}_0$ ) est rejetée, nous pouvons procéder à des tests de comparaisons multiples des différents effets des niveaux du facteur  $\alpha\beta$ .

## Cinquième test

Nous testons l'hypothèse nulle

$$(\mathcal{H}_0) : \sigma_{\alpha\tau}^2 = 0$$

contre l'hypothèse alternative

$$(\mathcal{H}_1) : \sigma_{\alpha\tau}^2 \neq 0.$$

Sous l'hypothèse nulle ( $\mathcal{H}_0$ ) précédente d'absence d'effet du facteur  $\alpha\tau$  et lorsque les conditions de validité du modèle sont respectées,  $F_{\alpha\tau, obs}$  est la réalisation d'une variable aléatoire qui suit une loi de Fisher à  $(I - 1)(S - 1)$  et  $(I - 1)(J - 1)(S - 1)$  degrés de liberté.



## Remarque

En fonction du contexte et de la problématique posée, vous pourrez être amené à vous en servir.

## Cinquième test

Nous testons l'hypothèse nulle

$$(\mathcal{H}_0) : \sigma_{\beta\tau}^2 = 0$$

contre l'hypothèse alternative

$$(\mathcal{H}_1) : \sigma_{\beta\tau}^2 \neq 0.$$

Sous l'hypothèse nulle ( $\mathcal{H}_0$ ) précédente d'absence d'effet du facteur  $\beta\tau$  et lorsque les conditions de validité du modèle sont respectées,  $F_{\beta\tau,obs}$  est la réalisation d'une variable aléatoire qui suit une loi de Fisher à  $(J - 1)(S - 1)$  et  $(I - 1)(J - 1)(S - 1)$  degrés de liberté.

## Remarque

En fonction du contexte et de la problématique posée, vous pourrez être amené à vous en servir.

## Retour à l'exemple - Sortie avec MINITAB

Analyse de la variance pour Volume, avec utilisation de la somme des carrés ajustée pour les tests

Sour.	DL	SC séq	CM ajust	F	P
A	1	1587,00	1587,00	775,87	0,000
B	1	2028,00	2028,00	524,89	0,000
Suj.	11	258,50	23,50	4,92	0,005 x
A*B	1	147,00	147,00	129,36	0,000
S*A	11	22,50	2,05	1,80	0,172
S*B	11	42,50	3,86	3,40	0,027
Err.	11	12,50	12,50	1,14	
Tot.	47	4098,00			

x n'est pas un test F exact.

S = 1,06600

R carré = 99,69% R carré (ajust) = 98,70 %

## Remarque

Il n'est pas nécessaire dans le cas présent de procéder à des comparaisons multiples pour le facteur  $\alpha$  et le facteur  $\beta$  puisque que chacun a deux niveaux.

## Remarque sur le logiciel MINITAB

Il resterait à faire des comparaisons multiples pour l'interaction entre la drogue A et la drogue B...Mais

\* ATTENTION \*

Aucune comparaison multiple n'est calculée pour les termes qui contiennent (ou interagissant avec) des facteurs aléatoires.