Modèles aléatoires et mixtes de l'analyse de la variance à deux facteurs

Frédéric Bertrand[1] & Myriam Maumy[1]

[1]IRMA, Université de Strasbourg
Strasbourg, France

DUS2
20-06-2011

Sommaire

Références

Ce cours s'appuie essentiellement sur

1. le livre David C. Howell, **Méthodes statistiques en sciences humaines** traduit de la sixième édition américaine aux éditions de Boeck, 2008.
2. le livre de Pierre Dagnelie, **Statistique théorique et appliquée**, Tome 2, aux éditions de Boeck, 1998.
3. le livre de Hardeo Sahai et Mohammed I. Ageel, **The Analysis of Variance : Fixed, Random and Mixed Models**, aux éditions Birkhäuser, 2000.

Sommaire

Quatre nouveaux modèles

Comme nous l'avons vu dans le chapitre « Compléments sur l'analyse de la variance à un facteur », il se peut que les effets d'un facteur ne puissent être modélisés par des effets fixes. Par conséquent, nous pouvons être confrontés à quatre autres types de modèles :

1. Un modèle avec deux facteurs à effets aléatoires, avec ou sans répétitions.
   Ces deux modèles sont appelés modèles à effets aléatoires.
2. Un modèle avec un facteur à effets aléatoires et un facteur à effets fixes, avec ou sans répétitions.
   Ces deux modèles sont appelés modèles à effets mixtes.

Sommaire

Exemple issu du livre de Dagnelie

Les responsables d'un laboratoire d'analyse chimique par spectrométrie dans le proche infrarouge se sont intéressés à la variabilité des résultats qu'ils obtenaient pour les mesures des teneurs en protéines du blé.
En particulier, ils se sont interrogés sur l'importance des différences qui pouvaient découler des étapes successives de préparation des matières à analyser.
Nous considérons ici le problème du broyage, en examinant les résultats obtenus à l'aide de trois moulins.

Suite de la mise en situation

Cinq échantillons de grains de blé ont été prélevés au hasard dans un arrivage relativement important, et divisés chacun en six sous-échantillons.
Pour chacun des échantillons, les sous-échantillons ont ensuite été affectés au hasard à trois moulins qui eux-mêmes ont été choisis au hasard dans une production de moulins.
Pour terminer, une analyse chimique a été effectuée dans chaque cas. Le tableau ci-dessous présente les résultats, à savoir les mesures des teneurs en protéines, exprimées en pourcentage de la matière sèche.

Tableau des données

Moulin/Échantillon	1	2	3	4	5
1	$13,33$	$13,62$	$13,53$	$13,60$	$13,97$
	$13,43$	$13,33$	$13,75$	$13,44$	$13,32$
2	$13,04$	$13,26$	$13,49$	$13,05$	$13,28$
	$13,34$	$13,49$	$13,59$	$13,44$	$13,67$
3	$13,24$	$13,33$	$13,07$	$13,47$	$13,46$
	$13,25$	$13,46$	$13,33$	$13,04$	$13,32$

Remarque

Le modèle d'analyse de la variance qui peut être envisagé pour analyser ces données est un modèle à deux facteurs aléatoires avec répétitions.

Le modèle

Le modèle statistique s'écrit de la façon suivante :

$$Y_{ijk} = \mu + A_i + B_j + (AB)_{ij} + \mathcal{E}_{ijk},$$

où $i = 1, \ldots, I, j = 1, \ldots, J, k = 1, \ldots, K$,
où Y_{ijk} est la valeur prise par la réponse Y dans les conditions (A_i, B_j) lors du $k-$ème essai.
Notons $n = I \times J \times K$ le nombre total de mesures ayant été effectuées.

Contexte

- Les termes A_i représentent un échantillon de taille I prélevé dans une population importante. Nous admettrons que les effets des A_i sont distribués suivant une loi normale centrée de variance σ_A^2.
- Les termes B_j représentent un échantillon de taille J prélevé dans une population importante. Nous admettrons que les effets des B_j sont distribués suivant une loi normale centrée de variance σ_B^2.
- Pour chacun des couples de modalités (A_i, B_j) nous effectuons $K \geqslant 2$ mesures d'une réponse Y qui est une variable continue.

Conditions liées à ce type d'analyse

Nous supposons que

$$\begin{aligned}
\mathcal{L}\left(A_i\right) &= \mathcal{N}(0,\sigma_A^2), \text{ pour tout } i, \quad 1 \leqslant i \leqslant I, \\
\mathcal{L}\left(B_j\right) &= \mathcal{N}(0,\sigma_B^2), \text{ pour tout } j, \quad 1 \leqslant j \leqslant J, \\
\mathcal{L}\left((AB)_{ij}\right) &= \mathcal{N}(0,\sigma_{AB}^2), \text{ pour tout } (i,j), 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J,
\end{aligned}$$

ainsi que l'indépendance des effets aléatoires :

- les effets aléatoires A_i sont indépendants
- les effets aléatoires B_j sont indépendants
- les effets aléatoires $(AB)_{ij}$ sont indépendants
- les effets aléatoires A_i et B_j sont indépendants
- les effets aléatoires A_i et $(AB)_{ij}$ sont indépendants
- les effets aléatoires B_j et $(AB)_{ij}$ sont indépendants.

Conditions classiques de l'ANOVA

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

1. les erreurs sont indépendantes
2. les erreurs ont même variance σ^2 inconnue
3. les erreurs sont de loi gaussienne.

Ajout de conditions

Nous ajoutons l'indépendance des effets aléatoires et des erreurs due à ce type d'analyse :

- les effets aléatoires A_i et les erreurs $\mathcal{E}_{ijk}$ sont indépendants
- les effets aléatoires B_j et les erreurs $\mathcal{E}_{ijk}$ sont indépendants
- les effets aléatoires $(AB)_{ij}$ et les erreurs $\mathcal{E}_{ijk}$ sont indépendants.

Relation fondamentale de l'ANOVA

Nous supposons que les conditions d'utilisation de ce modèle sont bien remplies.
Nous utilisons les quantités SC_A, SC_B, SC_{AB}, SC_R, SC_{TOT} déjà introduites au chapitre précédent.
Nous rappelons la relation fondamentale de l'ANOVA :

$$SC_{TOT} = SC_A + SC_B + SC_{AB} + SC_R.$$

Tableau de l'ANOVA

Variation	SC	ddl	CM	F_{obs}	F_c
Due au facteur A	sc_A	$I-1$	cm_A	$\dfrac{cm_A}{cm_{AB}}$	c
Due au facteur B	sc_B	$J-1$	cm_B	$\dfrac{cm_B}{cm_{AB}}$	c
Interaction	sc_{AB}	$(I-1)(J-1)$	cm_{AB}	$\dfrac{cm_{AB}}{cm_R}$	c
Résiduelle	sc_R	$IJ(K-1)$	cm_R		
Totale	sc_{TOT}	$n-1$			

Tests d'hypothèses

L'analyse de la variance à deux facteurs aléatoires avec répétitions permet trois tests de Fisher.

Premier test

Nous testons l'hypothèse nulle

$$\boxed{(\mathcal{H}_0) : \sigma_A^2 = 0}$$

contre l'hypothèse alternative

$$\boxed{(\mathcal{H}_1) : \sigma_A^2 \neq 0.}$$

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

Deuxième test

Nous testons l'hypothèse nulle

$$\boxed{(\mathcal{H}_0) : \sigma_B^2 = 0}$$

contre l'hypothèse alternative

$$\boxed{(\mathcal{H}_1) : \sigma_B^2 \neq 0.}$$

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

Troisième test

Nous testons l'hypothèse nulle

$$\boxed{(\mathcal{H}_0) : \sigma^2_{AB} = 0}$$

contre l'hypothèse alternative

$$\boxed{(\mathcal{H}_1) : \sigma^2_{AB} \neq 0.}$$

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

Retour à l'exemple - Sortie avec MINITAB

Analyse de la variance pour Teneurs en proteines, avec utilisation de la somme des carrés ajustée pour les tests

Source	DL	SomCar séq	CM ajust	F	P
Mou	2	0,29246	0,14623	8,70	0,010
Ech	4	0,20731	0,05183	3,08	0,082
Mou*Ech	8	0,13451	0,01681	0,38	0,917
Erreur	15	0,66840	0,04456		
Total	29	1,30268			

S = 0,211092 R carré = 48,69% R carré (ajust) = 0,80 %

Remarque

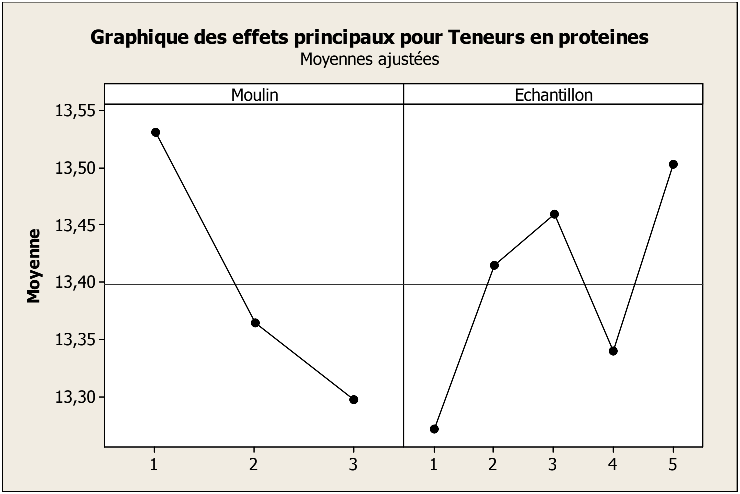
Nous supposons que les conditions du modèle sont bien remplies. Ce que nous vérifierons par la suite.

Analyse des résultats

1. Pour le premier test, P-value = 0,010, nous décidons, au seuil $\alpha = 5\%$, de refuser l'hypothèse nulle $(\mathcal{H}_0)$. Par conséquent, nous pouvons dire qu'il y a un effet significatif du facteur aléatoire « Moulin ».
   Le risque associé à cette décision est un risque de première espèce qui vaut 5%.

Analyse des résultats - Suite et fin

2. Pour le deuxième test, P-value = 0,082, 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 aléatoire « Échantillon ».
   Le risque associé à cette décision est un risque de deuxième espèce. Pour l'évaluer, il resterait à calculer la puissance de ce test.
3. Pour le troisième test, P-value = 0,917, 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 aléatoire « 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.


Graphique des effets principaux pour Teneurs en proteines
Moyennes ajustées
Moulin
Echantillon
Moyenne
13,55
13,50
13,45
13,40
13,35
13,30
1
2
3
1
2
3
4
5

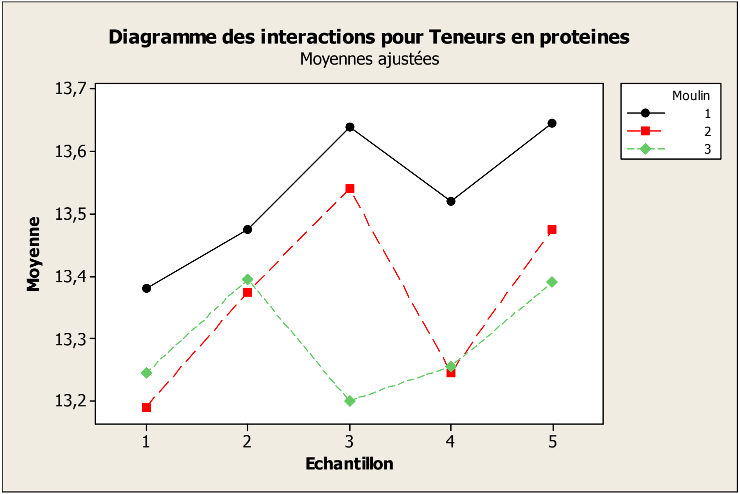


Diagramme des interactions pour Teneurs en proteines
Moyennes ajustées
Moyenne
13,7
13,6
13,5
13,4
13,3
13,2
1
2
3
4
5
Echantillon
Moulin
1
2
3

Sommaire

Exemple adapté du Diplôme Universitaire de Statistique

Nous étudions la dissolution du principe actif contenu dans un type donné de comprimé issu de lots de production distincts. Pour cela, six lots ont été sélectionnés au hasard parmi toute la production et la dissolution de quatre comprimés pris au hasard dans chacun des lots est observée. Après 15, 30, 45 et 60 minutes, un comprimé de chaque lot est sélectionné et le pourcentage de principe actif dissous, par rapport à la valeur titre, est déterminé. Ces valeurs sont données dans le tableau qui va suivre. Il est à noter que les temps d'observation à savoir, 15, 30, 45 et 60 minutes sont des temps qui ont été choisis aléatoirement par l'expérimentateur qui n'avait pas de connaissance a priori sur ces 24 comprimés.

Tableau des données

Lot/Temps	15 min	30 min	45 min	60 min
Lot 1	66	87	93	90
Lot 2	60	91	99	98
Lot 3	69	91	93	92
Lot 4	61	97	97	101
Lot 5	61	84	106	103
Lot 6	57	88	94	99

Question que se pose l'expérimentateur

À partir de quel instant pouvons-nous admettre qu'un comprimé est entièrement dissous ?

Modèle statistique

Le modèle statistique s'écrit de la façon suivante :

$$Y_{ij} = \mu + A_i + B_j + \mathcal{E}_{ij},$$

où $i = 1, \ldots, I, j = 1, \ldots, J,$
où Y_{ij} est la valeur prise par la réponse Y dans les conditions (A_i, B_j).
Notons $n = I \times J$ le nombre total de mesures ayant été effectuées.

Contexte

- Les termes A_i représentent un échantillon de taille I prélevé dans une population importante. Nous admettrons que les effets des A_i sont distribués suivant une loi normale centrée de variance σ_A^2.
- Les termes B_j représentent un échantillon de taille J prélevé dans une population importante. Nous admettrons que les effets des B_j sont distribués suivant une loi normale centrée de variance σ_B^2.

Conditions liées à ce type d'analyse

Nous supposons que

$$\begin{aligned} \mathcal{L}(A_i) &= \mathcal{N}(0, \sigma_A^2), \text{ pour tout } i, \quad 1 \leqslant i \leqslant I, \\ \mathcal{L}(B_j) &= \mathcal{N}(0, \sigma_B^2), \text{pour tout } j, \quad 1 \leqslant j \leqslant J, \end{aligned}$$

ainsi que l'indépendance des effets aléatoires :

- les effets aléatoires A_i sont indépendants
- les effets aléatoires B_j sont indépendants
- les effets aléatoires A_i et B_j sont indépendants.

Conditions classiques de l'ANOVA

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

1. les erreurs sont indépendantes
2. les erreurs ont même variance σ^2 inconnue
3. les erreurs sont de loi gaussienne.

Ajout de conditions

Nous ajoutons l'indépendance des effets aléatoires et des erreurs due à ce type d'analyse :

- les effets aléatoires A_i et les erreurs $\mathcal{E}_{ij}$ sont indépendants
- les effets aléatoires B_j et les erreurs $\mathcal{E}_{ij}$ sont indépendants.

Relation fondamentale de l'ANOVA

Nous supposons que les conditions d'utilisation de ce modèle sont bien remplies.
Nous utilisons les quantités SC_A, SC_B, SC_R, SC_{TOT} déjà introduites au chapitre précédent.
Nous rappelons la relation fondamentale de l'ANOVA :

$$SC_{TOT} = SC_A + SC_B + SC_R.$$

Tableau de l'ANOVA

Variation	SC	ddl	CM	F_{obs}	F_c
Due au facteur A	sc_A	$I-1$	cm_A	$\dfrac{cm_A}{cm_R}$	c
Due au facteur B	sc_B	$J-1$	cm_B	$\dfrac{cm_B}{cm_R}$	c
Résiduelle	sc_R	$(I-1)(J-1)$	cm_R		
Totale	sc_{TOT}	$n-1$			

Tests d'hypothèses

L'analyse de la variance à deux facteurs aléatoires sans répétition permet deux tests de Fisher.

Premier test

Nous testons l'hypothèse nulle

$$\boxed{(\mathcal{H}_0) : \sigma_A^2 = 0}$$

contre l'hypothèse alternative

$$\boxed{(\mathcal{H}_1) : \sigma_A^2 \neq 0.}$$

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

Deuxième test

Nous testons l'hypothèse nulle

$$\boxed{(\mathcal{H}_0) : \sigma_B^2 = 0}$$

contre l'hypothèse alternative

$$\boxed{(\mathcal{H}_1) : \sigma_B^2 \neq 0.}$$

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

Retour à l'exemple - Sortie avec MINITAB

Analyse de la variance pour Principe actif dissous, avec utilisation de la somme des carrés ajustée pour les tests

Source	DL	SomCar séq	CM ajust	F	P
Lot	5	83,21	16,64	0,68	0,647
Temps	3	4908,46	1636,15	66,6	0,000
Erreur	15	368,29	24,55		
Total	23	5359,96			

S = 4,95508 R carré = 93,13% R carré (ajust) = 89,46%

Analyse des résultats

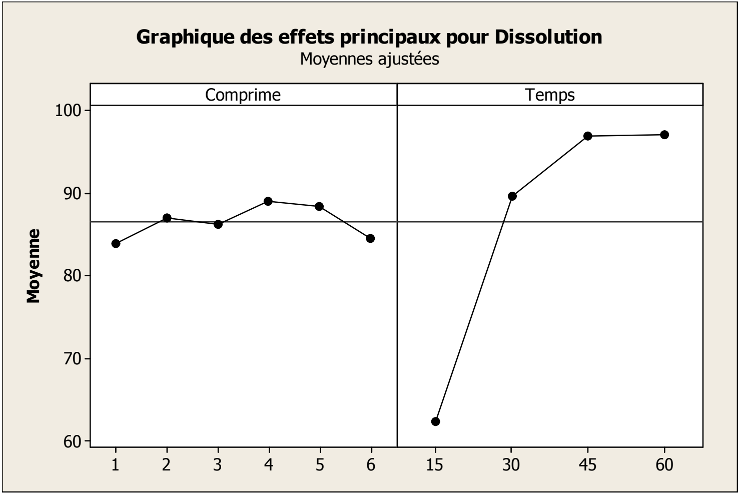
1. Pour le premier test, P-value = 0,647, nous décidons de ne pas refuser l'hypothèse nulle ($\mathcal{H}_0$). Par conséquent, nous n'avons pas réussi à mettre en évidence d'effet du facteur aléatoire « Lot ».
   Le risque associé à cette décision est un risque de deuxième espèce. Pour l'évaluer, il resterait à calculer la puissance de ce test.
2. Pour le deuxième 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 aléatoire « Temps ».

Analyse des résultats - Suite et fin

Nous ne sommes pas capables de répondre à la question de l'expérimentateur, à savoir :
« à partir de quel instant pouvons-nous admettre qu'un comprimé est entièrement dissous ? »
puisque nous ne pouvons pas faire de tests de comparaisons multiples, étant donné que le facteur « Temps » est à effets aléatoires.

Remarque

Bien sûr, nous pouvons faire cette analyse des résultats, si auparavant nous avons vérifié que les conditions du modèle sont bien remplies. Ce que nous vérifierons ultérieurement.


Graphique des effets principaux pour Dissolution
Moyennes ajustées
Comprime
Temps
Moyenne
100
90
80
70
60
1
2
3
4
5
6
15
30
45
60

Sommaire

Exemple issu du livre de Howell

Eysenck (1974) a mené une étude consacrée à la rétention de matériel verbal en fonction du niveau de traitement. Elle faisait varier aussi bien l'âge que la condition de rétention.

Le modèle de la mémorisation proposé par Craik et Lockhart (1972) stipule que le degré auquel un sujet se rappelle un matériel verbal est fonction du degré auquel ce matériel a été traité lors de sa présentation initiale. Ainsi, si l'on essaie de mémoriser une liste de mots, répéter simplement un mot pour soi-même (un niveau de traitement très bas) ne permet pas de le mémoriser aussi bien que si l'on y réfléchit en tentant de former des associations entre ce mot et un autre.

Exemple issu du livre de Howell (suite)

Eysenck (1974) voulait tester ce modèle et, plus important encore, examiner s'il pouvait contribuer à expliquer certaines différences relevées entre des sujets jeunes et âgés concernant leur aptitude à se rappeler du matériel verbal.

Eysenck a réparti aléatoirement 50 sujets âgés de 55 à 65 ans dans cinq groupes ; les quatre premiers impliquaient un apprentissage involontaire et le dernier un apprentissage intentionnel (l'apprentissage involontaire se caractérisait par le fait que le sujet ne savait pas qu'il devrait plus tard se rappeler le matériel appris).

Exemple issu du livre de Howell (suite)

Le premier groupe (addition) devait lire une liste de mots et se contenter de compter le nombre de lettres de chacun d'eux. Il s'agissait du niveau de traitement le plus bas, puisqu'il n'était pas nécessaire chaque mot autrement que comme une suite de lettres.

Le deuxième groupe (rimes) devait lire chaque mot et lui trouver une rime. Cette tâche impliquait de considérer la consonance de chaque mot, mais pas sa signification.

Le troisième groupe (adjectifs) devait donner un adjectif qui aurait pu être utilisé pour modifier chaque mot de la liste.

Exemple issu du livre de Howell (suite)

Le quatrième groupe (images) devait essayer de se former une image précise de chaque mot. Cette dernière tâche était supposée nécessiter le niveau de traitement le plus élevé parmi les quatre groupes d'apprentissage involontaire.

Aucun de ces groupes ne savait qu'il faudrait se rappeler les mots ultérieurement.

Exemple issu du livre de Howell (suite)

Enfin, le groupe d'apprentissage intentionnel devait lire la liste et mémoriser tous les mots. Après avoir passé trois fois en revue la liste de 27 mots, les sujets devaient retranscrire tous les mots dont ils se souvenaient.

Si l'apprentissage n'impliquait rien de plus qu'une exposition au matériel (soit la façon dont la plupart d'entre nous lisent le journal ou, pis encore, un devoir), les cinq groupes devaient obtenir des résultats identiques ; après tout, ils avaient tous vu tous les mots. Si le niveau de traitement était important, on devait constater des différences sensibles entre les moyennes des groupes.

Exemple issu du livre de Howell (suite)

L'étude incluait 50 participants dont l'âge se situait entre 18 et 30 ans, ainsi que 50 participants compris dans la tranche d'âge 55-65 ans. Pour plus de facilité, nous avons regroupé les 50 participants dont l'âge se situait entre 18 et 30 ans dans une classe que nous appellerons « sujets jeunes » et les 50 participants dont l'âge se situait entre 55 et 65 ans dans une classe que nous allons appeler « sujets âgés ».

Les données sont présentées dans le tableau suivant :

Exemple : **Sujets jeunes**

Addition	Rimes	Adjectifs	Images	Intentionnel
8	10	14	20	21
6	7	11	16	19
4	8	18	16	17
6	10	14	15	15
7	4	13	18	22
6	7	22	16	16
5	10	17	20	22
7	6	16	22	22
9	7	12	14	18
7	7	11	19	21

Exemple : **Sujets âgés**

Addition	Rimes	Adjectifs	Images	Intentionnel
9	7	11	12	10
8	9	13	11	19
6	6	8	16	14
8	6	6	11	5
10	6	14	9	10
4	11	11	23	11
6	6	13	12	14
5	3	13	10	15
7	8	10	19	11
7	7	11	11	11

Le modèle

Le modèle statistique s'écrit de la façon suivante :

$$Y_{ijk} = \mu + \alpha_i + B_j + (\alpha B)_{ij} + \mathcal{E}_{ijk}$$

où $i = 1, \ldots, I, j = 1, \ldots, J, k = 1, \ldots, K$,
avec les contraintes supplémentaires :

$$\sum_{i=1}^{I} \alpha_i = 0 \quad \text{et} \quad \sum_{i=1}^{I} (\alpha B)_{ij} = 0, \text{ pour tout } j \in \{1, \ldots, J\}$$

où Y_{ijk} est la valeur prise par la réponse Y dans les conditions (α_i, B_j) lors du $k-$ème essai.
Notons $n = I \times J \times K$ le nombre total de mesures ayant été effectuées.

Contexte

1. Un facteur contrôlé α se présente sous I modalités, chacune d'entre elles étant notée α_i.
2. Les B_j représentent un échantillon de taille J prélevé dans une population importante. Nous admettrons que les effets des B_j sont distribués suivant une loi normale centrée de variance σ_B^2.
3. Pour chacun des couples de modalités (α_i, B_j) nous effectuons $K \geqslant 2$ mesures d'une réponse Y qui est une variable continue.

Conditions liées à ce type d'analyse

Nous supposons que

$$\begin{aligned}\mathcal{L}(B_j) &= \mathcal{N}(0, \sigma_B^2), \text{ pour tout } j, 1 \leqslant j \leqslant J, \\ \mathcal{L}((\alpha B)_{ij}) &= \mathcal{N}(0, \sigma_{\alpha B}^2), \text{ pour tout } (i,j), 1 \leqslant i \leqslant I, \ 1 \leqslant j \leqslant J,\end{aligned}$$

ainsi que l'indépendance des effets aléatoires :

- les effets aléatoires B_j sont indépendants
- les effets aléatoires $(\alpha B)_{ij}$ sont indépendants
- les effets aléatoires B_j et $(\alpha B)_{ij}$ sont indépendants.

Conditions classiques de l'ANOVA

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

1. les erreurs sont indépendantes
2. les erreurs ont même variance σ^2 inconnue
3. les erreurs sont de loi gaussienne.

Ajout de conditions

Nous ajoutons l'indépendance des effets aléatoires et des erreurs due à ce type d'analyse :

- les effets aléatoires B_j et les erreurs $\mathcal{E}_{ijk}$ sont indépendants
- les effets aléatoires $(\alpha B)_{ij}$ et les erreurs $\mathcal{E}_{ijk}$ sont indépendants.

Relation fondamentale de l'ANOVA

Nous supposons que les conditions d'utilisation de ce modèle sont bien remplies.
Nous utilisons les quantités SC_α, SC_B, $SC_{\alpha B}$, SC_R, SC_{TOT} déjà introduites au chapitre précédent.
Nous rappelons la relation fondamentale de l'ANOVA :

$$SC_{TOT} = SC_\alpha + SC_B + SC_{\alpha B} + SC_R.$$

Tableau de l'ANOVA

Variation	SC	ddl	CM	F_{obs}	F_c
Due au facteur α	sc_α	$I-1$	cm_α	$\dfrac{cm_\alpha}{cm_{\alpha B}}$	c
Due au facteur B	sc_B	$J-1$	cm_B	$\dfrac{cm_B}{cm_R}$	c
Interaction	$sc_{\alpha B}$	$(I-1)(J-1)$	$cm_{\alpha B}$	$\dfrac{cm_{\alpha B}}{cm_R}$	c
Résiduelle	sc_R	$IJ(K-1)$	cm_R		
Totale	sc_{TOT}	$n-1$			

Tests d'hypothèses

L'analyse de la variance à un facteur fixe et à un facteur aléatoire avec répétitions permet trois tests de Fisher.

Premier test

Nous testons l'hypothèse nulle

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

contre l'hypothèse alternative

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

Sous l'hypothèse nulle $(\mathcal{H}_0)$ précédente d'absence d'effet du facteur α 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 $(I-1)(J-1)$ degrés de liberté.

Décision

Nous concluons alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α 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.

Tests de comparaisons multiples

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. Nous renvoyons au chapitre 1 qui traite des principales méthodes de comparaisons multiples.

Deuxième test

Nous testons l'hypothèse nulle

$$\boxed{(\mathcal{H}_0) : \sigma_B^2 = 0}$$

contre l'hypothèse alternative

$$\boxed{(\mathcal{H}_1) : \sigma_B^2 \neq 0.}$$

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

Troisième test

Nous testons l'hypothèse nulle

$$\boxed{(\mathcal{H}_0) : \sigma^2_{\alpha B} = 0}$$

contre l'hypothèse alternative

$$\boxed{(\mathcal{H}_1) : \sigma^2_{\alpha B} \neq 0.}$$

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

Retour à l'exemple - Sortie avec MINITAB

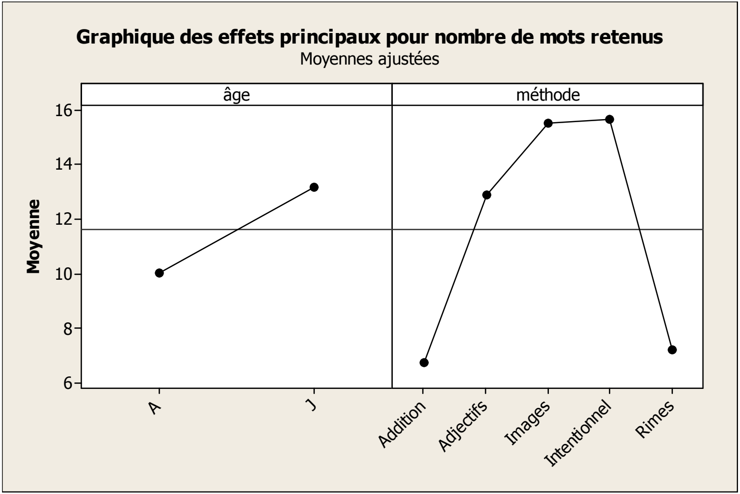
Source	DL	SomCar séq	CM ajust	F	P
age	1	240,25	240,25	5,05	0,088
met	4	1514,94	378,73	47,19	0,000
age*met	4	190,30	47,57	5,93	0,000
Erreur	90	722,30	8,03		
Total	99	2667,79			

Remarque

Nous allons faire une analyse des résultats, en supposant que les conditions du modèle sont bien remplies. Ce que nous vérifierons par la suite.
En Travaux Dirigés, vous apprendrez en particulier à vérifier la normalité du facteur à effets aléatoires.

Analyse des résultats

1. Pour le premier test, P-value = 0,088, 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 aléatoire « Âge ». Le risque associé à cette décision est un risque de deuxième espèce. Pour l'évaluer, il resterait à calculer la puissance de ce test.
2. Pour le deuxième test, P-value = 0,035, 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 « Méthode ».
3. 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 aléatoire « Interaction ».


Graphique des effets principaux pour nombre de mots retenus
Moyennes ajustées
âge
méthode
Moyenne
16
14
12
10
8
6
A
J
Addition
Adjectifs
Images
Intentionnel
Rimes

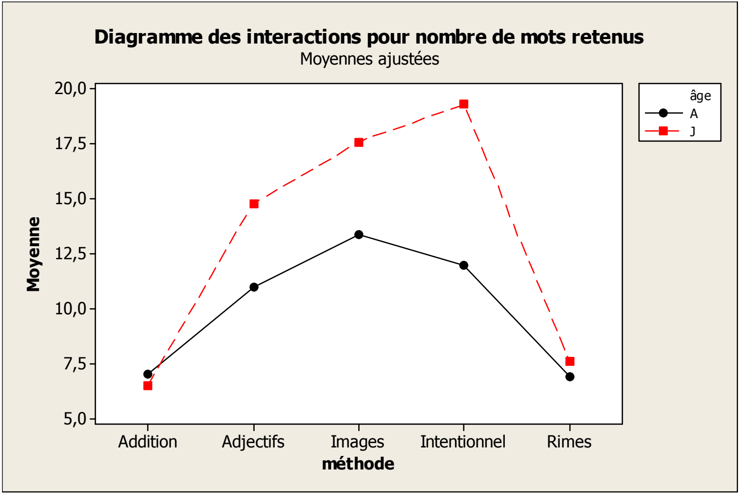


Diagramme des interactions pour nombre de mots retenus
Moyennes ajustées
Moyenne
20,0
17,5
15,0
12,5
10,0
7,5
5,0
Addition
Adjectifs
Images
Intentionnel
Rimes
méthode
âge
A
J

Sommaire

Exemple issu du Diplôme Universitaire de Statistique

Nous reprenons les données de l'exemple que nous avions étudié dans le cas de l'analyse à deux facteurs aléatoires sans répétition. Mais cette fois-ci, nous allons considérer le facteur « Temps » comme un facteur fixe. Par contre le facteur « Lot » reste toujours un facteur aléatoire.

Le modèle

Le modèle statistique s'écrit de la façon suivante :

$$Y_{ij} = \mu + \alpha_i + B_j + \mathcal{E}_{ij}$$

où $i = 1, \ldots, I, j = 1, \ldots, J,$
avec la contrainte supplémentaire :

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

où Y_{ij} est la valeur prise par la réponse Y dans les conditions (α_i, B_j).
Notons $n = I \times J$ le nombre total de mesures ayant été effectuées.

Contexte

1. Un facteur contrôlé α se présente sous I modalités, chacune d'entre elles étant notée α_i.
2. Les B_j représentent un échantillon de taille J prélevé dans une population importante. Nous admettrons que les effets des B_j sont distribués suivant une loi normale centrée de variance σ_B^2.

Conditions liées à ce type d'analyse

Nous supposons que

- $\mathcal{L}\left(B_j\right) = \mathcal{N}(0, \sigma_B^2)$, pour tout $j, 1 \leqslant j \leqslant J$,
- les effets aléatoires B_j sont indépendants.

Conditions classiques de l'ANOVA

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

1. les erreurs sont indépendantes
2. les erreurs ont même variance σ^2 inconnue
3. les erreurs sont de loi gaussienne.

Rajout de conditions

Nous ajoutons l'indépendance des effets aléatoires et des erreurs due à ce type d'analyse :

- les effets aléatoires B_j et les erreurs $\mathcal{E}_{ij}$ sont indépendants.

Relation fondamentale de l'ANOVA

Nous supposons que les conditions d'utilisation de ce modèle sont bien remplies.
Nous utilisons les quantités SC_α, SC_B, SC_R, SC_{TOT} déjà introduites au chapitre précédent.
Nous rappelons la relation fondamentale de l'ANOVA :

$$SC_{TOT} = SC_\alpha + SC_B + SC_R.$$

Tableau de l'ANOVA

Variation	SC	ddl	CM	F_{obs}	F_c
Due au facteur α	sc_α	$I-1$	cm_α	$\dfrac{cm_\alpha}{cm_R}$	c
Due au facteur B	sc_B	$J-1$	cm_B	$\dfrac{cm_B}{cm_R}$	c
Résiduelle	sc_R	$(I-1)(J-1)$	cm_R		
Totale	sc_{TOT}	$n-1$			

Tests d'hypothèses

L'analyse de la variance à un facteur fixe et à un facteur aléatoire sans répétition permet deux tests de Fisher.

Premier test

Nous testons l'hypothèse nulle

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

contre l'hypothèse alternative

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

Sous l'hypothèse nulle $(\mathcal{H}_0)$ précédente d'absence d'effet du facteur α 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 $(I - 1)(J - 1)$ degrés de liberté.

Décision

Nous concluons alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α 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.

Comparaisons multiples

Lorsque l'hypothèse nulle $(\mathcal{H}_0)$ est rejetée, nous pouvons procéder à des comparaisons multiples des différents effets des niveaux du facteur. Nous renvoyons au chapitre 1 qui traite des principales méthodes de comparaisons multiples.

Deuxième test

Nous testons l'hypothèse nulle

$$\boxed{(\mathcal{H}_0) : \sigma_B^2 = 0}$$

contre l'hypothèse alternative

$$\boxed{(\mathcal{H}_1) : \sigma_B^2 \neq 0.}$$

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

Retour à l'exemple - Sortie avec MINITAB

Analyse de la variance pour Principe actif dissous, avec utilisation de la somme des carrés ajustée pour les tests

Source	DL	SomCar séq	CM ajust	F	P
Lot	5	83,21	16,64	0,68	0,647
Temps	3	4908,46	1636,15	66,6	0,000
Erreur	15	368,29	24,55		
Total	23	5359,96			

S = 4,95508 R carré = 93,13% R carré (ajust) = 89,46%

Analyse des résultats

1. Pour le premier test, P-value = 0,647, 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 aléatoire « Lot ». Le risque associé à cette décision est un risque de deuxième espèce. Pour l'évaluer, il resterait à calculer la puissance de ce test.
2. Pour le deuxième 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 « Temps ».

Analyse des résultats - Suite

Nous sommes maintenant capables de répondre à la question de l'expérimentateur, à savoir
« à partir de quel instant pouvons-nous admettre qu'un comprimé est entièrement dissous ? »
puisque nous pouvons faire des comparaisons multiples, étant donné que le facteur « Temps » est maintenant fixe.

Remarque

Nous n'avons pas présenté de graphique des effets principaux pour la dissolution du comprimé, car le graphique est identique à celui du cas où les deux facteurs sont à effets aléatoires.

```
Tests de simultanéité de Tukey
Variable de réponse Principe actif dissous
Toutes les comparaisons deux à deux sur les
niveaux de Temps
Temps = 15 soustrait de :
                Erreur type
           Dif        de la   Valeur  Valeur de
 Temps  des moy  différence     de T  p ajustée
 30       27,33       2,861    9,554     0,0000
 45       34,67       2,861   12,118     0,0000
 60       34,83       2,861   12,176     0,0000
```

```
Temps = 30 soustrait de :
                 Erreur type
          Dif          de la   Valeur   Valeur de
 Temps  des moy   différence     de T   p ajustée
 45       7,333        2,861    2,563      0,0898
 60       7,500        2,861    2,622      0,0809
```

```
Temps = 45 soustrait de :
                    Erreur type
            Dif          de la     Valeur  Valeur de
 Temps   des moy   différence       de T  p ajustée
 60        0,1667        2,861   0,05826     0,9999
```

```
Tests de simultanéité de Dunnett
Variable de réponse Principe actif dissous
Comparaisons avec niveau de contrôle
Temps = 60 soustrait de :
```

Temps	Dif des moy	Erreur type de la différence	Valeur de T	Valeur de p ajustée
15	-34,83	2,861	-12,18	0,0000
30	-7,50	2,861	-2,62	0,0489
45	-0,17	2,861	-0,06	0,9999