Principes fondamentaux de l'analyse des expériences

Règles générales pour les modèles d'analyse de la variance

Frédéric Bertrand[1]

[1] IRMA, Université de Strasbourg
Strasbourg, France

ENSAI 3e Année
2015-2016

Sommaire

Sommaire

Référence

Ce cours s'appuie essentiellement sur

1. le livre de Hardeo Sahai et Mohammed I. Ageel, **The Analysis of Variance : Fixed, Random and Mixed Models**, aux éditions Birkhäuser, 2000.
2. le livre de Hardeo Sahai et Miguel Ojeda, **Analysis of Variance for Random Models**, **Volume I: Balanced data**, aux éditions Birkhäuser, 2004.
3. le livre de Hardeo Sahai et Miguel Ojeda, **Analysis of Variance for Random Models**, **Volume II: Unbalanced data**, aux éditions Birkhäuser, 2005.

Sommaire

Sommaire

Généralités

Introduction

Dans cette section, nous nous intéressons aux règles de construction d'un modèle d'analyse de la variance dans le cas équilibré. Ces règles s'appliquent aux plans qui vérifient les deux conditions suivantes :

1. si le plan comporte des facteurs croisés, alors pour chacune des combinaisons possibles de leurs modalités, le nombre des observations est identique ;
2. si le plan est complètement ou partiellement emboîté, alors un même facteur emboîté a un nombre de modalités constant lorsque les niveaux du ou des facteurs dans lequel il est emboîté varient.

Généralités

Cas des plans sans répétitions

Pour appliquer les résultats qui suivent lorsqu'il n'y a pas de répétition, il ne faut pas oublier de faire apparaître l'indice associé aux répétitions.

Il ne faudra pas être surpris par le fait que

- la somme des carrés des erreurs sera nulle (la valeur moyenne par cellule coïncide par la seule valeur observée !) ;
- il ne sera pas possible de tester tous les termes apparaissant dans le modèle puisque nous ne pouvons estimer la variance des erreurs.

Généralités

Cas des plans sans répétitions (suite)

En particulier, il n'est jamais possible de tester le terme de plus haut degré (celui qui implique tous les indices sauf celui associé aux répétitions).

Pour pouvoir se ramener à une situation plus favorable, nous devons **supposer** que le terme de plus haut degré n'a pas d'effet (hypothèse d'additivité). Le carré moyen qui lui est associé devient ainsi un estimateur sans biais de la variance des erreurs et prend la place du carré moyen résiduel dont nous nous servons en présence de répétitions.

Généralités

Modèles restreints et non-restreints

Les modèles **restreints** ne peuvent s'utiliser que dans le cas des plans **équilibrés** et dans cette situation ils correspondent naturellement avec la logique avec laquelle nous cherchons à modéliser les expériences et en particulier avec celle liée à l'introduction des termes d'erreurs dans le modèle.

Un modèle restreint est un modèle pour lequel les termes d'interaction entre des effets aléatoires et des effets fixes héritent des contraintes des effets fixes. Ces effets aléatoires ne sont donc pas indépendants puisque, pour toute combinaison fixée des modalités des facteurs à effets fixes, leur somme suivant les modalités des facteurs à effets aléatoires est nulle.

Généralités

Modèles restreints et non-restreints (suite)

Les modèles non **restreints** sont adaptés au cas des plans **déséquilibrés** et doivent nécessairement être utilisés dans cette situation (Searle 1971).

Un modèle non-restreint est un modèle pour le lequel les termes d'interaction entre des effets aléatoires et des effets fixes n'héritent pas des contraintes des effets fixes. Ces effets aléatoires sont donc indépendants.

Généralités

Modèles restreints et non-restreints (suite)

De nombreux logiciels de statistique utilisent par défaut et/ou exclusivement les modèles non-restreints. C'est le cas de la `PROC GLM` de **SAS** ou de la fonction `aov` de **R**.

Ce choix par défaut est compréhensible puisque les modèles restreints ne sont adaptés qu'au cas des plans équilibrés.

Généralités

Modèles restreints et non-restreints (suite)

Toutefois, la plupart des manuels consacrés au modèle linéaire et à l'analyse d'expérience, ainsi que de nombreux statisticiens de premier plan, recommandent l'utilisation des modèles restreints si le plan est équilibré.

Par conséquent, il nous faut apprendre à construire les tableaux d'analyse de la variance qui sont associés à ces modèles afin de pouvoir procéder à l'analyse effective des expériences puisque les logiciels de statistique ne vous les fourniront généralement pas.

L'objectif de ces supports est de vous fournir tous les éléments pour parvenir à réaliser cet objectif.

Généralités

Modèles restreints et non-restreints (suite)

La différence entre les modèles restreints et les modèles non-restreints se manifeste lors du calcul de l'espérance des carrés moyens.

Par conséquent, les statistiques de Fisher dédiées aux tests des différents termes du modèle peuvent différer d'un modèle à l'autre, les hypothèses testées ne s'interprétant pas dans la pratique de la même manière.

Généralités

Modèles restreints et non-restreints (suite)

Certaines des techniques détaillées par la suite pour le cas des modèles restreints sont également adaptées aux modèles non-restreints :

- trouver les termes du modèle ;
- trouver l'expression des sommes des carrés ;
- trouver les degrés de liberté des sommes des carrés.

Par contre le calcul de l'espérance des carrés moyens proposé, et tout ce qui en découle, tests et puissances, n'est adapté qu'au cas des modèles restreints.

Les coquillages

Exemple

Nous illustrerons les règles de construction avec l'exemple d'un modèle mixte d'analyse de la variance comportant trois facteurs à la fois croisés et emboîtés, c'est-à-dire un plan partiellement emboîté.

Les facteurs A et C sont croisés et le facteur B est emboîté dans le facteur A et croisé avec le facteur C.

1. Le facteur A a a niveaux et est à effets fixes.
2. Le facteur B a b niveaux et est à effets aléatoires.
3. Le facteur C a c niveaux et est à effets fixes.
4. Il y a n répétitions.

Les coquillages

Exemple

Le jeu de données, reproduit sur les vignettes suivantes, comporte, comme réponse expérimentale Y la densité de peuplement, exprimée en $\sqrt{\cdot} cm^{-2}$, de larves *Semibalanus balanoides*, un petit crustacé qui se trouve en abondance sur les plages rocheuses européennes.

Ces crustacés ne peuvent se déplacer qu'au stade larvaire.

La question est de savoir si le nombre de crustacés présents sur un rocher influe sur le nombre de larves venant s'y accrocher.

Il est également vraisemblable que le taux de peuplement des rochers voisins puisse influencer ce nombre.

Les coquillages

Exemple

Plusieurs rochers, ont été nettoyés des crustacés les habitant à l'exception d'un emplacement central où deux, huit ou 32 adultes n'ont pas été touchés. Ce facteur est le Facteur **Traitement**.

Pour déterminer la possible influence de la densité de peuplement des rochers voisins, qui est une caractéristique globale de la plage donc commune à tous les rochers de celle-ci, ces emplacements ont été préparés dans des conditions similaires sur différentes plages pour lesquels cette densité est élevée ou faible (Facteur **Recrutement**). Pour chacune de ces densités deux plages (Facteur **Plage**), ont été utilisées.

Les coquillages

Exemple

Cette situation expérimentale, qui s'appelle un plan **split plot** (en parcelles divisées), correspond bien au modèle ci-dessus puisque :

- les facteurs **Recrutement** et **Traitement** sont croisés à effets fixes ;
- le facteur **Plage** est à effets aléatoires et emboîté dans le facteur **Recrutement**.

Les coquillages

Exemple

T	Recrutement élevé					
	Cowes			Seaview		
2	0,386	0,397	0,432	0,279	0,411	0,260
8	0,484	0,482	0,514	0,625	0,531	0,478
32	0,484	0,520	0,569	0,738	0,570	0,620

T	Recrutement faible					
	Totland			Ventnor		
2	0,190	0,177	0,300	0,304	0,302	0,278
8	0,268	0,261	0,396	0,402	0,351	0,254
32	0,384	0,319	0,334	0,244	0,401	0,324

Sommaire

Règles de construction d'un modèle d'analyse de la variance

Objectif

Les règles qui vont suivre vont vous permettre de construire tous les termes du modèle à partir des observations faites sur les rapports entre les facteurs. Il suffit de connaître :

- les facteurs qui sont croisés ;
- les facteurs qui sont emboîtés et dans quels facteurs ceux-ci sont emboîtés.

Règle I.1

La constante

Tout modèle doit contenir un terme constant, qui s'interprète comme la moyenne générale, et qui est noté μ.

Exemple

Pour l'exemple précédent, le terme μ appartient au modèle.

Règle I.2

Les effets principaux

Tout modèle doit contenir un effet principal pour chacun des facteurs qui est désigné par la lettre minuscule grecque correspondante, s'il est à effets fixes, ou la majuscule romaine correspondante, s'il est à effets aléatoires, avec un indice indiquant la modalité du facteur à laquelle est associé cet effet. Si un facteur est emboîté dans un autre, cette relation est précisée par l'utilisation de parenthèses dans les indices décrivant les modalités.

Exemple

Pour l'exemple précédent, les effets principaux des facteurs A, B et C sont : α_i, $B_{j(i)}$, γ_k, $1 \leqslant i \leqslant a$, $1 \leqslant j \leqslant b$ et $1 \leqslant k \leqslant c$.

Règle I.3

Les interactions entre facteurs croisés

Tout modèle doit contenir tous les termes d'interaction associés à tous les facteurs croisés. Il n'y a pas de terme d'interaction entre un facteur emboîté et celui dans lequel ce facteur est emboîté.

Le terme d'interaction est noté dans le modèle par un ensemble de lettre minuscules grecques et/ou de majuscules romaines entre parenthèse avec un indice indiquant la combinaison des modalités des facteurs à laquelle est associé cet effet.

Règle I.3

Exemple

Le modèle contient donc les interactions $A \times C$ et $B \times C$ mais ne comporte ni le terme d'interaction $A \times B$ ni le terme d'interaction $A \times B \times C$ puisque le facteur B est emboîté dans le facteur A.

Pour l'exemple précédent, les termes d'interaction $A \times C$ et $B \times C$ sont : $(\alpha\gamma)_{ik}$ et $(B\gamma)_{jk}$, $1 \leqslant i \leqslant a$, $1 \leqslant j \leqslant b$ et $1 \leqslant k \leqslant c$.

Règle I.4

Les interactions avec des facteurs emboîtés

Les interactions entre un facteur emboîté et un facteur avec lequel ce facteur emboîté est croisé doivent toujours être considérées comme étant emboîtées.

Si un terme d'interaction est emboité dans un autre terme, l'emboîtement est indiqué par la présence de parenthèses dans l'indice associé à ce terme. Il est donc être nécessaire de compléter les indices de certains termes créés lors de l'application de la Règle I.3.

Règle I.4

Exemple

Le facteur B est emboîté dans le facteur A et est croisé avec le facteur C, ainsi l'interaction $B \times C$ doit être considérée comme étant, elle aussi, emboîtée dans le facteur A. Nous devons donc compléter l'indice du terme $(B\gamma)_{jk}$.

Pour l'exemple précédent, le fait que le terme d'interaction $B \times C$ est emboîté dans le facteur A est précisé par la présence de parenthèses dans la notation suivante : $(B\gamma)_{jk(i)}$, $1 \leqslant i \leqslant a$, $1 \leqslant j \leqslant b$ et $1 \leqslant k \leqslant c$.

Règle I.5

Le terme d'erreur

Le dernier terme à ajouter au modèle est le terme d'erreur qui est considéré comme étant emboité dans tous les facteurs.

Exemple

Pour l'exemple précédent, le terme d'erreur est $\varepsilon_{l(ijk)}$, $1 \leqslant i \leqslant a$, $1 \leqslant j \leqslant b$, $1 \leqslant k \leqslant c$ et $1 \leqslant l \leqslant n$.

Règle I.6

Le modèle final

Le modèle final s'écrit comme une équation reliant la variable réponse, généralement notée y ou x, comportant tous les indices utilisés précédemment et apparaissant dans le membre de gauche, à la somme de tous les termes construits lors de l'application des Règles I.1 à I.5.

Exemple

Pour l'exemple précédent, le modèle final retenu est :

$$y_{ijkl} = \mu + \alpha_i + B_{j(i)} + \gamma_k + (\alpha\gamma)_{ik} + (B\gamma)_{jk(i)} + \varepsilon_{l(ijk)}$$

avec $1 \leqslant i \leqslant a$, $1 \leqslant j \leqslant b$, $1 \leqslant k \leqslant c$ et $1 \leqslant l \leqslant n$.

Règle I.7 facultative

Différences entre blocs complets, split plot et mesures répétées

En présence de facteurs à effets aléatoires, il est souvent fait la distinction entre les deux situations expérimentales suivantes qui sont associées aux mêmes modèles statistiques.

- Les blocs complets, les split plots ;
- Les mesures répétées.

Exemple

Un effet « bloc » est par exemple celui qui affecte tous les traitements réalisés au sein d'une même parcelle, ou dans notre exemple tous les rochers d'une même plage. Il n'est pas contrôlable et ses causes ne peuvent pas être précisément déterminées.

Règle I.7 facultative

Différences entre blocs complets, split plot et mesures répétées

Certains manuels et statisticiens recommandent d'adopter le comportement suivant.

- **Blocs complets ou split plot**
  Lorsqu'un facteur à effets aléatoires est associé à un effet « bloc » ou « parcelle », il convient d'**envisager des interactions** entre ce facteur à effets aléatoires et les **autres facteurs** car l'effet du bloc est généralement la somme de divers effets qu'il est impossible de connaître et donc de maîtriser ;

Règle I.7 facultative

Différences entre blocs complets, split plot et mesures répétées

- **Mesures répétées**
  Lorsqu'un facteur à effets aléatoires est introduit dans le modèle pour modéliser des mesures répétées, les critères d'inclusion des sujets dans les études sont souvent suffisamment précis et strictes pour qu'il ne soit **pas nécessaire d'envisager des interactions** entre ce facteur à effets aléatoires et les **autres facteurs**.

Il n'y a pas de consensus réel sur ce plan au sein de la communauté statistique.

Sommaire

Expression des sommes des carrés et calcul des degrés de liberté

Introduction

Dans cette section, nous nous intéressons aux règles de construction des sommes des carrés et de calcul des degrés de liberté pour les modèles d'analyse de la variance. Ces règles s'appliquent aux plans équilibrés :

1. si le plan comporte des facteurs croisés, alors pour chacune des combinaisons possibles de leurs modalités, le nombre des observations est identique ;
2. si le plan est complètement ou partiellement emboîté, alors un même facteur emboîté a un nombre de modalités constant lorsque les niveaux du ou des facteurs dans lequel il est emboîté varient.

Expression des sommes des carrés et calcul des degrés de liberté

Exemple

Nous illustrerons les règles de construction avec l'exemple d'un modèle mixte d'analyse de la variance comportant trois facteurs à la fois croisés et emboîtés, c'est-à-dire un plan partiellement emboîté.

Les facteurs A et C sont croisés et le facteur B est emboîté dans le facteur A et croisé avec le facteur C.

1. Le facteur A a a niveaux et est à effets fixes.
2. Le facteur B a b niveaux et est à effets aléatoires.
3. Le facteur C a c niveaux et est à effets fixes.
4. Il y a n répétitions.

Règle II.1

Déterminer le modèle à utiliser

Écrire le modèle en suivant les règles établies à la Section précédente.

Exemple

Pour l'exemple précédent, le modèle final construit est :

$$y_{ijkl} = \mu + \alpha_i + B_{j(i)} + \gamma_k + (\alpha\gamma)_{ik} + (B\gamma)_{jk(i)} + \varepsilon_{l(ijk)}$$

avec $1 \leqslant i \leqslant a$, $1 \leqslant j \leqslant b$, $1 \leqslant k \leqslant c$ et $1 \leqslant l \leqslant n$.

Règle II.2

Produit symbolique

Pour chacun des termes du modèle, à l'exception de la constante, écrire un produit symbolique comportant chacun des indices, soit égal à lui-même si l'indice apparaît entre parenthèses, soit égal à lui-même moins un si l'indice n'apparaît pas entre parenthèses.

Développer le produit symbolique.

Exemple

Pour l'exemple précédent, le produit symbolique de α_i est $i-1$, celui de $(B\gamma)_{jk(i)}$ est $i(j-1)(k-1) = ijk - ij - ik + i, \ldots$

Règle II.3

Terme général des sommes des carrés

L'expression à mettre au carré pour obtenir le terme général de la somme des carrés associée à l'un des termes du modèle est formée de moyennes des valeurs de la réponse. Ces moyennes sont associées aux termes du développement du produit symbolique construit à la Règle II.2 : la moyenne est réalisée par rapport aux indices qui n'apparaissent pas dans le produit symbolique qui sont donc remplacés par des points $\bullet$. Le nombre 1 est associé à la moyenne générale c'est-à-dire celle pour laquelle tous les indices ont été remplacés par des points $\bullet$.

Règle II.3

Exemple

Pour l'exemple précédent, le produit symbolique :

1. de α_i est $i-1$ et l'expression à mettre au carré puis à sommer est $\overline{y_{i\bullet\bullet\bullet}} - \overline{y_{\bullet\bullet\bullet\bullet}}$;
2. de $(B\gamma)_{jk(i)}$ est $i(j-1)(k-1) = ijk - ij - ik + i$ et l'expression à mettre au carré puis à sommer est $\overline{y_{ijk\bullet}} - \overline{y_{ij\bullet\bullet}} - \overline{y_{i\bullet k\bullet}} + \overline{y_{i\bullet\bullet\bullet}}, \ldots$

Règle II.4

Somme des carrés

La somme des carrés pour l'un des termes du modèle s'obtient en mettant au carré l'expression obtenue à la Règle II.3, puis en sommant par rapport aux indices qui apparaissent dans cette expression et enfin en multipliant le résultat par le produit du nombre des niveaux des facteurs dont les indices n'apparaissent pas dans le terme du modèle considéré.

Règle II.4

Exemple

Pour l'exemple précédent, la somme des carrés :

1. pour α_i est obtenue en mettant au carré $\overline{y_{i\bullet\bullet\bullet}} - \overline{y_{\bullet\bullet\bullet\bullet}}$ puis en sommant par rapport à i et enfin en multipliant par bcn :

$$bcn \sum_{i=1}^{a} \left(\overline{y_{i\bullet\bullet\bullet}} - \overline{y_{\bullet\bullet\bullet\bullet}}\right)^2 ;$$

Règle II.4

Exemple

2. pour $(B\gamma)_{jk(i)}$ est obtenue en mettant au carré $\overline{y_{ijk\bullet}} - \overline{y_{ij\bullet\bullet}} - \overline{y_{i\bullet k\bullet}} + \overline{y_{i\bullet\bullet\bullet}}$ puis en sommant par rapport à i, j, k et enfin en multipliant par n :

$$n\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\left(\overline{y_{ijk\bullet}} - \overline{y_{ij\bullet\bullet}} - \overline{y_{i\bullet k\bullet}} + \overline{y_{i\bullet\bullet\bullet}}\right)^2, \ldots$$

Règle II.5

Somme des carrés de la moyenne générale

La somme des carrés pour la moyenne générale s'obtient en mettant au carré la moyenne générale $\overline{y_{\bullet\bullet\bullet\bullet}}$ puis en multipliant le résultat par le nombre total d'observations *abcn* :
$abcn\,(\overline{y_{\bullet\bullet\bullet\bullet}})^2$.

Cette somme des carrés ne figure généralement pas dans les tableaux d'analyse de la variance.

Règle II.6

Somme des carrés totale

La somme des carrés totale est la somme pour tous les indices, c'est-à-dire pour toutes les observations, des carrés des écarts entre les valeurs observées et la moyenne générale :

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{l=1}^{n}\left(y_{ijkl}-\overline{y_{\bullet\bullet\bullet\bullet}}\right)^2.$$

Règle II.7

Degrés de liberté

Les degrés de liberté associés à une somme des carrés s'obtiennent en remplaçant dans le produit symbolique, calculé à la Règle II.2, les indices par le nombre de modalités des facteurs qui leur sont associés.

Exemple

Pour l'exemple précédent, le produit symbolique de α_i est $i - 1$, les degrés de liberté sont donc $a - 1$, celui de $(B\gamma)_{jk(i)}$ est $i(j-1)(k-1) = ijk - ij - ik - i$, les degrés de liberté sont donc $a(b-1)(c-1)$, ...

Règle II.8

Degrés de liberté total

Le nombre de degrés de liberté pour la moyenne générale est égal à 1 et le nombre total de degrés de liberté est défini comme égal au nombre total d'observations moins un.

Expression des sommes des carrés et calcul des degrés de liberté

Relation fondamentale de l'ANOVA

Nous utilisons les quantités SC_α, $SC_{B|\alpha}$, SC_γ, $SC_{B\gamma|\alpha}$, $SC_{\alpha\gamma}$, SC_R et SC_{TOT} introduites précédemment.

Pour ce modèle, la relation fondamentale de l'ANOVA est la suivante :

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

Sommaire

Sommaire

Règles de calcul de l'espérance des carrés moyens

Introduction

Le calcul de l'espérance des carrés moyens dans les modèles d'analyse de la variance est crucial pour construire une statistique appropriée au test de Fisher de l'influence de l'un des termes du modèle sur la réponse. Ce calcul est également d'un intérêt majeur pour déterminer des estimateurs des composants de la variance. La difficulté mathématique associée à la détermination des espérances des carrés moyens n'est pas très élevée mais les calculs peuvent s'avérer longs et fastidieux.

Les règles définies aux deux sections précédentes fonctionnent de manière identique avec les facteurs à effets fixes et ceux à effets aléatoires. Il n'en va pas de même pour celles que nous allons voir maintenant.

Règles de calcul de l'espérance des carrés moyens

Introduction

Dans cette section, nous nous intéressons aux règles de calcul de l'espérance des carrés moyens pour les modèles d'analyse de la variance dans le cas équilibré. Ces règles s'appliquent aux plans qui vérifient les deux conditions suivantes :

1. si le plan comporte des facteurs croisés, alors pour chacune des combinaisons possibles de leurs modalités, le nombre des observations est identique ;
2. si le plan est complètement ou partiellement emboîté, alors un même facteur emboîté a un nombre de modalités constant lorsque les niveaux du ou des facteurs dans lequel il est emboîté varient.

Règles de calcul de l'espérance des carrés moyens

Exemple

Nous illustrerons les règles de construction avec l'exemple d'un modèle mixte d'analyse de la variance comportant trois facteurs à la fois croisés et emboîtés, c'est-à-dire un plan partiellement emboîté.

Les facteurs A et C sont croisés et le facteur B est emboîté dans le facteur A et croisé avec le facteur C.

1. Le facteur A a a niveaux et est à effets fixes.
2. Le facteur B a b niveaux et est à effets aléatoires.
3. Le facteur C a c niveaux et est à effets fixes.
4. Il y a n répétitions.

Règle III.1

Déterminer le modèle à utiliser

Écrire le modèle en suivant les règles établies à la première Section.

Exemple

Pour l'exemple précédent, le modèle final construit est :

$$y_{ijkl} = \mu + \alpha_i + B_{j(i)} + \gamma_k + (\alpha\gamma)_{ik} + (B\gamma)_{jk(i)} + \varepsilon_{l(ijk)}$$

avec $1 \leqslant i \leqslant a$, $1 \leqslant j \leqslant b$, $1 \leqslant k \leqslant c$ et $1 \leqslant l \leqslant n$.

Règle III.1

Exemple

Les α_i, γ_k et $(\alpha\gamma)_{ik}$ satisfont aux contraintes usuelles :

$$\sum_{i=1}^{a} \alpha_i = \sum_{k=1}^{c} \gamma_k = 0$$

$$\sum_{i=1}^{a} (\alpha\gamma)_{ik} = \sum_{k=1}^{c} (\alpha\gamma)_{ik} = 0.$$

Nous supposons de plus que :

1. $\mathcal{L}\left(B_{j(i)}\right) = \mathcal{N}(0, \sigma^2_{B|\alpha})$;
2. $\mathcal{L}\left((B\gamma)_{jk(i)}\right) = \mathcal{N}(0, \sigma^2_{B\gamma|\alpha})$;
3. $\mathcal{L}\left(\varepsilon_{l(ijk)}\right) = \mathcal{N}(0, \sigma^2_e)$.

Règle III.1

Exemple

4. Les trois groupes d'effets aléatoires $B_{j(i)}$, $(B\gamma)_{jk(i)}$ et $\varepsilon_{l(ijk)}$ sont deux à deux indépendants.
5. Les effets aléatoires $B_{j(i)}$ sont indépendants ainsi que les effets aléatoires $\varepsilon_{l(ijk)}$.
6. Par contre, les effets aléatoires $(B\gamma)_{jk(i)}$ sont corrélés à cause des contraintes :

$$\sum_{k=1}^{c}(B\gamma)_{jk(i)} = 0 \quad \text{pout tout } j(i).$$

Règle III.2

Matrice intermédiaire

Considérer un tableau où il y a une ligne pour chacun des termes du modèle différents de la constante et une colonne pour chacun des indices qui apparaissent dans le modèle. L'ordre des lignes et des colonnes n'a pas d'importance. Il est toutefois recommandé d'utiliser un « ordre naturel » pour éviter les erreurs.

Règle III.2

Exemple

Pour l'exemple précédent, le tableau à construire est donc le suivant :

	i	j	k	l
α_i	...	...	...	...
$B_{j(i)}$	...	...	...	...
γ_k	...	...	...	...
$(\alpha\gamma)_{ik}$	...	...	...	...
$(B\gamma)_{jk(i)}$	...	...	...	...
$\varepsilon_{l(ijk)}$	...	...	...	...

Règle III.3

Prise en compte des emboîtements

Dans chacune des lignes du tableau construit à la Règle III.2 où un ou plusieurs indices sont entre parenthèses, écrire 1 dans chaque colonne qui correspond à un indice entre parenthèses.

Règle III.3

Exemple

Pour l'exemple précédent, le tableau construit avec la Règle III.2 se complète donc de la manière suivante :

	i	j	k	l
α_i	$\dots$	$\dots$	$\dots$	$\dots$
$B_{j(i)}$	1	$\dots$	$\dots$	$\dots$
γ_k	$\dots$	$\dots$	$\dots$	$\dots$
$(\alpha\gamma)_{ik}$	$\dots$	$\dots$	$\dots$	$\dots$
$(B\gamma)_{jk(i)}$	1	$\dots$	$\dots$	$\dots$
$\varepsilon_{l(ijk)}$	1	1	1	$\dots$

Règle III.4

Prise en compte de la nature des effets

Dans chacune des lignes du tableau construit à la Règle III.2 où un ou plusieurs indices ne sont pas entre parenthèses :

1. écrire 1 dans chaque colonne qui correspond à un indice qui n'est pas entre parenthèses si l'indice représente un facteur à effets aléatoires ;
2. écrire 0 dans chaque colonne qui correspond à un indice qui n'est pas entre parenthèses si l'indice représente un facteur à effets fixes.

Règle III.4

Exemple

Pour l'exemple précédent, le tableau construit avec les Règles III.2 et III.3 se complète donc de la manière suivante :

	i	j	k	l
α_i	0	...	...	...
$B_{j(i)}$	1	1	...	...
γ_k	...	...	0	...
$(\alpha\gamma)_{ik}$	0	...	0	...
$(B\gamma)_{jk(i)}$	1	1	0	...
$\varepsilon_{l(ijk)}$	1	1	1	1

Règle III.5

Remplissage final du tableau

Dans chacune des cellules encore vides, reporter le nombre de niveaux du facteur dont l'indice correspond au nom de la colonne.

Règle III.5

Exemple

Pour l'exemple précédent, le tableau construit avec les Règles III.2, III.3 et III.4 se complète donc de la manière suivante :

	i	j	k	l
α_i	0	b	c	n
$B_{j(i)}$	1	1	c	n
γ_k	a	b	0	n
$(\alpha\gamma)_{ik}$	0	b	0	n
$(B\gamma)_{jk(i)}$	1	1	0	n
$\varepsilon_{l(ijk)}$	1	1	1	1

Règle III.6

Paramétrisation des effets

1. Pour chacun des termes à effets fixes, le paramètre quantifiant l'importance des effets est défini par la somme des carrés des effets divisée par le nombre de degrés de liberté associé au terme.
2. Pour chacun des termes à effets aléatoires, le paramètre quantifiant l'importance des effets est défini par le composant de variance associé au terme.

Règle III.6

Paramétrisation des effets

1. Pour chacun des termes associé à un facteur à effets fixes, $\lambda = \Phi$ désignera le paramètre quantifiant l'importance des effets c'est-à-dire la somme des carrés des effets divisée par le nombre de degrés de liberté associé au terme.
2. Pour chacun des termes associé à un facteur à effets aléatoires, $\lambda = \sigma^2$ désignera le paramètre quantifiant l'importance des effets c'est-à-dire le composant de variance associé au terme.

Ajouter les paramètres λ dans une colonne à la droite de celles du tableau créé à la Règle III.2 et complété aux Règles III.3, III.4 et III.5. Chaque paramètre λ sera reporté sur la ligne qui correspond au terme du modèle auquel il est associé.

Règle III.6

Exemple

Pour l'exemple précédent, le tableau construit avec les Règles III.2, III.3, III.4 et III.5 se complète donc de la manière suivante :

	i	j	k	l	λ
α_i	0	b	c	n	$\Phi(\alpha)$
$B_{j(i)}$	1	1	c	n	$\sigma^2_{B\|\alpha}$
γ_k	a	b	0	n	$\Phi(\gamma)$
$(\alpha\gamma)_{ik}$	0	b	0	n	$\Phi(\alpha\gamma)$
$(B\gamma)_{jk(i)}$	1	1	0	n	$\sigma^2_{B\gamma\|\alpha}$
$\varepsilon_{l(ijk)}$	1	1	1	1	σ^2_e

Règle III.7

Calcul de l'espérance des carrés moyens

L'espérance du carré moyen associé à l'un des termes du modèle s'obtient par une combinaison linéaire des paramètres λ introduits à la Règle III.6 avec les coefficients suivants :

1. le coefficient du paramètre λ est égal à zéro si les indices du terme du modèle dans la ligne associée à ce paramètre, qu'ils soient entre parenthèses ou non, n'incluent pas tous les indices, y compris ceux entre parenthèses, du terme pour lequel nous souhaitons déterminer l'espérance du carré moyen ;

Règle III.7

Calcul de l'espérance des carrés moyens

2. les coefficients des paramètres λ qui ne sont pas fixés à zéro par la règle 7(1) sont obtenus en commençant par supprimer les colonnes associées aux indices qui ne sont pas entre parenthèses dans le terme pour lequel nous souhaitons déterminer l'espérance du carré moyen puis en multipliant les valeurs présentes dans les colonnes restantes.

Règle III.7

Exemple

Pour l'exemple précédent, l'application de la règle 7(1) donne :

	Espérance du carré moyen du terme					
	A	$B(A)$	C	AC	$BC(A)$	Erreur
λ	α_i	$B_{j(i)}$	γ_k	$(\alpha\gamma)_{ik}$	$(B\gamma)_{jk(i)}$	$\varepsilon_{l(ijk)}$
$\Phi(\alpha)$	...	0	0	0	0	0
$\sigma^2_{B\|\alpha}$	...	...	0	0	0	0
$\Phi(\gamma)$	0	0	...	0	0	0
$\Phi(\alpha\gamma)$	...	0	...	...	0	0
$\sigma^2_{B\gamma\|\alpha}$	...	...	...	...	...	0
σ^2_e	...	...	...	...	...	...

Règle III.7

Exemple

Pour l'exemple précédent, les coefficients des paramètres λ sont les suivants :

	Espérance du carré moyen du terme					
	A	$B(A)$	C	AC	$BC(A)$	Erreur
λ	α_i	$B_{j(i)}$	γ_k	$(\alpha\gamma)_{ik}$	$(B\gamma)_{jk(i)}$	$\varepsilon_{l(ijk)}$
$\Phi(\alpha)$	bcn	0	0	0	0	0
$\sigma^2_{B\|\alpha}$	cn	cn	0	0	0	0
$\Phi(\gamma)$	0	0	abn	0	0	0
$\Phi(\alpha\gamma)$	0	0	0	bn	0	0
$\sigma^2_{B\gamma\|\alpha}$	0	0	n	n	n	0
σ^2_e	1	1	1	1	1	1

Règle III.7

Exemple

Pour l'exemple précédent, les espérances des carrés moyens sont donc :

$$\begin{aligned}
\mathbb{E}\left(CM_\alpha\right) &= \sigma_e^2 + cn\sigma_{B|\alpha}^2 + bcn\,\overbrace{\frac{1}{a-1}\sum_{i=1}^{a}\alpha_i^2}^{\Phi(\alpha)}, \\
\mathbb{E}\left(CM_{B|\alpha}\right) &= \sigma_e^2 + cn\sigma_{B|\alpha}^2, \\
\mathbb{E}\left(CM_\gamma\right) &= \sigma_e^2 + n\sigma_{B\gamma|\alpha}^2 + abn\,\overbrace{\frac{1}{c-1}\sum_{k=1}^{c}\gamma_k^2}^{\Phi(\gamma)},
\end{aligned}$$

Règle III.7

Exemple

$$\begin{aligned}
\mathbb{E}\left(CM_{\alpha\gamma}\right) &= \sigma_e^2 + n\sigma^2_{B\gamma|\alpha} + bn\overbrace{\frac{1}{(a-1)(c-1)}\sum_{i=1}^{a}\sum_{k=1}^{c}(\alpha\gamma)^2_{ik}}^{\Phi(\alpha\gamma)}, \\
\mathbb{E}\left(CM_{B\gamma|\alpha}\right) &= \sigma_e^2 + n\sigma^2_{B\gamma|\alpha} \\
\text{et} & \\
\mathbb{E}\left(CM_R\right) &= \sigma_e^2.
\end{aligned}$$

Sommaire

Estimateurs des composants de la variance

Exemple

Remarquons que les trois relations suivantes sont satisfaites par les espérances des carrés moyens.

$$\begin{aligned}
\mathbb{E}\left(CM_R\right) &= \sigma_e^2 \\
\mathbb{E}\left(CM_{B|\alpha}\right) &= \mathbb{E}\left(CM_R\right) + cn\sigma^2_{B|\alpha} \\
\mathbb{E}\left(CM_{B\gamma|\alpha}\right) &= \mathbb{E}\left(CM_R\right) + n\sigma^2_{B\gamma|\alpha}.
\end{aligned}$$

Estimateurs des composants de la variance

Exemple

Ce simple constat nous montre que

- mis à part le carré moyen résiduel et en présence de tous les effets du modèle, aucun autre carré moyen n'est un estimateur sans biais de la source de la variance à laquelle il est attaché ;
- il est facile de construire des estimateurs sans biais des composants de la variance.

Estimateurs des composants de la variance

Exemple

Les statistiques suivantes sont des estimateurs sans biais des composants de la variance :

$$
\begin{aligned}
\widehat{\sigma_e^2} &= CM_R \\
\widehat{\sigma_{B|\alpha}^2} &= \frac{1}{cn}\left(CM_{B|\alpha} - CM_R\right) \\
\widehat{\sigma_{B\gamma|\alpha}^2} &= \frac{1}{n}\left(CM_{B\gamma|\alpha} - CM_R\right).
\end{aligned}
$$

Il est malheureusement possible de montrer que ces deux derniers estimateurs prennent des valeurs strictement négatives avec une probabilité >0. Dans ce cas, l'estimation du composant de la variance est fixée égale à 0.

Estimateurs des composants de la variance

Synthèse

Les résultats que nous avons obtenus sur cet exemple sont en fait vrais pour tous les modèles d'analyse de la variance dans lesquels il y a au moins un facteur à effets aléatoires.

- mis à part le carré moyen résiduel et en présence de tous les effets du modèle, aucun autre carré moyen n'est un estimateur sans biais de la source de la variance à laquelle il est attaché ;
- il est facile de construire des estimateurs sans biais des composants de la variance ;
- ces estimateurs prennent des valeurs strictement négatives avec une probabilité >0. Dans ce cas, nous fixons à 0 l'estimation du composant de la variance.

Sommaire

Règles III.2 à III.7 simplifiées

Termes intervenant dans l'espérance des carrés moyens

Une rapide étude de l'algorithme précédent permet d'obtenir la méthode suivante qui pourra vous être utile si vous n'êtes intéressé que par les termes qui interviennent dans les espérances des carrés moyens, pour construire les rapports de Fisher, et non la valeur précise de ceux-ci, qui, elle, sert à estimer les composantes de variance ainsi qu'à déterminer la puissance des tests.

Règles III.2 à III.7 simplifiées

Paramétrisation des effets

1. Pour chacun des termes à effets fixes, le paramètre quantifiant l'importance des effets est défini par la somme des carrés des effets divisée par le nombre de degrés de liberté associé au terme.
2. Pour chacun des termes à effets aléatoires, le paramètre quantifiant l'importance des effets est défini par le composant de variance associé au terme.

Règles III.2 à III.7 simplifiées

Paramétrisation des effets

1. Pour chacun des termes associé à un facteur à effets fixes, Φ désignera le paramètre quantifiant l'importance des effets c'est-à-dire la somme des carrés des effets divisée par le nombre de degrés de liberté associé au terme.
2. Pour chacun des termes associé à un facteur à effets aléatoires, σ^2 désignera le paramètre quantifiant l'importance des effets c'est-à-dire le composant de variance associé au terme.

Nous obtenons la liste des composants de variation (Φ ou σ^2) estimés par chaque espérance du carré moyen d'un des termes du modèle de la manière décrite dans les vignettes suivantes.

Règles III.2 à III.7 simplifiées

Termes intervenant dans l'espérance des carrés moyens (suite)

1. Faire la liste des termes du modèle dans l'ordre naturel :
   - effets principaux simples ;
   - effets principaux emboîtés ;
   - interactions entre facteurs non-emboîtés ;
   - autres termes d'interactions ;
   - erreur.

Règles III.2 à III.7 simplifiées

Termes intervenant dans l'espérance des carrés moyens (suite)

2. Les contributions à l'espérance d'un carré moyen pour un terme X donné sont à choisir parmi les variations appartenant aux lignes situées en dessous de celle associée à X.
3. Commencer par la ligne du bas du tableau et remonter progressivement jusqu'à la ligne associée à l'effet X.

Règles III.2 à III.7 simplifiées

Termes intervenant dans l'espérance des carrés moyens (suite)

4. Inclure chaque terme dès que
   - Tous les **facteurs impliqués** dans X **apparaissent** dans la source de variation de la ligne considérée ;
   - Tous les **autres** facteurs apparaissant dans la source de variation, **hormis ceux entre parenthèses** qui décrivent les emboitements, sont à effets **aléatoires**. En l'absence d'emboîtement, tous les **autres** facteurs doivent être à effets **aléatoires**.

Règles III.2 à III.7 simplifiées

Exemple

Pour l'exemple précédent, les espérances des carrés moyens sont donc :

$$
\begin{aligned}
\mathbb{E}\left(CM_{\alpha}\right) &= \sigma_e^2 + K_1\sigma^2_{B|\alpha} + K_2\Phi(\alpha),\\
\mathbb{E}\left(CM_{B|\alpha}\right) &= \sigma_e^2 + K_1\sigma^2_{B|\alpha},\\
\mathbb{E}\left(CM_{\gamma}\right) &= \sigma_e^2 + K_3\sigma^2_{B\gamma|\alpha} + K_4\Phi(\gamma),\\
\mathbb{E}\left(CM_{\alpha\gamma}\right) &= \sigma_e^2 + K_3\sigma^2_{B\gamma|\alpha} + K_5\Phi(\alpha\gamma),\\
\mathbb{E}\left(CM_{B\gamma|\alpha}\right) &= \sigma_e^2 + K_3\sigma^2_{B\gamma|\alpha}\\
\text{et}\\
\mathbb{E}\left(CM_{R}\right) &= \sigma_e^2.
\end{aligned}
$$

Règles III.2 à III.7 simplifiées

Exemple

Certains auteurs préfèrent utiliser la notation suivante :

	Effet	Ddl	Composants $\mathbb{E}$[CM]	F
1	A	$a-1$	S′(C*B′(A))+B′(A)+A	1/2
2	B′(A)	$(b-1)a$	S′(C*B′(A))+B′(A)	2/6
3	C	$c-1$	S′(C*B′(A))+C*B′(A)+C	3/5
4	C*A	$(c-1)(a-1)$	S′(C*B′(A))+C*B′(A)+C*A	4/5
5	C*B′(A)	$(c-1)(b-1)a$	S′(C*B′(A))+C*B′(A)	5/6
6	S′(C*B′(A))	$(n-1)cba$	S′(C*B′(A))	-
	Total	$ncba-1$		

Le ′ désigna un facteur à effets aléatoires et $*$ une interaction.

Sommaire

Sommaire

Statistiques de test

Détermination

Les différents carrés moyens crées par les Règles II.1 à II.4 sont indépendants.

Les statistiques de test s'obtiennent grâce aux carrés moyens en formant, lorsque cela est possible, un rapport où l'espérance du numérateur ne diffère de celle du dénominateur que par le paramètre d'effets dont nous voulons tester la nullité.

Statistiques de test

Exemple

Pour l'exemple précédent, nous introduisons cinq statistiques de Fisher.

1. Pour le test de l'effet principal de A posons :

$$F_\alpha = \frac{CM_\alpha}{CM_{B|\alpha}}$$

car

$$\mathbb{E}\left(CM_\alpha\right) = \mathbb{E}\left(CM_{B|\alpha}\right) + bcn\Phi(\alpha)$$

et ainsi sous $\mathcal{H}_0 : \Phi(\alpha) = 0$, F_α suit une loi de Fisher à $a-1$ et $a(b-1)$ degrés de liberté.

Statistiques de test

Exemple

2. Pour le test de l'effet principal de B, $B(A)$, posons :

$$F_{B|\alpha} = \frac{CM_{B|\alpha}}{CM_R}$$

car

$$\mathbb{E}\left(CM_{B|\alpha}\right) = \mathbb{E}\left(CM_R\right) + cn\sigma^2_{B|\alpha}$$

et ainsi sous $\mathcal{H}_0 : \sigma^2_{B|\alpha} = 0$, $F_{B|\alpha}$ suit une loi de Fisher à $a(b-1)$ et $abc(n-1)$ degrés de liberté.

Statistiques de test

Exemple

3. Pour le test de l'effet principal C posons :

$$F_C = \frac{CM_\gamma}{CM_{B\gamma|\alpha}}$$

car

$$\mathbb{E}\left(CM_\gamma\right) = \mathbb{E}\left(CM_{B\gamma|\alpha}\right) + abn\Phi(\gamma)$$

et ainsi sous $\mathcal{H}_0 : \Phi(\gamma) = 0$, F_C suit une loi de Fisher à $c-1$ et $a(b-1)(c-1)$ degrés de liberté.

Statistiques de test

Exemple

4. Pour le test de l'effet de l'interaction $A * C$ posons :

$$F_{\alpha\gamma} = \frac{CM_{\alpha\gamma}}{CM_{B\gamma|\alpha}}$$

car

$$\mathbb{E}\left(CM_{\alpha\gamma}\right) = \mathbb{E}\left(CM_{B\gamma|\alpha}\right) + bn\Phi(\alpha\gamma)$$

et ainsi sous $\mathcal{H}_0 : \Phi(\alpha\gamma) = 0$, $F_{\alpha\gamma}$ suit une loi de Fisher à $(a-1)(c-1)$ et $a(b-1)(c-1)$ degrés de liberté.

Statistiques de test

Exemple

5. Pour le test de l'effet de l'interaction $B * C(A)$ posons :

$$F_{B\gamma|\alpha} = \frac{CM_{B\gamma|\alpha}}{CM_R}$$

car

$$\mathbb{E}\left(CM_{B\gamma|\alpha}\right) = \mathbb{E}\left(CM_R\right) + n\sigma^2_{B\gamma|\alpha}$$

et ainsi sous $\mathcal{H}_0 : \sigma^2_{B\gamma|\alpha} = 0$, $F_{B\gamma|\alpha}$ suit une loi de Fisher à $a(b-1)(c-1)$ et $abc(n-1)$ degrés de liberté.

Tableau d'analyse de la variance

Exemple

Variation	SC	ddl	CM	F_{obs}
Fact. α	sc_α	$a-1$	cm_α	$\dfrac{cm_\alpha}{cm_{B\|\alpha}}$
Fact. $B\|\alpha$	$sc_{B\|\alpha}$	$a(b-1)$	$cm_{B\|\alpha}$	$\dfrac{cm_{B\|\alpha}}{cm_R}$
Fact. γ	sc_γ	$c-1$	cm_γ	$\dfrac{cm_\gamma}{cm_{B\gamma\|\alpha}}$
Inter. $\alpha\gamma$	$sc_{\alpha\gamma}$	$(a-1)(c-1)$	$cm_{\alpha\gamma}$	$\dfrac{cm_{\alpha\gamma}}{cm_{B\gamma\|\alpha}}$

Tableau d'analyse de la variance

Exemple

Variation	SC	ddl	CM	F_{obs}
Inter. $B\gamma\vert\alpha$	$sc_{B\gamma\vert\alpha}$	$a(b-1)(c-1)$	$cm_{B\gamma\vert\alpha}$	$\dfrac{cm_{B\gamma\vert\alpha}}{cm_R}$
Résiduelle	sc_R	$abc(n-1)$	cm_R	
Totale	sc_{TOT}	$abcn-1$		

Sommaire

Puissance des tests

Introduction

En premier lieu, il faut distinguer si vous cherchez à tester un terme à **effets fixes** ou un terme à **effets aléatoires**. Nous nous concentrerons sur le cas des statistiques suivant exactement un loi de Fisher sous $\mathcal{H}_0$.

Dans le cas où nous aurions dû utiliser un pseudo-test F, basé sur l'approximation de Satterthwaite, pour tester l'existence de cet effet, il est également possible de calculer une puissance approchée, voir Scheffé, *The Analysis of Variance*, Wiley, 1959.

Puissance des tests

Effets fixes

Pour un terme à effets fixes, nous devons calculer la puissance :

$$\Pi = \mathbb{P}\left[F'(\nu_1, \nu_2; \lambda) > F(\nu_1, \nu_2; 1-\alpha)\right]$$

où

- $F(\nu_1, \nu_2; \alpha)$ est le quantile d'ordre α d'une loi de Fisher, centrale, à ν_1 degrés de liberté au numérateur et ν_2 degrés de liberté au dénominateur ;
- $F'(\nu_1, \nu_2; \lambda)$ est une variable aléatoire distribuée suivant une loi de Fisher **non centrale** à ν_1 degrés de liberté au numérateur, ν_2 degrés de liberté au dénominateur et de paramètre de non centralité λ ;

Puissance des tests

Effets fixes (suite)

- ν_1 est le degré de liberté du numérateur de la statistique de Fisher du test ;
- ν_2 celui du dénominateur ;
- λ est un paramètre de non centralité normalisé dont l'expression s'obtient à partir de celles que nous avons obtenues de l'espérance des carrés moyens.

Puissance des tests

Effets fixes (suite)

Comme le terme testé est à effets fixes, l'espérance du carré moyen apparaissant au numérateur comporte un terme du type $\alpha \times \Phi(\text{effet})$, où α est un produit d'effectifs.

Celui-ci est égal à la différence entre l'espérance du carré moyen du numérateur $\mathbb{E}(CM_{num})$ et celle du dénominateur $\mathbb{E}(CM_{denom})$. Le paramètre λ est ainsi égal à

$$\lambda = \frac{\nu_1(\mathbb{E}(CM_{num}) - \mathbb{E}(CM_{denom}))}{2\mathbb{E}(CM_{denom})} = \frac{\alpha \times \nu_1 \Phi(\text{effet})}{2\mathbb{E}(CM_{denom})}.$$

Puissance des tests

Remarque

La grande majorité des logiciels permet de calculer directement les quantiles des lois de Fisher non centrales.
Si ceux dont vous disposez ne le permettent pas, voici deux alternatives.

- Les abaques servant usuellement aux calculs de puissance pour les effets fixes permettent encore d'obtenir la valeur explicite de la puissance. Pour s'en servir, il faut utiliser le paramètre de non centralité normalisé ϕ :

$$\phi = \sqrt{\frac{2\lambda}{(\nu_1 + 1)}} = \sqrt{\frac{\alpha \times \nu_1 \Phi(\text{effet})}{(\nu_1 + 1)\mathbb{E}(CM_{denom})}}.$$

Puissance des tests

Remarque

- Il est également possible d'utiliser l'approximation normale suivante, Johnson at al. 1995 p. 491-492. La variable

$$Z = \frac{\sqrt{\nu_1 \nu_2^{-1}(2\nu_2 - 1) F[\nu_1, \nu_2; \lambda]} - \sqrt{2(\nu_1 + 2\lambda) - (\nu_1 + 2\lambda)^{-1}(\nu_1 + 4\lambda)}}{\sqrt{\nu_1 \nu_2^{-1} F[\nu_1, \nu_2; \lambda] + (\nu_1 + 2\lambda)^{-1}(\nu_1 + 4\lambda)}}$$

suit approximativement une loi normale. Ainsi

$$\Pi = 1 - \beta \simeq \mathbb{P}(Z \leqslant z_{1-\beta})$$

Puissance des tests

Remarque

où

$$z_{1-\beta} = \frac{\sqrt{\nu_1 \nu_2^{-1}(2\nu_2 - 1)F[\nu_1, \nu_2; 1-\alpha]} - \sqrt{2(\nu_1 + 2\lambda) - (\nu_1 + 2\lambda)^{-1}(\nu_1 + 4\lambda)}}{\sqrt{\nu_1 \nu_2^{-1} F[\nu_1, \nu_2; 1-\alpha] + (\nu_1 + 2\lambda)^{-1}(\nu_1 + 4\lambda)}}$$

avec $F[\nu_1, \nu_2; \alpha]$ le quantile d'ordre α d'une loi Fisher centrale à ν_1 degrés de liberté au numérateur et ν_2 degrés de liberté au dénominateur.

Puissance des tests

Effets aléatoires

Pour un terme à effets aléatoires, nous devons calculer la puissance Π :

$$\Pi = \mathbb{P}\left[F(\nu_1, \nu_2) > \lambda^{-2} F(\nu_1, \nu_2; 1-\alpha)\right]$$

où

- $F(\nu_1, \nu_2; \alpha)$ est le quantile d'ordre α d'une loi de Fisher, centrale, à ν_1 degrés de liberté au numérateur et ν_2 degrés de liberté au dénominateur ;
- $F(\nu_1, \nu_2)$ est une variable aléatoire distribuée suivant une loi de Fisher **centrale** à ν_1 degrés de liberté au numérateur, ν_2 degrés de liberté au dénominateur ;

Puissance des tests

Effets aléatoires (suite)

- ν_1 est le degré de liberté du numérateur de la statistique de Fisher du test ;
- ν_2 celui du dénominateur ;
- λ est un paramètre dont l'expression s'obtient à partir de celle que nous avons obtenue de l'espérance des carrés moyens :

$$\lambda = \sqrt{\frac{\mathbb{E}\left(CM_{num}\right)}{\mathbb{E}\left(CM_{denom}\right)}}.$$

Puissance des tests

Remarque

La grande majorité des logiciels permet de calculer directement les quantiles des lois de Fisher.

Si ceux dont vous disposez ne le permettent pas, les abaques servant usuellement aux calculs de puissance pour les effets aléatoires permettent encore d'obtenir la valeur explicite de la puissance.

Puissances des tests

Exemple

Nous devons calculer cinq puissances.

1. Pour le test de l'effet principal de A nous avons posé :

$$F_\alpha = \frac{CM_\alpha}{CM_{B|\alpha}}$$

et ainsi la puissance Π_A est

$$\Pi_\alpha = \mathbb{P}\left[F'(a-1, a(b-1); \lambda) > F(a-1, a(b-1); 1-\alpha)\right]$$

avec

$$\lambda = \frac{bcn(a-1)\Phi(\alpha)}{2(\sigma_e^2 + cn\sigma_{B|\alpha}^2)} \quad \text{et} \quad \phi = \sqrt{\frac{bcn(a-1)\Phi(\alpha)}{a(\sigma_e^2 + cn\sigma_{B|\alpha}^2)}}.$$

Puissances des tests

Exemple

2. Pour le test de l'effet principal de B, $B(A)$, nous avons posé :

$$F_{B|\alpha} = \frac{CM_{B|\alpha}}{CM_R}$$

et ainsi la puissance $\Pi_{B(A)}$ est

$$\Pi_{B|\alpha} = \mathbb{P}\Big[F(a(b-1), abc(n-1)) > \lambda^{-2} F(a(b-1), abc(n-1); 1-\alpha)\Big]$$

où le paramètre de non centralité λ est $\lambda = \sqrt{1 + \frac{cn\sigma^2_{B|\alpha}}{\sigma^2_e}}$.

Puissances des tests

Exemple

3. Pour le test de l'effet principal C nous avons posé :

$$F_\gamma = \frac{CM_\gamma}{CM_{B\gamma|\alpha}}$$

et ainsi la puissance $\Pi_{B(A)}$ est

$$\Pi_\gamma = \mathbb{P}\left[F'(c-1, a(b-1)(c-1); \lambda) > F(c-1, a(b-1)(c-1); 1-\alpha)\right]$$

avec $\lambda = \dfrac{abn(c-1)\Phi(\gamma)}{2(\sigma_e^2 + n\sigma_{B\gamma|\alpha}^2)}$ et $\phi = \sqrt{\dfrac{abn(c-1)\Phi(\gamma)}{c(\sigma_e^2 + n\sigma_{B\gamma|\alpha}^2)}}.$

Puissances des tests

Exemple

4. Pour le test de l'effet de l'interaction $A * C$ nous avons posé :

$$F_{\alpha\gamma} = \frac{CM_{\alpha\gamma}}{CM_{B\gamma|\alpha}}$$

et ainsi la puissance $\Pi_{\alpha\gamma}$ est

$$\Pi_{\alpha\gamma} = \mathbb{P}\left[F'((a-1)(c-1), a(b-1)(c-1); \lambda) > F((a-1)(c-1), a(b-1)(c-1); 1-\alpha)\right]$$

Puissances des tests

Exemple

4. avec

$$\lambda = \frac{bn(a-1)(c-1)\Phi(\alpha\gamma)}{2(\sigma_e^2 + n\sigma_{B\gamma|\alpha}^2)}$$

et

$$\phi = \sqrt{\frac{bn(a-1)(c-1)\Phi(\alpha\gamma)}{((a-1)(c-1)+1)(\sigma_e^2 + n\sigma_{B\gamma|\alpha}^2)}}.$$

Puissances des tests

Exemple

5. Pour le test de l'effet de l'interaction $B * C(A)$ nous avons posé :

$$F_{B\gamma|\alpha} = \frac{CM_{B\gamma|\alpha}}{CM_R}$$

et ainsi la puissance $\Pi_{B(A)}$ est

$$\Pi_{B\gamma|\alpha} = \mathbb{P}\Big[F(a(b-1)(c-1), abc(n-1)) > \lambda^{-2} \times F(a(b-1)(c-1), abc(n-1); 1-\alpha)\Big]$$

où le paramètre de non centralité λ est $\lambda = \sqrt{1 + \dfrac{n\sigma^2_{B\gamma|\alpha}}{\sigma^2_e}}$.

Sommaire

Approximation de Satterthwaite

Contexte

Pour $1 \leqslant i \leqslant p$, considérons CM_i et ν_i les carrés moyens et les degrés de liberté correspondants d'un modèle d'analyse de la variance, c'est-à-dire $CM_i = SC_i/\nu_i$ où SC_i est une somme de carrés de variables aléatoires, les SC_i sont indépendantes et ν_i est un entier naturel strictement positif tels que :

$$\frac{\nu_i CM_i}{\sigma_i^2} \sim \chi^2_{\nu_i},$$

où $\chi^2_{\nu_i}$ est la loi du chi-deux, centrée, à ν_i degrés de liberté.

Approximation de Satterthwaite

Contexte

Considérons la combinaison linéaire des carrés moyens CM_i donnée par l'expression suivante :

$$\eta = \sum_{i=1}^{p} l_i CM_i.$$

Approximation de Satterthwaite

Définition

L'approximation de Satterthwaite consiste à approcher la distribution de la variable :

$$\frac{\nu\eta}{\sum_{i=1}^{p} l_i\sigma_i^2} \tag{1}$$

par celle de g qui suit une loi du chi-deux à ν degrés de liberté où la valeur de ν est donnée au transparent suivant.

Approximation de Satterthwaite

Définition

La valeur ν est obtenue en faisant coïncider les deux premiers moments de g et de la variable définie en (1) :

$$\nu = \frac{\left(\sum_{i}^{p} l_i \sigma_i^2\right)^2}{\sum_{i}^{p} \frac{l_i^2 \sigma_i^4}{\nu_i}}. \tag{2}$$

Approximation de Satterthwaite

Espérance de (1)

1. $g \sim \chi^2_\nu$ donc $\mathbb{E}[g] = \nu$;
2. En ce qui concerne la variable η nous avons pour tout $\nu > 0$:

$$\begin{aligned}
\mathbb{E}\left[\nu\eta \Big/ \left(\sum_{i=1}^{p} l_i\sigma_i^2\right)\right] &= \nu\sum_{i=1}^{p} l_i\mathbb{E}\left[CM_i\right] \Big/ \left(\sum_{i=1}^{p} l_i\sigma_i^2\right) \\
&= \nu\sum_{i=1}^{p} l_i\sigma_i^2 \Big/ \left(\sum_{i=1}^{p} l_i\sigma_i^2\right) \\
&= \nu.
\end{aligned}$$

Les espérances de g et de la variable définie en (1) sont donc égales quelque soit la valeur de $\nu > 0$.

Approximation de Satterthwaite

Variance de (1)

1. $g \sim \chi^2_\nu$ donc $\mathrm{Var}[g] = 2\nu$;
2. En ce qui concerne la variable η, comme les CM_i nous avons pour tout $\nu > 0$:

$$\mathrm{Var}\left[\nu\eta\Big/\left(\sum_{i=1}^{p} l_i\sigma_i^2\right)\right] = \nu^2\sum_{i=1}^{p} l_i^2\mathrm{Var}\left[CM_i\right]\Big/\left(\sum_{i=1}^{p} l_i\sigma_i^2\right)^2$$

Par indépendance des CM_i

$$= \nu^2\sum_{i=1}^{p} l_i^2\frac{2\sigma_i^4}{\nu_i}\Big/\left(\sum_{i=1}^{p} l_i\sigma_i^2\right)^2.$$

Nous en déduisons la valeur de ν.

Approximation de Satterthwaite

Estimation de ν

Les valeurs σ_i^2 sont en général inconnues. Nous les remplaçons alors par leurs estimations sans biais cm_i obtenant ainsi un premier estimateur de ν :

$$\widehat{\nu} = \frac{\left(\sum_{i}^{p} l_i cm_i\right)^2}{\sum_{i}^{p} \frac{l_i^2 cm_i^2}{\nu_i}}.$$

Approximation de Satterthwaite

Estimation de ν

En procédant ainsi, nous utilisons un estimateur biaisé pour estimer σ_i^4. Un estimateur sans biais de σ_i^4 serait

$$\frac{\nu_i CM_i^2}{\nu_i + 2}.$$

Ainsi un second estimateur de ν, vraisemblablement meilleur, serait :

$$\widetilde{\nu} = \frac{\left(\sum_i^p l_i cm_i\right)^2}{\sum_i^p \frac{l_i^2 cm_i^2}{\nu_i + 2}}.$$

Approximation de Satterthwaite

Application à la construction de pseudo test F

L'approximation de Satterhwaite permet de construire un pseudo test F de Fisher dans les modèles d'analyse de la variance pour lesquels un test F exact ne peut s'obtenir comme rapport de deux carrés moyens.

Dans ces cas, il est possible d'utiliser une combinaison linéaire de carrés moyens au numérateur, au dénominateur ou aux deux endroits de telle sorte que les espérances du numérateur et du dénominateur soient égales sous l'hypothèse nulle $\mathcal{H}_0$.

Approximation de Satterthwaite

Application à la construction de pseudo test F

Dans la plupart des situations, il n'est pas nécessaire d'approcher à la fois le numérateur et le dénominateur lors de la construction de la statistique de test. Toutefois, procéder ainsi permet d'être assuré de ne considérer que des combinaisons linéaires à coefficients positifs des carrés moyens et de ce fait d'éviter de soustraire des carrés moyens ce qui résulte généralement en une approximation de mauvaise qualité.

Pour plus de détails sur ce problème, le lecteur pourra consulter le livre *Analysis of Variance for Random Models, Volume I: Balanced data* de Hardeo Sahai et Miguel Ojeda pages 417 et 418.

Approximation de Satterthwaite

Application à la construction de pseudo test F

Considérons par exemple :

$$CM' = l_r CM_r + \cdots + l_s CM_s$$

et

$$CM'' = l_u CM_u + \cdots + l_v CM_v$$

où les carrés moyens sont choisis tels que $\mathbb{E}[CM'] = \mathbb{E}[CM'']$ sous l'hypothèse nulle de l'absence de l'effet de l'un des termes du modèle.

Approximation de Satterthwaite

Application à la construction de pseudo test F

Ainsi, un pseudo test de Fisher pour cette hypothèse nulle $\mathcal{H}_0$ s'obtiendrait avec la statistique :

$$F = \frac{CM'}{CM''}$$

qui, sous l'hypothèse nulle $\mathcal{H}_0$, a une distribution proche d'une loi de Fisher avec ν' et ν'' degrés de liberté. Les valeurs de ν' et ν'' sont précisées au transparent suivant.

Approximation de Satterthwaite

Application à la construction de pseudo test F

Les nombres réels ν' et ν'' sont définis par les relations suivantes :

$$\nu' = \frac{(l_r CM_r + \cdots + l_s CM_s)^2}{\dfrac{l_r^2 CM_r^2}{\nu_r} + \cdots + \dfrac{l_s^2 CM_s^2}{\nu_s}}$$

et

$$\nu'' = \frac{(l_u CM_u + \cdots + l_v CM_v)^2}{\dfrac{l_u^2 CM_u^2}{\nu_u} + \cdots + \dfrac{l_v^2 CM_v^2}{\nu_v}}.$$

Approximation de Satterthwaite

Application aux intervalles de confiance

L'approximation de Satterthwaite est également souvent utilisée pour construire des intervalles de confiance pour la moyenne et la variance de composants de variance dans des modèles à effets aléatoires ou à effets mixtes.

Approximation de Satterthwaite

Application aux intervalles de confiance

Par exemple, si un composant de variance σ^2 est estimé par la combinaison linéaire $CM = \sum_{i=1}^{p} l_i CM_i$, alors un intervalle de confiance de niveau $100(1-\alpha)\%$ pour σ^2 est donné par :

$$\frac{CM}{\chi^2_{\nu,\alpha/2}} < \sigma^2 < \frac{CM}{\chi^2_{\nu,1-\alpha/2}},$$

où $\chi^2_{\nu,\alpha/2}$ et $\chi^2_{\nu,1-\alpha/2}$ sont les quantiles d'ordre $\alpha/2$ et $1-\alpha/2$ d'une loi du chi-deux à ν degrés de liberté où ν est calculé à l'aide de la formule (2).