Compléments sur l'analyse de la variance à un facteur

Myriam Maumy et Frédéric Bertrand[1]

[1]IRMA, Université de Strasbourg
Strasbourg, France

Module 2 du DU
18-06-2013

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

Sommaire

Cadre

Comme nous l'avons vu, l'analyse de la variance à un facteur se base sur les conditions d'application imposant

1. l'indépendance des variables « erreurs »,
2. la normalité des variables « erreurs »,
3. l'homogénéité des variances des variables « erreurs ».

L'indépendance des variables « erreurs »

Les sujets doivent être répartis aléatoirement dans les groupes. Enfreindre la condition d'application d'indépendance des variables « erreurs » nuit gravement à la santé de l'analyse de la variance.

Lorsque nous nous trouvons dans une situation où un même sujet aura subi plusieurs traitements, ou un même sujet aura été vu plusieurs fois, nous serons alors dans le cas où il faudra utiliser les plans à mesures répétées. Ce sujet sera traité dans ce module dans deux jours.

La normalité des variables « erreurs »

En ce qui concerne la condition d'application de normalité des variables « erreurs », nous sommes tout aussi stricts : il ne faut pas l'enfreindre.

Il faut savoir que d'autres techniques statistiques existent lorsque la condition de normalité n'est pas vérifiée.
Par exemple, nous pouvons envisager des transformations normalisantes. Nous renvoyons le lecteur au paragraphe suivant pour de plus amples renseignements à ce sujet.

L'homogénéité des variables « erreurs »

En ce qui concerne la condition d'application d'homogénéité, nous sommes encore aussi stricts : il ne faut absolument pas l'enfreindre.

Il faut savoir que d'autres techniques statistiques existent lorsque la condition d'homogénéité n'est pas vérifiée. Par exemple, Box (1954) a montré que dans le cas de variances hétérogènes, la distribution F adéquate à laquelle il faut comparer F_{obs} est une F régulière avec des *ddl* modifiés. Pour de plus amples détails sur cette procédure, nous renvoyons en première lecture au livre de Howell et en seconde lecture à l'article de Box. Mais l'approche de Box présente un inconvénient majeur : elle est extrêmement conservatrice. Il existe toutefois des alternatives.

Les alternatives existantes

Welch (1951) a proposé une autre approche, que nous ne présenterons pas ici par manque de temps. Le lecteur intéressé par ce sujet pourra en première lecture ouvrir le livre de Howell (sixième édition) à la page 327, puis aller lire l'article original de Welch. Il est à noter que la procédure de Welch se trouve dans la plupart des logiciels statistiques.

Enfin, à titre d'information, Wilcox (1987), dans son ouvrage « **New statistical procedures for the social sciences** » a un avis tranché sur les conséquences de l'hétérogénéité des variances. Il conseille d'utiliser la procédure de Welch, et en particulier lorsque les échantillons sont de tailles inégales.

Une dernière remarque

Lorsque l'une des deux conditions (la condition de normalité des variables erreurs ou la condition d'homogénéité des variables erreurs) n'est pas vérifiée au moyen d'un test statistique, il faut s'assurer que cela n'est pas dû à une valeur extrême ou aberrante. Par exemple, pour savoir si une des valeurs recueillies n'est pas représentative, nous pouvons par exemple utiliser **les tests de Grubbs ou de Dixon**. Pour ce sujet, le lecteur pourra consulter un cours recommandé par le rédacteur.

Sommaire

Transformations normalisantes

Il n'est pas conseillé dans un premier temps, d'utiliser les transformations normalisantes, mais plutôt d'avoir une réflexion profonde sur la nature des données à analyser et sur le modèle statistique à utiliser. En dernier recours, nous pourrons les envisager, comme nous l'avons conseillé dans le paragraphe précédent. Il en existe un certain nombre. Voici les principales :

$$
\begin{aligned}
y_i' &= \log(y_i + c) && \text{la transformation logarithmique,}\\
&= y_i^{\gamma} && \text{la transformation puissance,}\\
&= \Phi^{-1}(y_i) && \text{la transformation réciproque,}\\
&= \arcsin\left(\sqrt{y_i}\right) && \text{la transformation arc sinus}\\
& && \text{de la racine carrée,}\\
&= \frac{y_i^{\lambda}-1}{\lambda}, \quad \text{avec } \lambda > 0 && \text{la transformation Box-Cox}
\end{aligned}
$$

Remarque

Il est conseillé, au sujet de ces transformations, de lire les pages 327 à 334 du livre de Howell (sixième édition) pour savoir comment les appliquer et dans quel cas il faut utiliser celle-ci plutôt que celle-la.

Et si rien ne marche

Si nous n'avons toujours pas les conditions requises après ces transformations, il faut alors utiliser le test non paramétrique de Kruskal-Wallis. Ce test sera présenté ici ultérieurement.

Sommaire

Contexte

À l'heure actuelle, il existe au moins six mesures de la grandeur de l'effet expérimental. Elles sont toutes différentes et prétendent toutes être moins biaisées que les autres mesures. Ici, dans ce cours nous présenterons uniquement une des mesures les plus courantes : le eta carré.

Mesure de la taille de l'effet η^2

Quand le test d'égalité des moyennes est rejeté (l'hypothèse nulle $\mathcal{H}_0$ est rejetée), nous pouvons souhaiter donner une mesure de la taille de la différence entre moyennes.
Nous définissons η^2 comme le « pourcentage » de la variabilité des données Y_{ij} expliquée par la différence entre les groupes :

$$\begin{aligned} \eta^2 &= 1 - \frac{SC_R}{SC_{Tot}} \\ &= \frac{SC_F}{SC_{Tot}}. \end{aligned}$$

Exemple : D'après le livre de Georges Parreins, Techniques statistiques, moyens rationnels de choix et de décision

Nous voulons tester cinq types de carburateurs. Pour chaque type, nous disposons de six pièces que nous montons successivement en parallèle sur des voitures que nous supposons avoir des caractéristiques parfaitement identiques. Le tableau qui va suivre indique pour chaque essai la valeur d'un paramètre lié à la consommation.

Violation des conditions d'application
Transformation des variables
Grandeur de l'effet expérimental
Puissance
Facteur à effets aléatoires

Suite de l'exemple

	Car. 1	Car. 2	Car. 3	Car. 4	Car. 5
Essai 1	21	23	18	20	22
Essai 2	24	23	19	21	20
Essai 3	25	32	28	25	24
Essai 4	20	23	19	15	21
Essai 5	34	32	24	29	27
Essai 6	17	15	14	9	26

Violation des conditions d'application
Transformation des variables
Grandeur de l'effet expérimental
Puissance
Facteur à effets aléatoires

Suite de l'exemple

Afin de réaliser le test de Fisher de l'analyse de la variance à un facteur fixe, nous allons nous assurer que les conditions du modèle linéaire sont bien vérifiées. Pour cela, nous allons réaliser les différents tests grâce au logiciel **R**. Les lignes de commande à taper sont les suivantes :

```
> car<-rep(1:5,c(6,6,6,6,6))
> car<-factor(car)
> conso<-c(21,24,25,20,34,17,23,23,32,23,32,
15,18,19,28,19,24,14,20,21,25,15,29,9,22,20,
24,21,27,26)
> modele1<-aov(conso~car)
> shapiro.test(residuals(modele1))
```


Violation des conditions d'application
Transformation des variables
Grandeur de l'effet expérimental
Puissance
Facteur à effets aléatoires

Suite de l'exemple

```
Shapiro-Wilk normality test
data: residuals(modele1)
W = 0.9726, p-value = 0.6119
```

La *p*-valeur (0,6119) du test de Shapiro-Wilk étant strictement supérieure à $\alpha = 5\%$, le test n'est pas significatif. Vous conservez donc l'hypothèse nulle $\mathcal{H}_0$. Le risque d'erreur associé à cette décision est un risque de deuxième espèce β. Vous ne pouvez pas l'évaluer dans le cas d'un test de Shapiro-Wilk.

Suite de l'exemple

Il nous reste à vérifier la condition d'homogénéité des variances. Pour cela, nous allons utiliser le test de Bartlett, qui est un test paramétrique. La ligne de commande à taper est la suivante :

```
> bartlett.test(residuals(modele1)~car)
Bartlett test of homogeneity of variances
data: residuals(modele1) by car
Bartlett's K-squared = 3.9326, df = 4,
p-value = 0.4152
```


Violation des conditions d'application
Transformation des variables
Grandeur de l'effet expérimental
Puissance
Facteur à effets aléatoires

Suite de l'exemple

La *p*-valeur (0,4152) du test de Bartlett étant strictement supérieure à $\alpha = 5\%$, le test n'est pas significatif. Vous conservez donc l'hypothèse nulle $\mathcal{H}_0$. Le risque d'erreur associé à cette décision est un risque de deuxième espèce β. Vous ne pouvez pas l'évaluer dans le cas d'un test de Shapiro-Wilk.
Nous pouvons donc désormais construire le tableau de l'analyse de la variance et réaliser le test de Fisher.

Violation des conditions d'application
Transformation des variables
Grandeur de l'effet expérimental
Puissance
Facteur à effets aléatoires

Suite de l'exemple

Nous avons donc le tableau de l'analyse de la variance suivant :

Variation	*SC*	*ddl*	*CM*	F_{obs}
Due au facteur	108,3333	4	27,08332	0,8502509
Résiduelle	796,3333	25	31,85333	
Totale	904,6666	29		

Comme $F_{obs} < F_c(2,75871)$, le test de Fisher n'est pas significatif. Nous décidons de ne pas rejeter l'hypothèse nulle ($\mathcal{H}_0$) et par conséquent de l'accepter. Le risque d'erreur associé à cette décision est un risque de deuxième espèce β qu'il faudra évaluer (voir le paragraphe suivant).

Violation des conditions d'application
Transformation des variables
Grandeur de l'effet expérimental
Puissance
Facteur à effets aléatoires

Suite de l'exemple

Nous calculons η^2 à titre d'exemple. En principe, nous le calculerons lorsque le test est significatif :

$$\eta^2 = \frac{108,3333}{904,6666} \simeq 0,1197.$$

Remarque

Nous retrouvons cette valeur en face de `Multiple R-squared`, résultat obtenu en tapant les deux lignes de commande suivantes :

```
> modele2<-lm(conso car)
> summary(modele2)
```

Rappel

Dans l'analyse de la régression linéaire simple, nous utilisons le coefficient de détermination R^2 pour mesurer le pourcentage de la variance de la variable Y expliquée par le modèle. Rappelons ici sa définition :

$$R^2 = 1 - \frac{SC_R}{SC_{Tot}} = \frac{SC_{Regression}}{SC_{Tot}}.$$

Cette égalité ressemble beaucoup à celle qui définit le eta carré. Nous pouvons donc faire un parallèle entre ces deux mesures.

Sommaire

Cas de l'analyse de la variance à un facteur fixe

Nous nous intéressons à la puissance $1-\beta$, où β est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse nulle

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

contre l'hypothèse alternative

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

Calcul de la puissance

Cette puissance $1-\beta$ est donnée par la formule suivante :

$$1-\beta = \mathbb{P}\left[F'(I-1, I(J-1); \lambda) > F(I-1, I(J-1); 1-\alpha)\right],$$

où $F(I-1, I(J-1); 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $I-1$ et $I(J-1)$ degrés de liberté et $F'(I-1, I(J-1); \lambda)$ est une variable aléatoire qui suit une loi de Fisher non-centrale à $I-1$ et $I(J-1)$ degrés de liberté et de paramètre de non-centralité λ.

Calcul de la puissance - Suite

Ce paramètre de non-centralité λ est égal à :

$$\lambda = \frac{J}{2\sigma^2}\sum_{i=1}^{I}\alpha_i^2,$$

où J désigne la taille de chaque échantillon, I le nombre de modalités du facteur étudié, σ^2 la variance de la population et α_i les paramètres présents dans l'équation du modèle statistique.

Violation des conditions d'application
Transformation des variables
Grandeur de l'effet expérimental
Puissance
Facteur à effets aléatoires
Puissance a posteriori
Détermination du nombre de répétitions

Calcul du paramètre ϕ dans le cas équilibré

Lorsque nous utilisons une loi de Fisher non centrale à ν_1 et ν_2 *ddl* et de paramètre de non-centralité λ, nous introduisons le paramètre de non-centralité normalisé ϕ défini par :

$$\phi = \sqrt{\frac{2\lambda}{\nu_1 + 1}}.$$

Dans notre cas $\nu_1 = I - 1$, et donc nous obtenons après substitution et simplifications :

$$\phi = \frac{1}{\sigma}\sqrt{\frac{J}{I}\sum_{i=1}^{I}\alpha_i^2}.$$

Violation des conditions d'application
Transformation des variables
Grandeur de l'effet expérimental
Puissance
Facteur à effets aléatoires
Puissance a posteriori
Détermination du nombre de répétitions

Exemple : D'après le livre de Georges Parreins, Techniques statistiques, moyens rationnels de choix et de décision

Nous allons calculer la puissance *a posteriori* de l'analyse de la variance des carburateurs. Nous calculons dans un premier temps le paramètre de non-centralité λ.

$$\begin{aligned}\lambda &= \frac{6}{2\sigma^2}\sum_{i=1}^{5}\alpha_i^2 \\ &= \frac{3}{31,85338}\times 18,05478 \\ &= 1,700427\end{aligned}$$

Violation des conditions d'application
Transformation des variables
Grandeur de l'effet expérimental
Puissance
Facteur à effets aléatoires
Puissance a posteriori
Détermination du nombre de répétitions

Suite de l'exemple

Maintenant, il reste à calculer cette puissance en se servant de la formule :

$$1-\beta = \mathbb{P}\left[F'(I-1, I(J-1); \lambda) > F(I-1, I(J-1); 1-\alpha)\right].$$

Calculons d'abord $F(I-1, I(J-1); 1-\alpha)$.
Pour cela, il suffit de taper la ligne de commande suivante :

```
> qf(0.95,4,25)
```

et nous obtenons $F(4, 25; 0, 95) = 2, 75871$.

Suite et fin de l'exemple

Ensuite, il est facile de calculer $\beta = \mathbb{P}\left[F'(4,25;1,700427) < F(4,25;0,95)\right]$ et d'en déduire $1-\beta$.
Pour cela, il suffit de taper la ligne de commande suivante :

```
pf(qf(0.95,4,25),4,25,ncp=1.700427)
```

et nous obtenons $\beta = 0,8678891$.
Par conséquent, la puissance est égale à $0,1321109$, qui est une puissance très faible.

Calcul du paramètre ϕ dans le cas déséquilibré

Si le nombre de répétitions n_i effectué pour chaque modalité i du facteur α n'est pas constant, c'est-à-dire si le plan expérimental n'est pas équilibré, le paramètre de non-centralité λ devient :

$$\lambda = \frac{1}{2\sigma^2}\sum_{i=1}^{I} n_i \alpha_i^2.$$

Le paramètre de non-centralité normalisé ϕ est alors :

$$\phi = \frac{1}{\sigma}\sqrt{\frac{1}{I}\sum_{i=1}^{I} n_i \alpha_i^2}.$$

Remarques

1. Il faut avoir à l'esprit que nous sommes dans l'impossibilité de calculer exactement le paramètre ϕ ou le paramètre λ. (Il y a cette relation que nous venons d'exposer qui lie les deux paramètres.)
   Au mieux, nous serons capable de donner une estimation de ϕ car nous ne pourrons jamais connaître la variance σ^2 de la population.
2. Il est d'usage de travailler sur ϕ car les abaques que nous allons utiliser pour calculer les puissances se servent du paramètre ϕ et non du paramètre λ.

Puissance *a posteriori*

Nous obtenons la puissance ***a posteriori*** du test de l'absence d'effet du facteur α en remplaçant dans la formule appropriée ci-dessus. Le choix se fait en fonction du fait que le plan expérimental est équilibré ou non, les valeurs des paramètres par les estimations que nous avons obtenues en réalisant l'analyse de la variance. Généralement nous considérons qu'une puissance de 0,8 est satisfaisante et qu'alors la décision de ne pas rejeter l'hypothèse nulle ($\mathcal{H}_0$) est « vraiment » associée à l'absence d'effet du facteur considéré.

Détermination du nombre de répétitions

Une autre approche celle développée lorsque nous faisons de la planification expérimentale, serait de déterminer *a priori* le nombre de répétitions J nécessaires pour obtenir une valeur de puissance du test supérieure à un niveau fixé à l'avance.
L'intérêt de cette démarche réside dans le fait que nous ne connaissons pas a priori si le test que nous allons réaliser une fois que les expériences ont été réalisées sera significatif ou non à un seuil α fixé à l'avance.
Le fait de ne pas rejeter l'hypothèse nulle ($\mathcal{H}_0$) en ayant un risque élevé de commettre une erreur de deuxième espèce rendrait cette décision très peu fiable et ne permettrait pas de conclure avec une confiance suffisante à l'absence d'un effet du facteur étudié sur la réponse.

C'est pourquoi dans de nombreux domaines comme les études cliniques, où les expériences peuvent durer plusieurs années, il est primordial de s'assurer que si une différence existe il y aura un faible risque de ne pas la mettre en évidence. Généralement nous considérons qu'une puissance de 0,8 est satisfaisante ; dans certains cas nous visons même une puissance de 0,9.

Nous pouvons utiliser directement la formule ci-dessus pour déterminer le nombre de répétitions nécessaires à l'obtention d'une valeur minimale de puissance. Il faut néanmoins avoir une idée de la valeur minimale que peut prendre la somme $\sum_{i=1}^{I} n_i \alpha_i^2$ et la valeur maximale que peut avoir σ^2. Ces valeurs doivent être déterminées par un expert du domaine considéré.

Détermination du nombre de répétitions à l'aide de la plus petite différence détectable

Dans ce type d'étude prospective, la situation est compliquée par le fait qu'il est difficile d'évaluer le terme $\sum_{i=1}^{I} \alpha_i^2$. Nous introduisons alors le concept de plus petite différence détectable Δ, ce qui revient à évaluer la sensibilité du test en terme d'amplitude entre les effets des différents niveaux du facteur étudié. Ainsi nous chercherons à ce que la probabilité de détecter une amplitude $|\alpha_i - \alpha_j|$ entre les effets α_i et α_j de deux modalités i et j différentes du facteur α étudié strictement supérieure à Δ soit élevée.

Calcul de la puissance

Ainsi pour faire le calcul de la puissance nous nous plaçons dans le pire des cas, c'est-à-dire celui pour lequel tous les effets sont nuls sauf deux : α_{i_0} et α_{j_0}, pour lesquels il existe un écart en valeur absolue égal à Δ. Alors $|\alpha_{i_0}| = |\alpha_{j_0}| = \Delta/2$.
Nous obtenons alors :

$$\begin{aligned} \lambda &= \frac{J}{2\sigma^2} \sum_{i=1}^{I} \alpha_i^2 \\ &= \frac{J}{2\sigma^2} \left(\alpha_{i_0}^2 + \alpha_{j_0}^2 \right) \\ &= \frac{J}{4\sigma^2} \Delta^2. \end{aligned}$$

Calcul de la puissance - Suite et fin

Nous utilisons la formule ci-dessus pour déterminer les valeurs de J pour lesquelles la puissance $1-\beta$ est supérieure à une valeur $1-\beta_0$ fixée à l'avance, généralement 0,8 soit 80%. Remarquons que là encore il est nécessaire de connaître σ^2 ou au moins d'avoir une idée précise de la valeur de ce paramètre ce qui n'est malheureusement généralement pas le cas. Dans cette situation, nous considérons plutôt le paramètre de sensibilité Δ/σ à la place de Δ.

Sommaire

Rappel

Dans l'analyse de la variance à un facteur à effets fixes avec I modalités, nous observons pour chaque modalité du facteur n_i réalisations indépendantes d'une variable aléatoire Y. Nous savons que le modèle utilisé dans cette analyse s'écrit :

$$Y_{ij} = \mu + \alpha_i + \mathcal{E}_{ij}, \quad j = 1, \ldots, n_i, \quad i = 1, \ldots, I,$$

où $\mathcal{E}_{ij}$ sont indépendantes et $\mathcal{L}(\mathcal{E}_{ij}) = \mathcal{N}(0; \sigma^2)$. Cette variable représente l'erreur commise lors des observations, μ désigne l'effet « global » ou moyenne générale de la variable aléatoire Y et les effets α_i satisfont la contrainte : $\sum_{i=1}^{I} \alpha_i = 0$.

Un nouveau modèle

Mais ce modèle ne correspond pas toujours à la réalité. Dans certains cas, en particulier quand les modalités sont choisies au hasard, le fait de supposer que les effets sont fixes n'est pas adapté. Nous sommes amenés à considérer que chaque contribution α_i est une réalisation, indépendante des autres réalisations, d'une variable aléatoire A_i de loi $\mathcal{N}(0; \sigma_A^2)$, elle même indépendante de $\mathcal{E}$. Dans ces conditions, le modèle s'écrit :

$$Y_{ij} = \mu + A_i + \mathcal{E}_{ij}, \quad j = 1, \ldots, n_i, \quad i = 1, \ldots, I.$$

Mise en place du test de l'effet du facteur aléatoire

Nous nous proposons de tester l'hypothèse nulle

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

contre l'hypothèse alternative

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

Remarque

Ce test ne compare plus les moyennes mais teste au moyen de la variance du facteur A, si il y a un effet de ce facteur aléatoire.

Violation des conditions d'application
Transformation des variables
Grandeur de l'effet expérimental
Puissance
Facteur à effets aléatoires

Notations et propriétés

Si $\overline{Y}_i$ et $\overline{Y}$ désignent respectivement la moyenne des Y_{ij} où $j = 1, \ldots, n_i$ et la moyenne de toutes les variables Y_{ij}, un calcul simple nous montre que les lois des trois variables sont :

$$\mathcal{L}(Y_{ij}) = \mathcal{N}(\mu; \sigma^2 + \sigma_A^2), \quad \mathcal{L}(\overline{Y}_i) = \mathcal{N}\left(\mu; \frac{\sigma^2}{n_i} + \sigma_A^2\right),$$

$$\mathcal{L}(\overline{Y}) = \mathcal{N}\left(\mu; \frac{\sigma^2}{n} + \frac{\sigma_A^2}{n^2} \sum_{i=1}^{I} n_i^2\right).$$

Myriam Maumy et Frédéric Bertrand
Compléments sur l'analyse de la variance à un facteur

Violation des conditions d'application
Transformation des variables
Grandeur de l'effet expérimental
Puissance
Facteur à effets aléatoires

Tableau de l'ANOVA

Variation	SC	ddl	CM	F_{obs}	c
Due au facteur A	$\sum(\overline{y}_i - \overline{y})^2$	$I-1$	s_A^2	$\frac{s_A^2}{s_R^2}$	c
Résiduelle	$\sum(y_{ij} - \overline{y}_i)^2$	$n-I$	s_R^2		
Totale	$\sum(y_{ij} - \overline{y})^2$	$n-1$			

Remarque

Nous retrouvons strictement les mêmes formules que celles du cas de l'analyse de la variance à un facteur à effets fixes.

Myriam Maumy et Frédéric Bertrand
Compléments sur l'analyse de la variance à un facteur

Violation des conditions d'application
Transformation des variables
Grandeur de l'effet expérimental
Puissance
Facteur à effets aléatoires

Propriété

Si les trois conditions sont satisfaites et si l'hypothèse nulle $(\mathcal{H}_0)$ est vraie alors

$$F_{obs} = \frac{s_A^2}{s_R^2}$$

est une réalisation d'une variable aléatoire F qui suit une loi de Fisher à $I-1$ degrés de liberté au numérateur et $n-I$ degrés de liberté au dénominateur. Cette loi est notée $\mathcal{F}_{I-1,n-I}$.

Myriam Maumy et Frédéric Bertrand
Compléments sur l'analyse de la variance à un facteur

Décision et conclusion du test

Pour un seuil donné α (=5%=0,05 en général), les tables de Fisher nous fournissent une valeur critique c_α telle que $\mathbb{P}_{(\mathcal{H}_0)}(F(I-1, n-I) \leqslant c_\alpha) = 1-\alpha$. Si la valeur de la statistique calculée sur l'échantillon, notée F_{obs}, est supérieure ou égale à c_α, alors le test est significatif. Vous rejetez $\mathcal{H}_0$ et vous décidez que $\mathcal{H}_1$ est vraie avec un risque d'erreur de première espèce $alpha = 5\%$. Si la valeur de la statistique calculée sur l'échantillon, notée F_{obs}, est strictement inférieure à c_α, alors le test n'est pas significatif. Vous conservez $\mathcal{H}_0$ avec un risque d'erreur de deuxième espèce β qu'il faut évaluer.

Violation des conditions d'application
Transformation des variables
Grandeur de l'effet expérimental
Puissance
Facteur à effets aléatoires

Remarques importantes

1. La démarche pratique est donc la même que dans l'analyse à un facteur à effets fixes.
2. Cependant, les comparaisons multiples, lorsque l'hypothèse alternative $(\mathcal{H}_1)$ est acceptée, n'ont plus de sens et ne doivent pas être effectuées.
3. De même, le calcul de la puissance est différent.
4. De plus, la normalité des résidus ne peut plus être testée.
5. En revanche, la normalité des $Y_{ij} - \mu$, quantités qui sont estimées par $y_{ij} - \overline{y}$, peut être testée.