

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.

Sommaire

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

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 le cours qui est en ligne.

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 pourrions les envisager, comme nous l'avons conseillé dans le paragraphe précédent. Il en existe un certain nombre. Voici les principales :

$$\begin{aligned} y_i' &= \log(y_i) && \text{la transformation logarithmique,} \\ &= y_i^{\gamma} && \text{la transformation puissance,} \\ &= \Phi^{-1}(y_i) && \text{la transformation réciproque,} \\ &= \arcsin(\sqrt{y_i}) && \text{la transformation arc sinus,} \\ &= \dots \end{aligned}$$

Retour à l'exemple des laboratoires

Variation	SC	ddl	CM	F_{obs}	F_c
Due au facteur	118,467	2	59,233	9,49	3,35
Résiduelle	168,500	27	6,241		
Totale	286,967	29			

Ici nous décidons de rejeter (\mathcal{H}_0) et nous calculons η^2 :

$$\eta^2 = \frac{118,467}{286,967} \simeq 0,413.$$



Sommaire

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



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 :

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

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



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

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

contre l'hypothèse alternative

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



Remarque

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 pourrions jamais connaître la variance σ^2 de la population.

Remarque

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 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'un valeur minimale de puissance. Il faut néanmoins avoir une idée de la valeur minimale que peut prendre la somme $\sum_{i=1}^J n_i \alpha_i^2$ 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 $\sum_{j=1}^J \alpha_j^2$. Nous introduisons alors le concept de plus petite différence détectable Δ , ce qui revient à évaluer le 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 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 :

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

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 .

Exemple : D'après le livre de Georges Parreins.

On veut tester 4 types de carburateurs. Pour chaque type, on dispose de 6 pièces que l'on monte successivement en parallèle sur 4 voitures que l'on suppose avoir des caractéristiques parfaitement identiques. Le tableau indique pour chaque essai la valeur d'un paramètre lié à la consommation :

Essai/Carburateur	A ₁	A ₂	A ₃	A ₄
1	21	23	18	20
2	24	23	19	21
3	25	32	28	25
4	20	23	19	15
5	34	32	24	29
6	17	15	14	9

Sommaire

Exemple : D'après le livre de Georges Parreins

Nous voulons tester 4 types de carburateurs. Pour cela, nous avons réalisé une ANOVA à un facteur fixe et obtenu le tableau de l'ANOVA suivant :

Variation	SC	ddl	CM	F_{obs}	F_c
Due au facteur	100,83	3	33,61	0,888	3,10
Résiduelle	757,00	20	37,85		
Totale	857,83	23			

- 1 Violation des conditions d'application
- 2 Transformation des variables
- 3 Grandeur de l'effet expérimental
- 4 Puissance
 - Puissance a posteriori
 - Détermination du nombre de répétitions

5 Facteur à effets aléatoires

Rappel

Dans l'analyse de la variance à un facteur à effets fixes avec l 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 :

$$Y_{ij} = \mu + \alpha_j + \mathcal{E}_{ij}, \quad j = 1, \dots, n_i, \quad i = 1, \dots, l,$$

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_j satisfont la contrainte $\sum_{j=1}^l \alpha_j = 0$.

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_j est une réalisation, indépendante des autres réalisations, d'une variable aléatoire A_j de loi $\mathcal{N}(0; \sigma_A^2)$, elle même indépendante de \mathcal{E} . Dans ces conditions le modèle s'écrit :

$$Y_{ij} = \mu + A_j + \mathcal{E}_{ij}, \quad j = 1, \dots, n_i, \quad i = 1, \dots, l.$$

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.

Notations et propriétés

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

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

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

Tableau de l'ANOVA

Variation	SC	ddl	CM	F_{obs}	c
Due au facteur A	$\sum (\bar{Y}_i - \bar{Y})^2$	$l - 1$	cm_A	$\frac{cm_A}{cm_R}$	c
Résiduelle	$\sum (y_{ij} - \bar{Y}_i)^2$	$n - l$	cm_R		
Totale	$\sum (y_{ij} - \bar{Y})^2$	$n - 1$			

Remarque :

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

Propriété

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

$$F_{obs} = \frac{cm_A}{cm_R}$$

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

Décision

Pour un seuil donné α ($=5\%=0,05$ en général), les tables de Fisher nous fournissent une valeur critique c telle que

$\mathbb{P}_{(\mathcal{H}_0)}(F \leq c) = 1 - \alpha$. Alors nous décidons :

$$\begin{cases} \text{si } F_{obs} < c & (\mathcal{H}_0) \text{ est vraie,} \\ \text{si } c \leq F_{obs} & (\mathcal{H}_1) \text{ est vraie.} \end{cases}$$



Des 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 erreurs ne peut plus être testée.
- 5 En revanche, la normalité des $Y_{ij} - \mu_i$, quantités qui sont estimées par $y_{ij} - \bar{y}_i$, peut être testée.

