

# REVUE DE STATISTIQUE APPLIQUÉE

MOUNIR MESBAH

PIERRE DUCIMETIERE

## **Un test simple de comparaison de deux traitements dans un modèle hiérarchique mixte déséquilibré**

*Revue de statistique appliquée*, tome 30, n° 4 (1982), p. 5-12

[http://www.numdam.org/item?id=RSA\\_1982\\_\\_30\\_4\\_5\\_0](http://www.numdam.org/item?id=RSA_1982__30_4_5_0)

© Société française de statistique, 1982, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme  
Numérisation de documents anciens mathématiques  
<http://www.numdam.org/>

# UN TEST SIMPLE DE COMPARAISON DE DEUX TRAITEMENTS DANS UN MODELE HIERARCHIQUE MIXTE DESEQUILIBRE (1)

MESBAH Mounir (\*), DUCIMETIERE Pierre(\*)

## RESUME :

Dans un modèle hiérarchique à deux facteurs avec effets mixtes, le test F classique de comparaison des traitements n'est plus qu'approché quand le plan est déséquilibré. Un test alternatif basé sur l'estimation des composantes de variances est construit. Dans le but de comparer ces deux tests par simulation, une "mesure" du déséquilibre d'un plan est proposé. Le nouveau test semble être moins sensible au déséquilibre que le test F classique.

## I. INTRODUCTION

Dans le but de comparer deux ou plusieurs traitements, il est souvent nécessaire (par exemple pour éviter les "contagions") de randomiser les blocs et non les unités. Un traitement est affecté à chaque bloc et ainsi toutes les unités d'un même bloc reçoivent le même traitement.

L'analyse statistique est facile dans le cas équilibré. Par contre, quand le nombre d'observations par bloc diffère d'un bloc à l'autre, il n'existe plus de test exact et a fortiori optimal pour comparer les traitements.

En raison de sa simplicité, le test F conventionnel est généralement utilisé, même s'il n'est plus qu'approché, dans le cas non équilibré. Dans cet article nous proposons un test alternatif simple et comparons ses propriétés à celles du test F approché et ce par des simulations, dans le cas de deux traitements. Nous définissons d'abord une "mesure du déséquilibre" adapté au modèle et à notre problème de test.

*Modèle* : L'étude est limitée au cas simple de deux traitements (témoins-traités). La généralisation à plus de deux traitements est directe.

Le modèle s'écrit :

$$Y_{ijk} = t_i + B_{ij} + E_{ijk} \quad \text{avec } i = 1, 2 ; j = 1, \dots, m_i ; K = 1, \dots, n_{ij}.$$

---

(1) Communication présentée aux journées "Biométrie et Epidémiologie" de la Société Française de Biométrie à Nancy, en juin 1981, sous le titre : "Recherche d'un test simple de comparaison de deux traitements dans un modèle mixte déséquilibré".

(\*) Institut National de la Santé et de la Recherche médicale. Unité 169 Méthodes Statistiques et Epidémiologiques, 16 bis, avenue Paul VAILLANT Couturier, 94800 Villejuif.

$t_i$  est l'effet fixe du au traitement  $i$ .

$B_{ij} \sim N(0, \sigma_b^2)$  est l'effet aléatoire du bloc  $j$  qui reçoit le traitement  $i$  et  $E_{ijk} \sim N(0, \sigma^2)$  est l'erreur aléatoire. Les  $E_{ijk}$  et les  $B_{ij}$  sont tous supposés indépendamment distribués. Définissons  $\rho = \sigma_b / \sigma_b^2$ , l'intensité de l'effet bloc.

Le tableau d'analyse de la variance habituel est :

Source de variation	D.D.L.	Carrés moyens	Espérance du C.M.
Taitement	1	$\Sigma(y_{i..} - y_{...})^2 = MST$	$\sigma^2 + k_1 \sigma_b^2 + \phi(t)$
Entre blocs	$m_0 - 2$	$\Sigma(y_{ij.} - y_{i..})^2 / m_0 - 2 = MSB$	$\sigma^2 + k_2 \sigma_b^2$
Résiduelle	$n - m_0$	$\Sigma(y_{ijk} - y_{ij.})^2 / n - m_0 = MSR$	$\sigma^2$

avec

$$m_0 = \sum_{i=1}^2 m_i ; n_{i0} = \sum_{j=1}^{m_i} n_{ij} ; n = \sum_{i=1}^2 n_{i0} ; \phi(t) = \sum_{i=1}^2 n_{i0} (t_i - t.)^2$$

$$t. = \frac{n_{10} t_1 + n_{20} t_2}{n} ; k_1 = \sum_{i=1}^2 \frac{\sum_{j=1}^{m_i} n_{ij}^2}{n_{i0}} - \frac{\sum_{ij} n_{ij}^2}{n} ; k_2 = \frac{n - \sum_{ij} \frac{n_{ij}^2}{n_{i0}}}{m_0 - 2}$$

Les points remplaçant les indices représentent les moyennes dans les blocs ou groupes correspondants.

Les composantes de variances  $\sigma^2$  et  $\sigma_b$  sont les paramètres nuisibles du problème de test de  $H_0 : \phi(t) = 0$  ou  $t_1 = t_2$ .

Des estimateurs sans biais  $\hat{\sigma}^2 = MSR$  et  $\hat{\sigma}_b^2 = (MSB - MSR)/k_2$  peuvent être obtenu du tableau d'Analyse de Variances (estimateurs d'HENDERSON). Mais d'autres estimateurs peuvent être proposés (Minque, maximum de vraisemblance, etc.).

Dans le cas équilibré, le test optimal est basé sur la statistique  $U^2 = MST/MSB$  qui suit sous  $H_0$  une loi de FISHER à 1 et  $m_0 - 2$  degrés de liberté. Si le plan est "partiellement déséquilibré", avec  $n_{ij} = r, \forall i, \forall j$  mais  $m_1 \neq m_2$ , la même statistique peut être utilisée pour obtenir un test exact.

Dans le cas déséquilibré, le test le plus courant utilise la même statistique  $U^2$  et suppose qu'elle est distribuée sous  $H_0$  suivant un  $F(1, m_0 - 2)$ , ce qui est incorrect pour trois raisons : (Tietjen 1974)

1. Les deux carrés moyens ne sont plus indépendants ;
2. Ils ne suivent pas les lois de Khi-deux ;
3. Leur espérance sous  $H_0$  sont différentes ( $k_1 \neq k_2$ ).

Une modification du test F approché a été proposée (EISEN, 1966) dans laquelle MSB est modifiée de manière à ce que le numérateur et le dénominateur de  $U^2$  aient la même espérance sous  $H_0$ . On prend  $MSB' = \hat{\sigma}^2 + k_1 \hat{\sigma}_b^2$ . Les degrés de liberté de  $MSB'$  sont calculés grâce à la formule de Satterthwaite (GAYLOR et HOPPER, 1969). Mais en fait cette formule n'est valable que si  $MSB'$  est non négative ce qui n'est pas le cas ici. Des simulations (TIETJEN, 1974) ont montré que cette procédure n'était pas meilleure que le test F conventionnel.

## II. LE TEST ALTERNATIF

Quand  $\rho$  est connu, le vecteur  $Y$  des observations est tel que  $E(Y) = X_0 t$  et  $\text{Var } Y = \sigma^2 H = \sigma^2 (\rho X_1 X_1' + I_n)$  avec :

$$X_0 = \sum_{i=1}^2 J_{n_{i0}} \quad \text{et} \quad X_1 = \sum_{ij} J_{n_{ij}},$$

$\Sigma^+$  est le symbole de la somme directe de matrices et  $J_n$  est un  $n$ -vecteur dont tous les éléments sont égaux à 1.  $I_n$  est la matrice identité d'ordre  $n$ . Le test exact de  $H_0(t_1 = t_2)$  peut être obtenu après des calculs fastidieux mais, directs, sous la forme :

$$T^2(\rho) = \frac{\alpha_{10} \alpha_{20}}{\alpha_{00}} \frac{(\hat{t}_1 - \hat{t}_2)^2}{(\text{SCR} + \text{SC}(\rho))/n - 2}$$

$T^2(\rho)$  suit sous  $H_0$  une distribution de FISHER à 1 et  $n - 2$  degrés de libertés, on a :

$$\alpha_{ij} = \left( \rho + \frac{1}{n_{ij}} \right)^{-1}; \quad \alpha_{10} = \sum_j \alpha_{00} = \sum_i \alpha_{i0}; \quad \hat{t}_1 = \sum_j \alpha_{ij} Y_{ij} / \alpha_{i0};$$

$$\text{SCR} = \text{MRS} \times (n - m_0); \quad \text{SC}(\rho) = \sum_{ij} \alpha_{ij} (Y_{ij} - \hat{t}_i)^2.$$

Quand  $\rho$  est inconnu, le test proposé utilise la statistique  $T^2(\hat{\rho})$  avec  $\rho$  remplacé par son estimateur  $\hat{\rho}$ . Après quelques calculs, on peut montrer que  $T^2(\hat{\rho})$  se réduit dans le cas équilibré à la statistique MST/MSB avec  $\hat{\rho}$  choisi comme le rapport des deux estimateurs de Henderson  $\hat{\sigma}_b^2 / \hat{\sigma}^2$ . Le même résultat reste valable quand  $\hat{\rho}$  est l'estimateur du Minque.

Dans le cas non équilibré, nous supposons que  $T^2(\hat{\rho})$  suit une loi de  $F(1, m_0 - 2)$  sous  $H_0$ . Pour des raisons de simplicité, seul le test construit à partir de l'estimateur de HENDERSON de  $\rho$  sera étudié par simulation, ou plutôt nous prendrons l'estimateur permmissible  $\hat{\rho} = \sup(0, \hat{\rho})$ .

### Remarques :

- La formule donnant  $T^2(\rho)$  se généralise aisément à plus de deux traitements et le test alternatif est immédiat.

- Si toutes les quantités  $1/n_{ij}$  sont négligeables par rapport à  $\rho$ , un test conservatif peut être fait.  $\rho_1$  étant une borne supérieure de la vraie valeur  $\rho$ , on a  $T^2(\rho_1) < T^2(\rho)$ . La région critique du test  $T^2(\rho_1) > F_{n-2}^1(\alpha)$  est incluse dans la région critique du test exact  $T^2(\rho) > F_{n-2}^1(\alpha)$ .

## III. MESURE DU DESEQUILIBRE

Il n'existe pas de mesure universelle du déséquilibre des plans. Dans notre modèle et pour notre problème de test, une mesure possible du déséquilibre,

pourrait être la “distance” de la loi de  $U^2$  à celle de  $F$ . Les deux lois étant confondues à l'équilibre. Malheureusement la distribution de  $U^2$  n'est pas simple même sous  $H_0$  et de plus son espérance est inconnue. Le rapport des espérances des carrés moyens MST et MSB semble être le seul critère utilisable.

- Définissons l'efficacité du plan par :

$$I = \frac{E_{H_0}(\text{MST})/E_{H_0}(\text{MSB}) \text{ cas équilibré}}{E_{H_0}(\text{MST})/E_{H_0}(\text{MSB}) \text{ cas non équilibré}} = \frac{1 + k_2 \rho}{1 + k_1 \rho}$$

Supposons que les deux groupes ont le même nombre d'unités :  $n_{10} = n_{20}$ .

La quantité  $\mu = k_2/k_1$ , (on peut montrer facilement que  $0 \leq \mu \leq 1$ ) borne inférieure de l'efficacité est prise comme mesure du déséquilibre du plan. C'est “l'efficacité minimale” du plan, quand  $\rho$  est inconnu.

Supposons de plus, que  $m_1 = m_2 = m$ . Définissons la taille moyenne des blocs et leur variance par  $\bar{r} = n_{10}/m = n_{20}/m = n/2m$ , et :

$$V = \frac{1}{m_0} \left( \sum_{ij} n_{ij} - m_0 \bar{r}^2 \right) \cdot c^2 = \frac{V}{\bar{r}^2}$$

est le coefficient de variation des  $n_{ij}$ . Avec ces notations  $\mu = (m\nu - 1)/m - 1$  ou  $\nu = 1/(1 + c^2)$  est une mesure de déséquilibre proposée par AHRENS et PINCUS (1981) pour un modèle d'analyse de la variance à un facteur. Ces deux mesures  $\mu$  et  $\nu$  vérifient les “propriétés d'une mesure de déséquilibre” proposés par les auteurs précédents, à l'exception que  $\mu$  n'est pas invariante par une  $k$  replication du plan. Plus précisément, on peut montrer que  $\mu$  croît par replication. Ce qui reflète le fait que la distance entre MST et MSB décroît quand le nombre de blocs croît, les autres paramètres restant constant.  $\mu$  et  $\nu$  sont asymptotiquement égaux pour  $m$  grand. L'écart entre  $k_1$  et  $k_2$  a déjà été proposé par TIETJEN (1974) comme index du déséquilibre mais sans aucune justification.

#### IV. SIMULATIONS

L'exactitude des tests  $U$  et  $T$  est comparée par simulations pour différentes valeurs des paramètres, (degrés de déséquilibre, intensité de l'effet bloc et nombre de blocs).

##### Choix des plans

Dans une première étape, l'effet du déséquilibre du plan et de l'intensité des effets blocs ont été étudiés avec un nombre de blocs par groupe fixé à 10 et un effectif moyen des blocs fixé à approximativement 10. Trois plans ont été construits avec un degré de déséquilibre croissant (tableau 1, plans 3, 4 et 5) et pour chaque plan, la valeur de l'intensité de l'effet bloc fixé successivement à 0.25, 0.5 et 1.

Dans une deuxième étape, le nombre de blocs par groupe  $m$  a été fixé à 5 puis à 20, l'effectif total du plan étant fixé à 200. Pour chaque valeur de  $m$ , deux plans avec des degrés de déséquilibre différents ont été construits (tableau 1, plans 1, 2, 6 et 7), l'intensité de l'effet bloc prise est 0,5.

TABLE 1 – Plans

Plan n°	m	n <sub>1j</sub>	n <sub>2j</sub>	μ
1	5	20 20 20 18 22	20 20 18 26 16	0.980
2	5	10 10 10 20 50	10 10 10 15 55	0.491
3	10	10 10 10 10 10 10 10 10 9 9	10 10 10 10 9 9 9 9 9 9	0.958
4	10	6 7 21 17 25 10 21 21 4 11	7 10 18 14 7 4 4 4 22 8	0.583
5	10	40 40 10 2 2 2 1 1 1 1	40 20 20 10 5 1 1 1 1 1	0.272
6	20	5 5 5 5 5 5 5 5 5 5 5 5 5 5 4 4 4 6 6 6	5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 4 4 6 6	0.989
7	20	1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 10 10 10 20 30	1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 12 12 12 18 25	0.294

Le tableau 1 montre les sept plans utilisés avec leurs mesures de déséquilibre.

Les simulations ont été faites à l'aide d'un programme Fortran standart qui génère des variables normales. Pour étudier l'exactitude des tests et la puissance, on a simulé mille fois la variable Y, sous H<sub>0</sub> et calculé les statistiques U et T( $\hat{\rho}$ ), et ce pour chaque plan, chaque valeur de  $\rho$  et chaque valeur de  $t_1 - t_2$  (pour le calcul de la puissance).

Les valeurs successives de  $t_1 - t_2$  prises sont : 0.25, 0.5 et 1.

## Résultats et discussions

### • Exactitude

On a utilisé la statistique de Kolmogorov-Smirnov pour juger l'approximation des distributions empiriques de U et T à celle d'un Student (à 8, 18 ou 38 degrés de liberté). Le tableau 2 donne pour chaque plan et valeur de  $\rho$  la déviation maximum D entre les distributions empiriques et théoriques. A titre indicatif, quand D est utilisé pour faire le test de Kolmogorov-Smirnov au niveau 5 %, la valeur critique est égale à 0.043, pour un mille échantillon.

On peut voir que quand le plan est très déséquilibré (plans 2, 5 et 7) la distribution de U diffère considérablement de celle d'un Student, contrairement à

TABLE 2 – Valeurs de D

		m = 5		m = 10			m = 20	
U	$\rho$	1	2	3	4	5	6	7
U	0.25			0.028	0.061	0.134		
T				0.030	0.036	0.033		
U	0.50	0.049	0.152	0.023	0.045	0.145	0.021	0.127
T		0.047	0.054	0.020	0.021	0.045	0.027	0.026
U	1.00			0.037	0.048	0.169		
T				0.038	0.034	0.022		

TABLE 3 – Puissance des deux tests  
 Probabilité de rejeter  $H_0$  pour T (à gauche) et U (à droite) en fonction de la différence entre les traitements

Plan N°	$\rho$	$t_1 - t_2 = 0.25$		$t_1 - t_2 = 0.50$		$t_1 - t_2 = 1.00$	
1	0.50	0.045	0.045	0.170	0.163	0.485	0.475
2	0.50	0.065		0.175		0.430	
3	0.25	0.142	0.142	0.405	0.404	0.952	0.952
	0.50	0.124	0.122	0.279	0.278	0.775	0.779
	1.00	0.064	0.067	0.168	0.174	0.504	0.507
4	0.25	0.160		0.440		0.930	
	0.50	0.125		0.285		0.770	
	1.00	0.105		0.165		0.520	
5	0.25	0.145		0.325		0.840	
	0.50	0.110		0.285		0.685	
	1.00	0.060		0.160		0.505	
6	0.50	0.120	0.120	0.515	0.520	0.960	0.965
7	0.50	0.125		0.340		0.910	

celle de T. Les deux statistiques semblent être équivalentes quand le déséquilibre est faible ( $\mu > 0.9$ ).

Par contre l'écart pour la statistique U croit avec le déséquilibre et l'intensité de l'effet bloc, alors qu'il reste stable pour T.

Cet écart décroît pour les grandes valeurs de m, particulièrement pour T. Ce qui semble indiquer une convergence relativement rapide de la distribution de T vers la loi normale, pour de grandes valeurs de m. Remarquons que pour  $m = 5$ , la déviation D est plus grande que 0.043 pour les deux statistiques. Le faible nombre de blocs par traitement apparait comme une plus grande limitation du test T proposé ici, que le degré du déséquilibre.

• **Puissance**

Le tableau 3 donne les résultats des simulations, à l'exception des puissances de U pour les plans 2, 4, 5 et 7, sa distribution sous  $H_0$  étant trop éloigné de celle d'un Student.

Pour les plans peu déséquilibrés (1, 3 et 6) les puissances des deux tests sont comparables. La puissance du test T ne semble pas très influencée par le degré de déséquilibre du plan. Comme on aurait pu le prévoir, la puissance de T décroît avec  $\rho$  (plan 3, 4 et 5).

La figure 1 montre la variation de la fonction de puissance simulée du pour  $\rho = 0.5$  et n a peu près égal à 200, en fonction du nombre de blocs.

Pour chaque valeur de m, l'estimation de la puissance du test T semble être comparable à la puissance du test de Student exact du plan équilibré correspondant (voir par exemple SCHEFFE, 1969).

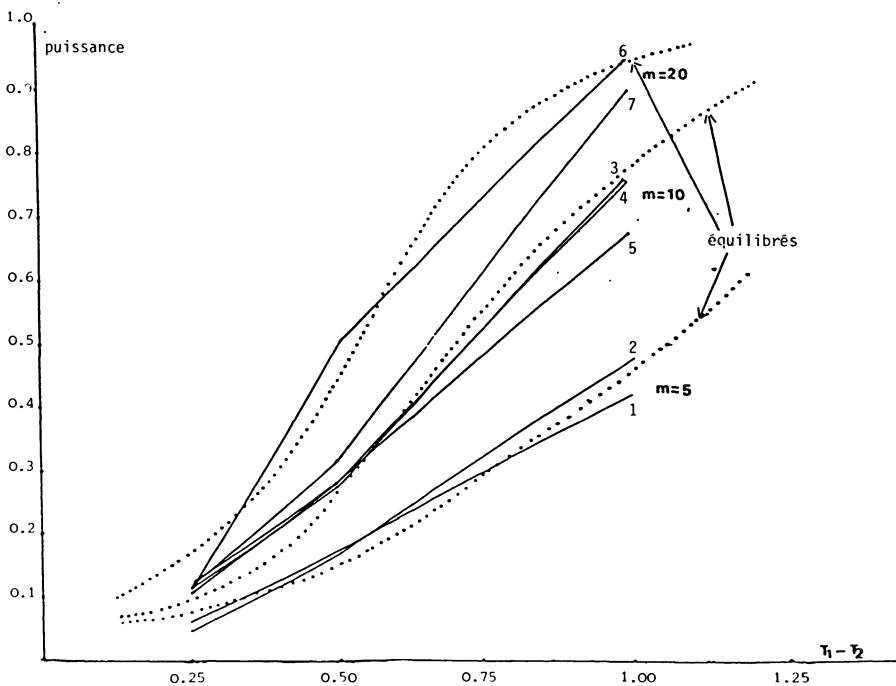


Figure 1. — Puissance du test T pour les divers plans de 1 à 7 et des plans équilibrés.



## CONCLUSION

Le test T proposé dans cet article semble être une alternative valable au test U approché, pour l'analyse d'un plan déséquilibré avec randomisation par blocs.

Si les deux tests sont équivalents quand le degré de déséquilibre est faible ( $\mu > 0.9$ ), l'exactitude du test T reste satisfaisante même si le degré de déséquilibre est élevé, du moins si le nombre de blocs par groupe n'est pas trop faible. Sa puissance ne paraît pas être influencée par le degré de déséquilibre et est comparable à celle du test exact du plan équilibré correspondant.

## REFERENCES

- H. AHRENS et R. PINCUS. — On two measures of imbalance in a one way model and their relation to efficiency. *Biom. J.* 23, n° 3, 227-235, 1981.
- E.J. EISEN. — The quasi F-test for an unnested fixed factor in an unbalanced hierarchal design with a mixed model. *Biometrics* 22, 937-942, 1966.
- O.W. GAYLOR et F.N. HOPPER. — Estimating the degrees of freedom for linear combination of mean squares by Satterthwaite's formula. *Technometrics*, 11, 691-706, 1969.
- SCHEFFE. — The analysis of variance New York John Wiley and Sons, In ,1959.
- G.L. TIEJEN. — Exact and approximate tests for unbalanced random effects designs. *Biometrics* 30, 573-581, 1974.