

REVUE DE STATISTIQUE APPLIQUÉE

A. VESSEREAU

L'analyse de la variance est-elle toujours la meilleure méthode ?

Revue de statistique appliquée, tome 36, n° 2 (1988), p. 37-52

<http://www.numdam.org/item?id=RSA_1988__36_2_37_0>

© Société française de statistique, 1988, 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/>

L'ANALYSE DE LA VARIANCE EST-ELLE TOUJOURS LA MEILLEURE MÉTHODE ?

A. VESSEREAU

1. Les hypothèses de l'analyse de la variance.
 - 1.1. L'hypothèse de normalité.
 - 1.2. L'hypothèse d'additivité.
2. Analyse non paramétrique de la méthode d'appariement.
 - 2.1. Le test des signes.
 - 2.2. Un exemple emprunté à R.A. FISHER.
3. Analyse non paramétrique de la méthode des blocs.
 - 3.1. Le test de FRIEDMAN.
 - 3.2. Deux exemples.
 - 3.2.1. Un exemple en expérimentation agronomique.
 - 3.2.2. Un exemple imaginé (donc imaginable).
4. Un test non prévu dans l'analyse de la variance.
 - 4.1. Le test de Page.
 - 4.2. Exemple.
5. Pour conclure.

Bibliographie.

Table Annexe A — Valeurs critiques de la statistique Φ de FRIEDMAN.

Table Annexe B — Valeurs critiques de la statistique P de Page.

L'analyse de la variance ⁽¹⁾ est l'un des instruments de routine qu'affectionnent, sans trop souvent se soucier de leur adéquation, beaucoup de ceux qui appliquent les méthodes statistiques à l'interprétation de résultats expérimentaux. Après avoir rappelé que, même dans les situations les plus simples, l'analyse de la variance présuppose des hypothèses généralement invérifiables, nous montrons que, dans certains cas, on peut s'affranchir de toute hypothèse en faisant appel à des tests non paramétriques (« free distribution tests »).

Nous aborderons à cet égard deux analyses très classiques :

- la « méthode d'appariement » : un seul facteur contrôlé (« traitement ») à 2 variantes et n répétitions.

(1) Traduction littérale de « analysis of variance » (ANOVA); il eut été préférable de dire « analyse de la variabilité ».

- l'analyse d'un plan factoriel : deux facteurs, dont le plus souvent l'un est un « traitement » à m variantes et l'autre un « facteur bloc » à n répétitions contrôlées.

Nous décrirons ensuite un test non paramétrique moins connu, qui convient à une situation non prévue dans l'analyse de la variance classique.

Certes, cette note n'apporte rien d'autre que de déjà connu. Si elle a quelque mérite, c'est de rassembler dans un texte court un certain nombre de méthodes éparées dans une abondante littérature, en suivant les idées directrices résumées dans l'Introduction à « Non parametric Statistical Methods » de M. HOLLANDER et D.A. WOLFE [1], que nous reproduisons ci-après :

1. Nonparametric methods require few assumptions about the underlying populations from which the data are obtained. In particular, nonparametric procedures forgo the traditional assumption that the underlying populations are normal.
2. Nonparametric techniques are often (although not always) easier to apply than their normal theory counterparts.
3. Nonparametric procedures are often quite easy to understand.
4. Nonparametric procedures are applicable in situations where the normal theory procedures cannot be utilized. For example, many of the procedures require not the actual magnitudes of the observations, but rather, their ranks.
5. Although at first most nonparametric procedures seem to sacrifice too much of the basic information in the samples, theoretical investigations have shown that this is not the case. More often than not, the nonparametric procedures are only slightly less efficient than their normal theory competitors when the underlying populations are normal (the home court of normal theory methods), and they can be middly and wildly more efficient than these competitors when the underlying populations are not normal.

1. Les hypothèses de l'analyse de la variance

1.1. L'hypothèse de normalité

Dans la méthode des blocs, que nous prendrons comme exemple de deux facteurs en combinaisons factorielles, ainsi que dans toute analyse de la variance, l'interprétation statistique utilise la « loi de F » (Snedecor); dans la méthode d'appariement, cette loi devient la loi de t^2 (carré de la variable de Student). Aussi bien t que F sont des variables dérivées de la variable normale réduite. Les tests classiques de l'analyse de la variance admettent donc implicitement l'hypothèse de normalité de ce qu'il est convenu d'appeler l'« erreur expérimentale ».

On ne s'est pas fait faute, au cours des dernières années, de brocarder la naïveté des statisticiens qui feraient de la loi de Laplace-Gauss un article de foi. Non sans raison sans doute, mais selon nous avec excès : les statisticiens « classiques » (de l'école FISHER-NEYMAN-PEARSON) se sont depuis longtemps, et à juste titre, préoccupés de la robustesse des tests qu'ils préconisent.

Les méthodes non paramétriques, qui n'utilisent que des propriétés combinatoires, constituent une autre réplique des anciens aux modernes.

1.2. L'hypothèse d'additivité

Dans l'analyse de la variance à effets fixes — la seule que nous considérons — les variantes de l'un des facteurs, appelé conventionnellement « traitement », sont représentées par m paramètres θ_i ($m = 2$ dans la méthode d'appariement); l'autre facteur (généralement facteur « bloc ») comporte n modalités représentées par n paramètres β_j .

Le modèle théorique auquel on ajuste les résultats expérimentaux, est un modèle additif :

$$X_{ij} = \theta_i + \beta_j + \varepsilon_{ij}$$

où ε_{ij} est aléatoire, distribué suivant une loi normale de variance σ^2 (variance de l'erreur expérimentale).

En plus de la normalité des ε , ce modèle admet qu'il n'y a pas d'interaction entre « effet traitement » et « effet bloc ». Cette deuxième hypothèse paraît raisonnable lorsque les traitements d'une part, les blocs d'autre part, ne sont pas trop différents; mais « pas trop » reste vague. Si, en fait, il y a interaction, celle-ci se trouve incluse (« confondue ») dans le terme résiduel de l'équation d'analyse de la variance où X_{ij} est remplacé par la valeur observée x_{ij} et les paramètres θ_i , β_j , par leurs estimations $\hat{\theta}_i$, $\hat{\beta}_j$. Une interaction éventuelle ne pourrait être mise en évidence qu'en procédant à une ou plusieurs répétition(s) de l'expérience. Mais répétitions dans quelles conditions : aussi identiques que possible, un peu différentes, largement différentes ?

On achoppe ici sur le problème (qui n'est pas propre à l'analyse classique, il existe aussi bien dans les méthodes non paramétriques) de l'extension du résultat d'une expérience unique à un univers de résultats potentiels. Problème qui se pose dès lors qu'on a l'ambition d'inférer du particulier au plus général.

2. Analyse non paramétrique de la méthode d'appariement

2.1. Le test des signes

On fait abstraction des valeurs numériques (x_i, y_i) obtenues pour les traitements (X, Y) dans les n paires, et l'on se borne à compter les nombres n^+ et n^- de différences $x_i - y_i$ positives et négatives ($n^+ + n^- = n$); on supposera dans ce qui suit qu'il n'y a pas d'ex-aequo $x_i = y_i$: s'il y en a très peu le plus simple est de supprimer les paires correspondantes, s'il y en a beaucoup, mieux vaut renoncer à interpréter les résultats.

L'hypothèse testée (« hypothèse nulle ») est que dans chacune des n paires, $+$ et $-$ apparaissent avec la même probabilité $p = 1/2$ (n parties de pile ou face). Dans un test *bilatéral*, cette hypothèse est rejetée au profit de l'alternative $X \neq Y$ si, en posant $n_+ = \max(n_+, n_-)$, n_+ est significativement supérieur à $n/2$.

Dans un *test unilatéral* où l'alternative est $X > Y$, l'hypothèse nulle est rejetée si, pour le traitement X, le nombre n^+ de différences $x_i - y_i$ positives est significativement supérieur à $n/2$. A significativement est associé un risque d'erreur α dont la valeur est choisie à l'avance.

Dans les deux cas (bilatéral et unilatéral) la valeur critique de n^+ s'obtient à partir d'une table des probabilités cumulées de la loi binomiale ($p = 1/2$, n), ou à l'aide d'un calculateur de bureau. La loi binomiale étant discontinue, on ne peut satisfaire exactement à la valeur α choisie; on convient généralement de prendre pour valeur critique la plus petite des valeurs n^+ telle que le risque réel α' soit au plus égal à α . Lorsque n est petit, α' peut être nettement inférieur à α — cet inconvénient existe dans tous les tests non paramétriques basés sur des statistiques d'ordre.

La méthode d'appariement est utilisée en particulier dans les tests organoleptiques (saveur, odeur, couleur ...) où toute mesure est généralement impossible. Les n membres d'un jury (experts qualifiés ou non suivant l'objectif poursuivi) sont invités à comparer deux produits et à exprimer leur préférence (test bilatéral de préférence) ou à reconnaître un ordre prédéterminé (test unilatéral d'identification).

2.2. Un exemple emprunté à R.A. Fisher

Malgré son ancienneté, et un style parfois assez ésotérique, « Statistical methods for research workers » [2] reste un ouvrage de référence que l'on a toujours intérêt à consulter.

L'exemple porte sur la comparaison de deux somnifères X,Y administrés à 10 sujets. Pour chacun de ceux-ci on a mesuré le nombre d'heures additionnel de sommeil obtenu par rapport à leur durée moyenne de sommeil naturel.

Les résultats ont été les suivants :

TABLEAU I

Somnifère → Sujet ↓	X x	Y y	z = y - z
1	+ 0,7	+ 1,9	+ 1,2
2	- 1,6	+ 0,8	+ 2,4
3	- 0,2	+ 1,1	+ 1,3
4	- 1,2	+ 0,1	+ 1,3
5	- 0,1	- 0,1	0,0
6	+ 3,4	+ 4,4	+ 1,0
7	+ 3,7	+ 5,5	+ 1,8
8	+ 0,8	+ 1,6	+ 0,8
9	0,0	+ 4,6	+ 4,6
10	+ 2,0	+ 3,4	+ 1,4
Moyenne	+ 0,75	+ 2,33	+ 1,58

Le test bilatéral classique des résultats appariés (colonne z du tableau) donne, pour le t de Student à 9 degrés de liberté, la valeur 4,06, comprise entre les valeurs 3,250 pour $\alpha = 0,005$ et 4,297 pour $\alpha = 0,001$. La conclusion est que Y est très significativement plus actif que X.

Si l'on applique le test des signes, le sujet 5 pose problème, et l'on décide de l'ignorer. On a alors 9 signes + sur 9 comparaisons, ce qui correspond, dans le test bilatéral ($p = 0,5$, $n = 9$) à la probabilité $1/256 = 0,0039$: la conclusion est donc la même.

Toutefois, si l'on y regarde de plus près, le test par la variable t est, comme on devait s'y attendre et comme l'écrit FISHER, « more sensitive » (nous dirions aujourd'hui « plus puissant », au risque de déplaire à FISHER).

Naturellement, le test par les signes ne dispense pas de calculer la différence des moyennes $\bar{y} - \bar{x} = 1,58$, estimation du supplément moyen d'heures de sommeil obtenu lorsqu'on substitue à X le somnifère Y.

3. Analyse non paramétrique de la méthode des blocs

3.1. Le test de Friedman

Comme dans la méthode d'appariement, on fait abstraction des valeurs numériques obtenues; à l'intérieur de chaque bloc on attribue aux traitements des rangs de classement de 1 à m, suivant les valeurs croissantes (ou décroissantes) des résultats. On obtient ainsi un tableau tel que le suivant, où le numérotage des blocs et des traitements sert simplement à les identifier.

TABLEAU II

Traitements →	T_1	----	T_i	----	T_m	Totaux
Blocs ↓						
B_1	< rangs de 1 à m >					$\frac{m(m+1)}{2}$
B_j	id					id
B_n	< rangs de 1 à m >					$\frac{m(m+1)}{2}$
Totaux	R_1		R_i		R_m	$\frac{nm(m+1)}{2}$

La moyenne des R_i étant $\bar{R} = n \frac{m+1}{2}$, il est logique de prendre comme mesure de l'accord entre les n classements la quantité :

$$S = \sum_{i=1}^m (R_i - \bar{R})^2 = \sum R_i^2 - \frac{m n^2 (m+1)^2}{4} \quad (1)$$

Si, à l'intérieur de chaque bloc, les rangs sont purement aléatoires, les totaux R_i se groupent plus ou moins étroitement autour du rang moyen $\bar{R} = n \frac{m+1}{2}$ (si $\bar{R}$ est un nombre entier), ou autour de $n \frac{m+1}{2} \pm \frac{1}{2}$ (si m est pair et n impair). La valeur minimale de S est, suivant l'un ou l'autre cas :

$$S_{\min} = 0 \quad \text{ou} \quad S_{\min} = \frac{m}{4} \quad (2)$$

A l'opposé si, dans tous les blocs, les rangs sont concordants, les totaux R_i sont, dans un ordre quelconque $n, 2n, \dots, mn$. On a alors :

$$\sum_{i=1}^m R_i^2 = \sum_{i=1}^m (i n)^2 = n^2 \sum_{i=1}^m i^2 = n^2 \frac{m(m+1)(2m+1)}{6}$$

d'où l'on déduit la valeur maximale de S

$$S_{\max} = \frac{n^2 m(m^2 - 1)}{12} \quad (3)$$

La *statistique de Friedman*, que nous désignerons par Φ , se déduit de S par la relation :

$$\Phi = \frac{12 S}{nm(m+1)} \quad (4)$$

Ses valeurs minimale et maximale sont :

$$\Phi_{\min} = 0 \quad \text{ou} \quad \frac{3}{n(m+1)} ; \quad \Phi_{\max} = n(m-1) \quad (5)$$

Par exemple, pour $m = 4$, on a les valeurs du tableau ci-dessous :

	$\Phi_{\min}$	$\Phi_{\max}$
$n = 3$	0,20	9
$n = 4$	0,—	12
$n = 5$	0,12	15
$n = 6$	0,—	18
$n = 7$	0,086	21
$n = 8$	0,—	24

Le nombre de tableaux distincts contenant dans leurs n lignes les m ! permutations des entiers de 1 à m est $(m!)^{n-1}$, nombre impressionnant dès que (m, n) ne sont pas trop petits (près de 200 millions pour $m = 4, n = 7$). Mais naturellement, le nombre de valeurs distinctes de S ou Φ est beaucoup plus petit (85 dans l'exemple précédent).

Pour $m = 3, n = 2$ [1] 15 et $m = 4, n = 2$ [1] 8 les fréquences et fréquences cumulées de la distribution de Φ sont données en détail dans « Owen-Handbook of Statistical Tables » (3).

Lorsque (n, m) sont suffisamment élevés, la distribution (discontinue) de Φ tend à se confondre avec la distribution (continue) de χ^2 à $\gamma = m - 1$ degrés de liberté; cette approximation est acceptable dès que $n > 15$ si $m = 3$, $n > 8$ si $m = 4$, $n > 5$ si $m = 5$, et quelque soit n si $m > 5$.

Dans le *test de Friedman*, l'hypothèse nulle est que les rangs figurant dans chacun des n blocs sont n permutations aléatoires des m premiers nombres entiers. Pour un risque au plus égal à une valeur α fixée, cette hypothèse est rejetée au profit de l'alternative « rangs globalement concordants » (ou « corrélation significative entre les rangs ») si la valeur obtenue pour la statistique de Friedman est au moins égale au fractile $\Phi_{1-\alpha}(m, n)$.

Pour les risques usuels $\alpha = 0,05$ ($\alpha' \leq 0,05$) et $\alpha = 0,01$ ($\alpha' \leq 0,01$), les valeurs critiques Φ sont données dans la Table A figurant en annexe de cette note (pour $m = 5$, les valeurs sont empruntés à (1)). Le cas $m = 2$ ne figure pas dans la Table, car il se ramène aux comparaisons par paires en *test bilatéral* (alors que le test de Friedman est *unilatéral*). On voit facilement qu'entre la valeur critique n_+ du test d'appariement et la valeur critique Φ du test de Friedman on a la relation :

$$\Phi = \frac{(2n^+ - n)^2}{n} = \frac{(n^+ - n^-)^2}{n}$$

Comme la méthode d'appariement, la « méthode des m classements selon Friedman » est utilisée dans les tests organoleptiques où n « sujets » sont invités à classer m « produits » dans l'ordre de leurs préférences. Mais, là encore, rien n'empêche de l'appliquer à la « méthode des blocs », et plus généralement à l'analyse d'un plan factoriel à deux facteurs contrôlés A, B : pour tester A on numérote les résultats de 1 à m pour chaque modalité de B, pour tester B on les numérote de 1 à n pour chaque modalité de A.

3.2. Deux exemples

3.2.1. Un exemple en expérimentation agronomique

Cet exemple est emprunté à [4] : comparaison en 4 blocs ($n = 4$) des rendements de 27 variétés ($m = 27$) d'orges de brasserie, à la station d'amélioration des plantes de Clermont-Ferrand. Dans ce qui suit :

- les noms, parfois compliqués, des variétés ont été remplacés par les 26 lettres de l'alphabet romain, complétées par la lettre Ω .
- l'affectation des lettres aux variétés a été choisie de façon que, dans l'analyse par le test de Friedman, les totaux de rang R_i soient en ordre non décroissant de la variété A à la variété Ω .
- lorsque, pour deux variétés, les rendements ont été identiques dans un même bloc (cas exceptionnels), on leur a attribué arbitrairement des rangs consécutifs.

Compte tenu de ces réaménagements, les tableaux III et IV ci-après donnent, d'une part les rendements obtenus, d'autre part les rangs de classement de 1 à 27 dans chaque bloc (1 correspond au rendement le plus élevé, 27 au rendement le plus faible).

Note : Le lecteur devra prendre garde que, pour des raisons d'ordre typographique, et contrairement à la présentation des précédents tableaux de cette note, les traitements (variétés) correspondent ici aux lignes des tableaux, et les répétitions (blocs) aux colonnes.

Utilisant les totaux R_i figurant dans la dernière colonne du tableau IV on a :

$$S = \sum_{i=1}^{i=27} (R_i - \bar{R})^2 = (8 - 56)^2 + (14 - 56)^2 + \dots + (106 - 56)^2 = 16\,430$$

$$\Phi = 12 \times \frac{16\,430}{4 \times 27 \times 28} = 65,20$$

Pour $m = 27$ on est très largement dans le cas où la loi de Φ est assimilable à la loi de χ^2 à 26 degrés de liberté; les valeurs critiques sont, pour $\alpha = 0,01$: 45,6, et pour $\alpha = 0,001$: 56,4. (65,2 correspond à un risque de l'ordre de 10^{-9}). On peut donc conclure avec quasi certitude qu'il existe des différences de rendement entre les 27 variétés.

TABLEAU III
(Rendements)

Variétés	Bloc				Moyennes $\bar{x}_v$
	B ₁	B ₂	B ₃	B ₄	
A	95	95	90	101	95,50
B	100	93	77	115	96,25
C	94	80	93	96	90,75
D	93	85	70	100	87,00
E	87	80	77	97	85,25
F	80	77	82	97	84,00
G	85	87	75	85	83,00
H	92	81	76	84	83,25
I	72	74	94	93	83,25
J	81	82	78	84	81,25
K	93	77	71	87	82,00
L	89	77	70	94	82,50
M	73	88	78	77	79,00
N	70	82	76	87	78,75
O	84	84	66	85	79,75
P	91	60	83	74	77,00
Q	77	72	73	83	76,25
R	83	74	73	73	75,75
S	70	69	79	77	73,75
T	78	69	72	82	75,25
U	73	73	70	79	73,75
V	60	71	83	67	70,25
W	69	60	71	80	70,00
X	71	65	62	75	68,25
Y	65	59	66	81	67,75
Z	63	66	67	68	66,00
Ω	52	53	51	68	56,00
Moyennes $\bar{x}_B$	79,30	75,30	74,93	84,78	$\bar{\bar{x}} = 78,57$

TABLEAU IV
(Classements)

Bloc	B ₁	B ₂	B ₃	B ₄	Totaux R _i
Variétés					
A	2	1	3	2	8
B	1	2	10	1	14
C	3	10	2	6	21
D	5	5	20	3	33
E	9	11	11	5	36
F	14	12	6	4	36
G	10	4	14	11	39
H	6	9	12	13	40
I	19	15	1	8	43
J	13	8	8	14	43
K	4	14	18	9	45
L	8	13	21	7	49
M	17	3	9	21	50
N	22	7	13	10	52
O	11	6	28	12	54
P	7	25	4	23	59
Q	16	18	16	15	65
R	12	16	15	24	67
S	21	20	7	20	68
T	15	21	17	16	69
U	18	17	22	19	76
V	26	19	5	27	77
W	23	24	19	18	84
X	20	23	26	22	91
Y	24	26	24	17	91
Z	25	22	23	26	96
Ω	27	27	27	25	106
Total	378	378	378	378	1 512
Rang moyen	14	14	14	14	$\bar{R} = 56$

TABLEAU V

Origine de la variation	Somme des carrés	Degrés de liberté	Carrés moyens
Blocs	1 766	3	$s_B^2 = 588,7$
Variétés	8 431	26	$s_V^2 = 324,3$
Résiduelle	3 978	78	$s_E^2 = 51,0$
Total	14 175	107	—

L'analyse de la variance classique, beaucoup plus laborieuse, est résumée dans le tableau V :

Le rapport $\frac{s_V^2}{s_E^2} = 6,35$ est très supérieur à la valeur 2,6 qui, dans la loi de F ($\gamma_1 = 26, \gamma_2 = 78$), correspond à un risque de l'ordre de 0,0005. La conclusion est donc la même que dans l'analyse non paramétrique.

Dans [4] la conclusion finale est que le « classement vraisemblable » des variétés est le suivant :

Rendements les plus élevés : B A C D

Rendements les plus faibles : V W X Y Z Ω

les 17 autres variétés étant « intermédiaires ». On aboutit à un classement quasi-identique en se basant sur les totaux de rang R_i ; seul D se trouve déclassé dans le groupe intermédiaire.

Le tableau d'analyse de la variance fait apparaître — subsidiairement — des différences hautement significatives entre les blocs.

On obtiendrait la même conclusion en appliquant le test de Friedman à un tableau déduit du tableau III, où les blocs recevraient les rangs de 1 à 4 pour chacune des 27 variétés.

3.2.2. *Un exemple imaginé (donc imaginable)*

La présentation est la même que dans les tableaux III et IV : pour 5 traitements (lignes) et 4 blocs (colonnes) données numériques dans le tableau VI, rangs des traitements de 1 à 5 à l'intérieur de chaque bloc dans le tableau VII.

Appliquant le test de Friedman on a :

$$S = (6 - 12)^2 + \dots + (17 - 12)^2 = 90 \quad \Phi = 12 \times \frac{90}{4 \times 5 \times 6} = 9,0$$

valeur significative au risque $\alpha = 0,05$ (la valeur critique est 8,8).

L'analyse de la variance classique est résumée dans le tableau VIII :

$\frac{s_T^2}{s_E^2} = 2,72$ est nettement inférieur à la valeur critique 3,26 de la variable F pour les degrés de liberté ($\gamma_1 = 4, \gamma_2 = 12$) et $\alpha = 0,05$. Contrairement au test de Friedman, et bien que moyennes et totaux de rang soient en étroite corrélation, l'analyse classique ne révèle pas de différences significatives entre les traitements.

TABLEAU VI
(Valeurs numériques)

Blocs Traitements	B ₁	B ₂	B ₃	B ₄	Moyennes
A	36	41	27	40	36,00
B	38	38	28	33	34,25
C	34	40	24	32	32,50
D	33	34	25	37	32,25
E	32	35	26	31	31,00
Moyennes	34,6	37,6	26,0	34,6	33,20

TABLEAU VII
(Classements)

Blocs Traitements	B ₁	B ₂	B ₃	B ₄	Totaux R _i
A	2	1	2	1	6
B	1	3	1	3	8
C	3	2	5	4	14
D	4	5	4	2	15
E	5	4	3	5	17
Total	15	15	15	15	60
					$\bar{R} = 12$

TABLEAU VIII

Origine de la variation	Somme des carrés	Degrés de liberté	Carré moyen
Blocs	375,6	3	$s_B^2 = 125,2$
Traitements	60,7	4	$s_T^2 = 15,175$
Résiduelle	66,9	12	$s_E^2 = 5,575$
Total	503,2	19	—

4. Un test non prévu dans l'ANALYSE de la VARIANCE

L'analyse de la variance classique ne retient contre l'hypothèse nulle que l'alternative « existence d'un ordre » (non précisé) entre les modalités du facteur traitement. Si ces modalités sont mesurables sur une échelle continue, ce n'est qu'après avoir rejeté l'hypothèse nulle qu'on peut ajuster les moyennes de traitement à un modèle de régression, linéaire le plus souvent.

Lorsqu'il existe un ordre préfixé des modalités du « traitement » — que cet ordre résulte d'un facteur objectif, mesurable ou repérable, ou d'une expérience acquise —, le test décrit ci-après (« test de Page ») permet, soit de confirmer cet ordre, ou bien de conclure que, dans les conditions expérimentales que l'on a fixées, il n'a pas été significativement reconnu.

4.1. Le test de Page

Les résultats ordonnés de 1 à m dans chaque bloc (répétition) sont inscrits dans un tableau analogue au tableau II, avec la seule différence que les colonnes sont elles-mêmes ordonnées selon l'ordre préfixé :

$$T_1 \leq T_2 \dots \leq T_i \dots \leq T_m \quad (6)$$

$R_1, R_2 \dots R_i \dots R_m$ désignant comme précédemment les totaux de colonnes, la statistique de Page est :

$$P = R_1 + 2 R_2 \dots + i R_i \dots + m R_m = \sum_{i=1}^m i R_i \quad (7)$$

La valeur maximale de P est obtenue lorsque tous les classements reproduisent exactement l'ordre fixé : $R_i = n_i$. D'où :

$$P_{\max} = \sum_{i=1}^m n_i^2 = \frac{nm(m+1)(2m+1)}{6} \quad (8)$$

A l'opposé, la valeur minimale correspond au cas où tous les classements reproduisent l'ordre inverse :

$$\begin{aligned} P_{\min} &= n[m + 2(m-1) + \dots + (i+1)(m-i) + \dots + m[m - (m-1)]] \\ &= nm \sum_{i=1}^m i - n \sum_{i=1}^m i(i-1) \end{aligned}$$

Un calcul élémentaire donne :

$$P_{\min} = \frac{nm(m+1)(m+2)}{6} \quad (9)$$

La valeur « neutre » de P (ou « valeur d'indifférence ») correspondrait au cas où tous les R_i seraient égaux au rang moyen $\bar{R} = n \frac{m+1}{2}$, ce qui ne peut avoir lieu que si $\bar{R}$ est un nombre entier; on obtient alors :

$$P_1 = nm \left(\frac{m+1}{2} \right)^2 \quad (10-a)$$

Dans le cas contraire (m pair, n impair), les « valeurs d'indifférence » sont comprises (limites incluses) entre P_{l_1} et P_{l_2} définies, comme le montre également un calcul élémentaire par :

$$\begin{aligned} P_{l_1} &= nm \left(\frac{m+1}{2} \right)^2 - \frac{m^2}{8} \\ P_{l_2} &= nm \left(\frac{m+1}{2} \right)^2 + \frac{m^2}{8} \end{aligned} \quad (10-b)$$

dont la moyenne arithmétique est P_l donné par (10-a).

Par exemple :

$$\begin{array}{llll} m = 9 & n = 5 & \bar{R} = 25 & P_{\min} = 825 \quad P_l = 1\,125 \quad P_{\max} = 1\,425 \\ m = 12 & n = 5 & \bar{R} = 32,5 & P_{\min} = 1\,820 \quad \left\{ \begin{array}{l} P_{l_1} = 2\,517 \\ P_{l_2} = 2\,553 \end{array} \right. \quad P_{\max} = 3\,250 \end{array}$$

Comme dans le test de Friedman, l'hypothèse nulle est que, dans chaque bloc, les rangs de 1 à m sont aléatoires. Pour un risque α donné, cette hypothèse est rejetée au profit de l'hypothèse alternative « l'ordre fixé a été globalement confirmé », si la valeur obtenue pour P est au moins égale au fractile $P_{1-\alpha}$ de la distribution de P pour le couple (m, n) : le test est évidemment *unilatéral*.

Plus précisément, l'hypothèse alternative acceptée est que, dans la suite d'inégalités $T_1 \leq T_2 \dots \leq T_i \dots \leq T_m$, l'une au moins des inégalités est stricte.

Pour les risques $\alpha = 0,05$ et $\alpha = 0,01$ et les valeurs les plus courantes de (m, n) les valeurs critiques de P sont données dans la table annexe B, empruntée à [1]. La distribution de P étant discontinue, le risque réel α' est au plus égal à α .

Sous l'hypothèse nulle, lorsque n est suffisamment grand, et quelque soit n si $m \geq 8$, la distribution de P devient pratiquement continue, assimilable à une distribution normale de moyenne $nm \left(\frac{m+1}{2} \right)^2$ et écart-type

$$\frac{m(m+1)}{12} \sqrt{n(m-1)},$$

ce qui permet d'effectuer le test pour toute valeur donnée α du risque.

Dans la Table annexe B ne figurent pas les valeurs critiques pour $m = 2$. Ce cas correspond au test unilatéral d'appariement : entre la valeur critique n^+ de ce test et la valeur critique de P on a la relation évidente $P = n + n^+$.

4.2. Exemple

Il est emprunté à [5] et reproduit dans [1]. On se propose de vérifier que la résistance à la rupture de fils de coton varie en sens inverse de la teneur en potasse du milieu de culture des cotonniers. Pour $m = 5$ taux de potasse (en unités anglaises lb/acre) on dispose de $n = 5$ résultats (répétitions); ceux-ci ne sont pas des « mesures » mais des « repères », l'appareil servant à éprouver la résistance étant « calibrated in arbitrary units ».

Le tableau ci-après donne les résistances ainsi repérées et dans chaque ligne, entre parenthèses, les rangs de 1 à 5 selon l'ordre croissant des résistances.

$$P = 5 + [5 \times 2] + [9 \times 3] + [14 \times 4] + [12 \times 5] = 158$$

La Table annexe B montre que cette valeur est significative au risque $\alpha = 0,05$ (la valeur critique est 155). On peut donc conclure avec une bonne sécurité, que la résistance des fils est fonction décroissante de la teneur en potasse.

TABLEAU IX

Potasse → Répétition ↓	144	108	72	54	36
1	7,46 (2)	7,17 (1)	7,76 (4)	8,14 (5)	7,63 (3)
2	7,68 (2)	7,57 (1)	7,73 (3)	8,15 (5)	8,00 (4)
3	7,21 (1)	7,80 (3)	7,74 (2)	7,87 (4)	7,93 (5)
Totaux R_i	5	5	9	14	12

5. Pour conclure

Le titre de cette note posait une interrogation : l'analyse de la variance est-elle toujours la meilleure méthode ? La réponse doit être nuancée.

Tout d'abord, on ne doit pas appliquer cette méthode lorsqu'elle est inapplicable (exemple de 4.2.). Cela paraît une évidence; mais n'arrive-t-il pas qu'on demande à un statisticien d'interpréter des résultats sans lui indiquer où, quand et comment ils ont été obtenus ? Il est vrai qu'un statisticien digne de ce nom doit exiger ces informations; disons plus nettement qu'il doit collaborer à l'organisation de l'expérience.

En second lieu, il est inutile d'entreprendre des calculs ennuyeux lorsque l'évidence crève les yeux. E. PEARSON aimait à dire — peut-être l'a-t-il écrit — qu'avant tout calcul il convenait, chaque fois que cela était possible, de faire un ou plusieurs graphiques afin de situer visiblement les résultats. Alternativement, leur classement préalable d'après des statistiques d'ordre peut, si l'on n'a pas besoin d'une analyse très fine, suffire à faire éclater une évidence. Dans l'exemple donné en 3.2.7, la dernière colonne des Tableaux III et IV montre clairement que les variétés A et B sont, sinon les meilleures, tout au moins parmi les meilleures et que c'est l'inverse pour Ω — et somme toute, le classement par les totaux de rang ne diffère guère du classement par les valeurs moyennes.

L'analyse de la variance est certes à conseiller lorsque toutes les précautions expérimentales ont été prises pour que les hypothèses sur lesquelles elle repose ont toute chance d'être suffisamment satisfaites. Si l'on a des doutes sérieux, les méthodes non paramétriques constituent un recours apparemment inattaquable.

Bibliographie

- [1] M. HOLLANDER et D.A. WOLFE. — *Non parametric Statistical Methods*. John Wiley & Sons. 1972.
- [2] R.A. FISHER. — *Statistical Methods for research Workers* Oliver and Boyd. London 1954 (2^e édition).
- [3] D.B. OWEN. — *Handbook of Statistical Tables*. Pergamon Press. London 1962.
- [4] A. VESSERAU. — *Méthodes statistiques en biologie et en agronomie*. J.B. Baillière. Ceresta — Tec et Doc-Lavoisier (réédition 1988).
- [5] W.G. COCHRAN et G. COX. — *Experimental Designs*. John Wiley & Sons. 1957 (2^e édition).

ANNEXE

TABLE A
Valeurs critiques
de la Statistique Φ de Friedman

m→ n ↓	3		4		5	
	α = 0,05	α = 0,01	α = 0,05	α = 0,01	α = 0,05	α = 0,001
2	—	—	6,00	—	7,60	8,00
3	6,00	—	7,40	9,00	8,53	10,13
4	6,50	8,00	7,80	9,60	8,80	11,00
5	6,40	8,40	7,80	9,96	8,96	11,52
6	7,00	9,00	7,60	10,20	9,49*	13,28*
7	7,14	8,86	7,84	10,37		
8	6,25	9,00	7,65	10,50		
9	6,22	9,67	7,81*	11,34*		
10	6,20	9,60				
11	6,55	9,45				
12	6,17	9,50				
13	6,00	9,39				
14	6,14	9,00				
15	6,40	8,93				

* Les valeurs « étoilées » résultent de l'approximation de loi de Φ par la loi de χ^2 à $(m - 1)$ degrés de liberté; cette approximation s'applique quelque soit n lorsque $m > 5$.

TABLE B
Valeurs critiques de la Statistique P de Page

Nb de rép. n	Nombre de termes de la comparaison (m)											
	3		4		5		6		7		8	
	$\alpha = 0,05$	$\alpha = 0,01$	$\alpha = 0,05$	$\alpha = 0,01$	$\alpha = 0,05$	$\alpha = 0,01$	$\alpha = 0,05$	$\alpha = 0,01$	$\alpha = 0,05$	$\alpha = 0,01$	$\alpha = 0,05$	$\alpha = 0,01$
2	28	—	58	60	103	106	166	173	252	261	362	376
3	41	42	84	87	150	155	244	252	370	382	532	549
4	54	55	111	114	197	204	321	331	487	501	701	722
5	66	68	137	141	244	251	397	409	603	620	869	893
6	79	81	163	167	291	299	474	486	719	737	1 037	1 063
7	91	93	189	193	338	346	550	563	835	855	1 204	1 232
8	104	106	214	220	384	393	625	640	950	972	1 371	1 401
9	116	119	240	246	431	441	701	717	1 065	1 088	1 537	1 569
10	128	131	266	272	477	487	777	793	1 180	1 205	1 703	1 736
11	141	144	292	298	523	534	852	869	1 295	1 321	1 868	1 905
12	153	156	317	324	570	581	928	946	1 410	1 437	2 035	2 072
13	165	169	*342	*350	*615	*627	*1 002	*1 021	*1 530	*1 560	*2 200	*2 240
14	178	181	*368	*376	*661	*674	*1 077	*1 096	*1 644	*1 676	*2 366	*2 407
15	190	194	*394	*402	*707	*721	*1 153	*1 173	*1 759	*1 792	*2 532	*2 574
16	202	206	*419	*427	*753	*767	*1 228	*1 249	*1 874	*1 907	*2 697	*2 740
17	215	218	*445	*453	*799	*813	*1 303	*1 325	*1 988	*2 023	*2 862	*2 906
18	227	231	*471	*479	*845	*860	*1 378	*1 400	*2 102	*2 138	*3 027	*3 073
19	239	243	*496	*505	*891	*906	*1 453	*1 476	*2 217	*2 254	*3 192	*3 239
20	251	256	*522	*531	*937	*953	*1 528	*1 551	*2 331	*2 369	*3 357	*3 405

Note : Pour les valeurs (n,m) supérieures à celles figurant dans la Table, les valeurs critiques s'obtiennent par les formules suivantes :

$$\alpha = 0,05, \quad P = \frac{m(m+1)}{4} [n(m+1) + 0,548 \sqrt{n(m-1)}]$$

$$\alpha = 0,01, \quad P = \frac{m(m+1)}{4} [n(m+1) + 0,776 \sqrt{n(m-1)}]$$

(valeurs arrondies aux entiers immédiatement supérieurs).

Dans la Table les valeurs « étoilées » ont été calculées par ces formules.