Journal de la société statistique de Paris

Robert Féron

Réflexions sur un test psychologique

Journal de la société statistique de Paris, tome 92 (1951), p. 152-155
<http://www.numdam.org/item?id=JSFS_1951__92__152_0>

Réflexions sur un test psychologique

Le but de cette note est de préciser les hypothèses faites dans les tests basés sur une analyse de variance et de montrer qu'on obtient des résultats sensiblement analogues en faisant des hypothèses sensiblement différentes mais à notre avis plus naturelles. Pour illustrer notre pensée prenons une application précise de l'analyse de la variance à l'arrangement en carré latin (1).

Description de l'expérience.

On a partagé une classe d'enfants en 4 groupes rigoureusement égaux que nous noterons 1, 2, 3, 4. On donne à chaque groupe d'enfants 4 listes de mots à écrire, groupes de mots que nous noterons I, II, III, IV. L'expérience d'autre part peut être effectuée de quatre manières différentes en plaçant l'enfant dans les conditions A, B, C, D.

On fait l'expérience de manière :

1° A placer une fois et une seule chaque enfant dans les conditions A, B, C, D.

2° A utiliser une fois et une seule chaque liste de mots dans les conditions A, B, C, D.

On note ensuite dans la case correspondante le nombre de mots écrits incorrectement par les enfants d'un groupe. Par exemple (tabl. 1) à l'intersection de la colonne 3 et de la ligne II on lit B 42, ceci signifie que 42 mots ont été écrits d'une manière défectueuse quand on a donné au 3e groupe d'enfants, la liste de mots II. (Le B signifie qu'on a alors placé les enfants dans les conditions B).

GROUPE d'enfants / LISTE de mots	1	2	3	4	TOTAL
I	A 81	B 41	C 44	D 53	219
II	D 38	A 97	B 42	C 49	226
III	C 31	D 43	A 67	D 36	177
IV	B 57	C 33	D 43	A 81	214
TOTAL. . .	207	214	196	219	836

Tableau 1.

Solution classique.

Dans la théorie classique on suppose que le nombre x_{ij} de mots écrits d'une manière incorrecte par le je groupe d'enfants dans la ie liste est :

1° La *somme* de 5 termes

$$x_{ij} = \mathcal{A} + \mathcal{L}_i + \mathcal{G}_j + \mathcal{C}_k + \xi$$

où $\mathcal{A}$, $\mathcal{L}_i$, $\mathcal{G}_j$, $\mathcal{C}_k$ ξ sont une constante, une fonction de i, une fonction de j, une fonction de k et où ξ est une variable aléatoire dont la distribution ne dépend pas de i, j ou k.

(1) Nous empruntons l'exemple ci-dessous et les calculs classiques à Thomson *Brit. J. Educ. Psy.* 1941, II, 135.

2º On suppose que ξ obéit à une loi de Laplace.

On se demande si on peut affirmer que les conditions dans lesquelles les enfants sont placés influent sur le résultat. Patnaik a prouvé en 1948 que le test le plus puissant pour répondre à cette question était l'analyse classique de variance tabl. 2.

ORIGINE DE L'ERREUR		SOMME des carrés	DEGRÉS de liberté	QUOTIENT
listes.	$4 \sum_i (x_i - \bar{x})^2$	359,5	3	119,83
groupes.	$4 \sum_j (x_j - \bar{x})^2$	74,5	3	24,38
conditions	$4 \sum_k (x_k - \bar{x})^2$	4626,4	3	1542,17
résidu	$\sum_i \sum_j (x_{ij} - x_i - x_j - x_k + 2\bar{x})^2$	606,5	6	101,08
TOTAUX. . .	$\sum_i \sum_j (x_{ij} - \bar{x})^2$	5667,0	15	

$x_i = \frac{1}{4} \sum_j x_{ij}$

$x_j = \frac{1}{4} \sum_i x_{ij}$

x_k = moyenne pour le traitement k

$\bar{x} = \frac{1}{16} \sum_i \sum_j x_{ij}$

Tableau 2.

Le quotient de l'erreur 101,08 fournit une estimation de la variance de ξ. Il en est de même du quotient des listes 119,83, si tous les k sont nuls. Ici il est impossible d'infirmer cette hypothèse. De même rien ne prouve que certaines listes de mots sont plus difficiles à écrire que d'autres. Par contre le quotient $F = \dfrac{1542}{101} = 15{,}3$ est anormalement grand (on trouve $F = 13{,}745$ pour le niveau de signification 0,01).

Modification des hypothèses précédentes.

Il nous a paru intéressant de regarder ce qui se passait si au lieu des hypothèses classiques nous faisons les hypothèses suivantes :

1º $$x_{ij} = \mathcal{A} \times \mathcal{G}_j \times \mathcal{L}_i \times \mathcal{C}_k \times \xi$$

où il est une constante Gi une fonction de i seul, etc...

2º Log ξ obéit à une loi de Laplace.

Ces deux hypothèses nous paraissent se rapprocher d'avantage de la réalité. En effet, dans toute expérience de biométrie, il est normal de considérer tout effet sur un être vivant comme une modification de son état cellulaire. On a donc affaire non à une somme mais bien à un produit. D'autre part il est bien connu que la loi log-normale fournit généralement des ajustements bien meilleurs que la loi de Laplace. (La loi de Laplace n'est acceptable en général que dans les cas où le coefficient de variation $\dfrac{\sigma}{\bar{x}}$ est petit. Dans ce cas on vérifie que la loi log-normale est très voisine d'une loi de Laplace.)

Sous ces hypothèses le test le plus puissant pour décider si l'hypothèse d'égalité de tous les C_k est admissible est le suivant :

poser

$$y_{ij} = \log \mathcal{A} + \mathrm{Log}\, \mathcal{G}_j + \mathrm{Log}\, \mathcal{L}_i + \mathrm{Log}\, \mathcal{C}_k + \mathrm{Log}\, \xi$$

et faire une analyse de variance sur le y_{ij}.

Le tableau 3 donne les diverses valeurs de y_{ij} et le tableau 4 l'analyse de variance correspondante.

GROUPES / LISTES	1	2	3	4	TOTAL
I	A 1,91	B 1,61	C 1,74	D 1,72	6,88
II	D 1,58	A 1,99	B 1,62	C 1,69	6,88
III	C 1,49	D 1,63	A 1,83	B 1,56	6,51
IV	B 1,76	C 1,51	D 1,63	A 1,91	6,81
TOTAL. . .	6,74	6,74	6,72	6,88	27,08

$y_A = 7,64$
$y_B = 6,55$
$y_C = 6,33$
$y_D = 6,56$

Tableau 3.

SOMME DES CARRÉS		DEGRÉS de liberté	QUOTIENT
listes.	234,5	3	76,16
groupes.	41	3	13,66
conditions	2607,5	3	869,16
résidu	465	6	77,5
TOTAUX. . .	3348	15	

Tableau 4.

On voit donc que nous n'avons sous ces hypothèses aucune raison de croire que les $\mathcal{L}_i$ ou les $\mathcal{G}_j$ sont différents. Par contre pour $\mathcal{C}_k$ nous avons $F = \dfrac{869,16}{77,5} = 11,2$ $\nu_1 = 3$ $\nu_2 = 6$. Ce résultat n'est plus significatif au niveau de signification 1 % mais il le reste toujours au niveau de signification 5 % (F = 4,7571).

Robert Féron.