

ANNALES DE L'I. H. P.

F. YATES

Bases logiques de la planification des expériences

Annales de l'I. H. P., tome 12, n° 2 (1951), p. 97-112

http://www.numdam.org/item?id=AIHP_1951__12_2_97_0

© Gauthier-Villars, 1951, tous droits réservés.

L'accès aux archives de la revue « Annales de l'I. H. P. » 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/>

Bases logiques de la planification des expériences

par

F. YATES.

La planification et l'analyse des expériences est la branche de la technique statistique moderne dans laquelle la logique, sur laquelle sont basées les inductions statistiques, est la plus simple et la plus nettement tranchée. C'est aussi, dans un certain sens, le progrès le plus révolutionnaire des récentes années, car c'est dans la théorie de la planification des expériences que la logique de l'induction statistique a été clarifiée pour la première fois. C'est aussi dans les expériences convenablement planifiées que l'édifice logique correspond vraiment à la situation pratique existante, de telle sorte que les inductions sont conformes à la réalité. Pour ces raisons j'ai toujours conseillé à ceux qui s'intéressent à la statistique d'étudier les principes de la planification des expériences, même s'ils ne doivent jamais s'intéresser directement à l'organisation de celles-ci.

Je n'ai, par conséquent, aucune hésitation à consacrer la première de ces deux conférences à l'étude théorique de cette question, d'autant plus que les explications données dans divers ouvrages classiques de statistique mathématique sont défectueuses à divers égards. Dans ma seconde conférence, je me propose de décrire quelques-uns des développements récents les plus importants, particulièrement du point de vue théorique.

Estimation et tests de signification. — La théorie fondamentale sera plus facile à décrire à partir d'un exemple concret :

I	F	B	E	C	F	E	II
	A	D	C	D	A	B	
III	C	E	F	D	C	A	IV
	A	D	B	F	B	E	

Fig. 1. — Blocs avec répartition au hasard pour quatre répétitions de six traitements

La figure 1 montre un arrangement pour une expérience en plein champ, de 6 traitements, ou variétés, avec 4 répétitions de chaque traitement. Le plan expérimental ainsi conçu est un exemple de ce que l'on désigne sous le nom de blocs avec répartition au hasard : l'ensemble des 24 parcelles est divisé en 4 blocs de 6 parcelles chacun, les 6 traitements considérés étant *répartis au hasard* dans les 6 parcelles de chaque bloc. L'importance de cette condition de répartition au hasard apparaîtra ultérieurement. L'analyse numérique des résultats d'une telle expérience est très simple. Les récoltes des 24 parcelles sont notées dans une table à double entrée (tableau I) :

TABLEAU I. — Récoltes dans une expérience avec répartition au hasard dans les blocs.

Bloc	A	B	C	D	E	F	Total
I	y_{11}	y_{12}	.	.	.	.	B_1
II	.	.	.	.	.	.	B_2
III	.	.	.	.	.	.	.
IV	.	.	.	.	.	.	.
Total	T_a	T_b	.	.	.	.	G

Les effets des traitements sont estimés à partir des différences entre les traitements totaux $T_a, T_b, \dots$, ou entre les moyennes correspondantes, exprimés en unités convenables. Les erreurs de ces estimations sont évaluées au moyen de la méthode connue sous le nom d'*analyse de la variance*. Dans notre exemple, celle-ci prend la forme indiquée dans le tableau II, dans lequel $\text{dev}^2 y$ représente la somme des carrés des déviations des y à partir de leur moyenne, etc.

TABLEAU II. — Analyse de la variance.

Degrés de liberté.		Somme des carrés.	Carré moyen.
Blocs.....	3	$(\text{dev}^2 B)/6$	P
Traitements.....	5	$(\text{dev}^2 T)/4$	Q
Erreur.....	15	(par différence)	R
Total.....	23	$\text{dev}^2 y$	

} par division

L'erreur quadratique moyenne R donne une estimation s^2 de σ^2 , variance de l'erreur par parcelle, de sorte que la variance de l'erreur

d'une moyenne de traitements (moyenne de 4 parcelles) est $\frac{1}{4}\sigma^2$, et celle de la différence de 2 moyennes de traitements est $\frac{1}{2}\sigma^2$.

Cette forme d'analyse peut facilement être justifiée par la méthode des moindres carrés, à partir de l'hypothèse que les récoltes des parcelles sont des fonctions de la forme

$$y_{rs} = \mu + \beta_r + \tau_s + \varepsilon_{rs},$$

dans laquelle μ est une constante pour l'ensemble, β_r est une constante relative à un bloc avec $S(\beta_r) = 0$, τ_s est une constante relative à un traitement, avec $S(\tau_s) = 0$, et ε_{rs} est une erreur aléatoire, tous les ε_{rs} étant normalement et indépendamment distribués, avec la même variance σ^2 . L'erreur quadratique moyenne R a pour espérance mathématique σ^2 , et, par conséquent, fournit une estimation correcte en moyenne de σ^2 , c'est-à-dire une estimation sans erreur systématique, ou sans biais.

La méthode des moindres carrés a été développée par Gauss. Elle peut aussi être considérée comme un cas particulier de la méthode proposée par R. A. Fisher, sous le nom de méthode de *maximum likelihood* (maximum de vraisemblance).

Fisher a étendu la théorie en observant que si les y sont normalement et indépendamment distribués avec la même variance σ^2 (tous les β et tous les τ étant nuls), les quantités $\frac{n_0 P}{\sigma^2}$, $\frac{n_1 Q}{\sigma^2}$ et $\frac{n_2 R}{\sigma^2}$ sont distribuées indépendamment comme χ^2 avec n_0 , n_1 et n_2 degrés de liberté respectivement [dans le cas actuel 3, 5 et 15 degrés de liberté ⁽¹⁾]. De plus le rapport des carrés moyens $\frac{Q}{R}$, qu'on appelle aussi le rapport des variances, est distribué comme le rapport $\frac{n_2 \chi_1^2}{n_1 \chi_2^2}$, qui est indépendant de σ^2 . Fisher a déterminé et mis en tables la distribution équivalente de la variable z , dans laquelle e^{2z} est le rapport des variances, fournissant ainsi un test exact de signification pour les traitements *considérés dans leur ensemble*. Si, par exemple, la valeur de z , calculée à partir de l'expérience, dépasse la valeur qui correspond au seuil 5 % de la distribution de z , on peut conclure que, s'il n'y a pas d'effet des traitements, les différences entre

⁽¹⁾ La distribution de χ^2 , pour n degrés de liberté, peut être définie comme la distribution de la somme des carrés de n quantités, chacune étant normalement et indépendamment distribuée avec un écart-type égal à l'unité.

les moyennes des traitements (qui doivent, dans ce cas, être dues entièrement aux fluctuations aléatoires), aussi grandes ou plus grandes que celles observées, ne se présenteraient, par hasard, que dans moins de 5 % de toutes les expériences analogues.

Le test z est une extension du test t qui fut proposé précédemment par Gosset (Student). Le test t fournit un test de signification pour la différence entre les résultats d'une paire quelconque de traitements et est équivalent au test z pour $n_1 = 1$. Son caractère révolutionnaire, en fournissant un test exact de signification pour les *petits échantillons* d'observations quantitatives, a été mis en évidence pour la première fois par Fisher.

Orthogonalité. — La structure de l'analyse de la variance, dans le cas d'une expérience avec répartition au hasard dans les blocs, est particulièrement simple, pour la raison que les blocs et les traitements sont *orthogonaux*. Dans la terminologie des moindres carrés, la réduction dans la somme des carrés des déviations due à l'ajustement des constantes relatives aux traitements est la même, que les constantes relatives aux blocs aient été aussi ajustées ou non. Cette propriété résulte du fait que chaque traitement se présente une fois, et une fois seulement, dans chaque bloc, de telle sorte que les différences entre blocs n'affectent pas les comparaisons entre traitements : des augmentations égales dans les récoltes de toutes les parcelles d'un même bloc, par exemple, augmentent également toutes les moyennes de traitements. C'est seulement lorsque deux groupés de degrés de liberté sont orthogonaux que les sommes correspondantes de carrés sont additives. En ce cas la somme des carrés relative à l'erreur peut être obtenue simplement par différence.

Si les blocs et les traitements n'étaient pas orthogonaux, ce qui serait le cas si les traitements étaient inégalement représentés dans les différents blocs, une solution par la méthode des moindres carrés serait encore possible, mais les constantes relatives aux blocs et aux traitements devraient être déduites d'équations linéaires simultanées au lieu d'être déduites directement des moyennes de blocs et de traitements. La somme des carrés, relative aux traitements, qui serait appropriée à l'emploi du test z , s'obtiendrait en déduisant la part de la somme des carrés provenant de l'ajustement des constantes relatives aux blocs seuls,

de celle provenant de l'ajustement simultané des constantes relatives aux blocs et aux traitements.

La symétrie, qui donne lieu à l'orthogonalité, a comme autre conséquence importante que les comparaisons entre traitements sont toutes de la même précision.

Les types les plus simples de planification des expériences sont complètement orthogonaux, mais un certain degré de non-orthogonalité peut être admis dans des types plus élaborés de plans d'expérience à condition qu'ils soient d'une nature telle que l'estimation des effets dus aux traitements et l'analyse de la variance ne soient pas compliquée à l'excès. Je donnerai ultérieurement quelques exemples de tels plans d'expérience. Pour l'instant, j'attire simplement l'attention sur le plan d'expérience connu sous le nom de *carré latin* qui est tel que chaque traitement se présente une fois, et une fois seulement, dans chaque rangée et dans chaque colonne d'un ensemble carré de parcelles. On voit aisément que ce plan est complètement orthogonal à la fois pour les rangs, les colonnes et les traitements. Par conséquent, l'analyse de la variance a la même simplicité que dans les blocs répartis au hasard.

Le rôle de la répartition au hasard. — Il y a lieu maintenant de considérer le rôle de la répartition au hasard. On voit immédiatement que l'hypothèse que les résidus ε_{rs} sont normalement et indépendamment distribués avec la même variance quel que soit l'arrangement des traitements, n'est pas justifiée. Dans l'expérimentation agricole, par exemple, il est bien connu que les récoltes des parcelles voisines ont tendance à être plus ou moins liées et que cet effet persiste même lorsque l'on calcule les résidus à partir des moyennes de blocs et de traitements.

Nous pouvons d'abord considérer le résultat d'une étude de l'analyse de la variance dans le cas où les traitements sont fictifs, c'est-à-dire dans le cas où toutes les parcelles ont subi le *même traitement*.

Dans notre exemple, il y a $(6!)^4$ arrangements possibles des symboles caractéristiques des traitements fictifs. Ceci donne lieu à $(6!)^3$ valeurs des carrés moyens relatifs aux traitements et à l'erreur, et aussi $(6!)^3$ valeurs de z . Il est facile de montrer que, si l'on considère tous les arrangements possibles, la valeur moyenne du carré moyen relatif à l'erreur est égale à la valeur moyenne du carré moyen relatif aux trai-

tements. Si le plan d'expérience est tel que cette condition soit réalisée, le carré moyen relatif à l'erreur peut être considéré comme donnant une estimation non systématiquement faussée de la variance relative à l'erreur. L'absence de déviation systématique est clairement une condition nécessaire pour que la distribution de z puisse être valablement employée.

Ces $(6!)^3$ valeurs de z , en elles-mêmes, fournissent un test de la valeur significative des effets des traitements, test qui est indépendant de toute hypothèse relative à la normalité.

Si la valeur de z , correspondant à l'arrangement effectif des traitements dans une expérience, se trouve être dépassée par moins de 5 % de toutes les valeurs de z , alors les effets des traitements peuvent être jugés significatifs au seuil de signification 5 %.

Ceci constitue ce que l'on appelle le test de répartition au hasard. On notera qu'avant de le réaliser il est nécessaire d'assigner préalablement un ordre aux différents arrangements. Il n'y a rien d'absolu dans l'ordre fourni par les valeurs de z (ou par le carré moyen relatif aux traitements); nous pouvons, par exemple, avoir considéré comme critère la différence entre la plus grande et la plus petite moyenne de traitements. En conséquence, le test de répartition au hasard est indépendant de toute hypothèse concernant la distribution des résidus, mais il n'est pas *unique*, ainsi qu'on le prétend quelquefois.

Des considérations générales nous conduisent à supposer que les probabilités données par le test de répartition au hasard se conformeront très exactement à celles données par le test z , pourvu que le nombre des parcelles ne soit pas trop petit et que la distribution des résidus à partir des moyennes de blocs ne soit pas trop éloignée de la distribution normale. Une des principales causes de désaccord ne réside pas dans le manque de normalité; mais provient du fait que les valeurs de z données par la distribution au hasard ont tendance à être groupées. Les $(6!)^3$ valeurs ci-dessus, par exemple, ne peuvent pas être regardées comme un échantillon au hasard de $(6!)^3$ valeurs prélevées dans une distribution continue.

Quelques comparaisons basées sur des observations numériques effectives ont montré un bon accord et elles ont été prises comme justification de l'emploi du test z sous condition d'une non-normalité modérée. Cependant, d'autres auteurs ont établi que dans des circonstances se

présentant assez communément, par exemple, dans une petite expérience, avec des récoltes parcellaires dont les résidus peuvent être considérés comme n'étant pas un échantillon trop improbable d'une population normale, le test de répartition au hasard ne s'accorde pas étroitement avec le test z et ces auteurs ont même suggéré qu'une correction devrait être faite à ce dernier test.

A mon avis, ceci est une erreur due au manque de connaissance du fait qu'il peut fréquemment y avoir plus d'un test raisonnable de signification pour un ensemble donné d'observations. Divers tests peuvent être en désaccord notoire, bien qu'ils soient d'accord en moyenne pour un grand nombre de groupes d'observations. Dans de telles circonstances, il n'est pas admissible de choisir le test qui donne, par exemple, la plus grande signification. Le test à employer doit être choisi d'avance, et l'on doit se soumettre à son verdict, qu'il soit favorable ou non.

En plus de sa commodité, on peut montrer que le test z basé sur la distribution normale est le plus efficace, si les résidus sont réellement distribués normalement. Il peut, par conséquent, être justifié sur ces bases.

D'ailleurs, le test z n'est pas utilisé en général, pour des observations trop éloignées de la distribution normale, puisque, dans la pratique, il est d'usage de transformer les observations qui sont vraiment éloignées de la distribution normale en données approximativement normales, par diverses transformations (racine carrée, logarithme ou transformation angulaire).

Estimation de l'erreur quand il y a des effets dus aux traitements. — L'emploi de tests de signification n'est pas le seul, ni même généralement le principal objet d'une expérience, puisque l'on sait fréquemment avant de faire l'expérience, que les traitements produiront un certain effet. En général, on a besoin d'estimations de la grandeur des différences entre les traitements, ainsi que de l'estimation des erreurs auxquelles ces différences sont sujettes.

Une analyse ordinaire de la variance est basée sur l'hypothèse que les résidus associés aux différents traitements peuvent être regardés comme ayant la même variance. L'analyse a aussi l'effet de grouper ensemble les estimations d'erreurs déduites des répétitions des divers traitements, mais ce genre de groupement n'est pas nécessairement approprié si

quelques-uns des traitements ont des effets beaucoup plus importants que certains autres.

Par exemple, si les traitements A et B ont pour effet une suppression complète de la récolte, et si nous nous proposons d'estimer les erreurs des comparaisons de C, D, E, F, les traitements A et B doivent être entièrement retirés de l'analyse de la variance. Ceci peut être fait simplement dans une expérience relative à des blocs répartis au hasard, mais dans des plans d'expérience plus compliqués, tels qu'un carré latin, l'orthogonalité aura disparu et l'analyse demande, par conséquent, à être modifiée. Cependant, dans des cas moins extrêmes, la difficulté peut souvent être surmontée par l'emploi d'une transformation appropriée.

En dehors de cet inconvénient qui se présentera seulement si les effets des traitements sont importants, la répartition au hasard, si elle est d'un type approprié, garantit que l'estimation de l'erreur, à partir de toutes les observations, est appropriée pour toutes les comparaisons. En particulier, dans une expérience de blocs avec répartition au hasard, n'importe quelle paire de traitements a les mêmes chances d'occuper n'importe quelle paire de parcelles dans l'un quelconque des blocs. Si les traitements ne produisent aucun effet, la moyenne des carrés des différences de n'importe quelle paire de traitements pour tous les arrangements possibles sera, après avoir été multipliée par un multiplicateur approprié, égale à la moyenne des carrés moyens relatifs à l'erreur. En d'autres termes, l'estimation de l'erreur peut être valablement utilisée pour la comparaison de n'importe quelle paire de traitements.

Refus d'arrangements non satisfaisants. — La répartition au hasard nous assure qu'aucun traitement n'est spécialement favorisé puisque chaque traitement a une égale probabilité de tomber sur chaque parcelle. En conséquence, les estimations des effets des traitements ne sont pas viciées, en ce sens que lorsqu'il n'y a pas d'effets de traitements, la valeur moyenne des estimations de ces effets, à partir de tous les schémas possibles de répartition au hasard, sera nulle. Cependant, dans un arrangement particulier quelconque, il peut arriver que certains traitements se trouvent mieux placés que d'autres. Ainsi, dans l'arrangement

de la figure 1, des traitements tels que B et E peuvent être désavantagés par rapport aux traitements C et D, si les bords de l'expérience ont tendance à avoir des récoltes plus faibles.

Dans une longue suite d'expériences semblables, les effets de cette espèce disparaîtront en moyenne, mais ceci ne règle pas la difficulté. A la vérité, des arrangements plus défavorables que ceux obtenus dans la figure 1 peuvent, occasionnellement, se produire par hasard. Ainsi, il y a une chance sur 216 qu'un traitement particulier se place aux quatre coins, et la même chance pour qu'un des traitements se place aux quatre parcelles du centre. Je pense que le praticien se rend compte que ni l'un ni l'autre de ces arrangements n'est un bon arrangement. Cependant si de tels arrangements sont rejetés lorsque la répartition au hasard y conduit, la propriété d'avoir une estimation correcte de l'erreur cesse d'exister. Aussi longtemps que l'expérimentateur réussit à rejeter les arrangements les moins précis, il obtiendra des résultats qui, en moyenne, seront plus précis que la moyenne générale relative à tous les arrangements au hasard, mais il les jugera moins précis, non seulement qu'ils le sont en réalité, mais encore moins précis que la moyenne de tous les arrangements au hasard.

Dans certains types de plans, les types les plus défavorables d'arrangements peuvent être exclus, cependant qu'il est possible de conserver une estimation correcte de l'erreur. Ceci se trouve réalisé au moyen de ce que j'ai appelé le principe de la *répartition au hasard restreint*. Cette méthode consiste à déterminer un sous-groupe d'arrangements pour lequel la propriété d'une estimation correcte de l'erreur est réalisée et à choisir au hasard un arrangement de ce sous-groupe. Je donnerai un exemple de l'application de ce principe dans ma seconde conférence.

Arrangements systématiques. — La position extrême à laquelle conduit le rejet d'arrangements considérés comme imprécis est l'abandon du principe de la répartition au hasard, remplacé par le choix de quelque arrangement particulier, lequel, en raison de ses propriétés d'équilibre ou de symétrie, apparaît à l'expérimentateur comme étant capable de donner des comparaisons particulièrement précises. Un tel arrangement s'appelle un arrangement systématique.

Un exemple classique d'arrangement systématique est constitué par

les carrés Knut Vik ou carrés de la marche du cavalier (*fig. 2*) qui sont des carrés latins particuliers à cinq lignes et cinq colonnes :

A	B	C	D	E	A	D	B	E	C
D	E	A	B	C	B	E	C	A	D
B	C	D	E	A	C	A	D	B	E
E	A	B	C	D	D	B	E	C	A
C	D	E	A	B	E	C	A	D	B

Fig. 2. — Carrés Knut Vik.

Des carrés latins particulièrement défavorables sont les carrés diagonaux (*fig. 3*) :

A	B	C	D	E	A	B	C	D	E
B	C	D	E	A	E	A	B	C	D
C	D	E	A	B	D	E	A	B	C
D	E	A	B	C	C	D	E	A	B
E	A	B	C	D	B	C	D	E	A

Fig. 3. — Les carrés diagonaux.

Tedin a testé ces quatre carrés sur 91 groupes de récoltes provenant d'expériences d'uniformité. Il a trouvé qu'en moyenne les carrés Knut Vik montraient un gain apparent de précision d'environ 9 %, par rapport à la moyenne de tous les arrangements au hasard et que les carrés diagonaux montraient une perte à peu près du même ordre, bien que, comme on pouvait le supposer, les précisions relatives montraient une variation considérable d'une expérience à l'autre.

En plus de la perte de tests de signification et d'estimations d'erreurs pleinement valables, les arrangements systématiques ont d'autres défauts qui, souvent, ne sont pas soupçonnés. Les variations périodiques dans la fertilité et la concurrence entre les traitements ou variétés peuvent toutes les deux être cause de complication. Ainsi, par exemple, dans la figure 2, si le haut du carré est orienté vers le Nord, la variété A portera ombre sur la variété C dans quatre parcelles sur cinq. Si A est une variété de grande taille, il pourra en résulter, pour la variété C, un désavantage appréciable.

Il est évident que des carrés latins obtenus occasionnellement, par répartition au hasard, peuvent être tels que la variété A porte ombre sur une même variété dans toutes les parcelles, mais cette situation est quelque peu exceptionnelle.

Des effets de cette espèce sont de nature à être particulièrement ennuyeux puisqu'ils peuvent donner un résultat paraissant hautement significatif, mais qui est entièrement artificiel. Un exemple a récemment attiré mon attention dans une expérimentation portant sur des plantes de serre à Rothamsted, expérience dans laquelle les traitements A, B, C, D, E, F représentaient des quantités croissantes d'un engrais chimique qui, en petites quantités, stimulait la croissance, mais était toxique en grandes quantités. Chaque répétition était arrangée dans l'ordre de A à F, de telle sorte que A était toujours placé à la suite de F (ou à la fin d'un rang). En conséquence B était plus ombragé que A dans toutes les répétitions. L'analyse statistique indiquait que l'effet de B était significativement moindre que celui de A ou de C, et l'expérimentateur qui étudiait la question pensait qu'une intéressante découverte avait été faite. Si un arrangement au hasard avait été utilisé (comme il le fut dans les expériences suivantes), la rivalité entre plantes n'aurait pas été entièrement éliminée, mais ses effets auraient été distribués sur l'ensemble de tous les traitements. L'effet général aurait été d'augmenter quelque peu les différences entre traitements, de manière approximativement proportionnelle à leur valeur absolue. En dehors de cette tendance à surestimer les véritables différences entre traitements, l'unique effet de la rivalité, dans un arrangement au hasard, est d'augmenter l'erreur, mais les tests de signification des différences entre traitements ne sont pas faussés.

Dans de nombreux arrangements systématiques, on peut envisager diverses méthodes d'estimation de l'erreur, aucune d'elles n'étant pleinement valable, mais toutes étant plausibles. Ceci conduit à des discussions fatigantes et sans profit sur les méthodes d'analyse. Un cas classique est fourni par la méthode des bandes d'un semoir double en agriculture, méthode qui fut, à une certaine époque, très populaire pour comparer deux variétés. Elle consiste dans l'arrangement de parcelles (ou bandes)

A B B A A B B A A B B A....

Chaque bande a une largeur égale à la moitié de celle du semoir et chaque paire AB ou BA est semée à un parcours aller ou retour du semoir, qui est partagé en deux moitiés, une pour chaque variété. Deux méthodes ont été recommandées pour l'estimation de l'erreur, la première étant basée sur la variation des différences $A - B$, calculées dans

chaque paire de bandes A B ou B A, d'un semoir double, la seconde étant basée sur la variation des différences A — B, calculées à partir des *sandwiches* A B B A. A première vue, la dernière estimation apparaît comme devant être la meilleure puisqu'elle élimine les effets d'une variation continue de fertilité dans l'estimation de l'erreur, ainsi que dans la comparaison entre variétés, donnée par l'arrangement. Cependant, il n'y a aucune justification réelle pour traiter les bandes contiguës d'un semoir double de la même variété comme indépendantes. Sauf le cas de défauts possibles dans une moitié du semoir, défauts qui affecteront toujours la même variété, l'arrangement est pratiquement équivalent à un arrangement de bandes alternées de semoir

A, B, A, B, A, B....

En conséquence, la deuxième méthode est, en fait, de nature à conduire à sous-estimer l'erreur en raison du fait que l'on néglige la corrélation des bandes contiguës, d'un semoir double, de l'une des variétés. On peut, en vérité, se demander pourquoi l'expérimentateur complique lui-même l'expérience en partageant le semoir en deux moitiés et en moissonnant séparément les bandes d'un semoir double contiguës et ensemencées avec une même variété.

L'autre alternative évidente relative à ce type d'arrangement systématique est l'emploi de paires de bandes A B ou B A au hasard, ou de *sandwichs* A B B A ou B A A B au hasard. Dans le dernier cas, chaque *sandwich* est considéré comme une unité dans l'estimation de l'erreur.

Restrictions donnant lieu à une estimation viciée de l'erreur. — A première vue, on peut penser que n'importe quel type de plan, donnant lieu à une solution basée sur la méthode des moindres carrés et pour lequel une estimation de l'erreur est obtenue simplement par l'analyse de la variance, peut être employé. Il n'en est cependant pas ainsi, car de nombreux plans qui se conforment à ces conditions sont tels qu'aucun procédé de répartition au hasard ne peut donner lieu à une estimation correcte de l'erreur. Un exemple simple est fourni par le plan suivant (*fig. 4*).

Il s'agit d'un carré latin à 4 rangées et 4 colonnes, dans lequel est imposée la restriction supplémentaire que chacun des 4 traitements figure dans chacun des blocs constitués par les 4 quartiers du carré. Des

plans semblables peuvent être aisément construits pour des carrés latins de diverses grandeurs. Une estimation de l'erreur peut être obtenue

I	C	D	A	B	II
	A	B	C	D	
III	B	A	D	C	IV
	D	C	B	A	

Fig. 4. — Carré 4×4 avec une restriction additionnelle relative à la subdivision en blocs.

simplement par la méthode de l'analyse de la variance, puisque le degré de liberté des blocs représenté par

$$I - II - III + IV$$

est orthogonal, relativement aux rangées, colonnes et traitements, et les deux autres degrés de liberté des blocs, c'est-à-dire

$$I + II - III - IV \quad \text{et} \quad I - II + III - IV$$

sont respectivement équivalents aux degrés de liberté des rangées et des colonnes.

Il y a $12 \times 4!$ carrés latins possibles qui satisfont à cette restriction. Parmi eux, $4 \times 4!$ carrés, dont l'un d'eux est montré dans la figure 4, sont d'une espèce pour laquelle la comparaison des traitements peut être partagée en deux types a et b , ainsi que le montre la figure 5 :

+	+	-	-	+	+	-	-	+	-	+	-	+	-	-	+
-	-	+	+	-	-	+	+	+	-	+	+	-	-	+	+
+	+	-	-	-	-	+	+	-	-	+	+	-	+	+	-
-	-	+	+	+	+	-	-	-	-	+	+	-	+	+	-
(Erreur)				(C + D - A - B)				(A + C - B - D)				(Erreur)			

Type a .

+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-
-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+
+	-	+	-	-	+	+	-	+	-	+	-	-	+	+	-
-	+	-	+	+	+	-	-	-	+	+	-	+	-	-	+
(Erreur)				(Erreur)				(B + C - A - D)				(Erreur)			

Type b .

Fig. 5. — Contrastes dans un carré 4×4 avec une restriction additionnelle relative aux blocs.

Dans ce cas, l'analyse de la variance peut être subdivisée ainsi que le montre la table 3 ci-après :

		Degrés de liberté.
Rangées		3
Colonnes.....		3
Blocs		1
Type <i>a.</i> {	Traitements.....	2
	Erreur.....	2
Type <i>b.</i> {	Traitement.....	1
	Erreur.....	3
TOTAL		15

Table 3. — Analyse de la variance dans un carré 4×4 avec une restriction additionnelle relative aux blocs.

Avec des récoltes données, l'espérance mathématique du carré moyen des traitements dans (*a*) est égale à celle du carré moyen de l'erreur dans (*a*) et celle du carré moyen relatif aux traitements dans (*b*) est égale à celle du carré moyen de l'erreur dans (*b*). Dans une analyse ordinaire de la variance, les erreurs relatives à (*a*) et (*b*) sont groupées ensemble et les traitements relatifs à (*a*) et (*b*) sont aussi groupés ensemble. Puisque les deux composantes entrent dans des proportions différentes dans les traitements et dans l'erreur, l'espérance mathématique du carré moyen de l'erreur dans cet ensemble ne sera pas égale à celle du carré moyen relatif aux traitements pris dans leur ensemble. Dans les $8 \times 4!$ modèles possibles restants, la subdivision ne peut pas être réalisée, de la même manière simple, mais les espérances mathématiques des sommes de carrés relatives aux traitements et à l'erreur sont les mêmes. Ainsi : l'estimation de l'erreur se trouve faussée. Le fait que cette déformation de l'erreur n'est pas, en pratique, de nature à être sans importance se voit si nous reconnaissons que le carré moyen de l'erreur dans (*a*) est de nature à être, en moyenne, notoirement plus grand que celui de l'erreur du type (*b*), puisque les erreurs du type (*a*) proviennent des contrastes des parcelles qui sont effectivement deux fois aussi grandes que celles du type (*b*).

De nombreux arrangements semblables sont sujets aux mêmes types de défauts, mais, en certains cas, il est possible d'utiliser des arrangements où les sommes de carrés relatifs à l'erreur et aux traitements peuvent être subdivisées comme dans l'analyse ci-dessus, de manière à fournir une estimation correcte de l'erreur. Évidemment, ceci ne

donnera pas un test α correct pour les traitements considérés dans leur ensemble.

Réduction de l'erreur par l'emploi d'informations supplémentaires. —

Pour ceux qui sont accoutumés aux solutions fournies par la méthode des moindres carrés, il apparaît qu'en plus des constantes ajustées pour représenter les blocs dans une expérience de blocs avec répartition au hasard, ou en plus des constantes ajustées pour représenter les rangées et les colonnes dans une expérience en carré latin, d'autres constantes peuvent être introduites, par exemple pour représenter une variation de fertilité au travers de l'expérience. Si une variation marquée existe, il est clair que l'on peut espérer que l'ajustement d'une telle constante conduira à des estimations plus précises des effets des traitements.

Incidemment, lorsque l'examen des résultats indique qu'il y a une importante variation de fertilité ou une autre irrégularité de fertilité simplement définie, qui n'est pas éliminée par le plan d'expérience, on peut avoir recours à l'ajustement de constantes additionnelles. Cependant, en général, ceci est à déconseiller, à la fois en raison du supplément de travail qui en résulte dans les calculs et aussi parce que les constantes choisies pour l'ajustement seront naturellement celles qui réduisent l'erreur apparente. Il peut en résulter une sous-estimation importante des véritables erreurs. Il y a aussi la possibilité que les résultats soient déformés en choisissant des constantes qui donnent lieu à des estimations paraissant les plus acceptables à d'autres égards.

L'ajustement de résultats expérimentaux par l'usage de régression basée sur une information supplémentaire, telle que le poids initial d'animaux étudiés dans des expériences de comparaisons de régimes alimentaires, est, au point de vue formel, analogue à l'ajustement d'une constante pour une variation de fertilité, l'un ou l'autre de ces problèmes peut être traité par une analyse de covariance.

Cependant, pourvu que la variable supplémentaire ne soit pas affectée par les traitements, cette forme d'ajustement est une procédure plus légitime, parce que le plan d'ajustement précède l'expérience; en fait, dans les expériences de comparaisons de nourritures, c'est la covariance même qui résout le problème de savoir si c'est le poids final (coefficient de régression d'ordre zéro) ou l'augmentation de poids (coefficient de régression d'ordre 1) qui doit être analysé, en

employant les observations elles-mêmes pour déterminer la valeur du coefficient de régression qui donne les résultats les plus précis.

On verra que l'emploi des méthodes ordinaires d'analyse, sans ajuster de constante supplémentaire pour tenir compte de la place sur le terrain, équivaut à ignorer complètement l'information fournie par la situation effective des parcelles. En ignorant cette information, nous obtenons une solution qui a des propriétés d'unicité et de simplicité et qui donne lieu à des tests de signification qui ont une base logique exacte. Mon expérience de la pratique de l'expérimentation me conduit à la conclusion que la petite perte de précision résultant du fait que nous disciplinons à notre ingéniosité d'analyse statistique est d'un prix peu élevé pour ces très réels avantages.

Sur ces bases, je condamne la tendance, encore souvent apparente, qui conduit à essayer d'améliorer la précision des expériences en plein champ par des techniques d'analyse statistique élaborées, telles que l'ajustement de courbes polynômes. Mais, en tout cas, si ceci doit être fait, que ce soit à partir de plans avec répartition au hasard.
