

PIERRE DAGNELIE

La planification des expériences : choix des traitements et dispositif expérimental

Journal de la société française de statistique, tome 141, n° 1-2 (2000), p. 5-29

<http://www.numdam.org/item?id=JSFS_2000__141_1-2_5_0>

© Société française de statistique, 2000, tous droits réservés.

L'accès aux archives de la revue « Journal de la société française de statistique » (<http://publications-sfds.math.cnrs.fr/index.php/J-SFdS>) 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/>

LA PLANIFICATION DES EXPÉRIENCES : CHOIX DES TRAITEMENTS ET DISPOSITIF EXPÉRIMENTAL

Pierre DAGNELIE *

RÉSUMÉ

On observe depuis plusieurs années une tendance à concentrer l'attention en matière d'expérimentation sur le choix des traitements, parfois à l'exclusion de toute autre considération (répétition, randomisation, dispositif expérimental, etc.). Cette tendance peut avoir des conséquences fâcheuses quand la variabilité résiduelle n'est pas négligeable (absence de connaissance de cette variabilité, faible précision des estimations, faible puissance des éventuels tests d'hypothèses, erreurs systématiques, etc.).

Nous identifions cette tendance en la plaçant dans un contexte historique, nous en illustrons les conséquences par l'examen de quelques exemples et nous formulons diverses recommandations à ce sujet.

ABSTRACT

Since several years, one may observe a tendency to stress mainly the choice of treatments (treatment design) in the field of experimentation, sometimes to the exclusion of any other issue (replication, randomization, experimental design, etc.). This tendency may have annoying consequences when the residual variability is not negligible (absence of knowledge of the variability, low precision of estimates, low power of possible tests, systematic errors, etc.).

We identify this tendency within a historical frame, we illustrate the consequences of this tendency by some examples, and we draw up some recommendations.

1. INTRODUCTION

Cette note a pour but de souligner le caractère essentiel de l'équilibre qu'il faut assurer entre les différents éléments constitutifs de tout plan d'expérience. Elle met l'accent en particulier sur la place qu'il faut accorder d'une part au choix des traitements (*treatment design*) et d'autre part au dispositif expérimental (*experimental design*), c'est-à-dire à la répartition des différents traitements et de leurs répétitions éventuelles dans l'espace et/ou dans le temps.

* Faculté universitaire des Sciences agronomiques, B-5030 Gembloux (Belgique), e-mail : dagnelie.p@fsagx.ac.be

Après un bref rappel historique (paragraphe 2), nous précisons ce que nous entendons par plan d'expérience ou protocole expérimental (paragraphe 3), nous observons l'évolution de la littérature au cours des dernières décennies (paragraphe 4), et nous en voyons les conséquences possibles (paragraphe 5). Nous illustrons ensuite notre propos en présentant quelques exemples numériques (paragraphe 6 à 9) et nous terminons par quelques conclusions (paragraphe 10).

2. BREF RAPPEL HISTORIQUE

Bien qu'on puisse trouver des prémisses des notions contemporaines d'expérimentation au cours des périodes antérieures [Cochran, 1976 ; Droesbeke *et al.*, 1997b ; Leclercq, 1960], nous traitons l'aspect historique du plan d'expérience en considérant presque exclusivement les années 1920 à 1960.

On peut en effet estimer que le point de départ des principes modernes d'expérimentation est l'engagement en 1919 de Ronald A. Fisher (1890-1962) à la *Rothamsted Experimental Station*¹ [Box, 1978 ; Preece, 1990 ; Yates, 1964]. Très rapidement, on trouve dans les travaux de Fisher les notions de répétition (*replication*), de répartition au hasard ou randomisation (*randomization*), de constitution de blocs (*blocking*), d'expérience factorielle (*factorial experiment*), de confusion d'interactions ou d'effets principaux et d'interactions (*confounding*) et d'expérience en parcelles divisées (*split-plot experiment*) [Fisher, 1925, 1926]².

En 1931, Fisher est rejoint à Rothamsted par Frank Yates (1902-1993). De leur collaboration, résultent entre autres de nouveaux développements en ce qui concerne les expériences factorielles, ainsi que les notions d'expérience en blocs aléatoires incomplets (*incomplete randomized blocks*) et d'expérience factorielle fractionnaire (*fractional factorial experiment*) [Yates, 1935, 1936, 1937].

Deux autres noms méritent d'être cités dans la même lignée, à savoir ceux de William G. Cochran (1909-1980) et de David J. Finney (1917-). On doit notamment à ces auteurs des contributions importantes en ce qui concerne les expériences croisées ou avec permutation des traitements (*cross-over* ou *change-over experiment*) [Cochran, 1939 ; Cochran *et al.*, 1941], ainsi que les expériences factorielles fractionnaires et les répétitions fractionnaires (*fractional replication*) [Finney, 1945, 1946].

A partir des années 1935-1940, les notions d'expérimentation, qui jusque là avaient été conçues et appliquées essentiellement dans le secteur agronomique, interviennent également dans le secteur industriel. Des concepts nouveaux

1. Centre de recherche agronomique situé à une cinquantaine de kilomètres au nord de Londres.

2. Dans les publications de Fisher, il est abondamment question de la notion de carré latin (*Latin square*), mais le principe en avait déjà été appliqué à diverses reprises longtemps auparavant [Droesbeke *et al.*, 1997b ; Federer et Balaam, 1972 ; Preece, 1990].

apparaissent alors, tels que les plans de Plackett et Burman (*Plackett-Burman design*) [Plackett et Burman, 1946], la notion de surface de réponse (*response surface*) [Box, 1952 ; Box et Wilson, 1951], l'application de cette notion au cas particulier des mélanges (*mixture design*) [Scheffé, 1958], les plans optimaux (*optimal designs*) [Kiefer, 1959] et les plans ou l'approche de Taguchi [1959, 1960, 1987].

Parallèlement, le secteur de la recherche médicale ou pharmaceutique (ou encore biomédicale ou biopharmaceutique) se développe considérablement et devient un des principaux domaines d'application des plans d'expériences, à la suite notamment des travaux de A.B. Hill [Ederer, 1998 ; Matthews, 1998].³

3. LE CONTENU DU PLAN D'EXPÉRIENCE OU PROTOCOLE EXPÉRIMENTAL

Nous considérons que les principaux éléments constitutifs d'un *plan d'expérience* (ou d'un *protocole expérimental*) sont normalement [Dagnelie, 1981, 1997] :

1. la définition du ou des *buts* et des *conditions* de l'expérience ;
2. la définition du ou des *facteurs* dont on désire étudier l'influence, de ses ou de leurs niveaux ou variantes ou modalités, et des combinaisons de ces niveaux, variantes ou modalités qui seront effectivement expérimentées ;
3. la définition des individus ou, d'une manière plus générale, des *unités expérimentales* qu'on se propose d'observer, tels que des patients ou des groupes de patients, des animaux ou des plantes ou des groupes d'animaux ou de plantes, ou diverses installations particulières que sont notamment des fermenteurs, des fours, des étuves, etc. ;
4. la définition des *observations* qui devront être réalisées et des modalités de collecte de ces observations ;
5. la manière dont les différentes modalités ou combinaisons de modalités des facteurs devront être affectées aux différentes unités expérimentales et la *répartition* des unes et des autres dans l'espace et/ou dans le temps ;
6. des informations assez détaillées relatives à l'*analyse des résultats*, tel qu'un schéma d'analyse de la variance et/ou la formulation d'une ou plusieurs équations de régression, y compris la spécification précise du ou des *modèles* envisagés.

Les différents éléments auxquels nous avons fait allusion au cours de notre bref rappel historique (paragraphe 2) ont essentiellement trait aux points

3. Des informations complémentaires relatives aux publications antérieures à 1968 peuvent être trouvées dans l'importante étude bibliographie de Federer et Balaam [1972], qui réunit plus de 8000 références.

2 et 5. Le point 2, auquel correspond la notion de *plan ou schéma des traitements (treatment design)*, fait intervenir les concepts d'expérience factorielle, d'expérience factorielle fractionnaire, de surface de réponse, de plan de Plackett et Burman, de plan optimal, etc. Le point 5, qui définit le *dispositif ou plan expérimental (experimental design)*, prend en considération les concepts de randomisation, de constitution de blocs, complets ou incomplets, de parcelles divisées, de cross-over, etc.

Nous entendons donc par *expérience* l'ensemble des opérations qui permettent d'étudier l'influence d'un ou plusieurs facteurs sur un phénomène donné, conformément notamment à l'esprit de la norme internationale ISO de 1985 :

Plan d'expérience, plan expérimental : ensemble des modalités selon lesquelles un programme expérimental doit être réalisé et choix des variantes (niveaux) d'un ou de plusieurs facteurs, ou des combinaisons de facteurs, à introduire dans l'expérience [X, 1985].

Cependant, le mot « expérience » est aussi utilisé fréquemment, en particulier en France dans le secteur industriel, pour désigner un élément isolé du dispositif global, à savoir la réalisation d'une seule manipulation ou d'un seul essai, relatif à l'application d'un traitement donné à une unité expérimentale donnée. Tel est d'ailleurs la conception qui a été adoptée dans le cadre de la norme française AFNOR :

Expérience ou essai

Une expérience est définie par l'application d'un traitement à une unité expérimentale [X, 1989].

L'expression « plan d'expériences » nécessite alors l'utilisation du pluriel, comme le montre d'ailleurs bien la même norme.

On notera aussi que, dans sa version de 1998, la norme internationale ISO précise :

Plan d'expérience

Affectation de traitements à chaque unité expérimentale ainsi que l'ordre temporel selon lequel les traitements doivent être appliqués.

et en introduction :

Les plans d'expérience constituent essentiellement une stratégie de planification d'expériences afin d'obtenir des conclusions solides et adéquates de manière efficace et économique. Il convient que le choix du plan d'expérience dépende de la nature des questions à traiter, du degré de généralité recherché pour les conclusions, et des ressources disponibles (matériau expérimental, personnel, contraintes de temps) [X, 1998].

Dans la suite de cet article, en ce qui concerne les termes spécifiques relatifs aux questions d'expérimentation, nous adopterons d'une manière générale les acceptions de cette dernière norme.

4. LA PLACE CONSACRÉE AU CHOIX DES TRAITEMENTS ET AUX DISPOSITIFS EXPÉRIMENTAUX

Dans l'ensemble des publications qui ont été citées au paragraphe 2, un certain équilibre existe globalement entre la place consacrée d'une part au choix des traitements et d'autre part aux dispositifs expérimentaux. Cet équilibre apparaît aussi dans la bibliographie de Federer et Balaam [1972] qui contient, pour les années 1950 à 1967, 1.939 références relatives à l'aspect *treatment design* et 1.683 références relatives à l'aspect *experimental design*.

Il en est de même encore en ce qui concerne les livres qui datent du milieu du xx^e siècle, avec cependant une prédominance parfois importante des dispositifs expérimentaux. Au cours des dernières années par contre, les problèmes de choix des traitements ont pris une place prépondérante, voire même exclusive dans certains cas, au détriment des divers autres aspects du plan d'expérience, et notamment des dispositifs expérimentaux. Les questions de répétition et de répartition au hasard ou de randomisation sont alors aussi largement ou totalement négligées.

On peut mettre cette évolution en évidence en considérant le contenu des livres publiés en français qui traitent d'expérimentation⁴.

Jusqu'au milieu des années 1980, ces livres abordaient le sujet uniquement ou presque uniquement sous l'angle des dispositifs expérimentaux [Douaire, 1985; Dugué et Girault, 1959; Vessereau, 1948] ou, du moins, accordaient une place souvent importante à cet aspect des choses [Dagnelie, 1981; Lellouch et Lazar, 1974; Lison, 1958; Philippeau, 1973; Schwartz et al., 1970]. Ultérieurement au contraire, l'accent a été mis plutôt, et parfois de façon quasi exclusive, sur le choix des traitements [Alexis, 1995; Alexis et Alexis, 2000; Benoist et al., 1995; Collombier, 1996; Demonsant, 1996; Goupy, 1988, 1999; Pillet, 1992, 1997; Sado et Sado, 1991; Vigier, 1988].

L'ouvrage collectif de Droesbeke *et al.* [1997a], de même que certains articles de synthèse, tels ceux de Goupy [1990] et Louvet [2000], et divers documents polycopiés, tels ceux de Louvet [1996] et Mozzo [1995], confirment cette tendance.

On peut penser que l'évolution ainsi observée est largement liée à une modification du domaine d'application pris en considération. Les premiers ouvrages que nous avons cités concernent en effet principalement les secteurs agronomique, biologique et médical, tandis que les suivants ont trait essentiellement aux applications industrielles⁵. Le livre de Bergonzini [1995], à orientation agronomique, est d'ailleurs le seul ou un des seuls ouvrages

4. En ce qui concerne les très nombreux livres publiés en anglais, on peut consulter, pour les années antérieures à 1982, les travaux de synthèse de Federer et Balaam [1972], Hahn [1982] et Posten [1974]. Le premier présente les tables des matières d'une vingtaine d'ouvrages; le deuxième donne des commentaires à propos d'une cinquantaine de livres; et le troisième fournit les références de plusieurs centaines de recensions d'ouvrages de statistique en général.

5. On pourrait mentionner aussi quelques autres livres, tels ceux de Benoist [1974], Chapouille [1973], Lochner et Matar [1992], Schimmerling *et al.* [1998], etc.

récents consacrés plus particulièrement, toujours en français, aux dispositifs expérimentaux.

On notera aussi que le secteur médical ou pharmaceutique reste toujours attaché à une vision globale de la notion de plan d'expérience, ainsi qu'en témoigne notamment le document adopté par les experts de l'*International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use* [X, 1999].

5. QUELQUES CONSÉQUENCES

Le fait de concentrer l'attention exclusivement ou presque exclusivement sur le choix des traitements lors de l'élaboration d'un protocole expérimental peut avoir les conséquences suivantes.

D'une part, on néglige la notion de répétition (ou réplique) et, de ce fait, on n'a pas de connaissance précise de la variabilité qui est relative aux conditions expérimentales et au matériel étudié, variabilité qui est souvent qualifiée de résiduelle ou expérimentale. On ne peut donc pas avoir d'idée précise de la qualité des estimations et des tests d'hypothèses auxquels on procède ou auxquels on pourrait procéder.

Dans certains cas, on dispose d'estimations basées sur un très petit nombre d'observations ou caractérisées par un très petit nombre de degrés de liberté (parfois 2 ou 3 degrés de liberté seulement). Les estimations et les éventuels tests d'hypothèses sont alors de très faible qualité.

D'autre part, on néglige la notion de randomisation, ce qui peut induire des erreurs systématiques importantes.

Enfin, on néglige les possibilités de réduction de la variabilité résiduelle, et donc d'amélioration des estimations et des éventuels tests d'hypothèses, par l'introduction de blocs ou d'autres dispositifs expérimentaux⁶.

Nous illustrerons ces différents points à l'aide de quelques exemples au cours des paragraphes suivants (paragraphes 6 à 9).

Les objections que nous venons de formuler sont sans importance aucune quand toutes les conditions de l'expérience sont parfaitement stables. Tel est le cas par exemple lors de certaines simulations réalisées sur ordinateur.

Ces objections peuvent aussi être d'importance relativement secondaire quand la variabilité résiduelle est très faible, par comparaison avec les effets des facteurs qu'on souhaite ou qu'on espère mettre en évidence. Le fait de considérer la variabilité résiduelle comme tout à fait secondaire explique d'ailleurs sans doute la différence d'approche que nous avons soulignée entre

6. Nous considérons les expressions *variabilité résiduelle*, *variance résiduelle*, *écart-type résiduel*, etc. comme englobant l'ensemble des écarts entre les valeurs observées au cours de l'expérience et le modèle théorique sous-jacent, y compris donc l'influence des facteurs qui ne sont pas pris en considération dans le modèle sous-jacent.

LA PLANIFICATION DES EXPÉRIENCES

les secteurs agronomique, biologique et médical d'une part et le secteur industriel d'autre part.

Nous pensons cependant que ce point de vue peut être fallacieux et dangereux. Notre expérience de la consultation statistique, très largement agronomique, mais orientée aussi vers la chimie et les industries agroalimentaires, nous a souvent montré que la variabilité résiduelle était en réalité largement supérieure à ce que les chercheurs en pensaient *a priori*.

Mais le problème se corse quand les mêmes principes sont transposés, par ignorance ou par facilité, du domaine industriel dans des domaines dont il est bien connu que la variabilité est importante. Or, il n'est pas rare à l'heure actuelle de voir les principes et les modèles développés dans le secteur industriel appliqués de façon inconsidérée dans le domaine biologique par exemple.

6. L'INFLUENCE DU NOMBRE DE DEGRÉS DE LIBERTÉ DE LA VARIANCE RÉSIDUELLE SUR LES ESTIMATIONS ET LES TESTS D'HYPOTHÈSES

6.1. Quelques chiffres

Le tableau 1 donne diverses informations relatives à l'influence que peut avoir le nombre de degrés de liberté de la variance résiduelle sur la qualité des résultats obtenus en matière d'estimations et de tests d'hypothèses. Le contenu des différentes colonnes de ce tableau est précisé ci-dessous.

TABLEAU 1. — Quelques résultats relatifs à l'influence du nombre de degrés de liberté de la variance résiduelle sur les estimations et les tests d'hypothèses.

Nb. de degrés de lib.	Estimations d'écart-types résiduels			Estimations de moyennes			Tests d'égalité de deux moyennes		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
2	12,07	3,07	4,81	4,30	1,67	1,93	0,197	4,94	10,15
3	6,58	1,68	2,62	3,18	1,24	1,43	0,092	2,30	4,73
4	4,80	1,22	1,91	2,78	1,08	1,25	0,056	1,39	2,86
5	3,93	1,00	1,56	2,57	1,00	1,15	0,040	1,00	2,06
6	3,42	0,87	1,36	2,45	0,95	1,10	0,032	0,79	1,63
7	3,08	0,78	1,23	2,36	0,92	1,06	0,027	0,67	1,37
8	2,84	0,72	1,13	2,31	0,90	1,04	0,023	0,58	1,20
9	2,65	0,68	1,06	2,26	0,88	1,02	0,021	0,53	1,09
10	2,51	0,64	1,00	2,23	0,87	1,00	0,019	0,49	1,00
20	1,89	0,48	0,75	2,09	0,81	0,94	0,013	0,34	0,69
30	1,67	0,43	0,67	2,04	0,79	0,92	0,012	0,30	0,61

6.2. Les estimations d'écarts-types résiduels et de variances résiduelles

La première colonne du tableau 1 contient les différents nombres de degrés de liberté que nous prenons en considération et les colonnes (1) à (3) concernent les estimations d'écarts-types résiduels.

La colonne (1), déduite des distributions χ^2 de Pearson, indique ce que vaut le rapport de la limite supérieure s_2 à la limite inférieure s_1 de l'intervalle de confiance classique de l'écart-type, pour un degré de confiance égal à 0,95, c'est-à-dire le quotient [Dagnelie, 1998, § 7.2.2] :

$$s_2/s_1 = \sqrt{\chi_{0,975}^2 / \chi_{0,025}^2}.$$

Ainsi, pour 2 degrés de liberté, la limite supérieure de l'intervalle de confiance est égale à environ 12 fois sa limite inférieure, et pour 10 degrés de liberté, le rapport de la limite supérieure à la limite inférieure est proche de 2,5. On notera aussi que les valeurs correspondantes relatives à la variance sont égales aux carrés des valeurs de la colonne (1), soit environ 146 pour 2 degrés de liberté et 6,3 pour 10 degrés de liberté.

La colonne (2) du tableau 1 exprime les valeurs de la colonne (1) en proportion de la valeur relative à 5 degrés de liberté, et la colonne (3) a été obtenue de la même manière en fonction de la valeur relative à 10 degrés de liberté. Il s'agit donc des quotients $(s_2/s_1)_k / (s_2/s_1)_5$ et $(s_2/s_1)_k / (s_2/s_1)_{10}$, les indices qui figurent en dehors des parenthèses désignant chaque fois le nombre de degrés de liberté, qui est représenté par k d'une manière générale.

Ces colonnes donnent une idée de l'imprécision, ou de la perte de précision, relative à de très petits nombres de degrés de liberté par comparaison, respectivement, avec 5 et avec 10 degrés de liberté.

Si, par exemple, on obtient une valeur estimée de l'écart-type résiduel égale à 2, les limites de l'intervalle de confiance qui concerne ce paramètre sont égales à 1,40 et 3,51 quand on dispose de 10 degrés de liberté, et à 1,04 et 12,6 quand on dispose de 2 degrés de liberté. On observe bien les rapports 2,5 (ou 2,51) dans le premier cas et 12 (ou 12,07) dans le deuxième cas, entre les limites supérieures et inférieures des intervalles de confiance. Et on vérifie aussi le fait que le rapport en question est multiplié par 4,8 (ou 4,81) quand on passe de 10 à 2 degrés de liberté, comme l'indique la colonne (3) du tableau 1.

Les résultats correspondants relatifs aux variances résiduelles sont par exemple, pour une valeur estimée de cette variance égale à 4, d'une part 1,95 et 12,3 avec 10 degrés de liberté, et d'autre part 1,08 et 158 avec 2 degrés de liberté. Le rapport entre les valeurs extrêmes de l'intervalle de confiance est donc bien égal à 6,3 dans le premier cas et à 146 dans le deuxième cas, le rapport étant multiplié par 23 (soit le carré de 4,8 ou 4,81) quand on passe de 10 à 2 degrés de liberté.

6.3. Les estimations de moyennes

Les colonnes (4) à (6) du tableau 1 concernent ensuite les estimations de moyennes, de différences de deux moyennes ou de contrastes, qui sont des fonctions linéaires de deux ou plusieurs moyennes.

La colonne (4) contient les valeurs $t_{0,975}$ de la variable t de Student relatives au quantile 0,975, c'est-à-dire les valeurs qui interviennent dans le calcul d'intervalles de confiance caractérisés par un degré de confiance égal à 0,95 [Dagnelie, 1998, § 8.2.2]. Quant aux colonnes (5) et (6), elles expriment, comme pour les écarts-types, les valeurs de la colonne (4) en proportion des valeurs relatives, respectivement, à 5 et à 10 degrés de liberté.

Ces colonnes illustrent la perte de précision qui est liée à un éventuel très petit nombre de degrés de liberté, dans l'estimation d'une moyenne ou d'une fonction linéaire de deux ou plusieurs moyennes, et cela indépendamment de la perte de précision qui peut résulter de différences éventuelles de nombres d'observations intervenant dans le calcul de la ou des moyennes elles-mêmes.

Supposons par exemple qu'on ait obtenu, pour une différence de deux moyennes, une valeur estimée égale à 4,5 et, avec 10 degrés de liberté, des limites de confiance égales à 3,1 et 5,9 (sans préciser ici les unités de mesure). La demi-longueur de l'intervalle de confiance, aussi appelée marge d'erreur, est donc égale à 1,4.

Comme l'indique la colonne (6), cette valeur est multipliée par 1,93 quand on passe, dans des conditions identiques, de 10 à 2 degrés de liberté. La marge d'erreur est alors égale à 2,7, de telle sorte que les limites de confiance sont égales à 1,8 et 7,2, au lieu de 3,1 et 5,9.

6.4. Les tests d'égalité de deux moyennes

Enfin, les colonnes (7) à (9) du tableau 1 sont relatives aux tests d'égalité de deux moyennes.

En vue d'isoler l'influence du nombre de degrés de liberté de la variance résiduelle, nous avons considéré qu'on souhaitait comparer deux moyennes, calculées chacune à partir de 8 observations, sur base d'une variance estimée avec 2, 3, ... degrés de liberté. Cette situation correspond par exemple à l'étude de l'influence d'un des facteurs d'une expérience factorielle complète de type 2^4 réalisée sans répétition, si on suppose qu'on possède par ailleurs une estimation de la variance résiduelle avec 2 ou 3 ou ... degrés de liberté. Il s'agit d'une situation comparable à celle des plans composites centraux relatifs à l'étude des surfaces de réponse où, en plus des points périphériques qui ne sont répétés qu'une seule fois, on dispose de plusieurs répétitions d'un point central [Dagnelie, 1981, § 2.3.4].

Nous considérons en outre à titre d'exemple que la différence théorique entre les deux moyennes est égale à deux fois l'écart-type, que le test est unilatéral (la moyenne relative à la deuxième modalité du facteur considéré étant supposée *a priori* supérieure à la moyenne relative à la première modalité), et que le niveau de signification du test est égal à 0,05.

La colonne (7) du tableau 1 donne dans ces conditions les valeurs du risque de deuxième espèce (probabilités de ne pas mettre en évidence la différence supposée), ces résultats étant déduits notamment des tables des distributions t non centrées de Bulgren [1974]. Comme dans les deux cas précédents, les dernières colonnes expriment les différentes valeurs de la colonne (7) en proportion, respectivement, des valeurs relatives à 5 et à 10 degrés de liberté. Pour un écart-type résiduel égal par exemple à 3 et dans les conditions définies ci-dessus, la probabilité de ne pas mettre en évidence une différence de moyennes de 6 unités est égale à 0,02 (ou 0,019) si on dispose de 10 degrés de liberté et à 0,20 (ou 0,197) si on dispose de 2 degrés de liberté. Le risque de deuxième espèce est donc multiplié par 10 (ou 10,15) quand on passe, dans des conditions identiques, de 10 à 2 degrés de liberté, comme l'indique la colonne (9).

6.5. Synthèse

Le tableau 1 montre ainsi clairement que, par comparaison avec les situations où on dispose de 5 ou de 10 degrés de liberté, le fait de n'avoir qu'un très petit nombre de degrés de liberté peut conduire à une augmentation considérable de la longueur de l'intervalle de confiance de l'écart-type résiduel (jusque 3 ou environ 5 fois plus) ou de la variance résiduelle (jusque 9 ou 23 fois plus), à une augmentation moindre mais néanmoins importante de la longueur des intervalles de confiance des moyennes et des fonctions linéaires des moyennes (jusque 70 à 90 % de plus), et à une augmentation importante du risque de deuxième espèce des tests d'égalité de deux moyennes (jusque 5 à 10 fois plus dans le cas envisagé)⁷.

La figure 1 présente sous forme graphique les différentes proportions des colonnes (3), (6) et (9) du tableau 1, ainsi que les valeurs correspondantes relatives aux variances, c'est-à-dire les carrés des valeurs de la colonne (3).

7. On notera que nous n'avons pas pris en considération le cas, qui serait encore plus « pathologique », où on ne disposerait que d'un seul degré de liberté. D'autre part, nous ne donnons ici, comme dans les paragraphes suivants, qu'une partie des résultats que nous avons obtenus par calcul ou par simulation et qui vont dans le même sens.

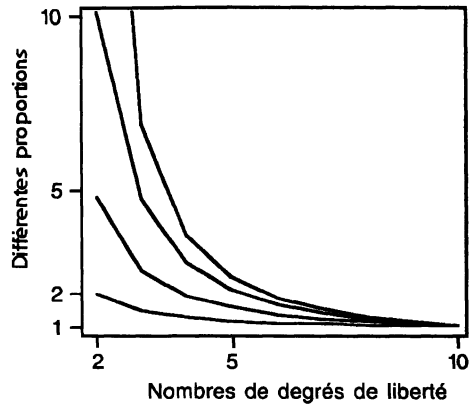


FIG 1. — Représentation graphique des proportions des colonnes (3), (6) et (9) du tableau 1, et des résultats correspondants relatifs aux variances, pour 2 à 10 degrés de liberté : les différentes lignes brisées concernent respectivement, de haut en bas, les estimations de variances, les tests d'égalité de deux moyennes, les estimations d'écarts-types et les estimations de moyennes ou de fonctions linéaires de moyennes.

7. LA RECHERCHE DU MAXIMUM D'UNE COURBE OU D'UNE SURFACE DE RÉPONSE

7.1. Le cas d'une courbe de réponse

En vue d'illustrer ce qu'on peut raisonnablement attendre d'expériences ne comportant qu'un petit ou un très petit nombre d'observations, nous considérons le problème relativement courant de la localisation du maximum de la réponse, dans le cas d'un seul facteur tout d'abord.

Plus concrètement, nous envisageons la situation où on réalise des observations à trois niveaux équidistants d'un seul facteur, désignés conventionnellement par -1 , 0 et $+1$. Nous supposons en outre que la valeur théorique de la variable (ou de la réponse) qui doit être observée est égale à 0 en -1 et en $+1$ et à une valeur positive k en 0 , la courbe de réponse théorique étant donc une parabole d'équation $y = k(1 - x^2)$.

Le tableau 2 est relatif à 20 cas différents. D'une part, le nombre d'observations est fixé à 4, 6, 9 ou 12, à savoir une observation à chacune des extrémités du domaine considéré ($x = -1$ et $x = +1$) et deux observations au centre de ce domaine ($x = 0$), deux observations en chacun des trois points (c'est-à-dire deux répétitions complètes), trois observations en chacun des trois points (trois répétitions) et quatre observations en chacun des trois points (quatre répétitions). D'autre part, nous supposons que la constante k , c'est-à-dire l'influence du facteur envisagé dans le domaine considéré, est égal à 10 fois, 5 fois ou 2 fois l'écart-type résiduel, ou à cet écart-type lui-même ou à la moitié de cet écart-type.

LA PLANIFICATION DES EXPÉRIENCES

TABLEAU 2. — Quelques résultats relatifs à la recherche du maximum d'une courbe de réponse parabolique.

Nb. d'obser- vations	Abscisses des maximums					Intervalles de confiance des abscisses				
	Influence du facteur					Influence du facteur				
	10	5	2	1	0,5	10	5	2	1	0,5
	Pourcentages de maximums non observés					Pourcentages de limites non obtenues				
4	0,0	0,0	2,2	16,0	31,2	43,4	69,3	86,4	91,2	93,8
6	0,0	0,0	1,2	12,6	28,2	0,0	3,6	58,0	82,2	89,9
9	0,0	0,0	0,1	7,6	24,4	0,0	0,0	27,6	70,1	85,2
12	0,0	0,0	0,1	5,2	20,3	0,0	0,0	13,6	59,7	81,8
	Médianes des abscisses					Médianes des longueurs				
4	0,024	0,048	0,124	0,265	0,561	1,181	-	-	-	-
6	0,017	0,034	0,088	0,185	0,394	0,148	0,331	-	-	-
9	0,014	0,028	0,070	0,146	0,315	0,096	0,203	0,763	-	-
12	0,012	0,024	0,060	0,127	0,264	0,078	0,161	0,516	-	-
	Neuvièmes déciles des abscisses					Neuvièmes déciles des longueurs				
4	0,059	0,123	0,471	-	-	-	-	-	-	-
6	0,041	0,086	0,286	-	-	0,253	0,933	-	-	-
9	0,034	0,068	0,203	1,538	-	0,138	0,329	-	-	-
12	0,029	0,059	0,169	0,742	-	0,105	0,235	-	-	-

Dans chacune des 20 situations, nous avons procédé à 10.000 simulations, en déterminant chaque fois, autant que possible, l'abscisse du maximum et les limites de confiance correspondantes, calculées par la méthode qui est décrite notamment par E.J. Williams [Dagnelie, 1998, § 15.2.2; Williams, 1959, § 6.11].

Au cours de l'application de cette procédure, des difficultés peuvent surgir à deux niveaux. Tout d'abord, on peut ne pas observer de maximum, parce que les valeurs simulées tendent à mettre en évidence un minimum (dérivée seconde positive) ou parce que la dérivée seconde est nulle. En outre, on peut ne pas obtenir de limites de confiance pour l'abscisse du maximum, en raison du fait que celui-ci n'est pas significatif.

Dans cette optique, la première partie du tableau 2 donne les pourcentages de cas où un maximum n'a pas été observé (première série de colonnes « 10 » à « 0,5 ») et les pourcentages de cas où les limites de confiance n'ont pas pu être obtenues (deuxième série de colonnes « 10 » à « 0,5 »). La deuxième partie de ce tableau donne les médianes des valeurs absolues des abscisses observées pour les maximums, c'est-à-dire les médianes des distances entre ces abscisses et la

valeur théorique 0, et les médianes des longueurs des intervalles de confiance, tandis que la troisième partie du tableau contient les neuvièmes déciles de ces distances et de ces longueurs^{8 9}.

Ainsi, pour 9 observations (trois en chacun des trois points -1 , 0 et $+1$) et avec une influence du facteur considéré égale à deux fois l'écart-type résiduel (colonnes «2»), on n'a pas observé de maximum dans 0,1 % des cas, et on n'a pas obtenu de limites de confiance dans 27,6 % des cas; la distance qui sépare l'abscisse du maximum de l'origine est inférieure ou égale à 0,070 dans un cas sur deux, et inférieure ou égale à 0,203 dans neuf cas sur dix; et la longueur de l'intervalle de confiance est inférieure ou égale à 0,763 dans un cas sur deux, le neuvième décile de la distribution des longueurs n'étant pas déterminé, puisque plus de 10 % des valeurs sont manquantes.

A la vue de ce tableau, on constate qu'on ne peut raisonnablement espérer mettre en évidence un maximum, même avec trois ou quatre répétitions complètes, que si l'influence du facteur étudié est au moins de l'ordre de grandeur de l'écart-type résiduel (colonnes «10» à «1»), dans le domaine considéré. Encore faut-il remarquer que, même dans ce cas, la position estimée du maximum peut être très éloignée de la réalité (neuvième décile égal à 1,5 pour trois répétitions et à 0,74 pour quatre répétitions) et le calcul de limites de confiance peut être fréquemment ou très fréquemment impossible (absence de limites de confiance dans 91 à 60 % des cas).

7.2. Le cas d'une surface de réponse

Nous envisageons aussi le même problème pour deux facteurs, dans le cas particulier d'un dispositif composite central formé, dans le plan des deux facteurs, des huit sommets d'un octogone régulier et d'un point central. Les sommets de l'octogone sont situés conventionnellement à une distance unitaire du point central, de telle sorte que les coordonnées des neuf points sont $(0, 0)$, $(0, 1)$, $(0, 707, 0, 707)$, $(1, 0)$, $(0, 707, -0, 707)$, $(0, -1)$, $(-0, 707, -0, 707)$, $(-1, 0)$ et $(-0, 707, 0, 707)$, comme le montre la figure 2.

Par analogie avec le cas d'un seul facteur, nous supposons que la valeur théorique de la variable qui doit être observée est égale à 0 aux huit sommets de l'octogone et à une valeur positive k au point central, la surface de réponse théorique étant donc un parabolôïde de révolution d'équation $y = k(1 - x_1^2 - x_2^2)$. Quant à la valeur k , elle est définie comme précédemment en proportion de l'écart-type résiduel.

Enfin, nous considérons le cas où on dispose d'une observation pour chacun des points périphériques et de six observations pour le point central (premier

8. L'utilisation des médianes et des neuvièmes déciles est justifiée par le fait que les distributions des distances par rapport à l'origine et des longueurs des intervalles de confiance sont fortement dissymétriques. De plus, nous entendons ici chaque fois par « médiane » l'observation de rang 5.000 et par « neuvième décile » l'observation de rang 9.000 (sur 10.000), les valeurs observées étant classées par ordre croissant et les valeurs manquantes étant rangées en fin de série.

9. Une partie des résultats du tableau 2 (principalement les pourcentages de maximums non observés) peuvent être recherchés de manière analytique. Par souci d'homogénéité, nous ne donnons cependant que des résultats obtenus par simulation.

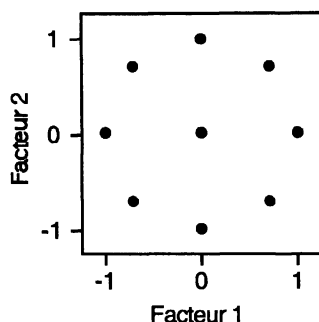


FIG 2. — Positions des neuf points expérimentaux dans le cas envisagé pour deux facteurs.

cas « 14 »), le cas où on dispose d'une observation pour un point périphérique sur deux et de deux observations pour le point central et chacun des autres points périphériques (deuxième cas « 14 »), ainsi que les cas où les neuf points sont tous l'objet de deux répétitions (cas « 18 ») ou de trois répétitions (cas « 27 »).

Dans chacune des 20 situations ainsi définies, nous avons également effectué 10.000 simulations.

Comme le tableau 2, le tableau 3 contient d'une part les pourcentages de cas où un maximum n'a pas été observé, puis les médianes des distances, dans le plan (x_1, x_2) , entre le point correspondant au maximum et l'origine, et enfin les neuvièmes déciles de ces distances. Le tableau 3 ne donne par contre aucune information quant à d'éventuelles régions de confiance des coordonnées des maximums.

Avec trois répétitions par exemple (lignes « 27 ») et pour une influence des facteurs considérés égale à deux fois l'écart-type résiduel (colonne « 2 »), on n'observe pas de maximum dans 0,4 % des cas, et la distance entre l'origine du système d'axes et le point du plan (x_1, x_2) correspondant au maximum est inférieure ou égale à 0,089 dans un cas sur deux, et inférieure ou égale à 0,212 dans neuf cas sur dix.

Globalement, les résultats obtenus sont très semblables à ceux qui concernent le cas d'un seul facteur. Ils conduisent pratiquement à la même conclusion, à savoir que dans les conditions envisagées, il ne faut pas espérer pouvoir mettre en évidence un maximum si l'influence des facteurs étudiés dans le domaine considéré n'est pas au moins de l'ordre de grandeur de l'écart-type résiduel.

Incidentement, on notera aussi que le deuxième dispositif « 14 » (une observation pour un point périphérique sur deux et deux observations pour le point central et chacun des autres points périphériques), dont il est en général très peu question, s'avère préférable au premier dispositif « 14 » (une observation pour chacun des points périphériques et six observations pour le point central), dont l'utilisation est fréquemment recommandée.

Enfin, de nombreuses autres simulations pourraient être réalisées également en s'écartant des conditions de normalité des distributions et d'égalité des

LA PLANIFICATION DES EXPÉRIENCES

TABLEAU 3. – Quelques résultats relatifs à l'estimation des coordonnées du maximum d'une surface de réponse parabolique.

Nb. d'observ.	Influence des facteurs				
	10	5	2	1	0,5
	Pourcentages de maximums non observés				
14	0,0	0,0	0,7	16,5	43,7
14	0,0	0,0	1,6	18,0	40,1
18	0,0	0,0	1,7	18,6	38,8
27	0,0	0,0	0,4	11,7	33,0
	Médianes des distances à l'origine				
14	0,030	0,061	0,170	0,500	3,352
14	0,024	0,049	0,136	0,342	1,346
18	0,021	0,042	0,114	0,281	1,007
27	0,017	0,035	0,089	0,207	0,657
	Neuvièmes déciles des distances				
14	0,054	0,112	0,467	-	-
14	0,044	0,095	0,480	-	-
18	0,039	0,080	0,344	-	-
27	0,031	0,065	0,212	-	-

variances des écarts par rapport au modèle théorique. Nous n'avons pas envisagé cette possibilité.

8. LA RÉPARTITION AU HASARD OU RANDOMISATION

8.1. Un cas d'école

En vue d'illustrer les problèmes qui peuvent résulter de l'absence de répartition au hasard ou randomisation, nous considérons tout d'abord le cas d'une expérience factorielle complète de type 2^4 , dont les 16 essais ou manipulations seraient réalisés de manière systématique dans le temps. Nous supposons que les effets des quatre facteurs considérés sont égaux respectivement à 0,5, 1, 1,5 et 2 fois l'écart-type résiduel.

Le tableau 4 contient ces valeurs théoriques et donne en outre les valeurs attendues correspondantes relatives à deux situations particulières.

La colonne « Sans dérive » de ce tableau contient les valeurs qui doivent être attendues en l'absence de toute modification des conditions expérimentales dans le temps.

LA PLANIFICATION DES EXPÉRIENCES

TABEAU 4. – Effets théoriques de quatre facteurs en l'absence de dérive et en présence de deux dérives différentes.

Facteurs	Sans dérive	Dérive négative	Dérive positive
1	0,50	0,10	0,55
2	1,00	0,80	1,10
3	1,50	1,40	1,70
4	2,00	1,95	2,40

La colonne «Dérive négative» correspond au cas où existerait, au cours du temps et indépendamment de l'effet des quatre facteurs, une dérive ou un gradient régulier consistant en une évolution linéaire négative des conditions expérimentales, dont l'ampleur serait égale à 5 % de l'écart-type résiduel d'un essai au suivant. Une telle évolution pourrait résulter aussi bien d'une modification de certaines caractéristiques des matières premières utilisées (augmentation ou diminution progressive du degré d'humidité par exemple), que d'une modification des conditions de réalisation des essais eux-mêmes (augmentation ou diminution progressive d'une température ou d'un débit ou modification progressive du réglage d'un appareil de mesure par exemple).

En outre, nous avons supposé que les 16 essais étaient réalisés dans l'ordre systématique suivant :

1111, 1112, 1121, 1122, 1211, 1212, 1221, 1222,
2111, 2112, 2121, 2122, 2211, 2212, 2221, 2222,

le premier chiffre désignant chaque fois, conformément aux conventions habituelles, le niveau inférieur (chiffre 1) ou supérieur (chiffre 2) du premier facteur, le deuxième chiffre désignant le niveau inférieur ou supérieur du deuxième facteur, etc. On observera que, dans ces conditions, le niveau du premier facteur ne doit être modifié qu'une fois au cours de l'expérience, après le huitième essai ; le niveau du deuxième facteur trois fois, après le quatrième, le huitième et le douzième essais ; etc.

La colonne «Dérive positive» du même tableau a été obtenue dans des conditions similaires, mais en supposant que la dérive temporelle, d'une même ampleur, est positive au lieu d'être négative, et que la succession des essais est la suivante :

1111, 2111, 1211, 2211, 1121, 2121, 1221, 2221,
1112, 2112, 1212, 2212, 1122, 2122, 1222, 2222,

le niveau du quatrième facteur n'étant donc modifié qu'une fois au cours de l'expérience, le niveau du troisième facteur trois fois, etc.

Le tableau 4 donne ainsi une idée des erreurs systématiques importantes qui peuvent intervenir dans l'estimation des effets des différents facteurs, quand une expérience factorielle est réalisée de manière non aléatoire en présence de modifications des conditions expérimentales : dans les cas envisagés, l'effet

observé du premier facteur, qui est théoriquement égal à 0,50, peut aller de 0,10 à 0,55 ; l'effet du deuxième facteur, théoriquement égal à 1,00, peut atteindre 0,80 ou 1,10 ; etc.

On notera aussi que, pour de telles répartitions systématiques, d'éventuelles dérives strictement linéaires n'ont au contraire aucune influence sur l'estimation des différentes interactions de deux ou de plusieurs facteurs.

8.2. Un cas réel

Nous considérons d'autre part le cas réel, et non plus hypothétique, d'une expérience factorielle complète, qui fait intervenir quatre facteurs dont les nombres de niveaux sont respectivement 2, 2, 3 et 4 (soit 48 conditions ou points expérimentaux), et qui a été réalisée sans répétition [Dagnelie, 1998, exemple 11.4.1.]¹⁰. Nous envisageons ici le problème sous l'angle des tests d'hypothèses auxquels conduit l'analyse de la variance.

Les colonnes « Sans dérive » du tableau 5 donnent, pour les quatre facteurs, les valeurs observées des variables F de Fisher-Snedecor, obtenues après avoir effectué un regroupement de diverses interactions, et les probabilités P (p -values) correspondantes.

TABLEAU 5. – Principaux résultats de diverses analyses de la variance réalisées en l'absence de dérive et en présence de deux dérives différentes.

Facteurs	Sans dérive		Dérives égales à 1%		Dérive égale à 2 %	
	F	P	F	P	F	P
1	2,39	0,134	0,82-2,30	0,141-0,375	0,07-2,22	0,147-0,796
2	7,96	0,009	4,76-7,82	0,009-0,038	2,37-7,67	0,010-0,135
3	16,4	0,000	16,6-20,6	0,000-0,000	16,9-25,4	0,000-0,000
4	52,1	0,000	52,6-57,9	0,000-0,000	53,0-64,1	0,000-0,000

Ici également, nous avons supposé que l'expérience en question aurait pu être réalisée de manière systématique en présence d'une dérive linéaire dans le temps et nous avons fixé l'ampleur totale de la dérive éventuelle à 1 % ou à 2 % seulement, du premier au dernier des 48 essais, alors que les résultats analysés (rendements de carbonisation du bois) varient d'un essai à l'autre de 55,1 à 80,5 %, soit une amplitude de plus de 25 %. Nous avons en outre considéré les 24 possibilités différentes d'organisation de l'expérience de manière systématique, définies comme dans l'exemple précédent en fonction de l'ordre dans lequel les niveaux des quatre facteurs sont maintenus constants le plus longtemps possible.

Les quatre dernières colonnes du tableau 5 donnent les valeurs extrêmes qui ont été obtenues de cette manière en ce qui concerne les variables F de Fisher-Snedecor et leurs probabilités P , pour les deux dérives considérées.

10. Les données dont il est question ici peuvent également être trouvées dans les pages Web www.fsagx.ac.be/aides_pdagogiques/livres_dagnelie (fichier S2E11041).

Le cas du deuxième facteur est le plus instructif. Pour ce facteur et avec une dérive d'une amplitude égale à 1 %, on peut observer des valeurs F allant de 4,76 à 7,82, au lieu de 7,96, et des probabilités P allant de 0,009 à 0,038, au lieu de 0,009. Les probabilités observées varient donc dans un rapport de 1 à 4 et le caractère hautement significatif du facteur considéré (significatif au niveau de probabilité 0,01) peut devenir simplement significatif (significatif au niveau 0,05).

Avec une dérive d'une amplitude égale à 2 %, les valeurs F varient de 2,37 à 7,67, c'est-à-dire dans un rapport de 1 à 3, et les probabilités P varient de 0,010 à 0,135, c'est-à-dire dans un rapport de 1 à 13 ou 14. L'examen détaillé des résultats des analyses de la variance relatives aux 24 situations envisagées montre aussi que l'influence du deuxième facteur reste proche du niveau de signification 0,01 dans 6 cas sur 24, devient simplement significative (significative au niveau 0,05) dans 12 cas sur 24, et est tout à fait non significative dans 6 cas sur 24 également. Comme nous l'avons dit ci-dessus, l'introduction d'une variation systématique linéaire du rendement dans le temps ne modifie en rien les résultats relatifs aux diverses interactions.

On constate ainsi l'importance que peuvent avoir des modifications des conditions expérimentales dont on ne tiendrait pas compte, même quand ces modifications sont d'une ampleur relativement limitée.

9. LA RANDOMISATION PARTIELLE ET LA CONSTITUTION DE BLOCS

9.1. La randomisation partielle

Dans de nombreux cas, la répartition complètement au hasard à laquelle nous nous sommes référés au cours du paragraphe précédent est difficilement réalisable, car elle peut impliquer par exemple de nombreux changements de réglages, longs ou délicats. D'autres solutions doivent donc être envisagées.

En ce qui concerne l'expérience 2^4 dont il a été question au début du paragraphe précédent, il se pourrait notamment que les modifications de niveaux de deux des facteurs (par exemple les deux premiers) soient sensiblement plus difficiles à réaliser que les modifications de niveaux des deux autres facteurs (par exemple les deux derniers). On peut alors envisager de réduire au minimum les modifications relatives aux deux premiers facteurs en adoptant la solution suivante :

**2112, 2122, 2121, 2111, 1121, 1122, 1111, 1112,
1222, 1221, 1211, 1212, 2211, 2221, 2212, 2222.**

Cette solution a été élaborée en répartissant tout d'abord au hasard les quatre combinaisons de niveaux relatives aux deux premiers facteurs (chiffres gras **11, 12, 21** et **22**), en décidant d'effectuer chaque fois de manière consécutive les quatre essais relatifs à chacune de ces combinaisons de niveaux, et en répartissant au hasard et de manière indépendante les deux niveaux de chacun

des deux autres facteurs au sein de chacun des quatre groupes d'essais ainsi définis (chiffres maigres 11, 12, 21 et 22).

Une telle randomisation partielle ou restreinte permet d'éviter toute erreur systématique dans l'estimation des effets des différents facteurs et de leurs interactions. En ce qui concerne les deux premiers facteurs, elle n'impose d'une manière générale qu'une, deux ou trois modifications de niveaux en cours d'expérience pour chacun des facteurs (dans l'exemple présenté ci-dessus, deux modifications pour le premier facteur et une seule modification pour le deuxième facteur), alors que les modifications de niveaux sont beaucoup plus fréquentes pour les deux autres facteurs (respectivement neuf et huit modifications pour le troisième et le quatrième facteurs, toujours dans l'exemple envisagé).

Le dispositif que nous venons de présenter s'apparente à la notion de parcelles divisées (ou subdivisées), qui est relativement courante en expérimentation agronomique¹¹. Il faut savoir qu'en présence de fluctuations des conditions expérimentales dans l'espace et/ou dans le temps, ce dispositif donne en général moins de précision pour les estimations et les comparaisons relatives aux facteurs qui sont l'objet de la première répartition au hasard (deux premiers facteurs) et plus de précision pour les estimations et les comparaisons relatives aux facteurs qui sont l'objet de la deuxième répartition au hasard (deux derniers facteurs)¹².

9.2. La constitution de blocs

La constitution de blocs ou de systèmes ou réseaux de blocs (lignes et colonnes par exemple) est un autre élément qui doit être envisagé dans l'élaboration d'un protocole expérimental. D'une manière générale, chacun des blocs est alors formé d'un ensemble d'unités expérimentales ou d'essais qui sont pris en considération dans des conditions aussi homogènes que possible, le but poursuivi étant de réduire autant que faire se peut la variabilité résiduelle.

A l'origine, la constitution de blocs impliquait l'existence de deux ou plusieurs répétitions de chacune des modalités ou combinaisons de modalités étudiées. Très rapidement cependant, la notion de blocs a été étendue au cas d'expériences réalisées en une seule répétition ou même en une fraction de répétition, et cela en particulier en ce qui concerne les expériences factorielles. Les blocs, d'abord complets, devenaient alors des blocs incomplets (équilibrés, partiellement équilibrés, avec *confounding*, etc.).

Nous n'envisagerons pas ici les diverses possibilités existantes, qui sont exposées notamment dans certains des ouvrages que nous avons cités au paragraphe 4, et nous nous limiterons à la présentation d'une solution qui s'adapte

11. On notera toutefois qu'il ne s'agit pas d'un dispositif en blocs aléatoires (ou randomisés) complets avec parcelles divisées, qui est le plus souvent envisagé, mais bien d'un dispositif complètement aléatoire (ou complètement randomisé) avec parcelles divisées.

12. On notera aussi que d'autres solutions ont été proposées d'une manière générale en vue de faire face notamment à l'existence de dérives ou de gradients linéaires dans le cas d'expériences factorielles [Bailey, 1985; Cheng et Steinberg, 1991; Coster, 1993; Dickinson, 1974; Goupy, 1989].

bien au cas du deuxième exemple du paragraphe 8 (expérience factorielle $2^2 \ 3 \ 4$). La constitution de blocs incomplets dans un tel cas n'est pas particulièrement aisée, mais le fait qu'on sache *a priori* qu'un des facteurs (le quatrième) doit avoir un effet très important permet d'élaborer une solution fort simple.

Dans la mesure où il n'est alors pas essentiel de chiffrer de façon tout à fait correcte l'effet de ce facteur, on peut le confondre avec un facteur blocs, en constituant quatre blocs de 12 essais. Le dispositif expérimental peut en conséquence être le suivant :

bloc 1 : 2234, 2224, 1234, 2124, 1114, 1134, 1214, 2214, 1124, 2134, 2114, 1224 ;
 bloc 2 : 2131, 1221, 2221, 1121, 2111, 1111, 2211, 1231, 1211, 1131, 2231, 2121 ;
 bloc 3 : 1123, 1223, 2113, 1113, 1233, 2233, 1133, 2123, 2213, 2133, 1213, 2223 ;
 bloc 4 : 1222, 2132, 2112, 1132, 1232, 2232, 2222, 2122, 1212, 1112, 2212, 1122.

Dans ce dispositif, le quatrième facteur (chiffres gras) a été l'objet d'une première répartition aléatoire et les trois autres facteurs (chiffres maigres) ont été l'objet, globalement, d'une deuxième répartition aléatoire et indépendante au sein des quatre blocs définis par les différents niveaux du quatrième facteur. On doit s'attendre dans ces conditions à d'éventuelles erreurs systématiques ou distorsions en ce qui concerne précisément le quatrième facteur, auquel on attache en fait moins d'intérêt, mais il ne faut craindre aucun problème en ce qui concerne les trois premiers facteurs.

Une nuance qui nous paraît importante, par comparaison avec le cas précédent, est le fait que dans le premier cas (expérience 2^4), chacun des niveaux des deux facteurs qui sont l'objet de la première randomisation sont répétés indépendamment à deux moments différents au cours de l'expérience, tandis que dans le deuxième cas (expérience $2^2 \ 3 \ 4$), chacun des niveaux du facteur qui est l'objet de la première randomisation ne figure qu'à un seul moment au cours de l'expérience. Ceci explique l'absence de toute erreur systématique dans le premier cas et la possible existence de certaines erreurs systématiques ou distorsions dans le deuxième cas, en ce qui concerne le quatrième facteur.

9.3. Synthèse

Le choix d'un dispositif expérimental doit ainsi toujours être l'objet d'une réflexion particulière, qui tient compte dans chaque cas de toutes les caractéristiques du problème considéré, le plus souvent dans une optique d'économie ou de moyens limités (économie ou moyens limités en temps, en personnel, en matériel, etc.).

A cet égard, on remarquera que les dispositifs en blocs, qui ont été développés essentiellement à propos des expériences à un seul facteur, des expériences factorielles et des expériences factorielles fractionnaires (ou fractionnées), peuvent être appliqués également au cas des surfaces de réponse, éventuellement en relation avec la notion de plans optimaux [Atkins et Cheng, 1999 ; Cook et Nachtsheim, 1989 ; Goos et Vandebroek, 1999].

10. CONCLUSIONS

Sauf dans certaines situations très particulières, les résultats d'expériences sont toujours fonction à la fois de l'effet du ou des facteurs qui sont étudiés et de fluctuations résiduelles liées notamment à d'autres facteurs qui ne sont pas maîtrisés. L'importance relative du *signal*, correspondant à l'effet du ou des facteur(s) étudié(s), et du *bruit de fond*, correspondant à l'effet des facteurs non maîtrisés et aux fluctuations propres du matériel expérimental, peut varier considérablement d'un cas à l'autre : le bruit de fond est par exemple sensiblement plus important, en général, dans le domaine agronomique, biologique ou médical que dans le domaine industriel. Mais quelle que soit leur importance, il nous semble que les variations résiduelles ne devraient jamais être perdues de vue.

Dans cette optique, l'évolution que nous avons mise en évidence et qui tend à concentrer souvent l'attention exclusivement ou presque exclusivement sur le choix des traitements nous semble particulièrement dangereuse, comme nous l'avons illustré par divers exemples.

En vue de remédier à cette situation, nous pensons devoir formuler les recommandations suivantes :

- les ouvrages généraux d'expérimentation, même si leur objectif est d'être entièrement consacrés aux problèmes de choix des traitements, devraient toujours faire état de l'importance des autres points à prendre en considération dans l'élaboration des protocoles expérimentaux, afin de ne pas laisser les lecteurs dans l'ignorance des différents aspects des choses ;
- les protocoles expérimentaux devraient toujours envisager les différents points que nous avons évoqués au paragraphe 3¹³ ;
- dans la conception des plans d'expériences, une place suffisante devrait toujours être consacrée aux questions de répétition, de randomisation, de constitution de blocs, etc. ;
- si aucune randomisation suffisante ne s'avère possible en vue de garantir la fiabilité des résultats, ou si certaines restrictions sont apportées à la randomisation, une justification devrait en être donnée et l'attention des lecteurs des conclusions de l'expérience devrait toujours être attirée sur ce point et sur les conséquences qui peuvent en résulter ;
- une attention suffisante devrait toujours être consacrée à l'estimation de la variabilité résiduelle ou de ses différentes composantes ;
- dans cette optique, un minimum de 10 degrés de liberté (ou éventuellement 5 dans les situations les plus contraignantes) devrait toujours être affecté autant que possible à l'estimation de cette ou de ces différentes sources de variation¹⁴.

13. Peut-être est-il opportun de rappeler aussi qu'un protocole expérimental devrait toujours être un document écrit (ou enregistré sur ordinateur), et non pas seulement quelques idées auxquelles on a songé un jour ou l'autre.

14. On notera que la norme française AFNOR relative aux cartes de contrôle de la qualité préconise, en matière d'estimation de l'écart-type, de disposer d'au moins 20 échantillons d'au moins 5 observations chacun, soit au moins 80 degrés de liberté [X, 1995].

En formulant ces recommandations, nous souhaitons simplement contribuer à établir (ou à rétablir) un équilibre raisonnable entre les différentes composantes du plan d'expérience, en particulier entre la place accordée aux facteurs étudiés et la place accordée aux variations résiduelles, entre le signal et le bruit de fond. Nous ne faisons aussi que rappeler les trois principes de R.A. Fisher : *replication*, *randomization*, *blocking*.

RÉFÉRENCES BIBLIOGRAPHIQUES

- ALEXIS J. [1995]. *Pratique industrielle de la méthode Taguchi : les plans d'expériences*. Paris, Association française de Normalisation, 171 p.
- ALEXIS J., ALEXIS P. [2000]. *Pratique industrielle des plans d'expériences – la qualité à moindre coût : l'approche Taguchi*. Paris, Association française de Normalisation, 276 p.
- ATKINS J., CHENG C.S. [1999]. Optimal regression designs in the presence of random block effects. *J. Statist. Plann. Infer.* **77**, 321-335.
- BAILEY R.A. [1985]. Restricted randomization versus blocking. *Int. Statist. Rev.* **53** (2), 171-182.
- BERNOIST D. [1974]. *Notions sur les plans d'expériences*. Paris, Technip, 135 p.
- BERNOIST D., TOURBIER Y., GERMAIN-TOURBIER S. [1995]. *Plans d'expériences : construction et analyse*. Paris, Tec et Doc, 700 p.
- BERGONZINI J.C. [1995]. *Analyse et planification des expériences : les dispositifs en blocs*. Paris, Masson, 353 p.
- BOX G.E.P. [1952]. Multi-factor designs of first order. *Biometrika* **39** (1), 49-57.
- BOX G.E.P., WILSON K.B. [1951]. On the experimental attainment of optimum conditions (with discussion). *J. R. Statist. Soc., Ser. B*, **13** (1), 1-45.
- BOX J.F. [1978]. *R.A. Fisher, the life of a scientist*. New York, Wiley, 512 p.
- BULGREN W.G. [1974]. Probability integral of the doubly noncentral *t*-distribution with degrees of freedom *n* and non-centrality parameters δ and λ . In : Harter H.L., Owen D.B. (ed.). *Selected tables in mathematical statistics (vol. 2)*. Providence, American Mathematical Society, 1-138.
- CHAPOUILLE P. [1973]. *Planification et analyse des expériences*. Paris, Masson, 190 p.
- CHENG C.S., STEINBERG D.M. [1991]. Trend robust two-level factorial designs. *Biometrika* **78** (2), 325-336.
- COCHRAN W.G. [1939]. Long-term agricultural experiments (with discussion). *J. R. Statist. Soc., Suppl.* **6** (2), 104-148.
- COCHRAN W.G. [1976]. Early development of techniques in comparative experimentation. In : Owen D.B. (ed.). *On the history of statistics and probability*. New York, Dekker, 3-25.
- COCHRAN W.G., AUTREY K.M., CANNON C.Y. [1941]. A double changeover design for dairy cattle feeding experiments. *J. Dairy Sci.* **24**, 937-951.
- COLLOMBIER D. [1996]. *Plans d'expérience factoriels : construction et propriétés des fractions de plans*. Berlin, Springer, 194 p.
- COOK R., NACHTSHEIM C. [1989]. Computer-aided blocking of factorial and response-surface designs. *Technometrics* **31** (3), 339-346.
- COSTER D.C. [1993]. Tables of minimum cost, linear trend-free run sequences for two- and three-level fractional factorial designs. *Comput. Statist. Data Anal.* **16**, 325-336.

LA PLANIFICATION DES EXPÉRIENCES

- DAGNELIE P. [1981]. *Principes d'expérimentation*. Gembloux, Presses agronomiques, 182 p.
- DAGNELIE P. [1997]. La planification des expériences et l'analyse de la variance : une introduction. In : Driesbeke J.J., Fine J., Saporta G. (éd.). *Plans d'expériences : applications à l'entreprise*. Paris, Technip, 13-67.
- DAGNELIE P. [1998]. *Statistique théorique et appliquée. Tome 2 : Inférence statistique à une et à deux dimensions*. Paris et Bruxelles, De Boeck et Larcier, 659 p.
- DEMONSANT J. [1996]. *Comprendre et mener des plans d'expériences*. Paris, Association française de Normalisation, 176 p.
- DICKINSON A.W. [1974]. Some run orders requiring a minimum number of factor level changes for the 2^4 and 2^5 main effect plans. *Technometrics* **16** (1), 31-37.
- DOUAIRE G. [1985]. *Les dispositifs expérimentaux en blocs incomplets*. Rennes, Ecole nationale supérieure agronomique, 174 p.
- DRIESBEKE J.J., FINE J., SAPORTA G. (éd.). [1997a]. *Plans d'expériences : applications à l'entreprise*. Paris, Technip, 509 p.
- DRIESBEKE J.J., FINE J., SAPORTA G. [1997b]. Le cheminement historique des plans d'expériences. In : Driesbeke J.J., Fine J., Saporta G. (éd.). *Plans d'expériences : applications à l'entreprise*. Paris, Technip, 1-12.
- DUGUÉ D., GIRAULT M. [1959]. *Analyse de variance et plans d'expérience*. Paris, Dunod, 68 p.
- EDERER F. [1998]. History of clinical trials. In : Armitage P., Colton T. (ed.). *Encyclopedia of biostatistics (vol. 3)*. Chichester, Wiley, 1936-1945.
- FEDERER W.T., BALAM L.N. [1972]. *Bibliography on experiment and treatment design pre-1968*. Edinburgh, Oliver and Boyd, 769 p.
- FINNEY D.J. [1945]. The fractional replication of factorial experiments. *Ann. Eugenics* **12** (4), 291-301.
- FINNEY D.J. [1946]. Recent developments in the design of field experiments. III. Fractional replication. *J. Agric. Sci.* **36** (3), 184-191.
- FISHER R.A. [1925]. *Statistical methods for research workers*. Edinburgh, Oliver and Boyd, 239 p.
- FISHER R.A. [1926]. The arrangement of field experiments. *J. Ministry Agric.* **33**, 503-513.
- GOOS P., VANDEBROEK M. [1999]. *Optimal response surface designs in the presence of random block effects* (Department of Applied Economics Research Report 9945). Leuven, Katholieke Universiteit, 20 p.
- GOUPY J. [1988]. *La méthode des plans d'expériences : optimisation du choix des essais et de l'interprétation des résultats*. Paris, Dunod, 303 p.
- GOUPY J. [1989]. Erreur de dérive et choix de l'ordre des essais d'un plan d'expériences factoriel. *Rev. Statist. Appl.* **37** (1), 5-22.
- GOUPY J. [1990]. Etude comparative de divers plans d'expériences. *Rev. Statist. Appl.* **38** (4), 5-44.
- GOUPY J. [1999]. *Plans d'expériences pour surfaces de réponses*. Paris, Dunod, 409 p.
- HAHN G.J. [1982]. Design of experiments : an annotated bibliography. In : Kotz S., Johnson N.L. (ed.). *Encyclopedia of statistical sciences (vol. 2)*. New York, Wiley, 359-366.
- KIEFER J. [1959]. Optimum experimental designs (with discussion). *J. R. Statist. Soc., Ser. B*, **21** (2), 272-319.

LA PLANIFICATION DES EXPÉRIENCES

- LECLERCQ R. [1960]. *Histoire et avenir de la méthode expérimentale*. Paris, Masson, 138 p.
- LELLOUCH J., LAZAR P. [1974]. *Méthodes statistiques en expérimentation biologique*. Paris, Flammarion, 283 p.
- LISON L. [1958]. *Statistique appliquée à la biologie expérimentale : la planification de l'expérience et l'analyse des résultats*. Paris, Gauthier-Villars, 346 p.
- LOCHNER R.H., Matar J.E. [1992]. *Conception de la qualité : les plans d'expériences*. Paris, Association française de Normalisation, 266 p.
- LOUVET F. [1996]. *Introduction aux plans d'expériences : description d'une approche méthodologique pour un mariage stratégique mais néanmoins heureux des mathématiques, des statistiques et ... des expériences*. Limoges, Ecole nationale supérieure de Céramique industrielle, 61 p.
- LOUVET F. [2000]. Plans d'expériences. In : Barlier C. et al. *Conception en mécanique industrielle : calculs, agencement, prototypage*. Paris, Dunod, 56 p.
- MATTHEWS J.R. [1998]. History of biostatistics. In : Armitage P., Colton T. (ed.). *Encyclopedia of biostatistics (vol. 3)*. Chichester, Wiley, 1931-1936.
- MOZZO G. [1995]. *Initiation aux plans d'expériences*. Paris, Elf Atochem, 47 p.
- PHILIPPEAU G. [1973]. *Théorie des plans d'expérience : application à l'agronomie*. Paris, Institut technique des Céréales et des Fourrages, 205 p.
- PILLET M. [1992]. *Introduction aux plans d'expériences par la méthode Taguchi*. Paris, Editions d'Organisation, 224 p.
- PILLET M. [1997]. *Les plans d'expériences par la méthode Taguchi*. Paris, Editions d'Organisation, 330 p.
- PLACKETT R.L., BURMAN J.P. [1946]. The design of optimum multifactorial experiments. *Biometrika* **33** (4), 305-325.
- POSTEN H.O. [1974]. An abbreviated list of book reviews 1960-1973. *Amer. Statist.* **28** (4), 138-142.
- PREECE D.A. [1990]. R.A. Fisher and experimental design : a review. *Biometrics* **46** (4), 925-935.
- SADO G., SADO M.C. [1991]. *Les plans d'expériences : de l'expérimentation à l'assurance qualité*. Paris, Association française de Normalisation, 266 p.
- SCHEFFÉ H. [1958]. Experiments with mixtures. *J. R. Statist. Soc., Ser. B*, **20** (2), 344-360.
- SCHIMMERLING P., SISSON J.C., ZAÏDI A. [1998]. *Pratique des plans d'expériences*. Paris, Tec et Doc, 526 p.
- SCHWARTZ D., FLAMANT R., LELLOUCH J. [1970]. *L'essai thérapeutique chez l'homme*. Paris, Flammarion, 297 p.
- TAGUCHI G. [1959]. Linear graphs for orthogonal arrays and their application to experimental design, with aid of various techniques. *Rep. Statist. Appl. Res.* **6** (1), 1-43.
- TAGUCHI G. [1960]. Table of orthogonal arrays and linear graphs. *Rep. Statist. Appl. Res.* **6** (5), 176-227.
- TAGUCHI G. [1987]. *System of experimental design : engineering methods to optimize quality and minimize costs* (2 vol.). White Plains, Unipub/Kraus, 552 + 656 p.
- VESSEREAU A. [1948]. *Recherche et expérimentation en agriculture. Tome 2 : Méthodes statistiques en biologie et en agronomie*. Paris, Baillière, 381 p.
- VIGIER M. [1988]. *Pratique des plans d'expériences : méthodologie Taguchi et compléments*. Paris, Editions d'Organisation, 191 p.
- WILLIAMS E.J. [1959]. *Regression analysis*. New York, Wiley, 214 p.

LA PLANIFICATION DES EXPÉRIENCES

- YATES F. [1935]. Complex experiments (with discussion). *J. R. Statist. Soc.*, Ser. B, **2** (2), 181-247.
- YATES F. [1936]. Incomplete randomized blocks. *Ann. Eugenics* **7**, 121-140.
- YATES F. [1937]. *The design and analysis of factorial experiments*. Harpenden, Imperial Bureau of Soil Science, 95 p.
- YATES F. [1964]. Sir Ronald Fisher and the design of experiments. *Biometrics* **20** (2), 307-321.
- X [1985]. *Statistique. Vocabulaire et symboles. Partie 3 : Plans d'expérience* (norme ISO 3534/3-1985). Genève, Organisation internationale de Normalisation, 32 p.
- X [1989]. *Application de la statistique. Plan d'expériences. Vocabulaire et indications générales* (norme AFNOR X 06-080). Paris, Association française de Normalisation, 25 p.
- X [1995]. *Application de la statistique. Cartes de contrôles. Partie 1 : Cartes de contrôle de Shewart aux mesures* (norme AFNOR X 06-031-1). Paris, Association française de Normalisation, 39 p.
- X [1998]. *Statistique. Vocabulaire et symboles. Partie 3 : Plans d'expérience* (norme ISO 3534/3-1998). Genève, Organisation internationale de Normalisation, 34 p.
- X [1999]. ICH harmonised tripartite guideline : statistical principles for clinical trials. *Statist. Med.* **1** (15), 1905-1942.