

Le plan d'expérience évolue...

Pierre Dagnelie

Faculté universitaire des Sciences agronomiques

B-5030 Gembloux (Belgique)

<mailto:pierre@dagnelie.be>

Résumé

Cet article, de type *tutorial* ou *review*, esquisse l'évolution du concept de plan d'expérience, essentiellement dans le domaine agronomique et dans le domaine industriel, en partant des principes développés par Ronald Aylmer Fisher au cours des années 1920.

Dans un premier temps, nous donnons quelques indications relatives aux années antérieures à 1920 (paragraphe 2.1), nous rappelons ce qu'a été l'apport de Fisher (paragraphe 2.2), et nous présentons les orientations qui ont fait suite à ses travaux (paragraphe 2.3 à 2.6). Nous voyons alors la manière dont les principes de Fisher ont été relativement négligés dans le domaine industriel, en matière de répétition (paragraphe 3.1), ainsi qu'en matière de randomisation et de *blocking* (paragraphe 3.2). Après quoi, nous mettons en évidence quelques pistes qui permettent de remédier dans une certaine mesure à la situation observée, à savoir l'utilisation du principe du *split-plot* (paragraphe 4.1 et 4.2), de nouvelles formes de *blocking* (paragraphe 4.3), et le recours à des dispositifs expérimentaux insensibles à certaines dérives (paragraphe 4.4). Nous terminons par quelques conclusions et recommandations (paragraphe 5), et par une assez importante bibliographie.

Diverses informations complémentaires sont également données en annexe.

Summary

This paper, of tutorial or review type, outlines the evolution of the concept of experimental design, essentially in agricultural research and industry, starting from the principles developed by Ronald Aylmer Fisher during the 1920s.

We begin by giving some information on the years former to 1920 (section 2.1), pointing out Fisher's contribution (section 2.2), and presenting the orientations which followed his work (sections 2.3 to 2.6). We then mention how Fisher's principles were relatively neglected in the field of industry, as regards replication (section 3.1), as well as randomization and blocking (section 3.2). After that we highlight a few tracks which allow to a certain extent to remedy the situation observed, namely the use of the *split-plot* principle (sections 4.1 and 4.2), some new forms of blocking (section 4.3), and resorting to experimental designs insensitive to certain drifts or trends (section 4.4). Some conclusions and recommendations (section 5), and a rather important bibliography end this paper.

Various additional information is also given as an appendix.

Mots-clés – Keywords

Plan d'expérience. Experimental design.

1. Introduction

Les principes modernes d'expérimentation sont nés au cours des années 1920, essentiellement dans la recherche agronomique, au départ des travaux de Ronald Aylmer Fisher. Dans cet article, nous nous efforçons de montrer comment ces principes ont évolué depuis lors, en fonction notamment des développements de la recherche industrielle.

Avant d'entrer dans le vif du sujet, il peut être utile de rappeler que les éléments de base de tout plan d'expérience ou protocole expérimental sont :

- un ou différents objectifs bien définis ;
- un ensemble de conditions particulières dans lesquelles l'expérience doit être réalisée ;
- un certain nombre d'unités expérimentales (par exemple un certain nombre de parcelles, dans un champ, ou de patients, dans le secteur médical, ou plusieurs utilisations successives d'un même fermenteur, dans le domaine industriel, etc.) ;
- un certain nombre de traitements, susceptibles d'être appliqués aux unités expérimentales ;
- une ou plusieurs variables ou caractéristiques des unités expérimentales, souvent appelées variables d'intérêt, au sujet desquelles on souhaite étudier l'influence des traitements.

Dans le cas le plus simple, les "traitements" sont au nombre de deux, à savoir un témoin (aucun traitement particulier n'étant appliqué) et un traitement proprement dit. Mais les traitements peuvent aussi correspondre à deux ou plusieurs niveaux ou modalités d'un même facteur (par exemple deux ou plusieurs doses d'un même engrais), ou à des combinaisons de niveaux ou de modalités de deux ou plusieurs facteurs (par exemple différentes pressions associées à différentes températures).

2. Perspective historique

2.1. Avant 1920

Les concepts d'expérience, d'expérimentation, etc. sont fort anciens, même si, pendant longtemps, ils n'ont pas été l'objet d'un formalisme rigoureux [Cochran, 1976 ; Droesbeke *et al.*, 1997 ; Leclercq, 1960 ; Ullrich, 2002].

En ce qui concerne le 18^{ème} siècle, on cite fréquemment, à titre d'exemples, les expériences de James Lind (1716-1794) et de François Cretté de Palluel (1741-1798). La première a trait à l'influence de la consommation d'oranges et de citrons sur le scorbut dont sont atteints un ensemble de marins. Les unités expérimentales sont les marins soumis à l'expérience, les traitements sont les différents compléments alimentaires qui leur sont attribués (dont des oranges et des citrons), et la caractéristique à laquelle on s'intéresse est l'évolution de la maladie [Lind, 1753, 1756]. La deuxième expérience citée comme exemple est relative à la comparaison de différentes alimentations données à des moutons. Il est intéressant de noter que cette expérience fait intervenir, sans en donner le nom, un dispositif en carré latin* (EXP 8.1.1) [Cretté de Palluel 1788, 1790].

* Les concepts marqués d'un astérisque sont explicités en annexe, le plus souvent par des exemples. En outre, les mentions du type "EXP 8.1.1" renvoient à l'un ou l'autre paragraphe (ici le paragraphe 8.1.1) du livre *Principes d'expérimentation : planification des expériences et analyse de leurs résultats* [Dagnelie, 2003], qui est disponible à la fois sous forme imprimée et sur internet (<www.dagnelie.be/exaces1.html>).

Ensuite, dans la lignée notamment des travaux du chimiste Justus von Liebig (1803-1873), le 19^{ème} siècle est marqué par la mise en place, dans plusieurs pays, d'expériences en champ destinées à comparer différentes fumures, certaines de ces expériences existant toujours à l'heure actuelle. Les plus remarquables sont sans doute celles de la *Rothamsted Experimental Station* (Harpenden, Grande-Bretagne), et en particulier l'expérience de Broadbalk, installée en 1843 [Anon., 2007].

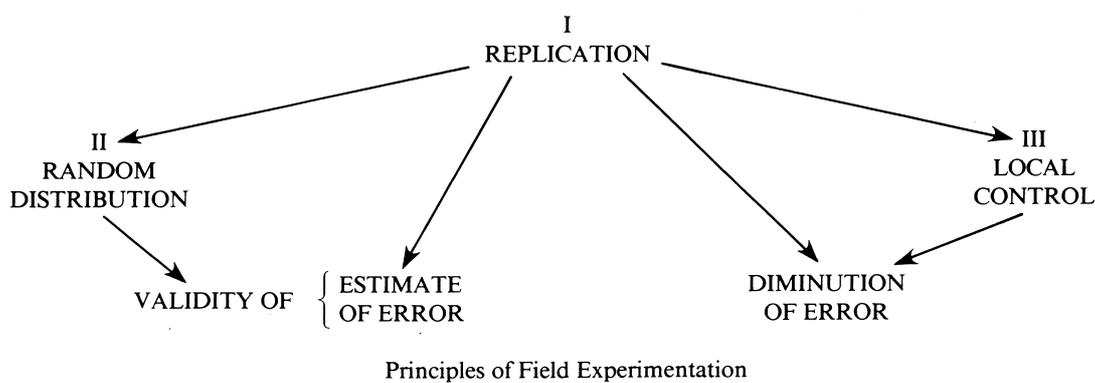
Toujours en ce qui concerne la même période, on peut mentionner également, entre autres, les travaux de Claude Bernard (1813-1878), en médecine, et de Johann Gregor Mendel (1822-1884), en génétique [Bernard, 1865 ; Mendel, 1866].

Au début du 20^{ème} siècle, l'interprétation des résultats des expériences en champ mises en place antérieurement retient l'attention des chercheurs, et est l'objet de diverses publications, telles que celles de Mercer et Hall [1911] et de Wood et Stratton [1910].

2.2. Les années 1920 : Ronald Aylmer Fisher

En 1919, Ronald Aylmer Fisher (1890-1962), diplômé de l'Université de Cambridge, en mathématiques et en physique, est engagé à la *Rothamsted Experimental Station*, en vue précisément de procéder à l'analyse de l'ensemble des données provenant des expériences du milieu du 19^{ème} siècle. Très rapidement, à partir de 1921, Fisher publie une série de notes intitulées notamment *Studies in crop variation* [Fisher, 1921 ; Fisher et Mackenzie, 1923 ; Fisher, 1926 ; etc.], ainsi que son livre mondialement connu *Statistical methods for research workers* [Fisher, 1925].

On peut considérer que ces travaux sont à la base de tout ce qui a été réalisé ultérieurement en matière d'expérimentation [Box, 1978, 1980 ; Preece, 1990 ; Yates, 1964]. Fisher y préconise le respect de trois principes, à savoir : *replication*, *randomization* et *local control*. Ces principes, qui n'étaient pas entièrement neufs à ce moment, sont bien mis en évidence dans le schéma suivant, publié en 1931 [Fisher, 1931 ; Preece, 1990] :



La répétition (*Replication*) de chacun des traitements considérés, un certain nombre de fois dans une même expérience, a pour objectif de permettre une estimation de la variabilité résiduelle (*Estimate of error*), c'est-à-dire de la variabilité qui n'est pas liée aux traitements étudiés, et simultanément, d'augmenter la précision de l'expérience (*Diminution of error*).

La randomisation (*Random distribution* ou *randomization*), c'est-à-dire la répartition "au hasard" des différents traitements au sein des différentes parcelles d'une expérience en champ ou, d'une manière plus générale, au sein des différentes unités expérimentales, permet d'obtenir des estimations non biaisées de la variabilité résiduelle et de l'influence des traitements (*Validity of estimate*).

Enfin, le contrôle local (*Local control*, aussi qualifié ailleurs de *blocking*) a pour but, comme la répétition, d'augmenter la précision de l'expérience. Dans l'esprit de Fisher, ce "contrôle local"

peut être réalisé par l'utilisation de divers dispositifs expérimentaux tels que des blocs* (EXP 6.1), c'est-à-dire des ensembles de parcelles ou d'unités expérimentales voisines ou aussi semblables que possible, des carrés latins* (EXP 8.1.1), etc.

Parmi les trois éléments envisagés, le principal apport de Fisher est sans doute la randomisation et, corrélativement, l'obtention d'estimations correctes.

Dans les publications de Fisher que nous avons citées, apparaissent en outre les concepts d'expérience factorielle* (EXP 2.3.2), c'est-à-dire d'expérience faisant intervenir deux ou plusieurs facteurs, chacune des modalités de chacun des facteurs étant associée à chacune des modalités du ou des autres facteurs, et d'analyse de la variance*, qui occupe une place centrale dans l'interprétation des résultats d'expériences.

Indépendamment de nombreuses autres contributions, souvent plus théoriques, on peut encore ajouter au crédit de Fisher, le *confounding** (EXP 10.1), c'est-à-dire, dans les expériences factorielles, la confusion éventuelle des interactions et des effets de deux ou plusieurs facteurs, et l'analyse de la covariance.

On notera aussi que, telles qu'elles sont présentées dans les premiers travaux de Fisher, l'analyse de la variance et l'analyse de la covariance ne font pas intervenir les distributions F de Snedecor ou de Fisher-Snedecor, qui ne seront définies qu'ultérieurement*.

2.3. Après 1930

Après son départ de *Rothamsted*, en 1933, Fisher publie son livre *The design of experiments*, qui est en quelque sorte une synthèse des études qu'il a réalisées au cours de ses 14 années consacrées à la recherche agronomique [Fisher, 1935].

Pour la suite, on peut mentionner quelques publications particulièrement importantes, qui constituent le point de départ des principaux développements nouveaux :

- les travaux de Yates, qui fut le collaborateur puis le premier successeur de Fisher à Rothamsted, relatifs aux expériences factorielles* (EXP 2.3.2) et aux expériences en blocs incomplets* (EXP 9.1) [Yates, 1935, 1936, 1937] ;
- les travaux de Cochran, relatifs notamment aux expériences en *cross-over* (ou *change-over*)* (EXP 8.1.2) [Cochran, 1939 ; Cochran *et al.*, 1941] ;
- le travail de Finney [1945], relatif aux expériences factorielles fractionnaires (ou incomplètes)* (EXP 2.3.3) ;
- le travail de Box et Wilson [1951], relatif au choix des traitements dans le but de déterminer au mieux des surfaces de réponse* (EXP 2.4.1), c'est-à-dire des relations exprimant la ou les variables étudiées en fonction des différents facteurs pris en considération ;
- le travail de Scheffé [1958], qui considère la question des surfaces de réponse dans le cas particulier des mélanges* de différentes substances (EXP 2.4.2) ;
- et le travail de Kiefer [1959], qui aborde, sous la dénomination de plans optimaux (EXP 2.4.3), les problèmes envisagés par Box et Wilson et par Scheffé, dans des situations plus complexes, liées souvent à l'existence de contraintes particulières relatives aux différents facteurs.

Les années ultérieures sont alors très largement des années d'approfondissement de ces diverses orientations, principalement en fonction des possibilités nouvelles offertes par le développement rapide des moyens de traitement de l'information, et cela tant en matière de planification des expériences que d'analyse des résultats obtenus. Une caractéristique essentielle de ces années ultérieures est la recherche de plans relatifs à l'étude d'un nombre important de facteurs,

à partir d'un petit nombre d'unités expérimentales, mais conservant diverses propriétés particulières, d'optimalité, d'orthogonalité, etc.

Les plans saturés (EXP 2.3.3.8°), qui permettent d'étudier quasi autant de facteurs qu'il y a d'unités expérimentales, dont déjà les plans de Plackett et Burman [1946], et les plans supersaturés, qui sont destinés à étudier plus de facteurs qu'il n'y a d'unités expérimentales, constituent le stade ultime de cette orientation.

Il est intéressant de mettre en évidence l'évolution générale qui apparaît ainsi en ce qui concerne l'intérêt porté, d'une part, au choix des traitements (*treatment design*) et, d'autre part, aux modalités d'affectation des traitements aux unités expérimentales, c'est-à-dire au dispositif expérimental proprement dit (*experimental design*), et aussi l'évolution relative aux méthodes d'analyse des résultats.

Fisher, puis Cochran, étudient en priorité la manière d'affecter les traitements aux unités expérimentales (randomisation, expériences en blocs, en carré latin, en *cross-over*, etc.), et en deuxième lieu seulement le choix des traitements (expériences factorielles). Yates et Finney inversent les rôles : expériences factorielles et factorielles fractionnaires tout d'abord, expériences en blocs incomplets ensuite. Mais, comme Fisher et Cochran, Yates et Finney envisagent principalement l'interprétation des résultats par l'analyse de la variance, en termes d'effets principaux et d'interactions des facteurs, avec subsidiairement des notions de régression (utilisation de polynômes orthogonaux en particulier).

Au contraire, Box et Wilson, de même que Scheffé, Kiefer et de nombreux autres auteurs qui ont abordé ultérieurement les mêmes problèmes, se consacrent exclusivement ou quasi exclusivement au choix des traitements, souvent en négligeant dans une large mesure les questions de randomisation, etc. En outre, ils travaillent dans une optique essentiellement de régression, la détermination de surfaces de réponse se substituant très largement à l'étude des effets des facteurs et de leurs interactions.

2.4. Quelques commentaires

L'évolution ainsi observée est étroitement liée à une évolution parallèle des domaines d'application : l'expérimentation du début du 20^{ème} siècle était essentiellement agronomique, tandis que l'expérimentation de la deuxième moitié du 20^{ème} siècle devient largement industrielle.

Un des éléments qui justifie le glissement progressif des expériences factorielles vers l'étude des surfaces de réponse est le fait que les expériences agronomiques concernent fréquemment des facteurs qualitatifs, éventuellement associés à des facteurs quantitatifs, alors que les expériences industrielles ont essentiellement traité à des facteurs quantitatifs.

L'évolution que nous avons esquissée se double en outre d'une évolution sémantique. Traditionnellement, dans le domaine agronomique, le mot "expérience" correspond à l'ensemble des opérations qui permettent d'étudier deux ou plusieurs traitements appliqués à un certain nombre d'unités expérimentales. Dans le domaine industriel par contre, le mot "expérience" désigne le plus souvent chacune des manipulations qui est relative à une seule application d'un traitement donné à une unité expérimentale.

Ainsi, dans le travail réalisé par Cretté de Palluel, dont il est question au paragraphe 2.1, l'ensemble serait considéré comme une seule expérience (ou un seul essai) dans l'optique agronomique, mais au contraire comme un groupe de 16 expériences (ou 16 essais) dans l'optique industrielle, chaque attribution d'une alimentation à un mouton étant considérée comme une expérience (ou un essai).

Cette nuance provient notamment du fait que, dans le premier cas, et en particulier pour les expériences en champ, toutes les opérations (installation des parcelles, application des traitements et collecte des informations) sont généralement réalisées simultanément, pour l'ensemble des unités expérimentales, alors que dans le deuxième cas, les diverses opérations sont réalisées le plus souvent en séquence, de façon consécutive.

Dans cette optique aussi, l'expression "plan d'expérience" (au singulier) cède souvent la place à "plan d'expériences" (au pluriel).

2.5. Les livres et les revues

L'évolution évoquée au cours des paragraphes précédents apparaît clairement dans les livres et les revues.

Jusqu'au milieu des années 1980, les livres relatifs à l'expérimentation consacraient une place importante, voire parfois quasi exclusive, à l'aspect "dispositif expérimental" (*experimental design*), tandis que, depuis lors, et surtout en ce qui concerne la littérature de langue française, la priorité, voire l'exclusivité, est généralement accordée à l'aspect "choix des traitements" (*treatment design*) [Dagnelie, 2000]. Cette assertion peut être illustrée, par exemple, en observant que le très classique *Experimental designs* de Cochran et Cox [1950] est relatif quasi exclusivement aux dispositifs expérimentaux et à l'analyse des résultats obtenus, tandis que les ouvrages récents, tels que ceux de Goupy [2005], et Goupy et Creighton [2006], sont relatifs à 90% ou plus au choix des traitements.

En ce qui concerne les revues, on peut noter, par exemple, que les titres *Biometrika* et *Journal of Agricultural Science*, qui publient dès le début du 20^{ème} siècle de nombreux articles traitant de l'expérimentation, ont été lancés respectivement en 1900 et 1905, alors que *Technometrics* et *Journal of Quality Technology* datent respectivement de 1959 et 1969.

2.6. La recherche médicale

Nous n'avons considéré jusqu'ici que l'expérimentation agronomique et l'expérimentation industrielle. Or, l'expérimentation médicale (ou pharmaceutique) s'est aussi développée, plus récemment, et a pris un essor considérable au cours des dernières décennies [Armitage, 1995 ; Bloom, 1986 ; Ederer, 1998]. En particulier, des revues nouvelles sont nées, telles que *Statistics in Medicine* et *Statistical Methods in Medical Research*, respectivement en 1982 et 1992.

La recherche médicale occupe en fait une position intermédiaire entre les deux situations extrêmes que nous avons évoquées. Tantôt, elle concerne un grand nombre de produits ou de molécules susceptibles d'avoir des effets thérapeutiques, et elle se rapproche alors de la recherche industrielle. Tantôt, et notamment dans les essais cliniques, elle porte sur un petit nombre de "traitements", voire fréquemment sur un seul traitement et un témoin (placebo), et elle se rapproche dans ces conditions de la recherche agronomique.

Aux problèmes évoqués précédemment, s'ajoutent évidemment ici des questions éthiques particulièrement importantes, et divers problèmes liés à la durée des expériences, à la disparition et au retrait de certains patients en cours d'expérience, et en conséquence, à l'existence de grands nombres de données manquantes, etc.

On peut souligner aussi le fait que les distinctions entre recherche agronomique, recherche industrielle et recherche médicale peuvent se réduire considérablement quand il s'agit de problèmes mitoyens tels que ceux du secteur alimentaire et des biotechnologies.

Enfin, d'autres domaines d'utilisation des plans d'expériences pourraient également être envisagés (sciences de l'éducation, marketing, etc.).

3. En oubliant Fisher...

3.1. L'absence de répétition

Une conséquence de la prédominance progressive de l'expérimentation industrielle par rapport à l'expérimentation agronomique est le fait que les principes développés par Fisher ont été souvent négligés. La question de savoir quel peut être l'impact d'une telle situation mérite d'être prise sérieusement en considération.

En ce qui concerne les expériences factorielles, le nombre de traitements, qui résulte de la combinaison des différentes modalités des facteurs, augmente très rapidement en fonction du nombre de facteurs. Très rapidement aussi, il n'est donc plus possible de prévoir plusieurs répétitions de l'ensemble des traitements. On réalise alors des expériences en "répétition unique" (EXP 2.3.2.8°) ou en "répétition partielle" (EXP 2.3.3), les expériences factorielles complètes cédant le pas aux expériences factorielles fractionnaires, par application du principe du *confounding* (souvent sous forme de demi-répétitions, de quarts de répétition, etc.).

Des méthodes particulières doivent être utilisées dans ces conditions pour procéder à l'analyse des résultats. Deux solutions sont fréquemment adoptées. D'une part, on peut considérer une ou plusieurs interactions d'ordre supérieur comme des substituts de la variation résiduelle, et effectuer les tests d'hypothèses et les éventuelles opérations complémentaires (estimations de moyennes, de différences de moyennes, etc.) sur cette base (EXP 5.2.3°). D'autre part, les facteurs et les interactions qui jouent un rôle prépondérant, souvent qualifiés de "facteurs actifs", peuvent être identifiés dans une certaine mesure par la réalisation de diagrammes de probabilité (EXP 5.2.4°).

Quant aux expériences non factorielles relatives à l'étude des surfaces de réponse, des mélanges, etc., elles sont très souvent organisées sans répétition ou avec un très petit nombre de répétitions d'un ou de quelques traitements.

Dans les différents cas, l'absence de répétitions ou l'existence de répétitions partielles et peu nombreuses a pour résultat qu'on ne dispose que de très petits nombres de degrés de liberté, dans l'estimation de la variabilité résiduelle. Cette situation peut ne pas présenter d'inconvénients majeurs quand la variabilité est très réduite, ce qui est assez fréquemment le cas dans le domaine industriel. Par contre, cette situation n'est pas sans inconvénients quand la variabilité résiduelle est importante, ce qui est fréquent lorsque des phénomènes biologiques sont en jeu.

Dans l'optique de l'étude des facteurs et de leurs interactions par l'analyse de la variance, on peut essayer de chiffrer à titre indicatif l'ampleur du problème, en considérant les précisions obtenues dans les estimations de variances, d'écart-types et de moyennes, ou de différences de moyennes, ou encore de fonctions linéaires de moyennes (contrastes).

Un principe assez fréquemment admis est de prévoir autant que possible un minimum de 10 degrés de liberté, en vue de l'estimation d'une variance résiduelle. Un simple examen des distributions khi-carré de Pearson et des distributions *t* de Student permet de voir comment évoluent les longueurs des intervalles de confiance, par rapport à cette situation de référence.

Le tableau suivant donne quelques indications à ce sujet, les longueurs des intervalles de confiance y étant exprimées par comparaison avec le cas de 10 degrés de liberté :

D.l.	Var.	Ec.t.	Moy.
2	23,1	4,8	1,93
3	6,9	2,6	1,43
4	3,6	1,9	1,25
5	2,4	1,6	1,15

Ce tableau montre que, pour la variance résiduelle, la longueur de l'intervalle de confiance est multipliée par 23 quand le nombre de degrés de liberté est limité à 2, au lieu de 10, par 7 quand le nombre de degrés de liberté est égal à 3, au lieu de 10, etc. En termes d'écart-types résiduels, ces rapports sont égaux à 4,8, 2,6, etc. Et en ce qui concerne les moyennes, les différences de moyennes et les fonctions linéaires de moyennes, ces rapports sont plus réduits, mais restent non négligeables, les longueurs des intervalles de confiance étant augmentées de 15 à 93%, dans les différents cas considérés, et toujours par comparaison avec la même situation de référence (10 degrés de liberté).

La même question peut être posée en ce qui concerne les comparaisons, et non plus les estimations de moyennes, mais cette question est plus difficile à résoudre, en raison de la diversité des situations (différents niveaux de signification des tests, différents degrés de fausseté des hypothèses nulles, etc.). On peut toutefois montrer que, dans bien des cas, et toujours par comparaison avec 10 degrés de liberté, le risque de deuxième espèce, c'est-à-dire la probabilité de ne pas mettre en évidence une différence de moyennes pourtant bien réelle, est multiplié par un facteur de l'ordre de 2 à 5, voire plus.

Des difficultés semblables se présentent dans l'optique de l'étude des surfaces de réponse, en vue de la recherche de conditions optimales (maximum ou minimum). Ces difficultés peuvent être considérées à deux niveaux différents. D'une part, un maximum (ou un minimum) peut ne pas être identifié à partir des données observées, alors qu'il existe cependant. D'autre part, un éventuel maximum (ou minimum) peut être identifié, mais sans que sa localisation ne soit très précise.

Cette dernière situation est relativement fréquente quand l'effet des facteurs n'est pas suffisamment important, par rapport à la variabilité résiduelle. Elle apparaît en fait lors de la recherche de limites ou de régions de confiance, car celles-ci ne sont pas définies. Il est difficile de donner des indications générales assez précises à ce sujet, en raison, ici également, de la diversité des situations possibles, mais on peut néanmoins constater, dans des cas très simples, que l'effet des facteurs doit souvent être du même ordre de grandeur que l'écart-type résiduel, ou parfois sensiblement plus élevé, pour pouvoir mettre valablement en évidence des maximums ou des minimums [Dagnelie, 2000].

3.2. L'absence de randomisation et de *blocking*

En ce qui concerne les expériences réalisées en séquence, ce qui est souvent le cas dans le domaine industriel, il peut être tentant de travailler de façon systématique, les modalités des différents facteurs étant modifiées progressivement, à tour de rôle. Cette façon de procéder permet en effet de réduire sensiblement le nombre de changements des modalités des facteurs, auxquels correspondent en général des réglages plus ou moins délicats.

En outre, il est fréquent que les expériences soient réalisées, en répétition unique ou partielle, sans aucun type de blocs.

À titre d'exemple, dans une expérience factorielle qui ferait intervenir quatre facteurs, chacun à deux modalités (expérience de type 2^4), et sans répétition, l'ordre systématique suivant :

1111 1112 1121 1122 1211 1212 1221 1222 2111 2112 2121 2122 ... 2222 ,

permettrait de ne modifier qu'une seule fois le niveau du premier facteur, auquel correspond le premier chiffre, trois fois le niveau du deuxième facteur (deuxième chiffre), sept fois le niveau du troisième facteur (troisième chiffre), et quinze fois le niveau du quatrième facteur (quatrième chiffre). L'organisation de l'expérience en respectant un tel ordre peut conduire à des gains parfois importants, si on considère comme premier facteur celui dont les modifications des modalités sont les plus coûteuses (en temps ou en moyens de quelque nature), ..., et comme dernier facteur celui dont les modifications sont les moins coûteuses.

Les résultats des expériences organisées sans randomisation ni *blocking* sont néanmoins analysés, le plus souvent, comme si une randomisation tout à fait normale avait été effectuée, ce qui n'est pas sans conséquences. Comme pour l'absence de répétition, qui a été abordée au cours du paragraphe 3.1, ces conséquences sont relativement mineures quand la variabilité résiduelle est faible, mais peuvent au contraire être importantes quand la variabilité résiduelle est élevée.

En vue d'illustrer cette situation, considérons le cas réel d'une expérience factorielle relative à quatre facteurs, les deux premiers présentant chacun deux modalités, le troisième trois modalités, et le quatrième quatre modalités, les 48 combinaisons de ces différentes modalités étant étudiées en séquence, avec une seule répétition [Dagnelie, 2006 ; Lacrosse, 1990].

Les valeurs observées au cours de cette expérience sont des rendements exprimés en pourcentage, la moyenne de ces résultats étant égale à 67,9 %, et les valeurs extrêmes étant 55,1 et 80,5 %, soit une amplitude de 25,4 %. La partie gauche du tableau suivant donne les résultats de l'analyse de la variance (valeurs *F* de Fisher-Snedecor et probabilités correspondantes) relative à cette expérience, ce tableau étant limité aux quatre facteurs principaux, et les différents tests de signification ayant été réalisés par rapport à un ensemble d'interactions :

Fact.	Sans dérive		Avec dérive	
	<i>F</i>	<i>P</i>	<i>F</i>	<i>P</i>
1	2,39	0,134	0,61–5,34	0,029–0,443
2	7,97	0,009	4,23–12,9	0,001–0,050
3	16,4	0,000	12,0–21,5	0,000–0,000
4	52,1	0,000	45,6–59,1	0,000–0,000

Nous supposons dans la suite qu'une modification progressive des résultats est intervenue, dans le temps, de manière linéaire, en raison de modifications des conditions expérimentales (modification des qualités de la matière première ou du réglage de certains instruments de mesure par exemple). Et nous supposons que cette dérive est égale à 1/1000 de l'amplitude des résultats, soit 0,0254 %, d'un essai à l'autre.

En considérant que les traitements ont été appliqués dans l'ordre systématique suivant :

1111 1112 1113 1114 1121 1122 1123 1124 1131 1132 1133 1134 1211 ... 2234 ,

une telle dérive linéaire aurait pour conséquence que la première observation, égale à 74,4, ne serait pas modifiée, que la deuxième observation, égale à 71,3, deviendrait 71,3254, que la troisième observation, égale à 69,6, deviendrait 69,6508, etc.

Dans les conditions envisagées, on peut constater que les valeurs *F* de Fisher-Snedecor seraient alors :

5,34 , 10,3 , 15,2 et 51,5 ,

au lieu de :

2,39 , 7,97 , 16,4 et 52,1 ,

et les probabilités correspondantes seraient évidemment modifiées en conséquence.

Mais les 48 traitements, qui concernent quatre facteurs, pourraient être étudiés de façon systématique de 24 manières différentes ($4! = 24$), et la dérive pourrait être soit positive soit négative, ce qui conduit à 48 situations et 48 analyses de la variance différentes. Le tableau présenté ci-dessus donne, dans sa partie droite, les valeurs extrêmes observées à l'issue de ces 48 analyses de la variance, tant pour les valeurs *F* que pour les probabilités qui leur sont associées, et toujours pour les seuls quatre facteurs principaux.

En ce qui concerne le premier facteur par exemple, les résultats qui correspondent à la valeur initiale 2,39 vont, pour les 48 analyses de la variance, de 0,61 à 5,34, soit un rapport de 1 à 9, l'effet de ce facteur pouvant tout aussi bien être considéré comme significatif (probabilité égale à 0,029), que comme totalement non significatif (probabilité égale à 0,443). De même, pour le deuxième facteur, les valeurs F vont de 4,23 à 12,9, et les probabilités de 0,001 à 0,050.

L'introduction, à titre d'illustration, de diverses dérives a évidemment aussi une incidence sur les valeurs des moyennes. Ainsi, la différence de moyennes existant entre les deux modalités du premier facteur, qui est initialement égale à 1,23 %, devient 0,62 et 1,83 %, soit un rapport de 1 à 3, pour les deux cas extrêmes. Et de même, en ce qui concerne le deuxième facteur, la différence initiale est égale à 2,25 %, et les valeurs extrêmes sont 1,64 et 2,86 %.

Cet exemple permet de constater qu'une éventuelle dérive, même très limitée, dont il ne serait pas tenu compte, peut avoir des conséquences très importantes quant à l'interprétation des résultats. Ces conséquences sont particulièrement marquées pour les facteurs dont les effets sont peu prononcés et qui sont à la limite des niveaux de signification habituels, c'est-à-dire précisément dans les cas les plus sensibles.

Comme nous l'avons signalé, nous n'avons pas envisagé la question des interactions. Cette attitude est justifiée par le fait que les interactions ne sont aucunement influencées par d'éventuelles dérives linéaires, mais elles pourraient l'être dans le cas de dérives non linéaires.

4. Quelques solutions

4.1. Le *split-plot* : quelques exemples

Face aux problèmes, particulièrement aigus dans certains cas, que peut soulever l'absence de répétition, de randomisation et de *blocking*, différentes pistes, qu'il n'est pas toujours facile de distinguer les unes des autres, ont été suivies. La principale d'entre elles est sans doute le *split-plot*.

À condition d'y introduire une certaine randomisation, les exemples du paragraphe 3.2 font en effet penser aux expériences en parcelles divisées ou en *split-plot** (EXP 7.1), d'un usage relativement courant dans le domaine agronomique.

Dans le cas d'une expérience organisée en séquence, qui ferait intervenir trois facteurs, l'un à deux modalités et les deux autres à trois modalités (expérience 2×3^2), on pourrait adopter par exemple l'ordre suivant :

231 233 232 211 213 212 221 223 222 123 121 122 131 132 133 113 111 112 .

Cet ordre a été obtenu en procédant à une première randomisation pour les deux modalités du premier facteur, puis à une deuxième randomisation pour les trois modalités du deuxième facteur, séparément pour chacune des deux modalités du premier facteur, et enfin à une troisième randomisation pour les trois modalités du troisième facteur, séparément pour chacune des six combinaisons des modalités des deux premiers facteurs. Cette organisation correspond au principe du *split-split-plot*, ou *split-plot* à trois niveaux au lieu de deux (EXP 7.1.1.2°).

Une autre solution pourrait être :

231 232 211 221 222 213 233 212 223 132 113 112 123 122 121 131 111 133 ,

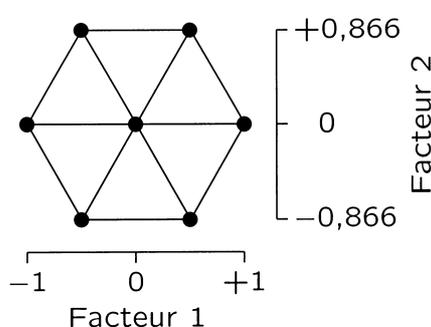
avec, comme ci-dessus, une première randomisation pour le premier facteur, puis globalement une deuxième randomisation pour les neuf combinaisons des modalités des deux autres facteurs. Il s'agirait alors d'un dispositif en *split-plot* simple.

Dans un cas comme dans l'autre, l'objectif de réduction du nombre de changements des modalités, qui est recherché dans l'organisation d'expériences systématiques, est atteint. Comme le montrent les deux exemples qui viennent d'être présentés, sauf pour le premier facteur, le nombre de changements des modalités peut même être inférieur à ce qu'il est dans le cas systématique.

Différentes distinctions apparaissent toutefois entre les situations que nous examinons ici et les expériences agronomiques classiques. D'une part, nous envisageons ici le cas de séries d'essais réalisés en séquence, alors que le *split-plot* classique concerne essentiellement des expériences en champ réalisées en une fois. D'autre part, nous avons fait allusion ci-dessus à des expériences organisées en une seule répétition, et sans aucun type de blocs, alors que, classiquement, le *split-plot* concerne des expériences en blocs aléatoires complets, comportant au moins deux répétitions.

En outre, dans le domaine industriel, le principe du *split-plot* a été étendu au cas des expériences factorielles fractionnaires et à certaines études de surfaces de réponse, les différentes randomisations successives ne portant plus, chaque fois, sur toutes les modalités du ou des facteurs concernés.

Par exemple, dans la recherche d'une surface de réponse relative à deux facteurs, à l'aide d'un plan de Doehlert (EXP 2.4.1.6°) se présentant comme suit :



on peut concevoir une organisation de type *split-plot* en prévoyant une première randomisation pour les différents niveaux du deuxième facteur (trois niveaux), puis une deuxième randomisation pour les différents niveaux du premier facteur (deux ou trois niveaux, selon les cas). On peut aussi mieux équilibrer le dispositif en prévoyant en outre, pour le niveau inférieur et pour le niveau supérieur du deuxième facteur, deux répétitions d'un des deux points expérimentaux, ce qui conduit à réaliser en séquence trois groupes de trois essais.

4.2. Le *split-plot* : quelques commentaires et compléments

Un point essentiel, qu'il y a lieu de ne jamais négliger, en ce qui concerne les expériences en *split-plot*, est le fait que les différents facteurs ne bénéficient généralement pas d'une même précision, les différences de précision pouvant d'ailleurs être considérables.

Ainsi, dans le premier exemple du paragraphe précédent, le premier facteur, dont les deux modalités sont, en termes agronomiques, affectées aux "grandes parcelles", est étudié de façon très grossière. Le deuxième facteur, dont les différentes modalités sont affectées à des "sous-parcelles", au sein des "grandes parcelles", bénéficie d'une plus grande précision, les "grandes parcelles" correspondant pour ce facteur à deux blocs aléatoires complets. Et le troisième facteur jouit d'une plus grande précision encore, dans la mesure où il est en fait étudié au sein de six blocs aléatoires complets, correspondant aux six "sous-parcelles" relatives aux deux premiers facteurs.

Dans le deuxième exemple, la situation du premier facteur est identique à celle du premier exemple, mais les deux autres facteurs se trouvent sur pied d'égalité, leurs neuf combinaisons de modalités étant réparties au sein des deux "grandes parcelles".

Ces différences de précision peuvent être à l'origine de dilemmes difficiles à résoudre. Au paragraphe 3.2, nous avons envisagé la détermination de l'ordre des essais dans une optique de réduction des coûts, la prise en considération des différents facteurs intervenant dans l'ordre décroissant des coûts de modification de leurs modalités, le premier facteur étant celui dont les modifications des modalités sont les plus coûteuses, et le dernier facteur celui dont les modifications des modalités sont les moins coûteuses.

Le principe du *split-plot* suggère au contraire d'étudier les facteurs dans l'ordre croissant de l'intérêt qu'on leur porte, un facteur qui serait sans grande importance ou dont on saurait a priori que l'effet est considérable pouvant, sans inconvénient, être pris en considération en premier lieu, tandis que le facteur auquel on attacherait le plus d'importance devrait être considéré en dernier lieu. Mais évidemment, l'ordre décroissant des coûts des modifications des modalités et l'ordre croissant de l'intérêt porté aux facteurs ne sont pas nécessairement concordants, ce qui peut soulever de délicats problèmes de décision.

L'application du principe des "parcelles divisées" dans le domaine industriel s'est surtout développée au cours des années 1990, comme l'illustre par exemple la publication de Box et Jones [1992]. De nombreuses possibilités nouvelles ont ensuite été envisagées, depuis le début du 21ème siècle, parfois sous d'autres appellations, telles que bi-randomisation et randomisation multistrat.

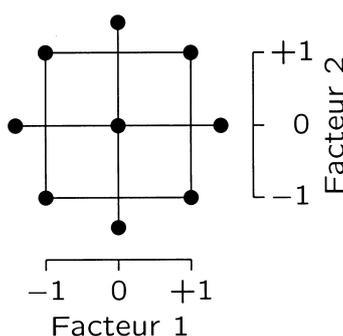
Parmi les publications du tout début du siècle, on peut citer celles de Bingham et Sitter [2001], Goos et Vandebroek [2001a], et Trinca et Gilmour [2001], et parmi les publications récentes, qui sont nombreuses, celles de Goos et Donev [2007a], Jones et Goos [2007], Kowalski *et al.* [2007], Kulahci [2007], Naes *et al.* [2007], Parker *et al.* [2007a, 2007b, 2007c], Smith *et al.* [2007], et Yang *et al.* [2007], ainsi que le livre de Federer et King [2007].

Quant aux principales orientations qui ont été envisagées, on peut mentionner la recherche de plans optimaux ou sous-optimaux en tenant compte à la fois des coûts relatifs des modifications des facteurs et des précisions relatives dont jouissent les différents facteurs, la recherche de plans optimaux ou sous-optimaux permettant de faire face à des contraintes relatives aux nombres de "grandes parcelles" et de "sous-parcelles", l'utilisation de dispositifs analogues au *split-plot*, tel que le dispositif en *split-block* ou *strip-plot* (EXP 7.1.3), et l'interprétation des résultats d'expériences par des méthodes plus élaborées que l'analyse de la variance classique (modèle linéaire général, etc.).

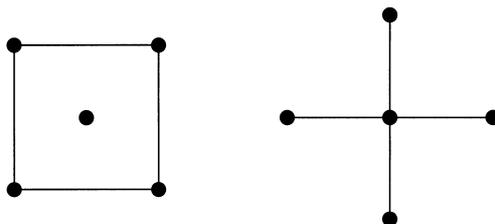
4.3. L'utilisation de blocs

L'utilisation de blocs, complets ou incomplets, est tout à fait classique en ce qui concerne les expériences factorielles et factorielles fractionnaires. Une telle utilisation a également été introduite dans l'étude des surfaces de réponse, mais semble avoir été souvent négligée, comme en témoignent nombre d'ouvrages consacrés à ce sujet.

À titre d'exemple, on peut considérer le cas du dispositif composite centré à deux facteurs* (EXP 2.4.1.3°), qui se présente de la manière suivante :



Ce dispositif peut être divisé comme suit en deux blocs de cinq unités expérimentales :



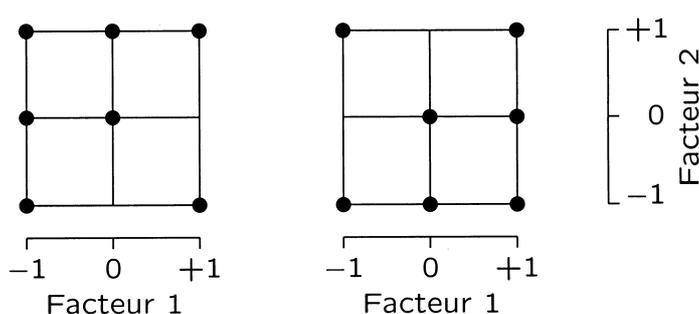
le point central étant alors l'objet de deux observations.

Une solution plus judicieuse pourrait être de répéter le point central deux fois dans chacun des deux blocs, le nombre d'unités expérimentales par bloc étant égal à six, et les quatre répétitions du point central donnant naissance à trois degrés de liberté, dont un degré de liberté lié à l'éventuelle différence entre blocs et deux degrés de liberté indépendants des blocs.

Un des éléments qui a longtemps limité l'utilisation de blocs, en particulier dans le cas des surfaces de réponse, est le fait que la définition des blocs est relativement délicate à réaliser quand le nombre de blocs et/ou le nombre d'unités expérimentales par bloc sont imposés, et surtout quand ces nombres sont relativement réduits.

Le développement des moyens modernes de traitement de l'information a cependant permis de lever dans une large mesure cette difficulté, par la recherche dans chaque cas de solutions optimales ou sous-optimales adaptées à la situation considérée, en donnant ainsi une nouvelle impulsion à l'utilisation des blocs. Une telle recherche de solutions optimales peut être réalisée notamment par des algorithmes d'échanges, en partant d'une solution initiale plus ou moins arbitraire. L'utilisation de blocs intervient alors souvent en relation avec l'application du principe du *split-plot*.

La solution suivante peut ainsi être obtenue dans le cas d'une expérience factorielle faisant intervenir deux facteurs à trois modalités (expérience 3^2) et devant être organisée en deux blocs de sept unités expérimentales [Cook et Nachtshiem, 1989] :



On remarquera que cinq points expérimentaux, régulièrement répartis, sont répétés chacun deux fois, ce qui permet de disposer d'un degré de liberté lié à l'éventuelle différence entre blocs et quatre degrés de liberté indépendants des blocs.

Comme au paragraphe précédent, nous donnons quelques références de publications datant des environs de l'année 1990 [Atkinson et Donev, 1989 ; Cook et Nachtshiem, 1989], quelques références intermédiaires [Goos et Vandebroek, 2001b ; Trinca et Gilmour, 2000], et diverses références récentes [Butler, 2006 ; Goos, 2006 ; Goos et Donev, 2006a, 2006b, 2007b ; Khuri, 2006 ; McLeod et Brewster, 2006], en y ajoutant le livre de Goos [2002].

4.4. Les dispositifs insensibles aux dérives

Des dispositifs particuliers ont également été proposés en vue de faire face à l'existence de dérives, notamment linéaires, dans les expériences réalisées en séquence.

Nous avons vu au paragraphe 3.2 que, dans une expérience factorielle organisée de manière systématique, une dérive linéaire peut avoir une influence importante sur l'estimation des effets des facteurs, mais n'a pas d'influence sur l'estimation des interactions. Tenant compte du fait que, fréquemment, les expérimentateurs souhaitent en réalité étudier les facteurs eux-mêmes, plutôt que leurs interactions, une solution peut être de reporter l'influence des dérives sur les interactions.

Ainsi, dans le cas d'une expérience qui fait intervenir trois facteurs présentant chacun deux modalités (expérience 2^3), on peut montrer que la séquence :

122 221 212 111 211 112 121 222 ,

est telle que les effets des trois facteurs, ainsi que l'interaction du deuxième et du troisième facteur ne sont pas influencés par une éventuelle dérive linéaire, alors que les trois autres interactions le sont, dans une mesure plus ou moins importante.

La recherche de tels dispositifs insensibles (ou partiellement insensibles) aux dérives peut être associée à la recherche d'un coût minimal de l'ensemble des modifications des facteurs et, aussi, à l'utilisation de blocs, les diverses approches pouvant être complémentaires.

L'exemple suivant :

212 122 222 111 221 121 211 111 222 112
221 111 211 122 212 112 222 122 211 121,

est également relatif à une expérience à trois facteurs présentant chacun deux modalités, mais qui serait organisée en quatre blocs de cinq unités expérimentales.

L'existence de 20 unités expérimentales pour les huit combinaisons des modalités des trois facteurs implique évidemment que le nombre de répétitions n'est pas constant, certains éléments (112, 121, 212 et 221) étant en fait répétés deux fois, et d'autres (111, 122, 211 et 222) trois fois. Mais la caractéristique remarquable de ce dispositif est que les effets des trois facteurs et les différentes interactions, de deux ou des trois facteurs, ne sont pas influencés par d'éventuelles dérives linéaires, tandis que les effets des trois facteurs et l'interaction des trois facteurs ne sont pas affectés non plus par d'éventuelles dérives quadratiques [Lin et Dean, 1991].

D'une manière générale, on peut faire remarquer en outre qu'en présence d'une dérive non linéaire mais monotone (dérive toujours positive ou toujours négative), l'utilisation d'un dispositif insensible aux seules dérives linéaires peut déjà rendre de grands services.

Enfin, il faut signaler que le problème considéré ici est parfois abordé sous l'angle des corrélations entre résultats successifs, notamment par l'étude de processus d'autocorrélation. Une dérive, linéaire par exemple, induit en effet automatiquement une certaine corrélation entre les observations successives.

La littérature relative aux dispositifs *trend-free* est moins abondante que celle qui concerne le *split-plot* et l'utilisation de blocs, mais nous pouvons néanmoins citer, comme précédemment, quelques références des environs de l'année 1990 [Goupy, 1989 ; Lin et Dean, 1991], des environs de l'année 2000 [Rohan et Jones, 2000 ; Tack et Vandebroek 2001], et des dernières années [Adekeye et Kunert, 2006 ; Carrano *et al.*, 2006 ; de León Adams *et al.*, 2005].

5. Conclusions et recommandations

La manière dont les principes développés par Fisher au cours des années 1920 (paragraphe 2.2) sont appliqués a sensiblement évolué au cours du temps. En particulier, la transition de l'expérimentation agronomique à l'expérimentation industrielle a conduit à négliger dans une certaine mesure – consciemment parfois, mais pas toujours – les principes de répétition, de randomisation et de *blocking* (paragraphe 2.3 à 2.5).

Un désintérêt pour ces principes est sans conséquence dommageable quand l'expérience peut être organisée dans des conditions parfaitement contrôlées, sans autre variabilité que celle liée aux facteurs étudiés. Par contre, les conséquences peuvent être importantes quand les conditions de l'expérience sont susceptibles d'évoluer, dans l'espace ou dans le temps, ou quand le matériel expérimental présente lui-même une certaine variabilité (paragraphe 3.1 et 3.2). Et notre pratique de la consultation statistique nous a montré que nombre d'expérimentateurs sous-estiment, parfois considérablement, la variabilité des conditions dans lesquelles ils travaillent.

Une telle sous-estimation provient notamment du fait que la variabilité qualifiée de "résiduelle" englobe généralement différents types de variation, tels que les variations liées à l'échantillonnage réalisé lors de la sélection du matériel expérimental, les effets des facteurs non contrôlés, les erreurs de mesure, et les erreurs dues à l'inadéquation du modèle utilisé (régression linéaire ou quadratique par exemple). Et assez fréquemment, seules les erreurs de mesure sont prises en considération.

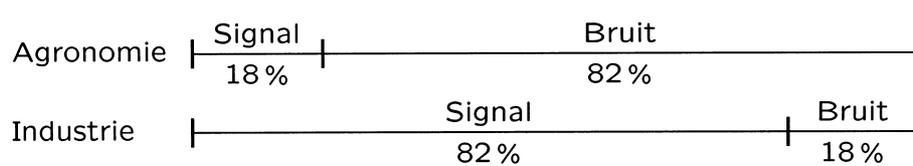
Nous pensons que les principes de répétition (ou de répétition partielle, en vue d'obtenir une estimation suffisamment précise de la variabilité des résultats), de randomisation (ou de randomisation restreinte, comme le permet notamment le *split-plot*) et de *blocking* ne devraient jamais être perdus de vue ou passés sous silence, ni dans les ouvrages généraux relatifs à l'expérimentation, ni dans les protocoles expérimentaux.

Nous préconisons en fait un juste équilibre entre les aspects "choix des traitements" (*treatment design*) et "choix du dispositif expérimental" (*experimental design*), ce point de vue pouvant être illustré concrètement de la manière suivante.

Une expérience de type agronomique, tout à fait classique, pourrait faire intervenir deux facteurs comportant chacun trois modalités, soit neuf traitements dans le cas d'une expérience factorielle 3^2 , et pourrait être organisée en cinq blocs complets (quatre à six répétitions étant de pratique courante), soit un ensemble de 45 unités expérimentales. Dans l'optique habituelle de l'analyse de la variance, une telle expérience consacrerait huit degrés de liberté aux deux facteurs auxquels on s'intéresse, quatre degrés de liberté au facteur blocs et 32 degrés de liberté à la "variation résiduelle", soit 18 % du nombre total de degrés de liberté au "signal" auquel on s'intéresse, et 82 % du nombre total de degrés de liberté au "bruit de fond".

Une expérience de type industriel, tout aussi classique, pourrait être constituée, pour trois facteurs, de neuf points expérimentaux, trois d'entre eux étant répétés chacun deux fois, soit un total de 12 unités expérimentales (ou 12 "essais" ou 12 "expériences", selon la terminologie habituelle du domaine considéré). Il pourrait s'agir par exemple d'une expérience factorielle fractionnaire, constituée d'un tiers de répétition d'un plan factoriel 3^3 , dont les trois sommets seraient répétés chacun deux fois (EXP 2.3.3.6°). Dans l'optique d'une régression qui ferait intervenir neuf variables, neuf degrés de liberté, soit 82 %, seraient consacrés au "signal", et deux degrés de liberté, soit 18 %, seraient consacrés au "bruit".

Schématiquement, la situation peut être résumée de la manière suivante :



Notre sentiment est que, dans le domaine agricole, l'attention portée au "signal" mériterait souvent d'être augmentée, le nombre de degrés de liberté consacrés au "bruit" étant souvent excessif, tandis que, dans le domaine industriel, c'est l'attention portée au "bruit" qui mériterait sans doute souvent d'être augmentée, ou en tout cas de ne pas être trop réduite (voire réduite à néant).

Dans la même optique, les schémas d'expériences qui ont été conçus dans des domaines caractérisés par une faible variabilité ne devraient jamais être transposés inconsidérément dans des domaines à plus forte variabilité, et leur utilisation ne devrait pas être encouragée d'une manière générale, sans restrictions.

En résumé, la "lutte contre la variabilité" – ou contre les effets néfastes de la variabilité –, par la répétition, la randomisation et le *blocking*, est toujours d'actualité.

Références

Les différents sites web qui sont mentionnés ont été consultés en dernier lieu le 20.10.2007.

Anon. [2007]. *Rothamsted's classical experiments*.

<www.rothamsted.ac.uk/resources/ClassicalExperiments.html>.

Adekeye K.S., Kunert J. [2006]. On the comparison of run orders of unreplicated 2^{k-p} designs in the presence of a time trend. *Metrika* **63** (3), 257-269.

Armitage P. [1995]. Before and after Bradford Hill : some trends in medical statistics. *J. R. Stat. Soc., Ser. A*, **158** (1), 143-153.

Atkinson A.C., Donev A.N. [1989]. The construction of exact *D*-optimum experimental designs with application to blocking response surface designs. *Biometrika* **76** (3), 515-526.

Bernard C. [1865]. *Introduction à l'étude de la médecine expérimentale*. Paris, Baillière, 400 p. ;
<gallica.bnf.fr/ark:/12148/bpt6k3812d>,
<web2.bium.univ-paris5.fr/livanc/?cote=31054&do=livre> et
<classiques.uqac.ca/classiques/bernard_claude/intro_etude_medecine_exp/intro_etude.html>.

Bingham D.R., Sitter R.R. [2001]. Design issues in fractional factorial split-plot experiments. *J. Qual. Technol.* **33** (1), 2-15.

Bloom B.S. [1986]. Controlled studies in measuring the effectiveness of medical care : a historical perspective. *Int. J. Technol. Assessm. Health Care* **2** (2), 299-310.

Box G.E.P., Jones S. [1992]. Split-plot designs for robust product experimentation. *J. Appl. Stat.* **19** (1), 3-26.

Box G.E.P., Wilson K.B. [1951]. On the experimental attainment of optimum conditions (with discussion). *J. R. Stat. Soc., Ser. B*, **13** (1), 1-45.

Box J.F. [1978]. *R.A. Fisher : the life of a scientist*. New York, Wiley, 512 p.

Box J.F. [1980]. R.A. Fisher and the design of experiments, 1922-1926. *Amer. Stat.* **34** (1), 1-7.

Butler N.A. [2006]. Optimal blocking of two-level factorial designs. *Biometrika* **93** (2), 289-302.

- Carrano A.L., Thorn B.K., Lopez G. [2006]. An integer programming approach to the construction of trend-free experimental plans on split-plot designs. *J. Manuf. Syst.* **25** (1), 39-44.
- Cochran W.G. [1939]. Long-term agricultural experiments (with discussion). *J. R. Stat. 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 change-over design for dairy cattle feeding experiments. *J. Dairy Sci.* **24**, 937-951.
- Cochran W.G., Cox G.M. [1950]. *Experimental designs*. New York, Wiley, 454 p.
- Cook R., Nachtsheim C. [1989]. Computer-aided blocking of factorial and response-surface designs. *Technometrics* **31** (3), 339-346.
- Cretté de Palluel F. [1788]. Mémoire sur les avantages et l'économie que procurent les racines employées à l'engrais des moutons à l'étable. *Mémoires d'Agriculture*, Trimestre d'été, 17-23.
- Cretté de Palluel F. [1790]. On the advantage and economy of feeding sheep in the house with roots. *Annals of Agriculture* **14**, 133-139.
- Dagnelie P. [2000]. La planification des expériences : choix des traitements et dispositif expérimental (avec discussion). *J. Soc. Franç. Stat.* **141** (1-2), 5-69.
- Dagnelie P. [2003]. *Principes d'expérimentation : planification des expériences et analyse de leurs résultats*. Gembloux, Presses agronomiques, et édition électronique, <www.dagnelie.be>, 397 p.
- Dagnelie P. [2006]. *Statistique théorique et appliquée. Tome 2. Inférence statistique à une et à deux dimensions*. Bruxelles, De Boeck et Larcier, 734 p.
- Dagnelie P. [2007]. *Statistique théorique et appliquée. Tome 1. Statistique descriptive et bases de l'inférence statistique*. Bruxelles, De Boeck et Larcier, 511 p.
- de León Adams G., Grima Cintas P., Tort-Martorell Llabrés X. [2005]. Experimentation order in factorial designs with 8 or 16 runs. *J. Appl. Stat.* **32** (3), 297-313.
- Droesbeke J.J., Fine J., Saporta G. [1997]. Le cheminement historique des plans d'expériences. In : Droesbeke J.J., Fine J., Saporta G. (éd.). *Plans d'expériences : applications à l'entreprise*. Paris, Technip, 1-12.
- 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., King F. [2007]. *Variations on split plot and split block experiment designs*. New York, Wiley, 270 p.
- Finney D.J. [1945]. The fractional replication of factorial arrangements. *Ann. Eugenics* **12** (4), 291-301.
- Fisher R.A. [1921]. Studies in crop variation. I. An examination of the yield of dressed grain from Broadbalk. *J. Agric. Sci.* **11**, 107-135 ; <digital.library.adelaide.edu.au/dspace/handle/2440/15170>.
- Fisher R.A. [1925]. *Statistical methods for research workers*. Edinburgh, Oliver and Boyd, 239 p. ; <psychclassics.yorku.ca/Fisher/Methods>.
- Fisher R.A. [1926]. The arrangement of field experiments. *J. Min. Agric.* **33**, 503-513 ; <digital.library.adelaide.edu.au/dspace/handle/2440/15191>.

- Fisher R.A. [1931]. Principles of plot experimentation in relation to the statistical interpretation of the results. In : *The technique of field experiments*. Harpenden, Rothamsted Experimental Station, 11-13 ; <digital.library.adelaide.edu.au/dspace/handle/2440/15210>.
- Fisher R.A. [1935]. *The design of experiments*. Edinburgh, Oliver and Boyd, 252 p.
- Fisher R.A., Mackenzie W.A. [1923]. Studies in crop variation. II. The manurial response of different potato varieties. *J. Agric. Sci.* **13**, 311-320 ; <digital.library.adelaide.edu.au/dspace/handle/2440/15179>.
- Goos P. [2002]. *The optimal design of blocked and split-plot experiments*. New York, Springer, 244 p.
- Goos P. [2006]. Optimal versus orthogonal and equivalent-estimation design of blocked and split-plot experiments. *Stat. Neerl.* **60** (3), 361-378.
- Goos P., Donev A.N. [2006a]. Blocking response surface designs. *Comput. Stat. Data Anal.* **51** (2), 1075-1088.
- Goos P., Donev A.N. [2006b]. The *D*-optimal design of blocked experiments with mixture components. *J. Qual. Technol.* **38** (4), 319-332.
- Goos P., Donev A.N. [2007a]. Tailor-made split-plot designs for mixture and process variables. *J. Qual. Technol.* **39** (4), 326-339.
- Goos P., Donev A.N. [2007b]. *D*-optimal minimum support mixture designs in blocks. *Metrika* **65** (1), 53-68.
- Goos P., Vandebroek M. [2001a]. Optimal split-plot designs. *J. Qual. Technol.* **33** (4), 436-450.
- Goos P., Vandebroek M. [2001b]. *D*-optimal response surface designs in the presence of random block effects. *Comput. Stat. Data Anal.* **37** (4), 433-453.
- Goupy J. [1989]. Erreur de dérive et choix de l'ordre des essais d'un plan d'expériences factoriel. *Rev. Stat. Appl.* **37** (1), 5-21.
- Goupy J. [2005]. *Pratiquer les plans d'expériences*. Paris, Dunod, 568 p.
- Goupy J., Creighton L. [2006]. *Introduction aux plans d'expériences*. Paris, Dunod, 336 p.
- Jones B., Goos P. [2007]. A candidate-set-free algorithm for generating *D*-optimal split-plot designs. *Appl. Stat.* **56** (3), 347-364.
- Khuri A.I. [2006]. Mixed response surface models with heterogeneous within-block error variances. *Technometrics* **48** (2), 206-218.
- Kiefer J. [1959]. Optimum experimental designs (with discussion). *J. R. Stat. Soc., Ser. B*, **21** (2), 272-319.
- Kowalski S.M., Parker P.A., Vining G.G. [2007]. Tutorial : industrial split-plot experiments. *Qual. Engin.* **19** (1), 1-15.
- Kulahci M. [2007]. Split-plot experiments with unusual numbers of subplot runs. *Qual. Engin.* **19** (4), 363-371
- Lacrosse L. [1990]. *Contribution à l'étude du comportement des combustibles ligno-cellulosiques en gazogènes à lit fixe et à tirage inversé*. Gembloux, Faculté des Sciences agronomiques, 132 p.
- Leclercq R. [1960]. *Histoire et avenir de la méthode expérimentale*. Paris, Masson, 138 p.
- Lin M., Dean A.M. [1991]. Trend-free block designs for varietal and fractional experiments. *Ann. Stat.* **19** (3), 1582-1596.
- Lind J. [1753]. *A treatise of the scurvy*. Edinburgh, Kincaid and Donaldson ; <www.jameslindlibrary.org/trial_records/17th_18th_Century/lind/lind_tp.html>.

- Lind J. [1756]. *Traité du scorbut (2 vol.)*. Paris, Ganeau.
- McLeod R.G., Brewster J.F. [2006]. Blocked fractional factorial split-plot experiments for robust parameter design. *J. Qual. Technol.* **38** (3), 267-279.
- Mendel G. [1866]. Versuche über Pflanzenhybriden. *Verhandlungen des naturforschenden Vereins in Brünn* **4**, 3-47 ;
 <www.esp.org/foundations/genetics/classical/gm-65-f.pdf>,
 <www.biologie.uni-hamburg.de/b-online/d08_mend/mendel.htm> et
 <www.mendelweb.org/MWpaptoc.html>.
- Mercer W.B., Hall A.D. [1911]. The experimental error of field trials. *J. Agric. Sci.* **4**, 107-132.
- Naes T., Aastveit A.H., Sahni N.S. [2007]. Analysis of split-plot designs : an overview and comparison of methods. *Qual. Reliab. Engin. Int.* **23** (7), 801-820.
- Parker P.A., Anderson-Cook C., Robinson T.J., Liang L. [2007a]. Robust split-plot designs. *Qual. Reliab. Engin. Int.* **24** (1), 107-121.
- Parker P.A., Kowalski S.M., Vining G.G. [2007b]. Construction of balanced equivalent estimation second-order split-plot designs. *Technometrics* **49** (1), 56-65.
- Parker P.A., Kowalski S.M., Vining G.G. [2007c]. Unbalanced and minimal point equivalent estimation second-order split-plot designs. *J. Qual. Technol.* **39** (4), 376-388.
- Plackett R.L., Burman J.P. [1946]. The design of optimum multifactorial experiments. *Biometrika* **33** (4), 305-325.
- Preece D.A. [1990]. R.A. Fisher and experimental design : a review. *Biometrics* **46** (4), 925-935.
- Rohan V.M., Jones G. [2000]. Efficient run orders for a two-factor response surface experiment on a correlated process. *Comm. Stat. Th. Meth.* **29** (3), 593-609.
- Scheffé H. [1958]. Experiments with mixtures. *J. R. Stat. Soc., Ser. B*, **20** (2), 344-360.
- Smith C.D., Johnson D.E. [2007]. Comparing analyses of unbalanced split-plot experiments. *J. Stat. Comput. Simul.* **77** (2), 119-129.
- Tack L., Vandebroek M. [2001]. (D_t , C)-optimal run orders. *J. Stat. Plann. Infer.* **98** (1-2), 293-310.
- Trinca L.A., Gilmour S.G. [2000]. An algorithm for arranging response surface designs in small blocks. *Comput. Stat. Data Anal.* **33** (1), 25-43.
- Trinca L.A., Gilmour S.G. [2001]. Multistratum response surface designs. *Technometrics* **43** (1), 25-33.
- Ullrich P. [2002]. Officers, playing cards, and sheep : on the history of Eulerian squares and of the design of experiments. *Metrika* **56** (3), 189-204.
- Wood T.B., Stratton F.J.M. [1910]. The interpretation of experimental results. *J. Agric. Sci.* **3**, 417-440.
- Yang J., Zhang R., Liu M. [2007]. Construction of fractional factorial split-plot designs with weak minimum aberration. *Stat. Prob. Lett.* **77** (15), 1567-1573.
- Yates F. [1935]. Complex experiments (with discussion). *J. R. Stat. Soc., Suppl.* **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.

Annexe

Cette annexe explicite un certain nombre de notions dont il est question dans le texte.
Les illustrations sont extraites ou inspirées du livre *Principes d'expérimentation : planification des expériences et analyse de leurs résultats* [Dagnelie, 2003].

Quelques dispositifs expérimentaux

Blocs aléatoires complets : exemple d'expérience permettant de comparer sept traitements au sein de quatre blocs complets, la répartition des traitements dans les blocs étant réalisée au hasard et indépendamment d'un bloc à l'autre.

3	4	4	6
6	3	5	3
1	5	2	1
2	1	7	5
7	7	3	2
4	2	1	4
5	6	6	7

Blocs incomplets équilibrés : exemple d'expérience permettant de comparer sept traitements au sein de sept blocs de quatre parcelles ou unités expérimentales seulement, la répartition des traitements dans les blocs étant réalisée au hasard et indépendamment d'un bloc à l'autre, mais de telle sorte que tous les couples de traitements soient présents un même nombre de fois dans les différents blocs (d'où le qualificatif "équilibrés").

4	2	7	3	6	6	1
2	1	6	1	5	2	3
5	7	3	2	4	3	4
7	6	5	5	1	4	7

Parcelles divisées : exemple d'expérience factorielle en blocs aléatoires complets avec parcelles divisées, permettant de comparer six traitements, relatifs à deux facteurs, l'un possédant trois modalités et l'autre deux modalités (traitements 11, 12, 21, 22, 31 et 32), au sein de quatre blocs complets, la répartition au hasard étant réalisée tout d'abord pour le premier facteur, au sein des quatre blocs, en définissant 12 "grandes parcelles", puis pour le deuxième facteur, au sein de ces "grandes parcelles", chacune d'entre elles étant constituée de deux "petites parcelles".

21	11	32	11
22	12	31	12
32	31	12	22
31	32	11	21
11	21	22	32
12	22	21	31

Carré latin : exemple d'expérience permettant de comparer quatre traitements dans un ensemble de 16 parcelles formant quatre lignes et quatre colonnes de quatre parcelles, chacun des traitements étant présent une et une seule fois dans chacune des lignes et dans chacune des colonnes, et la répartition aléatoire étant réalisée par une permutation au hasard des lignes d'une part et des colonnes d'autre part.

3	1	2	4
1	4	3	2
4	2	1	3
2	3	4	1

Les lignes et les colonnes peuvent aussi correspondre à certains facteurs particuliers. Ainsi, dans l'expérience de Cretté de Palluel, dont il est question au paragraphe 2.1, les traitements sont en fait les quatre alimentations étudiées, tandis que les lignes correspondent à quatre périodes successives et les colonnes à quatre races différentes de moutons, l'expérience portant sur un ensemble de 16 moutons.

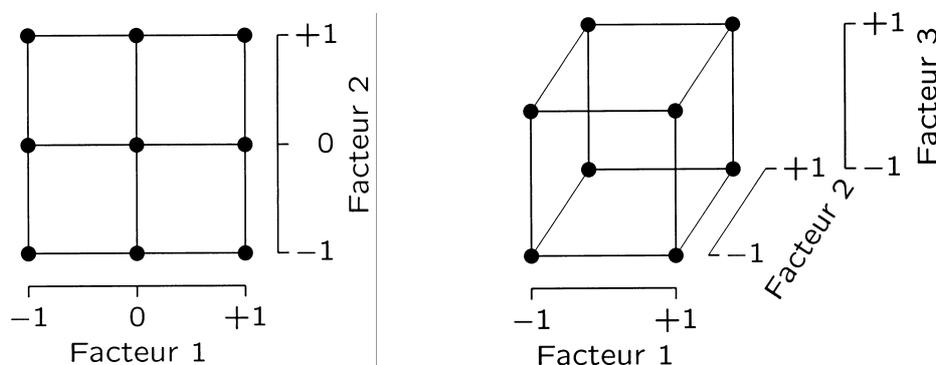
Cross-over : exemple d'expérience permettant de comparer deux traitements dans un dispositif comportant deux lignes et huit colonnes.

1	2	2	2	1	2	1	1
2	1	1	1	2	1	2	2

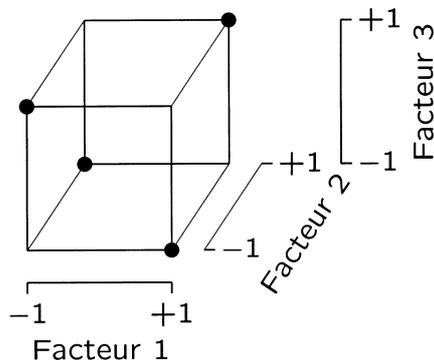
De tels dispositifs sont très fréquemment utilisés dans le domaine médical, où ils permettent de comparer deux ou plusieurs traitements appliqués successivement aux mêmes patients, l'exemple considéré correspondant au cas de deux traitements, deux périodes successives et huit patients.

Expériences factorielles et expériences factorielles fractionnaires

Expérience factorielle : représentation schématique d'une expérience factorielle complète à deux facteurs comportant chacun trois modalités (expérience 3^2), et d'une expérience factorielle complète à trois facteurs comportant chacun deux modalités (expérience 2^3) ; chaque point représente un traitement, les modalités des différents facteurs étant codées en -1 , 0 et $+1$, ou en -1 et $+1$.



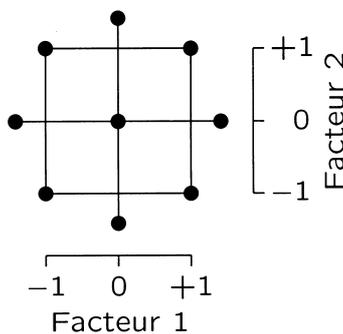
Expérience factorielle fractionnaire et confounding : représentation schématique d'une expérience factorielle fractionnaire (ou incomplète) à trois facteurs comportant chacun deux modalités, seules la moitié des combinaisons des trois facteurs étant prises en considération (demi-expérience 2^3 , aussi appelée expérience 2^{3-1}).



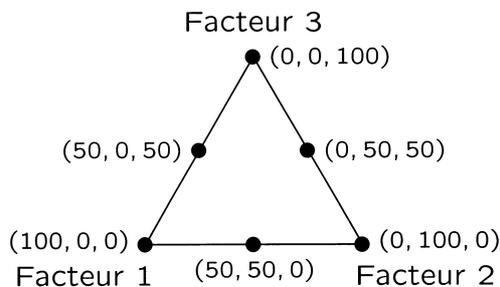
Dans ce cas, toute interaction éventuelle de deux facteurs est confondue (au sens du *confounding*) avec l'effet éventuel du troisième facteur : l'interaction des facteurs 1 et 2 est confondue avec le facteur 3, etc.

Surfaces de réponse et mélanges

Surfaces de réponse : exemple de traitements permettant d'estimer au mieux une surface de réponse, à savoir la relation qui lie la variable étudiée aux différents facteurs, dans le cas de deux facteurs qui ne sont soumis à aucune contrainte.



Mélanges : exemple de traitements permettant d'estimer au mieux une surface de réponse, dans le cas d'un mélange de trois composants, c'est-à-dire de trois facteurs tels que leurs différentes modalités soient toujours de somme égale à 100 (pour une composition du mélange exprimée en pour-cent).



Analyse de la variance

Le texte suivant, extrait de la publication de Fisher et Mackenzie de 1923, illustre les débuts de l'analyse de la variance, qui sont étroitement liés au développement de l'expérimentation agronomique à la *Rothamsted Experimental Station* au cours des années 1920.

In Table III is shown the analysis of the variation into these four classes; the mean square deviation is found by dividing the sum of squares in each class by the number of degrees of freedom, while the standard deviation is shown in the last column. When this value is significantly greater than the standard deviation of the differences between parallel plots, we may conclude that the corresponding effect is not due to chance.

Table III.

Variation due to	Degrees of freedom	Sum of squares	Mean square	Standard deviation
Manuring	5	6,158	1231.6	35.09
Variety	11	2,843	258.5	16.07
Deviations from summation formula	55	981	17.84	4.22
Variation between parallel plots ...	141	1,758	12.47	3.53
Total	212	11,740	—	—

In comparing the standard deviations in the last column we may use the fact that 3.53, for example, has the same accuracy as if it had been determined from a sample of 142; the variance of its natural logarithm is therefore $\frac{1}{2 \times 141}$. Thus, to test if the deviations from the summation formula are significantly greater than would occur by chance, we compare the difference of the logarithms with its standard error, namely $\sqrt{\frac{1}{282} + \frac{1}{110}}$:

$$\begin{array}{rcl} \frac{1}{282} & = & .003546 \\ \frac{1}{110} & = & .009091 \\ \text{Sum} & & .012637 \\ \text{Standard Error} & & .1124 \end{array} \qquad \begin{array}{rcl} \log_e 4.22 & = & 1.4398 \\ \log_e 3.53 & = & 1.2613 \\ \text{Difference} & & .1785 \pm .1124 \end{array}$$

The difference in the logarithm thus exceeds its standard error, but not sufficiently to be significant; while the effects both of manuring and of variety are very strongly marked, and clearly significant.

La comparaison qui est présentée de façon détaillée concerne les lignes *Deviations from summation formula* et *Variation between parallel plots*, qui correspondent à l'interaction et à la variation résiduelle d'une analyse de la variance tout à fait classique à deux critères de classification.

Les carrés moyens (*Mean square*) ne sont pas comparés en calculant leur rapport et en se référant aux distributions *F* de Snedecor ou Fisher-Snedecor, mais bien en calculant leurs racines carrées (*Standard deviation*) et en appliquant la méthode de l'erreur-standard. Fisher et Mackenzie se basent ainsi sur le fait que l'erreur-standard du logarithme de l'écart-type d'un échantillon d'effectif *n* suffisamment élevé est approximativement égale à $1/[2(n-1)]$, dans le cas d'une population possédant une distribution normale, d'où les quotients $1/282$ et $1/110$, respectivement pour 141 et 55 degrés de liberté, et l'erreur-standard 0,1124 pour la différence des logarithmes.

Cette procédure n'est évidemment applicable que pour des nombres de degrés de liberté suffisamment élevés. Dans son livre de 1925, Fisher, introduit en conséquence, pour de petits nombres de degrés de liberté, une variable et une distribution z , qui seront ultérieurement équivalentes à la variable et à la distribution F , la variable z étant égale à $(\log F)/2$.

On notera éventuellement, d'une part, que les 141 degrés de liberté de la variation résiduelle proviennent du fait que trois observations sont considérées pour chacune des 72 combinaisons des 6 fumures (*Manuring*) et des 12 variétés (*Variety*), trois observations étant toutefois manquantes, et d'autre part, que la propriété relative à l'erreur-standard du logarithme de l'écart-type peut être établie facilement sur base des relations qui ont trait aux transformations de variables et à la distribution d'échantillonnage de la variance [Dagnelie, 2007].