NAZIOARTEKO ESTATISTIKA MINTEGIA EUSKADIN **1984** SEMINARIO INTERNACIONAL DE ESTADISTICA EN EUSKADI

EXPERIMENTUEN DISEINUA

DISEÑO DE EXPERIMENTOS

DESING OF EXPERIMENTS

PLANIFICATION ET PLAN D'EXPERIENCES

D.J. FINNEY

EUSKO JAURLARITZA
Estatistika Zuzendaritza

GOBIERNO VASCO
Dirección de Estadística

NAZIOARTEKO ESTATISTIKA MINTEGIA EUSKADIN **1984** SEMINARIO INTERNACIONAL DE ESTADISTICA EN EUSKADI

EXPERIMENTUEN DISEINUA
●
DISEÑO DE EXPERIMENTOS
●
DESING OF EXPERIMENTS
●
PLANIFICATION ET PLAN D'EXPERIENCES

D.J. FINNEY

CUADERNO **5** KOADERNOA

EUSKO JAURLARITZA
Estatistika Zuzendaritza

GOBIERNO VASCO
Dirección de Estadística

Depósito Legal: S.S. 429-85
ISBN: 84-7542-127-10. Obra Completa.
ISBN: 84-7542-192-X. Tomo V.
Impreso en Itxaropena, S.A. - Errikobarra kalea, 2 - Zarautz

AURKEZPENA

Estatistikako Mintegi Internazionalak sustatzean, hainbat xederekin bete nahi luke Eusko Jaurlaritzaren Estatistika Zuzendaritzak, hala nola:

- Unibertsitatearekiko eta, bereziki, Estatistika Sailarekiko lankidetza bultzatu.
- Funtzionari, irakasle, ikasle eta estatistikaren alorrean interesaturik leudekeen guztien birziklapen profesionala erraztu.
- Estatistikako alorrean eta mundu-mailan irakasle prestu eta abangoardiako ikerlari diren pertsonaiak Euskadira ekarri, guzti horrek zuzeneko harremanei eta esperientzien ezagupenei dagokienez soposatzen duen ondorio positiboarekin.

Iharduketa osagarri bezala eta interesaturik leudekeen ahalik eta pertsona eta Erakunde gehienetara iristearren, Ikastaro hauetako txostenak argitaratzea erabaki da, beti ere txostenemailearen jatorrizko hizkuntza errespetatuz, horrela gure Herrian gai honi buruzko ezagutza zabaltzen laguntzeko asmoarekin.

PRESENTACION

Al promover los Seminarios Internacionales de Estadística, la Dirección de Estadística del Gobierno Vasco pretende cubrir varios objetivos:

- Fomentar la colaboración con la Universidad y en especial con los Departamentos de Estadística.
- Facilitar el reciclaje profesional de funcionarios, profesores, alumnos y cuantos puedan estar interesados en el campo estadístico.
- Traer a Euskadi a ilustres profesores e investigadores de vanguardia en materia estadística, a nivel mundial, con el consiguiente efecto positivo en cuanto a relación directa y conocimiento de experiencias.

Como actuación complementaria y para llegar al mayor número posible de personas e Instituciones interesadas, se ha decidido publicar las ponencias de estos Cursos, respetando en todo caso la lengua original del ponente, para contribuir así a acrecentar el conocimiento sobre esta materia en nuestro País.

PRESENTATION

La Direction de Statistique du Gouvernement Basque se propose d'atteindre plusieurs objectifs par la promotion des Séminaires Internationaux de Statistique:

- Encourager la collaboration avec l'université et spécialment avec les départements de statistique.
- Faciliter le recyclage professionnel des fonctionnaires, professeurs, élèves, et tous ceux qui pourraient etre intéressés par la statistique.
- Inviter en Euskadi des professeurs mondialement renommés et des chercheurs de premier ordre en matière de Statistique avec tout ce que cela pourrait entraîner comme avantage dans les rappots et l'échange d'expériences.

En outre, il a été décidé de publier les exposés de ces rencontres afin d'atteindre le plus grand nombre de personnes et d'institutions intéressées, et pour contribuer ainsi à développer dans notre pays les connaissances sur cette matière. Dans chaque cas la langue d'origine du conférencier sera respectée.

PRESENTATION

In promoting the International Seminars on Statistics, the Statistics Office of the Basque Goverment is attempting to achieve a number of objectives:

- Encourage joint working with the Basque University and, in particular, with its Department of Statistics.
- Facilitate the in-training of civil servants, teachers and students and of all those interested in the field of statistics.
- Bring to Euskadi distinguished academics and researchers in the front line of statistics work, at a wolrd-wide level, with all the benefits that this will bring through direct contacts and the interchange of experiences and ideas.

As an additional step this year, it has been decided to publish in advance the papers to be presented at these courses, respecting the native language of the speaker, in each case. This is in order that as many interested people and institutions as possible are made aware. In this way we hope to contribute to the growth and awareness concerning this topic in our country.

Vitoria-Gazteiz, Diciembre 1984 Abendua

JOSE IGNACIO GARCIA RAMOS
Estatistikako Zuzendaria
Director de Estadística

SARRERA

Liburu honek, Eusko Jaurlaritzaren Estatistika-Zuzendaritzak eta Euskal Herriko Unibertsitatearen Matematika Aplikatuko Departamentuak antolaturik, D.J. FINNEY Jaunak Euskadiko Estatistikako II. Nazioarteko Mintegiaren barruan "Design of Experiments" gaiari buruz eman duen ikastaroa laburbiltzen digu. II. Mintegi honek kontatzen du baita Madrileko Unibertsitateko F. AZORIN Jaunaren partaidetzarekin ere, "Aspectos de Teoría y Aplicaciones en el Muestreo" gaiari buruzko ikastaro batekin.

INTRODUCCION

Este libro resume el curso que sobre "Desing of Experiments" ha impartido D.J. FINNEY dentro del II Seminario Internacional de Estadística en Euskadi, organizado por la Dirección de Estadística del Gobierno Vasco y el Departamento de Matemática Aplicada de la Universidad del País Vasco. Este II Seminario cuenta además con la participación de F. AZORIN de la Universidad de Madrid con un curso sobre "Aspectos de Teoría y Aplicaciones en el Muestreo".

INTRODUCTION

Ce livre résume le cours donné par D.J. FINNEY sur "Design of Experiments" dans le cadre du II Séminaire International de Statistique d'Euskadi, organisé par le Conseil de Statistique du Gouvernement Basque et le Département de Mathématique Appliquée de l'Université du Pays Basque. Ce II Séminaire compte aussi avec la participation de F. AZORIN de l'Université de Madrid avec un cours sur "Aspectos de Teoría y Aplicaciones en el Muestreo".

INTRODUCTION

This paper summarises the lecture on "Desing of Experiments" presented by D.F. FINNEY for the II International Statistic Seminar in Euskadi, organised by the Statistics Office of the Basque Government and by the Department of Applied Mathematics at the University of the Basque Country. This II Seminar included also the participation of F. AZORIN of the University of Madrid with a series of lectures on "Aspectos of Theory and Applications in Sample Statistics".

BIOGRAFIA

David J. Finney *Estatistikako Katedraduna dugu. Estatistika Departamentuko Zuzendaria Edinburgoko Unibertsitatean, Britainiar Inperioko Gomendadorea. Britainiar Erret Akademiako kidea. Eskoziako Ikerketa Agronomikoen Kontseiluko Zuzendaria. Ronald Fisher Irakasle eta estatistikari ospetsuaren ikaslea. F.A.O. erakundearen adituetako bat Indiako Ikerketa Agronomikoen Kontseiluan (1952-1973). Unibertsitate-Konputuko Zentruetarako Britainiar Kontseiluko Lehendakaria. Biometric Society-ko Lehendakaria. Royal Statistical Society-ko Lehendakaria. F.A.O.ren Estatistika-Batzordeko kide iraunkorra. I.S.I.ko bazkide iraunkorra. Royal Statistical Society-ko partaidea. Honoris Causa Doktore Belgikako Gembloux-ko Estatu-Unibertsitatearen aldetik. Honoris Causa Doktore B.H.ko Londreseko City University-ren aldetik. Adolphe Quetelet Elkartearen Ohorezko Bazkide. Esperimentuen diseinuari buruzko hainbat libururen egile.*

David J. Finney *es catedrático de Estadística. Director del Dpto. de Estadística de la Universidad de Edimburgo. Comendador del Imperio Británico. Miembro de la Real Academia Británica. Director del Consejo de Investigaciones Agronómicas (Escocia). Alumno del prestigioso estadístico Profesor Ronald Fisher. Experto de la F.A.O. en el Consejo de Investigaciones Agronómicas de la India (1952-1973). Presidente del Consejo Británico para los Centros de Cómputo Universitario. Presidente de la Biometric Society. Presidente de la Royal Statistical Society. Miembro permanente del Comité Estadístico de la F.A.O. Miembro permanente del I.S.I. Fellow de Royal Statistical Society. Doctor Honoris Causa por la Université de l'Etat á Gembloux (Bélgica). Doctor Honoris Causa por la City University (Londres, UK.) Doctor Honoris Causa por la Universidad de Calcuta (India). Miembro Honorario de la Sociedad Adolphe Quetelet. Autor de numerosos libros sobre Diseño de Experimentos.*

David J. Finney *est professeur de Statistique. Directeur du Département de Statistique de l'Université d'Edimbourg. Commandeur de l'Empire Britannique. Membre de l'Académie Royale Britannique. Directeur du Conseil de la Recherche Agronomique (Ecosse). Il a été l'élève du prestigieux staticien, le Professeur Ronald Fisher. Il est actuellement Expert de l'O.A.A. dans le Conseil de la Recherche Agronomique de l'Inde (1952-1973). Président du Conseil Britannique pour les Centres de Computation Universitaires. Président de la Biometric Society. Président de la Royal Statistical Society. Membre permanent du Commité Statistique de la O.A.A. Membre permanent de l'I.S.I. Fellow de la Royal Statistical Society. Docteur Honoris Causa par l'Université de l'Etat a Gembloux (Belgique). Docteur Honoris Causa par la City University (Londres, UK). Docteur Honoris Causa par l'Université de Calcuta (Inde). Membre Honoraire de la Société Adolphe Quatelet. Auteur de nombreux ouvrages sur "Planification et Plan d'Experiences".*

David J. Finney *Is Professor of Statistics an Director of the Statistics Department at the University of Edinburgh. He is a Knight Commander of the British Empire and a menber of the Royal Academy. He is Director of the Council for Agronomic Research (Scotland). He has been a student of the distinguished statistician, Professor Ronald Fisher and he was an F.A.O. expert on the Indian Council for Agronomic Research (1952-1973). President of the British Council of University Computer Centres. Permanent member of the Statistics Committee of the F.A.O. Permanent member of the I.S.I. Fellow of the Royal Statistical Society. Honorary Doctor at the State University of Gembloux (Belgium). Honorary Doctor of the City University, London (U.K.). Honorary Doctor of the University of Calcuta (India). Honorary Member of the Adolphe Quetelet Society. Author of numerous books on Experiment Design.*

AURKIBIDEA

INDICE

SAIAKUNTZEN DISEINU ETA PLANIFIKAZIOA

D. J. Finney
Edinburgh University
SCOTLAND

LABURPENA

Finney Profesorearen hitzaldiak hamar kapitulutan banatuak doaz, eta esan dezakegu, kapitulu horietan zehar, esperimentuen diseinu eta planifikazioari buruzko ideiarik oinarrizkoenak lantzen direla bertan, autoreak berak bere ikerketa propioetan lortutako emaitzekin batera.

Lehen eta bigarren kapituluak esperimentuen diseinuaren funtsezko ideiak erakusten dizkigute. Hirugarrenak bariantzen analisia aztertzen du, R. A. Fisher-en ideien arabera. Lau eta bostgarrenetan, berriz, blokeen diseinuaz eta aldakuntzez iharduten da. Seigarrena diseinu faktorialei dagokie eta zazpigarrena bestelako diseinuei, hala nola, sekuentizalei, konbinatoriei, errepikakoak, etab. Zortzigarren eta bederatzigarren kapituluei dagokienez, entseiu biologikoen edo bio-entseiuen funtsezko kontzeptuak garatzen dituzte era labur batean, izan ere hauxe lantzen eman bait du Fisher Profesoreak bere bizitzaren zati handi bat.

Eta azkenik, hamargarren kapituluan, estatistikariak esperimentu baten planifikazioan eta garapenean topa ditzakeen problema desberdinak azaltzen dira.

D. J. FINNEY

DESING OF EXPERIMENTS

I. DESIGN AND ESTIMATION ; COMPLETE RANDOMIZATION

1. INTRODUCTION

My dictionary states that an *experiment* is an action undertaken in order to discover something that is unknown. Note the implication that the experimenter has some freewill: He determines the particular conditions of experimentation. Thus an early chemist wished to know whether burning material required a supply of air. His experiment was to place a lighted candle inside a container and then to seal the container so that no more air could enter: of course the candle was soon extinguished. The experimenter applied the *treatment* of the sealed container to the *subject*, a lighted candle. He had the power to withhold or to use the treatment, and indeed to define it exactly in terms of size, shape, materials, etc.

I shall be concerned solely with *comparative experiments*, that is to say experiments in which two or more treatments are to be compared with one another in respect of some measurable property. Obviously one wishes to make a fair comparison, ensuring that all other conditions relevant to the outcome are as similar as possible and subjects differ only in the treatments they receive. Simbolically (in a notation that I shall later replace by something more exact), the measurement y for two subjects might be represented by

$$y_1 = M + G + F + T_1 \tag{1.1}$$

and

$$y_2 = M + G + F + T_2 \tag{1.2}$$

for two treatments. each applied to one subject. Here M is a general mean level, G represents deviations from this associated with factors inherent in the subject -in a biological context, possibly the modification of M appropiate to an animal of a particular strain, genetic constitution, sex, age, etc. - and F represents deviations associated with environmental factors such as diet and temperature. Provided that G and F are identical for the two subjects,the difference between the measurements, $y_1 - y_2$, is equal to the difference $T_1 - T_2$ between the effects of the treatments.

Two assumptions are implicit here: (i) subjects can be found for which G and F are identical, and (ii) subjects identical in G, F and treatment will give exactly the same y measurement. These may be near enough to the truth for candles burning in sealed containers. They are certainly untrue in biological experimentation, where y may be a measurement of weight or of blood sugar or of survival time for an animal receiving a specified drug treatment. The equations need to be modified to

$$y_1 = M + G + F + T_1 + \varepsilon_1 \qquad (1.3)$$

$$y_2 = M + G + F + T_2 + \varepsilon_2 \qquad (1.4)$$

Here G, F now relate to *average* inherent and environmental effects for a class or population of very similar but inevitably not identical subjects; ε_1, ε_2 are measures of individual variation among subjects chosen to be as homogeneous as possible, combined with effects arising from variation in inherent and environmental characteristics relative to the mean.

Essentially the same situation arises in many other branches of science and technology: for example,

Subjects	**Treatments**	**Measurements**
Sufferers from a disease	Medicines	Time to recovery
Children	Methods of teaching	Test performance
Metal sheets	Methods of rust proofing	Amount of rust after 1 year
Automobiles	Types of petroleum	Consumption for 100 km
Plots of wheat	Fertilizers	Crop yield

is applicable.

We can no longer state that

$$T_1 - T_2 = y_1 - y_2 ,$$

and in general we know nothing about values of ε_1, ε_2. Two steps are open to us

(i) **Randomization**

By allowing chance, a fair lottery, to determine which subject has each treatment, we ensure that the error in using $y_1 - y_2$ as the value of $T_1 - T_2$ is equally likely to be $\varepsilon_1 - \varepsilon_2$ or $\varepsilon_2 - \varepsilon_1$. If the experiment were repeated many times, on average the value obtained for $T_1 - T_2$ would be correct.

(ii) **Replication**

The principle of randomization is one of the major contributions of statistics to research. Yet in itself it does not remove all difficulties. Replication provides the answer. By assigning several subjects to each treatment, the uncertainties arising from the "experimental errors", ε, are reduced. A treatment is now assessed in terms of the mean of y for the subjects, and the familiar result on variances:

Var(mean of r) = Var(single observation)/r (1.5)

Experimental design is concerned with exploiting these ideas so as to use available subjects and materials to best advantage in minimizing the variance of treatment means. In particular, relations among the subjects may be used to reduce the contribution to error arising from differences in the inherent and invironmental characters of the subjects.

2. PARAMETERS

Let us now formulate these ideas a little more exactly. Suppose that y is measured for subject number k of those from a group with genetic and environmental classification i that receive treatment j. We can rewrite earlier equations as

$$y_{ijk} = \mu + \beta_i + \tau_j + \varepsilon_{ijk} \qquad (2.1)$$

Here μ is a general mean, β_i embodies the previous G and F and is the deviation from μ corresponding to the average state of the particular inherent and environmental class used, and τ_j is the further deviation associated with treatment j (commonly termed the *effect* of treatment j); ε as before is the residual error arising from variability within the category i and from variability of individual subjects, and is labelled with ijk simply to show its co respondence with the observation y_{ijk}.

We refer to the τ_j (for j = 1,2,...,t if there are t treatments under discussion - t need not be 2) as *parameters*. (This is a much over-used word today, both in medical literature and by journalists. To the statistician it has long had a precise meaning as a numerical value characterizing a population or a theoretical formulation, usually unknown but requiring to be estimated from data). In addition to these treatment parameters, the β_i are also parameters relating to the environment or other background characteristic of the subjects; we may often refer to them as *block* parameters. Also μ is a parameter for the general mean.

We shall further suppose that the expectation of ϵ^2 is the same for all observations,

$$E(\epsilon^2) = \sigma^2 \qquad (2.2)$$

where the parameter σ^2 is the variance (σ the standard deviation) per observation. This assumption or constant variance is usually reasonable unless y is exceptionally variable, in particular more variable for some treatments than for others. An assumption that the ϵ have a Normal or Gaussian distribution is *not* needed for the main discussion of design.

3. ESTIMATION

In the simple situation envisaged in Section 1, no problem of estimation enters. It was implied there that all subjects have the same , say β_1 , so that for subjects on different treatments we have

$$\left.\begin{aligned} y_{11k} &= \mu + \beta_1 + \tau_1 + \epsilon_{11k}, \\ y_{12k} &= \mu + \beta_1 + \tau_2 + \epsilon_{12k}, \\ y_{13k} &= \mu + \beta_1 + \tau_3 + \epsilon_{13k}, \end{aligned}\right\} \qquad (3.1)$$
etc.

Evidently if we form means (averages) for subjects on each treatment in turn, differences between these means will estimate differences between the corresponding τ parameters, as the $(\mu + \beta_1)$ is common to all.

We shall see that we may want to involve more than one β in an experiment. There are two reasons - to reduce the error variance or to broaden the basis for inference. If β_1 represents a very broad class, for example all ages of subject, the ϵ will be much influenced by variation from sujbect to subject associated with age. By restricting subjects to a narrow age range, the variability of ϵ and hence the value of σ^2 is reduced. By retaining subjects from several distinct age groups within one experiment, we may hope to give results a broader validity. Thus in humans we might have subjects of ages 20-30, 30-40, 40-60, 60-80, each with its own β . Again, in a comparison of fertisizers for wheat, we might wish for an inference based on several different times of sowing.

Consider a very simple example. Suppose we wish to study 3 treatments; we might use 4 age groups,1 subject in each:

Treatment (τ)	Age (β) 1	2	3	4
1	(x)	x	x	x
2	x	x	x	x
3	x	x	x	(x)

In this symmetric situation, we can still estimate differences between treatment parameters just as before, averaging over all age groups. Treatments are balanced with equal numbers from each age group. The same will be true if we have, say, 9 subjects in age group 2 with 3 on each treatment. But now suppose only 10 subjects had been used - the two in ( ) do not exist. Then simple averaging of the remaining subjects may mislead. The average for treatment 2 is over all age groups, but that for treatment 1 omits the youngest subjects. Hence, if y is a measurement tending to increase with age irrespective of treatment (e. g. blood pressure), this simple comparision will be biased.

However, we have values of y for 10 subjects and can write the parametric formulation for each:

$$\left.\begin{aligned} y_{211} &= \mu + \beta_2 + \tau_1 + \epsilon_{211}, \\ y_{331} &= \mu + \beta_3 + \tau_3 + \epsilon_{331}, \end{aligned}\right\} \qquad (3.2)$$
etc.

(k = 1 everywhere, as we have only 1 subject in each "cell" of the table). How do we choose numerical values for the parameters so as to optimize agreement between observations and parameters? A widely accepted statistical principle is that of *Least Squares*: estimate the parameters in such a way as to make the sum of que squares of residuals as small as possible (Finney, 1980; Yates, 1933). If m, b_i, t_j are estimates of the corresponding μ, β_i, τ_j,residuals are defined as

$$\left.\begin{aligned} e_{211} &= y_{211} - m - b_2 - t_1 \ , \\ e_{331} &= y_{331} - m - b_3 - t_3 \ , \end{aligned}\right\} \quad (3.3)$$

etc.

The method of estimation is to minimize the sum of the e^2 by suitable choice of numerical values for the parameters. I cannot go into detail about the general principle, except to say that it has many desirable theoritical properties: it is unbiased, gives estimates with minimal variance, and, if errors have a Normal distribution, it is equivalent to the fully efficient maximum/ likelihood procedure.

Hence we should write

$$S = (y_{211} - \mu - \beta_2 - \tau_1)^2 + (y_{311} - \mu - \beta_3 - \tau_1)^2 + (y_{411} - \mu - \beta_4 - \tau_1)^2 + \ldots\ldots + \ldots\ldots + (y_{331} - \mu - \beta_3 - \tau_3)^2 \quad (3.3)$$

and then minimize S by appropriate choice of values/ for the parameters.

In the simple symmetric situation that I first mentioned, for each age group 1 subject (or more generally a constant number of subjects) on each treatment,we can easily prove that minimization of S reduces to simple averaging. If m, b_i, t_j now denote these least squares estimators, m is the mean of all observations, $(m + b_i)$ is the mean of all in age group i, $(m + t_j)$ is the / mean of all on treatment j. In the absence of symmetry and balance, the values of m, b_i, t_i are less obvious and must be obtained by solving sets of linear equa- / tions. But we now have a tractable and sensible numerical procedure for all cases, although I have left some details unexplained. These ideas will recur in later / lectures. The method of least squares underlies the whole of analysis of variance and multiple regression methodology, but we shall not pursue this very far.

4. COMPLETE RANDOMIZATION

Here and in Lecture II, I assume some familiarity/ with analysis of variance. I shall discuss the general structure of this analysis and the essential concept of orthogonality in Lecture III.

The simplest experimental design is that in which the available subjects are allocated completely at random among t treatment, the experimenter simply deciding how many shall go to each treatment. For example,/ with $t = 5$ and 43 subjects available for use, the experimenter might draw lots so as to allocate 11 subjects to the first treatment, 6 to the second, 14 to the third, 4 and 8 to the fourth and fifth. I shall discuss the choice of numbers in Lecture II.

Table 4.1 shows bone ash percentages in chickens on four different vitamin D preparations. You may think the conclusions obvious, but the data can illustrate the computations. We can calculate a sum of squares of deviations within each treatment:

$$\begin{aligned} T_1 &: \ 5.2^2 + 6.9^2 + \ldots\ldots + 6.1^2 - \frac{33.3^2}{7} \\ &\quad = 18.54 \quad (6\text{df}) \\ T_2 &: \ 43.56 \quad (5\text{df}) \\ T_3 &: \ 53.92 \quad (7\text{df}) \\ T_4 &: \ 29.71 \quad (6\text{df}) \end{aligned}$$

(I am deliberately not checking all my arithmetic; to / check and correct it, here and elsewhere, is a good exercise for the student!) Combining ("pooling") all evidence on variance gives the estimate

$$s^2 = (18.54 + 43.56 + 53.92 + 29.71)/(6 + 5 + 7 + 6) = 6.072 \quad (24\text{df})$$

The variance of a treatment mean is s^2/r_1, where r_1 / is the number of subjects. Table 4.2 shows means and / standard errors. We can form SEs of differences, estimate any $\tau_i - \tau_j$ and put probability limits on it, make significance tests, and so on at will. I comment only / that the difference between T_1, T_2 and T_3, T_4 is unquestionable, whereas differences within these pairs are appreciably less than twice their standard errors.

Consider an alternative method of computation, here having little advantage but important for the future:

(i) Form total sum of squares of deviations for all 28/ observations:

$$5.0^2 + 6.9^2 + \ldots\ldots + 13.1^2 - (33.3 + 42.9 + 106.0 + 106.5)^2/28 = 3635.69 - 2976.70 = 658.99$$

(ii)Form sum of squares "between treatments" (explana-/ tion in Lecture III):

$$\frac{33.3^2}{7} + \frac{42.9^2}{6} + \frac{106.0^2}{8} + \frac{106.5^2}{7} - \frac{288.7^2}{28} = 513.27$$

Insert these values into Table 4.3, the analysis of variance, and obtain the error sum of squares by subtraction. This is easily seen to give exactly the same error mean square, s^2, as before, except for arithmetical rounding.

Comparision of the treatments and error mean squares, by a variance ratio test, provides a test of significance of the null hypothesis "All τ_i are equal". Strictly, the test is valid only if the ϵ are Normally distributed, but in practice this restriction does not matter greatly. Certainly here, where

$$F = 28.2 \quad \text{with} \quad 3.24 \text{ df}$$

there is no doubt about significance. However this type of comprehensive test is seldom important.

In most experiments, the primary value of the analysis of variance is as a procedure for obtaining the error mean square, this being the basic variance for use in particular significance tests, in expressing the precision of estimates of parameters and comparisons among parameters, and in calculating limits of error at stated probabilities.

TABLE 4.1

Comparison of Four Vitamin D Preparations
(percent bone ash in chickens)

Treatments	T_1	T_2	T_3	T_4
Percentages	5.0	5.6	8.7	17.0
	6.9	9.0	11.9	16.3
	4.6	7.8	13.3	17.2
	5.7	2.2	15.9	14.8
	1.8	7.6	16.9	16.7
	3.2	10.7	15.8	11.4
	6.1		11.9	13.1
			11.6	
Totals	33.3	42.9	106.0	106.5

TABLE 4.2

Summary of Means for Table 1.1

Treatment	T_1	T_2	T_3	T_4
Mean	4.8	7.2	13.2	15.2
SE	±0.93	±1.01	±0.87	±0.93

TABLE 4.3

Analysis of Variance for Table 1.1

Adjustment for mean		2976.70	
Variation	df	Sum of squares	Mean square
Treatments	3	513.27	171.09
Error	24	145.72	6.072
Total	27	658.99	

II. RANDOMIZED BLOCKS; ACCIDENTS AND SALVAGE

1. COMPLETE RANDOMIZATION: THE EXPERIMENTAL PLAN

Section I 4 used entirely arbitrary numbers of subjects per treatment. How should these numbers be chosen? Many different situations can be envisaged. Only two will be discussed here.

Suppose that the experimenter is prepared to use N subjects in all, r_i to treatment T_i for i = 1, 2, ..., t. Thus

$$r_1 + r_2 + \ldots\ldots + r_t = N. \qquad (1.1)$$

If all treatments are equally interesting, one might wish to minimize the average variance of differences between pairs of treatments. This means that the average value of

$$\frac{1}{r_i} + \frac{1}{r_j} \quad \text{for all i,j pairs}$$

is to be minimized subject to equation (1.1). A little differential calculus shows the minimun to be given by

$$r_i = N/t \qquad (1.2)$$

for all i. That is to say, replicate all treatments equally. Of course, N may not be an exact multiple of t, in which case some treatments must have one extra subject.

Next suppose that T_1 is a standard treatment with which each other treatment is to be compared, but that no interest attaches to differences among the others. Then the need is to minimize the average value of

$$\frac{1}{r_1} + \frac{1}{r_j} \quad \text{for all i other than i=1,}$$

still subject to equation (1.1). The answer is less obvious, though a relatively simple proof leads to

$$r_1 = \frac{N}{1+\surd(t-1)}\,,$$
$$r_2 = r_3 = \ldots\ldots = r_1/\surd(t-1). \qquad (1.3)$$

In other words, T_1 should be replicated to a greater extent than the others by a factor $\surd(t-1)$. This can seldom be achieved exactly in integers, but a small departure makes little difference.

Thus we see that, for the Vitamin D experiment of Table I 4.1, had all comparisons been of equal interest the optimal design would have had 7 chickens on each treatment. On the other hand, had interest been solely in comparison between T_1 and each of the other three, $r_1 = 10$, $r_2 = r_3 = r_4 = 6$ would have been a better plan (10/6 is close to $\surd 3$). Of course there is no reason to expect s^2 to depend upon the r_j. The alternatives here are not very different. The first arrangement gives a variance $0.29\ s^2$ for every difference between two of the T_j ; the second makes the variance $0.27\ s^2$ for T_1 with any one of the others, $0.33\ s^2$ for any pair of T_2, T_3, T_4 . In practice, any advantage in departing from equality of the r_i is small unless there are ten or more treatments.

2. RANDOMIZED BLOCKS

Table 2.1 shows uterine weights of ovariectomized rats that had received one of four preparations of oestrone. The available animals were four females from each of seven litters. Randomization was within litters: the four treatments were randomly allocated each to one rat in litter I, each to one in litter II, and so on. This is a *randomized block* design. The appropriateness of equation I (2.1) should be obvious, except that k=1 throughout and can be omitted: for the rat on treatment j in litter i

$$y_{ij} = \mu + \beta_i + \tau_j + \varepsilon_{ij}\,. \qquad (2.1)$$

Why do this? One could not obtain 28 rats from a single litter! A completely randomized design would be legitimate, but then the block (litter) paraments would be combined with the error components and thus inflate the effective error.

Analysis of variance (Table 2,2) now begins to show its merits. The total sum of squares:

$$0.54^2 + 0.49^2 + \ldots + 1.08^2 - (24.19)^2/28 = 2.5757$$

is found without difficulty. (Again, I leave you to check my arithmetic!)

The sum of squares for treatments requires the same calculation as in Section II, now somewhat simpler because all are equally replicated:

$(4.64^2 + 7.39^2 + 5.19^2 + 6.97^2)/7 - 24.19^2/28 =$
$= 0.7671.$

Similarly a sum of squares for blocks can be formed:

$(3.59^2 + 2.57^2 + \ldots\ldots + 4.75^2)/4 - 24.19^2/28 =$
$= 0.9232.$

Because of the balance of the design - each treatment/ appears equally often in each block - these two sums / of squares are independent (in the sense to be ex-// plained in Lecture III, *orthogonal*): both can be sub-/ tracted from the total to leave the error with (27-6-/ -3) df, $s^2 = 0.04919$.

The comprehensive test of significance for treat-/ ments

$$F = 5.20 \text{ with } 3,\ 18 \text{ df}$$

leaves little doubt that differences are real. The // summary in Table 2.3 is much tidier than that in Table I.4.2 because of equi-replication,and each treatment / mean has the variance $s^2/7 = (0.083)^2$. A difference / between two treatment means has variance $2s^2/7$; multiplication of the corresponding SE by 2.10, the 0,95 // probability value for t with 18 df, gives 0.249 as an/ uncertainly to be attached to any estimated difference. For example, $\tau_2 - \tau_1$, the difference in parameters between T_2 and T_1, is estimated as 0.393, and with pro-/ bability 0.95 we assert that the truth lies between / 0.144 and 0.642.

The symmetry of this design is such that the estimates of parameters obtained by least squares are identical with those resulting from the obvious and uncri-/ tical averaging of data. Thus for a general randomized/ block design with b blocks of t treatments, for which/ (2.1) is still appropiate, the estimates are (as already stated in Section I 3)

m = mean of all bt values of y (2.2)

b_i = (mean of all t values of block i) - m (2.3)

t_j = (mean of all b values for treatment j) - m (2.4)

For some of the designs I discuss later, notably in / Lecture V, estimation is less obvious.

3. TWO COMMENTS ON ERROR VARIANCE

The analysis of variance permits s to be found / very easily, although the procedure may seem indirect./ It is possible to make a direct calculation, leading to exactly the same answer, but this is much more labo- / rious.

Suppose a completely randomized design had been used with these 28 rats. Inter-litter variability would/ then have gone into experimental error. We can estimate that the error mean square would have been

$$(0.9232 + 0.04919 \times 21)/27 = 0.07245,$$

nearly 50% greater than s^2 . This indicates taht about 1.5 times as many rats, say 11 per preparation, would / have been needed in order to obtain the same precision / (same SEs of means).

4. PAIRED COMPARISONS

One form of randomized blocks commonly used for experiments of simple structure is that with blocks of 2,/ known as paired comparisons. A treatment may be compared with the control, untreated, state on pairs of subjects such as two rabbits from the same litter, identical twin calves, left and right sides of the human body, pairs / of patients matches for age, sex and severity of disease and so on. Randomization is exactly as for other rando-/ mized blocks.

Of course, the results can be analyzed exactly as in Section 2. An alternative is to form the difference be-/ tween pairs of y values, "treated-control", for each / block, and then to do a direct calculation of variance / on the differences. This is described in many elementary text books.

The two methods of calculation lead to identical results, as may easily be verified algebraically.

5. GENERAL COMMENT

Completely randomized and randomized block designs / are undoubtedly the two most widely used and most important experimental designs. Both are of very wide appli-/ cability in almost every field of quantitative experimentation. Moreover, most other designs, some of which I // discuss in later lectures, are generalizations and extensions of these ideas.

6. ANALYSIS OF COVARIANCE

Suppose that corresponding to every measurement of / y there is also a measurement of a variate x that is // known to have been unaffected by treatment. The most satisfactory situation is that x was measured before the / randomized allocation of treatments. For example, x may/ be the weight of an animal before the start of an expe-/

riment, and y is to be perhaps the total weight 10 // weeks after treatment or the weight of a particular organ at this later time. Alternatively, x may be the// yield of fruit from a tree in 1983, up to which year / all trees were treated alike, and y is the yield in / 1984 after an experiment to compare methods of pruning/ has begun. Although such an x is often a measurement/ of the same kind as the subsequent y, it need not be - for example, it might be height of the fruit tree; the only essential feature is that casual connexion between x and treatment can be logically excluded.

If a suitable x can be chosen and measured, it / may contain information on that component of variability in y that is not due to treatment. One resonable approximation is to modify equation (2.1) to

$$y_{ijk} = \mu + \beta_i + \tau_j + \theta(x_{ijk} - \bar{x}) + \epsilon_{ijk} \qquad (6.1)$$

(and similar modifications for other experimental designs), where θ is an additional parameter (estimated by least squares, of course), and $\bar{x}$ is the mean of / all the x_{ijk}. The consecuence of including estimation/ of θ is to revise the estimation of the treatment parameters τ_j to values representing an equalization in respect of x.

This procedure is known as the *analysis of covariance*. Computationally, it is most easily handled by / extending the analysis of variance table to include two more columns: one is the analysis of x^2, exactly like/ that of y^2, and the second is an analysis of the same form made on the products xy instead of on x^2 or y^2. Then θ is estimated as

$\hat{\theta}$ = Error sum of products/Error sum of squares for x. (6.2)

Each t_j for y can now be adjusted by

$$t_j \text{ (adjusted)} = t_j \text{ (for } y) - \theta\, t_j \text{ (for } x). \qquad (6.3)$$

There is an obvious and close connexion with linear regression, which enables variances of comparisons of the adjusted t_j to be formed from consideration of the // appropriate linear contrasts (Lecture III). Cochran & / Cox (1957) show excellent examples. I shall not say more. The method is valuable, and insufficiently exploted for use with x-variates that are easily recorded as --/ part of the experimental process. By analogy with multiple linear regression, two or more distinct x-variates/ can be used in a multiple covariance analysis. I have / introduced the subject primarily in order to be able to utilize analysis of covariance in Sections 7, 8.

7. MISSING OBSERVATIONS

Even the most careful experimenter occasionally // losses the value of y from a plot. Perhaps grazing animals invade a plot of wheat, lightning strikes a tree, or an experimental rat dies accidentally - all inci--// dents that should be unconnected with the applied treatments. What is to be done? The method of least squares is the appropriate way of handing the lack of balance./ If the experiment is completely randomized, there is no problem: simply regard the treatment involved as having one less plot and perform the standard calculations.For randomized blocks and other designs, least squares can be handled in three ways, all equivalent.

(i) Write z for the "missing value" of y. Perfom an analysis of variance in terms of z and find / the error sum of squares in the form

$$Az^2 + Bz + C \quad (A > 0). \qquad (7.1)$$

This will be minimized by

$$\hat{z} = -B/2A. \qquad (7.2)$$

Insert this numerical value of z in place of the // missing y, reduce the number of degrees of freedom for error by 1, and the completed analysis of / variance will give the appropriate s^2 as an unbiased estimated of σ^2.

For a randomized block design with b blocks / of treatmnets, this leads to

$$\hat{z} = \frac{bB + tT - G}{(b-1)(t-1)} \qquad (7.3)$$

where B is the total of all other y in the block/ from which one is missing, T is the total of all / other y for the treatment from which one is mis/ sing, and G is the total of the (bt-1) recorded values of y. Analogous formulae can be found for o-/ ther designs.

(ii) Guess a value for z, and find the estimates / of all parameters m, b_i, t_j. Then determine

$$m + b_i + t_j$$

for the "missing plot", take this in place of the / previous z, and repeat the cycle until z re-- / mains constant. This ensures that the residual /// (Section 13) for the missing plot is zero. It will/ give exactly the same $\hat{z}$ as in (i).

Although (i) and (ii) lead to unbiased estimation of σ^2, variances and standard errors of // differences in the t_j require more care. The in-/ serted value of $\hat{z}$ is, as in (7.3), a linear func/ tion of all the other y, and in consequence // variances of treatment differences are greater // than if the data were complete. A crude adjust--/ ment, slightly too extreme, is simply to regard / the affected treatment as having one less repli-/ cate.

(iii) Insert an arbitrary value in the missing po-// sition. A suitable choice is the general mean of/ all y values from other plots, but the eventual result is the same whatever is chosen. Then define x as a "dummy variate" taking the value --// (N-1) for the missing plot, -1 for all others, // where N is the total number of plots. Make a co-/ variance analysis of y on x. The adjusted t_j / will have taken proper account of the missing value, and the standard covariance analysis - li--/ near regression procedures will attend to variances and standard errors appropriately.

Methods (i), (ii), (iii), can be generalized to deal // with two or more missing observations. One must have // z_1, z_2, ... separately defined, lose as many degrees of freedom from s^2 as there are missing plots, and introduce a separate dummy variate x_1, x_2, ... for each missing plot.

Although method (i) is convenient for simple cases/ where a formula such as (7.3) is known, method (iii) / has the advantage of generality. It applies to any design, however complex, and is easily handled by any general computer package that includes covariance analy-/ sis.

Note the logical necessity that loss of a value // shall be causally independent of the applied treatment/ If the yield of a fruit tree is lost because its encouraged early ripening of fruit and consequent destruc--/ tion by birds, or if a rat dies because its experimen-/ tal diet was deficient in some vital component, any use of the methods of this Section must produce biased re-/ sults. If a particular diet under test introduces a serious risk that rats will die before they are 6 months/ old, what meaning can be given to estimating the expected weight at 12 months for rats on this diet? In such/ circumstances, statistical method alone cannot help: / the whole concept and purpose of the experiment must be re-examined.

8. SOME OTHER ACCIDENTS

The method of least squares and the analysis of covariance are powerful aids to the salvage of other ex-/ periments that have gone wrong. For example, it has -/ been known for the produce of two plots of land to be / mixed at harvest, so that only the total ot two values/ of y is known. Any of methods (i), (ii), (iii) can be adapted to this problem.

Federer & Schlottfeldt (1954) and Outwaite & Rutherford (1955) discussed plant heights recorded for seven/ treatments in eight randomized blocks. The blocks were/ side by side, and the seven plots of each block were in a single line. Unfortunately, the planning of the --// blocks failed to take adequate account of trends in -// soil fertility, and the precesion of the experiment was reduced by a consistent fertility gradient from plot / to plot within blocks. The authors recovered informa--/ tion by covariance on a dummy variate. The main step // was to define x_1 as a variate increasing linearly from/ plot to plot within each block, say x_1 = 0, 1, 2, .../ ..., 6, and to use this as a covariate. By extending // the analysis to include a carefully chosen set of 6 -// dummy variates, all trend (linear or other) from plot / to plot along the blocks can be eliminated. The experiment would have been more precise had blocking been based on the actual trend (or had a two-way Latin square/ design been used), but if the trend was not known in advance the covariance technique provides an effective -/ salvage operation.

Some years ago I was asked for help with an experiment on varieties of cereal, consisting of 36 long na-/ rrow plots side (Finney, 1962). All plot corners were / marked by pegs in the ground. The peg at one corner of/ the whole experiment was accidentally removed. Before / the accident was discoverd, the crop was harvested as / 35 parallelograms (nearly rectangular)!. Each harvested plot contained triangular halves of two adjacent true / plots. This enabled each observed yield to be expressed in terms of a block parameter and a sum of two treat--/ ment parameters. Estimation by least squares was surprisingly simple (I had no computer then!). Not surprisin-/ gly, the loss of information relative to the full expe-/ riment was great, variances being about twice what they/ ought to have been, but the important fact is that something useful was saved from an experiment that otherwise would have been discarded.

TABLE 2.1

Responses to Four Oestrone Preparations

(Uterine weights of rats, mg per g body weight)

Litter	T_1	T_2	T_3	T_4	Totals
I	0,54	1,52	0,61	0,92	3,59
II	0,49	0,71	0,74	0,63	2,57
III	0,51	1,12	0,51	1,07	3,21
IV	0,40	0,58	0,60	1,02	2,60
V	0,81	1,02	1,07	1,20	4,10
VI	0,63	0,86	0,83	1,05	3,37
VII	1,26	1,58	0,83	1,08	4,75
Totals	4,64	7,39	5,19	6,97	24,19

(The results of this experiment have been modified slightly to make them more suitable as an illustrative example.)

TABLE 2.2

Analysis of Variance for Table 2.1

Variation	df	Sum of squares	Mean square
Adjustment for mean		20.8984	
Litters (blocks)	6	0.9232	
Treatments	3	0.7671	0.2557
Error	18	0.8854	0.04919
	27	2.5757	

TABLE 2.3

Summary of Means for Table 2.1

Treatment	T_1	T_2	T_3	T_4	
Mean	0.663	1.056	0.741	0.996	$\pm$ 0.084

III. ANALYSIS OF VARIANCE

1. INTRODUCTION

Presentation of the theory of the analysis of variance in terms of matrix algebra has merits both for conciseness of theory and for formal organization of computer programs. Whether it gives the same insight into the structure of the analysis as was possessed by those who first developed the subject of experimental design may be doubted. Early papers by Fisher and by Yates are remarkable for the facility with which they produced new families of designs, and then unhesitatingly wrote the arithmetic steps for a well-organized analysis of variance; these papers still deserve reading.

Scientists unskilled in the manipulation of matrices often have difficulty in understanding the formulae underlying the analysis of variance. Yet the theory can be expressed entirely in terms of fairly elementary algebra, within the comprehension of anyone who refuses to be intimidated by expressions and equations involving several subscripts. This lecture outlines rather formally a presentation that I have long found useful in teaching Though it contains nothing new, systematic statement and proof along these lines does not seem to be readily accessible elsewhere. The approach derives from an elementary lecture given by R. A. Fisher at Rothamsted Experimental Station about 1.942, but I have inserted much more detail. I am well aware that theory and proof can be expressed more concisely in matrix terms, but I have never found this so enlightening.

My concern is particularly for the non-mathematician who is reasonably competent in algebraic manipulation, but even the professional statistician may improve his management of the analysis of variance by familiarizing himself whit simple properties of contrasts. I assume that you know the numerical techniques, at least for the simpler analyses of one-way and two-way data, though you may be unsure why the methods work. If you find the many suffixes and summations confusing, you should write out the theory exactly as it is stated but with particular small integers in place of general numbers of observations.

Nothing that follows is intended to discourage the proper use of sophisticated algebraic techiniques. The aim is solely to supplement these by a more elementary approach that may help some people to appreciate the underlying logic. Note also that nothing in this Lecture depends upon any assumption of Normality.

2. NOTATION

Suppose y_i for $i = 1, 2, \ldots, n$ to be n independent observations on a variate y . Define U_j as linear function of the y_i, arbitrary except for the restriction that at least one of the coefficients a_{ji} differs from zero:

$$\left.\begin{aligned} U_j &= a_{j1}y_1 + a_{j2}y_2 + \cdots\cdots + a_{jn}y_n \\ &= \sum_{i=1}^{n} a_{ji}y_i \,. \end{aligned}\right\} \tag{2.1}$$

Until further notice, the summation $\sum_{i=1}^{n}$ will be written simply as Σ ; all other summations will have their limits stated.

One particularly important linear function in the simple sum of the y_i ; this will be referred to as U_o , where

$$U_o = \Sigma y_i \quad (a_{oi} = 1 \text{ for all } i) \,. \tag{2.2}$$

Clearly the mean of all observations is

$$\bar{y} = U_o/n \,. \tag{2.3}$$

3. A FUNDAMENTAL THEOREM

The analysis of variance depends upon the following general algebraic theorem:

> Suppose that p linear functions $U_1, U_2, \ldots, U_p$ are defined as in Section 2. The equation
>
> $$\Sigma\, y_i^2 = \sum_{j=1}^{p} U_j^2 \tag{3.1}$$
>
> will be true for every possible set of numerical values for the y_i *if and only if* three conditions are satisfied:

$$p = n \,, \tag{3.2}$$

$$\Sigma a_{ji}^2 = 1 \quad \text{for each of } j=1, 2, \ldots, p \,, \tag{3.3}$$

$$\Sigma a_{ji}a_{ki} = 0 \quad \text{for each pair of unequal } j, k \,. \tag{3.4}$$

The proof is not difficult, but is a piece of formal mathematics that I need nor give. The theorem is closely related to the well-known fact that a set of p independent linear equations in n unknowns will have an unique solution only if p=n . Note that (3.4) is the key . Equation (3.2) merely demands the right number of components. Equation (3.3) can be secured by a simple standardization, dividing all coefficients of U_j by an appropriate amount. Indeed, a possibly simpler statement of / the theorem is that

$$\sum_{j=1}^{n} (U_j^2/D_j) = \Sigma\, y_i^2 \quad , \qquad (3.5)$$

where D_j is defined by

$$D_j = \Sigma\, a_{ji}^2 \quad , \qquad (3.6)$$

will be true for all possible y_i if and only if (3.4)/ is satisfied.

Equation (3.4) is said to be the condition that U_j, U_k/ are *orthogonal*.

This terminology is geometrical in origin. Relative to axes y_i in n dimensions, $U_j = 0$ and $U_k = 0$ are the equations of two hyperplanes (straight lines if n=2 ordinary planes if n=3), and (3.4) is the condition / that the two shall intersect at right angles.

If **n>1** , a great deal of freedom exists in the definition of the U_j . One of them, U_n say, can be defined quite arbitrarily. Equation (3.4) whith n-1, n for j, k then gives one linear constraint on the coefficients of U_{n-1} : all but one of the $a_{n-1,i}$ can be chosen arbitrarily and the final one is determined. / Equation (3.4), first whit n-2, n for j, k and again whith n-2, n-1 for j, k, gives two linear constraints/ on the coefficients of U_{n-2} : (n-2) of the $a_{n-2,i}$ / can be chosen arbitrarily and the final two are determined. So one proceeds until permitted only one arbitrary/ choice of a coefficient in U_1 ; this last merely / amounts to the choice of a multiplicative factor for all a_{1i} , which leaves $U_1^2/\Sigma a_{j1}^2$ unaltered.

4. CONTRASTS

For statistical purposes, we always choose U_o in / equation (2.2) as one of the functions; I shall now use U_o in place of U_n . Then, with the aid of another well-known result.

$$\Sigma(y_i-\bar{y})^2 = \Sigma y_i^2 - \frac{U_o^2}{n}$$

$$= \sum_{j=1}^{n-1} (U_j^2/D_j) \qquad (4.1)$$

subject to the appropriate conditions. The orthogonality of each other U_j whit U_o requires that.

$$\Sigma a_{ji} = 0 \text{ for each of } j=1, 2, \ldots, (n-1) \; . \qquad (4.2)$$

Any linear function satisfying (4.2), having the sum of its coefficients zero, is said to be a *contrast* of the y_i . Hereafter the U_j notation (except U_o) will be / restricted to contrasts; D_j may be termed the *contrast divisor* and

$$Q_j = U_j^2 \,/\, D_j \qquad (4.3)$$

the *contrast square*.

The statistical form of the theorem can now be / stated:

Let U_j for j = 1, 2, ..., (n-1) be contrast among the y_i and let Q_j be the corresponding contrast/ square. Then the equation

$$\Sigma(y_i-y)^2 = \sum_{j=1}^{n-1} Q_j$$

will be true for every possible set of numerical values for the y_i if and only if every pair of contrast has coefficients satisfying the orthogonality condition (3.4)

The art of the analysis of variance consists in / choosing contrasts that are best suited to the objecti-/ves of a particular investigation. Note that each U_j and Q_j corresponds with *1 degree of freedom.*

5. SOME EXAMPLES

By an argument similar to that at the end of Section 3, if any number of mutually orthogonal contrasts less/ than (n-1) is specified, an addtitional one can be / found, and the process can be repeated until a set of (n-1) is obtained. For suppose orthogonal contrasts U_1, U_2, ..., U_j are specified. The conditions that a linear function U_{j+1} shall be a contrast and shall be orthogonal with U_1, U_2, ..., U_j give (j+1) linear equations to be satisfied by the $a_{j+1,i}$. If $j<n-1$, these will have a solution with some of the coefficients non-zero.

An example makes the process clear. Suppose $n = 4$. Take the arbitrary contrast

$$U_1 = 4y_1 - 3y_2 + 2y_3 - 3y_4 \tag{5.1}$$

Then U_2 must have

$$a_{21} + a_{22} + a_{23} + a_{24} = 0$$

$$4a_{21} - 3a_{22} + 2a_{23} - 3a_{24} = 0$$

Among the unlimited number of solutions, one possibility is

$$U_2 = 15y_1 + 5y_2 - 21y_3 + y_4 \tag{5.2}$$

any linear function with coefficients in the ratios 15 : 5 : 21 : 1 is essentially the same contrast. Next U_3 is required to satisfy the condition that it is a contrast:

$$a_{31} + a_{32} + a_{33} + a_{34} = 0$$

and that it is orthogonal with U_1 and U_2 :

$$4a_{31} - 3a_{32} + 2a_{33} - 3a_{34} = 0$$

$$15a_{31} + 5a_{32} - 21a_{33} + a_{34} = 0$$

The only solution, except for a factor running right through, is

$$U_3 = 5y_1 - 56y_2 - 7y_3 + 58y_4 \tag{5.3}$$

Thus U_1 is arbitrary, U_2 has been chosen with an arbitrary element, and U_3 is determinate; of course no U_4 can be found orthogonal to these three. Exactly the same process can be used for any n. Algebraic verification, or trial of arbitrary numerical values for the y_i, should convince the reader that

$$\Sigma(y_i-\bar{y})^2 = \frac{U_1^2}{38} + \frac{U_2^2}{692} + \frac{U_3^3}{6574} \tag{5.4}$$

with U_1, U_2, U_3 as defined by (5.1)-(5.3).

One general set of orthogonal contrasts, often useful as an example, is

$$\left.\begin{array}{lcl} U_1 & = & y_1 - y_2 \\ U_2 & = & y_1 + y_2 - 2y_3 \\ & \cdot & \\ & \cdot & \\ & \cdot & \\ U_j & = & y_1 + y_2 + \ldots + y_j - jy_{j+1} \\ & \cdot & \\ & \cdot & \\ & \cdot & \\ U_{n-1} & = & y_1 + y_2 + \ldots + y_{n-1} - (n-1)y_n \end{array}\right) \tag{5.5}$$

Every pair of these is easily seen to satisfy (3.4); they have

$$D_j = j\,(j+1)\,. \tag{5.6}$$

6. A SECOND THEOREM

Recall expressions such as II(2.1) for individual observations. They can be expressed here as

$$y_i = \eta_i + \epsilon_i \tag{6.1}$$

where η_i is the parametric expression corresponding to y_i and ϵ_i is the "error". Then the expectation of y_i is

$$E(y_i) = \eta_i\,, \tag{6.2}$$

and we usually take

$$E(\epsilon^2) = \sigma^2 \tag{6.3}$$

as in I(2.2). Evidently

$$E(U_j) = \Sigma a_{ji}\eta_i\,. \tag{6.4}$$

Moreover, we can write

$$U_j = \Sigma a_{ji}\eta_i + \Sigma a_{ji}\epsilon_i\,.$$

If we square this and take expectations, remembering that the η_i are constants and the different ϵ_i are independent of one another, we obtain

$$\begin{aligned} E(U_j^2) &= (\Sigma a_{ji}\eta_i)^2 + \sigma^2\Sigma a_{ji}^2 \\ &= [E(U_j)]^2 + \sigma^2 D_j\,. \end{aligned}$$

This result can be written

$$E(Q_j) = \frac{[E(U_j)]^2}{D_j} + \sigma^2\,. \tag{6.5}$$

Thus if U_j is a contrast whit zero expectation, Q_j

has expectation σ^2 ; otherwise, that σ^2 is increased by a component proportional to the square of $E(U_j)$. / This is the basis of estimation of σ^2 from suitable / mean squares in the analysis of variance, as well as / for "variance ratio" tests of significance.

7. COMPLETELY RANDOMIZED DESING

As in Section II 1, suppose the n observations to be r_h from treatment T_h for h=1, 2, ..., t . Write Y_h for the sum of the r_h values of y for treatment T_h . Now define

$$\left.\begin{aligned} U_1 &= r_2 Y_1 - r_1 Y_2 \\ U_2 &= r_3 (Y_1+Y_2) - (r_1+r_2) Y_3 \\ U_3 &= r_4 (Y_1+Y_2+Y_3) - (r_1+r_2+r_3) Y_4 \\ &\qquad\text{etc., up to } U_{t-1} \end{aligned}\right] \quad (7.1)$$

These expressions must be regarded as abbreviations for linear functions of the y_i . If each Y_h is written / in full as a sum of observations, each U_1, U_2, ..., / U_{t-1} is seen to be a contrast and every pair conforms to equation (3.4). These (t-1) contrasts may be termed *contrasts between treatments*, since they can be calculated from treatment totals (although for their proper--- ties as contrats one must always think of them as written in full in the y_i).

Now consider treatment T_1 alone. The total of its r_1 replicates is Y_1 , and the theorem of Section 4 shows that we can find (r_1-1) mutually orthogonal contrasts among the r_1 values of y (the a coeffi---- cients for all y not in T_1 will be zero). Any such contrast is obviously orthogonal to U_1, U_2, ..., U_{t-1}. Moreover the sum of the Q's for these (r_1-1) con---- trasts is "$\Sigma(y-\bar{y})^2$" within treatment T_1. We can argue similarly for each treatment in turn, noting that any contrast "within T_1" is orthogonal whit any " within T_2" etc. Thus we identify a total of (n-t) contrasts within treatments whose contrast squares sum to

$$\Sigma y_i^2 - \frac{Y_1^2}{r_1} - \frac{Y_2^2}{r_2} - \dots - \frac{Y_t^2}{r_t} \quad . \qquad (7.2)$$

If we also include the (t-1) contrasts between treat-- ments, we have a full set of (n-1) mutually orthogonal contrasts, and together the contrasts contrast squares must form $\Sigma(y-\bar{y})^2$ for the n observations. It immediately follows that

$$Q_1+Q_2 + \dots + Q_{t-1} = \frac{Y_1^2}{r_1} + \frac{Y_2^2}{r_2} + \dots + \frac{Y_t^2}{r_t} - \frac{(\Sigma y)^2}{n} \quad (7.3)$$

Hence we see how Table I (4.3) is constructed. / Moreover, all the within treatment contrasts have expectation zero, so that the mean square for these (n-t) / df estimates σ^2 . Each of U_1, U_2, ..., U_{t-1} has an expectation involving treatment parameters, as in / equation (6.5). Therefore the expectation of the mean / square from the (t-1) df for treatments exceeds by an average value of the (t-1) expressions / $\left[E(U_j)\right]^2/D_j$. The two mean squares have equal expecta-- tions if and only if all $E(U_j)$ are zero, which is equivalent to the condition that all treatment parameters / are equal. Hence it is appropiate to use the ratio of the mean squares as the basis for a test of significance and also to use the estimate of σ^2 from the error / line as a basis for attaching a standard error to any / contrast between treatment means.

8. RANDOMIZED BLOCKS

In a randomized block desing, contrasts between treatments can be isolated in exactly the same manner. But the t treatments and the b blocks enter symmetrically and so (b-1) orthogonal contrasts between blocks can / also be isolated. One then easily sees that any block / contrast is orthogonal with every treatment contrast. / There must remain

$$(n-1) - (t-1) - (b-1) = (t-1)(b-1)$$

contrasts orthogonal both with treatments and with / blocks; these must have zero expectation whatever the / treatment and block parameters may be, and therefore the mean square with (t-1)(b-1) df estimates .

As a simple illustration, easily extended to the experiment discussed in Section II 2, suppose we have 3 / treatments in 4 blocks. Then the set of coefficients

	T_1	T_2	T_3
I	-2	3	-1
II	-2	3	-1
III	-2	3	-1
IV	-2	3	-1

defines a treatment contrast: it is a contrast because / the 12 coefficients add to zero and it is a treatment / contrast because the 4 replicates of a treatment have the same numerical coefficient. Similarly

	T_1	T_2	T_3
I	-1	-1	-1
II	4	4	4

III	-2	-2	-2
IV	-1	-1	-1

is a block contrast. Verify that these two are orthogo--nal, if you do not find it obvius. Now considerer

	T_1	T_2	T_3
I	2	3	-5
II	-4	-6	10
III	6	9	-15
IV	-4	-6	10

This is a contrast, and is easily verified to be orthogo nal with each of the other two.

I shall not complete the argument in detail, but what has been said rapidly leads to Table II 2.2 and all that follows from it. A further result, requiring a little mo re algebra than I have shown but being an extension of / (6.5), is that the expectation of the treatment mean square is

$$\sigma^2 + \frac{b}{t-1} \sum_{h=1}^{t} (\tau_h - \bar{\tau}) \qquad (8.1)$$

9. GENERALIZATIONS (FOR FUTURE REFERENCE)

The two-way analysis of variance is the basis of the/ statistical analysis of most planned experiments. As described here, it relates to a randomized block design. If each block has p observations from each treatment / instead of only one, an estimator of σ^2 alone can be / found from a sum of squares within classes: the analysis discussed in Section 8 is then based on totals for each of the bt classes, division of all sums of squares by p gives expectations as above, and the intra-class mean / square with bt(p-1) degrees of freedom obviously has expectation σ^2 .

Generalizations to more complex designs follow the same pattern. A Latin square (b=t) has another sum of / squares exactly like those for rows and columns extracta-ble from what was previously "error". A factorial design allows the treatment sum of squares to be subdivided into components that can be separately interpreted, and each has an expectation involving the relevant parameters. / Incomplete block designs are more troublesome, as the expectation of a crude sum of squares for treatments in--volves block parameters and vice versa. However, the ana-lysis described in Section V 4 forms a sum of squares for treatments adjusted for blocks (or an intrablock compo---nent of the treatment sum of squares) for which the expec ted mean squares is again σ^2 plus a function of the / treatment parameters only. Designs for sample surveys may involve hierarchical analysis of variance, possibly with/ the pattern of Section 7 repeated at each successive pair of levels of the hierarchy and preferably (for simplicity al least) with all r_h equal.

IV. LATIN SQUARES AND HIGHER ORTHOGONALITIES

1. WHICH BLOCKS?

In some situations where a randomized block design is comtemplated the experimenter may recognize two (or more) different blocking systems as candidates for use. The classical instance is that of experimentation on / agricultural crops. Evidence on the fertility varia- / tions within a site may suggest that blocks should be separated by lines running east-west: fertility is / thought to decline steadily from north to south,so that such blocks will minimize intra-block variation. On the other hand, the pattern of new irrigation facilities ,/ or even the convenience of field operations,may suggest that block boundaries should run north-south. A *Latin square* design will permit the two systems to be used si multaneously; it does require that the number of treatments is the same as the number of each kind of block. For example, if A, B, C, D, E are 5 varieties of wheat that are to be compared, the field arrangement might/ be:

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

Note that both the rows and the columns of this arrange ment satisfy the block condition, in that each row and each column contains one "plot" of each treatment. The name "square" refers to the formation of the letters, / and in no way indicates that the actual plots of wheat must be square!

The use of Latin squares is not restricted to agricultural research. Exactly similar positional considera tions may arise in taking sample pieces of skin from a hide for comparison of tanning treatments, pieces of/ cloth from a large strip for comparison of dyes, or pie ces of metal from a sheet for comparisons of the qualities of applied paints. But rows and columns need not represent positions in this way. A comparison of me- / thods of inoculating leaves with a virus might have / plants as columns, order of leaf from the base as rows. A comparison of cell- counting ability among 5 technicians might use 5 microscopes and 5 slides as rows and columns.

Latin squares of any size can be constructed. If / one of the smaller squares (especially 2x2 and 3x3) does not give enough replication for an experiment, a set of 2, 3, or more squares can be used to form a larger experiment. Squares larger than 10x10 are seldom wanted for practical use, as their merits for controlling variabili ty may then be less, but particular cases can be conside red on their merits.

2. ANALYSIS OF A LATIN SQUARE

A Latin square for t treatments has t^2 observations. The balance of the design makes clear that the / sets of (t-1) contrasts for rows, columns, and treat- / ments are mutually orthogonal. Hence the analysis takes the form:

	df
Rows	t-1
Columns	t-1
Treatments	t-1
Error	(t-1)(t-2)
Total	t^2-1

in which the sum of squares for each of the first three/ lines is calculated exactly as was that for blocks or / treatments in Section II 2, using the appropriate set of t totals. The error sum of squares is again obtained by subtraction. No new features enter. Standard errors / follow in the usual way.

3. MATHEMATICAL FORMULATION

I think we should look at the kind of formulation of the observations that appears as equation II (2.1) a / little more carefully. That equation generalizes here to

$$y_{ijk} = \mu + \beta_i + \gamma_j + \tau_k + \epsilon_{ijk} \qquad (3.1)$$

for the observation in row i , column j , if this - / happens to relate to treatment k. However, there is an indeterminacy in the parameters: without altering re- / sults, one could add a fixed amount to μ and subtract that same amount from all the β_i or from all the γ_j or from all the τ_k . It is usual to overcome this by introducing the constraints

$$\text{Sum of all } \beta_i = 0 \,,$$

$$\left.\begin{array}{l}\text{Sum of all } \gamma_j = 0 \,, \\ \text{Sum of all } \tau_k = 0 \,,\end{array}\right\} \qquad (3.2)$$

(and of course similarly for randomized blocks). Thus in particular the τ_k are deviations from the general / mean, and have unique values. This becomes important for the handling of more complicated designs. The expected / mean square shown as III (8.1) becomes

$$\sigma^2 + \frac{b}{t-1} \sum_{k=1}^{t} \tau_k^2 \qquad (3.3)$$

and for a Latin square this is further modified to

$$\sigma^2 + \frac{t}{t-1} \sum_{k=1}^{t} \tau_k^2 \qquad (3.4)$$

4. A WARNING

Note that the validity of a Latin square experiment rests on equation (3.1), that is to say on the additivity of the various effects. If the deviations from the general mean associated with columns are not the same from row to row, the t quantities γ_j will need to be replaced by t^2 quantities γ_{ij}. Treatment effects may / than be masked by these "row-column interactions" and/ the experimenter is untrustworthy.

This is unlikely to be a serious worry when the / overall variability is small, but it may be if that variability is very large.

5. SOME PROPERTIES

One may easily verify that there are only 2 distinct 2x2 Latin squares and only 12 distinct 3x3 squares. Larger squares are far more numerous, 576 for 4x4 and more than 6×10^{13} for 7x7. Beyond that, no counts available.

A Latin square for use ought to be chosen at random from all squares of the desired size. In their *Statistical Tables*, Fisher & Yates (1964) describe how this can be done for the smaller squares.

For larger squares, it is adequate to take a speci-/ men square, rearrange the rows in a random order (e. g./ put rows 1, 2, 3, 4, 5 in the order 2, 4, 1, 3, 5), - / rearrange the columns in a random order, and assign / treatments to the letters of the square in a random order.

6. MODIFIED USES

The requirement that the number of rows, columns,and treatments shall be equal may seem severe. However, some variation is possible by including one treatment twice (or even 3 times) as though it were two different treatments. The only change in the analysis is that one or / more contrasts that apparently are between treatments / now have expectation zero, and the appropriate square / should be transferred to the error sum of squares. Sometimes this fits well with considerations of optimality / such as were mentioned in Section II 1.

It is also permissible to plan a Latin square and/ then deliberately to omit the final row (or the final column). This destroys orthogonality, but still leaves possible a least squares analysis such as that for incom-- plete block designs (Lecture V).

7. GRAECO-LATIN SQUARES

Consider an addition to the square in Section 1:

Cβ	Aδ	Eε	Dα	Bγ
Dγ	Eβ	Cα	Bδ	Aε
Bε	Cγ	Dδ	Aβ	Eα
Aα	Dε	Bβ	Eγ	Cδ
Eδ	Bα	Aγ	Cε	Dβ

Each Roman letter now has a Greek letter alongside it . Note that:

(i) The Greek letters themselves nave the latin /// square property, each occurs once in each row // and once in each column.

(ii) Every one of the 25 possible Roman-Greek pairs / occurs once.

This is a *Graeco-Latin square*. The Greek letters give a new classification of the observations, the contrasts / for which are orthogonal to rows, columns, and Roman letters. Consequently, formally, the data can be analyzed / by a further extension of the analysis of variance:

	df
Rows	t-1
Columns	t-1
Roman	t-1
Greek	t-1
Error	(t-1)(t-3)
Total	t^2-1

One could envisage an experiment in which one set of treatments (e.g. types of virus inoculation) is associated with Roman letters and a second set (e.g. concentration of inoculum) is associated with Greek letters. This is usually unwise, unless there is very good reason to believe that the effects of the two sets are truly additive. The design may occasionally be useful when a Latin square experiment has been conducted and subsequently the experimenter wishes to use the same material (plots of land, single trees, animals) for comparing a new set of t treatments; if he believes that effects induced by the first treatments are very likely to have disappeared, he may use a Greek pattern for the new treatments, confident that any departures from simple additivity will be small.

Graeco-Latin squares have further importance as aids to the construction of other designs (Lecture V).

Not every Latin square can have Greek letters added orthogonally. Indeed the proportion that can is quite small. Note the following pair of 4x4 Latin squares:

Aα	Bβ	Cγ	Dδ	and	A	B	C	D
Bγ	Aδ	Dα	Cβ		B	A	D	C
Cδ	Dγ	Aβ	Bα		C	D	B	A
Dβ	Cα	Bδ	Aγ		D	C	A	B

They differ only slightly, yet the first is shown with a Graeco solution and the second cannot be extended in this way. In fact only 1/4 of all 4x4 squares can have Greek letters added, and only 3/28 of all 5x5 squares. The 6x6 squares are remarkable in that no Graeco-Latin arrangements exist. For every number greater than 6,Graeco-Latins exist (a truth that was in doubt until 25 years ago), but they are relatively very scarce.

Still more complicated orthogonal structures exist. Consider the first of the two 4x4 squares still further extended:

Aα1	Bβ2	Cγ3	Dδ4
Bγ4	Aδ3	Dα2	Cβ1
Cδ2	Dγ1	Aβ4	Bα3
Dβ3	Cα4	Bδ1	Aγ2

The numerals 1, 2, 3, 4 give a further classification that is orthogonal with rows, columns, Romans, and Greeks. No more is possible for a 4x4 square. It is easily proved that, for a txt square, the number of distinct mutually orthogonal sets of symbols cannot exceed (t-1). It is also provable that this maximum can be achieved if t is a prime number or a power of a single prime (3,4, 5, 7, 8, 9, 11, 13, 16, 17, ...),but not much is known beyond this. Can 14 sets of symbols be put orthogonally into a 15x15 square? I do not know, and I do not think that anyone else knows!

8. ORTHOGONAL PARTITIONS OF LATIN SQUARES

A txt Graeco-Latin square may be regarded as a (1^t) partitioning of each row in such a way as to be simultaneously a (1^t) partition of columns and a (1^t) partition of the Roman letters. Occasionally less extreme partitions are useful. For example, if the treatments in a 5x5 square are thought to have ceased to affect the experimental units, a scientist might wish to superpose a new set of 4 treatments in a balanced manner, so that one new treatment has double replication in each row, in each column, and with each original treatment. This is easily done by making two of the Greek letters identical in the square at the beginning of Section 7: for example if each ε is replaced by δ , we have a $(1^3 2)$ orthogonal partition.

Many such partitions exist for Latin squares that do not have complete (1^t) Graeco partitions. The second 4x4 square shown in Section 7 has a (2^2) partition:

Aα	Bα	Cβ	Dβ
Bβ	Aβ	Dα	Cα
Cα	Dα	Bβ	Aβ
Dβ	Cβ	Aα	Bα

More useful are the many orthogonal partitions of 6x6 squares, especially because Graeco-Latin 6x6 arrangements do not exist. For example, there are 2^3 partitions such as

Bγ	Dβ	Aγ	Fα	Cβ	Eα
Dα	Bα	Eβ	Cγ	Fγ	Aβ
Aβ	Fγ	Bβ	Dα	Eα	Cγ
Fβ	Eγ	Cα	Bβ	Aα	Dγ
Cα	Aα	Dγ	Eβ	Bγ	Fβ
Eγ	Cβ	Fα	Aγ	Dβ	Bα

that could enable 3 new treatments to be put in a balanced arrangement on an existing 6x6 Latin square. All types or partition except (1^6), such as (3^2), (1, 2, 3), can be found for 6x6 squares. I have recently (1982b) published a very full account of these and of higher order orthogonalities for the system.

V. INCOMPLETE BLOCKS

1. THE NEED

In many circumstances in which an experimenter would like to use randomized blocks, the number of treatments exceeds the available or desirable block size - sometimes very substantially. I mention three contrasting examples:

(i) Field trials of new varieties of a major crop may need to include 20-50 varieties, yet experience suggests that the efficiency of blocks for the control of variability in agricultural experiments is much reduced for plots of more than 10 plots;

(ii) The block is to be litter-mate animals of one sex, and the number of treatments exceeds the number that can reasonably be expected to occur frequently (possibly 6 for rats or 3 for sheep);

(iii) Maximum block size is dictated by administrative convenience, such as the number of clases of some phenomenon (a disease, a type of accident, a meteorological state) occurring in one month, and the number of treatments exceeds this.

Randomized complete blocks will always be first choice unless there are strong reasons favouring smaller blocks: they are easy to interpret, efficient, and computationally simple. However, necessity may force the use of smaller blocks, or the gain in efficiency from reduced block size may be thought likely to be greater than the loss consequent upon incomplete replication within blocks.

2. THE ANSWER

Let us recall equation II (2.1):

$$y_{ij} = \mu + \beta_i + \tau_j + \epsilon_{ij} \qquad (2.1)$$

for the observation y on treatment j in block i; as in equation IV(3.2), we now adopt the constaints

$$\left.\begin{aligned} \sum_{i=1}^{b} \beta_i = 0 \\ \sum_{j=1}^{t} \tau_j = 0 \end{aligned}\right\} \qquad (2.2)$$

where b, t are the numbers of blocks and treatments. Suppose we have 5 treatments A,B,C,D,E and blocks of size 4,4,3,2. Then if we chose the block contents (before randomization of order) as

I :	A	B	C	E
II :	B	C	D	E
III :	A	C	E	
VI :	A	D		

we could estimate the parameters by least squares, determining values to minimize the sum of

$$(y_{ij} - \mu - \beta_i - \tau_j)^2$$

over all 13 observations. This may look a difficult task, but with modern computing facilities it can be done quite rapidly even for much larger experiments. There are two objections. First, the implicit assumption of constant variance, equation I(2.2), may be inappropriate. One might expect σ^2 to depend upon block size, and indeed the reduction in σ^2 as block size is reduced has already been stated as a major reason for using small blocks. This consideration is likely to be more important for plots of land than for litters of rats, and it may matter little for blocks that are nearly the same size (say 6 and 5, but perhaps not 4 and 2). Secondly, the arrangement shown is totally without symmetry; in consequence, variances for treatment means will depend upon which comparisons are being made (the difference between C and E will be much more precisely estimated than that between B and D).

Much better designs can be achieved if all blocks are of the same size. Look for example at arrangements for 3 treatments in 6 blocks of 2 and 5 treatments in 5 blocks of 4:

I	A	B
II	A	B
III	A	C
IV	A	C
V	B	C
VI	B	C

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

If we restrict ourselves to intra-block comparisons, that is to say to contrasts orthogonal with blocks, it is evident that (in the first design), $(\tau_1-\tau_2)$ can be estimated from blocks I and II. But further information is available: $(\tau_1-\tau_3)$ can be estimated from blocks III, IV, $(\tau_2-\tau_3)$ from blocks V, VI, and the difference between these again estimates $(\tau_1-\tau_2)$.

The symmetry of this design (and of the other example)/ ensures that all variances are equal and leads to a // tractable symmetric form of analysis.

3. BALANCED INCOMPLETE BLOCKS

The two examples in Section 2 are very simple e-// xamples of this imoortant class of designs. The defi-/ nition of a balanced incomplete block design is that / there shall be t treatments in b blocks each of k "plots" (the word generally used to mean animal or human subjects, field plots of a crop, or any other ex-/ perimental units) such that

(i) Each treatment occurs on r plots,

(ii) Each pair of treatments occurs in λ of/ the b blocks.

Of course, t, b, k, λ must be integers, as must

$$r = bk/t, \tag{3.1}$$

the number of replicates of each treatment, alternatively expressed as

$$N = bk = rt \tag{3.2}$$

the total number of plots. Since a block contains /// $k(k-1)/2$ pairs of treatments and in all there are /// $t(t-1)/2$ different treatment pairs, it follows that

$$\lambda = \frac{bk(k-1)}{t(t-1)},$$

which with the aid of (3.2) simplifies to

$$\lambda = \frac{r(k-1)}{t-1}, \tag{3.3}$$

so that the expresion on the right must be an integer.

The two simple examples in Section 2 had

$$k=2,\ b=6,\ t=3,\ r=4,\ \lambda=2$$

and

$$k=4,\ b=5,\ t=5,\ r=4,\ \lambda=3$$

respectively. A less trivial illustration is

I	ACEG
II	BCFG
III	ABEF
IV	DEFG
V	ACDF
VI	BCDE
VII	ABDG

which has $k=4$, $b=7$, $t=7$, $r=4$, $\lambda=2$. A design with // $k=4$, $b=20$, $t=16$, $r=5$, $\lambda=1$ can be constructed from the completely orthogonalized 4x4 square at the end of Section IV 7. Write the 16 treatments in random order on / top of the square. Take the rows as defining 4 blocks,/ the columns as defining 4 more, and similarly 4 blocks/ from Latin letters, Greek letters, and numerals.

For any t, k, one can always form a balanced incomplete block design by taking all possible selections of k from t , but often this gives an intolerably large/ numbers of blocks. Unfortunately, although the condi-/ tions that r, λ in (3.1), (3.3) are *necessary* for the / existence of a BIB design, they are not *sufficient*. / For example, $k=5$, $b=21$, $t=15$, $r=7$, $\lambda=2$ satisfy the conditions but no design corresponding to these exists. // Except when t and r are small, investigation of the existence of designs is difficult. Extensive catalogues are available; there are also various rules for cons--/ tructing particular subsets of designs. Among the ra--/ ther few general theorems known are:

(i) For any design, $r>k$ (and therefore $b>t$); hence
$k=6$, $b=8$, $t=16$, $r=3$ $\lambda=1$ is impossible;

(ii) If $t=b$ (and therefore $r=k$) and t is an / even integer, $(r-\lambda)$ must be a perfect / square; hence
$t=b=22$, $r=k=7$, $\lambda=2$ is impossible.

Simple and elegant mathematical proofs of these // theorems have been obtained, and I have reproduced them/ elsewhere (Finney, 1960, Chapter 6); I do not propose to include proofs in my lectures, but shall gladly provide/ them if asked.

4. STATISTICAL ANALYSIS

I shall not describe every detail of the statistical analysis of BIB designs, as this is better studied in / text books. However some account of a particular example with a few more general formulae, should help you to // read books easily; it is also instructive in relation to other types of design.

Table 4.1 shows subjective assessments of pain re-// corder by 30 subjects, each of whom received 3 doses of/ penicillin by injection at 3 sites. Pain was scored on a scale from 0 (none) to 4 (severe), separately at each // site. The allocation of doses to subjects was in accor-/ dance with a BIB design for which

$$k=3,\ b=30,\ t=6,\ r=15,\ \lambda=6.$$

The very small values of y may cast serious doubts on/ any assumptions of Normality, but the simple arithmetic/ is helpful in an illustrative example.

As usual, we assert

$$y_{ij} = \mu + \beta_i + \tau_j + \varepsilon_{ij} . \qquad (4.1)$$

Table 4.2 contains various preliminary calcula-- tions. The T_j are the treatment totals, transcribed from Table 4.1. The B_j are the totals of all blocks (subjects) that had treatment j ; for example, B_5 is the total for subjects I-V, X-XV, XX-XXV, each of which had treatment E. Look now at the quantities Q_j defined in general by

$$kQ_j = kT_j - B_j \qquad (4.2)$$

(Do not confuse the use of Q here with that in Lecture III). By careful examination of how the different values of (4.1) contribute to the T_j, B_j, and consequently Q_j, and remembering that

$$\sum_1^t \tau_j = 0 , \qquad (4.3)$$

we can show that, for example,

$$kQ_1-kQ_2 = [k\lambda + (k-1)(r-\lambda)](\tau_1-\tau_2)$$
$$\pm\ k\varepsilon \text{ from each of } 2\lambda \text{ plots}$$
$$\pm\ (k-1)\varepsilon \text{ from each of } 2(r-\lambda) \text{ plots}$$
$$\pm\ \varepsilon \text{ from each of } 2(k-1)(r-\lambda) \text{ plots} \qquad (4.4)$$

This follows because $(2r-\lambda)$ blocks contain plots that contribute to $(kQ_1 - kQ_2)$, of which blocks $(r-\lambda)$ contain treatment 1 but not 2, $(r-\lambda)$ contain 2 but not 1, and λ contain both treatments. The symbol ε re-represents any one of the ε_{ij} ; careful counting shows that 2λ of the ε_{ij} occur with multiplier k or $-k$ from the plots of the two treatments in the blocks where both occur, and so on. With the aid of (3.3), equation (4.4) leads to the expectation or mean value

$$E(kQ_1 - kQ_2) = t\lambda(\tau_1-\tau_2) . \qquad (4.5)$$

Treatments enter the design symmetrically, and the fact that differences between Q_j are independent of the β_i shows that contrasts among the Q_j are orthogonal with block contrasts. Moreover, from (4.4), the variance of $(kQ_1 - kQ_2)$ is

$$V(kQ_1-kQ_2) = \sigma^2[2\lambda k^2 + 2(r-\lambda)(k-1)^2 + 2(r-\lambda)(k-1)]$$
$$= 2\lambda kt\sigma^2 \qquad (4.6).$$

by further use of (3.1), (3.3). Because $(kQ_1 - kQ_2)^2/2$ would be the square for one contrast in the sum of squares of deviations $\sum_{j=1}^{t} [(kQ_j)^2]$,

it follows that, in units of a single plot (note that $\Sigma kQ_j = 0$),

$$\frac{\Sigma[(kQ_j)^2]}{\lambda kt} \qquad (4.7)$$

is a sum of squares with (t-1) df for insertion in the analysis of variance after inclusion of a simple sum of squares for blocks. Table 4.3 shows this analysis.

From (4.5), τ_j is estimated by $kQ_j/t\lambda$, and this can be added to the general mean of y (2.11 in the example), to give means appropriate to each treatment. From (4.6), for comparisons among treatment mean we can attribute a variance.

$$\frac{k\sigma^2}{t\lambda} \qquad (4.8)$$

to each mean. Using $s^2 = 0.4741$ from Table (4.3) as an estimate of σ^2, one obtains the SE 0.20, shown with the means in Table 4.4. Obviously there are very marked differences between E,F and the other 4 treatments.

This is not the end of the story. If subjects have been allocated by random choice to the various sets of three doses, there is additional information on the τ_j from block totals. For example,

$$B_1-B_2 = (r-\lambda)(\tau_1-\tau_2) \pm k\beta \text{ from each of } 2(r-\lambda) \text{ blocks}$$
$$\pm\ \varepsilon \text{ from each of } 2k(r-\lambda) \text{ plots.} \qquad (4.9)$$

But the randomization over blocks ensures that the β_j are random errors applicable to blocks, so that we can write

$$E(\beta) = 0,\ E(\beta^2) = \sigma_B^2 . \qquad (4.10)$$

From (4.9)

$$E(B_1-B_2) = (r-\lambda)(\tau_1-\tau_2) , \qquad (4.11)$$

and

$$V(B_1-B_2) = 2k^2(r-\lambda)\sigma_B^2 + 2k(r-\lambda)\sigma^2 . \qquad (4.12)$$

These equations are analogous to (4.5), (4.6), and they indicate that τ_j can be estimated by

$$(B_j-\bar{B})/(r-\lambda) \qquad (4.13)$$

with variance

$$k(k\sigma_B^2 + \sigma^2)/(r-\lambda) \qquad (4.14)$$

Moreover, consideration of other inter-block contrasts leads to estimation of $(k\sigma_B^2 + \sigma^2)$; the procedure is not obvious, but is presented in detail in my 1960 book and elsewhere, and on request I will explain it. In the penicillin experiment, the outcome is another set of / estimated treatment means with SE 0.74. This SE is so / much greater than the 0.20 in Table 4.4 that, in the // present instance, the inter-block analysis is scarcely worth having, but a weight combination of the two independent sets of estimates leads to a final set of means with SE 0.19.

You may like to consider the consequences of using/ a modified design. Suppose the three sites on each subject were identified as P_1, P_2, P_3, and suppose that // the experimenter suspected that these might differ in / sensitivity to pain. I would have been inclined to re-/ commend using each set of three subjects with the same/ block type (e.g. subjects IV, XIV, XXIV) in a Latin -/ square scheme such as:

	P_1	P_2	P_3
IV	B	D	E
XIV	E	B	D
XXIV	D	E	B

How would I have neede to modify the statistical ana-/ lysis?

5. OTHER INCOMPLETE BLOCK DESIGNS

Balanced incomplete blocks are perhaps the simplest of incomplete block designs, and where they are suita-/ ble they are ideal. Unfortunately many experimental situations impose conditions for which no BIB design -// exits. Fortunately, there are many other useful fami-// lies of designs with lesser symmetry than the BIB but / more than that in Section 2. Partially balanced designs which include the important class of lattice designs, / are possibly the most widely known and used. Some families, such as doubly balanced incomplete blocks, have / been developed to suit special experimental needs, some perhaps more because of a purely mathematical interest/ in their combinatoric properties. All require statistical analysis along similar lines, though the complexity increases when the symmetry is less. In particular, /// equation (4.2) is a standard device for producing estimators of treatment effects orthogonal with blocks. The management of the intra-block and inter-block analyses/ and combination of their estimates can involve very // complicated algebra, but once this is embodied in a // computer program the work can be executed easily and // quickly. In 1984, there is no excuse for choosing a design merely because it is algebraically simple or com-/ putationally familiar: emphasis should always be placed on finding a design best suited to the questions ---// needing to be answered, and doing this efficiently --// within the constraints of available resources.

One development particularly important to the ---/ testing of varieties of agricultural crops has been that of the **α-designs** (Patterson *et al.*, 1978). These were first produced to meet the needs of the British system / of coordinated variety trials (Patterson & Silvey, 1980; Patterson & Hunter, 1983). Rigidly imposed conditions / are that:

(i) The experiments shall have 2, 3 or 4 replicates;
(ii) The designs shall be *resolvable* (this means that the blocks can be grouped into complete replicates, a property not possessed by any of the ear-/ lier examples, although it is possible for some / BIB designs);
(iii) The designs shall have small blocks (preferable / not exceeding k=10), but shall accomodate large / numbers of varieties (t=50 would not be unusual).

Resolvability has two merits, one statistical and one / very practical. If the experiment can be laid out as // complete replicates, each replicate or full block being divided into the incomple e blocks of the design, the / option of analyzing as randomized blocks is created; / this can be helpful as a quick preliminary analysis, / and it ensures that even in the most unexpectedly un-/ favourable circumstances there can be no *loss* of precision relative to randomized blocks. Secondly, a compact single replicate is often valuable for visual inspec- / tion, demonstration to farmers, and so on.

An example of such a design is shown in Table 5.1. / The designs are defined by a method of generation rather than by conditions such as those stated in Section 3 to characterize BIB designs. The family of designs is very large, and it includes many lattice and cyclic designs. However, methods are known that enable a trial design / to be progressively modified until one of high efficiency is inversely related to the variance of comparisons/ between treatments from an intra-block analysis as compared with the variance for randomized blocks if there / were no additional variance among incomplete blocks. It measures the price that would be paid by using the in-// complete blocks when they fail to remove any additional/ variance; in practice, one hopes to use an incomplete // block design only when this loss of efficiency is more / than compensated by reduction in the effective error / mean square. If a BIB design exists for specified k, b,/ t, r, it will have maximal efficiency. However, the // best of the **α-designs** have efficiencies very little/

lees than would corresponding BIB designs if they are / also resolvable.

In the practice of variety trials, one relaxation // of conditions is sometimes acceptable in order to in--// crease efficiency. This is to permit experiments with // some blocks of k plots and some of (k-1). There is /// though to be little fear that the variance within blocks will depend appreciably upon whether the block size is,/ 6 or 7, and this extra freedom may allow efficiency to / approach closer to the maximun. Again extensive catalo-/ gues of designs exist.

TABLE 4.1

Results from a Balanced Incomplete Block Experiment on Pain from Intramuscular Injection of Penicillin.

Subject	Dose						Total
	A	B	C	D	E	F	
I	2		3		1		6
II		3	4		4		11
III	4				1	2	7
IV		1		2	1		4
V				4	1	3	8
VI	2		2	2			6
VII		2	1			1	4
VIII			4	2		1	7
IX	4	3				2	9
X	4	4		4			12
XI	1		2		0		3
XII		1	4		1		6
XIII	3				1	1	5
XIV		4		3	3		10
XV				3	1	1	5
XVI	3		3	3			9
XVII		3	2			1	6
XVIII			2	2		1	5
XIX	3	1				1	5
XX	3	2		2			7
XXI	2		1		1		4
XXII		3	3		1		7
XXIII	3				1	1	5
XXIV		4		2	1		7
XXV				1	1	1	3
XXVI	3		2	1			6
XXVII		3	3			1	7
XXVIII			2	1		1	4
XXIX	2	1				0	3
XXX	4	3		2			9
Total	43	38	38	34	19	18	190

TABLE 4.2
Intermediate Calculations from Table 4.1

Dose	A	B	C	D	E	F	Total
T	43	38	38	34	19	18	190
B	96	107	91	102	91	83	570
3Q=3T-B	33	7	23	0	-34	-29	0

TABLE 4.3
Analysis of Variance for Table 4.1

Variation	df	Sum of squares	Mean square
Adjustment for mean		401.1111	
Blocks (subjects)	29	52.8889	
Treatments (intra-block)	5	33.9259	6.7852
Error	55	26.0741	0.4741
Total	89	112.8889	

TABLE 4.4
Estimated Mean Pain Records for Table 4.1

Dose	A	B	C	D	E	F	
Mean	3.03	2.39	2.68	2.16	1.20	1.28	± 0.20

TABLE 5.1

Blocks of an α-Design for 30 Treatments in Four Replicates and Blocks of Five (treatments identified by numbers)

Replicate 1						Replicate 2					
I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
1	2	3	4	5	6	1	2	3	4	5	6
7	8	9	10	11	12	8	9	10	11	12	7
13	14	15	16	17	18	16	17	18	13	14	15
19	20	21	22	23	24	21	22	23	24	19	20
25	26	27	28	29	30	29	30	25	26	27	28

Replicate 3						Replicate 4					
XIII	XIV	XV	XVI	XVII	XVIII	XIX	XX	XXI	XXII	XXIII	XXIV
1	2	3	4	5	6	1	2	3	4	5	6
12	7	8	9	10	11	11	12	7	8	9	10
15	16	17	18	13	14	18	13	14	15	16	17
22	23	24	19	20	21	20	21	22	23	24	19
26	27	28	29	30	25	27	28	29	30	25	26

For use in the field, the numbers 1 to 30 would be allocated in random order to the actual treatments or varieties. Each replicate would be set out as a superblock of 30 plots. Within each superblock, the 6 // blocks (sets of 5 treatments) would be put in random/ order. Within each block, the order of treatments /// would be randomized.

VI. FACTORIAL DESIGN

1. INTRODUCTION

One of the great advances in experimental design was the realization of the gains to be achieved by having a factorial structure of treatments. That is to say, the/ full set of treatments in an experiment consists of all combinations of "levels" of two or more factors. Suppose that an investigator wishes to compare the growth of animals that receive two alternative diets and also wishes to compare the effects of two environmental temperatures Then a factorial experiment would define 4 treatments as the 2x2 combinations of diet and temperature. If there / were a third factor, perhaps two frequencies of feeding, there would be 8 combinations.

The advantajes of including several factors in one experiment have often been described. Primarily they / are:

(i) Economy of effort and material are achieved because the effects of each factor are estimated independently (orthogonally) within one experiment;

(ii) The basis for inference on any one factor is - / broadened because it is tested (in a balanced - / manner) on a range of combinations of other factors;

(iii) Only in this way can we study interactions be- / tween factors , such as the difference in growth/ rates of animals on the two diets being greater / in cold than in heated houses.

Of necessity, I restrict myself almost entirely to / factors at 2 levels. Everything I say can be generalized to other numbers of levels, say 3 temperatures, 3 diets, and 3 methods of feeding. Naturally greater complica- / tions enter when factors have 3 levels, but all that I shall say about factors at 2 levels will generalize to 3 or 4 or 5. There is of course no reason in principle / why an experiment should not include factors with different numbers of levels.

2. NOTATION

For formal discussion, I shall denote factors by / upper-case letters A, B, C,...; in practice I should try to use a letter suggestive of the factor (D: Diets, T : temperatures). I shall distinguish the two levels of a factor by presence or absence of the corresponding lower case letter. Where a factor is quantitative, say 20^{0} C, 25^{0} C, it is natural (not essential) to use presence of the letter for the higher level, but for a non-quantitative factor this identification can be arbitrary.

If we have 4 factors, A, B, C, D, the symbol ac will denote the treatment with the upper or positive states / of A and C, the lower or negative states of B and D. Similarly b represents the treatment in which only B is at the upper level. It is usual to represent the combina- / tion of lower levels of all factors by (1), or simply 1. Hence the 16 combinations are

1, a, b, ab, c, ac, bc, ..., cd, acd, bcd, abcd.

This will be termed a 2^4 (4 factors at 2 levels) facto-- rial set of treatments.

3. ORTHOGONAL CONTRASTS

Table 3.1 lists across the top the treatment combinations for a 2^3 factorial. Suppose a 2^3 factorial experiment is conducted as 3 randomized blocks of 8 plots.Construct a linear function of the 24 plots using the quantities shown in the line labelled A as the coefficients of individual plot "yields", in agreement with the treat- / ment symbols. This form the difference between the 12 / plots with a and the 12 without a: division of the constrast by 12 is a measure of the effect on the yield variate y of changing the level of factor A. With b, B in place of a, A, the same may be said of the next line. Moreover these two contrasts are easily seen to be orthogonal (III, Section 3). The line labelled C is similarly constructed; the contrast is orthogonal with A and with B.

Look now at the line labelled AB. It can be regarded as the difference between "plots with a -plots without a in the presence of b" and "plots with a - plots without/ a, in the absence of b". It is symmetric,in the sense / that "a" and "b" can be interchanged in this statement . It measures the extent to which the effect of factor A is modified by factor B (and *vice versa*). It is kown as the *interaction* of A and B, written A,B or AB. This contrast is orthogonal with each of A, B, C. Similar - / comments follow for AC and Bc. Finally the line ABC is a contrast for the three factor interaction, measuring the extent to which the AB interaction is modified by factor

C. We now have 7 mutually orthogonal contrasts, and consequently a way of subdividing the treatment sum of squares (7 d.f.) into single squares for the main effects A, B, C, and the interactions.

The analysis of variance can be presented as in Table 3.2. If required, a test of significance can be made on each of the components, but the important features are that the effects of three factors are estimated simultaneously from one experiment, and evidence is acquired on the various interactions without which understanding is incomplete. Note particularly that the effect of A is estimated from 12 replicates of plots with and without a, *and* the effects of B and C are estimated with the same replication from the same plots. Yet only 24 plots have been used in all.

4. CONFOUNDING

A good scientist or technologist is likely to think of many factors that he would like to include in his experiment. The total number of treatment combinations then becomes large (e.g. $2^5 = 32$), exceeding the size of block that is available or that is believed suitable for controlling variance by reducing intra-block variance.

If one or more members of the orthogonal set of treatment contrasts can be nominated as of little interest (perhaps because a multifactor interaction is thought likely to be negligibly small or because a main effect is already so well understood that further information on it is not wanted), we can adopt the device of confounding. For example, a 2^3 experiment can be conducted in blocks of 4 (with randomization of order within each block), by using the two types of block.

(i) 1, ab, ac, bc
(ii) a, b, c, abc.

In the experiment had 6 blocks of 4, 3 blocks of each type, the interaction ABC is *confounded* between blocks : the contrast labelled ABC in Table 3.1 can be estimated only by the difference between all blocks of type (ii) and all of type (i). All other treatment contrasts are still orthogonal with blocks. The analysis of variance takes the form of Table 4.1. The calculations involve nothing new; the 6 treatment contrasts are calculated exactly as before, and the sum of squares for blocks follows the standard rules.

Confounding becomes more important in larger experiments, but the same method can be applied and generalized. For example, a 2^5 experiment could be conducted in blocks of 16 by confounding ABCDE. One block type will consist of all treatment combinations with an even number of letters (1, ad, bcde, etc.), the other of all combinations with an odd number of letters. Alternatively, one may confound any other interaction such as ACD or even the main effect D if that were thought of little interest. A confounding in blocks of 8 can be obtained by simultaneous confounding of two contrasts. For example, ABCDE and ADE might be chosen but then necessarily BC is also confounded. The rule is that if any two contrasts are confounded so also is their "product" where in the product the square of any letter is deleted:

$$ABCDE.ADE = \cancel{A^2}BC\cancel{D^2}\cancel{E^2} = BC$$

If, as is commonly the case, the wish is to restrict confounding to the higher order interactions, a better choice will be ABCD and ACE, which must also confound A^2BC^2DE or BDE.

Of course this rule is not an arbitrary restriction; it is unavoidable. If you divide the combinations of a 2^n factorial into 4 types of block, with 2^{n-2} treatment combinations in each, in such a way that the contrast between types (i), (ii) and types (iii), (iv) confounds one nominated interaction (or main effect), and the contrast between (i), (iii) and (ii), (iv) confounds a second interaction, you will find that the constitution of the block types is uniquely determined; moreover, the contrast between (i), (iv) and (ii), (iii) necessarily confounds the "product" of the two interactions.

The blocks for design are easily constructed. First choose treatment symbols that contain an even number of lower case letters from the letters forming the confounded interactions. For ACE, BDE, ABCD, these are

(i) 1, ac, bd, abcd, abe, bce, ade, cde.
Note that these have the property that the product of any pair (where again any letter occurring twice is deleted) is another one of the set of eight (ab.ace = bce), so that only 3 independent combinations have to be found. These 8 forms blocks of type (i). For (ii), take any treatment not already included, say ab, and multiply all the 8 by it;

(ii) ab, bc, ad, cd, e, ace, bde, abcde.
Similarly, for (iii) multiply members of (i) by,

say, d;

(iii) d, acd, b, abc, abde, bcde, ae, ce.
The remaining 8 combinations form (iv), and can be obtained by multiplication by, say, acde;

(iv) acde, de, abce, be, bcd, abd, c, a.
You may easily verify that three orthogonal con-- trast concisely symbolized by

$$(i)+(ii)-(iii)-(iv)$$
$$(i)-(ii)+(iii)-(iv)$$
$$(i)-(ii)-(iii)+(iv)$$

form the interactions ABCD, ACE, BDE respectively.

If the experiment has several replicates, one may / confound the same set of interactions in each. An alternative is *partial confounding*, where a new set of confounded interactions is used for each replicate; each interaction is then estimated from all replicates in which it is not confounded.

5. SINGLE REPLICATION

If some high order interactions are negligible, the mean squares corresponding to them will have expecta--- tions that scarcely exceed σ^2 . A common practice is to take advantage of this by including as many factors / as possible and using only a single replicate; one way then allocate various interactions to form a sum of squares that should give only a very slightly biased estimate of σ^2 . For example, suppose that an experimenter were contemplating using two replicates of the design/ for 2^5 in blocks of 8 in Section 4. He might do better / do add F, an extra factor (almost certainly he has factors in mind that he would like to include), and use a single replicate confounding perhaps

ACE, BDE, ABCD, ADF, CDEF, ABEF, BCF.

The computations follow the standard pattern. For the / error sum of squares, 7 d. f. could be obtained from ABCDEF and the 6 five-factor interactions (ABCDE, etc.); a further 12 d.f. from unconfounded four-factor interac--- tions might be added to these. Thus the analysis of va-- riance would have 7 d.f. for blocks, 7 or 19 d.f. for/ error, and the remaining 49 or 37 d.f. for individual / main effects and interactions.

Thus with very little disadvantage an extra factor / and its potentially interesting interactions with the first five have been added to the information from the / experiment.

If originally 4 replicates of of the 2^5 had been intented, factors F and G might have been added so as to give a single replicate of 2^7.

6. FRACTIONAL REPLICATION

This idea can be carried further. If a high order interaction is negligible, not only might it be used for / estimating σ^2 but also no harm will come from confu--- sing it with another more interesting main effect or interaction. As a trivial case, consider a 2^4 factorial / structure in which results, y, are available for only 8 plots; they are

1	ab	ac	ad	bc	bd	cd	abcd
y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8

Then the contrast

$$-y_1+ y_2 + y_3 + y_4 - y_5 - y_6 - y_7 + y_8$$

estimates A (to be multiplied by 1/4). But if we attempt to estimate BCD we find that we need exactly the same contrast. We can write

$$A = BCD$$

Similarly

$$B = ACD$$

and other relations are obtained. All of these can be expressed symbolically by

$$ABCD = 1$$

with the understanding that again the product rule can be used, this time to identify *aliases*. Thus D.ABCD= ABC hence

$$D = ABC$$

and D is *aliased* with the interaction ABC. This means / that ABC cannot be estimated distinctly from D,a serious objection to the fractional replication of 2^4; still / worse are aliases such as

$$AD = BC$$

However 2^4 merely illustrates the method. With more factors, the situation is different. For 7 factors and

$$ABCDEFG = 1$$

typical aliases are

$$B = ACDEFG$$
$$DG = ABCEF$$
$$ABE = CDFG$$

all of which may be tolerated, in the sense that the resulting ambiguities of interpretation may be unimportant

When the number of factors is large, 1/4 or 1/8 replication may be practicable. Also fractionally replicated experiments can be confounded, but I must omit details.

7. FACTORS AT THREE LEVELS

Often it es desirable to have three levels for each factor - three temperatures, three diets, and so on. The notation and methods generalize. Whereas the special algebra of the 2^n designs is based upon

$$A^2 = B^2 = \cdots\cdots = 1$$

and

$$a^2 = b^2 = \ldots\ldots = 1$$

in the rules of multiplication, for 3^n designs we use

$$A^3 = B^3 = \ldots\ldots = 1$$
$$a^3 = b^3 = \ldots\ldots = 1$$

For example, the 9 treatment combinations for 3^2 are

$$1,\ a,\ a^2,\ b,\ ab,\ a^2b,\ b^2,\ ab^2,\ a^2b^2$$

where a, a^2 mean the middle and the upper levels of factor A with the lowest of B, a^2b is the combination of the upper level of A with the middle level of B, and so on. Note that the product of any two is also one these :

$$a^2b.a^2b^2 = a^4b^3 = a$$

The groups of treatments

(i) $1,\ a^2b,\ ab^2$
(ii) $a,\ b,\ a^2b^2$
(iii) $a^2,\ ab,\ b^2$

will confound 2 d.f. from the interaction that has in all 4 d.f. These d.f. are symbolized by AB, A^2B^2. The rules of orthogonality are better illustrated by the confounding of 3^3 in blocks of 9. We can choose a pair of d.f. from the ABC interaction (8 d.f. in all) such as AB^2C, A^2BC^2; note that each is the "square" of the other, since

$$(AB^2C)^2 = A^2B^4C^2 = A^2BC^2$$

To confound these, choose all elements of the form $a^{\alpha}b^{\beta}c^{\gamma}$ such that the sum of products of the indice of the confounded effects with the corresponding indices α, β, γ is a multiple of 3. More concisely

$$\alpha + 2\beta + \gamma \equiv 0 \pmod 3$$

and

$$2\alpha + \beta + 2\gamma \equiv 0 \pmod 3$$

The second condition is equivalent to the first. The elements are easily found as

(i) 1, ab, a^2b^2, ac^2, a^2c, a^2bc^2, ab^2c, bc, b^2c^2 and these constitute block type (i). For this principal block, the square of any element and the product of any two elements also belong to the set. As for 2^n designs, form another block type by multiplying each element of (i) by any new element, say c:

(ii) c, abc, a^2b^2c, a, a_2c_2, a^2b, ab^2c^2, bc^2, b^2, and then form type (iii) by one more multiplication, say by $a^2b^2c^2$:

(iii) $a^2b^2c^2$, c^2, abc^2, b^2c, ab^2, ac, b, a^2, a^2bc. This , and three similar designs confounding ABC^2, A^2B^2C or AB^2C^2, A^2BC or ABC, $A^2B^2C^2$, are of great practical value.

Another valuable set of designs confounds 3^4 in 9 blocks of 9, confounding for example

$$ABC,\ A^2B^2C^2,\ A^2BD,\ AB^2D^2,\ B^2CD,\ BC^2D^2,\ AC^2D\ ,\ A^2CD^2$$

You may like the exercise of constructing some of the 9 block types. There is indeed a useful fractional replicate design or 1/3 of 3^5 in blocks of 9. I shall not have time to discuss these in detail, but I will gladly talk further about them if asked.

8. A GENERAL THEOREM

An important general theorem is that, for any prime number π .

A π^n experiment can be arranged in π^{n-p} blocks of n^p plots each, without confounding either main effects or 2-factor interactions, if and only if

$$n \le (n^p-1)/(n-1)$$

An equivalent theorem relates to fractional replication I shall not prove this unless asked.

9. MIXED LEVELS

Factorial designs can have factors at different numbers of levels, such as 2^2x3, 2x3^3, or 2x3x4. Sometimes these are essential to the desired character of experiments. If they can be conducted in randomized complete blocks, they present no difficulty. Confounding, however, is more complicated than before, and usually results in partial loss of information on several contrasts. Examples can be found in the textbooks I have mentioned, but I unlikely to have time to discuss them in lectures. If all factors have either 2 or 4 levels, it may be practicable to regard each 4-level factor as a pair of factors at 2 levels in order to devise a confounding arrangement. Thus a 2^3x4^2 design might be confounded as though it was 2^7. This requires considerable care, as neglect of the true nature of the factors may prove disastrous.

TABLE 3.1

Orthogonal Contrasts for a 2^3 Experiment

Contrasts	Treatments 1	a	b	ab	c	ac	bc	abc
A	-1	+1	-1	+1	-1	+1	-1	+1
B	-1	-1	+1	+1	-1	-1	+1	+1
AB	+1	-1	-1	+1	+1	-1	-1	+1
C	-1	-1	-1	-1	+1	+1	+1	+1
AC	+1	-1	+1	-1	-1	+1	-1	+1
BC	+1	+1	-1	-1	-1	-1	+1	+1
ABC	-1	+1	+1	-1	+1	-1	-1	+1

TABLE 3.2

Form of Analysis of Variance for 3 Randomized Blocks of a 2^3 Factorial

Variation	d.f.	Sum of squares	Mean square
Blocks	2		
A	1		
B	1		
AB	1		
C	1		
AC	1		
BC	1		
ABC	1		
Treatments	9		
Error	14		
Total	23		

TABLE 4.1

Table 3.2 Modified for Confounding of ABC

Variation	d.f.	Sum of squares	Mean square
Blocks	5		
A	1		
B	1		
AB	1		
C	1		
AC	1		
BC	1		
Error	12		
Total	23		

VII. FURTHER DEVICES IN DESIGN

1. SPLIT PLOTS

Sometimes an experiment uses two (or more) different sizes of unit for the application of treatments. More particularly, levels of one factor may be allocated to plots, and levels of a second factor may be allocated to subdivisions of each plot (or *subplots*). For example dietary differences can be studied only on whole ani--- mals but skin reactions to an inoculation can be measured at several points on each body. Therefore an experiment on the extent to which skin reactions are modified by diet might use rats as plots in a randomized block / design for diets, with 4 types of inoculation on posi-- tions as subplots. In agricultural trials, irrigation / may have to be controlled over large areas, so that a comparison of irrigation rates must use large plots, / but varietal comparisons can be put on subplots of the same experiment.

The analysis of variance presents no difficulty, but it has one important new feature: there are two / distinct error lines corresponding to interplot and / intraplot variation. Table 1.1 illustrates this for an experiment on 5 diets tested on 6 litters of 5 rats, / with 4 inoculation treatments tested on each rat. The arithmetic should always be done in terms of the / smallest experimental units, the subplots, so that (in accordance with Lecture III) the sum of squares for / main plots (29 d.f.) is found as

$$\frac{\Sigma(\text{mainplot total})^2}{4} - \frac{(\text{grand total})^2}{120} .$$

Other parts of the analysis are calculated in the usual way, and take their places in Table 1.1. The symmetry should make clear that all the components are orthogo-- nal. Of curse means must be compared with the aid of of the appropiate error mean square.

There can be other reasons for using split plots. / For example, if an experiment continues for a long time the experimenter may wish to modify it by adding a factor; one way of doing this is to put the new factor on subplots, but the alternative of confounding may need / examination. Again, occasionally an experimenter wishes to achieve relatively greater precision on one factor / than on another: he can make use of the fact that / subplot variance is usually smaller (sometimes very / much smaller) than the main plot variance.

The split plot principle can be extended so as to / have more levels of splitting, giving split-split plots or split-split-split plots! This is seldom a desirable feature of design, but it can useful. The great mistake is to regard the splitting of plots as an easy way of fitting factors into an experiment, to be adopted / without thought of other types of confounding. Note that these designs can be alternatively described in / terms of confounding. For the experiment in Table 1.1,/ we could speak of diets as being confounded between rats whereas inoculations and the interaction are unconfounded.

2. REPEATED MEASUREMENTS

In other circumstances, each main unit of the experiment may be measured several times for rather diffe--- rent reasons. One situation is that a particular property must be studied by sampling. In a field experiment / on a ceral crop, the total weight of grain for each / plot will be measured. However, nitrogen content of the grain may be studied by analyzing several small subsamples from each plot; if interest lies in plant size or insect damage, subsamples of individual plants may be / have been measured or recorded within each plot. / Although there is an analogy with split plots, no treatments or other structure is imposed on the sampling / units. So far as the experiment is concerned, any interpretation of treatment effects on nitrogen content or / insect damage will be based upon plot means for the sampling. A complete analysis of variance, such as / Table 2.1, again calculated in terms of the smallest / units, is convenient, but only the plot analysis is relevant to assessment of treatment effects. The magnitude of the sampling error is useful only for indicating whether sampling was intensive enough so that variance/ from this source makes only a small contribution to the plot variance; it can help to indicate whether in a future similar experiment 3 o 6 sampling units should be taken from each plot instead of 4.

Similar sampling considerations arise in many / circumstances. In clinical medical studies, replicate / analyses of blood samples may be made for biochemical / or hormonal determinations. If alternative methods of manufacture of an electronic component are to be compared, the plot may be a manufacturing batch, from each of which a few sample components are selected for / measurement of quality or durability. In all these circumstances, to use as the error variance a mean /

square with 110 d.f in Table 2.1 or its equivalent / would be totally wrong, since this would be a composite of two variances that might be very unequal. This is a mistake that is rarely made, though I have seen it.

A much more common mistake, of the same kind and / equally serious, occurs where measurements on a plot / repeated in time. I shall illustrate by reference to a simplified account of an experiment on a marine snail / (Finney, 1978b, 1982c). The aim was to see whether modifying the number of its predators affected its numbers. The aim was to see whether modifying the number of its / predators affected its numbers. Six plots were marked in the intertidal zone of a beach on the Pacific coast of / the USA. Two plots, selected at random, were untouched ; from two more the members of a predatory species were re moved; on the remaining two, predators were added to tho se naturally there. Subsequently, the number of snails / on each plot was counted about once every two months for a year. (The counts were in reality on small sample /// areas, but the sampling is a distinct issue from the pre sent discussion and can be ignored). How should the 36 / "data", 6 counts on the 6 plots be analuzed? The temptation, totally wrong, is to extract 2 d.f for treatments/ and to attribute to treatment means standard errors ba-- sed upon a mean square from the remaining 33 d.f. Treatments have been applied to whole plots, and therefore on ly variability *between* whole plots is relevant to the question of whether treatments have affected snail num-- bers. We can indeed make use of the separate counts on each plot, as is shown in the analysis in Table 2.2.

First we can look at mean or total counts per plot. / Then, independently, we can look at the linear regres--- sions of counts on time, which could be at least as rele vant to the effects of the initial interference with the plots as are the means. A further analysis can be made of another contrast representing a quadratic component of trend. The remaining degrees of freedom, shown as / grouped together in Table 2.2, can be similarly subdivided for further components of trend. Other ways of / arranging the analysis can be based upon alternative / subdivisions of the degrees of freedom, but one princi-- ple must be maintained. Whatever mean or contrats among the 6 dates of counting is to be discussed, the error / variance must be obtained from the variability in this quantity among replicate plots (here the pairs of identi cally treated plots).

Unfortunately, many scientists make the mistake of / assuming that an analysis of such an experiment can be condensed into 2 d.f. for Treatments, 5 d.f. for Dates, 10 d.f. for the interaction TD, and the remainder for an/ error that is supposed homogeneous. In reality the errors appropriate to different contrats may differ widely. I would expect to find the error mean squares in Table 2.2 decreasing steadily as I read down the table (though with so few degrees of freedom the pattern may be irregular) . A totally false sense of security in conclusions can come from a misunderstanding of the variance structure of an experiment. Evidently this experiment is inadequately / replicated unless the effects of treating the plots are very large.

The actual experiment on which this discussions is / based was more complicated (Finney, 1982c). Counts were also made on all plots during the year before the treat-- ments were applied. For practical reasons, no more than 2 plots could ever be counted on one day and intervals were erratic. It was not practicable to have more than 6 plots The temptation to use the apparent replication of individual counts was great, but is not made more legitimate by the additional complications. Not surprisingly, the experiment gave no convincing evidence of any effect of / treatments.

3. CROSS-OVER DESINGS

An important experimental device, especially with / human or animal subjects, is that of changing treatments once or more during the course of an experiment. By intro ducing a suitable balance, average differences between / subjects can be eliminated and the variance relevant to treatment comparisons can become entirely intra-subject . Table 3.1 shows (a) the simplest form for 2 treatments / and (b) an elaboration of this. With the first desing, / simple totalling for A and for B gives a comparison that is balanced over subjects and over periods. The second desing needs more thorough discussion than I have time for, but even a cursory inspection shows the possibi lity of examining residual as well as direct effects: we can compare A after A with A after B, B after A with B after B.

Table 3.2 shows a more complicated cross-over (or change-over) desing for 4 treatments. Note that in each of periods 2, 3, 4, the three subjects on any one treat-- ment have received the other 3 treatments in the previous period, so that again we can estimate not only the / effects of treatments currently given but also residual / effects from previous period. Table 3.2 can be used in several variants. The experiment might be run with periods 1 and 2 alone, though 12 subjects would then / scarcely suffice and one would hope to repeat everything on a second set of 12. The allocation of treatments to / subjects then has a balanced incomplete block structure; this is also true if the experiment is run with periods 1, 2, 3, alone. Only if all 4 periods are included do we have full balance over subjects, with no partial / confounding of treatments between subjects. If time / permits, there can be advantages in having a fifth period that repeats the treatments of period 4, so that estimates are also available relating to A after A, B, after B, etc.

I shall not discuss the statistical analysis. It must again use least squares for estimation and develop an / analysis of variance in terms of appropriate orthogonal contrats.

4. SERIES OF EXPERIMENTS

In some branches of science and technology, an experi mental programme may need to comprise a series of unit experiments. A possibility is to repeat one standard / experiment, of fairly simple design, at many places or in successive years. This can be useful when the number of treatment combinations is small. Such a set of experi ments will allow treatment effects to be averaged over a range of conditions, as well as giving information on / the extent to which effects vary from one site to / another. This approach has been used for clinical compa- risons between drugs conducted in several hospitals or clinics, and for drug standardization in a set of labora tories.

In agricultural research, factorial design encourages interest in much larger numbers of treatment combina---- tions. However, if it is desired to study fertilizer / responses over a region by trials at a sample of widely distributed sites, a block of a suitable confounding / scheme might be allocated to each site; thus a group of sites might constitute one replicate. A comprenhensive / analysis of results would enable average unconfounded / effects to be estimated for the region; in addition, / there is the possibility of looking for any differences between sub-regions in the magnitudes of effects as well as obtaining some information on the confounded interac- tions. Each site can be regarded as a fractional replica te, though there may not be enough plots at a single si- te for the information from it to be much use alone. The 3^n confounded designs are especially suitable. Other / arrangements are possible, in which the treatment combi- nations at a site are not determined by blocks of a / confounding scheme. These can have practical advantages, but they are for specialized agricultural purposes and I shall not discuss them further.

When series of experiments are planned, one must be / clear whether the aim is to estimate effects under diffe- rent conditions or to use a sample of environments for / estimating average effects applicable to all environments For example, one might wish to estimate parameters corres ponding to dietary differences for an animal species, and to do so under various climatic conditions or for various ages and types of animal. Similar information may be / wanted on each category of animal, and pattern among cate gories may emerge, but the experiments are to be seen as leading to a broad understanding. On the other hand, one might wish to estimate effects of diet averaged over con- ditions and categories as the basis of advisory policy / applicable to all; ideally, the conditions, categories, / or sites should be a random selection from all available. Experiments repeated over years will usually have this se cond emphasis, since annual differences have no characte- ristics identifiable in advance. Repetitions over places or other contemporary subgroups may have either objective

The distinction is particularly clear and important in agricultural research. Are experiments on varieties of a crop or on amounts of fertilizer, conducted at many sites and for several years, intended for estimating what is / best separately for many places or sub-regions, though / necessarily averaging over years since seasonal condi--- tions cannot be predicted in advance? Or are they inten-- ded also to average over experimental sites so as to lead to a policy for the whole region studied? In the latter / circumstances, estimation of the relevant parameters must take account of treatmentxsite and treatmentxyear interac tions as components of error additional to the intra-expe rimental variance. These interactions measure the consis- tency of effects over sites and over years, and are there fore very relevant to the advisory policy. In designing / such series of experiments, the replication over sites / and over years must be carefully decided, since the requi red measure of precision may be far more dependent on the interactions than on replication within each site.

5. RESPONSE SURFACE DESIGNS

When all the factors for an experiment are measurable on continuous scales (weights, lengths, concentrations,/ times, etc.), a different approach to design is possible For two factors (generalization to more is obvious), the combinations of levels can be represented by coordinates (x_1, x_2). Then the expectation of y for fixed x_1, x_2 / can be expressed as

$$Y = F(x_1, x_2) \ , \tag{5.1}$$

where F( ) is a continuous function. Within a range of values of x_1 and x_2, this function should be adequately approximated by a polynomial. For example, we might try

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{12} x_1 x_2 + \beta_{22} x_2^2 \quad . \qquad (5.2)$$

Design then consists in specificacion of a set of coordinates (x_1, x_2) in such a way as to optimize estimation of the β parameters. The computations are essentially those of multiple regression, but a well balanced design may greatly simplify the structure of the $X'X$ matrix. The choice of design must be constrained by limitations on the ranges of x_1 and x_2 that can be used, for equation (5.2) cannot be valid or even a good approximation over unrestricted ranges.

These response surface designs have been found particularly useful in chemical engineering and other industrial contexts, where emphasis is on finding the combination of levels that will produce the highest yield. A guess at the optimal factor levels can be taken as the origin of coordinates, 0, for all factors, so that the design is centred on this point. The designs seem to have found less favour in pure and applied biological sciences, despite arguments that they avoid use of the extreme combinations that a conventional factorial will require. The reason may be in part that they seem more difficult to interpret in terms of effects of single factors; perhaps more importantly, non-quantitative factors are often wanted in biological experiments, and response-surface designs are less appropriate for a mixture of quantitative and qualitative factors. Cochran & Cox (1957) have a good elementary account of this topic.

6. SEQUENTIAL EXPERIMENTATION

In some types of experiment, individual "plots" or subjects present themselves in a temporal sequence, and in some of these the result for each subject may be available before the next is treated. Some clinical trials are of this kind, the effect of treatment being rapid relative to the rate at which new patients are encountered.

It is then possible to use ideas adapted from the theory of sequential sampling. The duration of the experiment, and perhaps even also the determination of treatment for each subject, can be made dependent upon all past results. Thus replication can be intensified for the apparently more successful and more interesting treatments, and the experiment can end as soon as the evidence suffices to demonstrate treatment differences. In practice, the methods are usually limited to comparisons of two treatments.

I am not familiar with this type of experimentation and do not propose to discuss it.

TABLE 1.1

Form of Analysis of Variance for a Split Plot Experiment

Variation	d.f.	Sum of squares	Mean square
Litters	5		
Diets (D)	4		
Error (main plot)	20		
Rats	29		
Inoculations (I)	3		
D.I	12		
Error (sub plot)	75		
Total	119		

TABLE 2.1

Form of Analysis of Variance for a Randomized Block Experiment on 5 Treatments in 6 Blocks, with 4 Sampling Units per Plot

Variation	d.f.	Sum of squares	Mean square
Blocks	5		
Treatments	4		
Error	20		
Plots	29		
Sampling error	90		
Total	119		

TABLE 2.2

Form of Analysis of Variance for an Experiment on a Marine Snail

Variation	d.f.	Sum of squares	Mean square
Mean counts			
Treatments	2		
Error (1)	3		
	5		
Linear trend			
Average	1		
Treatments	2		
Error (2)	3		
	6		
Quadratic trend			
Average	1		
Treatments	2		
Error (3)	3		
	6		
Other trend components			
Average	3		
Treatments	6		
Error (4-6)	9		
	18		

TABLE 3.1

Cross-over Designs for 2 Treatments

(a)

	Period	
Subject	1	2
I	A	B
II	B	A

(b)

		Period	
Subject	1	2	3
I	A	A	B
II	B	B	A
III	A	B	B
IV	B	A	A

For practical use, what is shown above would be repeated several times; that is to say, each treatment sequence would be assigned to several / subjects.

TABLE 3.2

Cross-over Design for 4 Treatments

		Period		
Subject	1	2	3	4
I	A	C	B	D
II	D	B	C	A
III	B	D	A	C
IV	C	A	D	B
V	D	C	A	B
VI	C	D	B	A
VII	A	B	D	C
VIII	B	A	C	D
IX	B	C	D	A
X	D	A	B	C
XI	A	D	C	B
XII	C	B	A	D

VIII. CLINICAL TRIALS; BIOLOGICAL ASSAY

1. ETHICS OF EXPERIMENTATION

Research in clinica medicine inevitably introduces questions of ethical behaviour by doctors and others. / To what extend is it ethically justifiable to give to a human being a drug (or other treatment) that may damage his health at the same time as benefitting it, or that/ may even produce damage without benefit? Under what -// circumstances (if any) is it permissible for a treat-// ment to be tested in some persons, with a risk of harm, in order that others may benefit? It is not my place to lecture on ethics, but the statistician cannot totally/ evade the issue.

The statement is often made that no new drug should be released for human use until it is known to be beneficial and without harmful side effects. Unfortunately/ no amount of laboratory testing or of trial in other animal species can guarantee good effects in man. Of /// course, experience of biology and pharmacology will indicate types of non-human study that commonly predict/ human response. Nevertheless, whatever the existing information and however good the intention, the first introduction into human subjects is experimental. If the/ doctrine of no human use without known safety is ---// accepted, no new drugs will be introduced scarcely an / acceptable situation. Surely the conclusion must be / that the necessarily experimental first human uses // shall be planned as *good* experiments. For this reason,/ statisticians have maintained that the critical time / for clinical experimentation is when there is confidence that a new drug is not seriously harmful but complete uncertainly as to whether or not it is an improve--/ ment on the drug it might replace.

If there is to be a formal planned experiment, it / must have a design that will use the available subjects and the information obtained from them as effectively / as possible. An ill-planned experiment is always unethical. Moreover, the design has to operate under the two/ constraints that:

(i) The experiment must stop if the evidence is - / clear that one treatment is superior to the - / others;

(ii) An individual patient will be withdrawn if his / physician is convinced that the treatment allo-/ cated to him is doing more harm that good.Course these remark of apply to experiments that /// must be conduced on sick persons whose chances / of recovery are good. Ethical considerations - / will be somewhat different for treatment of mi-/ nor ailments such as headaches, nausea or brui-/ ses. With treatments that may reduce pain or --/ even offer some hope of cure for a terminal illness, there may be greater willingness to risk / quite serious side effects. Questions of patient consent, volunteer subjects, and the like are -/ perhaps not very relevant to statistical matters, though the statistician must remember the possibility that patients (or healthy persons) -- // willing to participate in an experiment may not / be typical of the general population.

In recent years, the ethics of animal experimenta-/ tion have come under closer examination. I do not think that this is the place to discuss the legitimacy of - / using animals (and causing them to suffer) for the be-/ nefit of man; I believe that any such discussion should distinguish between experiments aimed at benefits to -/ human medicine and health, those for testing cosmetics/ or tobacco or other inessential human pleasures, those/ purely for the increase of knowlegde, and so on. I hope we may all again agree that, if there is to be an experiment, it should be planned so as to use resources and materials effectively with minimal distress to the animals. In particular, I believe it wrong to use more a-/ nimals than are needed for the precision appropriate to the experimenter's aims, and also wrong to use so few / animals that imprecision makes all results worthless.

If animal experiments call for ethical consideration what about insects, and plants?

Whatever the nature of experimentation, the statistician has a right to be informed on any matters that/ may to him introduce some ethical conflict. He has a // conscience no less important than that of others in a / research team.

2. CLINICAL EXPERIMENTS

In their combinatorial structure, clinical experi-/ ments are usually very simple. Often only two treatments are compared. This is because of the very substancial // organizational problems. In many instances, most of the/ mangement, measurement, and recording may have to be undertaken by people whose primary commitment is not to -/ research but to patient care, and the latter must take / priority.

Nevertheless, immense improvements in clinical // experimentation have been achieved in the last 30 -// years. Especially important is the acceptance of randomization. This should not encounter ethical diffi-/ culties. If an experiment is conducted when no treatment is known to be the best, there should be no ob-/ jection to random allocation of treatments to pa-- // tients; if there are strong reasons to believe that / drug B is better than drug A, the ethical issue is // not randomization but the experiment itself. The // process of randomization must be protected against // any adjustment or manipulation that may cause bias! / Ideally the order of events is that a patient is -- / accepted as suitable for inclusion and only then is a sealed envelope opened to tell which treatment he is/ to receive. Also vitally important is the recognition that some from of *control* treatment must be included/ for comparison with new treatments. According to circumstances, a control may be absence of any positive/ treatment, or a placebo beleived to have no true - / effect, or the existing standard treatment for a di-/ sease.

The danger that assessment of results may be -// affected by the patient, and by nurses or physicians/ who examine nim, is well-known. In order to avoid any subjective influences, it is desirable that neither / the patient nor those who care for him or who assess/ results should know which treatment he has had. This/ may be practicable where the treatments are similar / drugs, but is not in a comparison of chemotherapy - / with surgery. Of course there must be provision for / breaking the code in an emergency.

When treatments must be compared in patients suffering from a disease, subjects suitable for inclu- / sion are likely to be identified over a period of time. Randomized blocks can still be used. Blocks may / be defined in terms of sex, age, physiological characteristics, medical history, etc., and randomizations/ performed in advance. The first female aged 20-30 is/ given to the treatment first in randomized order for/ a block so defined, and so with subsequent patients./ There may be a number of blocks of "females, 20-30",/ or all such patients may be regarded as one block - / with many replicates.

If treatments are such that patients on different treatments will be in the care of different physi--// cians or different nursing staffs, care must be given/ to standardization of conditions; even then, the - // possibility must be kept in mind that apparent diffe-// rences in results are in part subjective or psychogenic in origin. In order to obtain more patients, collaboration among several hospitals may be arranged. Each hospital should be seen as having its own small experiment with complete blocks, and at all costs confounding of/ treatment with hospital differences must be avoided.

Simple factorial experiments have been used, and / probably ought to be used more often, but the number of treatment combinations is a limitation. I recall one 2^3 experiment, I think on the management of diabetics,that involved the collaboration of about 10 hospitals in the USA each of which studied equal numbers of patients on/ each of the 8 treatments. The experiment had to conti-/ nue over several years, and of course many different / measurements and records were made on each patient. // Inevitably there were losses by death and other causes. The results were controversial. largely because diffe-/ rent hospitals were inconsistent in their evidence. /// Despite very careful planning, possibly instructions on treatments were not always interpreted the same way or/ measuring processes were inadequately standardized. /// Nevertheless, the work was valuable: had the experiment been confined to one hospital, not only would replica-/ tion have been much less but treatmentxhospital inter-/ actions would have been unsuspected.

Sequential designs have been strongly advocated for clinical research, because they permit a decision and / ending of the experiment as soon as treatment is demonstrated superior to an alternative. They have been used,/ effectively, but the ethical attractions are perhaps /// less great than first appears. Various arbitrary deci-// sions must be taken at the start in order to define the/ rules for termination. Continous watching of results may introduce greater ethical uncertainties about whether // the experiment should terminate even earlier. A good experiment should give much more than a conclusion "B is / better than A", such as information on "how much better" and information on many aspects of the health of pa- // tients. Moreover, sequential design is practicable only/ when the time between administration of treatment of result is short relative to that between successive pa-// tient arrivals. An alternative procedure that has been / suggested is to modify treatment allocation as an expe-/ riment continues. Initially treatments A, B can be allocated to successive patients with equal probabilities; / if early results point towards a superiority for B, the/ probability for B can be increased. The object now is to research the end of the experiment with an adequately //

precise comparison between A and B but with a minimun / number of treatments having been allocated to the /// poorer treatment.

If patients require to be treated over a long pe- / riod but the effect of any one dose is short and ra-/ pidly cleared from the system, cross-over designs are / valuable. Drugs for relief of long-continuing or fre-// quently recurring pain (e.g. rheumatism, migraine) can/ be compared in this ways; the measure of succes then is likely to be a subjective judgement by the patient, and maintaining his ignorance of whether he is currently // receiving a placebo, a familiar remedy, or a new drug / will be important.

3. BIOLOGICAL ASSAY

This large subject is concerned with laboratory /// standardization of drugs, and introduces some special// problems of experimental design (Finney, 1978a). I can/ outline only briefly. Suppose that an animal receiving/ a dose z of a *standard* drug, S, produces a measurable response y such that

$$E(y/z) = Y = F(z) \tag{3.1}$$

where $F(z)$ is a monotonic function of z involving /// unknown parameters. A new but closely related drug, // or a newly manufactured batch of the old, can be /// *standardized* relative to S if and only if there is a // constant ρ such that a dose z of the new drug, T / (the *test preparation*), behaves exactly like a dose / ρz of S. That is to say, for T the dose-response / relation

$$Y = F(\rho z), \tag{3.2}$$

where F is the same function as for S in (3.1). If, as is often the case, T contains the same active principle as S but may be diluted to a different extent by mate-/ rials, the *relative potency*, ρ, is not just a pro- // perty of the animal system in which y is measured but is the ratio of the concentrations of active principle/ in S, T - a "rate of exchange" that is needed in human/ therapeutic practice for determining dosage of T.

In the simplest and most important case, over a wi- de range of doses $F(z)$ is a linear function of the lo-/ garithm of dose. Write

$$x = \ell nz , \tag{3.3}$$

and

$$\mu = \ell n\rho . \tag{3.4}$$

Then for S, T

$$Y_S = \alpha + \beta x , \tag{3.5}$$

$$Y_T = \alpha + \beta\mu + \beta x . \tag{3.6}$$

From a comparative experiment on various doses of S, T, two parallel reqression equations can be estimated:

$$\hat{T}_S = a_S + \beta x , \tag{3.7}$$

$$\hat{Y}_T = a_T + \beta x , \tag{3.8}$$

and then

$$M = (a_T - a_S)/b \tag{3.9}$$

is an estimator of μ . I shall comment on the / design of such experiments, under the assumption that / all responses, y, are Normally and independently dis-// tributed about Y with constant variance σ^2.

A second important case is that in which y is a / binomial variate taking only the values 0, 1. Such a // *quantal response* may be survival-death, absence or pre- sence of convulsions, death or germination of a seed or spore, etc. We can now interpret $F(z)$ as the probabi-/ lity of a positive response at dose z. Provided that / subjects are independent, if n subjects are tested at a dose the number responding will follow a binomial - / distribution. The two most useful and widely employed / formulations for $F(z)$ are

$$F(z) = \int_{-\infty}^{Y} \frac{1}{\sqrt{2\pi}} \exp(-{}^1/{}_2t^2)dt \tag{3.10}$$

and

$$F(z) = [1 + \exp(-2Y)]^{-1} \tag{3.11}$$

with x, Y as in (3.3), (3.5), (3.6). These correspond / with the well-known "probit" and "logit" methods. Both/ can be defended on theoretical grounds, and empirically both have been found satisfactory for many sets of data There is little theoretical or practical basis for pre- ferring one to the other. Indeed, with the definitions/ in (3.10), (3,11), $F(z)$ for a specified value of Y // will be much the same for both. Maximum likelihood me-/ thods are now appropriate to estimating values for a_T,/ a_S, b in (3.7), (3.8), and estimation of ρ proceeds as before by way of (3.9).

4. BIOASSAY DESIGN WITH QUANTITATIVE RESPONSES

In earlier lectures, I have been primarily concer-/ ned with experiments in which all treatment comparisons are of similar interest. I draw your attention to bio-/ logical assay as an example of more specific concern // for particular contrats, In an assay for which y is /

measured on a continuous scale (typically a weight of / an animal or of some organ of an animal), if there has/ been replication over several doses of S, T, standard / analysis of variance and regression procedures will // lead to (3.7), (3.8) with a_T, a_S, b being three linear/ contrasts among the observations. Interest is concen-// trated on the two contrasts $(a_T - a_S)$, b, on which / the experimenter will therefore be anxious to have high precision. In most symmetric designs, these two con- / trasts will be orthogonal, which simplifies the analy-/ sis. Either missing observations or deliberate choice / in respect of incomplete blocks may destroy orthogona-/ lity; this should not cause serious difficulty, but it/ will reduce precision.

Certainly that the two regressions are linear and / parallel is rare. In any assay, linearity is at best / an approximation over a range of doses; parallelism // may be disturbed by some unintended contamination of S or T, or by an unsuspecting attempt to estimate a re- / lative potency for a test preparation that does not - / obey the *similarity condition* expressed by (3.1), (3.2) Almost always, *tests of validity* of the linearity and / parallelism on which estiamtion of ρ depends will be/ wanted. The need then arises to balance the emphasis on precision of estimation and that on power for detection of invalidity.

Evidently a test of parallelism requires that at // least 2 doses of S and of T are included, and a test of linearity requires 3 doses. A popular design is the // (2.2), two doses of each preparation, but this allows / no test of linearity. The (3.3) design, 3 doses of S / and of T with equally spaced values of x, is far // better if validity tests are wanted, but for the same / total number of subjects it gives less precision for b. The range of doses for each preparation should be as // wide as may be risked without great risk or non-linea-/ rity, and any reasonable guess at μ (or ρ) should be used to choose doses of T at which the expected respon- ses will be close to those for the corresponding doses/ of S.

Interesting confounding schemes can arise. For - / example, a (3.3) design may need to be arranged in -/ blocks of 4. Tables 4.1, 4.2 show two ways of doing // this; a full experiment might repeat these sets of 3 // blocks several times. Although (3,3) is in one sense a/ 2x3 factorial design, in the general presentation of / Lecture VI I certainly would not have proposed these // designs. For bioassay, they can be very suitable. Table 4.1 has partial confounding of the parallelism and li-/ nearity tests. Table 4.2 avoids confounding of paralle- lism by reducing the precision of b. Both have merits/ according to needs. Table 4.3, despite unusual features avoids all confounding of b and of parallelism. Table 4.4., with a multiple of 6 blocks, reduces markedly the confounding of linearity and has less confounding of // parallelism and of b than in Tables 4.1, 4.2, respec- tively. I cannot discuss these designs in detail: they/ are here simply to illustrate my theme that highly spe- cific needs for estimation and testing call for spe-/ cially constructed designs, and that the statistician / must be prepared to evaluate the relative merits of al- ternatives.

In some circumstances, cross-over designs can be / used with great gains for precision.

Design has even wider connotations if the experimen ter has some choice of experimental conditions, source/ of subjects, and so on. He should then seek conditions/ such that the variance (σ^2) is small and the slope (β) large; smallness of the ratio σ/β is important to // high precision in the estiamte M.

5. BIOASSAY DESIGN WITH QUANTAL RESPONSES

Some design considerations are the same as for pa-/ rallel line assays in Section 4. A large value of β in (3.5) is still desirable, and each dose of T should be/ chosen with the aim of having the probability of respon se close to that for the corresponding dose of S. There is usually less concern for combinatorial patterns than with quantitative responses, largely because assays // require greater numbers of subjects and these tend to be more homogeneous in their quantal responsiveness // than they might be for a quantitative response.

Attention must be given to choice of doses and allo- cation of subjects to doses. A new difficulty arises /// from the binomial in the proportion responding: the fa-/ miliar expression "P(1-P)/n" has a maximum at P=0.5 and/ declines to zero when the response rate approaches 0 or/ 1. Combined with the flattening of the response curve // (3.10) or (3.11) at extremes of dose, this has two im-// portant consequences for an experiment with a fixed to-/ tal number of subjects:

(i) the precision of $a_S - a_T$ in (3.7), (3.8) is a maximum if all doses are close to that for /// $F(z) = 0.5$;

(ii) The precision of b is a maximum if subjects are about equally divided between a dose with $F(z) = 0.05$ to 0.10 and another with $F(z) =$ / $= 0.90$ to 0.95.

Now M involves the ratio of these two quantities. Algebraic study shows the ideal compromise to depend upon N, the total number of subjects, because the precision of b becomes less critical as N increases. / Although the assayist cannot act upon any exact rule/ since he does not know the dose-response relation, / he can be helped by having stated aims against which/ to interpret any advance information or even guesses. For example, when N=48 (a very small number), the // ideal probabilities are about 0.16, o.50, 0.84 for a/ (3,3) assay and 0.16, 0.36, 0.64, 0.84 for a (4,4). / When N=240, these are altered to 0.28, o.50, 0.72 and 0.28, 0.42, 0.58, 0.72. All these sets are close to / the optimal whether the response function be Normal / logistic, (3.10) or (3.11).

Occasionally an experimenter may have to choose / between alternative bioassay procedures, some using / quantal and some using quantitative responses. So far/ as precision is concerned, he can compare values of / $1/\beta$ for quantal responses and equation (3.10) with/ values of σ/β for quantitative responses. Of course, he will have at best numerical estimates of β, σ / from past experiments. He can use $1/\beta$, σ/β as/ though, to a first order, they are standard deviations of log potency estimates, per response measured. He / can then balance the costs of alternative assay procedures in respect of resource and time requirements / against these approximate deviations and make his --/ choice. The approximation may be very unsatisfactory / in small experiments.

TABLE 4.1

First 3 blocks for (3,3) parallel line assay in blocks of 4, with parallelism and linearity partially confounded.

	S_1	S_2	S_3	T_1	T_2	T_3
Block I	x		x	x		x
Block II	x	x			x	x
Block III		x	x	x	x	

TABLE 4.2

First 3 blocks for (3,3) parallel line assay in blocks of 4, with slope and linearity partially confounded.

	S_1	S_2	S_3	T_1	T_2	T_3
Block I	x		x	x		x
Block II	x	x		x	x	
Block III		x	x		x	x

TABLE 4.3

First 3 blocks (3,3) parallel line assay in blocks of 4, with linearity patially confounded.

	S_1	S_2	S_3	T_1	T_2	T_3
Block I	x		x	x		x
Block II	x		x		xx	
Block III		xx		x		x

TABLE 4.4

First 6 blocks for (3,3) parallel line assay in blocks of 4, with lesser confounding of slope, parallelism and linearity.

	S_1	S_2	S_3	T_1	T_2	T_3
Block I	x		x		xx	
Block II		xx		x		x
Block III	x		x	x	x	
Block IV	x		x		x	x
Block V	x	x		x		x
Block VI		x	x	x		x

D. J. FINNEY

IX. MULTISTAGE SELECTION AND SCREENING

1. INTRODUCTION

This lecture, based upon a lecture I gave to the American Statistical Association in August 1983, reviews another class of problems in the planning of experiments Though some of the mathematics goes back to Pearson 45 years earlier, my subject begins with Cochran (1951), who considered the problem of selection from a population in two stages. He referred especially to selection of good crop varieties from the many new candidates produced by plant breeders. He compared the consequences of achieving the same total intensity of selection by alternative pairs of selection intensities in the two years ; for example, an initial N varieties might be reduced to 0.01N by selecting a proportion 0.2 in the first year and further selecting a proportion 0.05 from these in the second year, or by using alternative pairs of intensities such as 0.1, 0.1 or 0.04, 0.25. Discussion of these and related options provides a good example of the planning of experiments.

2. THE GENERAL PROBLEM

Suppose that the "value" of an "entity" is represented by a variable x that cannot be measured directly (the yielding capacity of a wheat variety, the curative power of a bacteride), but estimates of x can be formed from experiments conducted in distinct stages successively in time (e.g. annually). For example, among a large number of wheat varieties, that labelled i might have a yielding capacity x_i. Observed yields $y_{i1}, y_{i2}, \ldots, y_{ik}$ in k successive stages (typically, a stage is a year) would have

$$y_{ir} = x_i + e_{ir} \qquad (r = 1,2,\ldots,k) \qquad (2.1)$$

where e_{ir} is a random error for which

$$E(e_{ir}) = 0, \quad E(e_{ir}^2) = \epsilon_r^2 \qquad (2.2)$$

with independent errors at different stages. With an agricultural crop, the expectation of yield would vary from year to year on account of weather and other conditions; all selection decisions, however, will be based on comparisons within a year, and provided that additivity obtains we need not complicate (2.1) by adding to x_i a parameter for "year" that is independent of i. The selection process begins with N_1 varieties or other entities, so that $i = 1,2,\ldots,N_1$. After each stage, some are discarded and he remainder are carried forward to the next stage, so that at the end of stage r only N_{r+1} remain. The aim is to optimize the values of the x_i for the final N_{k+1}, by choice of the experimental conditions and the selection intensity at each stage (Finney, 1964)

Discarded entities will not be measured in subsequent stages, so reducing experimental costs. At stage r selection can be based only on $y_{i1}, y_{i2}, \ldots, y_{ir}$ for the N_r entities still under trial. There will be restriction on the total resouces available, and ϵ_r^2 will be inversely related to the resources spent on the experiment at stage r. The selection at stage r might be based on one of the three sets of quantities:

(A) y_{ir} alone;

(B) $\bar{y}_{ir} = (y_{i1} + y_{i2} + \ldots + y_{ir})/r$, the unweighted mean;

(C) $\bar{y}_{irw} = (y_{i1}\epsilon_1^{-2} + y_{i2}\epsilon_2^{-2} + \ldots + y_{ir}\epsilon_r^{-2})/(\epsilon_1^{-2} + \epsilon_2^{-2} + \ldots + \epsilon_r^{-2})$ the weighted mean (or a similar mean using estimated variances).

The criterion of optimality for the N_{k+1} remains to be that maximizing the average of the values of x_i is more relevant to practical situations.

For simplicity hereafter, I omit the i subscript on x and y ; the context makes clear wheter particular x_i, y_{ir} or the sets of N_r values of these at stage r are intended.

3. CROP VARIETIES

Breeders of any species of crop plant produce large numbers of new genetic combinations (e.g. in the UK, about 500.000 annually for wheat), most of which will be rapidly discarded because of susceptibility to disease and pests, unsatisfactory habit of growth, and so on. After three seasons 1000 - 10000 may remain as potential varieties requiring to be assessed in terms of yield. The inherent yielding capacity of a variety, x, can be estimated (relative to others) by annual experiments giving the succession of means $y_1, y_2, \ldots, y_k$. Suppose that A measures total resources, and a portion A_r is used for experiments at stage r, where

$$\sum_{r=1}^{k} A_r = A \qquad (3.1)$$

If all experiments use the same small plots and resources are measured in terms of area of land, the number of plots per variety at stage r will be proportional to A_r/N_r, the share of resources available for one variety. Also, ε_r^2, the variance of the mean yield of a variety in the experiment at stage r, will be the variance per plot divided by the number of plots, and therefore

$$\varepsilon_r^2 \propto N_r/A_r \tag{3.2}$$

For various reasons, this relation is not exact, but it is a reasonable approximation. In a stable situation, a new cohort of N_1 enters the selection programme each year (of course N_1 is not exactly constant), and is passed through the stages independently of previous cohorts. Thus A_r may be regarded *either* as the share allotted to stage r from the total resources. A available to one cohort *or* as the share of total annual resources A that is allotted to the cohort then in stage r.

Cochran (1951) assumed the initial values of x to be a large sample from a Normal distribution (variance σ^2), and also assumed all error distributions to be Normal. His 2-stage selection used $\bar{y}_{rw}$; he noted that retaining the largest values must be optimal. If N_{r+1} is large, the mean of the N_{r+1} values of $\bar{y}_{rw}$ selected will approach the mean of the upper tail of the distribution for the fraction selected. For a N(0,1) distribution, the mean of the upper tail of area P is

$$\nu(P) = Z/P \tag{3.3}$$

where is the ordinate bounding the tail, and therefore $\sigma\nu(P)$ is the expected improvement if a fraction P of the largest x were to be selected without error. Thence in particular we easily derive the expected gain in x from a single stage of selection subject to error:

$$G = \sigma^2\nu(P)/(\sigma^2+\varepsilon_1^2)^{1/2} \tag{3.4}$$

Following Cochran, I studied (1958) selection of successive fractions $P_1, P_2, P_3, \dots, P_k$, where

$$P_r = N_{r+1}/N_r \tag{3.5}$$

and N_{k+1} is large. In accordance with (3.2), I wrote

$$\varepsilon_r^2 = \gamma\sigma^2 N_r A/(N_1 A_r) \tag{3.6}$$

γ being a dimensionless constant. I believed selection on y_r to be closer to the practice of plant breeders, and so at stage r I ignored information from earlier stages. Calculation could be based on the (k+1)- variate Normal distribution of $x, y_1, y_2, \dots, y_k$, with evaluation of the mean of x after truncations on each y_r separately. Because I wished to look at $k>2$, I developed an alternative procedure of obtaining the distribution of the x values remainig at each stage with the aid of cumulant transformations (Finney, 1956; 1961;1962 a), using expansions due to Cornish and Fisher (1937; Fisher and Cornish, 1960).
Writing

$$\Pi = P_1P_2\dots P_k \tag{3.7}$$

as the total selection fraction, I demonstrated that (when γ has a reasonable value 1.0) for $\Pi = 1/100$ nearly 90% of the maximum possible gain $\sigma\nu(0.01)$ could be achieved with $k = 2$; of course the position is less favourable for large γ. Optimal conditions are very flat in the neighbourhood of $P_1 = P_2$, $A_1 = A_2$. This suggests the more general rule that, for a total selection Π in k stages,

$$P_r = \Pi^{1/k} \text{ and } A_r = A/k \text{ for all } k \tag{3.8}$$

will be close to the maximizing conditions for gain. The rule is certainly not exact, and becomes a poorer aproximation for large k, but its operational ease must be attractive. Moreover, under most conditions, $k = 3$ seems large anough for all practical purposes. If Π is very small, an initial random discarding of a fraction $(1-P_0)$ may increase the eventual gain.

The advent of computers made practicable not only more extensive calculations on these lines but also comparison with finite samples. In 1966, I reported results with small samples from a Normal x-distribution. Agreement with the previous large sample calculations was good, except that for any value of Π the expected gain is smaller when N is small. For example, with $\Pi = 0.01$, the gain under perfect selection, using (3.3), is 2.665σ. Two-stage symmetric selection, equation (3.8), is always close to the best. With $\gamma = 1$, the expected gain for large N_1 is 2.345σ; this declines to 2.30σ for $N_1 = 500$, $N_3 = 5$ and to 2.13σ for $N_1 = 100$, $N_3 = 1$. Clearly this does not leave much opportunity for greater gains by allowing more stages: with $N_1 = 100$, the optimal appears to be 4 stages, for which the gain is 2.29σ, a small advantage to set against the delay of two seasons. A small study of variance functions used a generalization of (3.6):

$$\epsilon_r^2 = \gamma\sigma^2[N_rA/(N_1A_r)]^w \qquad (3.9)$$

and found both optimal conditions and expected gains remarkable insensitive to changes in w over a range from 0.1 to 2.0. Curnow (1960, 1961) found that replacing the Normal distribution of x by various beta and χ^2 distributions altered the magnitude of gains from selection / but made little difference to the optimal conditions and (3.8) remained a good aproximation. He confirmed that / even skew distributions give little advantage to a / fourth stage.

Table 3.1 illustrates results from simulations with $N_1 = 16$, $N_3 = 1$ for various N_2 and A_1/A, where $\gamma=1$ and

$$\epsilon_1^2 = \sigma^2A/A_1, \quad \epsilon_2^2 = \sigma^2N_2A/(N_1A_2) \qquad (3.10)$$

It shows clearly that the symmetric $N_2=4$, $A_1=A/2$ is close to the optimal, though the relation between the selected x and N_2, A_1 is very flat in this region. The general appearance of Table 3.1 is typical of what is found for larger and more interesting numbers. The scope for improvement is small: even perfect selection in this / example would yield a gain of only 1.766σ. Selection / based on $\bar{y}_{rw}$, however, is likely to make the result still less sensitive to changes in N_2 and A_1, since its benefits will be greatest when ϵ_2 is relatively large. Table 3.2 confirms this, showing only slight increases / in the maximum expected gain but an even larger region that is flat in respect of N_2 and A_1. Under realistic / conditions, most of the weight for $\bar{y}_{rw}$ comes from the current stage, so that results using y_r and $\bar{y}_{rw}$ will be highly correlated. Table 3.3 shows the variances among the individual simulations used for Table 3.1, and makes clear how easily an individual "16- N2-1" selection can give results far less (or far more) satisfactory than Table 3.1 indicates. The corresponding variances for selection based on $\bar{y}_{rw}$ are much the same except that they increase very little as A_1 approaches A.

A varieties x year interaction has the effect of increasing each ϵ_r^2 by a constant amount. If the interaction is appreciable, the ϵ_r^2 become more nearly / equal and for each variety $\bar{y}_{rw}$ moves closer to $\bar{y}_r$. The expected gain will be reduced by any interaction. Optimal conditions involve having the A_r monotonically de- / creasing instead of equal.

More recently, much has been done to systematize the whole process of running national crop variety trials. / Patterson & Silvey (1980) have described the British version of this. Current practice owes something to earlier ideas on multistage selection, but considerations of experimental design and of administrative practicability / raise new issues.

4. EXTERNAL ECONOMY

In 1960, I sought a basis for deciding the optimal / size of a selection scheme in relation to its benefits / for the national economy. What should the total resources be? How many varieties should enter stage 1 as an annual cohort? How many stages should be used, bearing/ in mind that increase in k delays the exploitation of gains? The principle is easy, but results depend upon parameters that are not easily guessed. Suppose replacement of a variety by a newer one that on average produces an increased yield of 1.0 per unit area will bring a monetary benefit of W from the total area under the / crop. Write U as the cost of increasing A by one unit / and V as the cost of producing one extra variety for testing. Then, with G still representing the increased ---/ yield per unit area, the net gain,

$$T = WG - UA - VN \qquad (4.1)$$

is to be maximized.

My calculations for various ratios U:V:W were restricted to k=1. They showed optimal conditions to give / values of T relatively insensitive to change in U : V, but over the range studied a 5-fold increase in W --- / brought a 6 to 7-fold increase in T. Curnow (1961) also studied k=2; his maxima for T were 15-20 percent greater than with k=1. There are reasons against using large k. Considerations of confidence in the farming community / show the dangers both of prematurity and of excessive delay in releasing a new variety. As k in increased, organization becomes more complicated, so that administrative expenses are increased and interest charges on a larger investment increase. Most important, for a continuing varietal improvement programme that starts a new cohort each year, is the fact that increasing k means / that a longer time elapses before release of a good variety makes its contribution to the term WG in (4.1).

5. ANIMAL SELECTION

When the "yield of an animal is measurable only in females, as in milk production, the value of a male for breeding has to be estimated by progeny testing. That is to say, he must be mated to a number of females and the yields of his female progeny must be measured under comparable conditions. Provided that each sire has mates representative of the same population, the means of the / progenies can guide the selection of sires. Constraints on the programme are likely to be the total number of/ progeny that can be tested in a generation and the number of sires to be selected as parents of a new generation. The variable factor is the number of sires to be / tested each generation and the consequent number of progeny per sire.

Robertson (1957, 1960) looked at this from a genetical standpoint and obtained an optimization rule that is essentially the same as for single-stage varietal selection as in Section 3. Multistage selection in animals -/ poses greater problems than in plants, because generations overlap and are longer.

6. DRUG SCREENING

I have discussed varietal selection at length(though still with little detail) because it is the application/ I know best. Logically similar problems arise in the - / screening of new chemical compounds for possible therapeutic activity. Again, large numbers are easily and - / cheaply produced in amounts adequate for testing, but very few will ameliorate a specified pathological condition. Davies (1958) and Armitage and Scheiderman (1958) initiated studies of this situation. Davies emphasized / that a very different distribution of X was appropriate. He proposed as an approximation a binomial distribution between a small proportion, θ , of effective and a large proportion of ineffective compounds, where θ / might be of the order of 0.01 or smaller. The aim of - / screening is to produce a much higher concentration of "good" compounds.

Dunnet (1961) discussed various criteria. He envisaged using a unit test (perhaps in an animal or a bacterial culture) to produce a response that has expectations x_0, x_1 for ineffective and effective compounds / respectively. Individual responses, y, are Normally distributed about x_0 or x_1 with variance . The probabilities P_0, P_1 that a set of means of y for a compound will exceed specified values are easily written as tail areas of a standardized Normal multivariate distribution. Suppose now that a series of "utilities" can be stated on an agreed scale. Let a be gain per compound / accepted, b_1 the loss from accepting an inactive compound b_2 the loss from rejecting an active compound, and c the cost per test unit. Then expected gain per compound tested is

$$G = aP_1\theta - b_1P_0(1-\theta) - b_2(1-P_1)\theta - c\bar{n} \qquad (6.1)$$

where $\bar{n}$ is the average number of unit tests per compound Dunnett proposed several forms for b_2. One can argue for

$$b_2 = [b_1P_0(1-\theta) + c\bar{n}]/P_1\theta \qquad (6.2)$$

the average cost incurred per active compound accepted ,/ or

$$b_2 = c\bar{n}/P_1\theta \qquad (6.3)$$

the average cost of the testing per active compound accepted, or

$$b_2 = [a(1-P_1)\theta + b_1P_0(1-\theta) + c\bar{n}]/P_1\theta \qquad (6.4)$$

which includes also the missed gain from rejection of active compounds. Any proposal for multistage selection / can be based on maximizing G with b_2 defined by one of (6.2), (6.3), (6.4). Dunnett mentioned other possibilities, such as maximizing the expected number of actives / accepted per unit cost of testing, $P_1\theta/c\bar{n}$, which is / the principle that Davies used.

The main difference from Section 3 in Dunnett's pattern of selection is that he advocated a predetermined / cut-off point at each stage. At stage r of his k - stage scheme, n_r responses are measured for each surviving compound; $\bar{y}_{rw}$ for any compound is now the simple mean response from the $(n_1 + n_2 + \ldots + n_r)$ tests made to date, and only those compounds for which

$$\bar{y}_{rw} > \eta_r$$

continue to the next stage, η_r being a fixed quantity . The formal problem to be solved is then that of choosing the n_r and the η_r $(r = 1, 2, \ldots, k)$ so as to optimize the criterion adopted.

7. THE BECHHOFER APPROACH

In a series of interesting papers, Bechhofer and his colleagues have studied selection with the different aim of identifying the best single entity. They began (Bechhofer, 1954; Bechhofer *et al.*, 1954) with theory relating populations (equivalent to "entities" in the present paper) on the basis of either a single sample or a sequence of /

two samples. Later (Bechhofer, 1958; Bechhofer and Blumenthal, 1962) they moved towards a classical sequential process for identifying the first in rank; they did not reject until the final stage, and ended with $N_{k+1} = 1$. Their criterion was the probability that the final selection is the entity with the maximum x from among the initial N_1. In a definitive account of these and related topics, Bechhofer *et al.* (1968) have generalized the theory to the case of $N_{k+1}>1$.

Bechhofer's work has not invoked any distribution of yielding capacity (x of earlier Sections), and has instead used minimax or similar considerations. But Bechhofer's idea is easily extended; why not proceed as in Sections 2, 3, with the same formulations of distributions and errors, but seek to maximize the probability that the final N_{k+1} includes the largest x? The question is closely related to that studied by Dunnett (1960), though he was concerned with choice of sample size for a single sample. At first sight, concentration of attention on the probability of selecting the best is attractive, but I doubt whether it has special merit for varietal selection, or even for drug screening. My reasons are:

(a) If N_{k+1}/N_1 is small, the probability of inclusion is likely to be so small (even at its maximum) as to make the chance of actually securing the largest x negligible; typically for varieties and for drugs N_{k+1}/N_1 is of the order 0.01, 0.001, or less.

(b) Alternatively, if a reasonable probability of success were demanded, the required replication would be unacceptably large or the required number of stages so great as to make the final selection out of date because of steady improvement among new input cohorts of N_1.

(c) If N_1 is moderately large, little will be lost if selection ends with only the second or third largest x, yet the criterion of maximization makes no allowance for this.

(d) The computations for finding the appropriate rejection rules would be intolerably laborious for any large N_1 ; Bechhofer's examples relate to $N_1 < 10$.

However, maximizing the probability of perfect selection and maximizing the mean of the final N_{k+1} values of x must require similar decisions at each stage so that a plan near to optimal for either will almost certainly be good for the other. In particular, for any specified N_r. A_r , it seems intuitively obvious that selecting the N_{r+1} largest values of y_r (or $\bar{y}_{rw}$) must be optimal for the probability criterion. Indeed, the probability criterion is equivalent to transforming values of x so as to replace the largest N_{k+1} values of x by 1 and the remaining (N_1-N_{k+1}) by 0. Tables 7.1, 7.2, with $N_{k+1}=1$, are from the same simulations that produced Tables 3.1, 3.2. The similarity of pattern confirms that for this small example, the optimal choice of N_2, A_1, is much the same for maximum probability of "correct" selection as for maximizing expectation.

8. HUMAN SELECTION

If a chemical compound shows little therapeutic benefit, it will be discarded; knowledge of how to re-synthesize it remains, but no notion of "fairness" to the compound impedes its removal from consideration in the current context. If a new wheat line does not achieve the selection level at any stage, it likewise will be discarded unless special features call for its retention as breeding material. In some circumstances, notably where human beings are selected, the position is very different.

Seeking an example, I attempted to study the selection of young people for educational patterns (1962) . I was concerned solely to recognize that some educational selection, or separation into different channels, is inevitable, and to explore the consequences of doing this in a manner that is not irrevocable and that seeks to optimize in respect of an explicit aim. The idea of "discarding" must be abandoned and any statistical solution must be interpreted flexibly and humanely. No country can afford to omit from its educational systems *either* measures for the fair treatment of each young person according to needs and potential *or* a planned policy for ensuring that the pool of ability is developed for the benefit of the community.

I discussed a system of 2-stage selection. Suppose that each child has an inherent ability, x and that some form of test at the end of primary education gives and estimate, y_1, that is subject to error. On the basis of Y_1, children are distributed between an academic and a non-academic secondary pattern in proportions P_1, $(1-P_1)$. Some years later, admission to university or other form of tertiary education is based upon a new estimate, y_2, and proportions P_2, P_2^* are accepted from the two streams. Here P_2^*, much smaller tha P_2, is intented as a "safety-net" for those who were erroneously placed in the non-academic stream at the first stage. The total proportion entering university is

$$\Pi = P_1P_2 + (1-P_1)P_2^* \qquad (8.1)$$

I regarded Π as fixed by external factors, and then enquired into the choice of P_1, P_2, P_2^* that would be - / "best", a word that remained to be defined. The obvious criterion was maximization of the mean value of x among those selected for university; this was used, but with full awareness of its faults.

I shall not present results of work that in retros-- pect can be seen to have rested on a vastly oversimpli-- fied formulation. One important difference from the types of selection discussed earlier is that any realistic approach would have to allow for changes in x between / stages. My purpose was to draw attention to a problem / and to the possibility of objective study. Even the ---/ attempt to formulate the problem in statistical terms is illuminating, for it discloses incertainties about aims and ignorance of relevant facts such as the size of - / correlations. My approach may have been inadequate. No country can afford to neglect the problem of ensuring / that those young people who are especially able to benefit shall be admitted to the highest levels of education

How this is to be achieved is debatable, but I am / convinced that the optimization of an appropriate multistage selection process must be a guiding principle. The major questions are: What "measurements" are to be made? What criterion is to be optimized?

TABLE 3.1

Mean Net Gains for Finite Sampling, $N_1 = 16$, $N_3 = 1$, $k = 2$, $\gamma = 1.0$, Selection Based on y_r

(expresed as multiples of σ, and estimated from 5000 simulations per entry)

Values of N_2

A_1/A	1	2	3	4	5	6	8	10	12	14	16
0.0	0	0.532	0.777	0.921	1.015	1.081	1.162	1.207	1.232	1.244	1.249
0.2	0.721	1.113	1.242	1.315	1.348	1.364	1.345	1.318	1.280	1.223	1.177
0.4	0.944	1.264	1.342	1.384	1.387	1.366	1.324	1.277	1.206	1.145	1.081
0.6	1.081	1.336	1.376	1.389	1.351	1.332	1.241	1.188	1.084	1.020	0.944
0.8	1.177	1.342	1.336	1.320	1.249	1.216	1.090	0.999	0.898	0.794	0.721
1.0	1.249	1.079	0.952	0.849	0.760	0.680	0.538	0.408	0.283	0.154	0.

All entries for $A_1/A = 0.0$, $A_1/A = 1.0$, $N_2 = 1$, $N_2 = 1$, $N_2 = 16$ are exact

Entry for $A_1/A = 0.5$, $N_2 = 4$ is 1.388 (10000 simulations)

TABLE 3.2

Mean Net Gains for Finite Sampling, $N_1 = 16$, $N_3 = 1$, $k = 2$, $\gamma = 1.0$, Selection Based on $\bar{y}_{rw}$

(expresed as multiples of σ, and estimated from 5000 simulations per entry)

	Values of N_2										
A_1/A	1	2	3	4	5	6	8	10	12	14	16
0.0	0	0.532	0.777	0.921	1.015	1.081	1.162	1.207	1.232	1.244	1.249
0.2	0.721	1.113	1.242	1.320	1.358	1.372	1.366	1.340	1.324	1.268	1.249
0.4	0.944	1.264	1.346	1.395	1.399	1.385	1.379	1.327	1.303	1.266	1.249
0.6	1.081	1.340	1.386	1.405	1.389	1.382	1.351	1.311	1.290	1.249	1.249
0.8	1.177	1.351	1.365	1.378	1.349	1.344	1.306	1.290	1.278	1.241	1.249
1.0	1.249	1.249	1.249	1.249	1.249	1.249	1.249	1.249	1.249	1.249	1.249

All entries for $A_1/A = 0.0$, $A_1/A = 1.0$, $N_2 = 1$, $N_2 = 16$ are exact

Entry for $A_1/A = 0.5$, $N_2 = 4$ is 1.402 (10000 simulations)

TABLE 3.3

Variance of Net Gains for Finite Sampling, $N_1 = 16$, $N_3 = 1$, $k = 2$, $\gamma = 1.0$, Selection Based on y_r

(expresed as multiples of σ^2, and estimated from 5000 simulations per entry)

	Values of N_2										
A_1/A	1	2	3	4	5	6	8	10	12	14	16
0.0	1.000	0.717	0.629	0.593	0.579	0.575	0.582	0.597	0.614	0.631	0.648
0.2	0.883	0.647	0.583	0.574	0.556	0.549	0.584	0.597	0.624	0.639	0.687
0.4	0.799	0.599	0.567	0.552	0.547	0.552	0.606	0.620	0.636	0.678	0.736
0.6	0.736	0.573	0.557	0.562	0.562	0.571	0.632	0.665	0.688	0.736	0.799
0.8	0.687	0.586	0.582	0.595	0.603	0.610	0.672	0.711	0.737	0.797	0.883
1.0	0.648	0.646	0.652	0.661	0.671	0.683	0.711	0.746	0.793	0.861	1.000

All Entries for $A_1/A = 0.0$, $A_1/A = 1.0$, $N_2 = 1$, $N_2 = 16$ are exact

Entry for $A_1/A = 0.5$, $N_2 = 4$ is 0.547 (10000 simulations)

TABLE 7.1

Proportion of Simulations in which Final Selection is the Largest x,

$N_1 = 16$, $N_3 = 1$, $k = 2$, $\gamma = 1.0$, Selection Based on y_r

(Estimated from 5000 simulations per entry)

A_1/A	Values of N_2: 1	2	3	4	5	6	8	10	12	14	16
0.0	0.064	0.124	0.171	0.215	0.242	0.280	0.330	0.351	0.363	0.368	0.375
0.2	0.191	0.311	0.361	0.407	0.420	0.427	0.425	0.412	0.382	0.354	0.341
0.4	0.253	0.384	0.421	0.450	0.444	0.438	0.424	0.388	0.340	0.318	0.302
0.6	0.302	0.421	0.444	0.449	0.420	0.415	0.377	0.346	0.288	0.276	0.253
0.8	0.341	0.428	0.419	0.407	0.365	0.350	0.300	0.264	0.223	0.204	0.191
1.0	0.375	0.285	0.220	0.199	0.164	0.152	0.122	0.097	0.079	0.078	0.064

Entry for $A_1/A = 0.5$, $N_2 = 4$ is 0.447 (10000 simulations)

TABLE 7.2

Proportion of Simulations in which Final Selection is the Largest x,

$N_1 = 16$, $N_3 = 1$, $k = 2$, $\gamma = 1.0$, Selection based on $\bar{y}_{rw}$

(Estimated from 5000 simulations per entry)

A_1/A	Values of N_2: 1	2	3	4	5	6	8	10	12	14	16
0.0	0.064	0.124	0.171	0.215	0.242	0.280	0.330	0.351	0.363	0.368	0.372
0.2	0.191	0.311	0.362	0.410	0.425	0.431	0.438	0.422	0.411	0.381	0.386
0.4	0.253	0.384	0.426	0.458	0.452	0.449	0.453	0.411	0.400	0.387	0.383
0.6	0.302	0.426	0.451	0.461	0.442	0.446	0.438	0.400	0.389	0.379	0.382
0.8	0.341	0.435	0.434	0.442	0.418	0.422	0.406	0.390	0.385	0.377	0.386
1.0	0.375	0.378	0.377	0.378	0.374	0.387	0.382	0.378	0.380	0.364	0.387

Entry for $A_1/A = 0.5$, $N_2 = 4$ is 0.458 (10000 simulations)

D. J. FINNEY

X. THE QUESTIONING STATISTICIAN

1. INTRODUCTION

I propose to end my lectures with a discussion of / the role of statisticians as consultants and collaborators in experimental design (Finney 1982a). I shall // do this by reference to 22 questions that are typical of those I might discuss with a colleague from another /// scientific discipline who wished for my help in this // experimental programme.

Too often the statistician is regarded as someone// who appears after data have been collected, performs // standard calculations, delivers a verdict "Significant" or "Not Significant", and departs. This ought to be totally false. A statistician needs to be involved at all stages of an investigation; unless he can also interact thoroughly with those who bring other expertise to the/ study, his own skills will not be used to best advantage. To be specific on details of such interaction is // not easy, for both the field of application and perso-/ nal experience are relevant. I shall restrict myself to the planning and design of *comparative experiments*, that is to say investigations for comparing two or more /// treatments or categories in terms of measurements or // observations on subjects, and shall illustrate by re-// ference to biological experiments. I shall have little/ to say on statistical analysis.

2. INTERACTION WITH THE INVESTIGATOR

A biologist may see himself as merely wishing to / ask a statistician a few critical questions. In prac- / tice, a more useful first stage may consist of ques-- / tions *from* the statistician, because he has a broader/ understanding of the relevant to effective deployment / of his discipline, such questions should be a basis for discussion rather than for exact answers. Certainly // there is no way in which a set of answers can automatically generate a plan for statistical activity!

Some of my 22 questions are easy and obvious, -- / others may require detailed study and complex answers. Some answers, and also particular circumstances, may // indicate further topics to be explored at the planning/ stage. Trials involving human subjects commonly intro-/ duce special difficulties in their ethical and organi-/ zational aspects; these may require deeper discussion / than, for example, Questions 5, 10 and 18 may seem to / imply. All the questions are important to good statis-/ tical collaboration. All should be clearly in the expe-/ rimental scientist and his statistician-colleague. Many/ are implicit in D. R. Cox's excellent book (1952), --/ which contains far more detail. I do not suggest that / the set of 22 is complete; a statistician with different experience might add more questions or group these 22 // differently.

3. WHAT EXPERIMENT?

QUESTION 1 IS WHAT YOU PLAN TRULY AN EXPERIMENT?

The essential feature is that the experimenter has / the power to determine which subjects shall receive - / which treatments.

He can choose which of 50 rabbits are to receive // each of several doses of an antibiotic; he cannot choose which of 50 young adults shall smoke 20 cigarettes a day for 10 years and which shall be non-smokers. He can --// choose which laboratory rats shall be exposed to a car-/ cinogen and which shall not, but he cannot choose which/ members of a population of rats in the wild shall be exposed to an accidental spillage of toxic wastes. To drow attention to the logical weaknesses inherent in various / types of "non-experiment" (retrospective case-control // studies, unplanned observational data, volunteered information) is not to condemn them. Armitage (1981) expresses the point succinctly: "When a randomized controlled --// trial is both practicable and ethical it is, in my view,/ a poor second best to rely on non-randomized comparisons. In the vast majority of aetiological studies, on the --// other hand, randomization is not practicable and it would be foolish to spurn a carefully controlled case-control / or cohort study": In epidemiology, as in ecology and astronomy (though for different reasons), the interpreta-/ tion of non-experimental data may offer the only hope of progress. But to analyze such data uncritically as though they come from a planned experiment invites fallacious // argument and misleading conclusions. The inferential /// problems contain many additional difficulties (Anderson / et al., 1980; Cochran, 1965, 1968).

QUESTION 2 WHY DO THE EXPERIMENT?

Four main classes of reason are worth distinguishing!

(A) Curiosity. What will happen if one substance,/ component, article, method, etc. is used in place of another?

(B) Direct interest of comparison in this class of / subject. Will a new vaccine protect more pigs / from disease than the existing standard vaccine?

(C) Intended transfer of conclusions to different // circumstances.
Anti-tumor drugs may be compared first in labo-/ ratory mammals, even though the sole purpose of/ the research is to improve human health and nu-/ trition.

(D) Subjects are instruments for calibrating treat-/ ments. In biological assay, animals or bacterial cultures are treated differentially not so as to/ measure treatment effects on subjects but to estimate how much of one material will produce the same response as 1 mg or 1 ml of another /Finney 1978).

At this point, an evaluation of previous work in the / same and closely related fields is needed.

4. THE EXPERIMENTAL UNITS

QUESTION 3 WHAT EXPERIMENTAL UNIT IS TO BE USED?

This is the unit to be differentiated for treat-/ ment. It is often termed the *plot*. It might be one patient, all patients in a ward, a single mouse, a group of animals or insects that must be treated alike, an / inoculation site on a live animal, the set of what - / plants on a 10mx2m area of land, a plate of bacterial/ culture, or even a specified time interval for any of/ these (where a time sequence of different treatments / can be applied to a biological unit).

Exact definition is essential. Gross errors arise/ if individual patients in a ward, animals in a pen, or insects in a colony are regarded as "plots" when in / fact the treatments are necessarily applied across the larger experimental unit.

The common environment of the ward or the colony / is likely to affect its individual members and so to / destroy the independence of their separate responses // to treatment.

QUESTION 4 ARE UNITS "OF EQUAL SIZE"?

If the units consist of different numbers of ani-/ mals or different lengths of time or if they have / different physical dimensions, the implications for // statistical analysis need early consideration. Techni-/ cal difficulties may require modifications to a compu-/ ter program or a minor development of statistical theory.

QUESTION 5 ARE UNITS GROUPED IN ANY WAY, SUCH // THAT MEMBERS OF THE SAME GROUP WILL BEHAVE MORE SIMILARY THAN MEMBERS OF /// DIFFERENT GROUPS?

Groupings might consist of neighbouring plants, a-/ nimals from one litter, strips of muscle from one ani-/ mal, and so on. This feature is relevant to block - // structure in design. There may be two potentially use-/ ful groupings, and consequently a need to assess their/ relative merits and make a choice, or to consider Latin squares.

QUESTION 6 DO SUCH GROUPS CONTAIN EQUAL NUMBERS OF UNITS?

If blocks of different sizes are to be used, greater complications may enter into discussion of good design. These are not insurmountable, but they must be // recognized.

QUESTION 7 CAN UNITS BE SUBDIVIDED, IN SPACE OR IN TIME, FOR FURTHER TREATMENT COMPARI- / SONS ?

For nutrition, the whole animal must be the unit, / but skin reactions to different materials may be tested on one animal. Treatment of plants against seed-borne / diseases must use whole plants or groups of plants as / plots, but it may be possible to compare fungicides on/ individual leaves. Before using such a *split plot desing* one must consider the possibility that the treatment // applied to one sub-unit may influence the measurements / recorded for other sub-units in the same main unit.

5. THE TREATMENTS

QUESTION 8 HOW ARE THE PROPOSED TREATMENTS STUCTURED?

There are four main types:

(P) No structure - e.g. many different proce-/ dures (perhaps varieties of wheat) to be compared, no / pattern or hierarchy among/ them.

(Q) Minimal structure- perhaps some grouping of // treatments e.g. several new drugs to be compared with a standard but only "new v. / standard" comparisons are / of interest.

(R) Linear structure - treatments cover a range in one dimension e.g. "slight", "mild", "severe" *or* quantitative, as in response curves studied and estimated / by using treatments 0,1,2,/ 3,4,...(or 2,5,10,25,...).

(S) Factorial structure- combinations of different / treatments. May take different sets of treatments of/ types P, Q, R, and use all/ combinations; e.g. one // standard drug and two new / ones, each tested at doses/ 0,1,2,4, that is to say one factor of type Q and one of type R.

With human subjects, the structure will usually be / simple, and the number of treatments will seldom ex-/ ceed four; a subsidiary feature of structure may per-/ mit planned cross-over of treatments during the ex-/ periment. With non-human subjects, and especially in/ agricultural and industrial research, the treatment / structure may involve far more complex combinatorial/ patterns.

QUESTION 9 HOW RIGID ARE THE REQUIREMENTS FOR // TREATMENTS ?

If 17 unstructured or minimally structured treat-/ ments are proposed, would there be serious objection / to including only 16 or increasing to 18 or 20? For/ type Q, might there be advantages in including a standard treatment with double or triple replication?

If the structure is linear, are the number and // spacing of levels open to discussion?

If several factors are to be included, is there any problem in accommodating all combinations? Thus 3 / drugs, 4 levels of dose, 2 methods of administration // give 3x4x2 = 24 combinations, possibly too many for the resources. Experimental design is easier,and in various ways more satisfactory, if the total number of treat- / ments (when large) is a square or factorizes easily // ($16=4^2$, $20=4x5$), and also for factorial structure if / all factors have the same number of levels; the statistician might prefer 3x3x3 to 3x4x2, but this is not // mandatory. Factors all at 2 or all at 3 levels are es-/ pecially suitable for confounding and fractional replication. The choice of number and spacing of levels can/ have large consequences for the precision of particular estimates. As in other matters, decision must come from the experimenter, but the statistician must accept responsibility for showing the merits and weaknesses of / various possibilities.

6. RANDOMIZATION

QUESTION 10 ARE THERE ANY CONSTRAINTS ON RANDOM ALLOCATION OF TREATMENTS TO UNITS ?

Randomization is one of the most important concepts that statisticians have introduced into experimentation. Any departure from completely random allocation of -- / treatments to units is a matter for careful examination at the planning stage. Block and confounding constraints adopted for excellent reasons related to the nature of / the units and the optimization of precision, demand // highly organized restrictions on randomization. A com-/ plex experiment may have independent but interlocking/ phases of randomization, in place of one total drawing/ of lots. These devices in no way destroy the validity / of inferences as long as it is remembered that the system of randomization determines the structure of the / statistical analysis.

For particular experiments, objections to randomi-/ zation may be raised. An ethical objection is often in/ reality an objection to experimentation rather than / only to randomization, possibly because leaving any subjects or experimental units as untreated controls is / deemed improper. Other objections may come from a wish/ to avoid putting certain pairs of treatments on adjacent units, or from an operational convenience if treatments are in a certain order. Any objection deserves dis-- // cussion: usually some compromise can maintain statis-/

tical validity within constrains that are inescapable. Ambiguities in the interpretation of experiments have/ often been caused by insufficient attention to randomization.

7. AIMS

QUESTION 11 WHAT ARE THE OBJECTIVES AND PRIORITIES?

Is the experiment primarily aimed at discovering / significant differences (e.g. is the effect of a drug/ sensitive to the vehicle in which it is administered?) or at estimating numerical properties (differences in/ yield potential between varieties of maize, relative / potency of two drugs, or other particular parameters)? If there are several objectives, how do they compare / in importance?

QUESTION 12 WHAT VARIATES ARE TO BE MEASURED ?

Is the experiment to be assessed solely, or primarily, in respect of one variate (quantity of milk, u-/ terine weight, survival time) or are several of comparable interest? If many, could some be omitted as --// nearly duplicating others (e.g. different techniques / for measuring what is qualitatively one characteris- / tic)? There is no point in collecting the same information from the same plots twice!

QUESTION 13 WHAT VARIATES ARE TO BE ANALYZED ?

Is each measured variate to be analyzed separately? Are combinations (percentages, ratios, indices, "corrected values", etc.) to be used? Are all these essen- // tially distinct? Why analyze both final weight and in-/ crease in weight if initial weights differ only negli-/ gibly? Must every variate be put through a full sta-// tistical analysis (e.g. analysis of variance) or will / simple tabulations suffice for some? A nutritional experiment, for example, can easily have variates rela-// ting to food intake, growth, perfomance, and various // aspects of metabolism, in all far more numerous than // the human or animal subjects under trial! Any experi-/ ment on animals or human subjects that involves conti-/ nous monitoring of certain functions can produce even / greater numbers of variates. However important an experiment may be, it should not be allowed to generate // large quantities of computer output that will never be/ read; not only does this waste effort, but excessive // output may distract attention from the most valuable // findings.

If a multivariate analysis is intended, what is its purpose? Is the interest in an estimated regression // function, in establishing a discrimination technique, / in seeking useful and interpretable canonical variates, in exploring classifications, or in something different?

8. PRECISION

QUESTION 14 HOW PRECISE OUGHT RESULTS TO BE ?

To say "Very" or "As precise as possible" is unhelpful! If the design is good, precision can be increased/ only by increasing the size of the experiment, that is / to say including more units and increasing the costs. // Therefore, at least for the most important variates, the minimal requirements for precision may need to be stated in terms of the standard error of a difference between / means, a range of error at a stated probability for an / estimated paramenter, or a power for a significance test. The experiment must then be planned either to satisfy all demands simultaneously or to conform to some agreed com-/ promise.

QUESTION 15 WHAT IS KNOWN OF THE VARIABILITY BETWEEN / UNITS ?

Previous similar experiments, or general notions on the range of replicate values, may facilitate a guess at the variance per unit for each major variate that must / eventually be analyzed. At best, this information can be accepted only as a rough guide, but it may aid planning.

QUESTION 16 CAN ANY CONCOMITANT BE USEFULLY MEASURED OR RECORDED ?

Records of initial weights or of some other perfo-/ mance or property of units *before treatments begin* can sometimes by used effectively in standardizing variates, improving comparability, and thereby reducing effective/ variance.

Even after 50 years, covariance analysis and // equivalent techniques for improving precision are / still inadequately exploited, except perhaps in agricultural field experiments.

9. SERIES OF EXPERIMENTS

QUESTION 17 IS THE EXPERIMENT AN ISOLATED STUDY OR / PART OF A SERIES ?

If an experiment is one of a series conducted at/
the same time in different places, coordination of //
design is desirable. Designs at different sites may /
be identical (except for randomization), or with mi-
nor variations on the treatments included, or inter-/
locking so that each represents part of an overall //
grand design. This can be particularly important in /
large-scale clinical studies and in agricultural re-/
search intended as the basis of advice for a region.

If an experiment is one of a sequence in time,///
many aspects of the design of each may need to be ///
considered in relation to what can be learned from //
its predecessors.

10. RESOURCES AND CONSTRAINTS

QUESTION 18 WHAT RESOURCES CAN BE USED ?

How many units can be used? For how long? What/
limitations are there on staff of all kinds? Are ma-
terials (e.g. supplies of scarce substances) a res-/
triction? Are there statutory requirements, such as/
"at least 15 mice on each drug"? Are any upper li- /
mits on resources absolute, costly to exceed, or //
merely convenient? In clinical research, what is the
approximate rate at which new cases will become avai-
lable?

Identification of the true limits on planning is/
not always obvious. A research student once asked my/
advice on an experiment on different varieties of - /
pasture grass. What size of plot and what replica- /
tion should she use in order to estimate rates of - /
growth by repeated cutting? Only at the end of an //
hour's discussion did reality emerge: the scale of //
the experiment was totally constrained by the dis-- /
tance that she personally could push a small lawn- /
mower in one afternoon!

QUESTION 19 ARE THERE ANY TIME CONSTRAINTS ?

Unless the design needed is obviously very sim-/
ple, a statistician will need time to absorb and di-/
gest the ideas and to produce a good proposal. He //
may make tentative suggestions during a first dis- /
cussion with the investigator, but he may need seve-/
ral days to explore these further. Indeed, he may ///
need more time than he woild have taken 30 years ago, /
since today he can make a computer study of the rela--
tive precision of various possibilities whereas pre--/
viously he might have had to rely on intuitive assess/
ment. In an emergency, he might produce a tolerable //
design very quickly; a good design may emerge only af/
ter weeks of intermittent attention to the problem, //
during which time new questions of feasibility and ///
appropriateness can be raised with the investigador.

He may need to examine timing in relation to availa-
bility of staff, equipment, and other resources. He may
ask about weather or other uncontrollable environmental
factors. Yet again, if the statistical analysis of re-/
sults is not going to be standard in type and moderate/
in quantity, he should ask whether the timing for the /
elaboration of analytical procedures and the execution/
of the analyses themselves (statistician's time and ///
computer availability) have been considered in relation
to any urgency in the production of reports.

QUESTION 20 IS THERE ANY PLACE FOR PLANNED SEQUENTIAL DESIGNS ?

If the time between administration of treatments and
completion of the measurement and assessment of each - /
subject is short relative to the rate at which new units
can be made ready for treating (in clinical trials, - /
effectively the rate of recruitment of new subjects), a/
formal sequential design may need consideration.

11. RECORDING AND ANALYSIS

QUESTION 21 HOW ARE THE RESULTS TO BE RECORDED ?

If recording is by pen-and-paper, from direct obser-
vation and counting or from reading of instruments, are/
adequate precautions taken against mistakes (gross mis-/
counting and misreading, biases due to misuse or mis- /
placement of instruments, etc.)? Will legibility be - /
preserved, especially if records are taken in the open /
air or under difficulties ? Will specially designed //
record forms be helpful and an aid to good organization?
Will the statistician receive the original records, a //
good photocopy, or (very undesirable) an unchecked ma- /
nual or typed copy?

If mechanical or electronic data-capture or data- //
logging equipment is used, is it of tested reliability ?
Will adequate information on its interpretation be avai-
lable. Are there problems of calibration, background //

rates, etc. ?

QUESTION 22 WHAT ARRANGEMENTS ARE NEEDED NOW FOR / SUBSEQUENT ANALYSIS ?

Possibly the only inmediate step is agreement that/ 1 week or 1 year from now results will be passed to the statistician. Possibly arrangements must be made for // further intensive discussions about what is to be ana-/ lyzed and how. Possibly there will be recognition that/ statistical algebra and computer programs must be stu-/ died *now*, in order to be ready for the eventual analy-/ sis. Is a suitable program available, must one be spe-/ cially written, or will an existing program with a /// little supplementary *ad hoc* computing suffice for a / job unlikely to recur? If a major series of experi- // ments is being planned, almost certainly attention to / providing comprehensive software for record storage, // data management, statistical analysis, and production / of graphical and tabular summaries will be essential.

12. GENERAL COMMENTS

More important than exact statement of questions is recognition, by the estatistician and by his colleagues of his role as a questioner over a wide field. For any complex problem, personal discussion spread over several occasions is essential; answers to a printed questio- // nnaire would be of far less value. I have known inves-// tigators who would regard some of these questions from a statistician as impertinent. Even Question 3 can evoke / the response "That is not your business"! I have known/ others who would bring ideas for a complicated experi- / ment to a statistician at 4.00 p.m. and request a design that can be put into operation at 9.30 a.m. the next day. These are not merely irritations to the statistician or / material for anecdotes. They are sure ways of obtaining/ advice less than the best and often totally inadequate.

Nothing I have said implies that the design and ana-/ lysis phases should be kept totally separate; commonly, / indeed, the eventual analysis and the computer facilities it needs should be discussed before the design is finally decided. Sometimes results from a well-designed experi-/ ment convey so clear a message that the only analysis /// needed is production of a few mean values. Sometimes an/ experienced statistician may see that standard techniques primarily analysis of variance, will extract almost all / relevant information from an exoeriment with minimal -// trouble. Sometimes a further extensive phase of questio-/ ning and interacting with the investigator is essential.

A few years ago, I heard of an important research / organization that used its one statistician solely as a computer with ears and vocal cords. He "analyzed" data/ put before him without regard to their nature, and re-/ turned formal summaries of his findings without concern for their comprehensibility. He was never involved in / planning and design. Where the blame lay is irrelevant: the correct role for a statistician is very different. He should be a collaborator, not a servant, participa-/ ting extensively and deeply in many aspects of a re- / search programme. In some investigations, he should be a full partner, contributing his expertise and sharing / the total responsibility at all stages from tentative // plans to publication of reports.

XI. REFERENCES

No good general book on experimental design has been published in recent years. There is great need for a text that takes account of developments in the past 25 years; designs for special purposes, sequential designs, integration of various ideas (cross-overs, nearest neighbours, etc.) with the mainstream of design planning for series of experiments in space and time and combination of their results. Above all, there is need for a book that recognizes the changes brought by computers in ease of generation of designs and especially the greater flexibility acceptable in combinatorial structure now that labour of computation is no longer a constraint. In connexion with my series of lectures I can recommend only five as broad enough in outlook: Cox (1959) as a superb account of topics such as I have mentioned briefly in Lecture X, Cochran & Cox (1957) and Finney (1960) for general presentations of design and analysis without full theory, Kempthorne (1952) and John (1971) for more detailed theoretical discussion.

The list of references that follows is not intended as a comprehensive bibliography. I have included only the books mentioned above and a short selection of publications that are relevant to topics in the lectures but are not necessarily original or primary sources of information.

ANDERSON, S., AUQUIER, A., HAUCK, W.W., OAKES, D., VANDEALE, V., and WEISBERG, H.I. (1980) *Statistical Methods for Comparative Studies*, New York: John Wiley & Sons.

ARMITAGE, P. (1981) Thrombosis and oral contraception. *British Journal of Hospital Medicine*, *25*, 485.

ARMITAGE, P. and SCHNEIDERMAN, M.A. (1958) Statistical problems in a mass screening program. *Annals of the New York Academy of Sciences*, *76*, 896-908.

BECHHOFER, R.E. (1954) A single-sample multiple decision procedure for ranking means of normal populations with known variances. *Annals of Mathematical Statistics*, *25*, 16-39.

BECHHOFER, R.E. (1958) A sequential multiple-decision procedure for selecting the best one of several normal populations with a common unknown variance, and its use with various experimental designs. *Biometrics*, *14*, 408-429.

BECHHOFER, R.E. and BLUMENTHAL, S. (1962) A sequential multiple-decision procedure for selecting the best one of several normal populations with a common unknown variance. II: Monte Carlo sampling results and new computing formulae. *Biometrics*, *18*, 52-67.

BOCHHOFER, R.E., DUNNETT, C.W., and SOBEL, M. (1954) A two-sample multiple decision procedure for ranking means of normal populations with a common unknown variance. *Biometrika*, *41*, 170-176.

BECHHOFER, R.E., KIEFER, J. and SOBEL, M. (1968) *Sequential Identificacion and Ranking Procedures*: Chicago: The University Press.

COCHRAN, W.G. (1951) Improvement by means of selection. *Proceedings of the Second Berkeley Symposium*, 449-470.

COCHRAN, W.G. (1965) The planning of observational studies of human populations. *Journal of the Royal Statistical Society*, A*128*, 234-265.

COCHRAN, W.G. (1968) The effectiveness of adjustments by subclassification in removing bias in observational studies. *Biometrics*, *24*, 295-313.

COCHRAN, W.G. and Cox, G.M. (1957) *Experimental Designs* (2nd edition), New York: John Wiley & Sons.

CORNISH, E.A. and FISHER, R.A. (1937) Moments and cumulants in the specification of distributions. *Revue of the International Statistical Institute*, 5, 307-320.

COX, D.R. (1959) *Planning Experiments*, New York: John Wiley & Sons.

CURNOW, R.N. (1960) The consequences of errors of measurement for selection from certain non-normal distributions. *Bulletin of the International Statistical Institute*, *37*, 291-308.

CURNOW, R.N. (1961) optima programmes for varietal selection. *Journal of the Royal Statistical Society*, B23, 282-318.

DAVIES, O.L. (1958) The design of screening tests in the pharmaceutical industry. *Bulletin of the International Statistical Institute*, 36, 226-241.

DUNNETT, C.W. (1960) On selecting the largest of k normal population means. *Journal of the Royal Statistical Society*, B22, 1-40.

DUNNETT, C.W. (1961) Statistical theory of drug screening. In *Quantitative Methods in Pharmacology* (ed. De Jonge). Amsterdam: North Holland Publishing Company, 212-231.

FEDERER, W.T. & SCHLOTTFELDT, C.S. (1954) The use of covariance to control gradients in experiments. *Biometrics*, 10, 282-290.

FINNEY, D.J. (1956) The consequences of selection for a variate subject to errors of measurement. *Revue of the International Statistical Institute*, 24, 22-29.

FINNEY, D.J. (1958) Statistical problems of plant selection. *Bulletin of the International Statistical Institute*, 36, 242-268.

FINNEY, D.J. (1960a) A simple example of the external economy of varietal selection. *Bulletin of the International Statistical Institute*, 37, 91-106.

FINNEY, D.J. (1960b) *An Introduction to the Theory of Experimental Design*, Chicago: University Press.

FINNEY, D.J. (1961) The transformation of a distribution under selection. *Sankhya*, A23, 309-324.

FINNEY, D.J. (1962a) Some propertues of a distribution specified by its cumulants. *Technometrics*, 5, 63-69.

FINNEY, D.J. (1962b) The statistical evaluation of educational allocation and selection. *Journal of the Royal Statistical Society*, A125, 525-564.

FINNEY, D.J. (1962c) An unusual salvage operation. *Biometrics*, 18, 247-250.

FINNEY, D.J. (1964) Screening processes: problems and illustrations. *Contributions to Statistics Presented to Professor* P.C. *Mahalanobis*, 73-100.

FINNEY D.J. (1966a) An experimental study of certain screening processes. *Journal of the Royal Statistical Society*, B28, 88-109.

FINNEY D.J. (1978a) *Statistical Method in Biological Assay*, (3rd edition), London: Charles Griffin & Co.

FINNEY, D.J. (1978b) Testing the effect of an intervention in sequential ecological data. *Biometric*, 34, 706-708.

FINNEY, D.J. (1980) The estimation of parameters by least squares from unbalanced experiments. *Journal of Agricultural Science*, 95, 181-189.

FINNEY, D.J. (1982a) The questioning statistician. *Statistics in Medicine*, 1, 5-13.

FINNEY, D.J. (1982b) Some enumerations for the 6x6 Latin squares. *Utilitas Mathematica*, 21A, 137-153.

FINNEY, D.J. (1982c) Intervention and correlated sequences of observations. *Biometric*, 38, 255-267.

FINNEY, D.J. (1984) Improvement by planned multistage selection. *Journal of the American Statistical Association* (in press).

FISHER, R.A. and CORNISH, E.A. (1960) The percentile points of distributions having known cumulants. *Technometrics*, 2, 209-225.

FISHER, R.A. and YATES, F. (1964) *Statistical Tables for Biological, Agricultural and Medical Research*, (6th edition), London: Oliver & Boyd.

FRANKLIN, M.F.F. and MANN, A. D. (1980) DSIGNX. *Inter-University/Research Councils Report*, no.48.

JOHN, P.W.M. (1971) *Statistical Design and Analysis of Experiments*, New York: The Macmillan Company.

KEMPTHORNE, O. (1952) *The Design and Analysis of Experiments*, New York: John Wiley & Sons.

OUTHWAITE, A.D. and RUTHERFORD, A.A. (1955) Covariance analysis as an alternative to stratificacion in the control of gradients. *Biometrics*, 11, 431-440.

PATTERSON, H.D. and HUNTER, E.A. (1983) The efficiency of incomplete block designs in National List and Recommended List cereal variety trials. *Journal of Agricultural Science*, 101, 427-433.

PATTERSON, H.D. and SILVEY, S.D. (1980) Statutory and reommended list trials of crop varieties in the United Kingdom. *Journal of the Royal Statistical Society*, A143, 219-252.

PATTERSON, H.D., WILLIAMS, E.R., and HUNTER, E.A. (1978) Block designs for variety trials. *Journal of Agricultural Science*, *90*, 395-400.

ROBERTSON, A. (1957) Optimum group size in progeny testing and family selection. *Biometrics*, *13*, 442-450.

ROBERTSON, A. (1960) On optimum family size in selection programmes. *Biometrics*, *16*, 296-298.

YATES, F. (1933) The principles of orthogonality and confounding in replicated experiments. *Journal of Agricultural Science*, *23*, 180-145.

D.J. FINNEY

DISEÑO DE EXPERIMENTOS

I. DISEÑO Y ESTIMACION: ALEATORIZACION COMPLETA

1. INTRODUCCION

Mi diccionario dice que un *experimento* es una acción emprendida con el fin de descubrir algo que es desconocido. Observen la implicación de que el experimentador // tiene cierto libre albedrío: determina las condiciones/ particulares de la experimentación. Por ejemplo, un antiguo químico deseaba saber si para que un material ardiera se necesitaba aire. Su experimento consistió en / colocar una vela encendida en el interior de un reci- / piente de manera que no pudiera entrar más aire: por supuesto la vela se apagó pronto. El experimentador aplicó el *tratamiento* del recipiente cerrado al *sujeto*, una vela encendida. Tenía el poder de impedir o utilizar o no el tratamiento, y definirlo exactamente en términos de tamaño, forma, materiales, etc.

Trataré sólo de *experimentos comparativos*, es decir, experimentos en que dos o más tratamientos deben compararse entre si, respecto a alguna propiedad mensurable. Obviamente, uno desea realizar una comparación impar- / cial, asegurando que todas las demás condiciones rela-/ cionadas con el resultado sean lo más semejante posible y que los sujetos difieran solamente en los tratamien-/ tos que reciben. Simbólicamente (con una notación que / substituiré más adelante por algo más exacto), la medición y para dos sujetos, pudiera representarse por

$$y_1 = M + G + F + T_1 \tag{1.1}$$

y

$$y_2 = M + G + F + T_2 \tag{1.2}$$

para dos tratamientos,aplicado cada uno de ellos a un sujeto.En este caso,M es un nivel medio general, G representa desviaciones de éste asociadas con factores inherentes en el sujeto -en un contexto biológico,posiblemente la modificación de M apropiada para un animal de una raza,constitución genética,sexo,edad,etc.,particulares- y F representa/ desviaciones asociadas con factores ambientales tales como la dieta y la temperatura.Siempre que G y F sean idénticos para los dos sujetos, la diferencia y_1-y_2 es igual a la diferencia T_1-T_2 entre los efectos de los tratamientos.

Aquí están implícitas dos suposiciones: (i) pueden encontrarse sujetos para los que G y F son idénticos y (ii)/ sujetos idénticos en G, F, y tratamiento darán exácta- // mente la misma medición y. Estas suposiciones se acercan/ bastante a la verdad en el caso de las velas que arden en recipientes cerrados. Son seguramente falsas en la experimentación biológica, en la que y puede ser una medi- / ción del peso o del azúcar en la sangre o del tiempo de supervivencia de un animal que recibe un tratamiento con una droga determinada. Las ecuaciones necesitan modifi-/ carse como:

$$y_1 = M + G + F + T_1 + \varepsilon_1 \qquad (1.3)$$

$$y_2 = M + G + F + T_2 + \varepsilon_2 \qquad (1.4)$$

Aquí G, F se relacionan ahora con los efectos inehrentes y ambientales *promedio* de una clase o población de sujetos muy similares pero inevitablemente no idénticos; /// ϵ_1, ϵ_2 , son medidas de la variación individual entre sujetos elegidos lo más homogéneo posibles, combinada con/ los efectos debidos a la variación en las característi-/ cas inherentes y ambientales relativas a la media.

Esencialmente, se presenta la misma situación en otras / muchas ramas de la ciencia y de la tecnología: por ejemplo,

Sujetos	Tratamientos	Mediciones
Pacientes de una enfermedad	Medicinas	Tiempo de recuperación
Niños	Métodos de enseñanza	Rendimiento en los test
Chapas metálicas	Métodos de protección	Cantidad de herrumbre después de un año
Automóviles	Tipos de gasolina	Consumo a los 100 km.
Parcelas de trigo	Fertilizantes	Cosecha

Ya no podemos afirmar que

$$T_1 - T_2 = y_1 - y_2 ,$$

y en general no sabemos nada acerca de los valores de // ϵ_1, ϵ_2. Se nos abren dos caminos:

(i) **Aleatorización**

Recurriendo al azar para determinar que sujeto recibe cada tratamiento, nos aseguramos que el error al utilizar $y_1 - y_2$ como el valor de T_1, T_2 es probablemente igual a $\epsilon_1 - \epsilon_2$ ó $\epsilon_2 - \epsilon_1$. Si el experimento fuera repetido muchas veces, como promedio sería correcto el valor obtenido para $T_1 - T_2$

(ii) **Replicación**

El principio de aleatorización es una de las mayo-/ res contribuciones de la Estadística a la investigación. Sin embargo, por sí mismo no elimina todas las dificultades. La replicación proporciona la respuesta. Asignando/ varios sujetos a cada tratamiento, se reducen las incertidumbres debidas a los "errores experimentales" . Un / tratamiento se evalúa ahora en términos de la media de y para los sujetos, y es aplicable el conocido resultado de las varianzas:

Var(media de r) = Var (observación individual)/r (1.5)

El diseño de experimentos, está relacionado con la explotación de estas ideas de modo que se utilicen los sujetos y materiales disponibles con el máximo aprovechamiento, reduciendo al mínimo la varianza de las medias de los tratamientos. En particular, las relaciones entre los sujetos pueden utilizarse para reducir la contribución del error debido a las diferencias en los caracteres inherentes y ambientales de los sujetos.

2. PARAMETROS

Formulemos ahora estas ideas de manera un poco más // exacta. Supongamos que se mide y para el sujeto número/ k de los de un grupo con clasificación genética y ambiental i que reciba el tratamiento j. Podemos volver a / escribir las ecuaciones anteriores como:

$$y_{ijk} = \mu + \beta_i + \tau_j + \varepsilon_{ijk} \qquad (2.1)$$

Aquí, μ es una media general, β_i abarca los G y F pre-/ vios y es la desviación de μ correspondiente al estado // promedio de la clase inherente y ambiental particular utilizada, y τ_j es la desviación adicional asociada con el / tratamiento j (denominado corrientemente *efecto* del tratamiento j); ϵ como antes es el error residual debido a la variabilidad dentro de la categoría i y a la variabilidad de los sujetos individuales, y se indica con ijk /

simplemente para mostrar su correspondencia con la observación y_{ijk}.

Nos referimos a τ_j (para j = 1, 2, ..., t si hay t tratamientos en discusión -es necesario que t no sea / 2) como *parámetros*.(Esta es una palabra utilizada en exceso hoy día tanto en la literatura médica como por los/ periodistas. Para el estadístico ha tenido desde hace // tiempo un significado preciso como un valor numérico que caracteriza una población o una formulación teórica u- / sualmente desconocida pero que necesita estimarse a partir de los datos). Además, de estos parámetros del tratamiento, los β_i son también parámetros relacionados con / el ambiente u otras características de antecedentes de / los sujetos; con frecuencia los denominaremos parámetros de *bloque*. También μ es un parámetro para la media general.

Supondremos además, que la esperanza de ϵ^2es la misma para todas las observaciones.

$$E(\epsilon^2) = \sigma^2 \tag{2.2}$$

siendo el parámetro de σ^2 la varianza (σ la desviación / típica) por observación. Esta suposición de varianza /// constante es usualmente razonable a menos que y sea excepcionalmente variable, en particular más variable para algunos tratamientos que para otros. Una suposición de / que ϵ tiene una distribución normal o Gaussiana *no* es necesaria para la exposición principal de diseño.

3. ESTIMACION

En la situación siempre prevista en la sección 1 no interviene el problema de la estimación. Allí se sobrentendía que todos los sujetos tienen el mismo β;es decir, β_1, por lo que para los sujetos en tratamientos diferentes tenemos

$$\left.\begin{aligned} y_{11k} &= \mu + \beta_1 + \tau_1 + \epsilon_{11k}, \\ y_{12k} &= \mu + \beta_1 + \tau_2 + \epsilon_{12k}, \\ y_{13k} &= \mu + \beta_1 + \tau_3 + \epsilon_{13k}, \end{aligned}\right\} \tag{3.1}$$

etc.

Evidentemente si formamos a su vez medias(promedios) para los sujetos en cada tratamiento, las diferencias en-/ tre estas medias estimarán las diferencias entre los parámetros τ correspondientes, ya que el $(\mu + \beta_1)$ es común a todos.

Veremos que podemos necesitar implicar a más de un / β en un experimento. Hay dos razones -reducir la varianza del error o ampliar la base de la inferencia. Si β_1 representa una clase muy amplia, por ejemplo, todas las edades del sujeto, ϵ se verá muy influenciado por la va-/ riación de sujeto a sujeto asociada con la edad. Restringiendo los sujetos de un recorrido de edad estrecho, se / reduce la variabilidad de σ^2y, por consiguiente, el valor de β. Reteniendo sujetos de varios grupos de edades distintos dentro de un experimento, esperamos proporcionar a los resultados una validez más amplia. Por ejemplo, en seres humanos podríamos tener sujetos de edades: 20-30, /// 30-40. 40-60, 60-80, cada uno con su propio β. Por otra / parte,en una comparación de fertilizantes para el trigo,/ podríamos desear una inferencia basada en varios tiempos/ diferentes de siembra.

Consideremos un ejemplo muy sencillo. Supongamos que deseamos estudiar tres tratamientos. Podríamos utilizar / cuatro grupos de edades 1 sujeto en cada uno de ellos:

Tratamiento (τ)	Edad (β) 1	2	3	4
1	(x)	x	x	x
2	x	x	x	x
3	x	x	x	(x)

En esta situación simétrica, podemos todavía estimar las diferencias entre los parámetros del tratamiento justo antes, promediando entre todos los grupos de edad. Los tratamientos están equilibrados con números iguales para cada grupo de edad. Lo mismo ocurrirá si tenemos, por ejemplo, 9 sujetos en el grupo de edad 2 con 3 en cada tratamiento. Pero supongamos ahora que se han utilizado sola-/ mente 10 sujetos -los dos entre ( ) no existen. Entonces el promediado simple de los restantes sujetos puede inducir a error. El promedio para el tratamiento 2, es para / todos los grupos de edad, pero el promedio para el tratamiento 1 omite los sujetos más jóvenes. Por consiguiente, si y es una medida que tiende a aumentar con la edad // con independencia del tratamiento (por ejemplo, la pre- / sión sanguínea), esta comparación simple será sesgada.

Sin embargo, tenemos valores de y para 10 sujetos y podemos describir la formulación parámetrica para cada // uno de ellos:

$$\left.\begin{aligned} y_{211} &= \mu + \beta_2 + \tau_1 + \epsilon_{211}, \\ &\;\;\vdots \\ y_{331} &= \mu + \beta_3 + \tau_3 + \epsilon_{331}, \end{aligned}\right\} \tag{3.2}$$

etc.

(k = 1 en todos los casos, ya que solamente tenemos un sujeto en cada "casillero" de la tabla). ¿Cómo elegimos valores númericos para los parámetros con el fin de optimizar la armonía entre las observaciones y los parámetros?. Un principio estadístico ampliamente aceptado es el de los Mínimos Cuadrados: estimar los parámetros de tal manera que se haga la suma de los cuadrados de los restos lo más pequeña posible (Finney 1980, Yates,1933). Si m, b_1, t_j son estimaciones de los μ, β_i, τ_j, correspondientes, los restos se definen como

$$\left.\begin{aligned} e_{211} &= y_{211} - m - b_2 - t_1 \;, \\ e_{331} &= y_{331} - m - b_3 - t_3 \;, \end{aligned}\right\} \qquad (3.3)$$

etc.

El método de estimación es minimizar la suma de e^2 mediante la elección adecuada de valores númericos para los parámetros. No puedo entrar en detalles sobre el principio general, salvo decir que tiene muchas propiedades teóricas deseables: es insesgado, proporciona estimaciones con varianza mínima, y , si los errores tienen una distribución normal, es equivalente al método totalmente eficiente de máxima verosimilitud.

Por consiguiente escribiríamos

$$\begin{aligned} S = (y_{211} - \mu - \beta_2 - \tau_1)^2 + (y_{311} - \mu - \beta_3 - \tau_1)^2 \\ + (y_{411} - \mu - \beta_4 - \tau_1)^2 + \ldots\ldots \\ + \ldots\ldots + (y_{331} - \mu - \beta_3 - \tau_3)^2 \end{aligned} \qquad (3.3)$$

y seguidamente minizariamos s mediante la elección apropiada de los valores de los parámetros.

En la sencilla situación simétrica que mencioné primero, para cada sujeto del grupo de edad 1 (o más en general para un número constante de sujetos) en cada tratamiento, podemos probar facilmente que la minimización de S_1 se reduce a un simple promediado. Si m, b_i, t_j representan ahora estos estimadores mínimos cuadráticos, m es la media de todas las observaciones, $(m + b_i)$ es la media de todas las del grupo de edad i, $(m + t_j)$ es la media general en el tratamiento j. En ausencia de simetría y equilibrio, los valores de m, b_i, t_j son menos obvios y deben obtenerse resolviendo conjuntos de ecuaciones lineales. Pero ahora tenemos un procedimiento numérico manejable y sensible para todos los casos, aunque haya dejado algunos detalles sin explicar. Estas ideas se repetirán en conferencias posteriores. El método de mínimos cuadrados sustenta todo el análisis de la varianza y la metodología de regresión múltiple, pero no insistiremos mucho en ello.

4. ALEATORIZACION COMPLETA

Aquí y en la Conferencia II, presumo que están algo familiarizados con el análisis de la varianza. Expondré la estructura general de este análisis y el concepto esencial de ortogonalidad en la Confernecia III.

El diseño experimental más sencillo, es aquel en que los sujetos disponibles se asignan completamente al azar entre los t tratamientos, decidiendo simplemente el experimentador cuantos corresponderán a cada tratamiento. Por ejemplo, con t = 5 y 43 sujetos disponibles el experimentador podría elegir al azar lotes con el fin de asignar 11 sujetos al primer tratamiento, 6 al segundo, 14 al tercero, 4 y 8 al cuarto y quinto. Explicaré la elección de los números en la Conferencia III.

La Tabla 4.1 muestra porcentajes de fosfato de calcio en pollos en cuatro preparados distintos de vitamina D. Las conclusiones parecen evidentes, pero los datos pueden explicar los cálculos. Podemos calcular una suma de los cuadrados de las desviaciones dentro de cada tratamiento:

$$T_1 : \; 5.2^2 + 6.9^2 + \ldots\ldots + 6.1^2 - \frac{33.3^2}{7}$$

$$= 18.54 \quad (6gl)$$

$$T_2 : \; 43.56 \quad (5gl)$$

$$T_3 : \; 53.92 \quad (7gl)$$

$$T_4 : \; 29.71 \quad (6gl)$$

(deliberadamente no estoy comprobando todo el cálculo aritmético comprobarlo y corregirlo aquí y en cualquier lugar, es un buen ejercicio para el estudiante). Combinando (agrupando) toda la evidencia sobre la varianza se obtiene la estimación:

$$s^2 = (18.54 + 43.56 + 53.92 + 29.71)/(6 + 5 + 7 + 6)$$
$$= 6.072 \quad (24gl)$$

La varianza de la media de un tratamiento es s^2/r_1, siendo r_i el número de sujetos. En la Tabla 4.2 se muestran las medias y los errores típicos. Podemos formar los errores típicos de las diferencias, estimar cualquier $\tau_i - \tau_j$ y poner límites de probabilidad en los mismos, hacer dócimas de significación y así sucesivamente a voluntad. Comento solamente que la diferencia entre T_1, T_2 y T_3, T_4 es incuestionable, mientras que las diferencias dentro de estos pares son apreciablemente menores que el doble de sus errores típicos.

Consideremos un método alternativo de cálculo que aquí es poco ventajoso, pero importante para el futuro.

(i) Formar la suma total de los cuadrados de las desviaciones para un total de 28 observaciones:

$$5.0^2 + 6.9^2 + ... + 13.1^2 - (33.3 + 42.9 + 106.0 + 106.5)^2/28 = 3635.69 - 2976.70 = 658.99 .$$

(ii) Formar la suma de los cuadrados "entre tratamien- / tos" (explicación en la Conferencia III):

$$\frac{33.3^2}{7} + \frac{42.9^2}{6} + \frac{106.0^3}{8} + \frac{106.5^2}{7} - \frac{288.7^2}{28} = 513.27 .$$

Insertar estos valores en la tabla 4.3, el análisis de la varianza, y obtener la suma de cuadrados del error por substracción. Se ve fácilmente que esto da exactamente el mismo cuadrado medio del error s^2 que antes, exceptuando el redondeo aritmético.

La comparación de los cuadrados medios de los trata-/ mientos y del error por medio de una dócima de razón de / varianzas, proporciona una dócima de significación de la hipótesis nula "todos los τ son iguales". Estrictamen-/ te, la dócima es válida solamente si las ϵ están distri-/ buidas normalmente, pero en la práctica esta restricción/ no importa mucho. Ciertamente aquí, donde SF = 28.2 con 3 y 24 grados de libertad no hay duda acerca de la significación. Sin embargo, este tipo de dócima rara vez es im-/ portante. En la mayoría de los experimentos el valor primario del análisis de la varianza es como procedimiento / para obtener el cuadrado medio del error, siendo esta la varianza básica para utilizarla en determinadas dócimas / de significación al expresar la precisión de las estima-/ ciones de los parámetros y las comparaciones entre los / parámetros, y al calcular los límites del error con las / probabilidades establecidas.

TABLA 4.1

Comparación de cuatro preparados de vitamina D
(Porcentaje de fosfato de calcio en pollos)

Tratamientos	T_1	T_2	T_3	T_4
Porcentajes	5.0	5.6	8.7	17.0
	6.9	9.0	11.9	16.3
	4.6	7.8	13.3	17.2
	5.7	2.2	15.9	14.8
	1.8	7.6	16.9	16.7
	3.2	10.7	15.8	11.4
	6.1		11.9	13.1
			11.6	
Totales	33.3	42.9	106.0	106.5

TABLA 4.2

Resumen de las medias para la Tabla 1.1

Tratamiento	T_1	T_2	T_3	T_4
Media	4.8	7.2	13.2	15.2
SE (Error típico)	± 0.93	± 1.01	± 0.87	± 0.93

TABLA 4.3

Análisis de la Varianza para la Tabla 1.1

Ajuste a la media		2976.70	
Variación	gl	Suma de los cuadrados	Media cuadrática
Tratamientos	3	513.27	171.09
Error	24	145.72	6.072
Total	27	658.99	

II. BLOQUES ALEATORIZADOS; ACCIDENTES Y PLANES DE SALVAMENTO

1. ALEATORIZACION COMPLETA: EL PLAN EXPERIMENTAL

En la Sección I 4, se utilizaban números totalmente/ arbitrarios de sujetos por tratamiento. ¿Cómo deberían / elegirse estos números?. Pueden contemplarse muchas situaciones diferentes. Aquí solamente se expondrán dos.

Supongamos que el experimentador está preparado para utilizar N sujetos en total, r_i para el tratamiento Ti para i = 1, 2, ... t. Por consiguiente

$$r_1 + r_2 + \ldots\ldots + r_t = N \qquad (1.1)$$

Si todos los tratamientos son igualmente interesan-/ tes se podría desear minimizar la varianza promedio de / las diferencias entre pares de tratamiento. Esto supone/ que el valor promedio de

$$\frac{1}{r_i} + \frac{1}{r_j} \qquad \text{para todos los pares } i,j$$

debe minimizarse sujeto a la ecuación (1.1). Un pequeño/ cálculo diferencial muestra que el mínimo está dado por

$$r_i = N/t \qquad (1.2)$$

para todo i. Es decir, repetir todos los tratamientos // igualmente. Por supuesto, N, puede no ser un múltiplo // exacto de t, en cuyo caso los tratamientos deben tener / un sujeto adicional.

Supongamos a continuación que T_1 es un tratamiento / estandar con el cual deben compararse cada uno de los // tratamientos restantes, pero que no presta atención a las diferencias entre los otros. Lo importante es minimizar/ el valor promedio de

$$\frac{1}{r_1} + \frac{1}{r_i} \qquad \text{para todo i excepto para i = 1,}$$

todavía sujeto a la ecuación (1.1). La respuesta es me-/ nos evidente aunque una prueba bastante sencilla conduce a

$$r_1 = \frac{N}{1+\sqrt{(t-1)}} \,,$$
$$r_2 = r_3 = \ldots\ldots = r_1/\sqrt{(t-1)}. \qquad (1.3)$$

En otras palabras T_1 debería repetirse en mayor medida que los otros por un factor $\sqrt{(t-1)}$. Esto rara vez / puede lograrse exactamente en números enteros, pero una / pequeña desviación supone poca diferencia.

Por consiguiente, en el experimento de la vitamina D de la Tabla I 4, vemos que si todas las comparaciones hubieran sido de igual interés, el diseño óptimo habría tenido 7 pollos en cada tratamiento. Por otra parte, si sólo hubiera habido interés en comparar T_1 con cada uno de/ los otros tres $r_1 = 10$, $r_2 = r_3 = r_4 = 6$ habría sido un / plan mejor (10/6 se aproxima a $\sqrt{3}$) Desde luego, no hay / razón para esperar que s^2 dependa de r_i. Las alternati-/ vas aquí no son muy diferentes. La primera da una varianza de 0,29 s^2 para cada diferencia entre dos de los T_j, la segunda de una varianza de 0,27 s^2 para T_1 con cualquiera de las otras, 0,33 s^2 para cualquier par de T_2, T_3 y T_4./ En la práctica, cualquier ventaja partiendo de la igual-/ dad de los r_i es pequeña a menos que haya diez o más tratamientos.

2. BLOQUES ALEATORIZADOS

En la Tabla 2.1, se muestran los pesos uterinos de ratas ovariectomizadas que habían recibido uno de cuatro / preparados de oestrona. Los animales disponibles eran cuatro hembras de cada una de siete camadas. La aleatoriza-/ ción se hizo dentro de las camadas: los cuatro tratamientos fueron asignados aleatoriamente cada uno a una rata / de la camada I, cada uno a una de la camada II y así sucesivamente. Este es un diseño de *bloque aleatorizado*. La // idoneidad de la ecuación I (2.1) sería evidente con la / excepción que siempre es k = 1 y puede omitirse: para la rata en el tratamiento j de la camada i,

$$y_{ij} = \mu + \beta_i + \tau_j + \varepsilon_{ij} \,. \qquad (2.1)$$

¿Por qué hacer esto? ¡Nadie podría obtener 28 ratas de // una sola camada!. Un diseño completamente aleatorizado sería legítimo, pero entonces los parámetros del bloque (camada) estarían combinados con los componentes de error y por consiguiente aumentarían el error efectivo.

El análisis de la varianza (tabla 2.2) comienza ahora a mostrar sus méritos. La suma total de cuadrados:

$$0.54^2 + 0.49^2 + \ldots + 1.08^2 - (24.19)^2/28 = 2.5757$$

se encuentra sin dificultad. (Otra vez les dejo que com-- prueben mi cálculo aritmético). La suma de cuadrados para los tratamientos requieren el mismo cálculo que en la Sección I 1, ahora algo más sencillo porque todos se repiten por igual:

$$(4.64^2 + 7.39^2 + 5.19^2 + 6.97^2)/7 - 24.19^2/28 = 0.7671$$

De manera semejante, puede formarse una suma de cuadrados para los bloques:

$$(3.59^2 + 2.57^2 + ... + 4.75^2)/4 - 24.19^2/28 = 0.9232$$

A causa del equilibrio del diseño -cada tratamiento aparece el mismo número de veces en cada bloque- estas dos sumas de cuadrados son independientes (en el sentido que se va a explicar en la Conferencia III, el ortogonal): Ambas pueden restarse del total para dejar el error con(27-6-3) gl., $s^2 = 0.04919$.

La dócima de significación para los tratamientos

$F = 5.20$ con 3 y 18 grados de libertad

deja poca duda acerca de que las diferencias sean reales. El resúmen de la tabla 2.3, es mucho más ordenado que el de la Tabla I4.2 debido a la equi-repetición, y la media de cada tratamiento tiene la varianza $s^2/7=(0.083)^2$. Una diferencia entre las medias de dos tratamientos, tiene como varianza $2s^2/7$; la multiplicación del error típico correspondiente por 2.10 da 0.249 como incertidumbre asociada a cualquier diferencia estimada con una probabilidad de 0.95 para t con 18 g.l. Por ejemplo, $\tau_2 - \tau_1$, la diferencia en los parámetros entre T_2 y T_1 se estima como 0.393 y con una probabilidad de 0.95 comprobamos que el verdadero valor está entre 0.144 y 0.642.

La simetría de este diseño es tal que las estimaciones de los parámetros obtenidas mediante mínimos cuadrados son idénticas a las resultantes del evidente y no crítico promediado de los datos. Por consiguiente, para los que (2.1) es todavía apropiado, los cálculos son (según se indicaba ya en la sección I 3)

m = media de todos los ht valores de y (2.2)

b_i = (media de todos los t valores del bloque i)-m (2.3)

t_i = (media de todos los h valores del tratamiento i) - m (2.4)

Para alguno de los diseños que explico, más adelante, en especial en la Conferencia V el cálculo es menos evidente.

3. DOS COMENTARIOS SOBRE LA VARIANZA DEL ERROR

El análisis de la varianza permite encontrar con mucha facilidad s^2, aunque el procedimiento pueda parecer indirecto. Es posible efectuar un cálculo directo, que conduzca a la misma respuesta exactamente, pero esto es mucho más laborioso.

Supongamos que se hubiera utilizado un diseño completamente aleatorizado con estas 28 ratas. La variabilidad entre camadas habría tenido entonces error experimental. Podemos estimar que el cuadrado medio del error habría sido:

$$(0.9232 + 0.04919 \times 21)/27 = 0.07245$$

aproximadamente un 50% mayor que s^2. Esto indica que habrían sido necesarios aproximadamente 1,5 veces más ratas, es decir, 11 por preparado para obtener la misma precisión . (Los mismos errores típicos de las medias).

4. COMPARACIONES EMPAREJADAS

Una forma de bloques aleatorizados utilizada corrientemente para experimentos de estructura simple es la de bloques de 2, conocida como comparaciones emparejadas.Un tratamiento puede ser comparado con el estado de control, sin tratar, sobre pares de sujetos tales como dos conejos de la misma camada, terneras gemelas idénticas, lados izquierdo y derecho del cuerpo humano, pares de pacientes de igual edad, sexo, y gravedad de enfermedad, y así sucesivamente. La aleatorización es exactamente como para otros bloques aleatorizados.

Por supuesto, los resultados pueden analizarse exactamente como en la sección 2. Una alternativa consiste en formar la diferencia entre los pares de los valores y, "control-tratado" para cada bloque, y a continuación efectuar un cálculo directo de la varianza de las diferencias. Esto se describe en muchos libros de texto elementales.

Los dos métodos de cálculo, conducen a resultados idénticos, como puede verificarse algebraicamente con facilidad.

5. COMENTARIO GENERAL

Los diseños completamente aleatorizados de bloques aleatorizados son indudablemente los dos diseños experimentales más importantes y más ampliamente utilizados.Ambos son de aplicabilidad muy amplia en casi todo campo de experimentación cuantitativa. Además, la mayoría de los otros diseños, de algunos de los cuales voy a hablar en conferencias posteriores, son generalizaciones y extensiones de estas ideas.

6. ANALISIS DE COVARIANZA

Supongamos que correspondiendo a cada medición de y, hay también una medición de una variable aleatoria x que se sabe que no ha sido afectada por el tratamiento. La / situación más satisfactoria es aquella en que se midió / antes de la asignación aleatorizada de los tratamientos. Por ejemplo, x puede ser el peso de un animal antes del comienzo de un experimento,y puede ser tal vez el peso / total 10 semanas después del tratamiento o el peso de un órgano determinado en este último momento. Alternativa-/ mente, x puede ser la cosecha de fruta de un árbol en 1984 después de haber comenzado un experimento para comparar los métodos de poda. Aunque dicha x es a menudo/ una medición de la misma clase que la y subsiguiente, no es necesario que lo sea -por ejemplo, pueda ser la altura del árbol frutal; la única característica esencial es que la conexión causal entre x y el tratamiento puede/ ser excluida logicamente.

Si puede elegirse y medirse una x adecuada, éste / puede contener información sobre dicho componente de / la variabilidad en y que no se debe al tratamiento. // Una aproximación razonable, consiste en modificar la /// ecuación (2.1) como

$$y_{ijk} = \mu + \beta_i + \tau_j + \theta(x_{ijk} - \bar{x}) + \epsilon_{ijk} \qquad (6.1)$$

(y modificaciones semejantes para otros diseños experi-/ mentales), siendo θ un parámetro adicional (estimado por mínimos cuadrados, por supuesto) y $\bar{x}$ la media de todos las x_{ijk}. La consecuencia de incluir la estimación de θ es la de corregir la estimación de los parámetros del // tratamiento τ_j para valores que representan una igualación con respecto a x.

Este procedimiento se conoce como *análisis de cova-/ rianza*. El cálculo se hace más fácilmente añadiendo a la tabla del análisis de varianza dos columnas más: Una es/ el análisis de x^2, exactamente como el de y^2, y la se-/ gunda es un análisis de la misma forma efectuado sobre / los productos xy en lugar x^2 ó y^2. Por consiguiente, θ se estima como

θ = Suma de productos del error / Suma de cuadrados del error para x. (6.2)

Cada t_j para y puede ajustarse ahora mediante:

t_j (ajustado) = t_j (para y) - θ t_j (para x) (6.3)

Hay una relación evidente y estrecha con la regre-// sión lineal, que permite que las varianzas de las comparaciones de las t_j ajustadas se formen considerando los contrastes lineales apropiados. (Conferencia III). Co-/ chran & Cox (1.957) dan ejemplos excelentes. No diré más. El método es valioso, e insuficientemente aprovechado para usarlo con variables aleatorias x que se registran / fácilmente como parte del proceso experimental. Por analogía con la regresión lineal múltiple, pueden utilizarse / dos o más variables aleatorias x distintas en un análisis de covarianza múltiple. He introducido el tema principalmente para poder utililizar el análisis de covarianza/ en las Secciones 7, 8.

7. OBSERVACIONES PERDIDAS

Incluso el experimentador más cuidadoso pierde ocasionalmente el valor de y en una parcela. Tal vez, animales/ que pacen invaden una parcela de trigo, un rayo cae sobre un árbol, o una rata de un experimento muere accidental-/ mente -incidentes todos que no deberían relacionarse con los tratamientos aplicados-. ¿Que es lo que hay que ha-// cer?. El método de mínimos cuadrados es la forma más apropiada de manejar la falta de equilibrio. Si el experimento está completamente aleatorizado, no hay problema: simplemente se considera que el tratamiento implicado tiene/ una parcela menos y se realizan los cálculos normales. Para bloques aleatorizados y otros diseños, el método de mínimos cuadrados puede utilizarse de tres maneras, todas / equivalentes.

(i) Escribir z para el "valor perdido" de y. Hacer un análisis de varianza en función de z, y hallar la / suma de cuadrados del error en la forma

$$Az^2 + Bz + C \qquad (A > 0) \qquad (7.1)$$

Este se minimizará mediante:

$$\hat{z} = -B/2A \qquad (7.2)$$

Introducir este valor numérico de z en el lugar de la y perdida, reducir en 1 el número de grados de/ libertad para el error, y el análisis de varianza // completo dará el valor apropiado de s^2 como una estimación insesgada σ^2.

Para un diseño de bloques aleatorizados con b blo-// ques de tratamientos esto conduce a

$$\hat{z} = \frac{bB + tT - G}{(b-1)(t-1)} \qquad (7.3)$$

siendo B el total de todas las demás y en el bloque/ del que falta una,T el total de todas las demás y para el tratamiento del que falta uno, y G el total de los (bt - 1) valores registrados de y. Pueden encontrarse fórmulas análogas para otros diseños.

(ii) Imaginar un valor para z, y hallar las estimaciones de todos los parámetros m, b_i, t_j. Luego determinar

$$m + b_i + t_j$$

para la "parcela perdida", colocándole en lugar del primitivo valor de z y repetir el ciclo hasta que z permanezca constante. Esto asegura que el residuo (sección 13) para la parcela perdida es cero. Se obtendrá exactamente la misma $\bar{z}$ que en (I).

Aunque (i) y (ii) conducen a la estimación insesgada de σ^2 las varianzas y errores típicos de las diferencias en el t_j requieren más cuidados. El valor de $\bar{z}$ es, como en (7.3), una función lineal de todas las demás y, y en consecuencia las diferencias de / las varianzas de los tratamientos son mayores que / si los datos estuvieran completos. Un ajuste rudi-/ mentario, aunque poco riguroso, consiste simplemente en considerar que el tratamiento afectado tiene/ una replicación menos.

(iii) Introducir un valor arbitrario en la posición perdida. Una elección adecuada es la media general de // todos los valores de y de las otras parcelas, pero el resultado consiguiente es el mismo sea cual / fuese el elegido. Después definir x como una "variable aleatoria ficticia" que toma el valor $(N-1)$/ para la parcela perdida, $y-1$ para todos las demás, siendo N el número total de parcelas. Hacer un análisis de covarianza de y sobre x. Al ajustar t_j se habrá tenido en cuenta el valor perdido, y al aplicar los métodos típicos de análisis de covarianza / -regresión lineal, se prestará la debida atención a las varianzas y errores típicos.

Los métodos (i), (ii), (iii) pueden generalizarse para / abordar los casos en que falten con dos o más observa-// ciones. Debemos tener definidos por separado z_1, z_2,..., perder tantos grados de libertad de s^2 como parcelas perdidas e introducir una variable aleatoria ficticia diferente x_1, x_2,... para cada parcela perdida.

Aunque el método (i) es conveniente para los casos sencillos en que se conoce la fórmula tal como (7.3), el métoto (iii) tiene la ventaja de su generalidad. Se aplica a cualquier diseño, por muy completo que sea, y se maneja fácilmente con la ayuda un fichero general de computa- / ción que incluya análisis de covarianza.

Obsérvese la necesidad lógica de que la pérdida de un valor será casualmente independiente del tratamiento/ aplicado. Si la cosecha de un árbol frutal se pierde porque su tratamiento ha estimulado la maduración temprana/ de la fruta y su consiguiente destrucción por los pája-/ ros, o si una rata muere porque su dieta experimental // era deficiente en algún componente vital, la utilización de cualquiera de los métodos de esta Sección debe producir resultados sesgados. Si una dieta particular bajo ensayo introduce un riesgo grave de que las ratas mueran / antes de que alcancen los seis meses de edad, ¿que significado tendría estimar el peso esperado a los doce meses para las ratas con dicha dieta?. En tales circunstan- / cias, el método estadístico solo no puede servir de ayuda: debe de ser examinado de nuevo el concepto y propósito completo del experimento.

8. ALGUNOS OTROS ACCIDENTES

El método de los mínimos cuadrados y el análisis de/ covarianza son ayudas poderosas para salvar aquellos experimentos que han ido mal. Por ejemplo, si se conoce // que la cosecha de dos parcelas de terreno se ha mezclado en la recolección, de manera que solamente se conoce el/ total de los dos valores de y, cualquiera de los métodos (i), (ii), (iii) puede adaptarse a este problema.

Federer & Schlottfeldt (1954) y Outhwaite & Rutherford (1955) estudiaron las alturas de las plantas registradas de siete tratamientos en ocho bloques aleatorizados. Los bloques estaban adosados y las siete parcelas / de cada bloque estaban en una sóla línea. Por desgracia, la planificación de los bloques no tuvo en cuenta adecuadamente las tendencias de la fertilidad del terreno,y la precisión del experimento se redujo por un gradiante // conveniente de la fertilidad de parcela a parcela dentro de los bloques. Los autores recuperaron información por/ covarianza sobre una varianza aleatoria ficticia. El siguiente paso consistió en definir x_1 como una variable/ aleatoria que aumentaba linealmente de una parcela a /// otra dentro de cada bloque, es decir, $x_1 = 0, 1, 2,...,6$ y utilizar ésta como covariable aletoria. Extendiendo el análisis para incluir una serie de seis variables aleatorias ficticias elegidas cuidadosamente, puede eliminarse toda tendencia (lineal o no), de parcela a parcela a lo/ largo de los bloques. El experimento habría sido más preciso, si los bloques se hubieran basado sobre la tendencia real (o se hubiera utilizado un diseño de cuadrado / latino de doble entrada, pero si la tendencia no se conocía con anticipación, el análisis de covarianza propor-/ ciona una operación de salvamento efectiva.

Hace algunos años se me pidió que ayudara en un experimento sobre variedades de cereales, formado por 36/ parcelas adosadas largas y estrechas (Finney 1.962). Todas las esquinas de las parcelas estaban marcadas por / estaquillas en el terreno. La estaquilla colocada en // una esquina del experimento completo se quitó accidentalmente. Antes de que se descubriera el accidente se / recolectó la cosecha como de 35 paralelogramos (¡Casi / rectangular!). Cada parcela recolectada contenía mitades triangulares de dos parcelas reales adyacentes, lo / cual permitió que cada cosecha observada se expresase en función de un parámetro de bloque y una suma de parámetros de dos tratamientos. La estimación por mínimos cuadrados fue sorprendentemente sencilla (¡yo no tenía un / computador entonces!. No es sorprendente que la pérdida/ de información relativa al experimento completo fuera // grande, siendo las varianzas aproximadamente el doble de lo que debieran haber sido, pero el hecho importante es/ que se salvo algo útil de un experimento que de otra manera tenía que haberse desechado.

TABLA 2.1

Respuesta a cuatro preparados de oestrona

(Pesos uterinos de ratas, mg por g de peso del cuerpo)

Camada	T_1	T_2	T_3	T_4	Totales
I	0,54	1,52	0,61	0,92	3,59
II	0,49	0,71	0,74	0,63	2,57
III	0,51	1,12	0,51	1,07	3,21
IV	0,40	0,58	0,60	1,02	2,60
V	0,81	1,02	1,07	1,20	4,10
VI	0,63	0,86	0,83	1,05	3,37
VII	1,26	1,58	0,83	1,08	4,75
Totales	4,64	7,39	5,19	6,97	24,19

(Los resultados de este experimento se han modificado ligeramente para hacerlos más adecuados como ejemplo ilustrativo).

TABLA 2.2

Análisis de la varianza para la Tabla 2.1

	g.l.	Suma de cuadrados	Cuadrado medio
Ajuste para la Variación media		20.8984	
Camadas (bloques)	6	0.9232	
Tratamientos	3	0.7671	0.2557
Error	18	0.8854	0.04919
	27	2.5757	

TABLA 2.3

Resumen de las medias para la Tabla 2.1

Tratamientos	T_1	T_2	T_3	T_4	
Medias	0.663	1.056	0.741	0.996	$\pm$ 0.084

III. ANALISIS DE LA VARIANZA

1. INTRODUCION

La presentación de la teoría del analísis de la / varianza en términos de álgebra matricial, es importante tanto por lo conciso de la teoría como por la organización formal de programas de ordenador. Es dudoso que / ésta proporcione el mismo conocimiento de la estructura del análisis que la que tenían los que primero desarrollaron el tema del diseño de experimentos. Los // primeros trabajos de Fisher y Yates son notables por / la facilidad con que presentaron nuevas familias de diseños, y luego, sin vacilar escribieron los pasos aritméticos para un análisis de la varianza bien organizado; merece la pena leer estos trabajos.

Los científicos no especializados en la manipulación de matrices tropiezan a menudo con dificultades para comprender las fórmulas en que se basa el análisis de la varianza. Sin embargo, la teoría puede expresarse por completo en términos de álgebra bastante / elemental, dentro de la comprensión de cualquiera que se niegue a ser intimidado por expresiones y ecuaciones que implican varios subíndices. Esta conferencia describe bastante formalmente una presentación que he comprobado desde hace tiempo que es útil en la enseñanza. Aunque no contiene nada nuevo, la exposición y demostraciones temáticas a lo largo de estas líneas, no parecen ser fácilmente accesibles en otras partes. El / enfoque se basa en una conferencia dada por R.A. /// Fisher en la Estación Experimental de Rothamsted hacía 1.942, pero he introducido muchos más detalles. Estoy/ muy al tanto de que teoría y demostración pueden expresarse más concisamente en términos matriciales, pero / jamás he encontrado que ésto sea tan instructivo.

Me preocupa en particular el no matemático que es razonablemente competente en la manipulación algebraica, pero incluso el estadístico profesional puede/ mejorar su manejo del análisis de la varianza familiarizandose con simples propiedades de los contrastes. Supongo que conocen las técnicas numéricas, por lo menos para los análisis más sencillos de datos de una y dos entradas, aunque puedan no estar seguros de cómo funcionan los métodos. Si los numerosos sufijos y sumas / les inducen a confusión, podrían escribir la teoría // exactamente como esta expuesta pero con valores enteros pequeños en lugar de números generales de observaciones.

Nada de lo que sigue pretende desaconsejar el uso apropiado de técnicas algebraicas más avanzadas, se intenta solamente suplementar éstas con un enfoque elemental que puede ayudar a algunas personas a apreciar la lógica subyacente. Tengan en cuenta también que nada en esta conferencia depende de ninguna hipótesis de normalidad.

2. NOTACION

Supongamos y_i para $i = 1, 2, \ldots n$ siendo n observaciones independientes de una variable aleatoria y. Definamos Uj como una función lineal de y_i, arbitraria excepto por la restricción de que al menos uno de los coeficientes a_{ji} sea distinto de cero.

$$\left.\begin{aligned} U_j &= a_{j1}y_1 + a_{j2}y_2 + \ldots\ldots + a_{jn}y_n \\ &= \sum_{i=1}^{n} a_{ji} y_i . \end{aligned}\right\} \qquad (2.1)$$

Hasta nueva advertencia , la suma $\sum\limits_{i=1}^{n}$ se escribirá simplemente como Σ ; todas las demás sumas tendrán sus límites establecidos.

Una función lineal particularmente importante es / la suma simple de y_i; esta se denominará U_0, siendo

$$U_0 = \Sigma y_i \quad (a_{0i} = 1 \text{ for all } i) . \qquad (2.2)$$

Es evidente que la media de todas las observaciones es

$$\bar{y} = U_0/n \qquad (2.3)$$

3. UN TEOREMA FUNDAMENTAL

El análisis de la varianza, depende del siguiente/ teorema algebraico general:

Supongamos que p funciones lineales $U_1, U_2, \ldots, U_p$ se definen como en la Sección 2. La ecuación

$$\Sigma y_i^2 = \sum_{j=1}^{p} U_j^2 \qquad (3.1)$$

será cierta para todo conjunto posible de valores númericos de y_i *si, y solamente si* se satisfacen / las tres condiciones

$$p = n , \quad (3.2)$$

$$\Sigma a_{ji}^2 = 1 \text{ para } j=1, 2, \ldots, p , \quad (3.3)$$

$$\Sigma a_{ji} a_{ki} = 0 \text{ para cada par de } j, k \text{ desiguales} \quad (3.4)$$

La demostración no es difícil, pero se trata de / un tipo de matemática formal que no necesito incluir.El teorema está estrechamente relacionado con el hecho // bien conocido de que un conjunto de p ecuaciones lineales independientes con n incógnitas, tendrá una solución única solamente si p = n. Tengan presente que(3.4) es la clave. La ecuación (3.2) exige simplemente el número correcto de componentes. La ecuación (3.3) puede / obtenerse por una simple normalización dividiendo todos los coeficientes de U_j por una cantidad apropiada. Realmente, una exposición posiblemente más sencilla del teorema es que

$$\sum_{j=1}^{n} (U_j^2/D_j) = \Sigma y_i^2 , \quad (3.5)$$

donde D_j se define por

$$D_j = \Sigma a_{ji}^2 , \quad (3.6)$$

que será verdad para todo valor de y_i posible si y solamente si se satisface (3.4). La ecuación (3.4) se dice que es la condición de que U_j, U_k sean *ortogonales*. Esta terminología es de orígen geométrico. Con relación a los ejes y_i en n dimensiones, $U_j = 0$ y $U_k = 0$ son las ecuaciones de dos hiperplanos (líneas rectas si n = 2, planos ordinarios si n = 3) y (3.4) es la condición // que los dos se corten en ángulos rectos.

Si $n>1$, existe mucha libertad en la definición / de U_j. Una de ellas, U_n por ejemplo, puede definirse / con bastante arbitrariedad. La ecuación (3.4) con n-1,n para j, k da una restricción lineal sobre los coeficientes de U_{n-1}: Todas las $a_{n-1,i}$ excepto una pueden ele-- girse arbitrariamente y se determina la última. La ecuación (3.4), primero con n-2, n para j, k y de nuevo con n-2, n-1 para j, k, da dos restricciones lineales sobre los coeficientes de U_{n-2}: (n-2) de los $a_{n-2,i}$ pueden // elegirse arbitrariamente y se determinan las dos últi-- mas. Así se procede hasta que solo puede hacerse una // elección arbitraria de un coeficiente en U_1; este último vale simplemente para elegir un factor miltiplicativo para todas las a_{1i}, que deja $U_1^2/\Sigma a_{j1}^2$ invariable

4. CONTRASTES

Para propósitos estadísticos, elegimos siempre U_o en la ecuación (2.2) como una de las funciones; ahora utilizaré U_o en lugar de U_n.Seguidamente, con la ayuda de otro resultado bien conocido

$$\Sigma(y_i-\bar{y})^2 = \Sigma y_i^2 - \frac{U_o^2}{n} = \sum_{j=1}^{n-1} (U_j^2/D_j) \quad (4.1)$$

sujeto a las condiciones apropiadas. La ortogonalidad de una y otra U_j con U_o requiere que

$$\Sigma a_{ji} = 0 \text{ para cada una de } j = 1, 2,\ldots,(n-1) \quad (4.2)$$

Cualquier función lineal que satisfaga (4.2), siendo suma de sus coeficientes cero, se dice que es un contraste de y_i. Por consiguiente la notación U_j (excepto U_o) se restringirá a los contrastes; D_j puede llamarse el "*divisor del contraste*" y

$$Q_j = U_j^2/D_j \quad (4.3)$$

el *cuadrado del contraste.*

La forma estadística del teorema puede exponerse /// ahora:

> Sean U_j, para j = 1, 2, ..., (n-1), contrastes entre las Y_i, y Q_j los cuadrados de los contrastes , correspondientes. Entonces la ecuación
>
> $$\Sigma(y_i-y)^2 = \sum_{j=1}^{n-1} Q_j$$
>
> se verificará para todo conjunto posible de valores numéricos para las y_i, si y solamente si cada par de contrastes tiene coeficientes que sa-- tisfacen la condición de ortogonalidad (3.4).

La técnica del análisis de la varianza consiste en elegir los contrastes que sean más adecuados para los objetivos de una determinada investigación. Obsérvese que a cada U_j y Q_j le corresponde *1 grado de libertad.*

5. ALGUNOS EJEMPLOS

Mediante un argumento semejante al del final de la sección 3, si se especifica cualquier número de contrastes mutuamente ortogonales i menor que (n-1) puede ha- / llarse uno adicional, y el proceso puede repetirse hasta que se obtenga un conjunto de (n-1). Para contrastes/ supuestamente ortogonales se escribe U_1, U_2,..., U_j. Las condiciones de que una función lineal U_{j+1} sea un con- / traste y sea ortogonal con U_1, U_2,..., da (j+1) ecuacines lineales que deben satisfacerse por $a_{j+1,i}$. Si / $j < n-1$, éstas tendrán una solución con algunos de los coeficientes no nulos.

Un ejemplo aclara el proceso. Supongamos $n = 4$. Tomemos el contraste arbitrario

$$U_1 = 4y_1 - 3y_2 + 2y_3 - 3y_4 \tag{5.1}$$

Entonces U_2 debe tener

$$a_{21} + a_{22} + a_{23} + a_{24} = 0\ ,$$
$$4a_{21} - 3a_{22} + 2a_{23} - 3a_{24} = 0$$

Entre las inumerables soluciones, una posible es:

$$U_2 = 15y_1 + 5y_2 - 21y_3 + y_4 \tag{5.2}$$

cualquier función lineal con coeficientes en la razón / 15: 5: 21: 1 es esencialmente el mismo contraste. Después se necesita que U_3 satisfaga la condición de ser un contraste

$$a_{31} + a_{32} + a_{33} + a_{34} = 0$$

y que sea ortogonal con U_1 y U_2:

$$4a_{31} - 3a_{32} + 2a_{33} - 3a_{34} = 0$$
$$15a_{31} + 5a_{32} - 21a_{33} + a_{34} = 0$$

La única solución, excepto un factor multiplicativo para el miembro de la derecha, es

$$U_3 = 5y_1 - 56y_2 - 7y_3 + 58y_4 \tag{5.3}$$

Por consiguiente, U_1 es arbitrario, U_2 se ha elegido con un elemento arbitrario y U_3 está determinado ; desde luego no puede encontrarse ningún U_4 ortogonal a estos tres. Puede utilizarse exactamente el mismo proceso para cualquier valor de n. La verificación algebraica o una prueba con valores númericos arbitrarios para las y_i convencería al lector de que

$$\Sigma(y_i-\bar{y})^2 = \frac{U_1^2}{38} + \frac{U_2^2}{692} + \frac{U_3^3}{6574} \tag{5.4}$$

con U_1, U_2, U_3 como los definidos por (5-1)-(5-3).

Una serie general de contrastes ortogonales, útil con frecuencia como ejemplo es:

$$\left.\begin{aligned} U_1 &= y_1-y_2 \\ U_2 &= y_1+y_2-2y_3 \\ &\ \vdots \\ U_j &= y_1 + y_2 + \dots + y_j - jy_{j+1} \\ &\ \vdots \\ U_{n-1} &= y_1 + y_2 + \dots + y_{n-1} - (n-1)y_n \end{aligned}\right\} \tag{5.5}$$

Cada par de estos se ve facilmente que satisface (3.4) ; tienen

$$D_j = j(j+1) \tag{5.6}$$

6. UN SEGUNDO TEOREMA

Recordemos las expresiones tales como II (2.1) para observaciones individuales. Pueden expresarse aquí como

$$y_i = \eta_i + \epsilon_i \tag{6.1}$$

siendo η_i la expresión paramétrica correspondiente a y_i/ y ϵ_i el error. Entonces, la esperanza de y_i es

$$E(y_i) = \eta_i\ , \tag{6.2}$$

y usualmente tomamos

$$E(\epsilon^2) = \sigma^2 \tag{6.3}$$

como en I (2.2). Evidentemente

$$E(U_j) = \Sigma a_{ji}\eta_i \tag{6.4}$$

Además, podemos escribir

$$U_j = \Sigma a_{ji}\eta_i + \Sigma a_{ji}\epsilon_i$$

Si lo elevamos al cuadrado y calculamos las esperanzas,/ recordando que las η_i son constantes y que las diferentes ϵ_i son independientes entre si, obtenemos

$$\begin{aligned} E(U_j^2) &= (\Sigma a_{ji}\eta_i)^2 + \sigma^2\Sigma a_{ji}^2 \\ &= [E(U_j)]^2 + \sigma^2 D_j\ . \end{aligned}$$

Este resultado puede escribirse

$$E(Q_j) = \frac{[E(U_j)]^2}{D_j} + \sigma^2\ . \tag{6.5}$$

Por consiguiente, si U_j es un contraste con esperanza / nula, Q_j tiene esperanza σ^2; salvo que σ^2 esté incrementado por un componente proporcional al cuadrado de $E(U_j)$. Esta es la base del cálculo de σ^2 a partir de los mínimos cuadrados en el análisis de varianza, así / como para las dócimas de significación de la "razón de la varianza".

7. DISEÑO COMPLETAMENTE ALEATORIZADO

Como en la sección III, supongamos que de las n / observaciones, r_h son del tratamiento T_h para h = 1, 2, ..., t. Escribamos Y_h para la suma de los r_h valores de y para el tratamiento T_h. Ahora definimos

$$\left.\begin{array}{l} U_1 = r_2Y_1 - r_1Y_2 \\ U_2 = r_3(Y_1+Y_2) - (r_1+r_2)\,Y_3 \\ U_3 = r_4(Y_1+Y_2+Y_3) - (r_1+r_2+r_3)\,Y_4 \\ \qquad\qquad\qquad \text{etc., hasta } U_{t-1} \end{array}\right] \qquad (7.1)$$

Estas expresiones, deben considerarse como abreviaturas/ para las funciones lineales de las y_i. Si cada y_h se escribe completamente como una suma de observaciones, se / ve que cada $U_1, U_2 \ldots, U_{t-1}$ es un contraste y cada par se ajusta a la ecuación(3.4). Estos (t-1) contrastes pueden denominarse *contrastes entre tratamientos*, ya que pueden calcularse a partir de los totales de los tratamientos/ (aunque para sus propiedades como contrastes deben considerarse siempre que estan introducidas todas las y_i).

Ahora consideremos el tratamiento T_1 a solas. El total de sus r_1 réplicas es Y_1, y el teorema de la Sección 4 muestra que podemos hallar (r_1-1) contrastes mutuamente ortogonales entre los r_1 valores de y, (los // coeficientes a para todas las y que no están en T_1 serán cero). Cualquiera de tales contrastes es evidente-- mente ortogonal a $U_1, U_2, \ldots, U_{t-1}$. Además la suma de las Q para estos r_1-1 contrastes es "$\Sigma(y-\bar{y})^2$" dentro del tratamiento T_1. Podemos razonar de forma semejante/ para cada tratamiento a su vez, observando que cual- / quier contraste "dentro de T_1" es ortogonal con cual- / quiera "dentro de T_2" etc. Por consiguiente, identificamos un total de (n-t) contrastes dentro de los trata- / mientos cuyos cuadrados de los contrastes suman

$$\Sigma y_i^2 - \frac{Y_1^2}{r_1} - \frac{Y_2^2}{r_2} - \ldots - \frac{Y_t^2}{r_t} \quad . \qquad (7.2)$$

Si también incluimos los (t-1) contrastes entre los tratamientos, tenemos un conjunto completo de n-1 contras-- tes mutuamente ortogonales, y juntos los cuadrados de // los contrastes deben formar $\Sigma(y-\bar{y})^2$ para las n observaciones. Se deduce inmediatamente que:

$$Q_1+Q_2 + \ldots + Q_{t-1} = \frac{Y_1^2}{r_1} + \frac{Y_2^2}{r_2} + \ldots + \frac{Y_t^2}{r_t} - \frac{(\Sigma y)^2}{n} \qquad (7.3)$$

De aquí, vemos como se construye la tabla I(4.3) . Además, todos los contrastes dentro de los tratamientos/ tienen esperanza nula, por lo que el cuadrado medio para estos (n-t) grados de libertad estima σ^2. Cada uno de // los $U_1, U_2, \ldots, U_{t-1}$ tiene una esperanza que implica parámetros de los tratamientos, como en la ecuación (6.5). Por eso, la esperanza del cuadrado medio de los (t-1) // grados de libertad para los tratamientos excede a σ^2 en un valor promedio de las (t-1) expresiones $[E(U_j)]^2/D_j$. / Los dos cuadrados medios tienen esperanzas iguales sí y solamente si todas las $E(U_j)$ son cero, que es equivalente a la condición de que todos los parámetros de los tratamientos sean iguales. De aquí que sea apropiado utilizar la razón de los cuadrados medios como la base para / una dócima de significación, y también utilizar la estimación de σ^2 del error como base para asignar un error típico a cualquier contraste entre las medias de los / tratamientos.

8. BLOQUES ALEATORIZADOS

En un diseño de bloques aleatorizados, los contrastes entre los tratamientos pueden ser aislados exactamente / de la misma manera. Pero los t tratamientos y los b / bloques entran simétricamente, y, por lo tanto, también pueden ser aislados los (b-1) contrastes ortogonales entre los bloques. Se ve facilmente que cualquier contraste de los bloques es ortogonal con cada contraste de los tratamientos. Deben quedar

$$(n-1) - (t-1) - (b-1) = (t-1)(b-1)$$

contrastes ortogonales tanto con los tratamientos como / con los bloques; estos deben tener esperanza nula, sean cuales fueran los parámetros del tratamiento y del blo-- que, y por lo tanto, el cuadrado medio con (t-1) (b-1) / grados de libertad se estima en σ^2.

Como ejemplo sencillo, que se aplica facilmente al experimento de la sección II 2, supongamos que tenemos 3 tratamientos en 4 bloques. Entonces, el conjunto de coeficientes:

	T_1	T_2	T_3
I	-2	3	-1
II	-2	3	-1
III	-2	3	-1
IV	-2	3	-1

define un contraste de tratamiento: es un contraste porque los 12 coeficientes suman cero y es un contraste de los tratamientos porque las cuatro repeticiones / de un tratamiento tienen el mismo coeficiente numérico. Análogamente

	T_1	T_2	T_3
I	-1	-1	-1
II	4	4	4
III	-2	-2	-2
IV	-1	-1	-1

es un contraste de los bloques. Verifíquese que estos / dos son ortogonales, si no lo ven evidente. Ahora /// consideremos:

	T_1	T_2	T_3
I	2	3	-5
II	-4	-6	10
III	6	9	-15
IV	-4	-6	10

Este es un contraste, y se verifica fácilmente que es ortogonal con cada uno de los otros dos.

No completaré el argumento con detalles, pero lo que se ha dicho conduce rápidamente a la Tabla II 2.2 y todo lo que sigue de la misma. Un resultado adicional, que requiere un poco más de algebra que la que he mostrado, pero que es una extensión de (6.5.), es que la esperanza del cuadrado medio de los tratamientos es

$$\sigma^2 + \frac{b}{t-1} \sum_{h=1}^{t} (\tau_h - \bar{\tau}) \qquad (8.1)$$

9. GENERALIZACIONES (PARA REFERENCIAS POSTERIORES)

El análisis de varianza de dóble entrada es la base del análisis estadístico en la mayoría de los experimentos planeados. Según se describe aquí, se relaciona con un diseño de bloque aleatorizado. Si cada bloque tiene p observaciones para cada tratamiento en lugar de una sola, un estimador de σ^2 solo puede hallarse a partir de una suma de cuadrados dentro de las clases. El análisis/ discutido en la Sección 8 se basa entonces en los totales para cada una de las bt clases, división de todas // las sumas de cuadrados por p esperanzas dadas como antes, y el cuadrado medio intra-clases con bt(p-1) grados de libertad tiene evidentemente como esperanza σ^2.

Las generalizaciones para diseños más complejos siguen el mismo patrón. Un cuadrado latino, (b=t) tiene // otra suma de cuadrados exáctamente como los de las filas y columnas extraibles de lo que era previamente "error". Un diseño factorial permite subdividir la suma de cuadrados de los tratamientos en componentes que pueden in--/ terpretarse por separado, y cada uno de ellos tiene una esperanza que implica los parámetros relevantes. Los di-- seños de bloques incompletos presentan más dificultades// ya que la esperanza de una suma bruta de cuadrados de // los tratamientos implica parámetros de los bloques y vice versa. Sin embargo, el análisis descrito en la sección V4 forma una suma de cuadrados para los tratamientos ajustada para los bloques (o un componente intrabloques de la / suma de cuadrados de los tratamientos) para la cual los / cuadrados medios esperados son de nuevo σ^2 más una fun- / ción de los parámetros de los tratamientos solamente. Los diseños para estudios de muestreo pueden implicar análisis de varianza jerárquico, posiblemente con el patrón de la Sección 7 repetido para cada par sucesivo de niveles / en la jerarquía y preferiblemente (por razones de simplicidad por lo menos) con todos los r_h iguales.

IV. CUADRADOS LATINOS Y ORTOGONALES MAS ALTOS

1.- ¿QUE BLOQUES?

En algunas situaciones cuando se examina un diseño / de bloques aleatorizados, el experimentador puede reconocer dos (o más) sistemas diferentes de bloques a utilizar. El caso clásico es el de la experimentación en cosechas agrícolas. La evidencia sobre las variaciones de / fertilidad dentro de un lugar puede sugerir que los bloques deberían estar separados por líneas que van de este a oeste: se considera que la fertilidad declina uniformemente de norte a sur de manera que tales bloques reducirian al mínimo la variación dentro de los bloques. Por / otra parte, el patrón de nuevos medios de riego o incluso la conveniencia de operaciones en el campo, puede sugerir que los límites de los bloques sean en dirección / norte-sur, Un diseño de *cuadrado latino* permitirá utilizar simultáneamente los dos sistemas; requiere que el / número de tratamientos sea el mismo que el número de cada clase de bloques. Por ejemplo, si A, B, C, D, E, son cinco variedades de trigo que deben compararse, la disposición del campo podría ser:

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

Obsérvese que tanto las filas como las columnas de esta disposición satisfacen la condición de bloque, ya que cada fila y cada columna contienen una "parcela" de cada / tratamiento. El nombre de "cuadrado" se refiere a la formación de las letras, ¡y no indica en absoluto que las parcelas reales de trigo deban ser cuadradas!

El uso de cuadrados latinos, no se restringe a la investigación agrícola. Consideraciones posicionales // exactamente semejantes pueden surgir al tomar muestras / de trozos de piel para comparar los tratamientos del curtido, trozos de tela de una pieza grande para comparar / los tintes, o trozos de chapa para comparar las calidades de la pintura aplicada. Ni las filas ni las columnas necesitan representar posiciones de esta forma. Al comparar los métodos de inoculación de un virus en las hojas, se podría considerar las plantas como columnas y las hojas partiendo de la base, como filas. Para comparar la habilidad de 5 técnicos para contar células se podrían / utilizar 5 microscopios como filas, y 5 platinas como columnas.

Pueden construirse cuadrados latinos de cualquier / tamaño. Si uno de los cuadrados menores (especialmente / 2 x 2 y 3 x 3) no proporciona repetición suficiente para un experimento, puede utilizarse un conjunto de 2, 3 o / más cuadrados para formas un experimento mayor. En la práctica, rara vez se necesitan cuadrados mayores de 10 x 10 ya que sus méritos para controlar la variabilidad podrían ser entonces menores, pero en determinados casos pueden / considerarse según su importancia.

2. ANALISIS DE UN CUADRADO LATINO

Un cuadrado latino para t tratamientos tiene t^2 observaciones. El equilibrio del diseño deja claro que los conjuntos de (t-1) contrastes para las filas, columnas y tratamientos son mutuamente ortogonales. Por consiguiente, el análisis toma la forma:

	g.l.
Filas	t-1
Columnas	t-1
Tratamientos	t-1
Error	(t-1)/(t-2)
Total	t^2-1

en que la suma de los cuadrados para cada una de las tres primeras líneas se calcula exáctamente como para los bloques o tratamientos de la sección II 2 utilizando el conjunto apropiado de t totales. La suma de cuadrados del error se obtiene también por resta. No se introducen nuevas características. Los errores típicos se deducen de la manera usual.

3. FORMULACION MATEMATICA

Creo que deberíamos observar la clase de formulación / de las observaciones que aparecen como ecuación II (2.1) un poco más cuidadosamente. Dicha ecuación se generaliza aquí como

$$y_{ijk} = \mu + \beta_i + \gamma_j + \tau_k + \epsilon_{ijk} \qquad (3.1)$$

para la observación en la fila, i, columna j, si esto ocurre al referirse al tratamiento k. Sin embargo, existe / una indeterminación en los parámetros. Sin alterar los resultados, se podría añadir una cantidad fija, a μ y restar la misma cantidad de todas las β_i o de todas las γ_j o de todas las τ_k . Es corriente superar esto, introduciendo todas las limitaciones

$$\left.\begin{array}{l}\text{Suma de todas las } \beta_i = 0,\\ \text{Suma de todas las } \gamma_j = 0,\\ \text{Suma de todas las } \tau_k = 0,\end{array}\right\} \quad (3.2)$$

(y por supuesto de manera semejante para los bloques aleatorizados). Por consiguiente, en particular las τ_k son las desviaciones de la media general, y tienen valores únicos. Esto resulta importante en la manipulación de diseños más complicados. El cuadrado medio esperado en III (8.1) se transforma en:

$$\sigma^2 + \frac{b}{t-1}\sum_{k=1}^{t} \tau_k^2 \quad (3.3)$$

y para un cuadrado latino se modifica como

$$\sigma^2 + \frac{t}{t-1}\sum_{k=1}^{t} \tau_k^2 \quad (3.4)$$

4. UNA ADVERTENCIA

Téngase en cuenta que la validez del experimento de un cuadrado latino se basa en la ecuación (3.1), es decir, en la aditividad de los diversos efectos. Si las desviaciones de la media general asociadas con las columnas no son las mismas de fila a fila, las t cantidades γ_j necesitarán substituirse por t^2 cantidades γ_{ij}. Los efectos del tratamiento, pueden quedar entonces enmascarados por estas "interacciones fila-columna" y el experimento no es fiable.

Es improbable que ésto sea una preocupación seria cuando la variabilidad total es pequeña, pero puede serlo si dicha variabilidad es muy grande.

5. ALGUNAS PROPIEDADES

Puede verificarse, fácilmente que hay sólo 2 cuadrados latinos 2 x 2 distintos y solamente 12 cuadrados 3 x 3. Los cuadrados de orden superior son mucho más numerosos, 576 para 4 x 4 y más de 6×10^{13} para 7 x 7. Para órdenes mayores, no se disponen de datos.

Un cuadrado latino debe elegirse al azar entre todos los cuadrados del tamaño deseado. En sus *Tablas Estadísticas* Fisher & Yates (1964) describen como pueden hacerse para los cuadrados más pequeños. Para cuadrados grandes lo apropiado es tomar un cuadrado de muestra, reordenar las filas en orden aleatorio (por ejemplo, ordenar filas, 1, 2, 3, 4, 5, en el orden 2, 4, 1, 3, 5), reordenar las columnas en un orden aleatorio y asignar tratamientos a las letras del cuadrado en un orden aleatorio.

6. USOS MODIFICADOS

La condición de que el número de filas, columnas y tratamientos deberá ser igual puede parecer severo. Sin embargo, es posible cierta variación incluyendo un tratamiento dos veces (o incluso tres veces) como si se trataran de dos tratamientos diferentes. El único cambio en el análisis es que uno o más contrastes que aparentemente están entre los tratamientos, tienen ahora esperanza nula y el cuadrado apropiado debería ser transferido a la suma de cuadrados del error. En ocasiones esto encaja bien con las consideraciones de optimalidad tales como las que se mencionaban en la sección II.1.

También es permisible diseñar un cuadrado latino y a continuación omitir deliberadamente la fila final (o la columna final). Esto destruye la ortogonalidad, pero todavía deja la posibilidad, de un análisis de mínimos cuadrados tal como el de los diseños para bloques incompletos (sección V).

7. CUADRADOS GRECO-LATINOS

Consideremos una adición al cuadrado de la Sección 1:

Cβ	Aδ	Eε	Dα	Bγ
Dγ	Eβ	Cα	Bδ	Aε
Bε	Cγ	Dδ	Aβ	Eα
Aα	Dε	Bβ	Eγ	Cδ
Eδ	Bα	Aγ	Cε	Dβ

Cada letra latina tiene ahora una letra griega junto a ella. Obsérvense que:

(i) Las letras griegas tienen la propiedad del cuadrado latino, cada una aparece una vez en cada fila y una vez en cada columna.

(ii) Cada uno de los 25 pares latino-griegos posibles se presenta una vez.

Este es un cuadrado *greco-latino*. Las letras griegas, proporcionan una nueva clasificación en las observaciones, los contrastes para las cuales son ortogonales con las filas, las columnas y las letras latinas. Por consiguiente, y de manera formal, los datos pueden analizarse por medio de una extensión del análisis de varianza:

	gl
Filas	t-1
Columnas	t-1
Romana	t-1
Griega	t-1
Error	(t-1)/(t-3)
Total	t^2-1

Se podría contemplar un experimento en el que un conjunto de tratamientos (por ejemplo tipos de inoculación de virus) está asociado con las letras romanas y otro conjunto (por ejemplo concentración de inoculante) está asociado con las letras griegas. En general, esto no es correcto a menos que existan razones suficientes/ para creer que los efectos de los dos conjuntos son verdaderamente aditivos. El diseño puede ser útil ocasionalmente cuando ha sido realizado un experimento de cuadrado latino y subsiguientemente el experimentador desea utilizar el mismo material (parcelas de terreno, árboles individuales, animales), para comparar un nuevo / conjunto de t tratamientos; si cree que los efectos / inducidos por los primeros tratamientos son muy probable que hayan desaparecido, puede utilizar un patrón / griego para los nuevos tratamientos en la confianza de que cualquier desviación de la simple aditividad será / pequeña.

Los cuadros greco-latinos son de gran ayuda en la construcción de otros diseños. (Conferencia V).

No pueden añadirse letras griegas ortogonalmente/ a todos los cuadrados latinos. En realidad, son pocos / en los que puede hacerse. Obsérvense el siguiente par de cuadrados latinos 4 x 4:

Aα	Bβ	Cγ	Dδ	y	A	B	C	D
Bγ	Aδ	Dα	Cβ		B	A	D	C
Cδ	Dγ	Aβ	Bα		C	D	B	A
Dβ	Cα	Bδ	Aγ		D	C	A	B

Difieren sólo ligeramente. Sin embargo, el primero se / muestra con una solución griega, y el segundo no puede extenderse de esta manera. De hecho, solamente 1/4 de / todos los cuadrados 4 x 4 pueden tener letras griegas / añadidas y sólo 3/28 de todos los cuadrados 5 x 5. Los cuadrados 6 x 6 son notables al no admitir ordenaciones greco-latinas. Para todo número superior a 6, existen / cuadrados greco-latinos (una verdad que estaba en duda hasta hace 25 años), pero son relativamente muy escasos.

Existen estructuras ortogonales todavía más complicadas. Consideremos el primero de los dos cuadrados 4 x 4 aún más amplio:

Aα1	Bβ2	Cγ3	Dδ4
Bγ4	Aδ3	Dα2	Cβ1
Cδ2	Dγ1	Aβ4	Bα3
Dβ3	Cα4	Bδ1	Aγ2

Los números 1, 2, 3, 4 dan una clasificación adicional / que es ortogonal con las filas y las columnas, latinas y griegas. No es posible más para un cuadrado 4 x 4. Se demuestra fácilmente que, para un cuadrado t x t, el número de conjuntos de símbolos distintos mutuamente ortogonales no puede exceder de (t-1). También se puede demostrar que este máximo puede alcanzarse si t es un número primo o una potencia de un solo número primo (3, 4,5, 7, 8, 9, 11, 13, 16, 17...) pero no se sabe mucho más, / aparte de esto. ¿Pueden ponerse ortogonalmente 14 conjuntos de símbolos en un cuadrado de 15 x 15? ¡No lo se, y no creo que nadie lo sepa!.

8. PARTICIONES ORTOGONALES DE CUADRADOS LATINOS

Un cuadrado greco-latino t x t puede considerarse como una (1^t) partición de cada fila de tal manera que sea simultáneamente una (1^t) partición de columnas y una / (1^t) partición de letras latinas. Ocasionalmente son útiles particiones menos extremas. Por ejemplo, si se considera que los tratamientos en un cuadrado 5 x 5, han cesado de afectar a las unidades experimentales, un científico podría desear superponer un nuevo conjunto de 4 tratamientos de una manera equilibrada, de forma que un nuevo tratamiento tenga doble réplica en cada fila, en cada columna y con cada tratamiento original. Esto se efectua / fácilmente haciendo idénticas dos de las letras griegas en el cuadrado del comienzo de la Sección 7: por ejemplo si cada ϵ se substituye por δ, se tiene una (1^t) partición ortogonal ($1^3 2$).

Existen muchas de tales particiones para los cuadrados latinos que no tienen (1^t) particiones griegas / completas. El segundo cuadrado 4 x 4 de la Sección 7 tiene una (2^2) partición:

Aα	Bα	Cβ	Dβ
Bβ	Aβ	Dα	Cα
Cα	Dα	Bβ	Aβ
Dβ	Cβ	Aα	Bα

Son más útiles las numerosas particiones ortogonales de los cuadrados 6 x 6, especialmente porque existen disposiciones greco-latinas 6 x 6. Por ejemplo, hay 2^3 particiones tales como

Bγ	Dβ	Aγ	Fα	Cβ	Eα
Dα	Bα	Eβ	Cγ	Fγ	Aβ
Aβ	Fγ	Bβ	Dα	Eα	Cγ
Fβ	Eγ	Cα	Bβ	Aα	Dγ
Cα	Aα	Dγ	Eβ	Bγ	Fβ
Eγ	Cβ	Fα	Aγ	Dβ	Bα

que podrían permitir poner 3 nuevos tratamientos en una disposición equilibrada en el cuadrado latino 6 x 6 / existente. Todos los tipos de partición excepto (1^6),ta les como (3^2), (1, 2, 3), pueden hallarse para cuadra-- dos 6 x 6. Recientemente (1982b) he publicado una relación muy completa de estas y de otras ortogonalidades / de orden más alto para el sistema 6 x 6.

V. BLOQUES INCOMPLETOS

1. LA NECESIDAD

En muchas circunstancias en que un experimentador // quiere utilizar bloques aleatorizados, el número de tratamientos excede el tamaño de los bloques disponible o / deseable -en ocasiones muy substancialmente. Menciono // tres ejemplos que contrastan entre si:

(i) Las pruebas de campo de nuevas variedades para una cosecha de gran magnitud pueden necesitar la inclusión de 20-50 variedades. Sin embargo, la experiencia sugiere que la eficiencia de los bloques, para el control de la variabilidad en los experimentos/ agrícolas, se reduce mucho para parcelas de más de 10 parcelas;

(ii) El bloque son animales de una misma camada y del / mismo sexo, y el total de tratamientos excede al número que puede esperarse razonablemente que ocurra con frecuencia (posiblemente 6 para ratas ó 3 para ovejas);

(iii) El tamaño máximo del bloque se impone por decisión administrativa, tal como el número de casos de algun fenómeno (una enfermedad, un tipo de accidente, un estado meteorológico) que ocurren en un mes y el número de tratamientos excede a este.

Los bloques completamente aleatorizados serán siempre la primera elección a menos que existan fuertes razones a favor de bloques menores: son fáciles de interpretar, eficientes, y su cálculo es sencillo. Sin embargo , la necesidad puede obligarnos a usar bloques menores o puede pensarse que usando bloques de tamaño reducido la ganancia en eficiencia es mayor que la pérdida que supone la replicación incompleta dentro de los bloques.

2. LA RESPUESTA

Recordemos la ecuación II (2.1):

$$y_{ij} = \mu + \beta_i + \tau_j + \varepsilon_{ij} \qquad (2.1)$$

para la observación y sobre el tratamiento j en el // bloque i; como en la ecuación IV (3.2) adoptamos ahora / las restricciones

$$\left.\begin{array}{l} \sum_{i=1}^{b} \beta_i = 0 \\ \sum_{j=1}^{t} \tau_j = 0 \end{array}\right\} \qquad (2.2)$$

donde b, t son el número de bloques y tratamientos. Supongamos que tenemos 5 tratamientos A, B, C , D, E y bloques (antes de la aleatorización del orden) como

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

podríamos calcular los parámetros mediante mínimos cuadrados, determinando los valores para minimizar la suma de

$$(y_{ij} - \mu - \beta_i - \tau_j)^2$$

sobre el total de las 13 observaciones. Esta puede parecer una tarea difícil, pero con las modernas facilidades de cálculo puede hacerse con bastante rapidez, incluso / para experimentos mucho más grandes. Hay dos objeciones. En primer lugar, la hipótesis implícita de varianza constante, ecuación I (2.2), puede ser inadecuada. Podría esperarse que σ^2 dependiera del tamaño de los bloques y / realmente la reducción de σ^2 al reducirse el tamaño de los b se ha considerado ya como una razón poderosa para utilizar bloques pequeños. Esta consideración es probable que sea más importante para parcelas de terreno que para camadas de ratas, y puede importar poco para bloques que son aproximadamente del mismo tamaño (por ejemplo 6 y 5, pero tal vez no 4 y 2). En segundo lugar, la distribución indicada es completamente asimétrica;por consiguiente, las varianzas para las medias de los tratamientos dependerán de las comparaciones que estén haciendo, (la diferencia entre C y E se calculará con mucha más precisión que la que hay entre B y D).

Pueden lograrse diseños mucho mejores si todos los bloques son del mismo tamaño. Observémos, por ejemplo, / las distribuciones para 3 tratamientos en 6 bloques de 2, y 5 tratamientos en 5 bloques de 4:

I	A	B
II	A	B
III	A	C
IV	A	C
V	B	C
VI	B	C

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

Si nos restringimos a las comparaciones intrabloques,es decir, a los contrastes ortogonales con bloques, es evidente que (en el primer diseño) $(\tau_1-\tau_2)$ puede estimarse a partir de los bloques I y II. Sin embargo, hay disponible información adicional:$(\tau_1-\tau_3)$puede estimarse a partir de los bloques III,IV, $(\tau_2-\tau_3)$ a partir de / los bloques V, VI y la diferencia entre éstos estima de nuevo $(\tau_1-\tau_2)$. La simetría de este diseño (y del otro ejemplo) asegura que todas las varianzas son iguales y da lugar a una forma de análisis simétrica manejable.

3. BLOQUES INCOMPLETOS EQUILIBRADOS (BIB)

Los dos ejemplos de la Sección 2 son ejemplos muy / sencillos de esta importante clase de diseño. La definición para un diseño de bloque incompleto equilibrado/ es que haya t tratamientos en b bloques cada uno de k "parcelas", (la palabra se utiliza en general para // nombrar animales o seres humanos, parcelas de una cosecha, o cualesquiera otras unidades experimentales) tal que:

(i) Cada tratamiento se presenta en r parcelas.

(ii) Cada par de tratamientos se presenta en λ de los b bloques. Desde luego, t, b, k, λ deben ser números enteros, al igual que

$$r = bk/t$$

el número de replicaciones de cada tratamiento, expresado alternativamente como:

$$N = bk = rt$$

el número total de parcelas. Puesto que un bloque contiene k(k-1)/2 pares de tratamientos y en todos hay t.(t-1)/2 diferentes pares de tratamientos, se sigue que

$$\lambda = \frac{bk(k-1)}{t(t-1)},$$

que por (3.2) se reduce a

$$\lambda = \frac{r(k-1)}{t-1}, \tag{3.3}$$

de donde el segundo miembro debe ser un entero. Los sencillos ejemplos de la Sección 2, tenían

$$k=2,\ b=6,\ t=3,\ r=4,\ \lambda=2$$

y

$$k=4,\ b=5,\ t=5,\ r=4,\ \lambda=3$$

respectivamente. Un ejemplo menos trivial es

I	ACEG
II	BCFG
III	ABEF
IV	DEFG
V	ACDF
VI	BCDE
VII	ABDG

que tiene: k = 4, b = 7, t = 7, r = 4, λ = 2. Un diseño k = 4, B = 20, t = 16, r = 5, λ = 1, puede construirse a partir del cuadrado 4 x 4 completamente ortogonalizado / del final de la Sección IV7. Escríbanse los 16 tratamientos en orden aleatorio en la parte superior del cuadrado. Elíjanse las filas como definición de 4 bloques, las columnas como definición de 4 más y análogamente 4 bloques a partir de las letras latinas, de las letras griegas, y de los números.

Para cualesquiera t, k, se puede formar siempre / un diseño de bloque incompleto equilibrado, tomando todas las selecciones posibles de k a partir de t, pero esto da con frecuencia un número de bloques demasiado grande. Por desgracia, aunque las condiciones de r, λ en // (3.1), (3.3) son *necesarias* para la existencia de un diseño BIB, no son *suficientes*. Por ejemplo, k = 5, b = 21, t = 15, r = 7, λ = 2 satisfacen las condiciones, pero no existe ningún diseño correspondiente a éstas. Excepto / cuando t y r son pequeñas, resulta difícil la investigación de la existencia de los diseños. Se dispone de catálogos extensos: también existen diversas reglas para la construcción de determinados subconjuntos de diseños. Entre los teoremas generales bastante poco conocidos estan:

(i) Para cualquier diseño, r $\geq$ k (y por consiguiente, / b $\geq$t); de donde k = 6, b = 8, t = 16, r = 3, λ= 1 que es imposible:

(ii) Si t = b y (por consiguiente r = k) y t es un entero par, (r - 1) debe ser un cuadrado perfecto; de donde t = b = 22, r = k = 7, λ= 2, que es imposible.

Se han obtenido demostraciones matemáticas sencillas y elegantes de estos teoremas, que se han reproducido en otro lugar (Finney 1960; capítulo 6;) no me propongo incluir demostraciones en estas conferencias, pero se las proporcionaré encantado a quién me las solicite.

4. ANALISIS ESTADISTICO

No descubriré todos los detalles del análisis estadístico de los diseños BIB, ya que esto se estudia mejor en los libros de texto. Sin embargo, la descripción de un caso, con unas pocas fórmulas más generales, les ayudará a leer los libros más fácilmente; es también instructivo con respecto a otros tipos de diseños.

La tabla 4.1., muestra las evaluaciones subjetivas del dolor registrado por 30 sujetos, a cada uno de los / cuales se les inyectó tres dosis de penicilina en tres / diferentes lugares. El dolor fue clasificado en una escala desde 0 (ninguno) hasta 4 (severo), independientemente para cada lugar. La asignación de las dosis en los sujetos se hizo de acuerdo con un diseño BIB para el cual

$$k=3,\ b=30,\ t=6,\ r=15,\ \lambda=6$$

Los valores muy pequeños de y pueden originar serias dudas sobre cualquier hipótesis de normalidad, aunque la sencilla aritmética sirve de ayuda para el ejemplo. Como es usual, establecemos

$$y_{ij} = \mu + \beta_i + \tau_j + \epsilon_{ij}\ . \qquad (4.1)$$

En la tabla 4.2 aparecen los cálculos preliminares. Los T_j son los totales de los tratamientos, transcritos de la Tabla 4.1. Los B_j son los totales de todos los bloques (sujetos) que recibieron el tratamiento j; / por ejemplo B_5 es el total para los sujetos I-V, X-XV, / XX-XXV, cada uno de los cuales recibió el tratamiento E. Veamos ahora las cantidades Q_j definidas en general como:

$$kQ_j = kT_j - B_j \qquad (4.2)$$

(No se confunda el uso de Q aquí con el de la conferencia III). Mediante un examen cuidadoso de cómo los diferentes valores de (4.1) contribuyen a T_j, B_j, y por consiguiente a Q_j, y recordando que

$$\sum_{1}^{t} \tau_j = 0\ , \qquad (4.3)$$

se puede demostrar que, por ejemplo:

$$kQ_1-kQ_2 = [k\lambda + (k-1)(r-\lambda)](\tau_1-\tau_2)$$

$\pm\ k\epsilon$ para cada una de las 2λ parcelas
$\pm\ (k-1)\ \epsilon$ para cada una de las $2(r-\lambda)$ parcelas
$\pm\ \epsilon$ para cada una de las $2(k-1)(r-\lambda)$ parcelas

(4.4)

Se deduce esto porque $(2r-\lambda)$ bloques contienen parcelas que contribuyen a (kQ_1-kQ_2), de los cuales $(r-\lambda)$ bloques contienen el tratamiento 1 pero no el 2, $(r-\lambda)$ / contienen el 2 pero no el 1, y λ contienen ambos tratamientos. El símbolo ϵ representa cualquiera de los ϵ_{ij}; un recuento cuidadoso prueba que 2λ de los ϵ_{ij} se presentan con el multiplicador k ó -k a partir de las parcelas de los dos tratamientos en los bloques en que se presentan ambos, y así sucesivamente. Con ayuda de (3.3), la ecuación (4.4) conduce a la esperanza o valor medio:

$$E(kQ_1 - kQ_2) = t\lambda(\tau_1-\tau_2)\ . \qquad (4.5)$$

Los tratamientos introducen simétricamente el diseño, y el hecho de que las diferencias entre las Q_j sean independientes de las β_i prueba que los contrastes entre las Q_j son ortogonales con los contrastes de los bloques. Además, de (4.4), la varianza de $(kQ_1 - kQ_2)$ es

$$V(kQ_1-kQ_2) = \sigma^2[2\lambda k^2 + 2(r-\lambda)(k-1)^2 + 2(r-\lambda)(k-1)]$$
$$= 2\lambda kt\sigma^2 \qquad (4.6)$$

utilizando (3.1), (3.3). Como $(kQ_1 - kQ_2)^2/2$ sería el cuadrado para un contraste en la suma de los cuadrados de / las desviaciones $\sum_{j=1}^{t} [(kQ_j)^2]$,

se sigue que, en unidades de una sola parcela obsérvese / que

$$\Sigma kQ_j = 0), \qquad \frac{\Sigma[(kQ_j)^2]}{\lambda kt} \qquad (4.7)$$

es una suma de cuadrados con (t-1) g.l. para su introducción en el análisis de varianza después de la inclusión / de una simple suma de cuadrados para los bloques. La tabla 4.3, expone este análisis.

De (4.5), τ_j se estima mediante, $kQ_j/t\lambda$, que puede añadirse a la media general de y (2.11 en el ejemplo), para obtener las medias correspondientes a cada tratamiento. De (4.6), para las comparaciones entre las medias de los tratamientos podremos asociar una varianza

$$\frac{k\sigma^2}{t\lambda} \qquad (4.8)$$

a cada media. Utilizando de la Tabla (4.3) $s^2 = 0,4741$ / como estimación de σ^2 se obtiene como SE 0,20, que aparece con las medias en la Tabla 4.4. Es evidente que hay diferencias muy marcadas entre E, F, y los otros cuatro tratamientos.

Este no es el fin de la historia, si los sujetos se han asignado aleatoriamente a los diversos conjuntos de / tres dosis, hay una información adicional sobre el τ_j para los totales de los bloques. Por ejemplo,

$$B_1-B_2 = (r-\lambda)(\tau_1-\tau_2) \pm k\beta \text{ de cada uno de } 2(r-\lambda) \text{ bloques}$$
$$\pm\ \epsilon \text{ de cada uno de } 2k(r-\lambda) \text{ bloques}$$
$$(4.9)$$

Pero la aleatorización sobre los bloques asegura que los β_j son errores aleatorios aplicables a los bloques, por lo que podemos escribir

$$E(\beta) = 0, \; E(\beta^2) = \sigma_B^2 \; . \qquad (4.10)$$

A partir de (4.9).

$$E(B_1 - B_2) = (r-\lambda)(\tau_1 - \tau_2) \; , \qquad (4.11)$$

y

$$V(B_1 - B_2) = 2k^2(r-\lambda)\sigma_B^2 + 2k(r-\lambda)\sigma^2 \; . \qquad (4.12)$$

Estas escuaciones son análogas a (4.5), 4.6), e indican/ que τ_j puede estimarse a partir de:

$$(\dot{B}_j - \bar{B})/(r-\lambda) \qquad (4.13)$$

con varianza

$$k(k\sigma_B^2 + \sigma^2)/(r-\lambda) \qquad (4.14)$$

Además, la consideración de otros contrastes interblo- / ques conduce al cálculo de $(k\sigma_B^2 + \sigma^2)$; el procedimiento no es evidente, pero se presenta con detalle tanto en mi libro de 1960 como en otros libros, y si me lo piden se lo explicaré. En el experimento de la penicilina, el resultado es otro conjunto de medias de los tratamientos estimadas con SE 0,74. Este SE (error típico) es, por lo tanto, mucho mayor que el 0,20 de la Tabla 4.4 por lo / que, en el caso presente, el análisis interbloques ape-- nas merece la pena, pero una combinación ponderada de // los dos conjuntos de estimaciones independientes conduce por último a un conjunto de medias con SE 0.19.

Quizás, se quisiera considerar las consecuencias / de usar un diseño modificado. Supóngase que los tres lugares en cada sujeto fueran identificados como P_1, P_2, / P_3 y que el experimentador sespechase que éstos pudieran diferir en sensibilidad al dolor. Yo me habría inclinado por recomendar el uso de cada conjunto de 3 sujetos con el mismo tipo de bloque (por ejemplo, los sujetos IV,XV, XXIV) en un esquema de cuadrado latino tal como:

	P_1	P_2	P_3
IV	B	D	E
XIV	E	B	D
XXIV	D	E	B

¿Cómo habría tenido que modificar el análisis estadístico?

5. OTROS DISEÑOS DE BLOQUES INCOMPLETOS

Los bloques incompletos equilibrados son tal vez los más sencillos de los diseños de bloques incompletos, y cuando son adecuados son ideales. Por desgracia, muchas/ situaciones experimentales imponen condiciones para las que no existe ningún diseño BIB. Afortunadamente, hay // otras muchas familias de diseños útiles con menor sime-- tría que la del el BIB pero con más que el de la sección 2. Los diseños parcialmente equilibrados, que incluyen/ la importante clase de los diseños reticulares, son posiblemente los más ampliamente conocidos y utilizados. Se han desarrollado algunas familias, tales como la de los bloques incompletos doblemente equilibrados para satisfacer necesidades experimentales especiales, algunas qui-- zás debido a un interés puramente matemático en sus propiedades combinatorias. Todas exigen un análisis estadístico semejante, aunque su complejidad aumenta cuanto menor es la simetria. En particular, la ecuación (4.2),es un método típico para obtener estimadores de los efectos de los tratamientos ortogonales con los bloques. El em-- pleo de los análisis intrabloques e interbloques y la / combinación de sus estimaciones puede implicar cálculos/ algebraicos muy complicados, pero una vez que se prepara un programa de ordenador el trabajo puede realizarse con facilidad y rapidez. En 1984, no hay excusa para elegir un diseño solo porque es sencillo algebraicamente o de / cálculo familiar: deberá ponerse un especial cuidado en hallar el diseño más adecuado a las preguntas que necesitan respuesta, y hacer ésto eficientemente dentro de / las limitaciones de los recursos disponibles.

Un desarrollo particularmente importante para el / análisis de variedades de cosechas agrícolas ha sido el de los α-diseños (Patterson *et alt*, 1978). Estos se diseñaron para satisfacer las necesidades del sistema británico de pruebas coordinadas de variedades, (Patterson & Silvey, 1980; Patterson & Hunter 1983). Las condiciones/ impuestas de forma rigurosa son:

(i) Los experimentos deberán tener 2, 3, ó 4 replicaciones.

(ii) Los diseños deberán ser resolubles (esto significa que los bloques pueden agruparse en replicaciones/ completas, una propiedad que no poseen ninguno de los ejemplos anteriores, aunque es posible para algunos diseños BIB);

(iii) Los diseños deberán tener bloques pequeños (preferiblemente no superiores a k = 10, pero tendrán //

capacidad para un gran número de variedades t=50, no sería raro).

La posibilidad de resolución tiene dos méritos, uno estadístico y otro muy práctico. Si el experimento puede/ proyectarse como replicaciones completas, dividiéndose/ cada replicación o bloque completo en los bloques incompletos del diseño, se crea la posibilidad de analizar / los bloques aleatorizados; esto puede servir de ayuda / como un análisis preliminar rápido y asegura que incluso en las circunstancias más inesperadamente desfavorables no puede haber *pérdida* de precisión con respecto a los bloques aleatorizados. Por otra parte, una sola replicación compacta es a menudo valiosa para inspección/ visual, demostración a los granjeros, etc.

En la Tabla 5.1. aparece un ejemplo de este tipo de diseño. Los diseños se definen mediante un método de generación más que por medio de condiciones tales como las establecidas en la Sección 3 para caracterizar los diseños BIB. La familia de diseños es muy grande, e incluye muchos diseños reticulares y cíclicos. Sin embargo, se conocen métodos que permiten modificar progresivamente un diseño de prueba hasta obtener uno de alta / eficiencia. El concepto de eficiencia está en relación/ inversa con la varianza de las comparaciones entre los tratamientos a partir de un análisis intra-bloques cuando se compara con la varianza para bloques aleatorizados si no existiera ninguna varianza adicional entre // los bloques incompletos. Mide el precio que se pagaría por utilizar los bloques incompletos cuando se deja de suprimir cualquier varianza adicional; en la práctica , se espera utilizar un diseño de bloques incompletos sólo cuando esta pérdida de eficiencia se compensa sobradamente por la reducción en el cuadrado medio del error efectivo. Si existe un diseño BIB para k, b, t, r especificados, tendrá la máxima eficiencia. Sin embargo, el mejor de los α-diseños tiene eficiencia muy poco más pequeña que la del diseño BIB correspondiente, si existiera, y es también resoluble.

Cuando se hacen pruebas de variedades , se acepta en ocasiones una relajación en las condiciones con / el fin de aumentar la eficiencia. Es decir, permitir experimentos con algunos bloques de k parcelas y algunos de (k - 1). Se piensa que es poco de temer que la va- / rianza dentro de los bloques dependerá apreciablemente/ del tamaño del bloque en sí, por ejemplo, 6 ó 7, y esta libertad extra puede permitir que la eficiencia se aproxime al máximo. Existen catálogos extensos de estos diseños.

TABLA 4.1

Resultados de un experimento de bloques incompletos equilibrados sobre el dolor de inyección intramuscular de penicilina.

Sujeto	Dosis A	B	C	D	E	F	Total
I	2		3		1		6
II		3	4		4		11
III	4				1	2	7
IV		1		2	1		4
V				4	1	3	8
VI	2		2	2			6
VII		2	1			1	4
VIII			4	2		1	7
IX	4	3				2	9
X	4	4		4			12
XI	1		2		0		3
XII		1	4		1		6
XIII	3				1	1	5
XIV		4		3	3		10
XV				3	1	1	5
XVI	3		3	3			9
XVII		3	2			1	6
XVIII			2	2		1	5
XIX	3	1				1	5
XX	3	2		2			7
XXI	2		1		1		4
XXII		3	3		1		7
XXIII	3				1	1	5
XXIV		4		2	1		7
XXV				1	1	1	3
XXVI	3		2	1			6
XXVII		3	3			1	7
XXVIII			2	1		1	4
XXIX	2	1				0	3
XXX	4	3		2			9
Total	43	38	38	34	19	18	190

TABLA 4.2

Cálculos Auxiliares de la tabla 4.1

Dosis	A	B	C	D	E	F	Total
T	43	38	38	34	19	18	190
B	96	107	91	102	91	83	570
3Q=37-B	33	7	23	0	-34	-29	0

TABLA 4.3

Análisis de la Varianza para la tabla 4.1

Ajuste para la		401.1111	
Variación media	g.l.	Suma de los cuadrados	Media cuadrática
Bloques (sujetos)	29	52.8889	6.7852
Tratamientos (intra-bloques)	5	33.9259	0.4741
Error	55	26.0741	
Total	89	112.8889	

TABLA 4.4

Registros estimados de la media del dolor para la Tabla 4.1

Dosis	A	B	C	D	E	F	
Media	3.03	2.39	2.68	2.16	1.20	1.28	$\pm$ 0.20

TABLA 5.1

Bloques de un α-Diseño para 30 tratamientos en
Cuatro Replicaciones y Bloques de 5
(tratamientos identificados por número)

Replicación 1

I	II	III	IV	V	VI
1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30

Replicación 2

VII	VIII	IX	X	XI	XII
1	2	3	4	5	6
8	9	10	11	12	7
16	17	18	13	14	15
21	22	23	24	19	20
29	30	25	26	27	28

Replicación 3

XIII	XIV	XV	XVI	XVII	XVIII
1	2	3	4	5	6
12	7	8	9	10	11
15	16	17	18	13	14
22	23	24	19	20	21
26	27	28	29	30	25

Replicación 4

XIX	XX	XXI	XXII	XXIII	XXIV
1	2	3	4	5	6
11	12	7	8	9	10
18	13	14	15	16	17
20	21	22	23	24	19
27	28	29	30	25	26

Para el uso in situ, los números 1 a 30, serían / asignados en orden aleatorio a los tratamientos o variedades reales. Cada replicación se consideraría como un superbloque de 30 parcelas. Dentro de cada superbloque, los seis bloques (series de 5 tratamientos) se pondrían en orden aleatorio. Dentro de cada bloque, se aleatorizaría el orden de los tratamientos.

VI. DISEÑOS FACTORIALES

1. INTRODUCCION

Uno de los grandes avances en el diseño de experimentos fue darse cuenta que los objetivos se lograban con una estructura factorial de los tratamientos. Es decir, el conjunto completo de los tratamientos en un experimento consta de todas las combinaciones de "niveles" de dos o más factores. Supóngase que un investigador quiere comparar el crecimiento de animales que reciben dos dietas alternativas y quiere comparar también los efectos de / dos temperaturas ambientales. Un experimento factorial / definiría cuatro tratamientos como las combinaciones // 2 x 2, de dieta y temperatura. Si existiera un tercer // factor, tal vez dos frecuencias de alimentación, habría ocho combinaciones.

A menudo se han descrito las ventajas de incluir / varios factores en un experimento.Las fundamentales son:

(i) Se economizan esfuerzo y material porque los efectos de cada factor se estiman independientemente / (ortogonalmente) dentro de un experimento;

(ii) La base para la inferencia sobre cualquier factor se amplia porque se docima (de una manera equilibrada) sobre una gama de combinaciones de otros / factores;

(iii) Sólo de este modo se pueden estudiar las interacciones entre los factores, tales como la diferencia en las tasas de crecimiento de animales con / las dos dietas que son mayores en establos sin calefacción que con ella.

Por necesidad, me limito casi totalmente a los factores en 2 niveles. Todo lo que se dice que puede generalizarse a otro múmero de niveles, por ejemplo 3 temperaturas, 3 dietas y 3 métodos de alimentación. Naturalmente, surgen mayores complicaciones cuando los factores // tienen 3 niveles, pero todo lo que se dirá acerca de los factores en 2 niveles, se puede generalizar a 3, 4 ó 5. Desde luego, en principio no existe ninguna razón que impida incluir en un experimento factores con diferente número de niveles.

2. NOTACION

Para su exposición se designarán los factores con letras mayúsculas A, B, C, D,... utilizando, siempre que sea posible, una letra que sugiera el factor (D: Dieta, T: Temperatura). Se distinguirán los dos niveles de un / factor por la presencia o ausencia de la letra minúscula correspondiente. Cuando un factor es cuantitativo (por / ejemplo 20ºC, 25ºC), es natural (no esencial) utilizar / la presencia de la letra para el nivel más alto, pero para un factor no cuantitativo esta identificación puede / ser arbitraria.

Si tenemos 4 factores A, B, C, D, el símbolo ac designará el tratamiento con los estados superiores o positivos de A y C, los estados inferiores o negativos de B y D. De forma análoga b representa el tratamiento en que solamente B está en el nivel superior. Es corriente representar la combinación de niveles más bajos de todos los / factores por (1), o simplemente 1. Por consiguiente las / 16 combinaciones son

1, a, b, ab, c, ac, bc,..., cd, acd, bcd, abcd.

Esto se denominará conjunto factorial de 2^4 tratamientos/ (4 factores en 2 niveles).

3. CONTRASTES ORTOGONALES

En la tabla 3.1, se indican en la parte superior las combinaciones del tratamiento para un factorial de 2^3. Supongamos que un experimento factorial de 2^3 se realiza como 3 bloques aleatorizados de 8 parcelas. Se construye // una función lineal de las 24 parcelas utilizando las cantidades indicadas en la línea rotulada A como los coeficientes de las parcelas individuales "cosechas" de acuerdo con los símbolos de los tratamientos. Esto indica la / diferencia entre las doce parcelas con a y las 12 sin a: la división del contraste por 12 es una media del efecto sobre la variable aleatoria de la cosecha y del cambio en el nivel del factor A. Con b, B en lugar de a, A, puede / afirmarse lo mismo de la siguiente línea. Además estos // dos contrastes se ve fácilmente que son ortogonales. (III Sección 3). La línea C se construye de manera análoga; el contraste es ortogonal con A y con B.

Observemos ahora la línea AB. Puede considerarse / como la diferencia entre "parcelas con a -parcelas sin a, en presencia de b" y "parcelas con a- parcelas sin a, en ausencia de b". Es simétrica, en el sentido de que "a" y "b" pueden intercambiarse en este contexto. Mide el grado hasta el cual el efecto del factor A es modificado /

por el factor B (y *viceversa*). Se conoce como la *interacción* de A y B, que se escribe A.B o AB. Este contraste es ortogonal con cada uno de los A, B, C. Se puede decir lo mismo de AB y BC. Por último, la línea ABC es un contraste para la interacción de tres factores, midiendo el grado hasta el cual la interacción AB es modificada por el factor C. Se tienen ahora siete contrastes mutuamente ortogonales, y por consiguiente, una forma de subdividir la suma de cuadrados de los tratamientos, (7 g. l.) en cuadrados individuales para los efectos principales A, B, C, y las interacciones.

El análisis de varianza puede presentarse como en la Tabla 3.2. Si se necesita puede hacerse una dócima de significación sobre cada uno de los componentes, pero las características importantes son que los efectos de tres factores se estiman simultánea e independientemente a partir de un experimento, y se adquiere evidencia sobre las diversas interacciones, sin la cual la comprensión resulta incompleta. Obsérvese en particular que el efecto de A se estima a partir de 13 replicaciones de parcelas con y sin a, y los efectos de B y C se estiman con la misma replicación a partir de las mismas parcelas. Sin embargo, sólo se han utilizado en total 24 parcelas.

4. CONFUSION

Un buen científico o técnico es probable que piense en los muchos factores que le gustaría incluir en su experimento. El número total de combinaciones de tratamientos resulta entonces grande (por ejemplo $2^5 = 32$), que excede del tamaño del bloque que está disponible o que se considera adecuado para el control de la varianza por reducción de la varianza intrabloques.

Si uno o más miembros del conjunto ortogonal de contrastes de los tratamientos pueden considerarse como de poco interés (quizás debido a que se considera probable que una interacción multifactorial sea despreciablemente pequeña o porque un efecto principal está ya tan bien comprendido que no se desea información adicional sobre el mismo), se puede adoptar el artificio de confusión. Por ejemplo, un experimento de 2^3 puede realizarse en bloques de 4 (con aleatorización del orden dentro de cada bloque), utilizando los dos tipos de bloque:

(i) 1, ab, ac, bc

(ii) a, b, c, abc.

Si el experimento tuviera 6 bloques de 4, 3 bloques cada tipo, la interacción ABC se *confunde* entre los bloques: el contraste ABC de la tabla 3.1 puede estimarse sólo por diferencia entre todos los bloques de tipo (ii) y todos los de tipo (i). Todos los demás contrastes de los tratamientos siguen siendo ortogonales con los bloques. El análisis de la varianza adopta la forma de la Tabla 4.1. Los cálculos no contienen nada nuevo; los 6 contrastes de los tratamientos se calculan exáctamente igual que antes, y la suma de los cuadrados para los bloques sigue las reglas usuales.

La confusión resulta más importante en los experimentos más grandes, pero puede aplicarse y generalizarse el mismo método. Por ejemplo, un experimento de 2^5 podría realizarse en bloques de 16 por confusión de A, B, C, D, E. Un tipo de bloque consistirá en todas las combinaciones de los tratamientos con un número par de letras (1, ad, bcde, etc.), el otro de todas las combinaciones con un número impar de letras. Alternativamente, se puede confundir cualquier otra interacción tal como ACD o incluso el efecto principal D si se considera de poco interés. Una confusión en bloques de 8 puede obtenerse por confusión simultánea de dos contrastes. Por ejemplo, podrían elegirse ABCDE y ADE, pero entonces necesariamente se confunde también BC. La regla es que si se confunden dos contrastes cualesquiera lo mismo ocurre con su "producto", cuando en el producto se suprime el cuadrado de cualquier letra:

$$ABCDE.ADE = A^2BCD^2E^2 = BC$$

Si, como ocurre generalmente, lo que se desea es restringir la confusión a las interacciones de orden más alto, la mejor elección será ABCD y ACE, que deben confundir también A^2BC^2DE ó BDE.

Por supuesto, esta regla no es una restricción arbitraria; es inevitable. Si se dividen las combinaciones de un factorial de 2^n en 4 tipos de bloque, con 2^{n-2} combinaciones de tratamientos en cada uno, de tal manera que el contraste entre los tipos (i), (ii) y los tipos (iii), (iv) confunda una interacción designada (o efecto principal), y el contraste entre (i), (iii) y (ii), (iv) confunda una segunda interacción, se comprobará que la constitución de los tipos de bloque se determina unívocamente; además, el contraste entre (i), (iv), y (ii), (iii) confunde necesariamente el "producto" de las dos interacciones.

Los bloques para el diseño se construyen fácilmente. Primero, se eligen símbolos de los tratamientos que contengan un número par de letras minúsculas a partir de las letras que forman las interacciones confundidas. Para

ACE, BDE, ABCE, éstas son:

(i) 1, ac, bd, abcd, abe, bce, ade, cde.
Obsérvese que éstas tienen la propiedad de que el producto de cualquier par (cuando de nuevo se su-- prime cualquier letra que aparece dos veces) es // otra del conjunto de ocho (ab. ace = bce), de mane ra que solo tienen que hallarse 3 combinaciones in dependientes. Estas 8 forman bloques del tipo (i). Para (ii), se elige cualquier tratamiento no in- / cluido todavía, por ejemplo ab, y se multiplican / los 8 por el mismo;

(ii) ab, bc, ad, cd, e, ace, bde, abcde.
De formaanáloga para (iii) se multiplican,por ejem plo los miembros de (i) por d;

(iii) d, acd, b, abc, abde, bcde, ae, ce
Las 8 combinaciones restantes forman (iv), y pue-- den obtenerse, por ejemplo, multiplicando por acde;

(iv) acde, de, abce, be, bcd, abd, c, a
Se pude comprobar fácilmente que los tres contrastes ortogonales simbolizados de forma abreviada // por

(i)+(ii)-(iii)-(iv)
(i)-(ii)+(iii)-(iv)
(i)-(ii)-(iii)+(iv)

forman las interacciones ABCD, ACE, BDE, respectivamente.

Si el experimento tiene varias replicaciones, se / puede confundir el mismo conjunto de interacciones en ca da uno. Una alternativa es la *confusión parcial* donde / se utiliza un nuevo conjunto de interacciones confundi-- das para cada replicación; cada interacción se estima se guidamente a partir de todos los elementos de replicacio nes en las cuales no se confunde.

5. REPLICACION UNICA

Si algunas interacciones de orden superior son despre ciables, los cuadrados medios correspondientes a las mis mas tendrán esperanzas que apenas superarán a σ^2. Una / práctica generalizada es aprovecharse de esto incluyendo tantos factores como sea posible y utilizando solamente una única replicación; luego se pueden asignar varias // interacciones para formar una suma de cuadrados que sólo daría una estimación ligeramente sesgada de σ^2. Por // ejemplo, supóngase que un experimentador estuviera pen-- sando utilizar dos replicaciones del diseño para 2^5 en / bloques de 8 de la Sección 4. Lo mejor que hacer sería // añadir F, un factor extra (casi con certeza tiene presente factores que quisiera incluir), y utilizar tal vez una confusión de una única replicación

ACE, BDE, ABCD, ADF, CDEF, ABEF, BCF

Los cálculos siguen el procedimiento normal. Para la suma de los cuadrados del error, podrían obtenerse 7g.l. a par tir de ABCDEF y las 6 interacciones de cinco factores. // (ABCDE, etc.); a éstas se les podría añadir 12 g.l. adi-- cionales de interacciones de cuatro factores inconfundi-- bles. Por consiguiente, el análisis de varianza tendría / 7 g.l. para los bloques, 7 ó 19 g.l. para el error, y los restantes 49 ó 37 g.l. para los efectos principales e interacciones individuales.

Por consiguiente, con escasos inconvenientes, un // factor extra y sus interacciones potencialmente interesan tes con las primeras cinco se han añadido a la informa- / ción del experimento.

Si originalmente se hubieran pensado en 4 repeticio nes del 2^5. Se podrían haber añadido los factores F y G./ con el fin de obtener una única replicación del 2^7.

6. REPLICACION FRACCIONARIA

Esta idea puede ampliarse. Si una interacción de orden superior es despreciable, no solo podría usarse para esti mar σ^2, sino que tampoco se produciría ningún peligro de confundirla con otro efecto principal ó interacción más / interesante. Como un caso trivial, considérese una estruc tura factorial de 2^4 en la que se puede disponer de los / resultados para 8 parcelas solamente que son;

1	ab	ac	ad	bc	bd	cd	abcd
y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8

Luego el contraste

$$-y_1 + y_2 + y_3 + y_4 - y_5 - y_6 - y_7 + y_8$$

estima A(al multiplicarse por 1/4). Pero si se intenta es timar BCD, se halla que se necesita exactamente el mismo contraste. Se puede escribir:

A = BCD

De manera análoga, se obtienen

B = ACD

y las otras relaciones. Todas estas pueden expresarse / simbólicamente como:

$$ABCD = 1$$

en el entendimiento de que de nuevo puede utilizarse la / regla del producto, esta vez para identificar los *apodos*. Por consiguiente D. ABCD = ABC; y de aquí

$$D = ABC$$

y D se *apoda* con la interacción ABC. Esto significa que ABC, no puede estimarse de forma distinta a D, una objeción sería a la replicación fraccionaria del 2^4; aún peores son los apodos tales como

$$AD = BC$$

Sin embargo, el 2^4 simplemente explica el método./ Con más factores la situación es diferente. Para 7 factores y

$$ABCDEFG = 1,$$

apodos típicos son:

$$B = ACDEFG,$$
$$DG = ABCEF,$$
$$ABE = CDFG,$$

todos los cuales pueden tolerarse, en el sentido de que las ambigüedades de interpretación resultantes pueden carecer de importancia.

Cuando el número de factores es grande, puede hacer una replicación de 1/4 ó 1/8. También los experimentos replicados fraccionalmente pueden confundirse, aunque omitiré sus detalles.

7. FACTORES A TRES NIVELES

A menudo se desea tener tres niveles para cada factor -tres temperaturas, tres dietas, etc.. La anotación y / los métodos se generalizan. Mientras que el algebra especial de los diseños de 2^n se basa en que:

$$A^2 = B^2 = \ldots\ldots = 1,$$

y

$$a^2 = b^2 = \ldots\ldots = 1,$$

en las reglas de multiplicación, para diseños de 3^n se / utiliza

$$A^3 = B^3 = \ldots\ldots = 1,$$
$$a^3 = b^3 = \ldots\ldots = 1.$$

Por ejemplo, las 9 combinaciones de tratamiento para el 3^2 son:

$$1,\ a,\ a^2,\ b,\ ab,\ a^2b,\ b^2,\ ab^2,\ a^2b^2.$$

donde a, a^2 son los niveles medio y superior del factor A con el inferior de B, a^2b es la combinación del nivel superior de A con el nivel medio de B, etc. Téngase en cuenta que el producto de dos cualesquiera es también uno de estos:

$$a^2b.a^2b^2 = a^4b^3 = a.$$

Los grupos de tratamientos

(i) $1,\ a^2b,\ ab^2$
(ii) $a,\ b,\ a^2b^2,$
(iii) $a^2,\ ab,\ b^2$

confundirán 2 g.l. de la interacción que tiene en total 4 g.l. Estos g.l. se designan por AB, A^2B^2. Las reglas de / ortogonalidad se explican mejor, por la confusión del 3^3 en bloques de 9. Se puede elegir un par de g.l. de la interacción ABC. (8 g.l. en total) tales como AB^2C, A^2BC^2 ; se observa que cada uno es el "cuadrado" del otro, ya que

$$(AB^2C)^2 = A^2B^4C^2 = A^2BC^2.$$

Para confundirlos, se eligen todos los elementos de la // forma $a^{\alpha}\ b^{\beta}\ c^{\gamma}$ tales que la suma de los productos de los índices de los efectos confundidos con los índices correspondientes α, β, γ sea un múltiplo de 3. Más concisamente

$$\alpha + 2\beta + \gamma \equiv 0 \pmod 3$$

y

$$2\alpha + \beta + 2\gamma \equiv 0 \pmod 3$$

La segunda condición es equivalente a la primera. Los elementos se hallan fácilmente como

(i) $1,\ ab,\ a^2b^2,\ ac^2,\ a^2c,\ a^2bc^2,\ ab^2c,\ bc,\ b^2c^2$
y estos constituyen el tipo de bloque (i). Para este bloque principal, el cuadrado de // cualquier elemento y el producto de dos elementos cualesquiera pertenece también al conjunto. Como para los diseños de 2^n, se forma otro tipo de bloques multiplicando cada / elemento de (i) por cualquier nuevo elemento por ejemplo por c;

(ii) $c,\ abc,\ a^2b^2c,\ a,\ a^2c^2,\ a^2b,\ ab^2c^2,\ bc^2,\ b^2$
y seguidamente se forma el tipo (iii) multiplicando una vez más, por ejemplo por $a^2b^2c^2$.

(iii) $a^2b^2c^2,\ c^2,\ abc^2,\ b^2c,\ ab^2,\ ac,\ b,\ a^2,\ a^2bc.$

Este, y tres diseños semejantes que confunden ABC^2, , A^2B^2C o AB^2C^2, A^2BC o ABC, $A^2B^2C^2$, son de gran valor práctico.

Otro conjunto valioso de diseños confunde el 3^4 en 9 bloques de 9, confundiendo por ejemplo

$$ABC,\ A^2B^2C^2,\ A^2BD,\ AB^2D^2,\ B^2CD,\ BC^2D^2,\ AC^2D,\ A^2CD^2.$$

Quizás les guste el ejercicio de construir algunos de los 9 tipos de bloque. Existe asimismo un diseño de replicación fracccionaria útil para 1/3 del 35 en bloques de 9. Aunque no tengo tiempo para exponerlo con detalle, después hablaré de ello con gusto si lo desean.

8. UN TEOREMA GENERAL

Un teorema general importante es que, para cualquier / número primo π, un experimento de π^n puede distribuirse en π^{n-p} bloques de π^p parcelas cada uno, sin confundir/ los efectos principales ni interacciones de dos factores, si y solamente si

$$n \leq (\pi^p-1)/(\pi-1)$$

Un teorema equivalente se refiere a la replicación fraccionaria. No lo demostraré a menos de que me lo pidan.

9. NIVELES MEZCLADOS

Los diseños factoriales pueden tener factores en diferentes número de niveles, tales como $2^2 \times 3$, 2×3^3 o // $2 \times 3 \times 4$. Algunas veces estos son esenciales para el carácter deseado de los experimentos. Si pueden realizarse/ en bloques completamente aleatorizados, no presentan ninguna dificultad. Sin embargo, la confusión es más complicada que antes y, en general, resulta en una pérdida parcial de información sobre varios contrastes. Pueden encontrarse ejemplos en los libros de texto que he mencionado, ya que es improbable que tenga tiempo para exponerlos en estas conferencias.

Si todos los factores tienen 2 ó 4 niveles, puede / ser posible en la práctica considerar cada factor de nivel 4, como un par de factores en 2 niveles para proyectar una ordenación confundiada. Por consiguiente, un diseño $2^3 \times 4^2$ podría confundirse como si fuera de 2^7. Esto / requiere mucho cuidado ya que el olvido de la verdadera / naturaleza de los factores puede resultar desastrosa.

TABLA 3.1

Contrastes Ortogonales para un experimento el 2^3

	Tratamientos							
Contrastes	1	a	b	ab	c	ac	bc	abc
A	-1	+1	-1	+1	-1	+1	-1	+1
B	-1	-1	+1	+1	-1	-1	+1	+1
AB	+1	-1	-1	+1	+1	-1	-1	+1
C	-1	-1	-1	-1	+1	+1	+1	+1
AC	+1	-1	+1	-1	-1	+1	-1	+1
BC	+1	+1	-1	-1	-1	-1	+1	+1
ABC	-1	+1	+1	-1	+1	-1	-1	+1

TABLA 3.2

Forma del Análisis de Varianza para 3 Bloques Aleatorizados de un Factorial de 2^3

Variación	g.l.	Suma de cuadrados	Cuadrado medio
Bloques	2		
A	1		
B	1		
AB	1		
C	1		
AC	1		
BC	1		
ABC	1		
Tratamientos	9		
Error	14		
Total	23		

TABLA 4.1

Tabla 3.2 Modificada para la confusión de ABC

Variación	g.l.	Suma de Cuadrados	Cuadrado medio
Bloques	5		
A	1		
B	1		
AB	1		
C	1		
AC	1		
BC	1		
Error	12		
Total	23		

VII. MAS FORMAS DE DISEÑOS

1. PARCELAS SUBDIVIDIDAS

A veces en un experimento se utilizan dos (o más) tamaños distintos de unidad para la aplicación de los tratamientos. Más específicamente, los niveles de un factor pueden asignarse a parcelas, y los niveles de otro factor pueden asignarse a subdivisiones de cada parcela (o *subparcelas*). Por ejemplo, las diferencias dietéticas sólo pueden estudiarse en animales completos, pero las reacciones de la piel a una inoculación pueden medirse en diversos puntos sobre cada cuerpo. Por consiguiente en un experimento sobre el grado hasta el cual la dieta modifica las reacciones cutáneas, podrían utilizarse ratas como parcelas en un diseño de bloques // aleatorizados para las dietas, con 4 tipos sobre posiciones de inoculación como subparcelas. En ensayos agrícolas, el riego quizás tenga que controlarse sobre /// áreas grandes, de manera que para una comparación de // los regímenes de riego deben utilizarse parcelas grandes, pero comparaciones de variedades pueden hacerse sobre subparcelas del mismo experimento.

El análisis de varianza no presenta dificultad, pero tiene una importante característica nueva. Hay dos / líneas de error distintas que corresponden a la variación interparcelas y la a intraparcelas. La tabla 1.1 / lo explica para un experimento sobre 5 dietas ensayadas sobre 6 camadas de 5 ratas con 4 tratamientos de inoculación ensayados sobre cada rata. El cálculo aritmético deberá hacerse siempre en función de las unidades experimentales más pequeñas, las subparcelas, de manera / que (de acuerdo con la conferencia III) la suma de los cuadrados para las parcelas principales (29g.1.) se encuentra que es:

$$\frac{\Sigma(\text{Total de parcelas principales})^2}{4} - \frac{(\text{Total general})^2}{120}$$

Las otras partes del análisis se calculan de la manera usual y ocupan sus lugares en la Tabla 1.1. La simetría debería dejar claro que todos los componentes son ortogonales. Desde luego, las medias deben compararse con la ayuda del cuadrado medio del error correspondiente.

Puede haber otras razones para utilizar parcelas divididas. Por ejemplo, si un experimento se realiza durante mucho tiempo, el experimentador puede desear modificarlo añadiendo un factor; una forma de hacerlo es poner el nuevo factor sobre subparcelas, aunque quizás necesita examinarse la alternativa de confusión. Asimismo a veces, un experimentador desee lograr una precisión relativamente mayor sobre un factor que sobre otro; se puede utilizar el hecho de que la varianza de la subparcela/ es usualmente menor (en ocasiones mucho menor) que la varianza de la parcela principal.

El principio de la parcela subdividida puede extenderse de modo que se tengan más niveles de división, obteniendo parcelas subdivididas-subdivididas o ¡Parcelas // subdivididas-subdivididas-subdivididas!. Rara vez esta es una característica deseable del diseño, pero puede resultad útil. El gran error es el de considerar la subdivi- / sión de parcelas como una manera fácil de ajustar los factores en un experimento, y adoptarse sin pensar en otros tipos de confusión. Obsérvese que estos diseños pueden // describirse alternativamente en función de la confusión . En el experimento de la Tabla 1.1, se podría hablar de // dos dietas como confundidas entre las ratas, mientras que no están confundidas las inoculaciones y la interacción.

2. MEDICIONES REPETIDAS

En otras circunstancias, cada unidad principal del experimento puede medirse varias veces por razones bastante diferentes. Una situación puede ser que una determinada/ propiedad deba estudiarse por muestreo. En un experimento in situ sobre una cosecha de cereales, se medirá el peso total del grano de cada parcela. Sin embargo, el contenido de nitrógeno del grano puede estudiarse analizando varias submuestras pequeñas de cada parcela; si el interés/ reside en el tamaño de las plantas o el daño de los insectos, pueden haberse medido o registrado submuestras de / plantas individuales dentro de cada parcela. Aunque existe analogía con las parcelas subdivididas, no se imponen ningún tratamiento ni ninguna otra estructura sobre las / unidades muestrales. En lo que se refiere al experimento, cualquier interpretación de los efectos del tratamiento / sobre el contenido de nitrógeno o sobre el daño de los insectos, se basará sobre las medias de las parcelas del // muestreo. Es conveniente un análisis completo de varianza tal como el de la Tabla 2.1, calculado de nuevo en fun- / ción de las unidades más pequeñas, pero sólo el análisis/ de las parcelas es importante para determinar los efectos de los tratamientos. La magnitud del error del muestreo / solo es útil para indicar si el muestreo fue lo suficiente completo para que la varianza de este origen sea solo una pequeña contribución a la varianza de la parcela; pue de servir de ayuda para indicar si en un experimento futuro semejante deberán tomarse 3 ó 6 unidades de muestreo para cada parcela en lugar de 4.

Consideraciones muestrales semejantes surgen en / muchas circunstancias. En los estudios clínicos-médicos pueden hacerse análisis replicados de muestras de sangre con fines de determinaciones bioquímicas u hormonales. Si deben compararse métodos alternativos de fabricación de un componente electrónico, la parcele puede / ser un lote de fabricación, de cada uno de los cuales / se seleccionan unos pocos componentes de muestra para / medir la calidad o la duración. En todos estos casos, / utilizar como varianza del error un cuadrado medio con 110 g.l. en la Tabla 2.1 o su equivalente sería totalmente erróneo, ya que éste sería una composición de dos varianzas, que podrían ser muy diferentes. Este es un / error que se comete rara vez, aunque a veces lo he observado.

Un error de la misma clase e igualmente serio, mucho más corriente, ocurre cuando las mediciones en una parcela se repiten con el tiempo. Lo explicaré refiriéndome al resumen de un experimento sobre un tipo de caracol marino (Finney, 1978b, 1982c). Su fin era ver si el modificar el número de sus depredadores afectaba a su / número. Se marcaron 6 parcelas en una zona afectada por las mareas de una playa de la costa del Pacífico de los Estados Unidos. Dos parcelas, seleccionadas al azar, no se tocaron; de otras dos se suprimieron las especies de depredadores; en las dos restantes, se añadieron depredadores a los que había allí normalmente. Posteriormente, se contó el número de caracoles en cada parcela aproximadamente una vez cada dos meses durante un año. (Los recuentos se hicieron en realidad sobre áreas muestrales/ pequeñas, pero el muestreo es un tema distinto de la presente discusión y se puede ignorar). ¿Cómo deberían analizarse los 36 "datos", 6 recuentos sobre las 6 parcelas?. La tentación, totalmente errónea, es escoger 2 g. l. para los tratamientos y atribuir a las medias de los tratamientos errores típicos basados en el cuadrado medio de los 33 g.l. restantes. Se han aplicado tratamientos a las parcelas enteras, y, por lo tanto, solamente/ la variabilidad *entre* las parcelas enteras es importante para la cuestión de si los tratamientos han afectado al número de caracoles. En realidad se pueden utilizar recuentos separados sobre cada parcela, como se indica en el análisis de la Tabla 2.2.

Primero se puede observar la media o los recuentos totales por parcela. A continuación, de manera independiente, se pueden mirar las regresiones lineales de los recuentos sobre el tiempo, que podrían ser por lo menos tan importantes para los efectos de la interferencia inicial con las parcelas, como lo son las medias. Puede / efectuarse un análisis adicional de otros contrastes que representan un componente cuadrático de la tendencia. Los restantes grados de libertad, que aparecen agrupados entre si en la Tabla 2.2., pueden subdividirse análogamente para componentes adicionales de la tendencia. Otras formas de disponer el análisis pueden basarse en subdivisiones alternativas de los grados de libertad, pero debe mantenerse un principio. Sea cual fuere la media o el con- / traste entre las 6 fechas de recuento a discutir, la varianza del error debe obtenerse a partir de la variabilidad de esta cantidad entre parcelas replicadas (aquí los pares de parcelas tratadas idénticamente).

Por desgracia, muchos científicos cometen el error de suponer que un análisis de tal experimento puede resumirse en 2 g.l. para los tratamientos, 5 g.l. para las fechas, 10 g.l. para la interacción TF y el resto para un / error que se supone homogéneo. En realidad, los errores / apropiados para diferentes contrastes pueden diferir grandemente. Se podría esperar que los cuadrados medios del / error en la Tabla 2.2. disminuyan uniformemente según se / va leyendo hacía abajo en la tabla (aunque con tan pocos grados de libertad el esquema puede ser irregular). Un sentido de seguridad en las conclusiones, totalmente falso, puede deberse a falta de comprensión de la estructura de la varianza de un experimento. Es evidente que este experimento no está adecuadamente replicado salvo que los // efectos de los tratamientos de las parcelas sean muy grandes.

El experimento real en que se basa esta discusión / fué más complicado, (Finney 1982c). Se hicieron recuentos también en todas las parcelas durante el año anterior al que se aplicaron los tratamientos. Por razones prácticas, no podían contarse nunca más de 2 parcelas al día y los / intervalos eran erráticos. No era posible en la práctica tener más de 6 parcelas. La tentación de utilizar la replicación aparente de recuentos individuales fué grande, pero no se hizo más genuina para evitar complicaciones // adicionales. No es sorprendente que el experimento no proporcionase evidencia convincente de ningún efecto de los tratamientos.

3. DISEÑOS CRUZADOS

Un procedimiento experimental importante, en especial con sujetos humanos o animales, es el de cambiar los tratamientos una vez o más durante el trascurso de un experimento. Introduciendo un equilibrio adecuado, las diferencias promedio entre los sujetos pueden eliminarse y la / varianza apropiada para las comparaciones de los trata-- mientos puede llegar a ser completamente intrasujetos.En la Tabla 3.1, se muestra (a) la forma más sencilla para dos tratamientos y (b) su elaboración. Con el primer di- seño, un total sencillo de A y de B da una comparación / que se equilibra sobre los sujetos y sobre los periodos. El segundo diseño necesita una expansión más extensa de la que puedo hacer en el tiempo de que dispongo, pero in cluso un breve examen muestra la posibilidad de observar los efectos directos tanto como los residuales: se puede comparar A después de A con A después de B, B después de A con B después de B.

En la Tabla 3.2, se expone un diseño cruzado más / complicado para 4 tratamientos. Obsérvese que en cada // uno de los periodos 2, 3, 4, los tres sujetos sobre // cualquier tratamiento han recibido los otros 3 tratamien tos en el periodo anterior, por lo que de nuevo podemos estimar no sólo los efectos de los tratamientos actual-- mente aplicados, sino también los efectos residuales del periodo anterior. La Tabla 3.2, puede utilizarse en va-- rias variantes. El experimento podría realizarse para // los periodos 1 y 2 solamente, aunque 12 sujetos apenas / serían entonces suficientes y se esperaría repetirlo con un segundo conjunto de 12. La asignación de los trata- / mientos a los sujetos, tiene entonces una estructura de bloque incompleto equilibrado que también es cierto si el experimento se realiza sólo para los periodos 1, 2,3. Solamente si se incluyen los 4 periodos se tiene el equi librio completo sobre los sujetos, sin confusión parcial de los tratamientos entre los sujetos. Si el tiempo lo / permite, pueden existir ventajas en contar con un quin-- to periodo que repita los tratamientos del periodo 4, de tal manera que las estimaciones estén también disponi- / bles relacionando a A después de A, a B después de B, / etc.

No se discutirá el análisis estadístico. Deben uti lizarse también en este caso los minímos cuadradcs para la estimación y desarrollar un análisis de varianza en función de los contrastes ortogonales correspondientes.

4. SERIES DE EXPERIMENTOS

En ciertas ramas de la ciencia y de la tecnología, un programa experimental puede necesitar incluir una serie de unidades experimentales. Una posibilidad consiste en repetir un experimento típico, de diseño relativamente / sencillo, en muchos lugares o en años sucesivos. Esto // puede ser útil cuando el número de combinaciones de tra- tamientos es pequeño. Dicho conjunto de experimentos per- mitirá hallar un promedio de los efectos de los tratamien tos sobre una variedad de condiciones, además de propor-- cionar información sobre hasta que punto varían los efec- tos de un sitio a otro. Este enfoque se ha utilizado para comparaciones clínicas entre medicamentos realizados en / varios hospitales o clínicas, y para la normalización de los fármacos en un conjunto de laboratorios.

En la investigación agrícola, el diseño factorial / estimula el interés en números mucho mayores de combina-- ciones de tratamientos. Sin embargo, si se desea estudiar respuestas de los fertilizantes sobre una región por prue bas en una muestra de lugares ampliamente distribuidos,po dría asignarse a cada lugar un bloque de un esquema de // confusión adecuado; por ejemplo, un grupo de lugares po-- dría constituir una replicación. Un análisis exhaustivo / de los resultados permitiría hallar el promedio de los // efectos no confundidos a estimar para la región; además, existe la posibilidad de buscar cualesquiera diferencias entre subregiones en la magnitud de los efectos así como obtener cierta información sobre las interacciones confun didas. Cada lugar puede considerarse como una replicación fraccionaria, aunque quizás no haya suficientes parcelas en un único lugar para que la información que se obtenga / pueda usarse mucho sola. Son especialmente adecuados los diseños confundidos de 3^n. Son posibles otras ordenacio-- nes en las cuales las combinaciones de los tratamientos / en un lugar no están determinadas por bloques de un esque ma de confusión. Estas pueden tener ventajas prácticas,pe ro como se utilizan con fines agrícolas especializados no hablaré más de ellas.

Cuando se planifican series de experimentos, debe / aclararse si su propósito es estimar los efectos bajo di- ferentes condiciones o usar una muestra de ambientes para estimar los efectos promedios aplicables a todos los am-- bientes. Por ejemplo, quizás se desee estimar los paráme- tros correspondientes a distintas diferencias dietéticas para una especie animal, y hacerlo así bajo diversas con- diciones climatológicas o para diversas edades y tipos de animal. Puede desearse información semejante sobre cada / categoría de animal,y puede surgir un patrón entre las / categorías, pero los experimentos deben considerarse como guía para una mayor comprensión. Asímismo, quizás se de- see estimar los efectos de la dieta promediados sobre con diciones y categorías como base de una politica asesora / aplicable a todos; idealmente, las condiciones, catego- / rías, o sitios deberán ser una selección aleatoria de to- dos los disponibles. Los experimentos repetidos a lo lar- go de los años tendrán usualmente este segundo sentido,ya que las diferencias anuales no tienen caracterfsticas // identificables por anticipado. Las repeticiones sobre lugares u otros subgrupos contemporáneos pueden tener cualquier objetivo.

La distinción es particular clara e importante en la investigación agrícola. ¿Están pensados los experimentos sobre las variedades de una cosecha o sobre las cantidades de fertilizante, realizados en muchos lugares y durante varios años, para estimar lo que es mejor por separado para muchos lugares o subregiones, aunque sea necesario calcular el promedio a lo largo de los años, ya que / las condiciones estacionales no pueden predecirse por anticipación? ó ¿Están pensados para calcular también el // promedio sobre los lugares experimentales con el fin de / establecer una política para toda la región estudiada? En este último caso, la estimación de los parámetros adecuados debe tener en cuenta las interacciones tratamiento x lugar y tratamiento x año como componentes del error adicional o la varianza intraexperimental. Estas interacciones miden la consistencia de los efectos sobre los lugares y sobre los años, y son, por lo tanto, muy interesantes para una política asesora. Al diseñar dichas series / de experimentos, la replicación sobre los lugares y sobre los años, debe decidirse con mucho cuidado, puesto que la medida de precisión necesaria, puede depender mucho más / de las interacciones que la replicación dentro de cada lugar.

5. DISEÑOS DE SUPERFICIE DE RESPUESTA

Cuando todos los factores de un experimento son mensurables sobre escalas continuas (pesos, longitudes, concentraciones, tiempos, etc.), puede enfocarse de otra / forma el diseño. Para dos factores, (la generalización/ para más factores es evidente), las combinaciones de niveles pueden representarse por las coordenadas (x_1, x_2). Entonces, la esperanza de y para x_1, x_2 fijos puede expresarse como

$$Y = F(x_1, x_2) , \tag{5.1}$$

siendo F( ) una función continua. Dentro del dominio de definición de los valores de x_1 y x_2, esta función podría aproximarse adecuadamente por medio de un polinomio. Por ejemplo, se podría ensayar:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{12} x_1 x_2 + \beta_{22} x_2^2 . \tag{5.2}$$

El diseño consiste en la especificación de un conjunto / de coordenadas (x_1, x_2) de tal manera que se optimice la estimación de los parámetros β. Los cálculos son esencialmente los de la regresión multiple, pero un diseño / bien equilibrado puede simplificar en gran manera la estructura de la matrix X'X. La elección del diseño ha de estar restringida por las limitaciones en los dominios / de definición de x_1 y x_2 que pueden utilizarse porque,para dominios de definición no restringidos, la ecuación / (5.2) no es válida, ni siquiera es una buena aproximación.

Estos diseños de superficie de respuesta son particularmente útiles en la ingenieria qufmica y en otros // procesos industriales, donde lo que interesa es hallar / la combinación de niveles que produzcan el rendimiento / más alto. Si se consideran los niveles óptimos de los // factores como origen de coordenadas, 0, para todos los factores, el diseño está centrado sobre este punto.Los / diseños parecen haber encontrado una acogida menos favorable en las ciencias biológicas puras y aplicadas, a pesar de los argumentos de que evitan el uso de algunas de las combinaciones extremas que requerirá un diseño factorial convencional. La razón puede ser en parte que parecen más difíciles de interpretar en función de los efectos de los factores individuales; tal vez lo más importante, es que en los experimentos biológicos, a menudo , se buscan factores no cuantitativos y los diseños de superficie de respuesta son menos adecuados para una mezcla de factores cuantitativos y cualitativos. Cochran & Cox (1957) ofrecen una buena explicación elemental sobre este tema.

6. EXPERIMENTACION SECUENCIAL

En ciertos tipos de experimentos, las "parcelas" o sujetos individuales se presentan en una secuencia temporal, y en algunos de éstos el resultado para cada sujeto puede estar disponible antes de tratar el siguiente. Algunas pruebas clínicas son de esta clase, siendo rápido el efecto del tratamiento con relación a la tasa en que / se encuentran nuevos pacientes.

Es posible entonces utilizar ideas adaptadas de la teoría del muestreo secuencial. La duración del experimento, y tal vez incluso también la determinación del // tratamiento para cada sujeto, pueden hacerse dependiente de todos los resultados anteriores. Por consiguiente, la replicación puede intensificarse para los tratamientos aparentemente más interesantes y de mayor éxito,y el experimento puede finalizar tan pronto como la evidencia baste para demostrar las diferencias entre los tratamien tos. En la práctica, los métodos se limitan normalmente a comparaciones de dos tratamientos.

No estoy familiarizado con este tipo de experimentación y no me propongo exponerla.

TABLA 1.1

Forma de Análisis de Varianza para un Experimento de Parcela Dividida

Variación	g.l.	Suma de Cuadrados	Cuadrado Medio
Camadas	5		
Dietas (D)	4		
Error (parcela principal)	20		
Ratas	29		
Inoculaciones (I)	3		
D-I	12		
Error (subparcela)	75		
Total	119		

TABLA 2.1

Forma de Análisis de Varianza para un Experimento de Bloques Aleatorizados de 5 Tratamientos en 6 Bloques, con 4 Unidades de Muestreo por parcela..

Variación	g.l.	Suma de cuadrados	Cuadrado Medio
Bloques	5		
Tratamientos	4		
Error	20		
Parcelas	29		
Error de muestreo	90		
Total	119		

TABLA 2.2

Forma de Analisis de Varianza para un Experimento sobre Caracoles Marinos

Variación	g.l.	Suma de cuadrados	Cuadrado Medio
Recuentos medios			
Tratamientos	2		
Error (1)	3		
	5		
Tendencia lineal			
Promedio	1		
Tratamientos	2		
Error (2)	3		
	6		
Tendencia cuadrática			
Promedio	1		
Tratamientos	2		
Error (3)	3		
	6		
Otros componentes de tendencia			
Promedio	3		
Tratamientos	6		
Error (4-6)	9		
	18		

TABLA 3.1

Diseños Cruzados para 2 Tratamientos

(a)

Sujeto	Periodo 1	2
I	A	B
II	B	A

(b)

Sujeto		Periodo	
I	1	2	3
II	A	A	B
III	B	B	A
IV	B	A	A

En la práctica, lo que se expone anteriormente se repetirá varias veces, es decir, cada secuencia de tratamientos sería asignada a varios sujetos.

TABLA 3.2

Cruzado para 4 Tratamientos

		Periodo		
Sujeto	1	2	3	4
I	A	C	B	D
II	D	B	C	A
III	B	D	A	C
IV	C	A	D	B
V	D	C	A	B
VI	C	D	B	A
VII	A	B	D	C
VIII	B	A	C	D
IX	B	C	D	A
X	D	A	B	C
XI	A	D	C	B
XII	C	B	A	D

VIII. PRUEBAS CLINICAS; ENSAYOS BIOLOGICOS

1. LA ETICA DE LA EXPERIMENTACION

La investigación en la medicina clínica introduce // inevitablemente cuestiones acerca del comportamiento // ético de los médicos y sus equipos. ¿Hasta que punto es justificable éticamente dar a un ser humano un medica-- mento (u otro tratamiento) que puede dañar su salud al mismo tiempo que la beneficia, o que incluso puede producir daño sin beneficio?. ¿Bajo que circunstancias (si las hubiere) se puede ensayar un tratamiento en algunas personas, con riesgo de daño, para que otras puedan beneficiarse?. No es asunto mio dar lecciones de ética,pero el estadístico no puede esquivar totalmente el tema.

A menudo se declara que ningún nuevo medicamento/ debe autorizarse para uso humano hasta que se sepa si / es beneficioso y sin efectos secundarios perjudiciales. Por desgracia, ni los ensayos de laboratorio ni las /// pruebas en otras especies animales pueden garantizar // buenos efectos en el hombre. Desde luego, la experien-- cia de la Biología y de la Farmacología indicará los tipos de estudios no humanos que predicen generalmente la respuesta humana. No obstante, sea cual fuere la información existente y por buena que sea la intención, el / primer ensayo en seres humanos es experimental. Si se / acepta la doctrina de no utilizar en seres humanos sin conocer su seguridad, no se introducirá ningún nuevo medicamento -situación que no puede considerarse acepta-- ble. Seguramente la conclusión debe ser que los prime-- ros usos humanos necesariamente experimentales deberán/ planificarse como experimentos *buenos*. Por esta razón , los estadísticos han mantenido que el tiempo crítico para la experimentación clínica es cuando existe confianza de que un nuevo medicamento no es seriamente perju-- dicial, pero existe incertidumbre completa acerca de si se trata o no de un perfeccionamiento del medicamento , al que pudiera substituir.

Para que un experimento planificado sea formal,debe diseñarse de tal forma que utilice los sujetos disponibles y la información obtenida de los mismos con la / mayor efectividad posible. Un experimento mal planificado es siempre antiético. Además el diseño ha de operar bajo las dos restricciones siguientes:

(i) El experimento debe interrumpirse si hay evidencia clara de que un tratamiento es superior a los // otros.

(ii) Se retirará el experimento a todo paciente al que / su médico juzga convenientemente que el tratamiento que se le aplica está haciendo más mal que bien.

Por supuesto, que estas observaciones se aplican a los experimentos que deben realizarse en personas enfer-- mas, cuyas posibilidades de recuperación son buenas. Las consideraciones éticas serán algo diferentes para el tratamiento de dolencias menores tales como jaquecas, naú- / seas o contusiones. Con tratamientos que puedan reducir / el dolor o incluso ofrecer cierta esperanza de curación para una enfermedad irreversible, puede haber mayor voluntad de correr el riesgo de efectos secundarios bastante / serios. Las cuestiones de consentimiento de los pacien- / tes, sujetos voluntarios y similares, tal vez no sean muy importantes en el aspecto estadístico, aunque el estadístico debe recordar la posibilidad de que pacientes (o personas sanas) que desean participar en un experimento pueden no ser representativos de la población en general.

En los últimos años se está examinando con más atención la ética de la experimentación con animales. No creo que ésta sea la ocasión de discutir la legitimidad de /// usar animales (haciendo que sufran) para beneficio del // hombre; considero que cualquier discusión de este tipo debería hacer una distinción entre los experimentos dirigidos a beneficiar la medicina y la salud humana de los destinados a las pruebas de productos cosméticos o tabaco u otros placeres humanos no esenciales, o de los que se // efectuan sencillamente para aumentar el conocimiento, u otros semejantes. Espero que todos estemos de acuerdo en que si tiene que haber un experimento, debe planificarse/ de modo que se utilicen los recursos y materiales de manera efectiva, con el sufrimiento mínimo para los animales. En particular, considero erróneo utilizar más animales de los necesarios para la precisión adecuada para los propósitos del experimentador, así como usar tan pocos anima-- les que la imprecisión haga que los resultados no tengan valor.

Si los experimentos con animales reclaman consideraciones éticas ¿qué ocurre con los insectos y las plantas?

Sea cual fuere la naturaleza de la experimentación, el estadístico tiene derecho a ser informado sobre cual-- quier cuestión que pueda plantearle algún conflicto ético. Su conciencia es tan importante como la de los demás miembros del equipo investigador.

2. EXPERIMENTOS CLINICOS

En su estructura combinatoria, los experimentos clínicos son normalmente muy sencillos. A menudo, sólo se / comparan dos tratamientos. Esto se debe a los muy subs-- tanciales problemas de organización. En muchos casos, la mayor parte de la dirección, medición y anotación tienen que realizarse por personas cuyo cometido principal no / es la investigación sino el cuidado de los pacientes, y estos últimos deben tener prioridad.

No obstante, en los últimos 30 años se han logrado adelantos inmensos en la experimentación clínica. Uno especialmente importante ha sido aceptar la aleatorización. Esta no debe encontrar dificultades éticas. Si un experimento se realiza cuando no se sabe cuál tratamiento es el mejor, no debe haber objeciones a la asignación aleatoria de los tratamientos a los pacientes. Si hay razones de peso para considerar que el medicamento B es meque el A, el problema ético no es el de la aleatorización, sino el propio experimento. ¡El proceso de la aleatorización debe protegerse contra cualquier ajuste o manipulación que ocasione sesgo! Idealmente, el orden de los eventos es abrir un sobre lacrado que diga cuál tratamiento debe recibir un paciente, una vez que es aceptado como idoneo para su inclusión en un tratamiento. También es vitalmente importante reconocer que debe incluirse cierta forma de tratamiento de *control* para compararlo con los nuevos tratamientos. De acuerdo con las circunstancias, un control puede ser la ausencia de cualquier tratamiento positivo o un "placebo", que se considera que no tiene efecto real, o el tratamiento clásico existente para una enfermedad.

Es bien conocido el peligro de que la evaluación / de los resultados pueda ser afectada por el paciente, y por las enfermeras o médicos que le examinan. Para evitar cualquier influencia subjetiva, es deseable que ni / el paciente ni los que le cuidan, ni los que evaluan los resultados sepan que tratamiento ha recibido. Esto puede ser posible en la práctica cuando los tratamientos son / medicamentos semejantes, pero no en una comparación de / quimioterapía con cirugía. Por supuesto, debe estar previsto romper el código en caso de necesidad.

Cuando los tratamientos deben compararse en pacientes que sufren de una enfermedad, es probable que los sujetos adecuados para su inclusión sean identificados a lo largo de un periodo de tiempo. También puede utilizarse bloques aleatorizados. Los bloques pueden definirse / en términos de sexo, edad, características fisiológicas, historial médico, etc y la aleatorización realizada por anticipado. La primera mujer de 20-30 años de edad se / asigna al primer tratamiento en orden aleatorizado para un bloque asi definido, y se procede igual con los pacientes subsiguientes. Puede haber un número de bloques de "mujeres de 20-30", o el total de tales pacientes pueden considerarse como un bloque con muchas replicaciones.

Si los tratamientos son tales que los pacientes con diferentes tratamientos estarán al cuidado de diferentes/ médicos o diferentes enfermeras, debe cuidarse la normalización de las condiciones; incluso entonces, debe tenerse presente la posibilidad de que las diferencias aparentes en los resultados sean en parte subjetivas o de origen // psicogénico. Para obtener más pacientes, puede gestionarse la colaboración entre varios hospitales. Cada hospital debe considerarse que tiene su propio experimento pequeño con bloques completos, y a toda costa debe evitarse la confusión del tratamiento con las diferencias de los hospitales.

Han sido utilizados, y probablemente debieran ser / utilizados con más frecuencia experimentos factoriales // sencillos, pero el número de combinaciones de tratamientos es una limitación. Recuerdo un experimento 2^3, creo / que se trataba de un tratamiento de diabéticos, en el // cual colaboraron unos 10 hospitales de Estados Unidos, cada uno de los cuales estudió igual número de pacientes en cada uno de los 8 tratamientos. El experimento hubo de / continuar durante varios años, y por supuesto se hicieron en cada paciente muchas mediciones y anotaciones diferentes. Inevitablemente se produjeron pérdidas por fallecimientos y otras causas. Los resultados fueron objeto de controversia, en gran parte porque los diferentes hospitales eran contradictorios en su evidencia. A pesar de una planificación muy cuidadosa, posiblemente las instrucciones sobre los tratamientos no fueron siempre interpretadas de la misma manera o los procesos de medición no fueron normalizados de forma adecuada. No obstante el trabajo, fué valioso; si el experimento se hubiera limitado a un hospital, la replicación no sólo habría sido mucho menor, sino que las interacciones tratamientos x hospital / habrían sido insospechadas.

Los diseños secuenciales han sido fuertemente defendidos para la investigación clínica, porque permiten terminar el experimento y tomar una decisión en cuanto se demuestra que un tratamiento es superior a uno alternativo. Se han utilizado de manera efectiva, pero los atractivos éticos son quizás menores que lo que parecen al principio.

Deben tomarse varias decisiones arbitrarias al comienzo/ para definir las reglas para su terminación. La vigilancia continua de los resultados puede introducir mayores incertidumbres éticas sobre si el experimento debe terminarse incluso antes. Un buen experimento debe proporcionar mucho más que la conclusión "B es mejor que A" así / como información sobre "cuánto mejor" y sobre muchos aspectos de la salud de los pacientes. Además, el diseño / secuencial sólo es practicable cuando el tiempo entre la administración del tratamiento y la evaluación del resultado es corto con relación al que transcurre entre las / llegadas de los pacientes sucesivos. Otro procedimento / es el modificar la asignación del tratamiento mientras / continua un experimento. Inicialmente, los tratamientos A y B, pueden asignarse a pacientes sucesivos con probabilidades iguales: si los primeros resultados indican / una superioridad de B, puede aumentarse la probabilidad para B. El objetivo ahora es alcanzar el fin del experimento con una comparación adecuadamente precisa entre A y B, pero habiéndose asignado un número mínimo de tratamientos al tratamiento peor.

Si los pacientes necesitan ser tratados durante / un periodo largo, pero el efecto de cualquier dosis es corto y eliminado rápidamente del sistema, son valiosos/ los diseños cruzados. Los medicamentos para el alivio / del dolor persistente o recidivo (por ejemplo reumatismo, migraña) pueden compararse de esta manera; la medida del éxito es probablemente un criterio subjetivo del paciente, y será importante mantenerle ignorante de si está recibiendo realmente un placebo, un remedio casero, o un nuevo medicamento.

3. ENSAYOS BIOLOGICOS O BIOENSAYOS

Este amplio tema se refiere a la normalización en laboratorio de los medicamentos y presenta ciertos problemas especiales en el diseño de experimentos (Finney, // 1978a). Voy a describirlo brevemente. Supongamos que un animal, que recibe una dosis z de un medicamento tipo S, origina una respuesta mesurable y, tal que

$$E(y|z) \equiv Y = F(z) \tag{3.1}$$

siendo F(z) una función monótona de z, que comprende parámetros desconocidos. Un nuevo medicamento pero estrechamente relacionado, o un lote de nueva fabricación del viejo, puede *normalizarse* con relación a S, si y solamente si existe una constante ρ tal que una dosis z del nuevo medicamento, T (*preparación de ensayo*) se comportan // exáctamente igual que una dosis ρz de S. Es decir, se tiene para T la relación dosis-respuesta:

$$Y = F(\rho z), \tag{3.2}$$

siendo F la misma función que para S en (3.1). Si, como / ocurre con frecuencia, T contiene el mismo principio activo que S, pero que puede estar diluido en diferente medida por materiales inactivos, *la potencia relativa*, ρ ,no es sólo una propiedad del sistema animal en que se mide / y, sino que es la razón de las concentraciones de principio activo en S y T-"la tasa del cambio" que se necesita/ en el campo de la práctica terapeútica humana para determinar la dosificación de T.

En el caso más sencillo e importante, en una amplia variedad de dosis, F(z) es una función lineal del logaritmo de la dosis. Se escribe:

$$x = \ell nz , \tag{3.3}$$

y

$$\mu = \ell n\rho . \tag{3.4}$$

Luego, para S y T

$$Y_S = \alpha + \beta x , \tag{3.5}$$

$$Y_T = \alpha + \beta\mu + \beta x . \tag{3.6}$$

A partir de un experimento comparativo sobre varias dosis de S y T,pueden estimarse dos ecuaciones de regresión paralelas:

$$\hat{T}_S = a_S + \beta x , \tag{3.7}$$

$$\hat{Y}_T = a_T + \beta x , \tag{3.8}$$

y por lo tanto

$$M = (a_T - a_S)/b \tag{3.9}$$

es un estimador de μ.Haré un comentario sobre el diseño de diseño de dichos experimentos, en el supuesto de que / todas las respuestas, y, estén distribuidas normal e independientemente respecto a Y, con varianza σ^2.

Un segundo caso importante es aquél en el cuál y es una variable aleatoria binomial que toma sólo los valores 0 y 1.Dicha *respuesta cuantal*,también denominada *reacción cuantal*, puede ser supervivencia-muerte, ausencia o presencia de convulsiones, muerte o germinación de una semilla o espora, etc. Se puede ahora interpretar F(z) como / la probabilidad de una reacción positiva a la dosis z. // Siempre que los sujetos sean independientes, si se ensaya una dosis con n sujetos, el número de reacciones obedecerá a una distribución binomial. Las dos expresiones de /

F(z), que son más útiles y se emplean más, son:

$$F(z) = \int_{-\infty}^{Y} \frac{1}{\sqrt{2\pi}} \exp(-^1/_2 t^2) dt \qquad (3.10)$$

Y

$$F(z) = [1 + \exp(-2Y)]^{-1} \qquad (3.11)$$

con x e y como en (3.3), (3.5) y (3.6). Estas correspon-- den a los bien conocidos métodos "probit"[Contracción de "*pro*bability un*it*"]Y "logit"[Contracción de "*log*istic // un*it*"]. Ambos pueden justificarse con argumentos teóri--- cos y empíricamente ambos son satisfactorios para muchos conjuntos de datos. No hay suficiente base teórica o // práctica para preferir uno al otro. Realmente, con las / definiciones de (3.10) y (3.11), F(z), para un valor determinado de Y, será casi igual para ambos. Los métodos de máxima verosimilitud son adecuados para estimar los / valores de a_T, a_S y b en (3.7) y (3.8) y para la estimación de ρ se procede como antes, por medio de (3.9).

4. DISEÑOS DE BIOENSAYO CON RESPUESTAS CUANTITATIVAS

En las conferencias anteriores, me he referido principalmente a experimentos en los cuales todas las comparaciones de tratamientos tienen un interés parecido. El ensayo biológico o bioensayo es un ejemplo que interesa // más específicamente para determinados contrastes. En un ensayo en el que y se mide en una escala contínua (usualmente el peso de un animal o de algún órgano de un ani- / mal), si ha habido replicación en varias dosis de S y T, el análisis clásico de varianza y los métodos de regre-- sión conducirán a (3.7) y (3.8), siendo a_T, a_S, b tres contrastes lineales entre las observaciones. El interés se concentra sobre los dos contrastes (a_t-a_s) y b, en / los cuales el experimentador querrá conseguir una gran / precisión. En la mayoría de los diseños simétricos estos dos contrastes serán ortogonales, lo que simplifica el análisis. Pueden destruir la ortogonalidad o las observaciones que faltan o la elección deliberada de bloques incompletos; esto no debe ser una dificultad seria, pero / reducirá la precisión.

Es raro tener la certeza de que las dos regresiones son lineales y paralelas. En un ensayo, la linealidad es como máximo una aproximación en una variedad de dosis;el paralelismo puede estar perturbado por alguna contaminación impensada de S o de T o por un intento insospechado de estimar una potencia relativa para la preparación de ensayo que no obedece a la *condición de semejanza* expresada mediante (3.1) y (3.2). Casi siempre, se necesita-- rán *dócimas de validez* de la linealidad y del paralelismo, de las que depende el cálculo de ρ. Entonces, surge la necesidad de equilibrar el énfasis en la precisión de la estimación y en la potencia para detectar la invali-- dez.

Es evidente que una dócima de paralelismo requiere que se incluyan como mínimo 2 dosis de S y T, y una dócima de linealidad exige 3 dosis. El diseño (2.2), dos / dosis de cada preparación, es muy popular, pero no permite ninguna dócima de linearidad. El diseño (3,3), tres / dosis de S y de T, con valores equidistantes de x, es mucho mejor si se desean dócimas de validez, pero para el mismo número total de sujetos proporciona menos preci- / sión para b. La variedad de dosis para cada preparación, debe ser tan amplia como se pueda sin que se corra un / gran riesgo de no linealidad, y cualquier suposición razonable para μ (or ρ) deberá utilizarse para elegir las / dosis de T, en la cual las respuestas esperadas se aproximarán a las de las dosis correspondientes de S.

Pueden surgir esquemas de confusión interesantes . Por ejemplo, un diseño (3,3) quizás necesite diseñarse / en bloques de 4. Las tablas 4.1 y 4.2, muestran dos maneras de hacerlo; un experimento completo podría repetir / varias veces estos conjuntos de 3 bloques. Aunque (3,3) es, en cierto sentido, un diseño factorial 2 x 3, en la presentación general de la Conferencia VI en verdad no / habría propuesto estos diseños. Para un bioensayo, pue-- den ser muy adecuados. La Tabla 4.1, tiene una confusión parcial de las dócimas de paralelismo y linealidad. La / Tabla 4.2 evita la confusión del paralelismo, reduciendo la precisión de b. Ambas tienen ventajas de acuerdo con las necesidades. La Tabla 4.3, a pesar de sus características poco corrientes, evita toda confusión de b y de paralelismo. La Tabla 4.4, con un múltiplo de 6 bloques,reduce notablemente la confusión de la linealidad y tiene menos confusión de paralelismo y de b que las Tablas 4.1 y 4.2, respectivamente. No puedo exponer estos diseños / con detalle: Los he incluido aquí simplemente para explicar mi tesis de que las necesidades altamente específi-- cas de estimación y docimasía exigen diseños especialmente construidos y de que el estadístico debe estar preparado para evaluar los méritos relativos de las alternativas.

En ciertas circunstancias pueden utilizarse los diseños cruzados con gran ganancia de precisión.

El diseño tiene incluso connotaciones más amplias si el experimentador puede elegir en cierta manera las / condiciones experimentales, la procedencia de los suje-- tos, etc. Deberá pues, buscar condiciones tales como que la varianza (σ^2) sea pequeña y la pendiente (β)grande:

La pequeñez de la relación σ/β es importante para una precisión grande en la estimación de M.

5. DISEÑOS DE BIOENSAYO CON RESPUESTAS CUANTALES

Algunas consideraciones de diseño son las mismas que para los ensayos de líneas paralelas de la Sección 4. / Sigue siendo deseable un valor grande β en (3.5), y cada dosis de T deberá elegirse con el propósito de tener la probabilidad de una respuesta próxima a la de la dosis correspondiente de S. Normalmente preocupan menos / los patrones combinatorios que los que tienen respuestas cuantitativas, en gran parte porque los ensayos requieren mayor número de sujetos y éstos tienden a ser / más homogéneos en su respuesta cuantal que lo que podrían ser para una respuesta cuantitativa.

Debe prestarse atención a la elección de dosis y la asignación de sujetos a las dosis. Una nueva dificultad surge de la distribución binomial para la proporción que responde: La expresión familiar "P(1-P)/n" tiene un máximo en P = 0,5 y disminuye a cero cuando la proporción de respuestas se aproxima a 0 ó 1. En combinación con el aplastamiento de la curva de respuesta (3.19) o (3.11) en los extremos de la dosis, esto tiene dos consecuencias importantes para un experimento con un número total fijo de sujetos:

(i) La precisión de a_S-a_T en (3.7) y (3.8) es máxima, si todas las dosis se aproximan a aquellas para / las que F(z) = 0,5;

(ii) La precisión de b es máxima, si los sujetos están divididos de forma aproximadamente igual entre dosis con F(z) = 0.05 a 0.10 y otra con F(z) = 0.90 a 0.95.

Ahora M implica la razón de estas dos cantidades. El estudio algebraico muestra que el compromiso ideal depende de N, del número total de sujetos, a causa de que la precisión de b resulta menos crítica según aumenta N. Aunque / el que hace el ensayo no pueda actuar de acuerdo con ninguna regla exacta, puesto que no conoce la relación dosis -respuesta, se le puede ayudar con objetivos establecidos contra los que puede interpretar cualquier información / anticipada o incluso suposiciones. Por ejemplo, cuando // N = 48 (un número muy pequeño), las probabilidades ideales son aproximadamente 0.16, 0.50, 0.80 para un ensayo / (3,3) y 0.16, 0.36, 0.64, 0.84 para uno (4,4). Cuando /// N = 240, éstas se modifican como 0.28, 0.50, 0.72 y 0.28, 0.42, 0.58, 0.72. Todos estos conjuntos se aproximan al óptimo si la función de respuestas es normal o logística, (3.10) ó (3.11).

En ocasiones, un experimentador puede tener que elegir entre métodos de bioensayo alternativos, unos utilizando respuestas cuantales y otros usando respuestas cuantitativas. En lo que se refiere a la precisión, puede comparar valores de $1/\beta$ para las respuestas cuantales, y la ecuación (3.10) con valores de σ/β para respuestas cuantitativas. Por supuesto, en el mejor de los casos tendrá estimaciones numéricas de y de experimentos pasados. / Puede utilizar $1/\beta$ y σ/β, en una primera aproximación, como si fueran desviaciones típicas de las estimaciones / de potencia logarítmica, por respuesta medida. A continuación, puede equilibrar los costos de los métodos de ensayo alternativos, respecto a los requisitos de tiempo y recursos, contra estas desviaciones típicas aproximadas y hacer su elección. La aproximación puede ser muy poco satisfactoria para experimentos pequeños.

TABLA 4.1

Primeros 3 bloques para ensayo de líneas paralelas (3,3), en bloques de 4, con paralelismo y linealidad parcialmente confundidos.

	S_1	S_2	S_3	T_1	T_2	T_3
Bloque I	x		x	x		x
Bloque II	x	x			x	x
Bloque III		x	x	x	x	

TABLA 4.2

Primeros 3 bloques para ensayo de líneas (3,3), en bloques de 4, con pendiente y linealidad parcialmente confundidas.

	S_1	S_2	S_3	T_1	T_2	T_3
Bloque I	x		x	x		x
Bloque II	x	x		x	x	
Bloque III		x	x		x	x

TABLA 4.3

Primeros 3 bloques para ensayo de líneas paralelas (3,3), en bloques de 4, con linealidad parcialmente confundida.

	S_1	S_2	S_3	T_1	T_2	T_3
Bloque I	x		x	x		x
Bloque II	x		x		xx	
Bloque III		xx		x		x

TABLA 4.4

Primeros 6 bloques para ensayo de líneas paralelas (3,3), en bloques de 4, con menor confusión de pendiente, paralelismo y linealidad.

	S_1	S_2	S_3	T_1	T_2	T_3
Bloque I	x		x		xx	
Bloque II		xx		x		x
Bloque III	x		x	x	x	
Bloque IV	x		x		x	x
Bloque V	x	x		x		x
Bloque VI		x	x	x		x

IX. SELECCION MULTIETAPICA Y CONTROL EXHAUSTIVO

1. INTRODUCCION

Esta conferencia, basada en la que pronuncié en la // American Statistical Association en Agosto de 1983, se / refiere a otra clase de problemas en la planificación de experimentos. Aunque parte del cálculo matemático se remonta a Pearson 45 años antes, mi tema comienza con /// Cochran (1951), que consideró el problema de la selección de una población en dos etapas. Se refería especialmente a la selección de buenas variedades de cosecha a partir de un gran número de nuevas variedades producidas por los agricultores. Comparaba las consecuencias de lograr la misma intensidad total de selección mediante pares alternativos de intensidades de selección en los dos años; por ejemplo, N variedades iniciales podrían reducirse a 0.01 N, seleccionando una proporción 0.2 en el / primer año, y seleccionando después una proporción 0.5 / de éstas en el segundo año, o utilizando pares alternativos de intensidad tales como 0.1, 0.1 ó 0.04, 0.25. La / discusión de estas opciones y otras relacionadas proporciona un buen ejemplo de la planificación de experimentos.

2. EL PROBLEMA GENERAL

Supongamos que el "valor" de una "entidad" se representa mediante una variable x que no puede medirse directamente. (La capacidad de producción de una variedad de trigo, el poder curativo de un bactericida), pero las estimaciones de x pueden obtenerse a partir de experimentos realizados en distintas etapas sucesivas en el tiempo (por ejemplo anualmente). Por ejemplo, entre un gran número de variedades de trigo, las denominadas i podrían tener una capacidad de producción x_i. Las cosechas y_{i1}, $y_{i2}, \ldots$, y_{ik} observadas en k etapas sucesivas, (típicamente una etapa es un año), serían:

$$y_{ir} = x_i + e_{ir} \qquad (r = 1,2,\ldots,k), \tag{2.1}$$

siendo e_{ir} un error aleatorio para el cual

$$E(e_{ir}) = 0, \quad E(e_{ir}^2) = \varepsilon_r^2 , \tag{2.2}$$

con errores independientes en las diferentes etapas. Con una cosecha agrícola, la producción esperada variaría de año a año debido al tiempo y a otras condiciones; sin / embargo, todas las decisiones de selección, se basarán / en comparaciones dentro de un año, y, supuesto que se // tiene aditividad, no necesitamos complicar (2.1) añadiendo a x_i un parámetro para "el año" que es independiente / de i. El proceso de selección comienza con N_1 variedades u otras entidades, de manera que i = 1, 2,... N_1. Después de cada etapa, algunas son desechadas y las que quedan / se pasan a la siguiente etapa, de manera que al final de la etapa r, sólo queda N_{r+1}. El própósito es optimizar // los valores de x_i para la N_{k+1} final, mediante la elección de las condiciones experimentales y la intensidad de selección en cada etapa (Finney, 1964).

Las entidades desechadas no se medirán en las etapas siguientes, reduciendo así los costes experimentales. En la etapa r, la selección puede basarse solamente en // y_{i1}, y_{i2}, ..., y_{ir} para las N_r entidades todavía a prueba. Habrá restricción sobre los recursos totales disponibles y ε_r^2 estará en relación inversa a los recursos gastados en la etapa r del experimento. La selección en la etapa r podría basarse en uno de los tres conjuntos de cantidades:

(A) y_{ir} sólo;

(B) $\bar{y}_{ir} = (y_{i1}+y_{i2}+\ldots+y_{ir})/r$, la media no // ponderada;

(C) $\bar{y}_{irw} = (y_{i1}\varepsilon_1^{-2}+y_{i2}\varepsilon_2^{-2}+\ldots+y_{ir}\varepsilon_r^{-2})/(\varepsilon_1^{-2}+\varepsilon_2^{-2}+\ldots+\varepsilon_r^{-2})$, la media ponderada (o // una media semejante que utilice varianzas estimadas)

El criterio de optimalidad para los N_{k+1} valores de x_i sigue siendo que maximizar el promedio de los valores / de x_i es más importante para situaciones prácticas.

En lo sucesivo, por razones de simplicidad, voy a // omitir el subíndice i en x e y; el contexto aclara si se aplican determinados x_i, y_{ir}, o los conjuntos de los N_r / valores de éstos en la etapa r.

3. VARIEDADES DE COSECHA

Los cultivadores de cualquier especie de cosecha producen grandes números de nuevas combinaciones genéticas(por ejemplo, en el Reino Unido, aproximadamente 500.000 anualmente para el trigo), la mayoría de las cuales serán rápidamente desechadas debido a su inclinación a las enfermedades y pestes, estado de crecimiento insatisfactorio, / etc. Después de tres temporadas, pueden quedar 1000- // 10.000 como variedades potenciales que requieren ser evaluadas en términos de rendimiento. La capacidad de rendimiento inherente de una variedad, x, puede estimarse(con relación a las otras) por experimentos anuales que pro-- porcionan la sucesión de medias $y_1, y_2 \ldots y_k$. Supongamos que A mide los recursos totales y se utiliza una parte A_r para la experimentación en la etapa r, siendo.

$$\sum_{r=1}^{k} A_r = A \qquad (3.1)$$

Si todos los experimentos utilizan las mismas parcelas / pequeñas y los recursos se miden en términos de área de terreno, el número de parcelas por variedad en la etapa r, será proporcional a A_r/N_r, la parte de recursos disponible para una variedad. También, ϵ_r^2, la varianza de la producción media de una variedad en la etapa del experimento, será la varianza por parcela dividida por el número de parcelas, y por lo tanto,

$$\epsilon_r^2 \propto N_r/A_r \qquad (3.2)$$

Por diversas razones, esta relación no es exacta, pero / es una buena aproximación. En una situación estable, una nueva cohorte de N_1 entra en el programa de selección cada año (por supuesto N_1 no es exactamente constante) y / se pasa a través de las etapas con independencia de las cohortes previas. Por consiguiente, A_r puede ser considerada *o bien* la parte asignada a la etapa r de los recursos totales A disponibles para una cohorte, *o* la parte / de los recursos anuales totales A que se asigna a la cohorte en la etapa r.

Cochran (1951) suponía que los valores iniciales / de x eran una muestra grande,una distribución normal (varianza σ^2), y que todas las distribuciones de los erro-- res eran normales. Su selección bietápica utilizaba $\bar{y}_{rw}$; observó que al retener los valores mayores era óptima.Si N_{r+1} es grande, la media de los N_{r+1} valores de $\bar{y}_{rw}$ se-- leccionados se aproximará a la media del extremo superior de la distribución para la fracción elegida. Para / una distribución N(0, 1), la media del extremo superior del área P es

$$\nu(P) = Z/P \ , \qquad (3.3)$$

siendo z la ordenada que acota el extremo, y por lo tanto $\sigma\nu$ (P) es el perfeccionameinto esperado si fuera a / seleccionarse sin error una fracción P de la x mayor. De aquí, en particular se deduce fácilmente que la ganancia esperada en x a partir de una selección monoetápica sujeta a error es:

$$G = \sigma^2\nu(P)/(\sigma^2+\epsilon_1^2)^{1/2} \qquad (3.4)$$

Siguiendo a Cochran, estudié (1958) la selección de fracciones sucesivas $P_1, P_2, P_3 \ldots P_k$ siendo

$$P_r = N_{r+1}/N_r \ . \qquad (3.5)$$

y N_{k+1} grande. De acuerdo con (3.2), escribí

$$\epsilon_r^2 = \gamma\sigma^2 N_r A/(N_1 A_r) \qquad (3.6)$$

siendo γ una constante sin dimensión. Consideré que la selección sobre y_r debía aproximarse mucho a la experiencia de los cultivadores de las plantas y, por lo tanto en la etapa r ignoré la información de etapas anteriores. El / cálculo podría basarse en la distribución normal de las / k+1 variables aleatorias sobre cada y_r, por separado. Como deseaba estudiar $k>2$, desarrollé un procedimiento alternativo para obtener la distribución de los valores x / que quedaban en cada etapa, con la ayuda de transformaciones de acumulantes (Finney 1956; 1961; 1962a), utilizando desarrollos debidos a Cornish y Fisher (1937; Fisher y // Cornish,1960).

Escribiendo:

$$\Pi = P_1 P_2 \ldots . P_k \qquad (3.7)$$

como la fracción de la selección total, demostré que(cuando γ tiene un valor razonable 1.0) para π = 1/100 aproximadamente el 90% de la ganancia máxima posible $\sigma\nu$ (0.01) podría alcanzarse con k = 2; desde luego, la posición es menos favorable para γ mayor. Las condiciones óptimas están muy extendidas en el entorno de $P_1 = P_2$, $A_1 = A_2$. Esto sugiere la regla más general de que, para una selec- / ción total π en k etapas.

$$P_r = \Pi^{1/k} \ \text{ and } \ A_r = A/k \ \text{ cualquiera que sea } k \qquad (3.8)$$

se aproximará a las condiciones que maximizan la ganan- / cia. La regla ciertamente no es exacta, y resulta una /// aproximación bastante pobre para k grande, pero su facilidad operativa debe ser atractiva. Además, en la mayoría / de las condiciones, k = 3 parece suficientemente grande / para todos los casos prácticos. Si π es muy pequeño, un / descarte aleatorio inicial de una fracción $(1-P_0)$ pueden aumentar la posible ganancia.

El advenimiento de los ordenadores ha hecho posible en la práctica no solo cálculo más extenso de tipo, sino también la comparación con muestras finitas. En 1966, informé acerca de resultados con muestras pequeñas de / una distribución normal de x. Su concordancia con los / cálculos anteriores con muestras grandes fue bueno, excepto que para cualquier valor de π la ganancia esperada es menor cuando N es pequeño. Por ejemplo, con //// π = 0.01, la ganancia con selección perfecta, utilizando (3.3) es 2.665σ.La selección simétrica bietápica, // ecuación (3.8), se aproxima siempre mucho a la óptima / con γ = 1, la ganancia esperada para N_1 grande es /// 2.345σ; que disminuye a 2.30σ para N_1 = 500, N_3 = 5 y a 2.13σ, para N_1 = 100, N_3 = 1. Es evidente que esto / no deja mucha oportunidad para mayores ganancias al utilizar más etapas: con N_1 = 100, lo óptimo parecen ser 4 etapas, para las cuales la ganancia es 2.29σ, una pequeña ventaja frente al retardo de dos temporadas. Un / pequeño estudio de las funciones de varianza utilizaba una generalización (3.6):

$$\epsilon_r^2 = \gamma\sigma^2[N_r A/(N_1 A_r)]^w \qquad (3.9)$$

y se encuentran condiciones óptimas y ganancias esperadas notablemente insensibles a los cambios en w en un recorrido de 0.1 a 2.0. Curnow (1960, 1961) comprobó que / substituyendo la distribución normal de x por varias distribuciones Beta y χ^2 se alteraba la magnitud de las ganancias de la selección, pero se lograban pocas diferencias para las condiciones óptimas y (3.8) seguía siendo una buena aproximación. Confirmó que incluso las distribuciones asimétricas proporcionaban poco beneficio para una cuarta etapa.

La Tabla 3.1, da los resultados de simulaciones // con N_1 = 16, N_3 = 1 para diversos N_2 y A_1/A siendo //// γ = 1 y

$$\epsilon_1^2 = \sigma^2 A/A_1, \quad \epsilon_2^2 = \sigma^2 N_2 A/(N_1 A_2) \qquad (3.10)$$

Esto demuestra claramente que el N_2 = 4, A_1 = A/2 simétrico se aproxima al óptimo, aunque la relación entre // la x elegida y N_2,A_1 es muy extensa en esta región. El / aspecto general de la tabla 3.1 es característico de lo que se encuentra para números mayores y más interesantes. El margen de perfeccionamiento es pequeño: incluso la selección perfecta en este ejemplo, proporcionaría sólo una ganancia de 1.766 σ. Sin embargo, la selección / basada en $\bar{y}_{rw}$ es probable que haga el resultado todavía menos sensible a cambios en N_2 y A_1, ya que sus beneficios serán los más grandes cuando ϵ_2 es relativamente // grande. La Tabla 3.2 confirma esto, al mostrar sólo ligeros aumentos en la ganancia máxima esperada, aún con una región todavía mayor que es extensa respecto a N_2 y A_1.En condiciones realistas, la mayor parte de la ponderación / para $\bar{y}_{rw}$ proviene de la etapa actual, de manera que los resultados que utilizan y e $\bar{y}_{rw}$ estarán grandemente correlacionados. La Tabla 3.3 muestra las varianzas entre / las simulaciones individuales utilizadas para la Tabla // 3.1, e indica con cuanta facilidad una selección individual "16-N2-1" puede proporcionar resultados mucho menos (o mucho más) satisfactorias que lo que expresa la Tabla 3.1. Las varianzas correspondientes para la selección basada sobre $\bar{y}_{rw}$ son en su mayoría las mismas, excepto que aumenta muy poco cuando A_1 se aproxima a A.

Una interacción de variedades x año tiene el efecto de aumentar cada ϵ_r^2 en una magnitud constante. Si la interacción es apreciable, la ϵ_r^2 se hace más aproximadamente igual y para cada variedad $\bar{y}_{rw}$ se aproxima más a // y_r. La ganancia esperada será reducida por cualquier interacción. Las condiciones óptimas implican que A_r disminuya monótonamente en lugar de ser igual.

Más recientemente, se ha hecho mucho para sistematizar el proceso entero de realización de pruebas nacionales de variedad de cosecha. Patterson y Silvey (1980) han descrito la versión británica de esto. La práctica actual debe algo a las ideas primitivas sobre la selección multiétapica, pero consideraciones de diseño experimental y de factibilidad administrativa plantean nuevos problemas.

4. ECONOMIA EXTERNA

En 1960, buscaba una base para decidir el tamaño óptimo de un esquema de selección en relación con sus beneficios para la economía nacional. ¿Cuáles deberían ser los recursos totales?. ¿Cuántas variedades deberían entrar en la etapa 1, como cohorte anual?. ¿Cuántas etapas deberían utilizarse teniendo presente que el aumento de k retrasa el aprovechamiento de las ganancias?. El principio es fácil, pero los resultados dependen de parámetros cuya predicción no es fácil. Supongamos que el remplazamiento de una variedad por otra nueva, que por término medio origina un incremento de producción de 1.0 por unidad de area, aportará un beneficio monetario W para el área total de / la cosecha. Escribamos U como el costo de aumentar A en / una unidad y V como el coste de producción de una variedad extra para ensayo. De aquí, representando todavía con G la producción incrementada por unidad de área,la ganancia neta

$$T = WG - UA - VN \tag{4.1}$$

aumentará al máximo.

Mis cálculos para diversas razones U:V:W se limitaron a k = 1.Mostraban condiciones óptimas para proporcionar valores de T, relativamente insensibles al cambio en U:V, pero para los valores estudiados un aumento de 5 veces en W, producía un aumento de 6 a 7 veces en T. Curnow (1961) estudió también k = 2; sus máximos para T fueron 15-20% mayores que para k = 1. Existen razones contra el uso de k grande. Las consideraciones de confianza en la comunidad agrícola, muestran los peligros tanto de prematuridad como de retraso excesivo en la distribución de una nueva variedad. Según aumenta k, la organización se hace más complicada, de manera que aumentan los gastos administrativos y los cargos por intereses sobre una inversión mayor. Lo que es más importante, para un programa contínuo de perfeccionamiento de variedades, que / empieza una nueva cohorte cada año, es el hecho de que / el aumento de k supone que transcurre un tiempo mayor antes de que la obtención de una buena variedad contribuya al término WG en (4.1).

5. SELECCION DE ANIMALES

Cuando la "producción" de un animal se mide sólo en / las hembras, como en la producción de leche, la aptitud de un macho para la reproducción ha de estimarse por medio de pruebas de progenie. Es decir, debe ser apareado con un número de hembras y la producción de sus crias-// hembras debe medirse en condiciones comparables.Supuesto que cada semental tenga hembras representativas de la misma población, las medias de la prole pueden guiar la selección de los sementales. Restricciones sobre el programa es probable que sean el número total de crias que puede ensayarse en una generación y el número de sementales a seleccionar como padres de una nueva generación . El factor variable es el número de sementales a probar / en cada generación y el número consiguiente de crias por semental.

Robertson (1057, 1960) estudió esto desde un punto de vista genético y obtuvo una regla de optimización que es esencialmente la misma que para la selección monoetápica de variedades de la Sección 3. La selección // multietápica en animales plantea mayores problemas que / en las plantas, debido al solapamiento de las generaciones y a que son mayores.

6. CONTROL EXHAUSTIVO DE MEDICAMENTOS

He discutido extensamente la selección de variedades / (aunque con poco detalle) porque es la aplicación que mejor conozco. Lógicamente surgen problemas semejantes en el control exahustivo de nuevos compuestos químicos en // cuanto a su posible actividad terapéutica. También en este caso, se producen muchos con facilidad y a bajo precio en cantidades adecuadas para análisis, pero muy pocos mejorarán una condición patológica especificada. Davies /// (1958) y Armitage y Scheneiderman (1958) iniciaron los estudios sobre esto. Davies hizo énfasis en que era conveniente una distribución muy diferente para x. Propuso como aproximación una distribución binomial entre una pequeña proporción, θ, de compuestos efectivos, pudiendo ser θ del orden de 0.01 o menor. El propósito del control exahustivo es producir una concentración mucho mayor de compuestos "buenos".

Dunnet (1961) discutió varios criterios. Previó la utilización de una unidad de análisis (tal vez en un animal o en un cultivo bacteriano) para producir una respuesta que tuviera esperanzas x_0, x_1 para los compuestos no / efectivos y efectivos, respectivamente. Las respuestas individuales, y, se distribuyen normalmente con respecto a x_0 ó x_1 con varianza σ^2. Las probabilidades P_0, P_1 de que un conjunto de medias de y para un compuesto sea superior a valores determinados se escriben fácilmente como / las áreas de los extremos de una distribución normal tipificada multivariante. Supongamos ahora que una serie de / "utilidades" puede establecerse sobre una escala convenida. Supongamos que a sea la ganancia por compuesto activo aceptado, b_1 la pérdida por aceptar un compuesto inactivo, b_2 la pérdida por rechazar un compuesto activo y c el coste por unidad de análisis. La ganancia esperada por // compuesto analizado será:

$$G = aP_1\theta - b_1P_0(1-\theta) - b_2(1-P_1)\theta - c\bar{n} , \tag{6.1}$$

siendo $\bar{n}$ el número promedio de unidades de análisis por compuesto. Dunnett propuso varias formas para b_2. Se puede suponer:

$$b_2 = [b_1P_0(1-\theta) + c\bar{n}]/P_1\theta , \tag{6-2}$$

para el coste promedio por compuesto activo aceptado, o

$$b_2 = c\bar{n}/P_1\theta , \tag{6.3}$$

el costo promedio de los ensayos por compuesto activo /// aceptado, o

$$b_2 = [a(1-P_1)\theta + b_1 P_0(1-\theta) + c\bar{n}]/P_1\theta \qquad (6.4)$$

que incluye también la ganancia perdida por rechazar compuestos activos. Toda propuesta de selección multietápica debe basarse en la maximización de G, con b_2 definido por (6.2), (6.3) ó (6.4). Dunnett mencionó otras posibilidades, tales como maximizar el número esperado de compuestos activos aceptados por coste unitario de los análisis, $P_1\ \theta/c\bar{n}$, que es el principio que utilizó Davies.

La principal diferencia de la Sección 3 en el patrón de selección de Dunnett es que defendía un punto de corte predeterminado en cada etapa. En la etapa r, en su esquema de k etapas, se miden n_r respuestas para cada // compuesto que sobrevive; $\bar{Y}rw$ para cualquier compuesto es ahora la respuesta media simple de los $(n_1 + n_2 + \ldots + n_r)$ análisis hechos hasta la fecha, y sólo aquellos compuestos, para los cuales

$$\bar{y}_{rw} > \eta_r$$

continuan en la siguiente etapa, siendo η_r una cantidad fija. El problema formal a resolver es, por consiguiente, el de la elección de n_r y de η_r (r = 1, 2,..., k) a fin de optimizar el criterio adoptado.

7. EL ENFOQUE DE BECHHOFER

En una serie de interesantes artículos, Bechhofer y / sus colegas han estudiado la selección con el propósito de identificar la mejor entidad individual. Comenzaron / (Bechhofer 1954; Bechhofer *et al* 1954) con una teoría relativa a la clasificación jerárquica de poblaciones /// (equivalente a "entidades" en la presente exposición) en base a una sola muestra o a una secuencia de dos muestras. Más adelante (Bechhofer 1958; Bechhofer y Blumenthal, // 1962) estudiaron un proceso secuencial clásico para identificar el primero en la clasificación; no rechazaban // hasta la etapa final y terminaban con $N_{k+1} = 1$. Su criterio era la probabilidad de que la selección final fuese la entidad con el máximo x de entre los N_1 iniciales. En una relación definitiva de estos temas y otros relacionados, Bechhofer *et alt* (1968) han generalizado la teoría al caso $N_{k+1} > 1$.

El trabajo de Bechhofer no ha implicado ninguna // distribución sobre la capacidad de producción (x de las secciones anteriores), y ha utilizado en cambio consideraciones mini-max o semejantes. Pero la idea de Bechhofer puede generalizarse fácilmente; ¿por qué no proceder como en las secciones 2, 3, con las mismas formulaciones de distribuciones o errores, pero buscando maximizar la probabilidad de que la N_{k+1} final incluya la x mayor?.La cuestión está estrechamente relacionada con la estudiada por Dunnett (1960), aunque éste se preocupaba de la elección de un tamaño de muestra para una sola muestra. A // primera vista, la concentración de la atención en la probabilidad de selección del mejor es atractiva, pero dudo que tenga importancia para la selección de variedades o incluso para el control exhaustivo de medicamentos. Mis razones son:

(a) Si N_{k+1}/N_1 es pequeño, la probabilidad de inclusión es probable que sea tan pequeña (incluso para su valor máximo) que haga despreciable la posibilidad de obtener realmente la x mayor; normalmente para variedades y para medicamentos N_{k+1}/N_1 es del orden de 0.01, 0.001, ó menos.

(b) Alternativamente, si se exigiera una probabilidad de éxito razonable, la replicación necesaria sería inaceptablemente grande o el número requerido de etapas tan grande, como para hacer la selección final desactualizada debido / al permanente perfeccionamiento entre nuevas / cohortes de entrada de N_1.

(c) Si N_1 es moderadamente grande, habrá poca pérdida si la selección termina sólo con la segunda o tercera x mayor, aunque el criterio de maximazación no tiene en cuenta esto.

(d) Los cálculos para hallar las adecuadas reglas de rechazo serían intolerablemente laboriosos para cualquier N_1 grande; el ejemplo de Bechhofer se refiere a $N_1 < 10$.

Sin embargo, la maximización de la probabilidad de selección perfecta y la maximización de la media de los N_{k+1} valores finales de x debe requerir decisiones semejantes en cada etapa, de manera que un plan casi óptimo para uno de ellos, será casi ciertamente bueno para / el otro. En particular, para cualesquiera N_r, A_r, parece intuitivamente obvio que la selección de los N_{k+1} valores mayores de y_r (θr $\bar{y}_{rw}$) debe ser óptimo para el criterio de la probabilidad. Realmente el criterio de la probabilidad es equivalente a la transformación de los valores de x con el fin de substituir los N_{k+1} valores mayores de x por 1 y los $(N_1 - N_{k+1})$ restantes por 0. Las Tablas 7.1 y 7.2 con $N_{k+1} = 1$ proceden de las mismas simulaciones que produjeron las Tablas 3.1 y 3.2. La semejanza del patrón confima que para este ejemplo pequeño, la elección óptima de N_2, A_1. es realmente igual para la / probabilidad máxima de selección "correcta", que para la maximización de la esperanza.

8. SELECCION HUMANA

Si un compuesto químico muestra poco beneficio tera-- peútico, será desechado; el conocimiento de cómo volverlo a sintetizar permanece, pero ninguna noción de "imparcialidad" para el compuesto impide su remoción a partir de consideraciones en el presente contexto. Si una nueva variedad de trigo no logra el nivel de selección enninguna etapa, probablemente será desechada a menos que características especiales exijan su retención como material de reproducción. En ciertas circunstancias, en especial cuando se seleccionan seres humanos, la situación es muy diferente.

Buscando un ejemplo, intenté estudiar la selección de jóvenes en los sistemas educativos (1962b). Me preocupaba sólo reconocer que cierta selección educativa, o separación en diferentes canales es inevitable y examinar / las consecuencias de hacerlo de una manera que no fuese / irrevocable, y que tratase de optimizar respecto a un propósito explícito. La idea de "destacar" debe abandonarse y cualquier solución estadística debe interpretarse flexible y humanamente. Ningún país puede permitirse omitir de su sistema docente *bien mediante* medidas para el trata- / miento justo de cada joven de acuerdo a las necesidades y al potencial, *bien mediante* una política planificada para asegurar que el agrupamiento de habilidad se desarrolla / para beneficio de la comunidad.

Discutí un sistema de selección bietápico. Supongamos que cada niño tiene una habilidad inherente x, y que algún test al final de la educación primaria proporciona una evaluación, y, que está sujeta a error. Sobre la base de y_1, los niños se distribuyen entre un sistema secundario académico y uno no académico en proporciones P_1, y // $(1-P_1)$. Unos años después, la admisión a la universidad u otra forma de educación terciaria se basa en una nueva // evaluación, y_2 y las proporciones P_2, P_2^* se aceptan para las dos corrientes. En este caso P_2^*, mucho menor que P_2 / se destina a "red de seguridad" para los que fueron puestos erróneamente en la corriente no académica en la primera etapa. La proporción total que entra en la universidad es:

$$\Pi = P_1 P_2 + (1-P_1) P_2^* \qquad (8.1)$$

Consideré que π estaba establecido por factores externos, y a continuación búsqué una elección de P_1, P_2 y P_2^* que fuese "óptima", una palabra que quedaba por definir. El / criterio obvio era la maximización del valor medio de x / entre los seleccionados para la universidad; se utilizó / éste, pero con completo conocimiento de sus fallos.

No presentaré resultados de un trabajo que, mirando hacia atrás, pueden parecer haberse basado en una formulación excesivamente simplificada. Una diferencia importante de los tipos de selección discutidos anteriormente es que cualquier enfoque realista tendría que permitir cambios en x entre las dos etapas. Mi propósito era llamar / la atención sobre un problema y sobre la posibilidad de un estudio objetivo. Incluso el intento de formular el problema en términos estadísticos es clarificador, ya que divulga incertidumbre acerca de los propósitos e ignorancia de hechos relevantes tales como el tamaño de las correlaciones. Mi enfoque puede haber sido inadecuado. Ningún // país puede permitirse despreciar el problema de asegurar que los jóvenes que están especialmente dotados sean admitidos a los niveles de educación más altos. Cómo debe lograrse esto es discutible, pero estoy convencido de que la optimización de un proceso de selección multietápico / apropiado debe ser un principio utilizado como guía. Los dilemas principales son: ¿Qué "mediciones" deben hacerse? ¿Qué criterio debe optimizarse?.

TABLA 3.1

Ganancias Netas Medias para Muestreo Finito, $N_1 = 16$, $N_3 = 1$, $K = 2$, $\gamma = 1.0$, Selección Basada sobre y_r (expresada como múltiplo de σ, y calculada a partir de 5000 simulaciones por entrada).

Valores de N_2

A_1/A	1	2	3	4	5	6	8	10	12	14	16
0.0	0	0.532	0.777	0.921	1.015	1.081	1.162	1.207	1.232	1.244	1.249
0.2	0.721	1.113	1.242	1.315	1.348	1.364	1.345	1.318	1.280	1.223	1.177
0.4	0.944	1.264	1.342	1.384	1.387	1.366	1.324	1.277	1.206	1.145	1.081
0.6	1.081	1.336	1.376	1.389	1.351	1.332	1.241	1.188	1.084	1.020	0.944
0.8	1.177	1.342	1.336	1.320	1.249	1.216	1.090	0.999	0.898	0.794	0.721
1.0	1.249	1.079	0.952	0.849	0.760	0.680	0.538	0.408	0.283	0.154	0.

Todas las entradas para $A_1/A = 0{,}0$, $A_1/A = 1{,}0$, $N_2 = 1$, $N_2 = 1$, $N_2 = 16$ son exactas.

Entrada para $A_1/A = 0.5$, $N_2 = 4$ es 1.388 (10000 simulaciones).

TABLA 3.2

Ganancias Netas Medias para Muestreo Finito, N_1 = 16, N_3 = 1, k = 2, γ = 1.0, Selección Basada en y_r (expresada como múltiplo de σ, y calculada a partir de 50000 simulaciones por entrada)

Valores de N_2

A_1/A	1	2	3	4	5	6	8	10	12	14	16
0.0	0	0.532	0.777	0.921	1.015	1.081	1.162	1.207	1.232	1.244	1.249
0.2	0.721	1.113	1.242	1.320	1.358	1.372	1.366	1.340	1.324	1.268	1.249
0.4	0.944	1.264	1.346	1.395	1.399	1.385	1.379	1.327	1.303	1.266	1.249
0.6	1.081	1.340	1.386	1.405	1.389	1.382	1.351	1.311	1.290	1.249	1.249
0.8	1.177	1.351	1.365	1.378	1.349	1.344	1.306	1.290	1.278	1.241	1.249
1.0	1.249	1.249	1.249	1.249	1.249	1.249	1.249	1.249	1.249	1.249	1.249

Todas las entradas para A_1/A = 0.0, A_1/A = 1.0, N_2 = 1, N_2 = 16 son exactas.

Entrada para A_1/A = o.5, N_2 = 4 es 1.402 (10000 simulaciones)

TABLA 3.3

Varianza de Ganancias Netas para Muestreo Finito, N_1 = 16, N_3 = 1, k = 2, γ = 1,0 Selección Basada en y_r (expresada como múltiplo de σ^2, calculada a partir de 5000 simulaciones por entrada)

Valores de N_2

A_1/A	1	2	3	4	5	6	8	10	12	14	16
0.0	1.000	0.717	0.629	0.593	0.579	0.575	0.582	0.597	0.614	0.631	0.648
0.2	0.883	0.647	0.583	0.574	0.556	0.549	0.584	0.597	0.624	0.639	0.687
0.4	0.799	0.599	0.567	0.552	0.547	0.552	0.606	0.620	0.636	0.678	0.736
0.6	0.736	0.573	0.557	0.562	0.562	0.571	0.632	0.665	0.688	0.736	0.799
0.8	0.687	0.586	0.582	0.595	0.603	0.610	0.672	0.711	0.737	0.797	0.883
1.0	0.648	0.646	0.652	0.661	0.671	0.683	0.711	0.746	0.793	0.861	1.000

Todas las entradas para A_1/A = 0.0, A_1/A = 1.0, N_2 = 1, N_2 = 16 son exactas.

Entrada para A_1/= 0.5, N_2 = 4 es 0.547 (10000 simulaciones)

TABLA 7.1

Proporción de simulaciones en las cuales la selección final es la x más grande, $N_1 = 16$, $N_3 = 1$, $k = 2$, $\gamma = 1.0$ Selección Basada en y_r (calculada a partir de 5000 simulaciones por entrada)

Valores de N_2

A_1/A	1	2	3	4	5	6	8	10	12	14	16
0.0	0.064	0.124	0.171	0.215	0.242	0.280	0.330	0.351	0.363	0.368	0.375
0.2	0.191	0.311	0.361	0.407	0.420	0.427	0.425	0.412	0.382	0.354	0.341
0.4	0.253	0.384	0.421	0.450	0.444	0.438	0.424	0.388	0.340	0.318	0.302
0.6	0.302	0.421	0.444	0.449	0.420	0.415	0.377	0.346	0.288	0.276	0.253
0.8	0.341	0.428	0.419	0.407	0.365	0.350	0.300	0.264	0.223	0.204	0.191
1.0	0.375	0.285	0.220	0.199	0.164	0.152	0.122	0.097	0.079	0.078	0.064

Entrada para $A_1/A = 0.5$, $N_2 = 4$, es 0.447 (10000 simulaciones)

TABLA 7.2

Proporción de simulaciones en las cuales la selección final es la x más grande

$N_1 = 16$, $N_3 = 1$, $k = 2$, $\gamma = 1.0$, Selección Basada en $\bar{Y}_{rw}$

(calculada a partir de 5000 simulaciones por entrada)

Valores de N_2

A_1/A	1	2	3	4	5	6	8	10	12	14	16
0.0	0.064	0.124	0.171	0.215	0.242	0.280	0.330	0.351	0.363	0.368	0.372
0.2	0.191	0.311	0.362	0.410	0.425	0.431	0.438	0.422	0.411	0.381	0.386
0.4	0.253	0.384	0.426	0.458	0.452	0.449	0.453	0.411	0.400	0.387	0.383
0.6	0.302	0.426	0.451	0.461	0.442	0.446	0.438	0.400	0.389	0.379	0.382
0.8	0.341	0.435	0.434	0.442	0.418	0.422	0.406	0.390	0.385	0.377	0.386
1.0	0.375	0.378	0.377	0.378	0.374	0.387	0.382	0.378	0.380	0.364	0.387

Entrada para $A_1/A = 0.5$, $N_2 = 4$ es 0.458 (10000 simulaciones)

X. EL ESTADISTICO PREGUNTON

1. INTRODUCCION

Les propongo terminar mis conferencias con una discusión del papel de los estadísticos como asesores y colaboradores en el diseño de experimentos (Finney 1982a).Lo haré refiriéndome a 22 preguntas que son las típicas que yo discutiría con un colega de otra disciplina científica que me pidiese ayuda para su programa experimental.

Con mucha frecuencia, el estadístico está considerado como alguien que aparece después de que han sido recopilados los datos, realiza cálculos típicos da un veredicto de "significativo" o "no significativo" y se va.Esto, debería ser totalmente falso. Un estadístico necesita estar implicado en todas las etapas de una investigación; a menos que pueda también trabajar codo a codo con los otros expertos en el estudio, no se aprovecharán al / máximo sus propios conocimientos. No es fácil ser específico en detalles de tal colaboración, ya que tanto el // campo de aplicación como la experiencia personal son importantes. Me referiré solo a la planificación y diseño de *experimentos comparativos*, es decir, a las investigaciones para comparar dos o más tratamientos o categorías en función de mediciones u observaciones sobre sujetos y lo explicará refiriéndome a experimentos biológicos.Tendré poco que decir sobre análisis estadístico.

2. COLABORACION CON EL INVESTIGADOR

Un biólogo puede verse a si mismo como una persona // que simplemente está deseando hacer a un estadístico // unas cuántas preguntas críticas. En la práctica, una primera etapa más útil puede consistir en *preguntas del estadístico*, porque él tiene un conocimiento más amplio de los temas de discusión interesantes para el desarrollo / efectivo de su disciplina. Dichas preguntas, deberían // ser una base para la discusión más que para respuestas exactas. ¡Ciertamente no hay forma de que un conjunto de respuestas puedan generar automáticamente un plan de trabajo estadístico.

Algunas de mis 22 preguntas son fáciles y evidentes, otras quizás requieran un estudio detallado y respuestas complejas. Algunas respuestas, y también determinadas circunstancias puedan sugerir más temas a examinar en la etapa de planificación. Las pruebas que implican a seres humanos introducen generalmente dificultades especiales en sus aspectos éticos y organizativos, las cuales pueden exigir una discusión más profunda que las Preguntas 5, 10 y 18. Todas deben estar claras en las mentes del científico experimental y de su colega estadístico.Muchas están implícitas en el excelente libro de D.R. Cos / (1952), que contiene muchos más detalles. No sugiero que el conjunto de las 22 sea completo: un estadístico con diferente experiencia podría añadir más preguntas o agrupar estas 22 de otra manera.

3. ¿QUE EXPERIMENTO?

PREGUNTA 1 ¿ES REALMENTE UN EXPERIMENTO LO QUE SE -- PLANEA?

La característica esencial es que el experimentador tiene el poder de determinar qué sujetos deberán recibir/ qué tratamientos.

Para elegir entre 50 conejos cuáles recibirán cada una de varias dosis de un antibiótico; no puede elegir entre 50 jóvenes adultos quiénes fumarán 20 cigarrillos al día durante 10 años y quiénes no serán fumadores. Puede / elegir qué ratas de un laboratorio serán expuestas a un / cancerígeno y cuáles no lo serán, pero no puede elegir // qué miembros de una población de ratas en estado salvaje estarán expuestos a un vertido accidental de residuos tóxicos. Llamar la atención sobre las debilidades lógicas / inherentes a varios tipos de "no experimento" (estudios / de control de casos retrospectivos, datos de observaciones no planificados, información voluntaria) no es condenarlas. Armitage (1981) expresa este punto sucíntamente : "Si existe una prueba controlada aleatorizada que pueda / ponerse en práctica y es ética, en mi opinión es una mala segunda opción confiar en comparaciones no aleatorizadas. En la mayoría de los estudios etiológicos, por otra parte, no puede realizarse la aleatorización y sería un disparate despreciar un control de casos controlados cuidadosamente o un estudio de cohortes". En Epidemiología, como en Ecología y Astronomía (aunque por diferentes razones), la interpretación de datos no experimentales puede ofrecer la única esperanza de progreso. Pero no analizar tales datos críticamente como si procedieran de un experimento planificado incita al sofisma y a conclusiones erróneas. Los problemas de inferencia implican muchas más dificultades (Anderson et alt, 1980; Cochran 1965, 1968).

PREGUNTA 2 ¿POR QUE HACER EL EXPERIMENTO?

Se pueden distinguir cuatro principales razones:

(A) Curiosidad. ¿Qué ocurrirá si se utiliza una substancia, componente, artículo, método etc., en lugar de otro?

(B) Interés claro en la comparación de esta clase de sujetos. ¿Protejerá más de una enfermedad a los cerdos una nueva vacuna que la existente?

(C) Intento de transferir conclusiones, circunstancias. Los medicamentos antitumorales pueden compararse / primero en mamíferos de laboratorio, aunque el único propósito de la investigación es el de mejorar / la salud y la nutrición humana.

(D) Los sujetos son instrumentos para calibrar los tratamientos. En los ensayos biológicos, los animales/ o cultivos bacterianos se tratan diferencialmente / no con el fin de medir los efectos del tratamiento/ sobre los sujetos, sino para calcular qué cantidad de un material producirá la misma respuesta que 1 / mg ó 1 ml de otro (Finney 1978).

En este punto, es necesaria una evaluación del trabajo / anterior en los mismos campos y otros estrechamente relacionados.

4. LAS UNIDADES EXPERIMENTALES

PREGUNTA 3 ¿QUE UNIDAD EXPERIMENTAL DEBE UTILIZARSE?

La unidad que debe seleccionarse para el tratamiento se denomina comunmente *parcela*. Puede tratarse de un / paciente, de todos los pacientes de una sala, de un solo ratón, de un grupo de animales o insectos a los que debe aplicarse el mismo tratamiento, del lugar de inoculación/ en un animal vivo, del conjunto de plantas de trigo en un área de terreno de 10 m x 2 m, de una placa de cultivo // bacteriano, o incluso de un intervalo de tiempo determinado para cualquiera de ellas (donde una secuencia de tiempo de tratamientos diferentes puede aplicarse a una unidad biológica).

Es esencial una definición exacta. Surgen errores de bulto, si los pacientes individuales de una sala, los animales de un corral o los insectos de una colonia se consideran como "parcelas", cuando de hecho los tratamientos / se aplican necesariamente a través de la unidad experimental más grande. El ambiente común de la sala o la colonia es probable que afecte a sus miembros individualmente y / destruya así la independencia de sus respuestas por separado al tratamiento.

PREGUNTA 4 ¿SON LAS UNIDADES "DE IGUAL TAMAÑO"?

Si las unidades constan de diferente número de animales o de longitudes de tiempo diferentes o si tienen / diferentes dimensiones físicas, las implicaciones del // análisis estadístico necesitan considerarse por anticipado. Las dificultades técnicas pueden obligar a modificar un programa de ordenador a un desarrollo secundario de / la teoría estadística.

PREGUNTA 5 ¿ESTAN AGRUPADAS LAS UNIDADES DE TAL MANERA QUE EL COMPORTAMIENTO DE QUE LOS MIEMBROS DE UN MISMO GRUPO ES MÁS PARECIDO QUE EL DE LOS MIEMBROS DE GRUPOS DIFERENTES?

Las agrupaciones podrían ser plantas próximas, animales de una camada, tiras del músculo de un animal, y // así sucesivamente. Esta característica es importante para la estructura de los bloques en el diseño. Puede haber // dos agrupaciones potencialmente útiles y, por consiguiente, existe la necesidad de evaluar sus méritos relativos/ y hacer una elección, o considerar cuadrados latinos.

PREGUNTA 6 ¿CONTIENEN DICHOS GRUPOS IGUAL NUMERO DE / UNIDADES?

Si deben utilizarse, bloques de tamaños diferentes, puede haber mayores complicaciones en la discusión de un buen diseño. Estas no son insuperables, pero se deben tener presentes.

PREGUNTA 7 ¿PUEDEN SUBDIVIDIRSE LAS UNIDADES EN EL ESPACIO O EN EL TIEMPO, PARA COMPARACIONES / POSTERIORES DEL TRATAMIENTO?

Para la nutrición, el animal entero debe ser la unidad, pero las reacciones de la piel a diferentes materiales pueden ensayarse en un animal. Para el tratamiento de las plantas contra las enfermedades transmitidas por / la simiente deben utilizarse plantas completas o grupos / de plantas como parcelas, pero pueden compararse fungicidas en hojas individuales. Antes de utilizar el diseño de *parcelas divididas*, debe considerarse la posibilidad de / que el tratamiento aplicado a una subunidad pueda influir en las mediciones hechas en otras subunidades de la misma unidad principal.

5. LOS TRATAMIENTOS

PREGUNTA 8 ¿COMO ESTAN ESTRUCTURADOS LOS TRATAMIENTOS PROPUESTOS?

Hay cuatro tipos principales:

(P) Sin estructura - por ejemplo, muchos procedimientos diferentes (tal vez variedades trigo),sin patrón o jerarquía entre ellos, que deben compararse.

(Q) Estructura mínima - tal vez alguna agrupación de tratamientos, por ejemplo, varios medicamentos/ nuevos que deben compararse con uno típico, aunque sólo son de interés comparaciones de "nuevo versus típico".

(R) Estructura lineal - Los tratamientos abarcan un dominio unidimensional cualitativo, por ejemplo/ "ligero", "suave", ó cuantitativo como en las curvas de respuesta estudiadas y estimadas utilizando los tratamientos 0, 1, 2, 3, 4,... (o 2, 5, 10 , 25,...)

(S) Estructura factorial - combinaciones de diferentes tratamientos. Se pueden tomar diferentes conjuntos de tratamientos de los tipos P, Q, R, y utilizar todas las combinaciones; por ejemplo, un / medicamento típico y dos huevos, ensayados, cada / uno de ellos, con las dosis 0, 1, 2, 4, es decir, un factor de tipo Q y // otro de tipo R.

Con los seres humanos, la estructura será realmente sencilla y el número de tratamientos rara vez pasará de cuatro; una característica subsidiaria de la estructura puede permitir el diseño cruzado de los tratamientos durante el experimento. Con seres no humanos y, en especial , en la investigación agrícola e industrial, la estructura del tratamiento puede implicar el uso de patrones combinatorios mucho más complejos.

PREGUNTA 9 ¿SON RIGIDOS LOS REQUISITOS PARA LOS TRATAMIENTOS?

Si se proponen 17 tratamientos no estructurados o / escasamente estructurados ¿habría una objeción seria para incluir sólo 16 ó aumentarlos hasta 18 ó 20?. Para el tipo Q, ¿tendría alguna ventaja incluir un tratamiento típico con doble o triple replicación?

Si la estructura es lineal, se pueden discutir el / número y la separación de los niveles?

Si deben incluirse varios factores ¿existe algún // problema al ajustar todas las combinaciones?. Por ejem- / plo, 3 medicamentos, 4 niveles de dosis, y 2 métodos de administración dan 3 x 4 x 2 = 24 combinaciones, posiblemente demasiadas para los recursos. El diseño experimen-- tal es más facil, y de diversas maneras más satisfacto- / rio, si el número total de tratamientos (cuando es gran-- de), es un cuadrado o se descompone fácilmente en facto-- res 16 = 4^2, 20 = 4 x 5), y también para la estructura // factorial si todos los factores tienen el mismo número de niveles; quizás el estadístico prefiera 3 x 3 x 3 a 3 x 4 x 2, pero esto no es obligatorio. Los factores todos a 2 ó todos a 3 niveles, son especialmente adecuados para la confusión y la replicación fraccionaria. La elección del número y de la división en niveles puede tener grandes // consecuencias para la precisión de determinadas estimaciones. Como en otras cuestiones, la decisión debe provenir del experimentador, pero el estadístico debe aceptar la / responsabilidad de mostrar los méritos y debilidades de / las varias posibilidades.

6. ALEATORIZACION

PREGUNTA 10 ¿EXISTEN RESTRICCIONES EN LA ASIGNACION // ALEATORIA DE LOS TRATAMIENTOS A LAS UNIDADES?

La aleatorización es uno de los conceptos más importantes que los estadísticos han introducido en la experimentación. Cualquier desviación de la asignación completamente aleatoria de los tratamientos a las unidades es una cuestión que debe examinarse cuidadosamente en la etapa / de planificación. Las restricciones de confusión y blo- / ques adoptadas por excelentes razones relacionadas con la naturaleza de las unidades y la optimización de la precisión exigen restricciones muy organizadas sobre la aleatorización. Un experimento complejo puede tener fases de // aleatorización independientes pero entrelazadas, en lugar de una extracción total de los lotes. Estos recursos no destruyen en absoluto la validez de las inferencias siempre que se recuerde que el sistema de aleatorización determina la estructura del análisis estadístico.

Para determinados experimentos particulares, pue-- den surgir objecciones acerca de la aleatorización. Una objeción ética es a menudo, en realidad, una objeción a la experimentación más que a la aleatorización, posiblemente porque se considera inadecuado dejar a algunos sujetos o unidades experimentales como controles no tratados. Otras objeciones pueden proceder de un deseo de evitar poner ciertos pares de tratamientos en unidades adyacentes, o de conveniencias operativas, si los tratamientos están en un cierto orden. Toda objeción merece discusión: normalmente algún compromiso puede mantener la validez estadística dentro de restricciones inevitables. / Ambigüedades en la interpretación de los experimentos // han sido ocasionadas a menudo por no prestar suficiente atención a la aleatorización.

7. OBJETIVOS

PREGUNTA 11 ¿CUALES SON LOS OBJETIVOS Y PRIORIDADES?

¿Está el experimento destinado principalmente a // descubrir diferencias significativas (por ejemplo, ¿es el efecto de un medicamento sensible al vehículo en el / que se administra?) o a estimar propiedades numéricas // (diferencias en la producción potencial entre variedades de maiz, potencia relativa de dos medicamentos u otros / parámetros?). Si hay varios objetivos ¿cómo compararlos/ en importancia?.

PREGUNTA 12 ¿QUE VARIABLES ALEATORIAS DEBEN MEDIRSE?

Debe evaluarse sólo el experimento, o principalmente con respecto a una variable aleatoria (cantidad de leche, peso uterino, tiempo de supervivencia) o hay varias de interés comparable. Si hay muchas, ¿Podría omitirse / alguna o casi duplicarse otras (por ejemplo, diferentes técnicas para medir lo que es cualitativamente una característica)?. ¡No tiene sentido recopilar la misma información de las mismas parcelas dos veces!

PREGUNTA 13 ¿QUE VARIABLES ALEATORIAS DEBEN ANALIZARSE?

¿Debe analizarse por separado cada variable aleatoria medida? ¿Deben utilizarse combinaciones (porcenta- / jes, razones, índices, "valores corregidos", etc.)? ¿Son todas esencialmente distintas? ¿Por qué se analizan tanto el peso final como el aumento de peso si los pesos // iniciales no difieren significativamente? ¿Debe someterse cada variable aleatoria a un análisis estadístico completo (por ejemplo, análisis de varianza), o serán suficientes sencillas tabulaciones para algunas? Un experi-- mento nutricional, por ejemplo, puede tener fácilmente / variables aleatorias relacionadas con la ingestión de comida, crecimiento, comportamiento, y diversos aspectos / del metabolismo),¡muchas más que los seres humanos o animales sometidos a la prueba!. Cualquier experimento en animales o seres humanos que impliquen la vigilancia continua de ciertas funciones puede producir todavía un número mucho mayor de variables aleatorias. Sin embargo, / por muy importante que sea un experimento no debe permitirse que genere grandes cantidades de salidas de ordenador que no se leerán nunca;esto,no sólo supone un derroche de esfuerzo,sino que una información excesiva para / distraer la atención de los hallazgos más valiosos.

Si se piensa hacer un análisis multivariante ¿cuál es su propósito? ¿Reside su interés en una función de regresión estimada, en el establecimiento de una técnica / de discriminación, en la búsqueda de variables aleato- / rias canónicas útiles e interpretables, en el exámen de clasificaciones o en algo diferente?

8. PRECISION

PREGUNTA 14 ¿CUAL DEBE SER LA PRECISION DE LOS RESULTADOS?

¡Decir "Muy" o "lo más preciso posible" no sirve / de ayuda!. Si el diseño es bueno, la precisión puede incrementarse solamente aumentando el tamaño del experimento, es decir, incluyendo más unidades y aumentando los / costos. En consecuencia, por lo menos para las variables aleatorias más importantes,los requisitos mínimos de precisión quizás necesiten establecerse en función del error típico de una diferencia entre las medias, del recorrido del error con una probabilidad predeterminada para un parámetro estimado, o de una potencia para una dócima de / significación. El experimento, pues, debe planificarse o bien para satisfacer todas las demandas simultáneamente, o bien para que esté de acuerdo con algún compromiso previo.

PREGUNTA 15 ¿QUE SE SABE DE LA VARIABILIDAD ENTRE LAS UNIDADES?

Experimentos semejantes, o nociones generales so-- bre el recorrido de los valores de replicación, pueden /

facilitar una estimación de la varianza por unidad para cada variable aleatoria importante que deba analizarse/ eventualmente. Como máximo,esta información puede acep-- tarse sólo como guía aproximada, pero puede ayudar en / la planificación.

PREGUNTA 16 ¿PUEDE MEDIRSE O REGISTRARSE UTILMENTE / CUALQUIER CONCOMITANTE?

La anotación de los pesos iniciales o de algunas / características funcional o propiedad de las unidades *antes de que comiencen los tratamientos* pueden utilizarse en ocasiones de manera efectiva en la tipificación de // las variables aleatorias, mejorando la comparabilidad y reduciendo, por lo tanto, la varianza efectiva. Incluso después de 50 años, el análisis de covarianza y las técnicas equivalentes para mejorar la precisión, siguen explotándose insuficientemente, excepto tal vez en experimentos agronómicos.

9. SERIES DE EXPERIMENTOS

PREGUNTA 17 ¿ES EL EXPERIMENTO UN ESTUDIO AISLADO O / FORMA PARTE DE UNA SERIE?

Si un experimento forma parte de una serie realizada al mismo tiempo en diferentes lugares, es conveniente la coordinación del diseño. Los diseños en lugares diferentes, pueden ser idénticos (excepto para la aleatorización) o con pequeñas variaciones en los tratamientos incluidos, o entrelazados de tal manera que cada uno represente parte del diseño total completo. Esto puede ser // particularmente importante en estudios clínicos a gran escala, y en la investigación agronómica destinada a servir de base de asesoramiento para una región.

Si un experimento forma parte de una secuencia en el tiempo, muchos aspectos del diseño de cada uno de /// ellos pueden ser necesario considerarlos en relación con lo que pueda aprenderse de sus predecesores.

10. RECURSOS Y RESTRICCIONES

PREGUNTA 18 ¿QUE RECURSOS PUEDEN UTILIZARSE?

¿Cuántas unidades pueden utilizarse? ¿Durante cuanto tiempo? ¿Qué limitaciones hay con todo el personal // del equipo? ¿Son los materiales (por ejemplo, suministros de substancias raras)una restricción? ¿Hay requisitos preestablecidos tales como "15 ratones como mínimo / bajo cada medicamento"? ¿Son absolutos, costosos de superar, o simplemente adecuados, los límites superiores de los recursos?. En investigaciones clínicas ¿Cuál es la / tasa aproximada a la que se dispondrán de nuevos casos?

La identificación de los límites reales en la planificación no es siempre evidente. Una estudiante de doctorado, me pidió una vez consejo para un experimento sobre las diferentes variedades de la hierba de los pastos ¿Qué tamaño de parcela y qué replicación debería utilizar para estimar las tasas de crecimiento por medio de cortes cí-- clicos?.Sólo al final de una discusión de una hora se vió la realidad. ¡La escala del experimento estaba totalmente restringida por la distancia que ella podría recorrer en una tarde empujando personalmente una pequeña cortadora / de cesped!

PREGUNTA 19 ¿HAY RESTRICCIONES DE TIEMPO?

A menos que el diseño requerido sea muy sencillo un estadístico necesitará tiempo para coger y digerir las // ideas y para hacer una buena propuesta. Puede hacer sugerencias provisionales durante una primera discusión con el investigador, pero quizás necesite varios días para // examinarlas con más detenimiento. En realidad, puede necesitar más tiempo que el que habría necesitado hace 30 // años, ya que hoy en día puede hacer un estudio con ordenador de las precisiones relativas de varias posiblidades , mientras que antes habría tenido que fiarse de su intui-- ción. En una emergencia, podría hacer un diseño aceptable muy rápidamente; un buen diseño sólo puede surgir después de semanas de atención intermitente al problema, durante este tiempo pueden plantearsele al investigador nuevas // cuestiones de factibilidad e idoneidad.

Puede ser necesario examinar el tiempo en relación con la disponibilidad de personal, equipo y otros recur-- sos. Puede preguntar acerca del tiempo atmosférico u /// otros factores ambientales incontrolables. Sin embargo, / también en este caso,si el análisis estadístico de los resultados no va a ser del tipo normal y moderado en cantidad,debe preguntar si el tiempo para la elaboración de // procedimientos analíticos y la ejecución de los análisis mismos (tiempo del estadístico y disponibilidad de ordenador) se han considerado en relación con cualquier urgen-- cia en la preparación de los informes.

PREGUNTA 20 ¿HAY LUGAR PARA DISEÑOS SECUENCIALES PLA-- NEADOS?

Si el tiempo entre la administración de los trata--

mientos y la terminación de las mediciones y evaluaciones de cada sujeto es corto, con relación a la tasa a la cual las nuevas unidades pueden estar preparadas para el tratamiento, (en pruebas clínicas, la tasa de reclutamiento de nuevos sujetos), puede considerarse la necesidad de un diseño secuencial formal.

11. REGISTRO Y ANALISIS

PREGUNTA 21 ¿COMO DEBEN REGISTRARSE LOS RESULTADOS?

Si las anotaciones son manuales, a partir de la observación y recuento directos o a partir de la lectura de instrumentos, ¿se toman las precauciones adecuadas contra los errores (errores de recuento y errores de lectura grandes, sesgos debido a mal uso o mala colocación de los instrumentos, etc.)?. ¿Se preservará la legibilidad, especialmente si las anotaciones se hacen al aire libre o con dificultades?. ¿Servirán de ayuda para una buena organización los formularios diseñados especialmente para hacer las anotaciones? ¿Recibirá el estadístico las anotaciones originales, una buena fotocopia, o (lo que es muy poco deseable) un manual o copia mecanografiada sin corregir?

Si se utiliza un equipo de registro de datos manual, mecánico o electrónico, ¿se ha comprobado su fiabilidad? ¿Se dispondrá de información adecuada para su interpretación? ¿Existen problemas de calibrado, porcentaje del error sistemático, etc.?

PREGUNTA 22 ¿QUE MAS SE NECESITA AHORA PARA EL ANALISIS SUBSIGUIENTE?

Es posible que el único paso inmediato sea el acuerdo de que dentro de 1 semana o de un 1 año a partir de ahora, se entregarán los resultados al estadístico. Posiblemente, deban hacerse preparativos para más discusiones sobre lo que debe ser analizado y cómo debe analizarse, y se admitirá que el algebra estadística y los programas de ordenador deben estudiarse *ahora* para que estén listos para el consiguiente análisis ¿Se dispone de un programa adecuado, debe escribirse uno especialmente, o bastará un programa existente con un poco de cálculo *ad-hoc* suplementario para un trabajo que es improbable que se repita?. Si se está planificando una serie importante de experimentos, casi con certeza, será esencial disponer de un amplio software para el almacenamiento de la información, manejo de datos, análisis estadístico y producción de resúmenes gráficos y tablas.

12. COMENTARIOS GENERALES

Más importante que la exposición exacta de las preguntas es el reconocimiento, por parte del estadístico y de sus colegas, de su papel de inquisidor en un campo amplio. Para cualquier problema complejo, es esencial una conversación personal desarrollada en varias ocasiones: las respuestas a un cuestionario impreso serían de mucho menos valor. He conocido investigadores que considerarían algunas de estas preguntas de un estadístico como impertinentes. Incluso la pregunta 3 puede evocar la respuesta ¡"esto no es cosa suya?!. Se de otros que presentarían ideas a un estadístico para un experimento complicado a las 4 de la tarde y querrían que el diseño pudiera ser operativo a las 9,30 de la mañana del día siguiente. Estas no son simplemente irritaciones para el estadístico ni material anecdótico. Son medios seguros obtener un asesoramiento pobre y a menudo totalmente inadecuado.

Nada de lo que he dicho implica que las fases de diseño y análisis deban mantenerse totalmente separadas; en realidad, comunmente, el posible análisis y las facilidades de ordenador que necesita deberían discutirse antes de que el diseño se decida totalmente. En ocasiones, los resultados de un experimento bien diseñado contienen un mensaje tan claro que el único análisis necesario es la producción de unos pocos valores medios. A veces, un estadístico experimentado puede apreciar que las técnicas usuales, principalmente el análisis de varianza, permitirán obtener casi toda la información importante de un experimento con el mínimo de problemas. En otras ocasiones es esencial una fase más extensa de preguntas y colaboración con el investigador.

Hace pocos años, supe que una importante organización de investigación utilizaba a su único estadístico, sólo como un ordenador con oídos y cuerdas vocales. El "analizaba" los datos que se le presentaban sin tener en cuenta su naturaleza, y devolvía resúmenes formales de sus hallazgos, sin preocuparse de su comprensibilidad. Jamás estaba implicado ni en la planificación ni en el diseño. No importa de quien es la culpa: el papel correcto de un estadístico es muy diferente. Debe ser un colaborador, no un servidor, participando extensa y profundamente en muchos aspectos de un programa de investigación. En ciertas investigaciones debe ser un socio cabal, contribuyendo con sus conocimientos y compartiendo la responsabilidad total en todas las etapas desde los planes provisionales hasta la publicación de los informes.

XI. REFERENCIAS

En los últimos años, no se ha publicado ningún // buen libro de carácter general sobre diseño de experimentos. Se necesita con urgencia un texto que tenga en cuenta los progresos de los últimos 25 años; diseños con fines especiales, diseños secuenciales, integración de diversas ideas (cruzamientos, elementos más próximos,etc.) con la corriente principal del diseño, planificación de series de experimentos en el espacio y el tiempo, y combinación de sus resultados. Por encima de todo, es necesario un libro que reconozca los cambios producidos por los ordenadores en la facilidad de generación de diseños y especialmente la mayor flexibilidad aceptable en la estructura combinatoria ahora que el trabajo del cálculo / no es ya una limitación. En relación con mi serie de conferencias, puedo recomendar solo cinco de perspectivas / suficientemente amplias. Cox (1959) soberbia relación de temas como los que mencioné brevemente en la Conferencia X, Cochran y Cox (1957) y Finney (1960), presentación general de diseños y análisis sin teoría completa,Kempthorne (1952) y John (1971) exposición teórica más detallada.

La lista de referencias que sigue no pretende ser una bibliografía comprensiva. He incluido solo los li- / bros mencionados anteriormente y una breve selección de publicaciones que son importantes para los temas de las conferencias pero que no son necesariamente fuentes de / información originales ni primarias.

ANDERSON, S., AUQUIER, A., HAUCK, W.W., OAKES, D., / VANDEALE, V., and WEISBERG, H.I. (1980) *Statistical Methods for Comparative Studies*, New York: John / Wiley & Sons.

ARMITAGE, P. (1981) Thrombosis and oral contraception./ *British Journal of Hospital Medicine*, 25, 485.

ARMITAGE, P. and SCHNEIDERMAN, M.A. (1958) Statistical problems in a mass screening program. *Annals of the New York Academy of Sciences*, 76, 896-908.

BECHHOFER, R.E. (1954) A single-sample multiple decision procedurefor ranking means of normal populations with known variances. *Annals of Mathematical Statistics*, 25, 16-39.

BECHHOFER, R.E. (1958) A sequential multiple-decision / procedure for selecting the best one of several normal populations with a common unknown variance, and its use with various experimental designs. / *Biometrics*, 14, 408-429.

BECHHOFER, R.E. and BLUMENTHAL, S. (1962) A sequential multiple-decision procedure for selecting the best / one of several normal populations with a common unknown variance. II: Monte Carlo sampling results and new computing formulae. *Biometrics*, 18, 52-67.

BECHHOFER, R.E., DUNNETT, C.W., and SOBEL, M. (1954) A two-sample multiple decision procedure for ranking / means of normal populations with a common unknown / variance. *Biometrika*, 41, 170-176.

BECHHOFER, R.E., KIEFER, J. and SOBEL, M. (1968) *Sequential Identificacion and Ranking Procedures:* Chicago: The University Press.

COCHRAN, W.G. (1951) Improvement by means of selection. *Proceedings of the Second Berkeley Symposium*, 449-470.

COCHRAN, W.G. (1965) The planning of observational studies of human populations. *Journal of the Royal Statistical Society*, A128, 234-265.

COCHRAN, W.G. (1968) The effectiveness of adjustments / by subclassification in removing bias in observational studies. *Biometrics*, 24, 295-313.

COCHRAN, W.G. and Cox, G.M. (1957) *Experimental Designs* (2nd edition), New York: John Wiley & Sons.

CORNISH, E.A. and FISHER, R.A. (1937) Moments and cumulants in the specification of distributions. *Revue of the International Statistical Institute*, 5, 307-320.

COX, D.R. (1959) *Planning Experiments*, New York: John / Wiley & Sons.

CURNOW, R.N. (1960) The consequences of errors of measurement for selection from certain non-normal distributions. *Bulletin of the International Statistical / Institute*, 37, 291-308.

CURNOW, R.N. (1961) optima programmes for varietal selection. *Journal of the Royal Statistical Society*, B23, 282-318.

DAVIES, O.L. (1958) The design of screening tests in / the pharmaceutical industry. *Bulletin of the International Statistical Institute*, 36, 226-241.

DUNNETT, C.W. (1960) On selecting the largest of k / normal population means. *Journal of the Royal Statistical Society*, *B22*, 1-40.

DUNNETT, C.W. (1961) Statistical theory of drug screening. In *Quantitative Methods in Pharmacology* (ed. / De Jonge). Amsterdam: North Holland Publishing Company, 212-231.

FEDERER, W.T. & SCHLOTTFELDT, C.S. (1954) The use of covariance to control gradients in experiments. / *Biometrics*, *10*, 282-290.

FINNEY, D.J. (1956) The consequences of selection for a variate subject to errors of measurement. *Revue of the International Statistical Institute*, *24*, 22-29.

FINNEY, D.J. (1958) Statistical problems of plant selection. *Bulletin of the International Statistical / Institute*, *36*, 242-268.

FINNEY, D.J. (1960a) A simple example of the external economy of varietal selection. *Bulletin of the International Statistical Institute*, 37, 91-106.

FINNEY, D.J. (1960b) *An Introduction to the Theory of Experimental Design*, Chicago: University Press.

FINNEY, D.J. (1961) The transformation of a distribution under selection. *Sankhya*, *A23*, 309-324.

FINNEY, D.J. (1962a) Some properties of a distribution specified by its cumulants. *Technometrics*, 5, / 63-69.

FINNEY, D.J. (1962b) The statistical evaluation of educational allocation and selection. *Journal of the / Royal Statistical Society*, *A125*, 525-564.

FINNEY, D.J. (1962c) An unusual salvage operation. / *Biometrics*, *18*, 247-250.

FINNEY, D.J. (1964) Screening processes: problems and illustrations. *Contributions to Statistics Presented to Professor* P.C. *Mahalanobis*, 73-100.

FINNEY D.J. (1966a) An experimental study of certain / screening processes. *Journal of the Royal Statistical Society*, *B28*, 88-109.

FINNEY D.J. (1978a) *Statistical Method in Biological / Assay*, (3rd edition), London: Charles Griffin & Co.

FINNEY, D.J. (1978b) Testing the effect of an intervention in sequential ecological data. *Biometric*, *34*, / 706-708.

FINNEY, D.J. (1980) The estimation of parameters by / least squares from unbalanced experiments. *Journal / of Agricultural Science*, *95*, 181-189.

FINNEY, D.J. (1982a) The questioning statistician. / *Statistics in Medicine*, *1*, 5-13.

FINNEY, D.J. (1982b) Some enumerations for the 6x6 Latin squares. *Utilitas Mathematica*, *21A*, 137-153.

FINNEY, D.J. (1982c) Intervention and correlated sequences of observations. *Biometric*, *38*, 255-267.

FINNEY, D.J. (1984) Improvement by planned multistage/ selection. *Journal of the American Statistical Association* (in press).

FISHER, R.A. and CORNISH, E.A. (1960) The percentile / points of distributions having known cumulants. *Technometrics*, *2*, 209-225.

FISHER, R.A. and YATES, F. (1964) *Statistical Tables / for Biological, Agricultural and Medical Research*, / (6th edition), London: Oliver & Boyd.

FRANKLIN, M.F.F. and MANN, A. D. (1980) DSIGNX. *Inter-University/Research Councils Report*, no.48.

JOHN, P.W.M. (1971) *Statistical Design and Analysis of Experiments*, New York: The Macmillan Company.

KEMPTHORNE, O. (1952) *The Design and Analysis of Experiments*, New York: John Wiley & Sons.

OUTHWAITE, A.D. and RUTHERFORD, A.A. (1955) Covariance analysis as an alternative to stratificacion in the control of gradients. *Biometrics*, *11*, 431-440.

PATTERSON, H.D. and HUNTER, E.A. (1983) The efficiency of incomplete block designs in National List and Recommended List cereal variety trials. *Journal of / Agricultural Science*, *101*, 427-433.

PATTERSON, H.D. and SILVEY, S.D. (1980) Statutory and recommended list trials of crop varieties in the // United Kingdom. *Journal of the Royal Statistical Society*, *A143*, 219-252.

PATTERSON, H.D., WILLIAMS, E.R., and HUNTER, E.A. (1978) Block designs for variety trials. *Journal of Agricultural Science*, *90*, 395-400.

ROBERTSON, A. (1957) Optimum group size in progeny / testing and family selection. *Biometrics*, *13*, 442-450.

ROBERTSON, A. (1960) On optimum family size in selection programmes. *Biometrics*, *16*, 296-298.

YATES, F. (1933) The principles of orthogonality and confounding in replicated experiments. *Journal of Agricultural Science*, *23*, 180-145.