

# ON $4^3$ FACTORIAL DESIGNS WITH CONFOUNDING <sup>1)</sup>

N. H. KUIPER

Division of Mathematics, Landbouwhogeschool, Wageningen

## SUMMARY

A design is given for three factors each at four levels in three blocks of  $8 \times 8$  plots. In each block in the field, fertility in the direction of columns and of rows is not neglected. The main effects are not confounded. The interactions are partially confounded. The part is  $\leq 1/3$  for any component of the interaction. Formulae for the estimation of main effects and (all) interactions and for the analysis of variance are given in vector notation.

Recently HAMMING, VENEKAMP and VERVELDE (1952) obtained as a special case some designs for testing and/or estimating the influence of three factors (N, P, K) (and their interactions), each at four levels ( $n_i, p_i, k_i, i = 1, 2, 3, 4$ ), on the yields in a field divided in four blocks of  $8 \times 8$  plots, on which variability of fertility in some sense was not neglected.

Here we present an alternative construction of a  $4^3$ -design which obeys some of their conditions.

As an introduction we first consider one set of 64 elements (field-plots). Our conditions are the following:

1. Any combination of levels occurs once. This implies orthogonality of pure main effects and pure interactions. This property is sufficiently visualized in the following  $8 \times 8$  scheme (which is not a map of the actual field).

	$n_1$	$n_2$	$n_3$	$n_4$	$n_1$	$n_2$	$n_3$	$n_4$
P1								
P2			$k_1$				$k_2$	
P3								
P4								
P1								
P2			$k_3$				$k_4$	
P3								
P4								

FIG. 1.

2. Two classifications in classes of 8 elements are to be presented. The first classification is a division in sets which will be the "columns on the map of the field". The second classification is a division in sets, which will be the "rows on the map of the field". These two classifications are mutually orthogonal and they are orthogonal to the main effects. They are assumed to cover essential variability in fertility in the field.

Pairs of mutually orthogonal classifications, orthogonal to rows and columns

<sup>1)</sup> Received for publication January 7, 1953.

of scheme fig. 1 are called pairs of mutually orthogonal  $8 \times 8$  Latin squares (Latijnse vierkanten). In the Tables of FISHER-YATES we find seven mutually orthogonal Latin squares  $8 \times 8$ .

The first of these is

				I				
1	2	3	4	5	6	7	8	
2	1	4	3	6	5	8	7	
3	4	1	2	7	8	5	6	
4	3	2	1	8	7	6	5	
5	6	7	8	1	2	3	4	
6	5	8	7	2	1	4	3	
7	8	5	6	3	4	1	2	
8	7	6	5	4	3	2	1	

FIG. 2.

As we observe, (the space of) this classification contains (the space of) comparisons  $k_1 + k_4 - k_2 - k_3$  (see fig. 1) completely. It follows that the other six Latin squares in the tables are orthogonal to *all* pure main effects in fig. 1.

If, for example, we choose the squares no. II en III:

a	b	c	d	e	f	g	h		p	q	r	s	t	u	v	w
e	f	g	h	a	b	c	d		v	w	t	u	r	s	p	q
b	a	d	c	f	e	h	g		t	u	v	w	p	q	r	s
f	e	h	g	b	a	d	c		r	s	p	q	v	w	t	u
g	h	e	f	c	d	a	b		w	v	u	t	s	r	q	p
c	d	a	b	g	h	e	f		q	p	s	r	u	t	w	v
h	g	f	e	d	c	b	a		s	r	q	p	w	v	u	t
d	c	b	a	h	g	f	e		u	t	w	v	q	p	s	r

FIG. 3.

and we arrange the plots of the original scheme (notation:  $ijm$  instead of  $n_i p_j k_m$ ):

111	211	311	411	112	212	312	412
121	221	321	421	122	222	322	422
131	231	331	431	132	232	332	432
141	241	341	441	142	242	342	442
113	213	313	413	114	214	314	414
123	223	323	423	124	224	324	424
133	233	333	433	134	234	334	434
143	243	343	443	144	244	344	444

FIG. 4.

according to columns and rows on the map of the field, then we obtain the map (eventually before randomization of rows and columns) fig. 5:

	a	b	c	d	e	f	g	h
p	111	414	322	223	433	132	244	341
q	314	211	123	422	232	333	441	144
r	122	423	311	214	444	141	233	332
s	323	222	114	411	241	344	432	133
t	434	131	243	342	112	413	321	224
u	231	334	442	143	313	212	124	421
v	443	142	234	331	121	424	312	213
w	242	343	431	134	324	221	113	412

FIG. 5.

If no interactions are assumed, then the main effects can be tested and estimated with this plan. If interactions exist then part of the space of interactions is also used for rows and columns-effects; interactions are partly confounded with rows and columns in the map.

Next we consider a design with replications from which also the interactions can be tested and estimated. The map of the field will consist of three  $8 \times 8$  plot-plans. The first of these is the one we obtained above. The second is analogous but obtained with the Latin  $8 \times 8$  squares IV and V in FISHER's tables. The third is analogous but obtained with the Latin squares VI and VII.

The experiment will give a set of  $3 \times 64$  numbers, which form a set X of three  $8 \times 8$  schemes  $x, y, z$ , each of which will be of the form of fig. 4. The vector  $x$  can be decomposed as follows:

$$x = \bar{x} + x_{N*} + x_{P*} + x_{K*} + x_A + x_B + x_C + x_R.$$

$\bar{x}$  is the scheme or *vector*, which we get by replacing each number in the scheme  $x$  by the average of all 64 numbers.

$x_{N*} = x_N - \bar{x}$ .  $x_N$  is obtained from  $x$  by replacing each number by the average in its class under classification N. It is the maximum likelihood estimate for the expected N-main-effect (from the first  $8 \times 8$  field).

$x_A$  is obtained as follows: Let  $x_{Ai}$  be obtained from  $x$  by replacing each number by the mean in its class under classification  $i$ ;  $i = II, III$  indicates the numbers of the  $8 \times 8$  - Latin squares in FISHER's tables.

$$x_A = x_{AII} + x_{AIII} - 2\bar{x}.$$

$x_B$  and  $x_C$  are obtained in an analogous way.  $x_R$  is the residual-vector.

In this way  $x$  is decomposed in mutually orthogonal vectors and the same can be done with  $y$  and  $z$ .

An unbiased estimate for the set consisting of mean level, main effects and interactions is as follows:

$u$  (a vector!) =

$$\frac{1}{3} (\bar{x} + \bar{y} + \bar{z} + x_{N*} + y_{N*} + z_{N*} + \dots + z_{K*} + x_R + y_R + z_R) \\ + \frac{1}{2} (y_A + z_A + x_B + z_B + x_C + y_C).$$

It is not the maximum likelihood estimate because we have neglected any information that was confounded<sup>2)</sup> with fertility (differences in columns or in rows in the field) as this leads to difficult computations (non-orthogonal projections).

The vector  $u$  is easily decomposed in level, main-effects and interactions:

Level: 
$$\bar{u} = \frac{1}{3} (\bar{x} + \bar{y} + \bar{z}).$$

Main effects: 
$$u_{N*} = \frac{1}{3} (x_{N*} + y_{N*} + z_{N*}) \text{ etc.}$$

If we write

$$w = \frac{1}{3} (x_R + y_R + z_R) + \frac{1}{2} (y_A + z_A + x_B + z_B + x_C + y_C)$$

then unbiased estimates for the pure interactions are given by

$$w_{N \times P}, w_{N \times K}, w_{P \times K}, w_{N \times P \times K} - w_{N \times P} - w_{N \times K} - w_{P \times K}.$$

2) With respect to the system of all  $3 \times 8 \times 8$  plots it is more appropriate to use the word non-orthogonality than the word confounding.

$w_{N \times P}$  is obtained from  $w$  by replacing each number by the average in its class in the classification under which a class is a set of elements which have got a particular N level and also a particular P level, etc.;  $w_{N \times P \times K} = w$ .

Hence estimates of all interesting effects can be computed.

*Analysis of variance over all  $3 \times 8 \times 8$  plots*

*Dimensions (degrees of freedom)*

Mean .....	1
Main effects $3 \times 3$ .....	9
Blocks .....	2
Latin squares $3 \times (7 + 7) =$ .....	42
Interaction first order $3 \times 9$ .....	27
Interaction second order $64 - 1 - 9 - 27 =$ .....	27
Residual variance .....	84
	—
$3 \times 8 \times 8 =$	192

In accordance with our neglect of some partially confounded information on interactions, we will not use all residual degrees of freedom for the estimation of the variance of error. An estimate for  $\sigma^2$  from 60 degrees of freedom is the following:

$$S = S(\sigma^2) = \left\{ \left\{ x_N^2 * + \dots + z_K^2 * - 3 u_N^2 * - 3 u_K^2 * \right\} + \left[ \frac{1}{2} \left\{ (y_A - z_A)^2 + (x_B - z_B)^2 + (x_C - y_C)^2 \right\} \right] : \left[ 2(3 + 3 + 3) + (14 + 14 + 14) \right] \right\}$$

(The product of two vectors is the scalar product).

Of course, with so many degrees of freedom it is allowed to use (in tests)  $S(\sigma^2)$  instead of  $\sigma^2$ .

The hypothesis of non-existence of particular expected effects or interactions can now be tested with the  $\chi^2$ -table: If for example (nullhypothesis) no N-main effect is expected, then

$$3(u_{N*})^2 / S(\sigma^2)$$

can be considered as a sample from a  $\chi^2$ -distribution with 3 degrees of freedom. So this nullhypothesis can be tested in the customary way.

If (nullhypothesis) no  $N \times P$  - interaction is expected, then

$$t. w_{N \times P}^2 / S(\sigma^2)$$

is approximately a sample from a  $\chi^2$ -distribution with 9 degrees of freedom,  $t$  is a constant between 2 and 3 (part of the interaction about 2 is once confounded). So we are on the safe side in our conclusions if we take  $t = 2$ .

LITERATURE

VENEKAMP, J. T. N., G. HAMMING and G. J. VERVELDE: A  $4^3$  factorial design with confounding in  $8 \times 8$  quasi-Latin squares. *Landbouwkundig Tijdschrift* 64 (1952) 325.  
 FISHER, R. A. and F. YATES: Statistical Tables for biological, agricultural and medical research. 3rd ed. 63. Oliver & Boyd, London (1949).  
 KUIPER, N. H.: A forthcoming paper on analysis of variance. *Statistica*, 1953.  
 KUIPER, N. H.: *Mededelingen Medisch-Biologische Sectie, Vereniging voor Statistiek*, February, 1953.