

Kansas State University Libraries

New Prairie Press

Conference on Applied Statistics in Agriculture

1994 - 6th Annual Conference Proceedings

DESIGN FOR THE EXPERIMENTER

O. B. Allen

D. A.J. Ryan

Follow this and additional works at: <https://newprairiepress.org/agstatconference>



Part of the [Agriculture Commons](#), and the [Applied Statistics Commons](#)



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License](#).

Recommended Citation

Allen, O. B. and Ryan, D. A.J. (1994). "DESIGN FOR THE EXPERIMENTER," *Conference on Applied Statistics in Agriculture*. <https://doi.org/10.4148/2475-7772.1357>

This is brought to you for free and open access by the Conferences at New Prairie Press. It has been accepted for inclusion in Conference on Applied Statistics in Agriculture by an authorized administrator of New Prairie Press. For more information, please contact cads@k-state.edu.

DESIGN FOR THE EXPERIMENTER

**O.B. Allen and D.A.J. Ryan
University of Guelph
Guelph, Canada N1G 2W1**

Abstract

Three research problems, for which there was no obvious textbook design available, are presented. Each of the three involve constructing incomplete block designs for factorial treatment arrangements. While the designs are not likely optimal by any of the classical criteria, they meet the objectives of the research projects. The constructions involved confounding, with incomplete blocks, those effects which were of least interest. However, effects of interest were also allowed to be slightly non-orthogonal to blocks in order to be able to examine a larger number of effects of interest.

Key Words: incomplete blocks, factorial designs, confounding, row-column designs, non-orthogonal designs.

Introduction

One is occasionally faced with choosing a design for a research problem where no suitable textbook design is available. A generation ago, one might have forced the problem to fit a standard design since the analysis of a more appropriate non-standard design would have been daunting. This is no longer a valid excuse (if it ever was). Statisticians should not feel constrained to use published designs when they are not appropriate. By using the concepts of orthogonality and confounding, one can construct designs which suit the research problem, although they may not be optimal by any of the classical criteria. Three examples are given below of research problems which came to our attention through statistical consulting, for which there was no obvious textbook design. All involve the blocking of factorial treatment arrangements. By confounding factors of least interest with blocks and maintaining factors of most interest orthogonal to blocks (or nearly so), we were able to construct designs which effectively addressed the issues of interest to the research scientist. Methods for constructing designs, often assume that the investigator is equally interested in all effects to be estimated. Thus, a subset of the effects of interest are required to be orthogonal to blocks and some of the remaining effects are completely confounded with blocks.

However, investigators will often be more interested in some effects than others, particularly at the exploratory stage of an investigation. For example, an investigator may be primarily interested in main effects but would also like to be alerted to any two factor interactions which are important. By relaxing the requirement that effects be either completely confounded or orthogonal to blocks, we are able to estimate effects (such as two-factor interactions) which are of some interest but which would otherwise be completely confounded with blocks.

There are a variety of computer algorithms available which will construct experimental designs. Within SAS/QC (SAS Institute Inc., 1989), the FACTEX procedure generates orthogonally confounded factorial designs, including fractional factorials. It will not generate designs with partial confounding. PROC OPTEX searches for A or D-optimal designs for a specified model from among a dataset of potential design points. The procedure does not allow one to specify different precisions for the various effects in the model. It also implicitly assumes that there is only one model of interest. It is unlikely, at the exploratory stage, that this is true.

Design 1

Introduction. The corn rootworm is a major pest of corn in Ontario and the US midwest. The adult oviposits in the soil of maturing corn fields between August and October each year. The larvae hatch in June of the following year and attack the root system of the growing corn plants. Effective insecticides can be applied at planting; however, this is expensive and environmentally unfriendly. The application is unnecessary if the economic damage done by the larvae is less than the cost of applying the insecticide. What is needed is an accurate estimate of the larval density in the soil. However, soil sampling is very expensive. This problem could be overcome if adult females could be enticed to oviposit in a monitoring device. Mulock's (1992) objective was to design an effective oviposition trap (essentially a can set into the soil). Six factors were considered:

- (A) trap volume (380, 870, 1740 ml)
- (B) moisture level, determined by the vertical distance from the top of the can to a drainage hole (4, 8, 12cm)
- (C) surface covering (no cover, foil, corn plant material)
- (D) trap opening (38, 78 cm²)
- (E) particle size of the substrate (Turface) in the can (medium, coarse)
- (F) trap location (within row, between rows)

Choice of a design. This is a $2^3 \times 3^3$ factorial. The experimenter was prepared to run one replicate of these 216 treatments. The traps were set out for 1 week and then examined and the number of eggs counted. However, she was not able to manage more than 24 containers at a time. Since the intensity of gravid beetles may vary over the season, it was desirable to block on weeks. Thus, we would like a design which provides 216 treatments in 9 incomplete blocks of size 24, in such a way that only higher order interactions are confounded with blocks.

Let A, B, and C denote the three level factors and D, E, and F the two level factors. The levels will be denoted 0, 1 or 2 (for the three level factors) and 0 or 1 (for the two level factors). Two factor interactions among factors A, B and C can be partitioned into two orthogonal components with 2 df each. For example, the AB interaction can be partitioned into AB and AB^2 , each with 2 df. If an ABC factor combination is denoted (i, j, k), then AB represents comparisons among the three groups for which $i+j \pmod{3}$ equals 0, 1 or 2. Similarly, AB^2 represents comparisons among the three groups for which $i+2j \pmod{3}$ equals 0, 1 or 2. The ABC interaction can be orthogonally partitioned into four components, each with 2 df, denoted ABC, ABC^2 , AB^2C and AB^2C^2 where, for example, AB^2C represents the comparisons among those groups for which $i+2j+k \pmod{3}$ equals 0, 1 or 2.

Standard references do not provide a design which meets the above specifications. However, using the principles of orthogonality and confounding, we can construct a design which meets the needs of the investigator. In order to construct such a design, focus first on the 3 level factors. Kempthorne (1952) provides a design for a 3^3 factorial in blocks (groups) of size 3 with AB, AC, BC^2 , AB^2C^2 confounded:

GROUPS								
I	II	III	IV	V	VI	VII	VIII	IX
000	012	021	002	011	020	001	022	010
122	101	110	121	100	112	120	111	102
211	220	202	210	222	201	212	200	221

Now, form a Youden square with the 2^3 factor combinations forming the rows, the 9 blocks forming the columns and the above groups forming the letters.

	BLOCKS								
	1	2	3	4	5	6	7	8	9
000	I	II	III	IV	V	VI	VII	VIII	IX
001	II	III	IV	V	VI	VII	VIII	IX	I
010	III	IV	V	VI	VII	VIII	IX	I	II
011	IV	V	VI	VII	VIII	IX	I	II	III
100	V	VI	VII	VIII	IX	I	II	III	IV
101	VI	VII	VIII	IX	I	II	III	IV	V
110	VII	VIII	IX	I	II	III	IV	V	VI
111	VIII	IX	I	II	III	IV	V	VI	VII

Each of the cells of this Youden square represents three treatments. For example, the upper left-hand cell represents 000000, 122000, and 211000. Thus, each of the blocks (columns) has 24 treatments. The row effects represent the main effects and interactions among the factors D, E, and F. The column effects represent the incomplete blocks and the Roman numeral effects represent the AB, AC, AB^2 and AB^2C^2 interactions. We know, by construction, that the Roman numerals are orthogonal to the rows and nearly orthogonal to the columns. Also, the rows are orthogonal to the columns. Thus, main effects and interactions among the factors D, E, and F are orthogonal to blocks and the AB, AC, BC^2 and AB^2C^2 interactions are nearly orthogonal to blocks. The relative information (Cochran and Cox, 1956, page 186) on these effects is 98%. Furthermore, the remaining main effects and interactions among the factors A, B and C are estimated within the cells of the Youden square (This follows from the first construction) and hence are orthogonal to the blocks. Thus, by elimination, blocks are primarily confounded with interactions of AB, AC, BC^2 , and AB^2C^2 with D, E, F, DE, DF, EF and DEF. Thus, the smallest interactions which are substantially confounded with blocks are three factor interactions, such as ABD.

Thus, this design allows the estimation of all main effects and two factor interactions. It also allows the estimation of selected higher order interactions (ABC, DEF and all interactions of A, B and C with the main effects and interactions of D, E and F). If the investigator is comfortable in assuming that the remaining three factor and higher order interactions are negligible, there remains 132 df to estimate experimental error. A two stage randomization is recommended; randomize the blocks and randomize the treatments within each block. A linear models program should be used to analyse the data from this experiment, since AB, AC, BC^2 and AB^2C^2 are not orthogonal to blocks.

Design 2

Introduction. A potential process for producing margarine involves the synthesis of esters and ester-ester interchange reactions (interesterification) on blends of fats. Lipase, of microbial origin, is a favoured catalyst. A moisture and temperature controlled bench-top continuous flow-through reactor, consisting of two water jacketed glass columns, was set up to examine the factors which affect this reaction. The pre-column, which was packed with biobone, regulated the moisture content of the fats and the enzyme column was packed with Lipozyme to catalyse the reaction. From start-up, the yield of the column increases at a decreasing rate, reaching an equilibrium after about 3 hours. The investigator wished to examine how the yield and rate of these reactions at equilibrium, are affected by 1) moisture content in the pre-column (7.4, 13.3, 19.2%) , 2) the flow rate through the column (11.4, 20.5, 28.3 g/h) and 3) the fat blend used (X, Y, Z). The runs were done sequentially; three runs could be done per day. It is likely that each of the columns will "age"; that is, the water content of the pre-column and the activity of the Lipozyme will likely decrease with time, affecting the rate and yield of the reaction. It was not feasible to change the water column or the enzyme column after each run. It was decided to change the enzyme column after every 9 runs and the water column after every 3 runs (ie each day). However, the age of the water column and age of the enzyme column can be treated as blocking factors. The task is to choose a design which allows the investigator to examine the effects of flow rate, moisture content and fat blend and their two factor interactions, while blocking for the age of the enzyme column, age of the water column and the repacking of the enzyme column. The investigator was prepared to do 50 to 70 runs.

Constructing a design. There are three factors of interest, each at three levels. We use the same notation as set out in design 1. We might start by partitioning the 27 treatments into three incomplete blocks by confounding the ABC component of the three factor interaction (Rep IV, plan 6.7, Cochran and Cox, 1957). Each block represents a repacking of the enzyme column. Secondly, confound the ABC^2 component with age of enzyme column. Note that this will partition the 9 treatments in each incomplete block into three groups of three, corresponding to ages 1, 2 and 3 of the enzyme. Finally, if we confound age of water column with the AB^2C component, the three treatments in a group will be associated with age 1, 2 or 3 of the water column. This determines a unique design. We can randomize the blocks, age of enzyme and age of water column to the three incomplete blocks, in each case. We can also randomize the order of the factors and the labels for the levels in the notation for the treatments. The investigator ran two replicates. The first replicate of the design is laid

out below where X, Y and Z denote the fat blends and the subscripts 11, 20 and 28 denote the flow rate. The runs are conducted sequentially, by rows.

Replicate 1

Repack enzyme col.	Age of enzyme col.	% water	Age of water column		
			1	2	3
1	1	7	X ₁₁	Z ₂₀	Y ₂₈
	2	13	Z ₁₁	Y ₂₀	X ₂₈
	3	19	Y ₁₁	X ₂₀	Z ₂₈
2	1	19	X ₂₈	Z ₁₁	Y ₂₀
	2	7	Z ₂₈	Y ₁₁	X ₂₀
	3	13	Y ₂₈	X ₁₁	Z ₂₀
3	1	13	X ₂₀	Z ₂₈	Y ₁₁
	2	19	Z ₂₀	Y ₂₈	X ₁₁
	3	7	Y ₂₀	X ₂₈	Z ₁₁

Analysis of the design: By construction, the blocking factors age of enzyme column, age of water column and repacking of enzyme column are all confounded with components of the three factor interaction of percent water, flow rate and fat blend. Consequently, these blocking factors are orthogonal to the main effects and two factor interactions. The analysis below is based on two replicates. All terms in the following analysis of variance are orthogonal and consequently linear models software is not required. The results of using this design are reported in Cho, deMan and Allen (1994).

<u>SOURCE OF VARIANCE</u>	<u>DF</u>
Replicate (R)	1
Block within Replicate (Bl/R)	4
Age of Enzyme within Rep (E/R)	4
Age of Water within Rep (W/R)	4
Fat Blend (A)	2
Flow Rate (B)	2
Moisture Content (C)	2
A x B	4
A x C	4
B x C	4
Error	22
Total	53

This design is reminiscent of a split plot, in which water content is randomized to days, but fat blends and flow rates are randomized to individual runs. The error term is primarily made up of interactions of replicate with the three treatment

factors and their two factor interactions. The error term could be partitioned up to examine these replicate interactions. If they exhibit heterogeneity then each treatment effect should be assessed using the particular treatment effect by replicate interaction.

Design 3

Introduction. Design 3 also dealt with developing effective technology to combat corn rootworm. The technology involved injecting insecticide into the soil near the corn root at the time of hatching in the spring. There were six factors of interest to the research scientist:

- A. Form of insecticide (water soluble, insoluble, gaseous)
- B. Rate of application (low, high)
- C. Depth of application (shallow, deep)
- D. Distance to the corn row (closer, further)
- E. Frequency of injection (1 or 2 injections)
- F. Applied to one or both sides of row (1 or 2 sides)

The treatment structure is thus a 3×2^5 and the incomplete blocks were of size 24. The research scientist was prepared to run 3 replicates.

Design of the experiment. We will design this experiment in a manner similar to the first experiment. We first partition the 2^5 treatment combinations corresponding to the factors B through F into 4 groups of size 8, by confounding BCD, BEF and CDEF. The result is given below:

GROUPS			
I	II	III	IV
00000	11000	10000	01000
01100	10100	11100	00100
11010	00010	01010	10010
10110	01110	01001	11110
11001	00001	00101	10001
10101	01101	10011	11101
00011	11011	11111	01011
01111	10111	00110	00111

We now write these groups into a 3×4 Youden square, with the 4 columns being the blocks and the 3 rows representing the levels of factor A.

A	BLOCKS			
	1	2	3	4
0	III	I	II	IV
1	IV	II	I	III
2	II	IV	III	I

Then, by construction, we know that the main effect of A is orthogonal to blocks and to the main effects and interactions of the factors B, C, D, E and F. The BCD, BEF and CDEF interactions have a balanced incomplete block arrangement with blocks. Hence, there is 89% relative information on these interactions. Finally, the remaining main effects and interactions among the factors B, C, D, E and F and their interactions with A are orthogonal to blocks since they are all estimated within groups. Thus, by elimination, what is primarily confounded with blocks are components of the ABCD, ABEF and ACDEF interactions. This design thus provides full information on all main effects and two factor interactions. It provides 89% of relative information on BCD and BEF and full information on the remaining three factor interactions.

If, as was the case for this experiment, several replicates of the experiment can be run, one can confound different sets of interactions with groups in each replicate. For instance, one could confound BCE, CDF and BDEF with groups in the second replicate, and BDF, CDE and BCEF in the third replicate. Then, one would have 89% relative information on BCD from replicate 1 and full information from replicates 2 and 3. Three replicates would thus provide 96% relative information overall, for BCD, BEF, BCE, CDF, BDF and CDE and full information on the remaining three factor interactions. We have chosen the first three replicates from Cochran and Cox (1956), plan 6.5, who provide a design for a 2^5 with balanced partial confounding, in 6 replicates. The effects which are primarily confounded with blocks will be different in each replicate, as well. This means that two of the three replicates provide full information on these confounded effects.

Acknowledgements

Design problem 1 was brought to D.A.J. Ryan by Barbara Mulock, a graduate student of Dr. C.R. Ellis, Department of Environmental Biology, University of Guelph. Design problem 2 was brought to O.B. Allen by France Chu, a graduate student of J.M. deMan, Department of Food Science, University of Guelph. Design problem 3 was brought to O.B. Allen by Dr. A. Schaafsma, Ridgetown College of Agricultural Technology.

References

- Cho, F., J. M. deMan and O.B. Allen. 1994. Modification of hydrogenated canola oil/palm stearin/canola oil blends by continuous enzymatic interesterification. To appear *Elaeis* 6:
- Cochran, W.G. and G.M. Cox. 1956. *Experimental Designs*. Second Edition. N.Y.: Wiley.
- Kempthorne, O. 1952. *The Design and Analysis of Experiments*. N.Y.: Wiley.
- Mulock, B.S. 1992. Evaluation of an ovaposition trap for monitoring egg populations of corn rootworm (Coleoptera: Chrysomelidae) in field corn. Ph.D. thesis, University of Guelph. 131 pp.
- SAS Institute Inc. 1989. *SAS/QC Software: Reference, Version 6*, First Edition. Cary, NC: SAS Institute Inc.