Selected Papers on Probability and Statistics

VOLUME 1

Experiments With Mixtures

Selected Papers on
Probability &
Statistics

Vol. 1

Experiments with Mixtures

May 1979

前　言

在化学、炼钢等许多科学实验中，因素成份是表为百分数，并且在每次试验中这些百分数的总和恰好等于1，我们称具有这二特征的试验为混料试验。数理统计学中所研究的各种试验设计，如正交试验设计，对减少试验次数，得到较可靠的结论是众所周知的，但并不适用于混料试验，这是因为通常用的几种试验设计中诸因素所取的数值是没有限制的，而混料试验中诸因素成份是受有总和必须等于1的限制的，换句话说，通常用的试验设计中的因素是独立的变量，而混料试验中的因素是不独立的变量。设以Xi记第i个因素成份，则几个因素成份受有 $\sum_{i=1}^{n} Xi$ 的约束。因此，通常用的几种试验设计不适用于混料试验。

然而，混料试验在生产斗争和科学实验中却占很大的比重，对它进行研究很有必要。1955年P. J. Claringbold在实践中第一次应用单形设计于混料试验，随后，H. Scheffé (1958, 1963) 系统地研究了混料试验，并提出了单形格子和单行重心两类设计，为混料试验设计奠定了基础。到了六十年代末叶，这方面的文章开始多起来了，但大多应侧重于如何应用。自从进入了七十年代，对于模型的形式、设计和分析、优良性的讨论等有了进一步的发展。

我国在推广正交试验设计的基础上，觉得对混料试验有实际地需要。目前由于我国工业发展的需要，关心混料试验的人日益增多。为了帮助大家对混料试验在理论上和实践上作深入的研究，我们这里收集了一部分原始论文，间附一些与混料试验间接有关的一些理论性文章。由于登载这些文章的刊物在国内并不容易找到，将它们汇编成集加以出版也许是有益的。

在选编过程中悉凭管见，漏掉重点文章和不当之处事所难免，竭诚欢迎读者予以批评指正。

上海师范大学数学系

概率统计教研组

CONTENTS

目　次

USE OF THE SIMPLEX DESIGN IN THE STUDY OF JOINT ACTION OF RELATED HORMONES

P. J. CLARINGBOLD

Department of Veterinary Physiology.
University of Sydney, N.S.W., Australia

INTRODUCTION

In certain toxicological studies the term joint action has come to take a special meaning. If members of a group of related compounds all cause death of an organism when administered separately the simultaneous action of these substances is called their joint action. Bliss (1939) first discussed the analysis of data obtained in this manner. From this time the problem has been examined in terms of tolerance distribution theory and developed in relation to probit analysis (Finney, 1952). Plackett and Hewlett (1951) have extended the tolerance distribution theory to different theoretical forms of joint action and developed a set of mathematical models, each of which is based on many assumptions and is very difficult to fit to experimental data.

Fisher (1954) has shown that parameters of the binomial distribution may be estimated without tolerance distribution assumptions. The aim of the present paper is to show that the study of joint action by means of an appropriate experimental design—the simplex design—allows ready interpretation of experimental data with no reference to a joint tolerance distribution, and no further assumptions than normally required in quantal analysis. The method is also appropriate without modification to the study of joint action of substances eliciting a graded response simply by applying the standard estimation procedures.

Examples will be drawn from the study of the action of oestrogens on the vagina of the ovariectomized mouse. The quantal response in this case is cornification of the vaginal epithelium.

MATHEMATICAL METHODS

The simplex design.

Suppose A_j, $(j = 1, 2, \cdots, k)$ are the doses of k hormones which, when administered separately, elicit approximately the same percentage response. A joint dose, D, may be defined in terms of k coordinates, X_j, which take positive values, thus,

$$D = \sum A_j X_j , \qquad (1)$$

where the coordinate values are restricted by,

$$\sum X_i = 1. \tag{2}$$

The experimental region is therefore restricted by the above to a $(k - 1)$ dimensional simplex with vertices at the points on the coordinates $X_i = 1$. This method of approach allows all different types of joint doses to be uniquely specified. Thus if $X_i = 1$, the ith hormone is administered separately. If $X_i + X_j = 1$, then some mixture of the ith and jth hormones are administered together. In both the study of experimental designs in this region and the analysis of experimental data it is essential that a $(k - 1)$ dimensional coordinate system be introduced to the simplex. This may be done in two stages, (1) shift the origin of the X system to the centroid of the simplex, i.e. the point where every coordinate has the value $1/k$, (2) rotate the axes so that (say) the kth is orthogonal to the simplex.

The first is accomplished by the simple transformation (3) which at the same time changes the scale of measurement so that the vertices have non-fractional coordinates in the new system, $\overline{X}$.

$$\overline{X}_i = k(X_i - 1/k) = kX_i - 1, \tag{3}$$

where 1 is a vector of length k all elements of which are unity.

The second stage is carried out by an orthogonal transformation (4) of rank k with matrix, Θ. The scale is also modified so the vertex-centroid distance becomes $(k - 1)$ units.

$$\overline{\overline{X}}_i = k' \cdot \overline{X}_i \cdot \Theta \tag{4}$$

where $k' = k(k - 1) \cdot / k$ is a scale factor.

$$\Theta = 1/k(k-1) \cdot \begin{bmatrix} k-1 & 0 & 0 & \cdot & 0 & s \\ -1 & (k-2)l & 0 & \cdot & 0 & s \\ -1 & -l & (k-3)m & \cdot & 0 & s \\ -1 & -l & -m & \cdot & 0 & s \\ -1 & -l & -m & \cdot & 0 & s \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ -1 & -l & -m & \cdot & n & s \\ -1 & -l & -m & \cdot & -n & s \end{bmatrix}$$

The additional letters in this matrix are determined from the fact that the sum of the squares of the elements of each column is $k(k - 1)$.

After this transformation all points of the simplex take the value 0 for the coordinate $\overline{\overline{X}}_k$ which may therefore be ignored or used to describe another experimental variable, say different equivalent levels of dose. An experimental design consists of N points of the experimental region and may be summarized in a matrix called the *design matrix*, Box & Wilson (1951). The N rows of this matrix give the values of the coordinates at each of the N experimental points. In the present case the design matrix is of order N by $(k - 1)$.

An example when $k = 2$.

In this case equations 1 and 2 become

$$D = X_1A_1 + X_2A_2\,, \quad X_1 + X_2 = 1, \quad 0 \leq X_1\,, \quad X_2 \leq 1 \qquad (5)$$

The experimental region is a line. For illustrative purposes a design matrix consisting of 5 experimental points, including the vertices, the centroid and two intermediate points will be transformed using the appropriate forms of equations 3 and 4.

$$\begin{matrix} X_1 & X_2 \end{matrix} \quad \begin{matrix} \overline{X}_1 & \overline{X}_2 \end{matrix} \quad \begin{matrix} \overline{\overline{X}}_1 & \overline{\overline{X}}_2 \end{matrix}$$

$$\begin{bmatrix} 1 & 0 \\ \frac{3}{4} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{3}{4} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ \frac{1}{2} & -\frac{1}{2} \\ 0 & 0 \\ -\frac{1}{2} & \frac{1}{2} \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \frac{1}{2} & 0 \\ 0 & 0 \\ -\frac{1}{2} & 0 \\ -1 & 0 \end{bmatrix}$$

where $\overline{X}_1 = 2X_1 - 1$, $\overline{X}_2 = 2X_2 - 1$.

$$[\overline{\overline{X}}_1\,, \overline{\overline{X}}_2] = \tfrac{1}{2}[\overline{X}_1\,, \overline{X}_2] \cdot \begin{bmatrix} 1, & 1 \\ -1, & 1 \end{bmatrix}$$

Transformation to a log dose scale.

Often in biological work response is linearly related to log dose. In studies on joint action it is of interest to test if this relationship still holds with respect to log joint dose. Since equation 1 is not linear in log joint dose a series of transformations are made so that this equation holds for the logarithms of the equivalent doses and the log joint dose in terms of different coordinates. These transformations are all based on the simple case $k = 2$.

Suppose equation 5 be written,

$$d = pa_1 + (1 - p)a_2 \qquad 0 \leq p \leq 1. \qquad (6)$$

This equation may be written,

$$\text{if} \qquad \log d = q \log a_1 + (1 - q) \log a_2 , \tag{7}$$

$$p = (r^{1-q} - r)/(1 - r), \qquad \text{where} \qquad r = a_2/a_1 . \tag{8}$$

Some values of this transformation are given in Table 1. Equation 7 may be put in the form of equation 5 by simple definition of terms, i.e. $D = \log d$, $A_1 = \log a_1$ and so on.

TABLE 1

Table of the transformation,

$$p = (r - r^{1-q})/(r - 1),$$

for equidistant sets of values of q. Geometric intervals of r are tabulated since the changes in p are more linear with this scale. (Figures in table are all × 10,000).

	r										
q	$\sqrt{2}$	2	$2\sqrt{2}$	4	$4\sqrt{2}$	8	$8\sqrt{2}$	16	32	64	128
1/2	5432	5858	6271	6667	7040	7388	7708	8000	8498	8889	9188
1/3	3725	4126	4531	4934	5330	5714	6083	6434	7071	7619	8079
2/3	7044	7401	7735	8042	8321	8571	8793	8987	9298	9524	9682
1/4	2834	3182	3541	3905	4271	4633	4988	5333	5983	6567	7082
3/4	7815	8108	8377	8619	8836	9026	9191	9333	9555	9710	9865
1/5	2286	2589	2904	3229	3558	3889	4217	4540	5161	5737	6260
2/5	4420	4843	5263	5675	6074	6454	6813	7148	7742	8234	8632
3/5	6410	6805	7180	7530	7853	8147	8411	8646	9032	9321	9530
4/5	8267	8513	8736	8935	9111	9263	9395	9506	9677	9794	9871
1/6	1916	2182	2461	2751	3047	3347	3648	3947	4529	5079	5589
5/6	8564	8775	8965	9134	9281	9408	9517	9608	9748	9841	9902
1/10	1163	1339	1528	1726	1933	2146	2363	2583	3023	3457	3875
3/10	3372	3755	4145	4537	4925	5304	5672	6024	6673	7241	7728
7/10	7354	7689	7998	8281	8536	8763	8962	9135	9410	9606	9741
9/10	9149	9282	9401	9514	9594	9670	9734	9787	9866	9918	9951

In more general cases the joint dose may be looked on as a series of equations 6. For example if $k = 3$ equation 1 may be written,

$$d = p\{p'a_1 + (1 - p')a_2\} + (1 - p)a_3$$

where $0 \leqq p, p' \leqq 1$. The quantity in braces may be regarded as a quantity, say b, and two transformations of the form of 8 made. Thus

$$d = a_1^{qq'} \cdot a_2^{q(1-q')} \cdot a_3^{(1-q)} \tag{9}$$

where

$$p' = (r_1^{(1-q')} - r_1)/(1 - r_1), \qquad r_1 = a_2/a_1$$
$$p = (r_2^{(1-q)} - r_2)/(1 - r_2), \qquad r_2 = a_3/b.$$

Equation 9, on taking logarithms and using obvious definitions may be written,

$$D = X_1A_1 + X_2A_2 + X_3A_3\,, \qquad X_1 + X_2 + X_2 = 1. \qquad (10)$$

In this form response may be related to log joint dose and its components in a simple manner. When equivalent doses are equal, logarithmic transformations need not be made, since in this case log joint dose is unaffected by variations in the coordinates subject to the restriction 2.

Extension of the experimental region.

Different equivalent levels of dose may be chosen for study. The method by which this is carried out depends on the form of the relationship of response to dose. In the case where this relationship is loglinear it is convenient to define each equivalent dose (A_i) as a function of an exponent (n) in terms of constants.

$$A_i = a_i r_i^n \qquad (11)$$

The values of the constants chosen depend on the Median effective dose (M.E.D.) and slope of the jth dose response line. Substituting the logarithm of these equations in equation 10 yields a function linear in n if the X's are held constant and linear in the X's if n is held constant. It is also useful to choose the levels of the constants so that the values of A_i chosen for study correspond to a set of equally spaced symmetric values of n centered at zero.

Other experimental variables may be introduced into the design in a factorial or other manner. In practice, however, if many points in the simplex are chosen for study this will lead to very large numbers of treatment combinations.

Analysis of variance.

Suppose a mixed level factorial experiment consists of the combinations of three factors denoted S, L and A at s, l and a levels respectively. The factor S is somewhat unusual and consists of s points of the simplex design, the factor L of l different levels of equivalent dose and A an additional factor at a levels. The complete design matrix therefore has $N = s \cdot l \cdot a$ rows and $(k - 1) + 2$ columns where k is the number of substances entering the simplex design. For each factor an orthogonal

set of comparisons including the identity may be drawn up and tabulated as an orthogonal matrix, or as the product of an orthogonal matrix and a diagonal matrix to preserve round numbers. Suppose these matrices be arranged so the columns present comparisons, the first column consisting only of unity i.e. the identity. Successive columns or comparisons may be numbered S^0, S^1, S^2, $\cdots$, where the superscript 0 denotes the identity and the other superscript has a currency of at most the number of degrees of freedom of the factor levels under consideration. The full set of orthogonal comparisons appropriate to the N treatment combinations may be obtained by the *direct product* (see Tocher, 1952 for a definition) of these matrices followed by an appropriate permutation of the columns. Since in general only main effects and first order interactions are required, other degrees of freedom going into an estimate or error or being isolated (see Fisher, 1951) only part of this product need be carried out. Main effect degree of freedom comparisons are obtained by the direct product of the column under consideration with all other identity columns. First order interaction comparisons are obtained by the direct product of the two individual main effect comparisons under consideration with the remaining identities. The matrix resulting from these direct products will consist of the first columns of an orthogonal matrix or the product of an orthogonal matrix with a diagonal matrix since the direct product of orthogonal matrices is orthogonal. The sum of squares attributable to the individual comparisons may be determined in the standard manner. For a binomial variable the appropriate procedures have been described by Claringbold, Biggers and Emmens (1953).

When $k = 2$, several sets of orthogonal comparisons have been determined for the purpose of detecting departures from linearity of response on dose. These are given for three cases, namely where one, two and three points are equally spaced on the line joining the two vertices.

Name													
$S^{0'}$	1	1	1	1	1	1	1	1	1	1	1	1	1
$S^{1'}$	−1	0	1	−1	0	0	1	−1	0	0	0	1	
$S^{2'}$	1	−2	1	−1	1	1	−1	1	0	−2	0	1	
$S^{3'}$				0	−1	1	0	2	−3	2	−3	2	
$S^{4'}$								0	−1	0	1	0	

The first row (since the matrices have been transposed for convenience) is the identity. The second tests whether equivalent doses

were given. The third tests whether the mid-point response(s) falls on the line joining the control responses. The last comparison in each case is determined by those already made. An example of the use of these coefficients is given by Claringbold and Biggers (1955).

Sets of comparisons may be determined for other cases where k is greater than two and when symmetric arrays of points in the simplex have been chosen.

Regression analysis.

A response transformate may be directly related to functions of the coordinates of the design matrix by a weighted regression analysis. The information matrix in this case is not diagonal since the sums of squares and cross-products of the coordinates of the design matrix are not in general independent.

EXAMPLE

The data are summarised in Table 2 together with the coordinates of the experimental design. The plan of the simplex design used in this experiment is shown in Fig. 1. The complete design is in the form

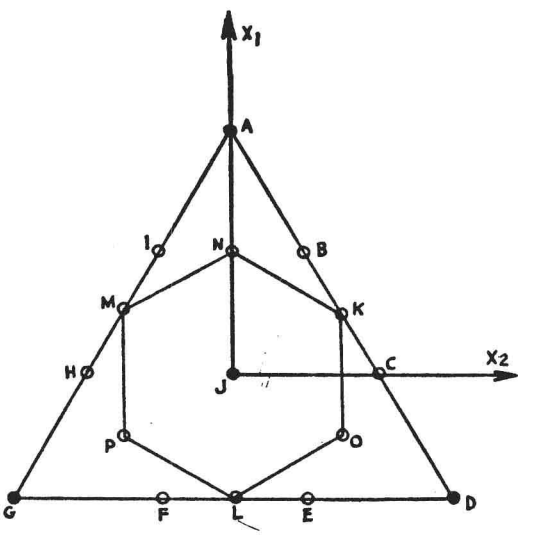

FIG. 1.

Plan of the two-dimensional simplex design used in the example. Points *A*, *D* and *G* correspond to the administration of oestrone, oestradiol-3:17β and oestriol alone, respectively. Points on the lines joining these vertices correspond to the administration of two oestrogen mixtures, while points within the triangle correspond to mixtures of three oestrogens.

of two equilateral triangular prisms, one for each replicate. Each prism has experimental points on three equidistant triangular planes. Since equal doses of oestrone, oestradiol-3:17β and oestriol are approximately equivalent in their effect on response when administered intravaginally no logarithmic transformations are used. The empirical angular response (Y) is related to ten functions of the coordinates of

TABLE 2

Percentage response of groups of 12 ovariectomized mice to joint intravaginal administration of oestrone, oestradiol and oestriol. The equivalent doses of these oestrogens denoted A_1, A_2, and A_3 were chosen so that

$A_1 = A_2 = A_3 = 0.75 \times 10^{-4}\mu g.$ when $X_L = -1$
$= 1.50 \times 10^{-4}\mu g.$ when $X_L = 0$
$= 3.00 \times 10^{-4}\mu g.$ when $X_L = 1.$

Original coordinates			Point	Coordinates in simplex		Response		
X_1	X_2	X_3		X_1	X_2	$X_L=-1$	$X_L=0$	$X_L=1$
First replicate—$X_R = -1$								
1	0	0	A	2	0	17	42	83
2/3	1/3	0	B	1	$t/3$	0	33	75
1/3	2/3	0	C	0	$2t/3$	33	33	75
0	1	0	D	-1	t	58	58	100
0	2/3	1/3	E	-1	$t/3$	17	33	67
0	1/3	2/3	F	-1	$-t/3$	33	33	58
0	0	1	G	-1	$-t$	25	50	42
1/3	0	2/3	H	0	$-2t/3$	25	42	42
2/3	0	1/3	I	1	$-t/3$	0	25	75
1/3	1/3	1/3	J	0	0	17	25	58
Second replicate—$X_R = 1$								
1	0	0	A	2	0	42	50	75
1/2	1/2	0	K	1/2	$t/2$	17	33	83
0	1	0	D	-1	t	75	67	83
0	1/2	1/2	L	-1	0	33	42	67
0	0	1	G	-1	$-t$	50	42	100
1/2	0	1/2	M	1/2	$-t/2$	17	42	58
2/3	1/6	1/6	N	1	0	33	33	58
1/6	2/3	1/6	O	$-1/2$	$t/2$	50	50	58
1/6	1/6	2/3	P	$-1/2$	$-t/2$	33	33	50
1/3	1/3	1/3	J	0	0	17	42	42

where $t = \sqrt{3}$

the design, by the following regression equation,

$$Y = \beta_0 + \beta_R X_R + \beta_1 \overline{\overline{X}}_1 + \beta_2 \overline{\overline{X}}_2 + \beta_L X_L + \beta_{12} \overline{\overline{X}}_1 \overline{\overline{X}}_2 + \beta_{1L} \overline{\overline{X}}_1 X_L + \beta_{2L} \overline{\overline{X}}_2 X_L + \beta_{11} \overline{\overline{X}}_1^2 + \beta_{22} \overline{\overline{X}}_2^2 .$$

The information matrix was determined for these ten parameters and was inverted to give the variance-covariance matrix (Table 3). The

TABLE 3

Variance-covariance matrix for the experimental data and design given in Table 2. The theoretical variance used in its formation is that tabulated by Claringbold, Biggers and Emmens (1953) for the empirical angular transformation.

3.281	−0.106	.	.	.	.	.	.	−1.066	−1.066
−0.106	1.261	.	.	.	.	.	.	0.056	0.056
.	.	3.182	.	.	.	.	.	−1.473	1.473
.	.	.	2.860	.	2.436	.	.	.	.
.	.	.	.	1.883	.	.	.	.	.
.	.	.	2.436	.	3.857	.	.	.	.
.	.	.	.	.	.	1.981	.	.	.
.	.	.	.	.	.	.	1.981	.	.
−1.066	0.056	−1.473	.	.	.	.	.	1.727	−0.605
−1.066	0.056	1.473	.	.	.	.	.	−0.605	1.727
X_0	X_R	$\overline{\overline{X}}_1$	$\overline{\overline{X}}_2$	X_L	$\overline{\overline{X}}_1\overline{\overline{X}}_2$	$\overline{\overline{X}}_1 X_L$	$\overline{\overline{X}}_2 X_L$	$\overline{\overline{X}}_1^2$	$\overline{\overline{X}}_2^2$

matrix inversion was carried out using the method of Fox (1950) and Fox and Hayes (1952). In Table 4 the estimates of regression co-

TABLE 4

Regression analysis of the data of Table 2 following the empirical angular transformation.

Regression coefficient	Least square estimate	$t(\infty)$	P
β_0	35.39		
β_R	2.61 ± 1.12	2.3	$0.02 > P > 0.01$
β_1	1.93 ± 1.78	1.1	$0.3 > P > 0.2$
β_2	2.84 ± 1.69	1.7	$0.1 > P > 0.05$
β_L	12.30 ± 1.37	9.0	$P < 0.001$
β_{12}	−0.90 ± 1.96	0.5	$0.7 > P > 0.6$
β_{1L}	3.12 ± 1.41	2.2	$0.05 > P > 0.02$
β_{2L}	0.54 ± 1.41	0.4	$0.7 > P > 0.6$
β_{11}	3.20 ± 1.32	2.4	$0.02 > P > 0.01$
β_{22}	4.21 ± 1.32	3.2	$0.01 > P > 0.001$

Deviations from regression: $\chi^2_{(50)} = 49.7$, $0.7 > P > 0.5$.

efficients are tabulated together with their standard errors and test of significance. Both estimates of regression on the quadratic functions of the simplex coordinates are significantly positive. This indicates that the response to mixtures becomes smaller as the centroid of the simplex, which corresponds to a 1/3: 1/3: 1/3 mixture of the three oestrogens, is approached, and shows that the oestrogens have a mutually antagonistic action. The physiological significance of these findings is discussed by Claringbold (1955).

DISCUSSION

The simplex design in itself is a non-factorial design and may be criticised on these grounds. Factorial experiments in joint action studies lead to complex response surfaces even if one drug behaves simply as a dilution of the other (i.e. similar action, see Finney, 1952). Suppose a factorial experiment is designed for two factors (A, A') each at three levels. Suppose as a theoretical example both factors are simply doses of the one hormone, i.e., similar action must hold, and also suppose that response is linearly related to log dose. A possible design could be : —

		Dose of A (units)					
		1	2	4	Log$_2$ total dose		
Dose of A' (units)	1	2	3	5	1.00	1.59	2.32
	2	3	4	6	1.59	2.00	2.58
	4	5	6	8	2.32	2.58	3.00

The total dose administered to each animal in the nine groups of animals is shown in the body of the table, while the log total dose is shown as a subsidiary block of mixtures in one-one correspondence to the first block. If response is linear to log dose it must be proportional to these elements apart from some constant. Thus in the simplest case a curved response surface must be evaluated. Also if treatments consisting of one substance or control treatments are included they create difficulties since the log of zero is $-\infty$. The data must be analysed, therefore, in a number of disconnected steps. Plackett and Hewlett (1951) use this method and their analysis takes the following form:

1. Fit one substance dose response lines.
2. Predict on basis of alternative models the response to joint doses.
3. Choose the hypothesis which describes the observed data best.

Using the method described in this paper in the theoretical example,