STATISTICAL DESIGNS AND RESPONSE MODELS FOR MIXTURES OF CULTIVARS[1]

W. T. Federer[2]

ABSTRACT

Concepts, definitions, statistical designs, and statistical analyses are presented for experiments involving mixtures or composites of k of v cultivars, lines, species, etc. General mixing (blending) effects, and various types of interaction (bi-specific, tri-specific, · · ·, n-specific mixing) effects are defined; various response model equations are developed. The statistical designs given are derived from weighing, balanced incomplete block, and supplemented block designs. It is noted that solutions for parameters in the response model equations are dependent upon the number and kind of mixtures in the design. The design and analysis of a particular example of mixtures of size two of eight bean cultivars and of the cultivars themselves, are described in detail. Finally, it is noted that the results of this paper may be applied to other than experiments on cropping.

---

[1] Contribution from the Biometrics Unit, Dep. of Plant Breeding and Biometry, Cornell University, Ithaca, N. Y. 14853. Rec'd. ____________

[2] Liberty Hyde Bailey Professor of Biological Statistics.

<u>List of symbols other than Roman alphabet or Arabic numbers</u>

μ	mu
ρ	rho
Σ	summation sign (sigma)
τ	tau
ϵ	epsilon
δ	delta
γ	gamma
π	pi
λ	lamba
β	beta
α	alpha
$>$	greater than
$<$	less than
$\leq$	less than or equal to

$\binom{v}{3}$ $\binom{v}{4}$ $\binom{v}{k}$

$\underline{\mu+\tau+\delta}$

$\underline{\beta}$

$\underline{\gamma}$

List of tables

Additional index words: Treatment design, Balanced incomplete block designs, General mixing or blending effects, Biblend or bispecific mixing effect, k-blend or k-specific mixing effect.

INTRODUCTION

The growing of mixtures (composites) of cultivars, lines, or species either in adjoining rows or plots, on the same area of land, or in successive (multiple) cropping systems has been and is of interest to agricultural researchers. The beneficial or detrimental effects of using mixtures as compared to monoculture (solid seeding, sole cropping) needs to be assessed. Various aspects of statistical design and analysis have been considered by many authors under the topics of diallel crossing or competition (e.g., see Hanson et al. (1961), Jensen and Federer (1964, 1965), Rawlings (1974), Kawano et al. (1974), Khalifa and Qualset (1974), Jensen (1978), Laskey and Wakefield (1978), to list a few). Most reported work relates to mixtures of two entities either in adjoining rows or as a composite. Statistical designs and analyses are required for $k \geq 2$ entities in each of c different mixtures. The present study is a contribution in this direction.

Statistical designs for evaluating various effects involved in growing mixtures, composites, or blends were considered to a limited extent by Aiyer (1949), Jensen (1952), and Federer et al. (1976). To extend and enhance these notions, definitions and concepts for mixtures of size k from a set of v lines are presented in the second section. Various forms of response and response models are discussed. Among the important definitions are those relating to a general mixing effect, which is, in some sense, comparable to general combining ability in genetic breeding experiments, and specific mixing effects. The latter effects, when confined to specific mixing effects between two lines, correspond to the concept of specific combining ability.

In the third section, treatment designs are presented for obtaining solutions for general and specific mixing effects. Some discussion is given relating to the nature of the parameters and restraints on parameters. In addition to considering mixtures of size k from v lines, two particular situations are discussed. The first relates to screening new lines for general mixing ability and the second situation relates to finding k_1 new lines out of a set of v which mix well with k_2 available lines.

Statistical analyses may be formulated under general linear model theory, but details need to be worked out for each specific experiment. Some aspects of the statistical analysis are considered and some solutions are given for specific cases. Statistical analyses for an actual experiment involving 28 mixtures of size two of eight bean cultivars plus the eight treatments of solid seeded single cultivars are discussed in the fifth section.

In the last section it is noted that the treatment designs could be used in other than cropping studies.

DEFINITIONS AND FORM OF RESPONSE

A treatment is a single entity in an experiment, and a treatment design refers to the selection of the treatments to be included in an experiment and is one of the components of statistical design. (See Federer (1955), Federer and Balaam (1972), and Federer and Federer (1973).) For a forage crop experiment where v legume lines are overseeded with a grass line, the herbage yield from the plot is composed of legume, grass and weeds. Total sward weight is obtained from each experimental unit, the smallest unit receiving one treatment, and the herbage is sampled to determine, by hand-separation, the relative proportions of legume, weeds and grass. The treatment design for $v = 7$ legume lines, A,B,C,D,E,F, and G, might be depicted for treatment design I in Table 1. Each of the 7 lines are grown with grass, which forms treatments 1 to 7. Treatment 8 is grass only. In some cases, treatment number 8 would be omitted from the treatment design.

A second treatment design might be the one listed as treatment design II in Table 1. The 7 lines are mixed together to form treatment 1, and 3 of the 7 lines are mixed together to form treatments 2 to 8. All mixtures are overseeded with grass. In this treatment design, plants from all $k = 3$ lines in each experimental unit appear at random; there is no separation into subplots for each line.

A third treatment design and arrangement is to divide the experimental unit into $1/v^{th} = 1/7^{th}$ or $1/k^{th} = 1/3^{rd}$ subplots and to keep the $k = 3$ or $v = 7$ lines completely separate from each other in subplots. This is treatment design III in Table 1. The $k = 3$ lines are randomly allotted to the $k = 3$ subplots within each experimental unit. The treatment arrangement is that for an optimal weighing design for weighing 7 objects (see Federer (1955), section XV.4, and Raghavarao (1971), chapter 17) in groups of 3 or 7. Here the subplot yields may be obtained as a composite by cutting a swath through the experimental unit and obtaining the total experimental unit yield, <u>or</u> they may be obtained individually. The form of the statistical analysis will depend upon which method of obtaining yields is utilized. If treatment number 1 is omitted this corresponds to a symmetrical balanced incomplete block design with $v = b = 7$, $k = r = 3$, and $\lambda = 1$, where v is the number of entries, b is the number of blocks, k is the block size, r is the number of replicates of each entry, and λ is the number of times any pair of varieties occurs together in the b blocks.

Suppose that a randomized complete block design is to be utilized. Then, for treatment design I, the yield response could be of the form:

$$Y_{gi} = \mu + \rho_g + \tau_i + \epsilon_{gi} \quad , \qquad [1]$$

where Y_{gi} is the yield of the i^{th} treatment in the g^{th} block, $g = 1,2 \cdots,r$, $i = 1,2,\cdots,v$, τ_i is the effect of the i^{th} line and is a fixed effect, ρ_g is the effect of the g^{th} replicate and the ρ_g are identically and independently distributed (IID) with zero mean and variance σ^2_ρ, and the ϵ_{gi} are IID with zero mean and common variance σ^2_ϵ and independent of the ρ_g. Many other models are of course possible, but we shall confine our attention to this simplistic model. Now, when the proportion of legume is estimated by $\hat{p}_{gi}$, say, then the estimated legume weight is $\hat{p}_{gi}Y_{gi}$. The estimated variance of the $\hat{p}_{gi}Y_{gi}$ will certainly be different than the variance of the Y_{gi}. One could perform a statistical analysis on the Y_{gi} to obtain an estimate of the variance for hay yields and then on the $\hat{p}_{gi}Y_{gi}$ to obtain an estimate of the variance for legume, grass, or week yields. This requires that a separate $\hat{p}_{gi}$ be made for each experimental unit.

Again, suppose that treatment design III is used in a randomized complete block design. The response equation could be of the linear, additive form as follows:

$$Y_{gi} = \mu + \rho_g + \Sigma^v_{j=1} n_{ij}\tau_j/k + \epsilon_{gi} \quad , \qquad [2]$$

where $n_{ij} = 1,0$, depending upon whether or not the j^{th} line is included in the i^{th} treatment composite of k lines and the other symbols are as defined in [1]. This model presumes no border or competitive effects and that only the total yield for treatment i in block g is available.

Treatment design II brings up a variety of new concepts and problems. For the sake of clarity, we shall consider mixtures of two and of three lines before generalizing to k lines and shall consider the simple models of the form of [1] and [2]. For the first situation consider composites of two lines corresponding to the diallel crossing design in genetic studies. We shall call the treatment design with mixtures of $k = 2$ lines as a biblend mixing design corresponding to the term diallel crossing design. For $k = 3$ lines in a mixture call this a triblend mixing design, and in general for k lines denote this as a k-blend mixing design.

Also, different types of response may be expected using treatment design II rather than treatment designs I or III. If a line generally performs differently when surrounded by individuals of different lines as compared to individuals of the same line, we denote this as a general mixing (or blending) effect to correspond to the term general combining ability. In any particular experiment with v lines, the estimates of general mixing effects will be relative only to those v lines in the experiment. To obtain estimates of these effects it will be necessary to include both treatment designs I and II, or II and III, in an experiment. It could be that all general mixing effect estimates are positive, that all are negative, or that some are positive and some are negative. Differences in estimated line effects of treatment design II and treatment design I (or III) will provide estimates of general mixing. One form of a yield equation when general mixing effects are present and when k lines are in the mixture, would be:

$$Y_{gi} = \mu + \rho_g + \Sigma_{j=1}^{v} n_{ij}(\tau_j+\delta_j)/k + \epsilon_{gi} , \qquad [3]$$

where δ_j is the parameter associated with the general mixing effect of line j and the remaining elements are as described for [2]. Note that $\Sigma_1^v \delta_j$ should not be taken as zero as all δ_j could be positive (or negative).

The situation becomes more complex when lines in combination interact to produce specific mixing effects. For example, for v lines in mixtures of k = 2 lines, there could be a specific mixing effect (interaction) between the two lines in the mixture. For lines h and i, say, denote the specific mixing effect by γ_{hi}. Then, [3] would be modified as follows:

$$Y_{ghi} = \mu + \rho_g + (\tau_h + \delta_h)/2 + (\tau_i + \delta_i)/2 + \gamma_{hi} + \epsilon_{ghi} \quad . \qquad [4]$$

Note that this form of the response equation with δ_h and δ_i omitted was used by Hanson et al. (1961), Jensen and Federer (1965), and Rawlings (1974). Since the h^{th} line occupies one-half of the experimental unit (plot), the line effect and the general mixing effect of the line are divided by two. The specific mixing effect (interaction) between two lines is not divided by two since two lines must be present for the interaction to occur. (Note that the above would be a useful concept for specific and general combining ability effects (see Eberhart and Gardner (1966)). For k = 3 lines in a mixture, the following yield equation for lines h, i, and j could represent the response:

$$\begin{aligned} Y_{ghij} = \mu + \rho_g &+ [(\tau_h + \delta_h) + (\tau_i + \delta_i) + (\tau_j + \delta_j)]/3 \\ &+ 2[\gamma_{hi} + \gamma_{hj} + \gamma_{ij}]/3 + \pi_{hij} + \epsilon_{ghij} \quad , \end{aligned} \qquad [5]$$

where π_{hij} is a three-line interaction effect for lines h, i, and j, and the other symbols are as defined previously. Note that each line occupies one-third of a plot and hence the line effect plus the line general mixing effect is divided by $k = 3$. Likewise, the specific mixing effect for any two lines is from plants occupying only two-thirds of the area and hence the multiplier 2/3 in [5]. An interaction effect of the form γ_{hi} in [4], as opposed to no such term in [3], is denoted as a <u>bispecific</u> mixing (or blending) effect to correspond to the term specific combining ability effect in genetic experiments. Likewise, the three-line interaction effect will be denoted as a <u>trispecific mixing effect</u>, and, in general, the interaction effect among n lines will be denoted as the <u>n^{th}-specific mixing effect</u>. A form of the response equation for the i^{th} treatment in the g^{th} block and including the interaction effects could be:

$$
\begin{aligned}
Y_{gi} = \mu + \rho_g + \Sigma_{j_1=1}^{v}\, n_{ij_1}(\tau_{j_1} + \delta_{j_1} &+ \Sigma_{j_2} n_{ij_2}(2\gamma_{j_1 j_2} \\
&+ \Sigma_{j_3} n_{ij_3}(3\pi_{j_1 j_2 j_3} + \Sigma_{j_4}(4\beta_{j_1 j_2 j_3 j_4} + \cdots \qquad [6] \\
&+ n_{ij_k}\, \alpha_{j_1 j_2 \cdots j_k})\cdots)/k + \epsilon_{gij} \;,
\end{aligned}
$$

where $j_1 < j_2 < j_3 < \cdots < j_k$ and the other elements are as defined previously. Note that from the above definitions the <u>unispecific mixing effect</u> for $n = 1$ becomes the general mixing effect (δ_h).

In order to have a short and concise notation, denote the general mixing effect as <u>gme</u> and the n^{th}-specific mixing effect as <u>n^{th}-sme</u>.

CONSTRUCTION OF TREATMENT DESIGNS FOR n^{th}-sme's

Federer et al. (1976) have shown how to construct treatment designs for estimating mean plus line effect plus general mixing effect of a line, and for estimating line effect plus gme of line. They used weighing designs and balanced incomplete block design theory. Minimal designs to estimate gme's and bi-sme's for v lines grown in mixtures of k lines have been considered in a M.S. thesis by D. B. Hall, Cornell University, 1976. We shall consider designs for additional situations.

Consider the particular example for v = 7 lines and for k = 3 lines in the composite or mixture. Treatment designs II and III in the preceding section consisted of a particular subset of all possible combinations of 7 items taken 3 at a time, i.e., $\binom{7}{3} = 7!/3!4! = 35$. These 35 combinations are given in Table 2. Blocks 1 to 7 form a symmetrical balanced incomplete block design (SBIBD) with parameters v = b = 7, r = k = 3, and $\lambda = 1$. Blocks 8 to 14 also form a SBIBD with the same parameters. Blocks 15 to 35 form a BIBD with parameters v = 7, k = 3, b = 21, r = 9, and $\lambda = 3$.

If the response model is given by [3], then blocks 1 to 7, blocks 8 to 14, blocks 15 to 35, or any combination of these may be used to estimate the $(\mu + \tau_j + \delta_j)$ effects given that all plants (seeds) are randomly mixed within the experimental unit as in treatment design II. If both treatment designs I and II, or II and III, arrangements are used then solutions for τ_j and δ_j are obtainable.

If response model [5] holds except that $\pi_{hij} = 0$ for all hij, it has been shown that blocks 15 to 35 form the minimal sized treatment design allowing unique solutions for gme and bi-sme effects. (D. B. Hall, loc. cit.) If constraints such as $\Sigma_{i=1}^{v} (\tau_i+\delta_i) = 0$, and $\Sigma_{h \text{ or } i}\, \gamma_{hi} = 0$

for all h and i are used, then solutions are possible for $(\mu+\rho_g)$, $(\tau_i+\delta_i)$, and γ_{hi}. One may use all 35 treatments (blocks) or blocks 15 to 35 plus any subset of blocks 1 to 14. To obtain solutions for all parameters in [5] under the constraints

$$\Sigma_{h=1}^{v} (\tau_h+\delta_h) = \Sigma_{h \text{ or } i}\, \gamma_{hi} = \Sigma_{h,i, \text{ or } j}\, \pi_{hij} = 0 , \qquad [7]$$

it is necessary to have the entire set of 35 treatments composed of blocks of k = 3 lines.

In general, for blocks of k = 3 lines and response model [5], the number of parameters for which solutions are to be obtained, the number of constraints placed on solutions to obtain unique solutions for the parameters, and the number of degrees of freedom associated with the mean, gme, bi-sme, and tri-sme are given in the top part of Table 3 for k = 3 lines in a mixture. In the middle part of Table 3, degrees of freedom for values of v = 3 to 9 are given. It is impossible to obtain solutions for all effects from all combinations of v lines taken 3 at a time unless $v > 5$.

In the bottom part of Table 3 the number of parameters, the number of independent constraints required to obtain unique solutions for the parameters, and the degrees of freedom for the mean, gme, bi-sme, tri-sme, and quater-sme are given for mixtures of k = 4 lines. From the total number of combinations of v lines taken 4 at a time, we note that v must be greater than 7 to obtain solutions for all effects.

If solutions for tri-sme's from mixtures of k = 3 and $v < 6$ or for quater-sme's from mixtures of k = 4 and $v < 8$ are desired, one procedure is to include mixtures of 2 and 3 lines for the former case and mixtures of 2, 3, and 4 lines for the latter case. This considerably increases the number of treatments in an experiment. If one does this, another point needs to be considered. D. B. Hall (loc. cit.) has pointed out that a bi-sme in mixtures of $k = 2$ may be different than the same bi-sme evaluated in mixtures of $k = 3$. This may also be true for gme's. If this situation holds, then it is necessary to use treatment designs for $k = 1,2,3,\cdots,n$ when solutions for effects up to the n^{th}-sme are required. When it is desired to use only a block size of $k = n$ and to obtain solutions up to the n^{th}-sme, the following procedure is suggested as an alternative to the case of variable k. The case of $v = 7$ and $k = 4$ is considered first.

To obtain solutions for quater-sme effects, we shall use several sets of mixtures of 4 lines as follows for $v = 7$. There are $(v-1)(v-2)(v-3)/6 = 20$ possible sets of 3 lines among the 6 lines not involving line h, and there are $(v-1)(v-2)(v-3)(v-4)/24 = 15$ sets of 3 lines which involve line h. There will be 7 sets of 20 mixtures of 3 lines which do not involve line h, $h=1,2,\cdots,7$. In each set of 20 mixtures for a given h, solutions for tri-sme's are obtainable. A given tri-sme, say π_{123}, will have solutions obtainable from $v-3 = 4$ lines, i.e., 4, 5, 6, and 7. Note that since solutions are obtained from a different set of 20 combinations for each of the lines 4, 5, 6, and 7 that independent solutions are available. Then, a solution for π_{123} can be obtained from all 140 entries. The deviation of the solution for π_{123} for line 4, say, from the average π_{123} is a solution for the quater-sme β_{1234}. Solutions for β_{1235}, β_{1236}, and β_{1237} can be obtained in a similar manner. The remaining quater-sme's can be obtained by the same procedure. Quater-sme's for $v = 6$ lines in mixtures of $k = 4$ may be obtained similarly. For smaller v, quater-sme's will need to be obtained from mixtures of 2, 3, and 4, whereas for $v > 7$, solutions for these effects are obtainable from all possible mixtures of v lines taken 4 at a time. (See bottom part of Table 3.) For example, with $\binom{9}{4} = 126$ combinations, the estimable functions are one for mean, 8 for gme's, 36 - 9 = 27 for bi-sme's, 84 - 36 = 48 for tri-sme's, and this leaves 126 - 84 = 42 for quater-sme's.

Proceeding in the same manner as from k = 3 to k = 4, one can go from k = 4 to k = 5 and so forth. As can be seen, the number of treatments becomes large. The problem of finding minimal designs for general k and v given that p^{th} and higher effects are all zero in response model [6] is an unresolved problem. If such effects were present, an experimenter could use mixtures sizes of (p-1) lines and could transfer the results to mixtures of k lines $p \leq k$, since none of the higher-ordered sme's are present.

In evaluating mixtures of lines, cultivars, or species for gme's and sme's, many situations arise. Two of these will be discussed. For the first case, consider the situation wherein the experimenter desires to screen lines for gme in a similar manner to screening lines for general combining ability in genetic studies. A procedure suggested is similar to that for top-crossing. First select (k-1) tester cultivars; second plant the area to tester cultivars and to a new line in equal proportions in a mixture, i.e., the seed (or plants) of the new line is equal to the total of the seed (or plants) of the k-1 tester cultivars. For k - 1 = 2 (tester cultivars X and Y, say) and v = 8 (new lines A, B, C, D, E, F, G, and H, say) the treatment design would be:

Treatment number for combination or mixture

1	2	3	4	5	6	7	8
X,Y, and A	X,Y, and B	X,Y, and C	X,Y, and D	X,Y, and E	X,Y, and F	X,Y, and G	X,Y, and H

In addition, a ninth treatment could be X and Y together in a mixture. The reason for having the seed (or plants) constitute one-half of the plot is to better evaluate the line. If only one-third of the plot were devoted to the new line, its effect would be diminished and would be more difficult to measure statistically. Using the above type of treatment design, lines could be screened for general mixing ability.

In a second situation, the experimenter might be searching for k_1 new lines to combine with k_2 standard lines to form a mixture of $k_1 + k_2 = k$ lines. Instead of adding individual lines as in the above procedure, all possible combinations of v new lines taken k_1 at a time could be used. A treatment consisting of the k_2 standard lines could be included as a check treatment. Supplemented block experiment design theory (Raghavarao (1971)) is usable for the above and other related mixture designs.

STATISTICAL ANALYSES

If only gme's are to be estimated, i.e., response model [2] or [3] holds, and if a symmetrical balanced incomplete block design (SBIBD) with incidence matrix N is used to obtain the mixtures of k lines, then the yield equations in usual matrix form are:

$$(N'_{v\times v} \quad kI_{v\times v}) \begin{pmatrix} \mu+\tau+\delta_{v\times 1} \\ \beta_{v\times 1} \end{pmatrix} = Y_{B,v\times 1}/r \qquad [8]$$

or

$$NN'(\underline{\mu+\tau+\delta}) + kN\underline{\beta} = N\underline{Y}_B/r \qquad [9]$$

where N is the line by block matrix of zeros and ones denoting whether or not a given line occurs in a given block, $N'N = NN' = (r^*-\lambda)I + \lambda J$, given that the parameters of the SBIBD are $v = b$, $r^* = k$, and λ, r = the number of replicates of a mixture in the experiment, J is a $v \times v$ matrix of ones, $\underline{Y}_B$ is a $v \times 1$ vector of totals of mixtures of k lines. Note that the vector $\underline{Y}_T$ of line totals is not available in this type of experiment and that the elements of $\underline{\beta}$ have zero expectation and variance σ^2_ϵ, since blocks of size k are the experimental units in the experiment. Hence, the solutions are:

$$\underline{\widehat{\mu+\tau+\delta}} = \frac{(NN')^{-1}}{r} N\underline{Y}_B = \frac{1}{r}\left(\frac{1}{r^*-\lambda} I - \frac{\lambda}{(r^*-\lambda)(r^*+(v-1)\lambda)} J\right) N\underline{Y}_B \qquad [10]$$

and

$$\underline{\widehat{\tau+\delta}} = \frac{1}{r}(NN')^{-1}N\underline{Y}_B - (\underline{1}'\underline{Y}_B/rvr^*)\underline{1} \quad , \qquad [11]$$

where $\underline{1}$ is a $v \times 1$ vector of ones, and $\underline{1}'(\underline{\tau+\delta})$ is taken to be zero, and hence $\hat{\mu} = \underline{1}'\underline{Y}_B/rvr^*$. For the general BIB design with parameters v, r^*, k, b, and λ, $NN' = (r^*-\lambda)I + \lambda J$ but $N'N \neq NN'$ (see Raghavarao (1971), chapter 10). However, [10] still gives the solution for $\underline{\widehat{\mu+\tau+\delta}}$.

When sme's are present, little is known about the situation; this problem has been investigated to some extent by Hall (loc. cit.) and minimal treatment designs have been determined to obtain solutions for gme's and bi-sme's for a specific case. The nature of the solution matrix when bi-sme's are present is unresolved.

As another special example, consider that the experimenter used blocks 1 to 14 of the 35 blocks given in Table 2. Note that under response model [3] blocks 1 to 7 provide one set of estimates of $\mu + \tau_h + \delta_h$ while blocks 8 to 14 provide a second set. Likewise, blocks 1 to 14 provide combined estimates of $\mu + \tau_h + \delta_h$. If one takes the sums of squares of differences between the two estimates for each of the 7 lines and divides by 6r, and if the γ_{hi} are considered to be $IID(0,\sigma^2_\gamma)$, this sum of squares has expectation $7(\sigma^2_\epsilon + r\sigma^2_\gamma)$. Thus, a variance component estimate for bi-sme's can be obtained using only blocks 1 to 14, even though solutions for the bi-sme's are not obtainable.

STATISTICAL ANALYSES FOR AN EXAMPLE

In 1967 Professor Neil Rutger (formerly of Cornell University but now at the University of California at Davis) conducted an experiment designed as a randomized complete blocks design with $v(v+1)/2 = 36 = c$ treatments, where eight of the treatments represent single cultivar mixtures and $8(7)/2 = 28$ treatments represent mixtures of two bean cultivars. All eight bean cultivars had different colored seeds so that it was possible to separate the seeds and to obtain yields for each of the two cultivars in a mixture. A total of $r(2(8)(7)/2+8) = 64r$ observations was available from the experiment. Note that for the individual subplot yields, the statistical analysis will take on aspects of a split-plot design analysis (see, e.g., Federer (1975)). An analysis of variance for the data from experiments of this type for v cultivars in r blocks of a randomized complete block design would be as given in Table 4. The analysis above the dotted lines is performed on sums or totals from experimental units while the analysis below the dotted line is carried out on differences between yields of cultivars h and i in each of the experimental units containing a mixture of two bean cultivars, i.e., $r(v^2-v)/2$ experimental units. Note that sums and differences are orthogonal and that one could extend the analysis to mixtures of k cultivars in the same manner. Note also that the sources of variation in the split plot part of the analysis of variance (i.e., below the dotted line) are put in quotes to indicate that these are interaction terms with the source of variation in quotes (see example VIII.1 of Federer (1955)).

An appropriate way to observe the nature of the sources of variation below the dotted line is to set up a model for the yield equations for this experiment. Let the individual yields for cultivar i in a mixture with cultivar h be denoted by

$$Y_{g(h)i} = (\mu+\rho_g)/2 + (\tau_i+\delta_i)/2 + \gamma_{hi}/2 + \epsilon_{g(h)i} , \qquad [12]$$

and let the individual yields of cultivar h in a mixture with cultivar i be denoted by

$$Y_{gh(i)} = (\mu+\rho_g)/2 + (\tau_h+\delta_h)/2 + \gamma_{hi}/2 + \epsilon_{gh(i)} , \qquad [13]$$

and let the yield of cultivar h only plots be denoted by

$$Y_{ghh} = \mu + \rho_g + \tau_h + \epsilon_{ghh} , \qquad [14]$$

where the symbols are as defined for response model [5]. Note that these response equations may be extended to mixtures of k lines. For example, for k = 3 let the response equations be:

$$Y_{g(hi)j} = (\mu+\rho_g)/3 + (\tau_j+\delta_j)/3 + (\gamma_{hj}+\gamma_{ij})/3 + \pi_{hij}/3 + \epsilon_{g(hi)j} ,$$

$$Y_{g(hj)i} = (\mu+\rho_g)/3 + (\tau_i+\delta_i)/3 + (\gamma_{hi}+\gamma_{ij})/3 + \pi_{hij}/3 + \epsilon_{g(hj)i} ,$$

$$Y_{gh(ij)} = (\mu+\rho_g)/3 + (\tau_h+\delta_h)/3 + (\gamma_{hi}+\gamma_{hj})/3 + \pi_{hij}/3 + \epsilon_{gh(ij)} ,$$

and

$$Y_{ghhh} = \mu + \rho_g + (\tau_h+\delta_h) + \epsilon_{ghhh} .$$

Consider now the differences of [12] and [13],

$$Y_{g(h)i} - Y_{gh(i)} = (\tau_i+\delta_i)/2 - (\tau_h+\delta_h)/2 + \epsilon_{g(h)i} - \epsilon_{gh(i)} , \quad [15]$$

which are utilized to obtain the last part of the analysis of variance in Table 4. Performing the same type of analysis on these differences (or single degree-of-freedom contrasts for mixtures of k cultivars), we obtain the bottom part of the table.

Alternatively, let us approach this analysis in the manner described by Federer (1975), section 3. Perform analyses of variance for each mixture of k of v cultivars in r complete blocks, after first constructing the two-way tables in the top part of Table 5. The analyses of variance for each of these tables are presented in the bottom half of Table 5, where c is the total number of combinations of mixtures of k cultivars. For our case, $c = 8(7)/2 = 28$ and $k = 2$. The sum of the 28 sums of squares for block X cultivars is equal to the "blocks" plus "blocks X treatments" in the analysis of variance in Table 4. Note that since these are differences, additive block effects are not present and hence, the "blocks" should be pooled with the "blocks X treatments", and that this corresponds to the "error (b)" sum of squares in a split plot analysis. Also, the sum of the sums of squares for cultivars for $k = 2$ and $c = v(v-1)/2$ is that for "treatments" plus the "correction for the mean" in the previous analysis; this sum of squares represents variation among cultivar yields within a specific mixture. Contrasts of the form of c_5 in Table 4 represent a general effect of a cultivar from the mean difference of all cultivars, and c_6 represents an interaction effect of a cultivar with individual cultivars.

A statistic of interest would be to compare the mean yield of sole cropped cultivar plots with the cultivar yield when grown in a mixture. The estimated difference is obtained as the difference of the means as follows for cultivar i:

$$2Y_{\cdot(\cdot)i}/r(v-1) - Y_{\cdot ii}/r = \bar{y}_{\circ(\cdot)i} - \bar{y}_{\cdot ii} = \hat{\delta}_i \,, \qquad [16]$$

where $Y_{\cdot ii}$ is the sum over blocks for the response in [14] and $Y_{\cdot(\cdot)i}$ is the sum of the yields in [12] over blocks and over the v-1 other cultivars with which it appears. If mixtures are to be beneficial, the $\Sigma_1^v \hat{\delta}_i$ must be a positive quantity (see Jensen and Federer (1964)), and it cannot be estimated unless both mixtures and solid stands are present.

One could question the splitting of the $(\mu+\rho_g)$ effect and the γ_{hi} into two equal parts in [12] and [13]. The same question could arise in the equation for mixtures of three cultivars. The justification for this is that γ_{hi} is a component of the particular combination of cultivars h and i in the blend and since <u>equal amounts</u> of seed were used, it would appear justifiable to split this effect. The same is true for the $\mu + \rho_g$ component since it is an experimental unit, not a sub-experimental unit, component. The yields then are on a $1/k^{th}$ experimental unit basis. Likewise, this is the reason for including τ_h in [14]; it is a total of k $1/k^{th}$ units.

APPLICATIONS OF THE RESULTS IN OTHER AREAS

In herbicide studies involving composites of k chemical units, one might wish to assess the effect of each of v chemical units as well as n-sme effects. The preceding discussion applies directly to such areas and has been used in chemical research by Free and Wilson (1964). As is frequently the case, statistical procedures developed for one type of experimentation have usefulness in other areas. Results for diallel crossing experiments have uses in competition studies between pairs of cultivars (Hanson et al. (1961), Jensen and Federer (1965), Rawlings (1974)). The concepts, designs, and analyses described herein have usefulness in research involving nutrition, medicine, recreation, education, surveys (Smith et al. and Raghavarao and Federer (1979)), and other areas involving mixtures of items where gme, bi-sme, etc. effects are present. There appears to be a large number of areas involved with studies on composites of items.

LITERATURE CITED

1. Aiyer, A. K. Yegna Narayan. 1949. Mixed cropping in India. Indian J. Agr. Sci. 19:439-543.

2. Eberhart, S. A. and C. O. Gardner. 1966. A general model for genetic effects. Biometrics 22:864-881.

3. Federer, W. T. 1955. Experimental Design – Theory and Application (2nd printing 1963). The Macmillan Co., New York. (Reprinted by the Oxford and IBH Publ. Co., Calcutta, India, in 1967 with 2nd printing in 1974.)

4. -------, W. T. 1975. The misunderstood split plot. Applied Statistics Proceedings of the Conference at Dalhousie University, Halifax, May 2-4, 1974. R. P. Gupta (ed). North Holland Publ. Co. Amsterdam, Oxford.

5. -------, W. T. and L. N. Balaam. Bibliography on Experiment and Treatment Design Pre-1968. Published for the International Statistical Institute by Oliver and Boyd, Edinburgh, 769 pp.

6. -------, W. T. and A. J. Federer. 1973. A study of statistical design publications from 1968 through 1971. The Am. Stat. 27:160-163.

7. -------, W. T., A. Hedayat, C. C. Lowe, and D. Raghavarao. 1976. Application of statistical design theory to crop estimation with special reference to legumes and mixtures of cultivars. Agron. J. 68:914-919.

8. Free, Jr., S. M. and J. W. Wilson. 1964. A mathematical contribution to structure-activity studies. J. Med. Chem. 7:395-399.

9. Hanson, W. D., C. A. Brim and Kuell Hinson. 1961. Design and analysis of competition studies with an application to field plot competition in the soybean. Crop Sci. 1:255-258.

10. Jensen, N. F. 1952. Intra-varietal diversification in oat breeding. Agron. J. 44:30-34.

11. ------, N. F. 1978. Seasonal competition in spring and winter wheat mixtures. Crop Sci. 18:1055-1057.

12. ------, N. F. and W. T. Federer. 1964. Adjacent row competition in wheat. Crop Sci. 4:641-645.

13. ------, N. F. and W. T. Federer. 1965. Competing ability in wheat. Crop Sci. 5:449-452.

14. Kawano, K., H. Gonzalez and M. Lucena. 1974. Intraspecific competition, competition with weeds, and spacing response in rice. Crop Sci. 14:841-845.

15. Khalifa, M. A. and C. O. Qualset. 1974. Intergenotypic competition

between tall and dwarf wheats. Crop Sci. 14:795-799.

16. Laskey, B. C. and R. C. Wakefield. 1978. Competitive effects of several grass species and weeds on the establishment of birdsfoot trefoil. Agron. J. 70:146-148.

17. Raghavarao, D. 1971. Constructions and Combinatorial Problems in Design of Experiments. John Wiley and Sons, Inc., New York, London.

18. ----------, D. and W. T. Federer. 1979. Block total response as an alternative to the randomized response method in surveys. J. Roy. Stat. Soc., Ser. B, 41: (to be published).

19. Rawlings, J. O. 1974. Analysis of diallel-type competition studies. Crop Sci. 14:515-518.

20. Smith, L. L., W. T. Federer and D. Raghavarao. 1974. A comparison of three techniques for eliciting truthful answers to sensitive questions. Social Statistics Section Proceedings of the American Statistical Association, pp. 447-452.

Table 1. Three possible treatment designs for mixtures of cultivars.

Treatment design and arrangement I

Treatment number

1	2	3	4	5	6	7	8
grass plus line A	grass plus line B	grass plus line C	grass plus line D	grass plus line E	grass plus line F	grass plus line G	grass only

Treatment design and arrangement II

Treatment number

1	2	3	4	5	6	7	8
grass plus all 7 lines	grass plus lines A,B,D	grass plus lines B,C,E	grass plus lines C,D,F	grass plus lines D,E,G	grass plus lines E,F,A	grass plus lines F,G,B	grass plus lines G,A,C

Treatment design and arrangement III (with all experimental units over-seeded with grass)

Treatment number

1	2	3	4	5	6	7	8
A B C D E F G	A B D	B C E	C D F	D E G	E F A	F G B	G A C

Table 2. Thirty-five possible combinations (blocks) of size $k = 3$ for $v = 7$ treatments.

block		block		block		block		block	
1	1 2 4	8	3 5 6	15	1 2 3	22	1 4 7	29	2 5 7
2	2 3 5	9	4 6 7	16	1 2 5	23	1 6 7	30	3 4 5
3	3 4 6	10	5 7 1	17	1 2 7	24	2 3 4	31	3 4 7
4	4 5 7	11	6 1 2	18	1 3 5	25	2 3 6	32	3 5 7
5	5 6 1	12	7 2 3	19	1 3 6	26	2 4 6	33	3 6 7
6	6 7 2	13	1 3 4	20	1 4 5	27	2 4 7	34	4 5 6
7	7 1 3	14	2 4 5	21	1 4 6	28	2 5 6	35	5 6 7

Table 3. Numbers of parameters and constraints on solutions with associated degrees of freedom for effects from [5] (see text) for k = 3 and 4.

Mixtures of k = 3

Source	Number of Parameters	Number of Independent Constraints	Degrees of Freedom
Total	$(v^3+5v+6)/6$	$(v^2+v+2)/2$	$\binom{v}{3}$
mean	1	0	1
gme	v	1	v-1
bi-sme	v(v-1)/2	v	v(v-3)/2
tri-sme	v(v-1)(v-2)/6	$\Sigma_{i=1}^{v}(v-i) = v(v-1)/2$	v(v-1)(v-5)/6

Degrees of Freedom From Above

Source	v = 3	v = 4	v = 5	v = 6	v = 7	v = 8	v = 9
Total	1	4	10	20	35	56	84
mean	1	1	1	1	1	1	1
gme	2	3	4	5	6	7	8
bi-sme	0	2	5	9	14	20	27
tri-sme	-2	-2	0	5	14	28	48

Mixtures of k = 4

Source	Number of Parameters	Number of Independent Constraints	Degrees of Freedom
Total	$(v^4-2v^3+11v^2+14v+24)/24$	$(v^3+5v+6)/6$	$\binom{v}{4}$
mean	1	0	1
gme	v	1	v-1
bi-sme	v(v-1)/2	v	v(v-3)/2
tri-sme	v(v-1)(v-2)/6	v(v-1)/2	v(v-1)(v-5)/6
quater-sme	v(v-1)(v-2)(v-3)/24	v(v-1)(v-2)/6	v(v-1)(v-2)(v-7)/24

Table 4. All possible mixtures of size 2 of 8 bean cultivars plus the cultivars themselves to form 36 treatments and an analysis of variance table for these entries from a randomized complete block design.

Source of variation	Degrees of freedom	Sum of squares
Total	rv^2	
Correction for mean	1	standard randomized complete block analysis
Blocks	r-1	
Treatments	(v+2)(v-1)/2	
c_1 = Among single cultivar yields	v-1	
c_2 = Single versus mixtures of 2	1	
c_3 = General mixing effects (gme)	v-1	see, e.g., section VIII.5 of Federer (1955)
c_4 = Specific mixing effects (sme)	v(v-3)/2	
Blocks × treatments	(r-1)(v+2)(v-1)/2	
c_1 × blocks	(r-1)(v-1)	see example VIII.1, e.g., of Federer (1955)
c_2 × blocks	(r-I)	
c_3 × blocks	(r-1)(v-1)	
c_4 × blocks	(r-1)(v)(v-3)/2	
Within experimental units	rv(v-1)/2	see above
"Correction for mean"	1	
"Blocks"	r-1	
"Treatments"	(v-2)(v+1)/2	
"c_5 = gme"	v-1	
"c_6 = sme"	v(v-3)/2	
"Blocks × Treatments"	(r-1)(v-2)(v+1)/2	
"c_5 × blocks"	(v-1)(v-1)	
"c_6 × blocks"	$(r-1)(v^2-3v)/2$	

Table 5. Two-way tables and analyses of variance for $\binom{v}{k} = c$ mixtures of k cultivars and v sole crop cultivars in a randomized complete block design.

Two-way Tables of Yields

Blocks	Cultivar yield $Y_{gj_1(j_2 j_3 \cdots j_k)}$	$Y_{g(j_1)j_2(j_3 \cdots j_k)}$	$\cdots$	$Y_{g(j_1 j_2 \cdots j_{k-1})j_k}$	Total
1					
2					
⋮					
r					
Total					

Degrees of Freedom in the Analyses of Variance

Source of variation	Treatment number and combination 1	2	...	c	Total
Total	rk	rk		rk	rkc
Correction for mean	1	1		1	c
Blocks	r-1	r-1		r-1	c(r-1)
Cultivars	k-1	k-1		k-1	c(k-1)
Blocks × Cultivars	(r-1)(k-1)	(r-1)(k-1)		(r-1)(k-1)	c(r-1)(k-1)