STATISTICAL ANALYSES FOR MULTISTAGE DESIGNS

Walter T. Federer
Cornell University, U.S.A.

BU-743-M July 1981

Abstract

Statistical analyses are discussed for the p^{th} experiment of p successive experiments conducted on the same set of experimental units. Response model equations for individual curvatures for the previous set of treatments and for additional nonadditivity parameters are presented. Then, statistical analyses are presented for p experiments conducted simultaneously on the same set of experimental units. Univariate and multivariate analyses are discussed. Examples for this latter situation in marketing and in intercropping investigations are presented.

Biometrics Unit Mimeo Series, Cornell University.

STATISTICAL ANALYSES FOR MULTISTAGE DESIGNS

WALTER T. FEDERER
Cornell University, U.S.A.

Experiment designs for conducting successive or simultaneous experiments on the same set of experimental units have been available since L. Euler constructed sets of orthogonal Latin squares in the eighteenth century. Statistical analyses for successively conducted experiments other than rotation experiments considered in the writings of F. Yates and W. G. Cochran, began to appear in the mid-twentieth century in the writings of S. C. Pearce and G. H. Freeman. The latter works were concerned with the effects of a previous set of treatments on the treatments in the present experiment for the same set of experimental units. Most of the interest in this class of experiment designs appears to have been centered on construction of designs suitable for this type of experimentation. Examples of this may be found in, e.g., Afsarinejad and Hedayat (1975), Anderson (1972), Anderson and Federer (1976), Bose and Srivastava (1964), Federer (1972), Freeman (1958), Hedayat et al. (1972), Hedayat and Raghavarao (1975), Hoblyn et al. (1954), Potthoff (1962a, 1962b), Preece (1966), Singh et al. (1981), and Srivastava and Anderson (1970, 1971).

Statistical analyses for successively conducted experiments have been considered by some authors, e.g., Bose and Srivastava (1964), Freeman and Jeffers (1962), Preece (1966), and Singh et al. (1981). Basically, all analyses are an extension of the general linear model of the form $E(\underset{\sim}{Y}) = \underset{\sim}{X}\underset{\sim}{\beta}$ Additional parameters are included in the parameter vector to take into account stratification of treatments in the present experiment due to the previous set of treatments. Whenever the treatments in the previous experiment are in a balanced or orthogonal arrangement to the treatments and stratification variables of the present experiment, relatively simple statistical analyses are possible.

To illustrate consider a Latin square design for both a past and the present experiment and consider that the two Latin squares are orthogonal. A response model equation of the following form is postulated for each of the Latin square experiments:

$$Y_{hij} = \mu + \rho_h + \gamma_i + \tau_j + \epsilon_{hij} \quad , \tag{1}$$

where μ, ρ_h, γ_i, τ_j, and ϵ_{hij} are a common mean, a row h effect, a column i effect, a treatment j effect, and an identically and independently normally distributed random error effect, respectively. For treatments of the present experiment taking into account stratification by treatments in the previous experiment on these same experimental units, the response model equation is extended to be:

$$Y_{hijk} = \mu + \rho_h + \gamma_i + \tau_j + \pi_k + \epsilon_{hijk} \quad , \tag{2}$$

where π_k is the stratification effect of the previous treatment k and ϵ_{hijk} has the same definition as ϵ_{hij} in Equation (1). For a Latin square of order r, the normal equations in vector form are:

$$\left.\begin{array}{ll} \underset{\sim}{R} = r\mu\underset{\sim}{1} + r\underset{\sim}{\rho} & \underset{\sim}{T} = r\mu\underset{\sim}{1} + r\underset{\sim}{\tau} \\ \underset{\sim}{C} = r\mu\underset{\sim}{1} + r\underset{\sim}{\gamma} & \underset{\sim}{T}_p = r\mu\underset{\sim}{1} + r\underset{\sim}{\pi} \end{array}\right\} \qquad (3)$$

where $\underset{\sim}{R}$, $\underset{\sim}{C}$, $\underset{\sim}{T}$, and $\underset{\sim}{T}_p$ represent totals for rows, columns, treatments in the present experiment, and treatment arrangement for the previous experiment, respectively, $\underset{\sim}{1}$ is an n × 1 vector of ones, $\underset{\sim}{\rho}$ is a vector of ρ_h, $\underset{\sim}{\gamma}$ is a vector of γ_i, $\underset{\sim}{\tau}$ is a vector of τ_j, and $\underset{\sim}{\pi}$ is a vector of π_k, under the usual constraints $\underset{\sim}{1}'\underset{\sim}{\rho} = \underset{\sim}{1}'\underset{\sim}{\gamma} = \underset{\sim}{1}'\underset{\sim}{\tau} = \underset{\sim}{1}'\underset{\sim}{\pi} = 0$. If the experiment design were of the O:OT:OTT type [see Preece (1966)], the normal equations would take the form:

$$\left.\begin{array}{l} \underset{\sim}{R} = c(\mu\underset{\sim}{1} + \underset{\sim}{\rho}) \\ \underset{\sim}{C} = r\mu\underset{\sim}{1} + r\underset{\sim}{\gamma} + N_{1c}'\underset{\sim}{\tau} + N_{2c}'\underset{\sim}{\pi} \\ \underset{\sim}{T} = r\mu\underset{\sim}{1} + N_{2c}\underset{\sim}{\gamma} + N_{21}\underset{\sim}{\pi} + r\underset{\sim}{\tau} \\ \underset{\sim}{T}_p = r\mu\underset{\sim}{1} + N_{2c}\underset{\sim}{\gamma} + r\underset{\sim}{\pi} + N_{21}'\underset{\sim}{\tau} \\ \text{Grand total} = rc\mu \text{ , for r rows and c columns} \end{array}\right\} \qquad (4)$$

where N_{1c} is the past treatment-column design matrix composed of zeros and ones, N_{2c} is the present treatment-column design matrix, and N_{21} is the present treatment-past treatment design matrix. It is a straightforward extension to add terms to Equation (2) to consider p successive experiments in row by column designs on the same set of experimental units.

Many other response model equations are possible for the many types of experiments that are conducted. One such equation that would have application in certain types of experiments would be to include differential regressions for treatments in the past experiment. Equation (2) could be altered in the same manner that Cox (1958) altered the classical Latin square response model, Equation (1), to include differential regressions within columns in place of row effects. Still another set of response model equations would be to add terms for non-additivity similar to those discussed by J. W. Tukey and others.

So far, discussion has been confined to consideration of successive experiments with only responses on the last experiment in the sequence. Consider now the situation wherein the p experiments are conducted simultaneously, e.g., each experiment is a Latin square design and the p designs are mutually orthogonal Latin squares. In place of one response equation as, for example, an extension of Equation (3) for (p-1) previous experiments, there are now p responses for each experimental unit. To illustrate, suppose that one has r grocery stores and r time periods in which to conduct r marketing procedures on each of p vegetables such as carrots, celery, potatoes, etc., and that the p experiments are to be carried out simultaneously in the stores and time periods; pr^2 responses are obtained. If possible, one would select p Latin square designs of order r which are mutually pairwise orthogonal. Note that if p = r - 1, then no degrees of freedom would be left for the error mean square for responses from each experiment. Also, note that if the design is not a Latin square, then one of the Potthoff (1962b), Hedayat _et al._ (1972), etc., designs may be suitable.

Instead of considering univariate analyses for each experiment, it is recommended that a p variate multivariate analysis of variance be performed on the prc responses from the r-row by c-column design. Also, it is recommended that univariate analyses of variance using equations similar to Equation (3) also be performed. A store manager could be interested in a linear combination of products sold in addition to considering sales for each product separately.

A similar situation arises in intercropping experiments as one obtains p responses for the p intercrops in a mixture. A grower is not solely interested in what each crop does individually, but is additionally interested in what the combination of p crops yields in terms of profit, protein, calories, or other combination of responses. Some form of analysis utilizing all prc responses is required.

To further illustrate the need for new statistical theory and methods, consider the following situation wherein a plant breeder wishes to screen maize lines and bean lines and to select lines which generally perform well in a mixture. A large number of lines of each crop, say m maize lines and b bean lines, are generally available and seed, space, and personnel may be limiting such that even if it is possible to use all mb mixtures, it is impossible to replicate the experiment at a single location. If mb experimental units can be handled at a given location, then the experimenter may use an m-row by b-column design. Further, if m = b, then the m maize lines may be arranged in a Latin square and the m bean lines arranged in a Latin square orthogonal to the first. To illustrate, let m = b = 5, then the 25 intercrops may be arranged as:

	columns				
rows	1	2	3	4	5
1	1a	2b	3c	4d	5e
2	2c	3d	4e	5a	1b
3	3e	4a	5b	1c	2d
4	4b	5c	1d	2e	3a
5	5d	1e	2a	3b	4c

where the maize lines are numbered 1, 2, 3, 4, and 5, and the bean lines are numbered a, b, c, d, and e. Equation (2) could be used for maize yields and for bean yields separately, and one could also perform a bivariate analysis on the bean and maize yields simultaneously using Equation (1). If, in addition, the five maize lines were grown without beans (sole crop) and likewise for the five bean lines, then there would be 35 combinations which one could design in a 7-row by 5-column design if there were two-way variation in the experimental material that needed to be controlled. Multivariate analyses would need to be altered to take into account the sole crop responses. Additional problems are encountered in multivariate analyses for intercropping experiments.

REFERENCES

Afsarinejad, K. and A. Hedayat (1975). Some contributions to the theory of multistage Youden design. Annals of Statist. 3, 707-711.

Anderson, D. A. (1972). Designs with partial factorial balance. Annals Math. Statist. 43, 1333-1341.

Anderson, D. A. and W. T. Federer (1976). Multidimensional balanced designs. Comm. In Statist. A 5, 1193-1204.

Bose, R. C. and J. N. Srivastava (1964). Multidimensional partially balanced designs and their analyses, with applications to partially balanced factorial fractions. Sankhyā A 26, 145-168.

Cox, C. P. (1958). The analysis of Latin square designs with individual curvatures in one direction. J. Roy. Statist. Soc. B 20, 193-204.

Federer, W. T. (1972). Construction of classes of experimental designs using transversals in Latin squares and Hedayat's sum composition method. Statistical Papers in Honor of George W. Snedecor (editor, T. A. Bancroft), Iowa State University Press, Ames, Iowa, pp. 91-114.

Freeman, G. H. (1958). Families of designs for two successive experiments. Annals Math. Statist. 29, 1063-1078.

Freeman, G. H. and J. N. R. Jeffers (1962). Estimation of means and standard errors in the analysis of non-orthogonal experiments by electronic computer. J. Roy. Statist. Soc. B 24, 435-446.

Hedayat, A. and D. Raghavarao (1975). 3-way BIB designs. J. Comb. Theory A 18, 207-209.

Hedayat, A., E. Seiden, and W. T. Federer (1972). Some families of designs for multistage experiments: Mutually balanced Youden designs when the number of treatments is prime power or twin primes. I. Annals Math. Statist. 43, 1517-1527.

Potthoff, R. F. (1962a). Three-factor additive designs more general than the Latin square. Technometrics 4, 187-208.

Potthoff, R. F. (1962b). Four-factor additive designs more general than the Greco-Latin square. Technometrics 4, 361-366.

Preece, D. A. (1966). Some row and column designs for two sets of treatments. Biometrics 22, 1-25.

Singh, N. P., A. K. Nigam, and M. Singh (1981). Analysis and characterization of row-column experiments involving two non-interacting sets of treatments. (Unpublished paper.)

Srivastava, J. N. and D. A. Anderson (1970). Some basic properties of multidimensional partially balanced designs. Annals Math. Statist. 41, 1438-1445.

Srivastava, J. N. and D. A. Anderson (1971). Factorial association schemes with applications to the construction of multidimensional partially balanced designs. Annals Math. Statist. 42, 1167-1181.

SUMMARY

Statistical analyses and response model equations are discussed for the p^{th} of p successive experiments and for p simultaneous experiments carried out on the same experimental units.

RÉSUMÉ

Analyses statistiques et equations models des responses sont discuté pour le p^{th} des p expériences successives et pour p expériences simultanés qui sont conduiu sur les memes unités experimentales.