ROTATION SAMPLING DESIGNS

E.E. Gbur and R.L. Sielken, Jr.
Texas A&M University

An important recent application of space technology to agriculture is in the area of crop acreage and yield prediction using satellite based data. In this application, a small area of land called a segment is "photographed" periodically during the growing season. At specific time points during the season, an estimate of the at-harvest acreage of a particular crop in the sampled segment is made. These segment estimates are, in turn, used to estimate a mean crop acreage per segment in a larger area (e.g., a state). At-harvest estimates are required for the larger area for several consecutive years.

In general, we have a population of units (segments) which is to be sampled for T consecutive periods (years). In any proposed sampling design, the units to be sampled can change from period to period but not at time points within the period. In addition, there is a positive correlation between the responses from a unit in consecutive time periods which can be utilized to reduce the standard errors of the estimators of the end of period means. The problem is to determine a T period sampling scheme which is optimal in some sense. Rotation designs are a natural starting point in the search for a solution of this problem.

2. An Extended Class of Rotation Designs

Rotation designs have been used for many years in sample surveys involving a wide variety of target populations; c.f., Jessen (1942), Hansen et al. (1955), and Cochran (1977). A prototype of a rotation design is shown in Figure 1 where an x indicates that the unit is to be sampled in that period. All such designs have a "smooth" rotation pattern; i.e., a unit is sampled for several consecutive periods and then removed from the survey, perhaps to be reintroduced in some later period. The use of smooth rotation patterns represents an attempt to reduce administrative and respondent burdens and response bias introduced by sampling the same unit in consecutive periods.

In the crop acreage estimation problem, there is no difficulty with administrative and respondent burdens and response bias. Rotation designs are employed primarily so that the positive correlation of the responses from the same unit in consecutive time periods can be used to reduce the standard errors of the estimators. Since there is very little administrative and respondent burden in the sampling, there is no need to restrict consideration to smooth rotation patterns. If we are willing to consider "non-smooth" patterns and, in the extremes, to allow for sampling the same set of units in consecutive periods and for sampling completely different sets of units in consecutive periods, then any sampling design in which the same number of units are sampled in each period can be considered as a rotation design. Although this extended class of rotation designs is much larger than the class usually considered, each design retains the essential characteristic of a rotation design; viz., the set of units to be sampled in each period can be partitioned into a subset (perhaps empty) of units which were sampled in the preceeding period and a subset (perhaps empty) which consists of new units and/or units which have been sampled in a period prior to the preceeding period. An example of one such non-smooth rotation pattern is shown in Figure 2.

Figure 1

Unit	Period 1	2	3	4	5
1	x			x	x
2	x				x
3	x	x			x
4	x	x			
5		x	x		
6		x	x		
7			x	x	
8			x	x	
9				x	x

Figure 2

Unit	Period 1	2	3	4
1	x		x	x
2	x			x
3	x			
4	x	x	x	x
5	x	x	x	
6		x	x	
7		x		x
8		x		
9			x	x

The requirement of within period estimation of the end of period mean and the possible selection of a non-smooth rotation design make it necessary to replace the usual composite estimator of the period means by an estimator obtained from a more flexible model. We shall use an analysis of variance model to provide a framework for the estimation of the end of period means and as a basis for the definition of design optimality criteria.

3. An Analysis of Variance Model

A simple analysis of variance model which can be used to describe the response y in a rotation sampling design is given by

$$y_{ts\ell} = \alpha_t + \delta_\ell + b_s + e_{ts\ell} \quad . \qquad (1)$$

$t = 1, \ldots, T \;, \; s = 1, \ldots, S, \; \ell = 1, \ldots, L,$

where

α_t is the mean response at the end of the time period t,

δ_ℓ is the bias associated with the estimation of α_t at the ℓ^{th} time within the period(δ_L=0),

b_s is the departure of the s^{th} sampled unit from the population mean, and

$e_{ts\ell}$ is the error.

We shall assume that $b_1, \ldots, b_S$ are iid $(0, \sigma_b^2)$, $e_{111}, \ldots, e_{TSL}$ are iid $(0, \sigma_e^2)$, and that b_s and and $e_{ts\ell}$ are **independent.**

Model (1) can be expressed in matrix form (using the obvious notation) as

$$Y = X\beta + Ub + e, \qquad (2)$$

where $\beta' = [\alpha', \delta'] = [\alpha_1, \ldots, \alpha_T, \delta_1, \ldots, \delta_{L-1}]'$. Using (2), the covariance matrix of the least squares estimator of β is given by

$$W\,\sigma_e^2 = (X'V^{-1}X)^{-1}\,\sigma_e^2, \qquad (3)$$

where $V = I + \gamma\, U\, U'$ and $\gamma = \sigma_b^2 \,/\, \sigma_e^2$. Let

$$W = \begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix}$$

be the partition of W corresponding to the partition of the parameter vector $\beta' = [\alpha', \delta']$. Then $W_{11}\sigma_e^2$ is the covariance matrix of $\hat{\alpha}$.

4. Optimality Criteria

In the notation of model (1), the parameters of primary interest are $\alpha_1, \ldots, \alpha_T$, the mean responses for the T periods. Therefore, any reasonable measure of design optimality should be based on the submatrix W_{11} associated with α . Two commonly used experimental design optimality criteria which are relevant in this context are D-optimality and A-optimality. Recall that a design will be A-optimal if it mimizes $tr(W_{11}) = \Sigma \, var(\hat{\alpha}_t) \,/\, \sigma_e^2$ and will be D-optimal if it minimizes $det(W_{11})$ which is proportional to the generalized variance of $\hat{\alpha}$ (c.f., Silvey 1980).

The repeated sampling of a population for several time periods allows for estimation of changes in the population as well as estimation of the mean response at the end of each period. For applications in which the estimation of change over time is of primary importance, D-optimality would be an appropriate design selection criterion since estimation of change involves linear combinations of the $\hat{\alpha}_t$. On the other hand, A-optimality would be appropriate when the emphasis is on the estimation of the mean response for each period since it requires minimizing the average of the variances of $\hat{\alpha}_1, \ldots, \hat{\alpha}_T$.

For given values of the number of periods T, the number of within period estimation times L, and the number of units sampled in each period R, each rotation design will have the same fixed effects design matrix X in (2). Thus, the value of the optimality measure of a rotation design for (2) is determined by the design matrix U, which describes the sampling pattern of the units.

5. Optimal Two and Three Period Designs

A study was undertaken to characterize A-optimal and D-optimal rotation designs for the analysis of variance model (1). Analytical results were obtained for two period designs and numerical results were obtained for three period designs having selected parameter combinations. A by-product of these calculations was the numerical values of the optimality measures for the corresponding two period designs. Numerical results were obtained by enumeration of all possible designs for T = 2, 3 periods, L = 2, 3 estimation points within each period, R = 2, 3, 4, 5 units sampled per period, and variance component ratios γ = 0.25, 0.5, 1.0, 2.0, 4.0.

For two and three period designs and each of the above parameter configurations in the case of T=3, the D-optimal design for the analysis of variance model (1) is the "no-rotation" design; i.e., the design in which the same units are sampled in each period. This leads to the following

Conjecture: For the model (1), the no-rotation design is D-optimal for T periods, T = 2, 3, ...

Note that the results of this section and the above conjecture are consistent with the recommendation of D-optimality as a design selection criterion when changes in the population over time are of primary interest.

For A-optimality the results are not as simple. Several examples will suffice to illustrate the nature of the results. For T=2 periods and R=4 units sampled per period, the A-optimal design is shown in Figure 3 for $\gamma \leq 1$ and in Figure 4 for $\gamma > 1$. For T=3 and R=4, the A-optimal design for $\gamma \leq 1$ is shown in Figure 5 and for $\gamma > 1$ in Figure 6.

Figure 3

Unit	Period 1	Period 2
1	x	
2	x	
3	x	x
4	x	x
5		x
6		x

Figure 4

Unit	Period 1	Period 2
1	x	
2	x	
3	x	
4	x	x
5		x
6		x
7		x

Figure 5

Unit	Period 1	Period 2	Period 3
1	x		
2	x		
3	x	x	x
4	x	x	x
5		x	
6		x	
7			x
8			x

Figure 6

Unit	Period 1	Period 2	Period 3
1	x		x
2	x		
3	x		
4	x	x	
5		x	
6		x	
7		x	x
8			x
9			x

From these examples and the remaining numerical work, the following points were observed.

(i) As γ increases there is a tendency to sample more distinct units.

(ii) For some parameter configurations (c.f., Figure 5), the extended class of rotation designs provides designs which are better than the usual smooth rotation designs.

(iii) The sampling pattern in the first two periods of A-optimal three period designs coincides with the corresponding A-optimal two period designs (Compare Figure 5 with Figure 3 and Figure 6 with Figure 4.).

This last point leads to the following

Conjecture: For the model (1) and fixed γ, the A-optimal T period design is obtained from the A-optimal T-1 period design.

If the above conjecture is true, then A-optimal designs can be constructed sequentially, thereby eliminating the need for complete enumeration of T-period designs in the search for optimal designs.

It should be noted that the value of γ=1 in the above examples is not always a critical value for γ ; in fact, for some configurations of the parameters T, R, and L, the range of γ must be partitioned into more than two intervals in order to indicate the A-optimal designs.

6. Concluding Remarks

We have investigated an extension of the usual concept of a rotation design which can be applied to the repeated sampling of populations in which administrative and respondent burden and response bias are not major concerns. An analysis of variance model has been used as a basis for parameter estimation and optimal design definition and selection.

The following areas need further research.

(1) The proof or disproof of the conjectures concerning A-optimal and D-optimal designs stated in Section 5.

(2) An investigation of the extent to which D-optimal and A-optimal designs depend on the particular analysis of variance model used. (In light of the comments in Section 4, the addition of fixed effects to model (1) may not change the optimal designs.)

(3) General theoretical results for optimal experimental designs for mixed models.

References

Cochran, W.G. (1977). *Sampling Techniques*, third edition. John Wiley and Sons, New York.

Hansen, M.H., W.H. Hurwitz, H. Nisselson, and J. Steinberg (1955). The redesign of the Census Current Population Survey. *J. Amer. Stat. Assoc.* 50, 701-709.

Jessen, R.J. (1942). Statistical investigation of a sample survey for obtaining farm facts. *Iowa Ag. Exper. Station Research Bull.* 304, 54-59.

Silvey, S.D. (1980). *Optimal Design*. Chapman and Hall, London.