THE MODEL BASED (PREDICTION) APPROACH TO FINITE POPULATION SAMPLING THEORY

Richard M. Royall, Department of Biostatistics,
The Johns Hopkins University

Introduction

Estimating a finite population mean from a sample is equivalent to predicting the mean of the non-sample values. This view, that finite population inference problems are actually prediction problems, leads naturally to a theory in which prediction models, not sample selection probabilities, are central. This paper is an informal survey of that theory.

The first section describes the model-based approach and attempts to make clear how and why it differs from the prevailing (randomization-based) theory. This section is built around a simple example, which is used to illustrate various facets of the approach. The second section addresses the question "What has the model-based approach accomplished?" This is not an attempt to catalog significant contributions to model-based sampling theory, but to describe and interpret the general kinds of developments that have occurred. Finally, the third section consists of some brief observations on current research.

What Is Model-Based Sampling Theory?

Model-based sampling theory begins by recognizing that problems of estimating finite population characteristics are naturally expressed as prediction problems (Kalbfleisch and Sprott, 1969; Geisser, 1986, p. 163). For example, Figure 1 shows the data for a sample of $n = 32$ hospitals. For each sample hospital we know the number of beds (x) and we have observed the number of patients discharged (y) during a given month. If we must estimate how many patients were discharged from another hospital, say one with x = 400 beds, we might fit the dotted line in Figure 1. The slope of that line, the ratio of total sample discharges to total sample beds, shows that in sample hospitals there were 3.1 patients discharged per bed. Thus we might estimate that there were about 3.1 × 400 = 1240 patients discharged from the other hospital. More generally, to estimate how many patients were discharged from a set r of non-sample hospitals having a total of $\Sigma_r x_i$ beds, we might use 3.1 $\Sigma_r x_i$. Then to estimate the patient total for the entire population composed of the thirty-two hospitals in the sample s as well as those in r, we would simply add the observed total for the thirty-two sample hospitals, $\Sigma_s y_i$ to our estimate for those not observed, 3.1 $\Sigma_r x_i$.

Clearly this estimate of the population total is reasonable only if it is reasonable to assume that the hospitals in r are "like" the ones in s: if the sample hospitals are in the eastern United States while the r-hospitals are in France, then this estimate is certainly questionable. How can we formalize this reasoning,

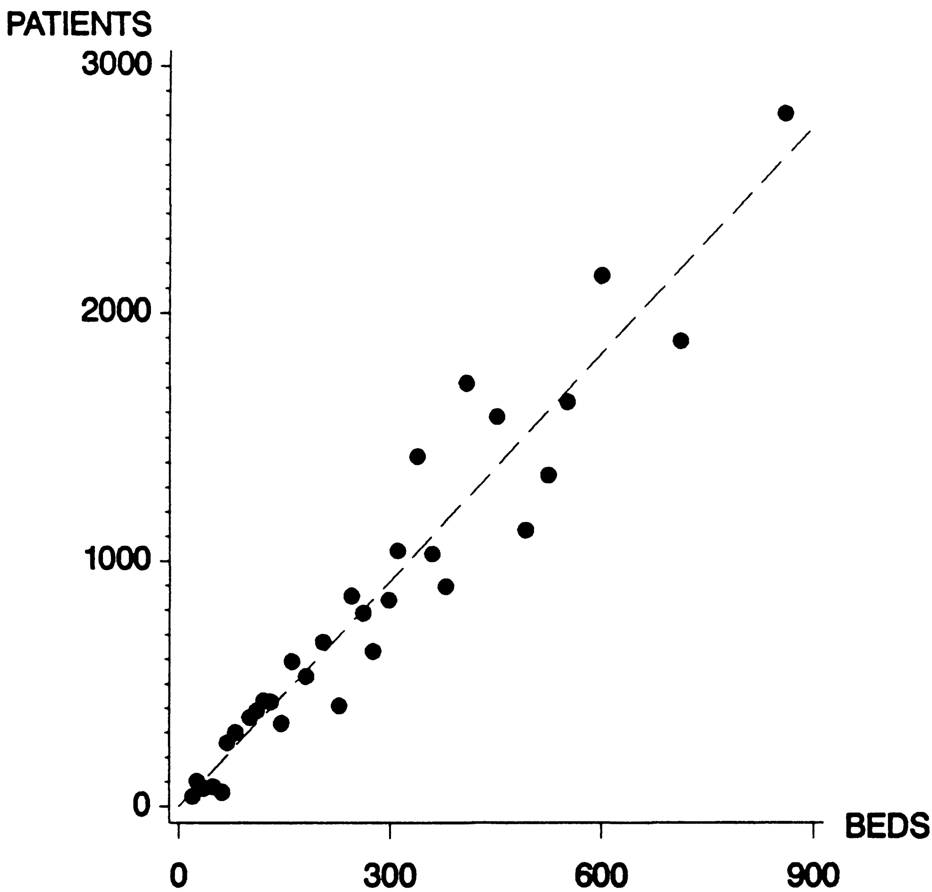

Figure 1. Number of patients discharged and number of beds in 32 short-stay U.S. Hospitals, June 1968.

exposing and clarifying the underlying assumptions and explaining when the estimate is a good one, and when it is not?

A natural way to express the assumptions is through a probability model for the numbers of patients discharged from each of the hospitals, both those in the sample s and those in r. The model represents these numbers, $y_1, y_2, \ldots, y_N$, as realized values of independent random variables $Y_1, Y_2, \ldots, Y_N$, where N is the total number of hospitals.

Model M. $$E(Y_i) = \beta x_i \,, \; var(Y_i) = \sigma^2 x_i \,,$$

$$cov(Y_i, Y_j) = 0 \,, \; i \neq j.$$

Under model M the y's will tend to be roughly proportional to the x's, with more variability about the expected value, βx, in large hospitals than in small ones. This model is consistent with the thirty-two observations shown in Figure 1, but is not unique in this respect. We must be alert to the possibility that other models might be more appropriate. Nevertheless, analysis under model M can explain much of what our informal look at the problem has already suggested.

First we note that the model represents a link between the two sets of numbers, $\{y_i;\ i \in s\}$ and $\{y_i;\ i \in r\}$, that enables us to learn about the second set by studying the first. Now the problem of estimating $T = \Sigma_s y_i + \Sigma_r y_i$ is evidently equivalent to the problem of predicting the value, $\Sigma_r y_i$, of the random variable, $\Sigma_r Y_i$. The estimate that we derived intuitively, $\hat{T}_R = \Sigma_s y + b\Sigma_r x$, where $b = \Sigma_s y / \Sigma_s x = 3.1$, is the best linear unbiased (BLU) estimator of T under model M, because $b\Sigma_r x$ is the BLU predictor of $\Sigma_r y$. Note that this is actually the popular *ratio estimator*, $(\Sigma_s y/\Sigma_s x)\Sigma_1^N x$. The reason that we would not use this estimate if the non-sample hospitals were in France is that we would be unwilling to apply the same model (with the same value of the expected number of patients per bed, β) to both the sample and non-sample facilities. Note that this conclusion would apply even if we had decided at random which ones to exclude from the sample and had chosen the overseas hospitals by bad luck. Our reluctance to use the sample ratio, 3.1 discharges per bed, to estimate for those not in the sample arises from unwillingness to make the assumptions expressed in the model, not from the process used to choose which hospitals to put in the sample and which ones to leave out.

The model also provides guidance in sampling. For a given split of the population into sample s and non-sample r hospitals, the estimation (prediction) error in the ratio estimate of T is $\hat{T}_R - T = b\Sigma_r x - \Sigma_r y$. Its expected value under M is zero and its variance is $var(\hat{T}_R - T) = (N/f)(1 - f)(\bar{x}\bar{x}_r/\bar{x}_s)\sigma^2$, where $\bar{x}$ is the population mean, $\bar{x}_s$ and $\bar{x}_r$ are sample and non-sample means, n is the sample size, and f is the sampling fraction, n/N. This variance decreases as $\bar{x}_s$ increases, so it is minimized when s consists of the n largest hospitals. Equally important, it is *maximized* when s consists of the n smallest. Although we will find that robustness considerations imply that it is often unwise to choose the largest units for s, the smallest units represent the worst possible sample under a wide variety of conditions.

Another role of the model is to validate large sample confidence intervals: if the population is enlarged, so that both sets of hospitals, s and r, grow in a stable way, then $(\hat{T}_R - T)/[var(\hat{T}_R - T)]^{1/2}$ converges in distribution to the standard normal. Because $v = \Sigma_s(y_i - bx_i)^2/nx_i$ is a consistent estimator of σ^2, an approximate confidence interval for T is given by $\hat{T}_R \pm z[(N/f)(1 - f)(\bar{x}\bar{x}_r/\bar{x}_s)v]^{1/2}$ when n and $N - n$ are both large (Royall and Cumberland, 1978).

Although the ratio estimator is BLU under model M, other estimators might also be considered, because of robustness, simplicity, or other criteria. Analysis under M remains critical: the estimator we choose must at least have reasonable properties under this model if it is to be appropriate for estimating the total number of patients discharged. For example, the simple expansion estimator $\hat{T}_E = \Sigma_s y + (N - n)\bar{y}_s = N\bar{y}_s$, which estimates the non-sample mean $\bar{y}_r$ by the sample mean $\bar{y}_s$, would be inappropriate here in any sample s of hospitals whose mean size $\bar{x}_s$ is not very close to the population mean $\bar{x}$. This is because the estimator is biased under M:

$$E(\hat{T}_E - T) = N\beta(\bar{x}_s - \bar{x}).$$

This expression shows that the expansion estimator will tend to underestimate T if the average size of sample hospitals, $\bar{x}_s$, is smaller than the population average, $\bar{x}$, and to overestimate when $\bar{x}_s$ is larger. By contrast, the linear *regression estimator* $\hat{T}_{RG} = N[\bar{y}_s + b_1(\bar{x} - \bar{x}_s)]$ where $b_1 = \Sigma_s(x_i - \bar{x}_s)y_i/\Sigma_s(x_i - \bar{x}_s)^2$, is, like the ratio estimator, unbiased under M in any sample s: $E(\hat{T}_{RG} - T) = 0$.

Thus we can evaluate estimators in terms of bias and variance under M, study how these properties are affected by characteristics of the sample, like $\bar{x}_s$, and find approximating distributions for setting confidence intervals. If M were known to be true, then this body of theoretical results might be satisfactory for guiding us in selecting a sample and making inferences from observations.

But M is not true. A sufficiently large sample of hospitals would surely reveal that M, like any mathematical model, is at best an approximation. Although we have adopted M as a working model for this population, we remain skeptical, aware that theoretical results derived under M have practical value only if they are robust in the face of plausible departures from this model.

Robustness to departures from M can be studied by changing the model. For example if we generalize by relaxing the restriction that $var(Y_i)$ be proportional to x_i, we see that the ratio estimator remains unbiased and consistent. But the large-sample confidence interval is no longer valid, because the estimator of $var(\hat{T}_R - T)$ is no longer consistent. Fortunately there are variance estimators that are consistent under the generalized model, providing robust large sample confidence intervals (Royall and Cumberland, 1981a).

To study the effects of errors in the working model's regression function, $E(Y_i) = \beta x_i$, we might consider a sequence of generalizations, first adding a constant term, then a quadratic, etc. Each term added to the regression model introduces bias in the ratio estimator. For example, if $E(Y_i) = \alpha + \beta x_i$, then the bias is $E(\hat{T}_R - T) = N\alpha(\bar{x} - \bar{x}_s)/\bar{x}_s$. Protection against this bias can be achieved by choosing a sample that is balanced on x: $\bar{x}_s = \bar{x}$. Protection against a quadratic term's bias can be achieved by balancing on x^2 as well: $\Sigma_s x^2/n = \Sigma_1^N x^2/N$. And balancing on other powers of x protects against bias caused by the presence of corresponding terms in the true regression function (Royall and Herson, 1973).

Thus in order to protect against the bias that can be caused by departure from the working model's regression function, we might choose a balanced sample in preference to the optimal (minimum variance) sample composed of the n largest hospitals. The same type of trade-off, efficiency for robustness, might apply to other aspects of the problem as well, such as the choice of an estimator for T and an estimator of the error variance, $var(\hat{T} - T)$. The model-based theory does not assume that a particular model is correct and proceed blindly under that assumption: alternative models are used to examine the key practical issue of robustness.

The main features of model-based sampling theory have appeared in our look at the hospital discharge population:

(i) representing the unknown numbers of interest as realized values of observable random variables,

(ii) recognizing that estimating a population value from an observed sample is a prediction problem, and

(iii) using probability models as the formal basis for prediction and for determining the primary statistical properties of samples and predictors.

The use of probability models as the basis for inference from sample to population, (iii), is the critical feature distinguishing the model-based theory from the prevailing one. Although a random sampling plan may be used for choosing which hospitals will be observed (and for which hospitals the number of discharges must be estimated), the basic inference framework is the probability model, not the random sampling plan. By contrast, the prevailing theory of finite population sampling reverses the priority, avoiding probability models in favor of distributions created by random sampling plans as the formal basis for inference.

Conventional theory defines bias, for example, with respect to the probability distribution generated by the random sampling plan. Thus the expansion estimator, $N\bar{y}_s$, is an unbiased estimator of T if every set of n hospitals is given the same probability of being selected as the sample. But the same estimator is biased if the sample is chosen by another selection scheme. The bias in $N\bar{y}_s$ is determined, not by relationships between the hospitals in s and those not in s, but by the probabilities with which other samples might have been selected. Recall that the model-based theory under model M said that this estimator has a positive bias if the sample consists of hospitals that are larger, on average, than those not in the sample, a negative bias if the sample hospitals are smaller, and no bias only if the sample is balanced on size. Although both definitions of bias are mathematically valid, for the purpose of inference from a given sample of hospitals the model-based one is clearly relevant and informative while the conventional one is misleading.

Conventional theory defines variance also as an average value over all possible samples. Again this is in contrast to model-based theory, which, because it defines the variance for a specific sample with respect to a prediction model for the unobserved variates, conditions on the characteristics of the sample actually observed as well as on those of the non-sample units whose values must be predicted.

Model-based theory, by insisting that inferences should be based on prediction models, not on probability distributions created by randomly choosing which units to observe, does not preclude the use of random sampling plans. It is not the presence or absence, but the *role*, of random sampling that distinguishes model-based from conventional finite population sampling theory. The terminology invites misunderstanding on this point: because the word *sampling* in the name suggests only the design phase–choosing samples – *model-based sampling theory* is easily misinterpreted as signifying a theory for choosing samples using models, whereas the critical feature is the use of models in inference.

There are other model-based approaches. The one sketched above is developed in terms of bias, variance, and approximate normality under linear models. Alternatives include approaches based on fiducial (Kalbfleisch and Sprott, 1969), likelihood (Royall, 1976b), and Bayesian prediction models. Ericson (1988) has recently surveyed the Bayesian theory. We will focus on the linear prediction approach, because it has seen the most vigorous development, empirical testing, and critical discussion.

What Has the Model-Based Approach Accomplished?

The model-based approach has bridged the gap between finite population problems and the rest of statistics. Before the model-based approach, finite population sampling was an eccentric realm where many of the basic concepts and tools of statistics were curiously inapplicable. Statisticians skilled in designing experiments and in applying linear models to make inferences from experimental and observational data found that finite population problems were apparently beyond the scope of their techniques. Although there were some familiar-looking formulas, such as the linear regression estimator shown in Section 1, these statistics lacked the familiar rationale and properties. Not only was the linear regression estimator biased (and therefore certainly not a BLU estimator) it was not even linear, because the random choice of observation points turned the denominator of the estimated slope into a random variable. To make matters appear utterly hopeless to one interested in statistical theory, Godambe (1955) proved that the BLU estimator for a finite population average does not exist and furthermore (1966) showed that the likelihood function generated by a random sample from a finite population is, for all practical purposes, totally uninformative. Attempts to fill the theoretical vacuum were uniformly unsuccessful (e.g., Godambe, 1966; Hanurav, 1968; Hartley and Rao, 1969; Royall, 1969).

The prediction approach revealed that the problem was rooted, not in esoteric aspects of finite population problems that invalidated the methods applicable to the rest of statistics, but in the attachment of those who worked in finite population sampling theory to a restrictive statistical doctrine based on a dubious principle. This is the Randomization Principle, proclaimed and then renounced by Fisher (1935 §21, 1960 §21.1), which asserts that the only probability distributions appropriate for statistical inference are those created by deliberate randomization.

A particularly clear statement of the Randomization Principle in the finite population setting was given by Stuart (1962):

> If you feel at times that the statistician, in his insistence on random sampling methods, is merely talking himself into a job, you should chasten yourself with the reflection that in the absence of random sampling, the whole apparatus of inference from sample to population falls to the ground, leaving the sampler without a scientific basis for the inferences which he wishes to make.

This Principle has had its champions in experimental statistics (Kempthorne, 1955), where it underlies the curious claim that no valid statistical inferences are possible in observational studies. (This last point is discussed in Royall (1976a), with references.) But in that area the Principle faced strong opposition, from "Student" (1937) and Neyman and Pearson (1937, p. 384) for example, and it never held sway. The Principle's unchallenged domination of finite population theory is thus curious; it is doubly curious because this domination is credited to Neyman (1934) (ref. Smith, 1976; O'Muircheartaigh and Wong, 1981).

The theoretical vacuum in finite population sampling was an inevitable consequence of the Randomization Principle. If the Principle is applied in other areas of statistics, entirely analogous results follow: if all inferences must be based on the probability distribution created by artificial randomization, so that all variables that have not been *made* random by the experimenter's actions must be treated as fixed (possibly unknown) constants, then the likelihood function for randomized comparative experiments is just like the finite population likelihood function—uninformative (Cornfield, 1966). Likewise, if inferences about regression coefficients must be based on the distribution created by using deliberate randomization to select material for observation, then the

Gauss-Markov theorem can justify least-squares estimators only in those cases where at each value of the regressor x the average response $\bar{y}$ over all units actually available for observation falls precisely on the regression line: thus the Principle would imply the non-existence of BLU estimators in essentially all real-world applications, certainly including all problems where each potential sample unit is characterized by a unique vector of regressor values.

Deliberate randomization is a valuable statistical tool (for protecting against unconscious bias, for example). Few statisticians would deny this. But the Randomization Principle claims much more: the only biases, standard errors, significance levels, and confidence coefficients acceptable for inference are those defined and justified in terms of deliberate randomization. The model-based prediction approach to finite population sampling consists of nothing more radical than taking the concepts, techniques, and tools that form the familiar core of applied statistics and using them where previously they had been precluded by acceptance of the Randomization Principle. This has had several important effects:

(i) providing techniques for systematic study of some finite population sampling problems that the randomization approach is ill-equipped to address,

(ii) bringing an alternative theoretical perspective to finite population methods that have been analyzed previously in terms of randomization theory,

(iii) revitalizing conventional randomization-based finite population theory,

(iv) providing a new context for studying the model-based methods that are standard outside of finite populations, and

(v) testing general statistical concepts and principles in a new setting.

Examples in the first category – problems that are difficult to address in terms of deliberate randomization alone – include non-response (Särndal, 1981; Little, 1982; Chiu and Sedransk, 1986), small area estimation (Laake, 1979; Holt, Smith, and Tomberlin, 1979; Royall, 1979), and inference from non-random samples (Smith, 1983; Kott, 1984). This is not to say that there was no methodology for these problems before the model-based approach came along. There were various techniques that had been derived intuitively and developed by trial and error. What models did was to provide a theoretical framework for studying the methods (such as *synthetic estimates* for small area estimation) and for describing the implicit assumptions behind them, as well as for suggesting alternatives.

Of greater theoretical interest are activities of the second type – applications of the model-based approach to problems where the old randomization approach had already generated a body of results. In some cases the prediction approach simply provided a new explanation and interpretation of conclusions that had been reached by conventional sampling theory. An example is the finding that the Yates-Grundy estimator is better than the Horvitz-Thompson estimator for the variance of the mean-of-ratios statistics, $N\bar{x}\Sigma_s(y_i/x_i)/n$, in samples chosen by *a probability-proportional-to-x* sampling plan (Cumberland and Royall, 1981).

In other cases the prediction approach revealed a clear preference for one of two procedures where the randomization approach had been noncommittal. One example is in post-stratification, where some followers of randomization theory had chosen to condition the variance on the actual stratum sample sizes, while others had chosen to use the unconditional variance. The deadlock was described by Holt and Smith (1979), whose prediction theory analysis made clear the need to condition.

Variance estimation for the ratio estimator provides another example of the activities in category (ii). Randomization theory had been unable to choose between two proposed variance estimators, yet model-based analyses revealed that the more popular of the two has a severe conditional bias. This bias is positive in some samples, leading to overly conservative confidence intervals, and negative in others, producing undercoverage. The second statistic is free of these biases. It is worth noting that empirical comparisons of these two variance estimators had also been inconclusive, because the investigators, guided by randomization theory, had averaged the results over all of the values of the conditioning variable, and had thereby averaged out the biases (Rao and Rao, 1971). Empirical studies guided by prediction theory exposed the biases clearly enough to inspire efforts to accommodate the conditional results within randomization theory (Fuller, 1981; Robinson, 1987).

The model-based approach has stimulated conventional sampling theory in other ways as well. For example, model-based results on variance estimation (Royall and Cumberland, 1981a) have inspired significant developments in conventional theory (Wu and Deng, 1983; Deng and Wu, 1987). At a more general level, the model-based approach has forced those who object to it to examine and articulate the reasons for their opposition (e.g., Hansen, Madow, and Tepping, 1983) and to extend and adapt the conventional theory to accommodate those model-based results that they find compelling (e.g., the above-cited attempts to develop a conditional randomization theory for the ratio estimator). Another general effect on conventional sampling theory has been to create a greater awareness of models and willingness to use them in analyses. Very important work has been done in studying the effects of using standard computer packages (i.e., analyses based on simple models) to analyze sample survey data when the models do not adequately describe the process generating the observations. Some of this work has been model-based and some has been based on random sampling distributions, but stimulated by the model-based activity, and using models in the analysis (e.g., Holt, Smith, and Winter, 1980; Skinner, Holmes, and Smith, 1986).

Developments in category (iv) are of very general importance. The model-based approach brings new statistical methods to finite populations, methods that are widely used in other areas of statistics. These new applications represent important tests cases for the methods, which are now used in real samples from real populations that can be examined *in toto* to determine exactly how large the estimator error is, whether the true mean actually lies within the confidence interval, etc. Studying statistical methods in finite populations entails a degree of realism and relevance to real-world phenomena that is hard to achieve in other contexts, where the object of estimation is an unobservable (usually purely conceptual) model parameter, or where the test data are generated artificially.

This is illustrated by the finite-population tests of the standard variance estimates in linear regression models (Royall and Cumberland, 1981a, b). These empirical studies showed that the estimates are much more sensitive to errors in the models' variance structure than had been generally acknowledged (see e.g. Efron, 1979 §7). This suggests that more attention should be paid to bias-robust alternatives. But further finite population studies have produced frightening examples showing that confidence intervals based on bias-robust estimates, although better than those based on the standard variance estimates, can also perform very poorly under conditions that, though not uncommon, are difficult to recognize when they occur (Royall and Cumberland, 1985).

Finally, the model-based approach to finite population sampling has also helped to clarify the basic concepts and principles of statistics. Stimulated by the good advice "Look at the data," along with exciting computer capabilities for display and analysis of samples, statisticians now rely heavily on the data to suggest and criticize models. Finite population studies have helped to emphasize the limitations of this sort of empiricism: model failure that is not apparent in the sample can produce seriously misleading inferences (e.g., Royall and Cumberland, 1981a; Rubin, 1983). Thus, robustness is vitally important even when the model fits the observed data well. Other important general issues that have been emphasized and illustrated in the model-based approach to finite population sampling include the critical distinction between probabilistic and inferential validity and the need for conditioning on ancillary statistics to achieve the latter (see Royall, 1976a, for discussion; ref. also Hinkley, 1983), the inferential inadequacy of probability distributions generated by artificial randomization, and the fundamental importance of likelihoods (Royall, 1976b, discusses the last two points).

Some Current Developments

The role of randomization in a model-based approach to finite population sampling is a subject of continuing research. Randomization is certainly valuable at the sampling stage. For example, it can ensure that the chances are good that the sample selected will be well balanced, so that in that sample a given estimator is robust with respect to variables that are not adequately accounted for by the prediction model (Royall and Herson, 1973). But just when and how random sampling probabilities should influence inferences from a given sample has proved to be a difficult issue. On one hand, the set of labels identifying the sampled units is an ancillary statistic, so that the Conditionality Principle evidently precludes any role for the random sampling distribution in inference (Basu, 1971). On the other hand, the expected balance associated with simple random sampling is a characteristic whose statistical relevance does not seem to vanish entirely when the perspective shifts from (i) choosing which units to observe to (ii) making estimates from an observed sample (ref. Royall, 1976a, p. 471). Thus there are continuing efforts to formalize and explain the precise role of random sampling in finite population inference (e.g., Sugden and Smith, 1984; Pfefferman and Holmes, 1985; Cumberland and Royall, 1988; Kott, 1988; and Tam, 1988) and to reconcile the prediction and randomization approaches (Brewer, Hanif, and Tam, 1988).

But recent progress in model-based theory has not been limited to the interface with randomization theory. Tam (1986) has given an elegant extension and unification of earlier work on robust estimation. Chambers (1988) has contributed both theoretical and empirical results on model-based estimation for domains within a larger population. And Valliant has used the prediction approach to analyze the statistical properties of a widely-used method of variance estimation (1987a), to discover critical conditional properties of estimators in stratified samples (1987b), and to study an important problem in economic statistics (1988).

References

Basu, D. (1971): An Essay on the Logical Foundations of Survey Sampling, Part I, in *Foundations of Statistical Inference*, eds. Godambe and Sprott, Holt, Rinehart and Winston of Canada, Toronto, 203-233.

Brewer, K. R. W., Hanif, M., and Tam, S. M. (1988): How nearly can model-based prediction and design-based estimation be reconciled?, *Journal of the American Statistical Association* 83, 128-131.

Chambers, R. L. (1988): Limited information estimation of domain means, manuscript submitted for publication.

Chiu, H. Y., and Sedransk, J. (1986): A Bayesian procedure for imputing missing values in sample surveys, *Journal of the American Statistical Association* 81, 667-676.

Cornfield, J. (1966): A note on the likelihood function generated by randomization over a finite set, *Bulletin of the International Statistical Institute* 41, Book 1, 79-80.

Cumberland, W. G., and Royall, R. M. (1981): Prediction models and unequal probability sampling, *Journal of the Royal Statistical Society*, Ser. B, 43, 353-367.

Cumberland, W. G., and Royall, R. M. (1988): Does simple random sampling provide adequate balance?, *Journal of the Royal Statistical Society*, Ser. B, 50, 118-124.

Deng, L.-Y., and Wu, C. F. J. (1987): Estimation of variance of the regression estimator, *Journal of the American Statistical Association* 82, 568-576.

Efron, B. (1979): Bootstrap methods: another look at the jackknife, *Annals of Statistics* 7, 1-26.

Ericson, W. A. (1988): Bayesian inference in finite populations, Chapter 9 in *Handbook of Statistics, Volume 6, Sampling*, eds. Krishnaiah and Rao, North-Holland, New York, 213-246.

Fisher, R. A. (1935): *The Design of Experiments*, first edition, Oliver and Boyd, Edinburgh.

Fisher, R. A. (1960): *The Design of Experiments*, seventh edition, Hafner Publishing Company, New York.

Fuller, W. A. (1981): Comment, *Journal of the American Statistical Association* 76, 76-80.

Geisser, S. (1986): Predictive analysis, in *Encyclopedia of Statistical Sciences*, Volume 7, eds. Kotz and Johnnson, John Wiley & Sons, Inc., New York.

Godambe, V. P. (1955): A unified theory of sampling from finite populations, *Journal of the Royal Statistical Society*, Ser. B, 17, 267-278.

Godambe, V. P. (1966): A new approach to sampling from finite populations I: Sufficiency and linear estimation, *Journal of the Royal Statistical Society*, Ser. B, 28, 310-219.

Hansen, M. H., Madow, W. G., and Tepping, B. J. (1983): An evaluation of model-dependent and probability-sampling inferences in sample surveys, *Journal of the American Statistical Association* 78, 776-793.

Hanurav, T. V. (1968): Hyper-admissibility and optimum estimators for sampling finite populations, *Annals of Mathematical Statistics* 39, 621-642.

Hartley, H. O., and Rao, J. N. K. (1968): A new estimation theory for sample surveys, *Biometrika* 55, 547-557.

Hinkley, D. V. (1983): Can frequentist inferences be very wrong? A conditional yes, in *Scientific Inference, Data Analysis, and Robustness*, eds. Box, Leonard, and Wu, Academic Press, New York, 85-103.

Holt, D., and Smith, T. M. F. (1979): Post stratification, *Journal of the Royal Statistical Society*, Ser. A, 142, 33-46.

Holt, D., Smith, T. M. F., and Tomberlin, T. J. (1979): A model-based approach to estimation for small subgroups of a population, *Journal of the American Statistical Association* 74, 405-410.

Holt, D., Smith, T. M. F., and Winter, P. D. (1980): Regression analysis of data from complex surveys, *Journal of the Royal Statistical Society*, Ser. A, 143, 474-487.

Kalbfleisch, J. D., and Sprott, D. (1969): Applications of likelihood and fiducial probability to sampling finite populations, in *New Developments in Survey Sampling*, eds. Johnson and Smith, John Wiley & Sons, New York, 358-389.

Kempthorne, O. (1955): The randomization theory of experimental inference, *Journal of the American Statistical Association* 50, 946-967.

Kott, P. S. (1986): Some asymptotic results for the systematic and stratified sampling of a finite population, *Biometrika* 73, 485-491.

Kott, P. S. (1988): Model-based finite population correction for the Horvitz-Thompson Estimator, *Biometrika* 75, to appear.

Laake, P. (1979): A predictive approach to subdomain estimators in finite populations, *Journal of the American Statistical Association* 74, 355-358.

Little, R. J. A. (1982): Models for nonresponse in survey samples, *Journal of the American Statistical Association* 77, 237-250.

Neyman, J. (1934): On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection, *Journal of the Royal Statistical Society* 97, 558-606.

Neyman, J., and Pearson, E. S. (1937): Note on some points in 'Students' paper on 'Comparison between balanced and random arrangements of field plots', *Biometrika* 29, 380-388.

O'Muircheartaigh, C., and Wong, S. T. (1981): The impact of sampling theory on survey sampling practice: A review, *Bulletin of the International Statistical Institute* 49, 465-493.

Pfefferman, D., and Holmes, D. J. (1985): Robustness considerations in the choice of a method of inference for regression analysis of survey data, *Journal of the Royal Statistical Society*, Ser. A, 148, 268-278.

Rao, P. S. R. S., and Rao, J. N. K. (1971): Small sample results for ratio estimators, *Biometrika* 58, 625-630.

Robinson, J. (1987): Conditioning ratio estimates under simple random sampling, *Journal of the American Statistical Association* 82, 826-831.

Royall, R. M. (1968): An old approach to finite population sampling theory, *Journal of the American Statistical Association* 63, 1269-1279.

Royall, R. M. (1976a): Current advances in sampling theory: Implications for human observational studies, *American Journal of Epidemiology* 104, 463-474.

Royall, R. M. (1976b): Likelihood functions in finite population sampling theory, *Biometrika* 63, 605-614.

Royall, R. M. (1979): Prediction models in small area estimation, in *Synthetic Estimates for Small Areas*, ed. J. Steinberg, NIDA Monograph 24, National Institute on Drug Abuse, Washington, D.C., 63-87.

Royall, R. M., and Cumberland, W. G. (1978): Variance estimation in finite population sampling, *Journal of the American Statistical Association* 73, 351-358.

Royall, R. M., and Cumberland, W. G. (1981a): An empirical study of the ratio estimator and estimators of its variance, *Journal of the American Statistical Association* 76, 66-77.

Royall, R. M., and Cumberland, W. G. (1981b): The finite-population linear regression estimator and estimators of its variance—An empirical study, *Journal of the American Statistical Association* 76, 924-930.

Royall, R. M., and Cumberland, W. G. (1985): Conditional coverage properties of finite population confidence intervals, *Journal of the American Statistical Association* 80, 355-359.

Royall, R. M., and Herson, J. H. (1973): Robust estimation in finite populations I, *Journal of the American Statistical Association* 68, 880-889.

Rubin, D. B. (1983): A case study of the robustness of Bayesian methods of inference: Estimating the total of a finite population using transformations to normality, in *Scientific Inference, Data Analysis, and Robustness*, eds. Box, Leonard, and Wu, Academic Press, New York, 213-244.

Särndal, C. E. (1981): Frameworks for inference in survey sampling with applications to small area estimation and adjustment for nonresponse, *Bulletin of the International Statistical Institute* 49, Book 1, 494-513.

Skinner, C. J., Holmes, D. J., and Smith, T. M. F. (1986): The effect of sample design on principal component analysis, *Journal of the American Statistical Association* 87, 789-798.

Smith, T. M. F. (1976): The foundations of survey sampling: A review, *Journal of the Royal Statistical Society*, Ser. A, 139, 183-195.

Smith, T. M. F. (1983): On the validity of inferences from non-random samples, *Journal of the Royal Statistical Society*, Ser. A, 146, 394-403.

Sugden, R. A., and Smith, T. M. F. (1984): Ignorable and informative designs in survey sampling inference, *Biometrika* 71, 495-506.

Stuart, A. (1962): *Basic Ideas of Scientific Sampling*, Hafner Publishing Company, New York.

"Student" (1937): Comparison between balanced and random arrangements of field plots, *Biometrika* 29, 363-379.

Tam, S M. (1986): Characterization of best model-based predictors in survey sampling, *Biometrika* 73, 232-235.

Tam, S. M. (1988): Asymptotically design-unbiased predictors in survey sampling, *Biometrika* 75, 175-177.

Valliant, R. (1987a): Some prediction properties of half-sample variance estimators in single-stage sampling, *Journal of the Royal Statistical Society*, Ser. B, 49, 68-81.

Valliant, R. (1987b): Conditional properties of some estimators in stratified sampling, *Journal of the American Statistical Association* 82, 509-519.

Valliant, R. (1988): Estimation of the Laspeyres price indexes using the prediction approach for finite population sampling, *Journal of Business and Econometric Statistics* 6, 189-196.

Wu, C. F. J., and Deng, L. Y. (1983): Estimation of variance of the ratio estimation: An empirical study, in *Scientific Inference, Data Analysis, and Robustness*, eds. Box, Leonard, and Wu, Academic Press, New York, 245-277.