The Annals of Statistics
1979, Vol. 7, No. 3, 531–536

THE BEST STRATEGY FOR ESTIMATING THE MEAN OF A FINITE POPULATION[1]

BY V. M. JOSHI

University of Western Ontario

If the finite population is homogeneous and the statistician's resources permit him to take a sample size m at the most, it is shown here that his best strategy for estimating the population mean is to draw a sample of size m by simple random sampling without replacement and to take the sample mean as the estimate. The strategy is the best in the sense that in the entire class of unbiased strategies subject to the restriction that the size of any observed sample does not exceed m, this strategy minimizes for any convex loss function both the maximum and the average risks over the set of parameter points arising from permutations of the labels of the population units. Similar decision-theoretic justification is also derived for the customary strategy for a two-stage cluster-sampling design.

1. Introduction. Suppose we have a finite population for which the mean value of some variate is to be estimated and, according to the available knowledge, the population is 'homogeneous' in respect to the variate. Suppose further that the resources at the statistician's disposal allow him to take a sample of size m at the most. What procedure should he adopt? Almost invariably the procedure adopted in practice is to draw a sample of size m by simple random sampling without replacement, and to take the sample mean as the estimate. This procedure is intuitively appealing, but a nonintuitive justification for preferring it over all others has not been given. In this note we provide this justification by showing that the above stated strategy (procedure) is the best in the entire class of unbiased strategies in the sense that it minimizes for a convex loss function both the average and the maximum risks over the set of parameter points obtained by permuting the labels of the population units. Here strategy means the choice of a sampling design and an estimator and it is said to be unbiased if, for the chosen sampling design, the estimator is unbiased. In deriving the result we use two previous results of Royall ([4], [5]).

2. Preliminaries. The finite population U consists of units labelled in some order as $u_1, u_2, \cdots, u_N$. With each unit u_i is associated an unknown real variate value x_i. $\mathbf{x} = (x_1, x_2, \cdots x_N)$ is the parametric vector. The parametric space is a specified subset $\bar{R}_N$ of the N-space. A sample s means any subset of U. S denotes the set of all samples s. A sampling design is any probability distribution p on S.

Received November 1977; revised July 1978.
[1]This research was partially supported by a grant from the National Research Council of Canada.
AMS 1970 *subject classifications.* 62D05, 60F05.
Key words and phrases. Homogeneous population, simple random sampling, sample mean.

An estimator e is a real valued function on $S \times \bar{R}_N$ which for each s, depends on $\mathbf{x}$ through only those x_i for which $u_i \in s$. We write $i \in s$ for brevity for $u_i \in s$.

For any sample s, the sample size $n(s)$ is the number of units contained in s. Let, for any s, the co-ordinates x_i, $i \in s$ arranged in nondecreasing order, be $y_1, y_2, \cdots, y_{n(s)}$. The latter are called the order statistics.

DEFINITION 2.1. An estimator e is symmetric if it depends on s and $\mathbf{x}$ only through $n(s)$ and the order statistics.

Thus the value of a symmetric estimator depends only on the observed values x_i, $i \in s$ and not on the units observed in the sample.

3. Main result. Let π denote a permutation of the label numbers $\{1, 2, \cdots, N\}$ of the units. Following Royall [5], assume that a unit which gets the label i after the permutation had the label $\pi(i)$ before. Let (p, e) be any given strategy, i.e., pair of sampling design and estimator. From e a symmetric estimator $\bar{e}$ (Definition 2.1) is obtained by averaging as follows: if $p(n(s)) > 0$, let

$$\bar{e}(s, \mathbf{x}) = \Sigma_\pi e(\pi^{-1}(s), \mathbf{x}_\pi)p(\pi^{-1}(s))/ (n(s))!\,(N - n(s))!\,p(n(s)) \tag{1}$$

where $\pi^{-1}(s) = \{j, \pi(j) \in s\}$, $\mathbf{x}_\pi$ is the coordinate vector resulting from the vector $\mathbf{x}$ by the permutation π, and $p(n(s))$ is the total probability under p of samples of the size $n(s)$, i.e.,

$$p(n(s)) = \Sigma_{s' : n(s') = n(s)} p(s'). \tag{2}$$

If $p(n(s)) = 0$, assign to $\bar{e}(s, \mathbf{x})$ any arbitrary value. It is seen that $\bar{e}$ is a symmetric estimator.

Following Royall [5] we assume a loss function $l(a, \theta)$ of real variables a and θ which is convex in a for every θ. For any strategy (p, e), $R(e, p, \mathbf{x})$ denotes the risk-function at $\mathbf{x}$, and $\bar{R}(e, p, \mathbf{x})$ denotes the mean value of the risk-function over the $N!$ parameter points (not necessarily all distinct), obtained by permuting the coordinates of $\mathbf{x}$. The parameter function under estimation is the population mean $\theta(\mathbf{x})$. We assume that the parameter space $\bar{R}_N$ is such that if $\mathbf{x} \in \bar{R}_N$, all the points arising from permutation of the coordinates of $\mathbf{x}$ also belong to $\bar{R}_N$. Then

PROPOSITION 3.1. *For all* $\mathbf{x} \in \bar{R}_N$

$$\bar{R}(\bar{e}, p, \mathbf{x}) \leqslant \bar{R}(e, p, \mathbf{x}). \tag{3}$$

PROOF. This result is proved in the theorem in [5].

Next, for any sampling design p, let $\bar{p}$ be the symmetric sampling design defined as follows: for any sample s

$$\bar{p}(s) = \binom{N}{n(s)}^{-1} p(n(s)) \tag{4}$$

where $p(n(s))$ is given by (2). Thus $\bar{p}$ assigns equal probabilities to all samples having the same size in such a way that the total probability for each size coincides with the corresponding probability under p. Next, PROPOSITION 3.2. $\bar{p}$ being the design defined by (4),

$$\bar{R}(\bar{e}, \bar{p}, \mathbf{x}) = \bar{R}(\bar{e}, p, \mathbf{x}). \tag{5}$$

PROOF.

$$\begin{aligned} N!\bar{R}(\bar{e}, p, \mathbf{x}) &= \Sigma_\pi \Sigma_{s \in S} p(s) l(\bar{e}(s, \mathbf{x}_\pi), \theta(\mathbf{x})) \\ &= \Sigma_{s \in S} p(s) \Sigma_\pi l(\bar{e}(s, \mathbf{x}_\pi), \theta(\mathbf{x})) \\ &= \Sigma_{m=1}^N \Sigma_{s\,:\,n(s)=m} p(s) \Sigma_\pi l(\bar{e}(s, \mathbf{x}_\pi), \theta(\mathbf{x})) \\ &= \Sigma_{m=1}^N p(m) m!\,(N-m)! \Sigma_{s\,:\,n(s)=m} l(\bar{e}(s, \mathbf{x}), \theta(\mathbf{x})) \\ &= N! \Sigma_{m=1}^N \Sigma_{n(s)=m} \bar{p}(s) l(\bar{e}(s, \mathbf{x}), \theta(\mathbf{x})) \\ &= N! \Sigma_{s \in S} \bar{p}(s) l(\bar{e}(s, \mathbf{x}), \theta(\mathbf{x})) \\ &= N! \bar{R}(\bar{e}, \bar{p}, \mathbf{x}). \end{aligned}$$

[REMARK 3.1. The proof essentially follows from $R(\bar{e}, \bar{p}, \mathbf{x}_\pi)$ not depending on π, so that $\bar{R}(\bar{e}, \bar{p}, \mathbf{x}) = R(\bar{e}, \bar{p}, \mathbf{x}) = \bar{R}(\bar{e}, p, \mathbf{x})$.] Next for $k = 1, 2, \cdots$, let $\bar{p}_k$ denote a symmetric sampling design of fixed sample size k, i.e., $\bar{p}_k$ is simple random sampling without replacement of size k. Let the sampling design p in (3) be such that the maximum sample size for any sample with positive probability does not exceed a given number m. For $k = 1, 2, \cdots, m$, let s_k denote a sample of size k and let e^* be the estimator defined by

$$e^*(s_m, \mathbf{x}) = \Sigma_{k=1}^m p(k) \binom{m}{k}^{-1} \Sigma_{s_k\,:\,s_k \subset s_m} \bar{e}(s_k, \mathbf{x}) \tag{6}$$

in which $\bar{e}$ is as in (1) and $p(k)$ as in (2). For $n(s) \neq m$, e^* may be assigned arbitrary values. Then

PROPOSITION 3.3. *For all* $\mathbf{x} \in \bar{R}_N$

$$\bar{R}(e^*, \bar{p}_m, \mathbf{x}) \leqslant \bar{R}(\bar{e}, \bar{p}, \mathbf{x}) \tag{7}$$

where $(\bar{e}, \bar{p})$ *is the strategy in equation* (5).

PROOF. By the convexity of $l(a, \theta)$ it follows from (6) that

$$l(e^*(s_m, \mathbf{x}), \theta(\mathbf{x})) \leqslant \Sigma_{k=1}^m \binom{m}{k}^{-1} p(k) \Sigma_{s_k\,:\,s_k \subset s_m} l(\bar{e}(s_k, \mathbf{x}), \theta(\mathbf{x})). \tag{8}$$

Multiply both sides of (8) by $\binom{N}{m}^{-1}$ and sum over all samples s_m. Noting that $\bar{p}_m(s_m) = \binom{N}{m}^{-1}$ and that in the summation over all s_m, each particular s_k occurs $\binom{N-k}{m-k}$ times, we obtain

$$\begin{aligned} R(e^*, \bar{p}_m, \mathbf{x}) &\leqslant \Sigma_{k=1}^m \Sigma_k \bar{p}(s_k) l(\bar{e}(s_k, \mathbf{x}), \theta(\mathbf{x})) \\ &= R(\bar{e}, \bar{p}, \mathbf{x}). \end{aligned} \tag{9}$$

Since both sides of (9) are invariant under permutations of $\mathbf{x}$, (7) follows.

Next, let the strategy (p, e) be unbiased, i.e.,

$$\Sigma_{s \in S} p(s) e(s, \mathbf{x}) = \theta(\mathbf{x})) \tag{10}$$

for all $\mathbf{x} \in \bar{R}_N$. Then,

PROPOSITION 3.4. *If the strategy (p, e) in (3) is unbiased for $\theta(\mathbf{x})$, the strategy $(\bar{p}_m, e^*)$ in (7) is also unbiased for $\theta(\mathbf{x})$ and further, provided the parametric space satisfies the condition in equation (11), e^* is the same as the sample mean.*

PROOF. From (1) and (4)

$$\begin{aligned}\Sigma_{s \in S} \bar{p}(s) \bar{e}(s, \mathbf{x}) &= (N!)^{-1} \Sigma_{s \in S} \Sigma_\pi e(\pi^{-1}(s), \mathbf{x}_\pi) p(\pi^{-1}(s)) \\ &= (N!)^{-1} \Sigma_\pi \Sigma_{s_1 \in S} e(s_1, \mathbf{x}_\pi) p(s_1) \\ &= (N!)^{-1} \Sigma_\pi \theta(\mathbf{x}) = \theta(\mathbf{x}),\end{aligned}$$

and then, from (6),

$$\begin{aligned}\Sigma_{s \in S} \bar{p}_m(s) e^*(s, \mathbf{x}) &= \Sigma_{s_m} \bar{p}_m(s_m) e^*(s_m, \mathbf{x}) \\ &= \Sigma \bar{p}(s) \bar{e}(s, \mathbf{x}) = \theta(\bar{x}).\end{aligned}$$

Hence the strategy $(\bar{p}_m, e^*)$ is unbiased.

Next, let ξ be any arbitrary set of real numbers and let $\bar{R}_N$ satisfy, for some ξ,

$$\mathbf{x} \in \bar{R}_N \quad \text{iff} \quad x_i \in \xi \quad \text{for} \quad i = 1, 2, \cdots, N. \tag{11}$$

((11) means that each x_i assumes in $\bar{R}_N$ all the values in the set ξ.)

It then follows, from the completeness of the order statistics, (cf. [4]), that the sample mean, $1/m \Sigma_{i \in s_m} x_i$ is the unique symmetric unbiased estimator of the population mean $\theta(\mathbf{x})$. Hence, in (6) $e^*(s_m, \mathbf{x}) =$ the sample mean. This completes the proof of Proposition 3.4.

Let D_m denote the class of all unbiased strategies (p, e) with the restriction that the maximum sample size for any sample with positive probability does not exceed m. Then (3), (5), (7) and Proposition 3.4 imply that the strategy $(\bar{p}_m, e^*)$ minimizes the average risk $\bar{R}$ for each $\mathbf{x} \in R_N$ in the class D_m, and further, since $R(e^*, \bar{p}_m, \mathbf{x})$ is invariant under the permutations π, the strategy (p_m, e^*) also minimizes max $R(e, p, \mathbf{x}_\pi)$ for $(p, e) \in D_m$. Hence,

THEOREM 3.1. *For estimating the population mean the strategy of selecting the sample of fixed size m by simple random sampling with the sample mean as the estimator is the best in the entire class of unbiased strategies with the maximum sample size $\leqslant m$, in the sense that it minimizes for each $\mathbf{x}$, both the average and the maximum risks over the set of (not necessarily distinct) $N!$ parameter points obtained by permuting the coordinates of $\mathbf{x}$.*

COROLLARY 3.1. *The strategy $(\bar{p}_m, e^*)$ minimizes in the class D_m the Bayes risk under any prior distribution symmetrical in the x_i for $i = 1, 2, \cdots, n$.*

COROLLARY 3.2. *For a sampling design with fixed sample size m, the sample mean is the best estimator in the sense defined in Theorem* 3.1.

Corollary 3.1 is obtained by integration over the space of all points (**x**). Corollary 3.2 is an obvious consequence of Theorem 3.1.

NOTE 3.1. If (p, e) is unbiased, the intermediate strategy $(p, \bar{e})$ in Proposition 3.1 is not necessarily unbiased. But this does not affect the subsequent argument.

REMARK 3.2. It was shown in [3] that, excluding trivial cases, there does not exist a UMVU estimator for the population mean. (cf. also [1] for an alternative, short and simple proof.) The argument is here completed by showing that the sample mean is the best estimator provided the variance is averaged (or maximized) over the set of parameter points obtained by permutation of the coordinates of **x**. The assumption that the finite population is 'homogeneous' in respect of x_i implies that all such parameter points are 'equally possible' and hence it is reasonable for the statistician to prefer the estimator which minimizes his maximum risk on this set and also the average risk on it.

4. Extension to other sampling designs. Sampling designs used in practice are generally more complex. It is, therefore, of interest to consider to what extent the decision-theoretic justification in Theorem 3.1 extends to other designs. The extension to a design of stratified simple random sampling is straightforward. Consider, therefore, a design of cluster sampling. For simplicity consider a two-stage procedure in which in the first stage only one cluster is chosen by random sampling from the clusters constituting the population and, in the second, a sample of size $\leqslant m$ is drawn from the chosen cluster. Let C denote the class of all unbiased strategies (p, e) in which the sampling design p is a two-stage procedure of the specified type. Then it turns out that, in this class, the customary strategy is the best one in the sense of Theorem 3.1.

For, suppose the given finite population $\{u_1, u_2, \ldots, u_N\}$ is divided into k clusters of sizes $M_1, M_2, \cdots, M_k$, the assumed condition of homogeneity—according to all available knowledge—holding in respect of the entire population of all clusters taken together. (The actual labelling of the population units may have been done in a different manner, but that does not affect the argument.) Let (p, e) be any arbitrary strategy in the class C and let $(\bar{p}_m, \bar{e})$ $(\bar{p}_m, e^*)$ be the strategies as defined by (4), (1) and (6). Then, as (p, e) is unbiased, $(\bar{p}_m, e^*)$ is unbiased by Proposition 3.4. Assuming that the parametric space R_N is sufficiently large, i.e., that (11) holds, it follows from the completeness of the order statistics that

$$e^* = \text{sample mean}. \tag{12}$$

Next let $\bar{\bar{p}}_m$ denote the sampling design in which in the first stage one cluster is chosen by sampling with probability proportional to cluster-size and, in the second, a sample of fixed size m drawn from the chosen cluster by simple random, without replacement sampling. Then the customary strategy is to use the sampling design $\bar{\bar{p}}_m$, with the sample mean as the unbiased estimate of the population mean (cf. [2], page 295). Hence by (12) the customary strategy is $(\bar{\bar{p}}_m, e^*)$. Application of Proposition (3.3) to $(\bar{\bar{p}}_m, e^*)$ yields

$$\bar{R}\left(e^*, \bar{\bar{p}}_m, \mathbf{x}\right) = \bar{R}(e^*, \bar{p}_m, \mathbf{x}) \leqslant \bar{R}(e, p, \mathbf{x}) \tag{13}$$

by (9), (5) and (3). Since the risk of $(\bar{\bar{p}}_m, e^*)$ is constant over the subset of parameter points obtained by permutations of the x_i for units within the same cluster, $(\bar{\bar{p}}_m, e^*)$ is minimax in class C on each such subset and hence, necessarily, also over the entire set of $N!$ (not necessarily distinct) points arising from all permutations of the x_i. As the average risk is also minimized by (13), the customary strategy is the 'best' in the class C.

The extensions suggest that over the major portions of the field of survey-sampling, it may be possible to justify customary procedures within the N-P-W approach.

Acknowledgment. I wish to thank a referee and the Associate Editor for suggestions regarding presentation.

REFERENCES

[1] Basu, D. (1971). An essay on the logical foundations of survey sampling, Part I. *Proc. Symp. Foundations Statist. Inference*. 203–242. Holt Rinehart and Winston, Toronto.

[2] Cochran, W. G. (1977). *Sampling Techniques, 3rd ed.* Wiley, New York.

[3] Godambe, V. P. and Joshi, V. M. (1965). Admissibility and Bayes estimation in sampling from finite populations I. *Ann. Math. Statist.* **36** 1707–1722.

[4] Royall, R. M. (1968). An old approach to finite sampling theory. *J. Amer. Statist. Assoc.* **63** 1269–1279.

[5] Royall, R. M. (1970). Finite population sampling—on labels in estimation. *Ann. Math. Statist.* **41** 1774–1779.

Department of Mathematics
University of Western Ontario
London, Canada N6A 5B9