Resource Inventory Notes

BLM-25 September 1979

SAMPLING OF NATURAL RESOURCE POPULATIONS: COMPARISON OF ESTIMATORS

by

Gary W. Fowler[1]

ABSTRACT: The mean square error is advocated as a measure of the accuracy of estimators of parameters of natural resource populations. The accuracies of two estimators are compared by determining the relative efficiency of one estimator to the other. For biased estimators, the effect of the bias on probability statements used to construct confidence intervals and test hypotheses is examined. Applications are made to two known forest populations.

\- -

INTRODUCTION

An estimator is a mathematical rule for calculating an estimate of a population characteristic (parameter) from sample observations. The estimate is a number calculated by substituting the n sample observations taken from the population into the estimator. For example, assume that a forest can be divided into a conceptual population of N equal-area mutually exclusive square plots and that the forest sampler is interested in estimating basal area (BA) per acre. A random sample of n plots is taken from the forest with BA/acre being determined on each plot. The estimator is

$$(1) \quad \bar{x} = \sum_{i=1}^{n} x_i/n$$

where $\bar{x}$ is the mean BA/acre for the n plots and x_i is the BA/acre determined on the ith plot in the sample. A random sample of 5 plots with BA/acre values of 60, 100, 80, 90, and 70 sq. ft. yields the estimate $\bar{x}$ = 80.0 sq. ft./acre by substituting the 5 sample values into equation 1.

Most samplers would like to choose estimators that give unbiased estimates of population parameters. The estimator $\bar{x}$ is unbiased if

$$(2) \quad E(\bar{x}) = \mu,$$

---

[1] The author is Associate Professor of Biometrics, School of Natural Resources, The University of Michigan, Ann Arbor, Michigan 48109

USDI, Bureau of Land Management, D 460
Denver Service Center, Denver Federal Center, Bldg. 50
Denver, Colorado, 80225

which means that the average of the sample means from all possible samples of size n from the population is equal to the population mean μ. While unbiasedness is a desirable characteristic of an estimator, it is not the only measure of the "goodness" of an estimator. The sampler would also like to have an estimator that is both precise and accurate.

The objective of this paper is to present the mean square error as a measure of the accuracy of an estimator, compare the accuracies of two estimators by determining the relative efficiency of one estimator to the other, and examine the effect of bias on probability statements used to construct confidence intervals and test hypotheses. Applications will be made to two known forest populations.

THE MEAN SQUARE ERROR AND RELATIVE EFFICIENCY

The precision of an estimator refers to the size of the deviations of sample estimates (x) from their mean and is usually measured by the variance (or standard deviation) of the estimator. The variance of the estimator is

$$(3)\quad \sigma_{\bar{x}}^{2} = \sum_{j=1}^{J} (\bar{x}_j - E(\bar{x}))^2/J$$

where J is the number of possible samples of size n that can be taken from the population, $\bar{x}_j$ is the sample mean associated with the jth sample, and $E(\bar{x}) = \sum_{j=1}^{J} \bar{x}_j/J$ is the mean of the J sample estimates. $J = \binom{N}{n}$ for simple random sampling without replacement.

The accuracy of an estimator refers to the size of the deviations of sample estimates from the population parameter μ of the population being sampled and is usually measured by the mean square error (Cochran 1965, Lindgren 1962, Raj 1968):

$$(4)\quad MSE_{\bar{x}} = E(\bar{x} - \mu)^2 = E[(\bar{x} - E(\bar{x})) + (E(\bar{x}) - \mu)]^2 = \sigma_{\bar{x}}^{2} + B^2$$

where $B = E(\bar{x}) - \mu$ is the bias of the estimator.

If an estimator is unbiased, $E(\bar{x}) = \mu$. In such cases, the mean square error is equal to the variance of the estimator $\sigma_{\bar{x}}^{2}$ and its accuracy and precision are identical. Thus, precision is a measure of the variability of an estimator around its mean while accuracy is a measure of the variability of an estimator around the true parameter of the population.

The sampler would like to choose the most accurate estimator for estimating a population parameter. If two estimators are unbiased, the one with the smallest variance $\sigma_{\bar{x}}^{2}$ is the most accurate. However, if one or both estimators are biased, the one with the smallest mean square error is the most accurate.

The relative efficiency of estimator 1 compared to estimator 2 is

(5) $e(1,2) = MSE_2/MSE_1$ (Lindgren 1962).

If $e(1,2) > 1$, estimator 1 is more accurate than estimator 2. The reverse is true if $e(1,2) < 1$. Both estimators have the same accuracy if $e(1,2) = 1$. If estimators 1 and 2 are unbiased, $e(1,2)$ would be the ratio of their variances. For example, if estimator 1 were unbiased with variance $\sigma_1^2 = 100$ and estimator 2 were biased with variance $\sigma_2^2 = 64$ and bias $B = 4$, then

$$MSE_1 = \sigma_1^2 = 100$$

$$MSE_2 = \sigma_2^2 + B^2 = 64 + 16 = 80$$

$$e(1,2) = MSE_2/MSE_1 = 80/100 = 0.8$$

Even though estimator 2 is biased, it is more accurate than estimator 1 because it is considerably more precise (has a much smaller variance) than estimator 1.

APPLICATIONS

Conceptual populations of mutually exclusive square plots were obtained for two completely enumerated forest populations:

1. conceptual populations of $N = 112$ and 28 plots for .031 and .126 acre plots, respectively, for a 3.52 acre eastside pine forest in California.[2/]
2. conceptual population of $N = 988$ one-fortieth (.025) acre plots for a 24.7 acre white pine forest in Michigan.

Conceptual populations of .025 and .020 acre fixed-area mutually exclusive circular plots for the 3.52 and 24.7 acre forests, respectively, were obtained by inscribing the circular plots inside the 112 and 988 square plots. The stem maps of the two forests from which these conceptual populations were constructed are described in Fowler and Davis (1979).

Assume that a random sample of n plots is taken from a conceptual population of N plots with the purpose of estimating the BA/acre for a forest population. Since the conceptual populations of square plots completely cover the forest, estimators based on square plots are unbiased estimators ($\bar{x}_{UB}$) of forest parameters. On the other hand, estimators based on samples of circular plots taken from a conceptual population of circular plots inscribed inside square plots are biased estimators ($\bar{x}_B$) of the forest parameters as only 78.54% of

---

[2/] The stem map from which this stand was constructed was kindly provided by Dr. William G. O'Regan, former Mathematical Statistician, Pacific Southwest Forest and Range Experiment Station, Berkeley, California.

the forest area is covered by the conceptual population (Fowler and Davis 1979).

The variances of the estimator $\bar{x}$[3/] for the conceptual populations of the .031 and .126 acre square plots for the 3.52 acre forest are

$$\sigma_1^2 = \frac{\sigma^2}{n} = \frac{39{,}109.71}{n} \text{ and } \sigma_2^2 = \frac{\sigma^2}{n} = \frac{9{,}765.31}{n}, \text{ respectively,}$$

where σ^2 is the variance of BA/acre for the populations of N = 112 and 28 plots, respectively. Since both estimators are unbiased,

$$MSE_1 = \sigma_1^2 = 39{,}109.71/n$$

$$MSE_2 = \sigma_2^2 = 9{,}765.31/n$$

$$e(1,2) = \frac{\sigma_2^2}{\sigma_1^2} = 0.25$$

As expected, the estimator based on the larger plot size is considerably more accurate at any sample size n. In fact, estimator 1 would need a sample size four times that of estimator 2 for the same level of accuracy.

The mean square errors based on the estimators $\bar{x}_{UB}$ and $\bar{x}_B$ for the conceptual populations of .025 acre square plots and .020 acre circular plots inscribed inside square plots, respectively, for the 24.7 acre forest are:

$$MSE_{UB} = \sigma_{UB}^2 = \sigma_s^2 = \sigma_s^2/n = 7{,}836.95/n, \text{ and}$$

$$MSE_B = \sigma_B^2 + B^2 = \sigma_c^2/n + B^2 = 9{,}684.85/n + 2.76$$

where σ_s^2 and σ_c^2 are the variances of BA/acre for the populations of 988 square and circular plots. The bias $B = \mu_c - \mu_s = 85.62 - 87.28 = -1.66$ where μ_s and μ_c are the means of the conceptual populations of square and circular plots, respectively. μ_c is the mean BA/acre for the forest population as $\bar{x}_{UB}$ is an unbiased estimator.

The mean square errors based on the estimators $\bar{x}_{UB}$ and $\bar{x}_B$ for the conceptual populations of .031 acre square plots and .025 acre circular plots inscribed inside square plots, respectively, for the 3.52 acre forest are:

[3/] For the purposes of this paper, sampling is assumed without replacement and the finite population correction (N-n)/(N-1) associated with the variance of the estimator is disregarded.

$$MSE_{UB} = \sigma_{UB}^2 = \sigma_s^2/n = 39{,}109.71/n, \text{ and}$$

$$MSE_B = \sigma_B^2 = \sigma_c^2/n + B^2 = 59{,}184.36/n + 585.64$$

where σ_s^2 and σ_c^2 are the variances of BA/acre for the population of 112 square and circular plots. The bias $B = \mu_c - \mu_s = 165.78 - 141.58 = 24.2$, where μ_c and μ_s are as defined above.

Table 1 includes σ_{UB}^2, σ_B^2, MSE_B, and $e(UB,B) = MSE_B/MSE_{UB}$ for five sample sizes for the 24.7 and 3.52 acre forests.

Table 1. σ_{UB}^2, σ_B^2, MSE_B, e(UB,B), and $|B|/\sigma_B$ for five sample sizes for the 24.7 and 3.52 acre forests.

n	24.7 ACRE FOREST					3.52 ACRE FOREST				
	σ_{UB}^2	σ_B^2	MSE_B	e(UB,B)	$\|B\|/\sigma_B$	σ_{UB}^2	σ_B^2	MSE_B	e(UB,B)	$\|B\|/\sigma_B$
5	1,567.39	1,936.97	1,939.73	1.238	.038	7,821.94	11,836.87	12,422.51	1.588	.222
10	783.70	968.48	971.24	1.239	.053	3,910.97	5,918.44	6,504.08	1.663	.315
20	391.85	484.24	487.00	1.243	.075	1,955.49	2,959.22	3,544.86	1.813	.445
50	156.74	193.70	196.46	1.253	.119	782.19	1,183.69	1,769.33	2.262	.703
100	78.37	96.85	99.61	1.271	.169	391.10	591.84	1,177.48	3.011	.995

For both forests, the estimator based on circular plots has a larger variance than the estimator based on square plots since the circular plots are only 78.54% as large as the square plots. $\sigma_c^2/\sigma_s^2 = 1.24$ and 1.51 for the 24.7 acre and 3.52 acre forests, respectively, which shows that estimates based on square plots are considerably more precise. However, since the estimator based on circular plots ($\bar{x}_B$) is biased, mean square errors must be used to compare the accuracies of the two estimators.

The relative efficiency of the unbiased estimator ($\bar{x}_B$) is a function of sample size. e(UB,B) increases as the sample size increases for both forests (Table 1) because the variances of both estimators decrease proportionally but the bias remains constant. In all cases, the estimator based on square plots is more accurate than the estimator based on circular plots and it becomes more accurate as the sample size increases.

Since the bias for the 3.52 acre forest is relatively larger than the bias for the 24.7 acre forest, the estimator based on square plots is relatively more accurate than the estimator based on circular plots for the 24.7 acre forest. It should be remembered that the circular plots are smaller than the square plots.

PROBABILITY STATEMENTS

If the natural resource sampler wants only to make a point estimate (e.g., $\bar{x}$) of a population parameter (e.g., μ), the estimator with the smallest mean square error is best. However, if probability statements about population parameters are also needed to construct confidence intervals or test hypotheses, it should be realized that bias distorts such probability statements (Cochran 1965, Raj 1968).

For an unbiased estimator, assuming normality, variance known, and level of significance $\alpha = .05$,

$$P\,\{\bar{x}_{UB} - 1.96\,\sigma_{UB} \leq \mu \leq \bar{x}_{UB} + 1.96\,\sigma_{UB}\} = .95,$$

which yields the 95% confidence interval for μ:

$$(6)\quad \bar{x}_{UB} \pm 1.96\,\sigma_{UB}.$$

However, for a biased estimator, assuming the bias unknown,

$$P\,\{\bar{x}_{B} - 1.96\,\sigma_{B} \leq \mu \leq \bar{x}_{B} + 1.96\,\sigma_{B}\} \leq .95,$$

which yields an actual α larger than .05 and less than 95% confidence interval for μ:

$$(7)\quad \bar{x}_{B} \pm 1.96\,\sigma_{B}.$$

Since $E(\bar{x}_B) = \mu + B$, the effect of bias on the probability statement depends on $|B|/\sigma_B$. The larger $|B|/\sigma_B$, the more the actual α will be than the nominal α and the smaller the actual confidence coefficient $(1 - \alpha)$ will be compared to the nominal confidence coefficient $(1 - \alpha)$. The actual values of α for various values of $|B|/\sigma_B$ and a nominal $\alpha = .05$ are shown in Table 2. If $|B|/\sigma_B \leq .1$, the effect of bias on the probability statement is negligible. Even with $|B|/\sigma_B = .3$, the effect is quite modest.

$|B|/\sigma_B$ associated with the estimators based on circular plots is shown in Table 1 for sample sizes of 5, 10, 20, 50, and 100 for the 24.7 and 3.52 acre forests. The effect of bias on probability statements related to these estimators becomes more serious as sample size increases. The effect for the 25.7 acre forest is negligible to modest for all sample sizes while the effect for the 3.52 acre forest is modest for smaller sample sizes but substantial for large sample sizes.

Table 2. Actual values of the level of significance α for selected values of $|B|/\sigma_B$ given the nominal value of $\alpha = .05$.

$\lvert B\rvert/\sigma_B$	α	$\lvert B\rvert/\sigma_B$	α	$\lvert B\rvert/\sigma_B$	α
.00	.0500	.10	.0511	.70	.1077
.01	.0500	.20	.0546	.80	.1259
.03	.0501	.30	.0604	.90	.1467
.05	.0503	.40	.0685	1.00	.1700
.07	.0506	.50	.0790	1.50	.3231
.09	.0509	.60	.0921		

CONCLUDING REMARKS

The natural resource sampler should use estimators of population parameters that are not only precise but also accurate. The mean square error can be used as a measure of the accuracy of an estimator. The relative efficiency of one estimator to another can be determined by comparing their mean square errors. The estimator with the smallest mean square error is the best one. A biased estimator might be the most accurate estimator, but the sampler should be aware that bias distorts probability statements associated with confidence intervals and hypothesis tests. The sampler must decide whether the distortion of probability statements caused by the bias expected in a given sampling problem is important and if so, whether the distortion is acceptable.

In choosing an estimator based on a specific sampling method, the costs of sampling, sampling errors, associated measurement errors, edge effect bias, and ease of use in the field must, of course, also be considered. Sampling errors can be controlled by determining the appropriate sampling intensity (Fowler and Hauke 1979).

LITERATURE CITED

Cochran, W. G. 1965. Sampling Techniques. 2nd Edition. John Wiley and Sons, Inc., New York. 413 p.

Fowler, G.W., and C.F. Davis. 1979. Sampling natural resource populations: mutually exclusive fixed-area sampling units. Resource Inventory Notes. BLM 23:1-7.

Fowler, G.W., and D. Hauke. 1979. A distribution-free method for interval estimation and sample size determination. Resource Inventory Notes. BLM 19:1-8.

Lindgren, B.W. 1962. Statistical Theory. The MacMillan Co., New York. 427 p.

Raj, D. 1968. Sampling Theory. McGraw-Hill Book Co., New York. 302 p.

Erratum

Res. Inv. Notes. BLM-15. Page 6, Formula under C should read:

$$n = \frac{[\Sigma(As) \text{ x } \Sigma(As \text{ x } Ss^2)\ (t)^2]}{[\Sigma(As)^2 \text{ x } E^2) + \Sigma(As \text{ x } Ss^2)}$$

where t = students t. In the example given, t is assumed to be 1.

Current Literature

Please order directly from sources given.

GENERAL

Technical Note 333. "Integrating Inventories: An Annotated Bibliography", by Gyde Lund and Elise McNutt. BLM (D-460), Denver Federal Center (Bldg. 50), Denver, Colorado, 80225.

Technical Paper 356. "An Approach to Assessment of Relief Formats for Hardcopy Topographic Maps". U.S. Army Research Institute for the Behavioral and Social Sciences, 5001 Eisenhower Ave., Alexandria, Virginia, 22333.

PB 291 786/AS. *A Handbook for the Assessment of Environmental Benefits and Pollution Control Costs*. National Technical Information Service, 5825 Port Royale Road, Springfield, Virginia, 22161. $6.75.

An Appraiser Looks at Multiple Regression, by D.W. Smith. *Appraisal Journal* 47(2): 248-252.

"Land Classification Series." (Reprint from the *Journal of Forestry*.)

Gen. Tech. Report WO-12. "Strategies for Protection and Management of Floodplain Wetlands and Other Riparian Ecosystems."

Publications Office, Rocky Mountain Forest and Range Experiment Station, 240 West Prospect, Fort Collins, Colorado, 80526.

Res. Report 323. "Land-Use Planning: What, Why, When, and By Whom." Bulletin Office, Department of Agric. Information, New Mexico State University, Drawer 3A1, Las Cruces, New Mexico, 88003.

"NAWDEX: A Key to Finding Water Data", National Water Data Exchange, U.S. Geol. Survey, 421 National Center, Reston, Virginia, 22092.

FISH AND WILDLIFE

"Aerial Survey of the Bottlenosed Dolphin and the West Indian Manatee in the Indian and Banana Rivers, Florida", by Stephen Leatherwood. Fisheries Bulletin, 77(1): 47-59.

"Method of Sampling Seaward Migrations of Juvenile Salmon", by Richard W. Tyler. The Progressive Fish Culturist 41(2): 78-81.

Information Report Series (Fisheries No. 78-2). "Effect of Number of Marked Fish and Years of Repetition on Precision in Studies of Contribution to a Fishery."

Information Report Series (Wildlife No. 78-1). "Characteristics of Shrubland Habitat Associated with Mule Deer Fawns at Birth and During Early Life in Southeastern Oregon."

Oregon Department of Fish and Wildlife, Research and Development Section, 303 Extension Hall, O.S.U., Corvallis, Oregon, 97331.

Gen. Tech. Report. RM-15A. "Run Wild--For the UNIVAC 110 Series Implementation and Maintenance." Publications, Rocky Mountain Forest and Range Experiment Station, 240 W. Prospect Street, Fort Collins, Colorado, 80526.

FORESTRY

"Elementary Timber Measurements", by Harry Wiant. Vandalia Press, 113 Scenery Place, Morgantown, West Virginia, 26505. 134 pages. $9.00.

AFRI Res. Report 39. "Location Key for Best Management Practices--Silviculture." Applied Forestry Research Institute, State University of New York, College of Environmental Sciences and Forestry, Syracuse, New York, 13210.

"Increasing Estimation Efficiency in 3P Cruises", by Kim Iles in Forestry Chronicle, Feb. 1978, pgs. 42 and 43.

AFRI Res. Report 43. "Stand Volume Tables for Second-Growth Northern Hardwoods in New York." Applied Forestry Research Institute, State University of New York, Syracuse, New York, 13210.

Gen. Tech. Report NC-45. "The Extent and Characteristics of Low Productivity Aspen Areas in Wisconsin."

Res. Paper NC-166. "Estimating Site Potential From the Early Height Growth of Red Pine in the Lake States."

Res. Paper NC-167. "A Comparison of Site Index Curves for Northern Hardwood Species."

North Central Forest Experiment Station, 1992 Folwell Ave.,
St. Paul, Minnesota, 55108.

Res. Paper RM-204. "Timber Yield and Financial Return Performance of the 1974 Forestry Incentives Program." Publications, Rocky Mountain Forest and Range Experiment Station, 240 W. Prospect Street, Fort Collins, Colorado, 80526.

REMOTE SENSING

Res. Report No. WI-221-047-78. "The Inventory and Distribution of Water and Associated Land Resources in the Garrison/Devils Lake Region of North Dakota: An Application of Resource Data Acquired by ERTS", by Roland Mower. Water Resources Research Institute, North Dakota State University, Fargo, North Dakota, 58102.

"Land Resource Inventory Demonstration Project." NASA, Ames Research Center, Moffet Field, California, 94035.

Resource Analysis Branch Catalogue. The Librarian, Resource Analysis Branch, Ministry of the Environment Parliament Buildings, Victoria, British Columbia, Canada, V8V 1X4. (A compilation and description of available publications covering terrain, soils, vegetation, climate, aquatic systems, agricultural capability, forest capability, outdoor recreation capability, wildlife capability, land use, and various special and miscellaneous reports.)

Meetings

1979

September 10-14	Symposium at University of Idaho, "Remote Sensing for Natural Resources--An International View of Problems, Promise and Accomplishments." Sponsors: Society of American Foresters, ASP, IUFRO and Univ. of Idaho. Contact: Robert C. Heller, College of Forestry, Wildlife and Range Sciences, Univ. of Idaho, Moscow, Idaho, 83843.

September 17-21	Fall Technical Meeting, American Society of Photogrammetry and American Congress on Surveying and Mapping. Location: Sioux Falls, South Dakota. Contact: Frederida Simon, P. O. Box 1837, Sioux Falls, South Dakota, 57101.
September 24-28	Vegetation/Terrain Analysis: Remote Sensing Workshop. Sponsors: College of Forestry, Wildlife, and Range Sciences and Office of Continuing Education, University of Idaho; Geological Survey; Pacific Northwest Regional Commission; Earth Resources Observation Systems (EROS) Data Center. Location: Coeur d'Alene, Idaho. Contact: Joseph J. Ulliman, College of Forestry, Wildlife and Range Sciences, University of Idaho, Moscow, Idaho, 83843.
October 23-25	The 3rd Conference on the Economics of Remote Sensing Information Systems. Focus on cost-effective implementation of remote sensing information systems. Sponsors: San Jose State Univ.; Univ. of Nevada Resources Dev. Assoc.; NASA; U.S.D.A.; and U.S. AID. Location: Lake Tahoe, California. Contact: Ms. Terri Wise, 3rd Conference Coordinator, P. O. Box 239, Los Altos, California, 94022 (415-961-7477).
November 5-9	Workshop on "Applied Remote Sensing for Soil Inventory and Assessment." Sponsors: University of California, Berkeley and the California Soil Survey Committee. Location: Forest Service Engineering Center, Pleasant Hill, California. Contact: Sharon Arce, University of California Extension, 2223 Fulton Street, Berkeley, California, 94720 (415-642-1061).
November 26-30	Second Conference on Scientific Research in the National Parks. Location: San Francisco, California. Contact: G. Jay Gogue, Southeast Region, National Park Service, 1895 Phoenix Boulevard, Atlanta, Georgia, 30349 (404-996-2520).

UNITED STATES
DEPARTMENT OF THE INTERIOR
BUREAU OF LAND MANAGEMENT
DENVER FEDERAL CENTER
BUILDING 50
DENVER, COLORADO 80225

OFFICIAL BUSINESS
PENALTY FOR PRIVATE USE $300

POSTAGE AND FEES PAID
U. S. DEPARTMENT OF THE INTERIOR
Int 415

U.S.MAIL

T GREGOIRE
I N E R JAMES HALL
UNIVERSITY OF NEW HAMPSHIRE
DURHAM NH 03824

01 002841