

Technical Notes: Double Sampling Revisited

ROBIN M. REICH, CHARLES D. BONHAM, AND KIMBERLY K. REMINGTON

Authors are associate professor, Forest Science Department, professor and graduate research assistant, Range Science Department, Colorado State University, Fort Collins 80523.

Abstract

The decision to use double sampling with a regression or ratio estimator is not a simple task. This study was conducted to determine whether a ratio or regression estimator should be used to estimate aboveground biomass of stands dominated by blue grama (*Bouteloua gracilis* (H.B.K.) Lag ex Steud.) in eastern Colorado. One hundred 0.25-m² circular plots were systematically located in a homogeneous stand of blue grama, and on each plot biomass was estimated visually and then clipped. Three methods (classical, jackknife, and bootstrap) of estimating the variance for double sampling with regression and ratio estimator were compared in a simulation study using sample sizes 10, 20, 30, 40, and 50 clipped plots. The ratio estimator consistently had smaller bias and should be used for estimating average clipped weight of blue grama. For n = 10 clipped plots, the jackknife variance estimator is recommended for constructing confidence intervals. For n ≥ 20 clipped plots, the classical variance estimate should be used to obtain reliable estimates of the population variance and in estimating confidence intervals.

Key Words: biomass, blue grama, bootstrap, double sampling, jackknife, monte carlo simulation

Pechanec and Pickford (1937) were among the first to estimate weight of vegetation by a combination of guessing and clipping. This method became known as double sampling. It has been described by several authors, and Francis et al. (1979) provided an extensive review of literature on the method. One of 2 statistical procedures can be used to express the relationship between clipped (y) and estimated (x) values: (1) linear regression, or (2) ratio estimation. Francis et al. (1979) compared the 2 methods, but their study was limited to analysis of variance estimates.

A major problem encountered in the combined use of clipping and estimating biomass is determination of the proportion of clipped plots to estimated plots. Of special interest is the ratio needed to obtain the desired accuracy of estimated biomass for the

greatest economy. It is well known that efficiency of double sampling depends on precision of the ratio or regression estimator, and on the relative cost of clipping compared to that of estimation. The objective of this study was to evaluate the use of ratio and regression estimators for determining the aboveground biomass of stands dominated by blue grama [*Bouteloua gracilis*, (H.B.K.) Lag. ex Steud.] in eastern Colorado.

Methods

One-hundred 0.25-m² circular plots were systematically located in a homogeneous stand of blue grama at 3-m spacings along twenty 15-m transects, which were spaced 5 m apart. Aboveground biomass of blue grama in each plot was estimated visually using 1 observer with 2 years of experience to eliminate any possible bias in estimated weights. Biomass in each plot was then clipped and weighed in the field. Samples were oven-dried in the laboratory at 60° C to a constant weight. The assumption that biomass of blue grama was normally distributed was evaluated with the Anderson-Darling test statistic at α = 0.05 (Stevens 1974).

A Monte-Carlo simulation of 10,000 samples of size n = 10, 20, 30, 40, and 50 clipped plots were drawn from the n = 100 plots by simple random sampling without replacement. For each sample, estimates of the average clipped biomass (g 0.25 m²) were computed using regression and ratio estimation.

In regression, the estimate of the clipped mean over all plots, \bar{y}_r is given by

$$\bar{y}_r = \bar{y} + B(\bar{x}' - \bar{x}) \quad (1)$$

where \bar{y} is the mean of clipped subsample plots, \bar{x} is the mean of the visually estimated subsample, and \bar{x}' is the mean of visually estimated weights in all 100 plots.

In ratio estimation,

$$\bar{y}_r = R\bar{x}' \quad (2)$$

Research was funded by the Colorado State University Agricultural Experiment Station, Projects 155453 (2202) and 156602 (4241).
Manuscript accepted 7 Mar. 1992.

Table 1. Descriptive statistics for blue grama biomass (g • 0.25 m²) on 100 sample plots.

Aboveground biomass	Mean	Median	Minimum	Maximum	Variance	Anderson-Darling test statistic for normality
	----- (g/0.25 m ²) -----					
Visually estimated (fresh wt.)	4.29	3.0	0	19.0	17.28	7.059*
Clipped (fresh wt.)	4.95	4.0	0	23.0	23.62	6.162*
Oven-dry weight	3.36	3.0	0	15.5	9.40	2.864*

*Significant at α = 0.05.

where the ratio R is obtained from

$$R = \frac{\bar{y}}{\bar{x}} \quad (3)$$

The decision to use the ratio or regression estimator depends on the relationship between clipped and visually estimated biomass. If a line fitting the data does not pass through the origin, and the distribution of clipped biomass is about the same throughout the range of visually estimated biomass, the regression estimator is most appropriate. If a line fitting the data passes through or very close to the origin, and the distribution of clipped biomass is proportional to visually estimated biomass, the ratio estimator is more appropriate (Cochran 1977).

The bias for the 2 double-sample estimators was computed as the difference between the average of the 10,000 estimates of clipped biomass and the average clipped biomass on the 100, 0.25-m² circular plots. Classical estimates of variance for both procedures, given by Cochran (1977) and Francis et al. (1979), were computed. In addition, variance estimates for (1) and (2) were computed using jackknife and bootstrap procedures (Schreuder and Ouyan 1992).

The jackknife procedure begins by removing one of the observations (i.e., plot) from the sample data. The desired statistic is then computed each time, with one of the observations eliminated (Smith and van Belle 1984). Standard error is then computed from the variability among these estimated values (Sokal and Rohlf 1981).

The bootstrap procedure requires random observations to be generated from the sample data. A subsample size of K is randomly selected with replacements from the n sample plots. This creates the bootstrap sample. The desired statistic is estimated based on the bootstrap sample. This is repeated N times to obtain N estimates of the average clipped biomass (Smith and van Belle 1984). The variance of the mean is then computed as the variance among the N estimates of an average clipped biomass.

The average variance of the 2 estimation methods was computed by averaging the variance estimates associated with the classical, jackknife, and bootstrap procedures. In addition, the variance of the mean (i.e., simulation variance) was computed as the variance among the 10,000 estimates of average clipped biomass using the 2 double-sample estimators. By definition, the variance of the mean is the best estimate of variability and can be used to evaluate whether variance formulae provide unbiased estimates. A 95% confidence interval for each estimate with the respective appropriate standard error of estimate was computed, and the proportion of confidence intervals enclosing the true population mean was determined.

Results and Discussion

Basic statistics for the data are given in Table 1. Tests for equality of variances (F-test) indicated that estimated and clipped weights had similar variances, while the variance of oven-dry weights was significantly smaller ($\alpha=0.05$). The Anderson-Darling test statistic confirmed that all 3 biomass distributions (estimated, clipped, oven-dried) departed significantly from normality ($\alpha=0.05$) (Table 1). The relationship between clipped and visually estimated biomass was linear through the origin. However, the ratio of clipped to visually estimated biomass varies somewhat instead of being constant over the range of estimated biomass, indicating that either the ratio or regression estimates could be used for estimating average aboveground biomass.

Both ratio and regression estimators underestimated clipped biomass of blue grama. The regression estimator consistently had a larger bias than the ratio estimator (Table 2). However, the bias for

Table 2. Effects of sample size on percent bias of estimates of aboveground biomass of blue grama for double sampling with ratio and regression estimators.

Clipped sample size	Ratio	Regression
10	-2.7	-4.4
20	-3.2	-3.6
30	-2.7	-3.0
40	-2.4	-2.6
50	-1.9	-1.9

¹Population mean = 4.95 g • 0.25 m².

the regression estimator decreased linearly as the clipped sample size increased, while the bias for the ratio estimator was non-linear over sample size.

Because both ratio and regression estimators were biased, one way to compare the variance of 2 estimators with different amounts of bias is to use the mean squared error (MSE), which is defined as the "variance + bias²" (Cochran 1977). Thus, an estimator with a smaller MSE is considered more precise than one with a larger MSE, even though the latter may have a smaller variance.

In terms of the MSE, the ratio estimator was always a better estimator than the regression estimator (Table 3). Both the jack-

Table 3. Mean squared error of the classical, jackknife (J), and bootstrap (B) variance estimates of aboveground biomass of blue grama.

Estimator	Clipped sample size				
	10	20	30	40	50
Ratio	0.923	0.573	0.442	0.368	0.311
Ratio-J	1.221	0.709	0.545	0.472	0.413
Ratio-B	1.085	0.705	0.524	0.467	0.410
Regression	1.038	0.598	0.447	0.379	0.314
Regression-J	2.025	0.771	0.566	0.485	0.418
Regression-B	1.335	0.756	0.542	0.480	0.414

knife and bootstrap estimates of the sample variance yielded larger MSE than the classical method for variance estimates. The jackknife and bootstrap estimates of the variance had a smaller MSE for the ratio method than for their regression counterparts.

Another method of evaluating the effect of bias on the statistical properties of the ratio and regression estimator is to examine the relative error, which is defined as the bias divided by the root mean squared error (bias/√MSE). A large relative error has the effect of distorting confidence probabilities (Cochran 1977). For example, with a bias/√MSE = 0.20, the actual confidence level associated with a nominal 95% confidence interval is about 0.9454.

The relative error was less than 0.15 in 5 out of 15 cases for the ratio estimator (Table 4), and less than 0.20 in 13 out of 15 cases (Table 4). The regression estimator had only 2 out of 15 variance

Table 4. Sample bias relative to the root mean square error of the classical, jackknife (J), and bootstrap (B) variance estimates of aboveground biomass of blue grama.

Estimator	Clipped sample size				
	10	20	30	40	50
Ratio	0.138	0.208	0.204	0.199	0.168
Ratio-J	0.120	0.187	0.184	0.176	0.146
Ratio-B	0.127	0.187	0.187	0.176	0.140
Regression	0.215	0.232	0.224	0.208	0.169
Regression-J	0.154	0.205	0.199	0.184	0.146
Regression-B	0.190	0.207	0.204	0.185	0.147

estimates with a relative error less than 0.15, and 8 out of 15 less than 0.20. The bias relative to the root mean square error also approximates a curvilinear relationship with respect to the number of clipped plots. In addition, the jackknife and bootstrap variance estimates had a smaller relative error across all sample sizes compared to the classical variance estimators (Table 4). The lower relative error results from variance overestimation by the jackknife and bootstrap estimates, thus reducing the effect of the bias on the width of the confidence interval. Similarly, the lower relative error associated with the ratio estimator results from the smaller bias and MSE, compared to the regression estimator.

Ratio of classical, jackknife, and bootstrap variance estimates to the simulation variance are given in Table 5. A ratio less than 1

Table 5. Ratio of mean variance of the classical, jackknife (J), and bootstrap (B) variance estimates to variance of the mean for the ratio and regression estimator.

Estimator	Clipped sample size				
	10	20	30	40	50
Ratio	0.674	0.982	0.916	0.936	1.202
Ratio-J	0.898	1.233	1.144	1.216	1.611
Ratio-B	0.796	1.224	1.097	1.210	1.601
Regression	0.624	0.991	0.896	1.015	1.164
Regression-J	1.265	1.303	1.154	1.320	1.569
Regression-B	0.817	1.277	1.103	1.307	1.553

indicates an underestimation of the variance, while a ratio greater than 1 indicates an overestimation. Using this as a guideline, one can see that the classical regression estimator provides the best estimate of the variance across all sample sizes tested. Next best estimates were provided by the classical ratio estimator. In particular, for sample sizes of $n \geq 20$, the jackknife and bootstrap variance estimate overestimated the true variance for both the ratio and regression estimators. On the other hand, for a sample size, $n = 10$, the jackknife ratio estimate provided the best estimate of the variance, while the bootstrap regression estimate provided the second best estimate.

Francis et al. (1979) also used simulation to compare sample variances of regression and ratio estimators to theoretical variances and found that the regression estimate provided the minimum variance. Their finding is the opposite of what we observed in our study. This difference is due to the relationship between clipped and estimated biomass. In our study, this relationship was linear through the origin, and the variance of clipped biomass tended to be proportional to estimated biomass. In their study, the relationship between clipped and estimated biomass was also linear, but not necessarily one that passed through the origin, which resulted in the regression estimator being more efficient than the ratio estimator (Cochran 1977). The previous authors also noted that if bias is present in the estimation process, the difference between the theoretical and sample variance is more pronounced when the ratio estimate is used than when the regression estimate is used.

Furthermore, these authors noted that when the variance of the weight estimate is constant over the range of clipped values, double sampling significantly reduces the variance of the estimate over that of clipping only if the cost is constant. In their study, for sample sizes $n \leq 30$, double sampling reduced the variance by a factor of 2 over that of clipping only, and by a factor of 1.5 for sample sizes $n > 30$.

Table 6. Relative coverage of population mean for 95% confidence limits for the classical, jackknife (J), and bootstrap (B) variance estimates of aboveground biomass of blue grama.

Estimator	Clipped sample size				
	10	20	30	40	50
Ratio	0.946	0.942	0.893	0.930	0.946
Ratio-J	0.946	0.969	0.955	0.953	0.986
Ratio-B	0.865	0.969	0.933	0.953	0.986
Regression	0.919	0.942	0.899	0.953	0.933
Regression-J	0.946	0.960	0.955	1.000	0.986
Regression-B	0.892	0.973	0.944	0.977	0.986

The actual coverage rates of nominal 95% confidence intervals for the ratio and regression estimator, using classical, jackknife, and bootstrap variance estimates, are given in Table 6. All variance estimates provided reasonable coverage rates, especially for small sample sizes ($n = 10$). The classical variance estimates consistently had coverage rates less than 0.95, while the jackknife estimator had coverage rates exceeding 0.95. The bootstrap estimate had the lowest coverage rate for both the ratio and regression estimator at $n = 10$.

Finally, it should be mentioned that the results of this study may not be applicable in stands in which blue grama is a minor component of the species mix. In such instances, the regression estimator may be more appropriate.

Recommendations

1. Fresh weights, instead of oven-dry weights, should be used with visually estimated weights in double sampling because of the similarities in their distribution. If needed, average fresh weights can be converted to oven-dry weights.
2. The ratio estimator should be used for estimating average clipped weight in stands dominated by blue grama in eastern Colorado in preference to the regression estimator.
3. For $n = 10$ clipped plots, the jackknife variance estimate is recommended for estimated confidence intervals. For $n \geq 20$ clipped plots, the classical variance estimates should be used to obtain reliable estimates of the population variance and to estimate confidence intervals.

Literature Cited

- Cochran, W.G. 1977. Sampling techniques. John Wiley & Sons, Inc., N.Y.
- Francis, R.C., G.M. Van Dyne, and B.K. Williams. 1979. An evaluation of weight estimation double sampling as a method of botanical analysis. *J. Environ. Manage.* 8:55-72.
- Pechanec, J.F., and G.D. Pickford. 1937. A weight estimate method for determination of range or pasture production. *J. Amer. Soc. Agron.* 29:894-904.
- Smith, E.P., and G. van Belle. 1984. Nonparametric estimation of species richness. *Biometrics* 40:119-129.
- Sokal, R.R., and F.J. Rohlf. 1981. *Biometry*. Freeman, San Francisco, p. 795.
- Stevens, M.A. 1974. EDF statistics for goodness of fit and some comparisons. *J. Amer. Sta. Assoc.* 69:730-737.
- Scheaffer, R.L., W. Mendenhall, and L. Ott. 1990. *Elementary survey sampling*. DWS-Keet Pub. Co., Boston, Mass.
- Schreuder, H.T., and Z. Ouyang. 1992. Optimal sampling strategies for weighted linear regression estimation. *Can. J. Forest. Res.* (In press).