

Combination of Weight Estimates with Clipped Sample Data

VILMA CARANDE AND DONALD A. JAMESON

Abstract

Weight estimation is a common procedure to determine range forage production. In this method clipped samples are used to train an observer, to make periodic checks on observer performances, and to compute estimated/clipped conversion factors. The clipped sample data are then discarded. This is in contrast to a double sampling procedure where both clipped values and estimated values are used in computation of sample variance. However, if both estimated values and clipped plots are taken at random, they can be combined to compute sample means and variances by using techniques appropriate to combining data of different variances. The efficiency of the combined sample appears to be greater than that of formal double sampling and also has the advantage that plots clipped for training and checking can also be used as part of the sample.

Sampling for forage production occupies a very important position in range research studies. The choices of basic approaches are few: (1) clipping, (2) some "indirect" method such as weight estimation, or (3) a combination of (1) or (2).

Clipping is generally considered to be the superior method but has 2 commonly understood drawbacks: (1) it may be very time consuming compared to indirect methods and (2) it is destructive. In addition, clipping itself, although commonly accepted as the least prone to measurement errors, does, in fact, contain such errors (in addition to sampling errors).

For indirect methods this discussion will focus on weight estimation, although the concepts also apply to other "indirect" methods. Weight estimation is often preferred to clipping because it is both less time consuming and nondestructive (Pechanec and Pickford 1937). It is generally understood that the weight estimates are less exact than clipped samples.

In order to capture some of the advantages of both, weight estimation and clipping are sometimes combined in "double sampling." Double sampling was first discussed in the range management literature by Wilm et al. (1944) and more recently by Francis et al. (1979) and Ahmed et al. (1983). Double sampling uses an optimum ratio of estimated to clipped plots, considering both the relative costs and relative variances of the 2 methods.

Whether the method used is double sampling or weight estimation alone, the weight estimates are converted to either dry or wet clipped weight equivalent by computing a linear regression or ratio. For linear regression

$$y = a + bx \quad (1)$$

where y is the clipped weight, x is the estimated weight, a is the value of y when $x = 0$, and b is the slope of the regression line. For ratio sampling a is equated to 0 and

$$y = Rx \quad (2)$$

where R is the ratio estimator.

Equation (1) always results in a lower variance estimate than Equation (2) (Ahmed et al. 1983), but nonzero values of the constant a may create difficulties. Therefore, linear regression is recommended for species (or other taxa) with few zero estimates, and ratio sampling is recommended when zero estimates are com-

mon enough that assigning the value a to plots estimated as zero would be bothersome (Francis et al. 1979).

Although theoretically sound, the double sampling procedure has some drawbacks. If the optimum ratio of estimated plots to clipped plots is to be calculated, a presample must be collected to compute the needed variances. Generally a standardized ratio rather than an optimum ratio of estimated plots per clipped plot is used for each study; this will, of course, result in some loss in efficiency. The computations require a strict discipline in proportion of plots clipped, and plots clipped for training purposes are not used in the statistical calculations.

Perhaps more commonly used than double sampling is weight estimation where the estimates are controlled by reference to clipped plots. In this method some plots are both clipped and estimated in order to train the observers. After training is judged to be sufficient, the clipped and estimated values are used to compute a linear regression or ratio correction. Although the clipped plots are used for training and for calculating the correction factor, they are not considered part of the actual sample. The sampling variance from the corrected estimates is considered to be appropriate for determining sample size.

Although clipping is usually considered to be the superior method, it is usually the case that for a set of plots that are both clipped and estimated the variance of the estimated plots S_e^2 is less than the variance of the clipped plots S_y^2 . Primarily, this is caused by the fact that most observers estimate a smaller range of values than the corresponding clipped values. In the extreme, an observer could always estimate a constant value, resulting in a zero variance for the corrected values, which is not a superior result. If low variance is equated with a superior method, an anomaly is thereby created: the use of weight estimates may result in a procedure where poor estimates are rewarded by a lower variance.

A major concern is to compute an appropriate variance for the estimated plots, designated here as S_e^2 . The variance for estimated plots S_e^2 should not be less than S_y^2 . An appropriate method for developing S_e^2 which meets this requirement is shown in the Appendix.

In the usual weight estimate method the clipped values are discarded once the correction factors are computed. This could represent a considerable loss of information, particularly if a great deal of clipping is needed for adequate training and control. If both the clipped values and the estimated values are random samples, both could be used; but each set of values may have a different variance. If the clipped plots are from a sample with variance S_y^2 and estimated plots are from a sample with variance assigned as S_e^2 , a variance appropriate to the pooled value of a single estimated plot and a single clipped plot can be computed by

$$\frac{1}{S_p^2} = \frac{1}{S_e^2} + \frac{1}{S_y^2} \quad (3)$$

where S_p^2 is the pooled variance (Maybeck 1979). For more than one datum value, S_e^2 and S_y^2 would be replaced by the appropriate square of the standard error of the mean. A weighting factor g can be calculated from S_e^2 and S_y^2 :

$$g = \frac{S_e^2}{S_e^2 + S_y^2} \quad (4.1)$$

Authors are instructor and professor, Universidad Nacional de Cordoba, CC509, 5000 Cordoba, Argentina; Department of Range Science, Colorado State University, Fort Collins 80523.

Manuscript accepted 28 May 1985.

The weighted mean using the two methods is computed by

$$x' = g\bar{y} + (1-g)\bar{x} \quad (4.2)$$

where \bar{y} is the mean of clipped plots, \bar{x} the mean of the estimated plots, and x' the combined best estimate (Gelb 1974, Maybeck 1979). To avoid double counting of plots that are both estimated and clipped, \bar{x} should include only those plots which do not have clipped values.

Numerical Example

Data for this example were collected in the oakbrush type of southwestern Colorado. One hundred eighteen plots were both clipped and estimated, and an additional 600 plots were estimated only. A ratio correction (i.e., $y = Rx$) was used.

For this example:

$$\text{variance of clipped plots; } S_y^2 = 1311 \quad (5.1)$$

$$\text{variance of estimated plots; } S_x^2 = 1221 \quad (5.2)$$

$$\text{variance of estimated plots independent of clipped plots; } S_{xy}^2 = 219 \quad (5.3)$$

calculated variance applied to estimated plots (using the Appendix);

$$S_o^2 = S_y^2 + S_{xy}^2 = 1530 \quad (5.4)$$

Based on the equations of Francis et al. (1979) and Cochran (1977), the optimum ratio of estimated to clipped plots was 6.8. Taken to the nearest integer, the optimum ratio would have been 7:1. The appropriate variance from Francis et al. (1979) is 389. (The equations for double sampling calculations are not shown here because they are not part of the method being described; double sampling results are only included for comparison.)

For comparison purposes the same ratio of clipped to estimated plots will be used in the combined method (in practice a constant ratio does not have to be maintained. For plots estimated but not clipped, the variance S_o^2 is 1530 (Equation 5.4). For 6 such plots the appropriate square of the standard error of the mean is $1530/6 = 255$. From Equation (4.1) the weighting factor g to combine the 6 estimated (but not clipped) plots with 1 clipped plot is

$$g = \frac{255}{255 + 1311} = 0.163 \quad (6)$$

The pooled variance of clipped and estimated plots from Equation (3) is

$$\frac{1}{S_p^2} = \frac{1}{255} + \frac{1}{1311} = .004684 \quad (7.1)$$

$$S_p^2 = 213 \quad (7.2)$$

If all 7 plots had been clipped, the appropriate square of the standard error of the mean would have been $1311/7 = 187$. Assuming that the cost of clipping a plot was 10 units of time and the cost of estimating a plot was 1 unit of time, the following error terms and costs can be compared:

Type of Sampling	Error Term	Cost
7 estimated plots using S_x^2/n	175	7
7 clipped plots using S_y^2/n	187	70
1 clipped plot and 6 estimated plots combined	213	17
(the clipped plot is also estimated)		
double sampling, 7:1 ratio	389	17

Conclusion

Both the estimated values and clipped control plots of the weight estimation procedure can be used to compute sample means and variances using techniques appropriate to combining data of different variances. In the numerical example analyzed here, the

combination of the clipped and estimated plots resulted in an error term which is only slightly higher than that of clipping alone and considerably less than the error term from double sampling. Since the variance of the estimated plots is assumed to be $(S_y^2 + S_{xy}^2)$, the error term is more conservative than would have been the case had only clipping been used. Thus, the combination of clipped and estimated plots results in a procedure that has a relatively low cost and a conservative estimate of variance.

Appendix

Proportioning S_x^2 for the linear regression method is shown by:

$$S_{xy}^2 = (1-r^2)S_x^2 \quad (A1)$$

where S_{xy}^2 is the variance of x independent of y and r is the correlation coefficient between clipped and estimated plots. For the ratio method the equations of Francis et al. (1979) and Cochran (1977) are modified to find the variance of x independent of y :

$$S_{xy}^2 = S_x^2 - 2RS_{yx} + R^2S_y^2 \quad (A2)$$

where R is the ratio estimator for $x = Ry$ and S_{xy}^2 is the covariance between x and y .

We propose that an appropriate calculation of variance for estimated plots is a combined variance term S_o^2 , where S_o^2 is defined by

$$S_o^2 = S_y^2 + S_{xy}^2 \quad (A3)$$

Equation (A3) will have the following results for a "perfect" estimate:

$$S_{xy}^2 \rightarrow 0 \text{ and } S_o^2 \rightarrow S_y^2 \quad (A4)$$

for constant clipped plot weights:

$$S_y^2 \rightarrow 0, S_{xy}^2 \rightarrow S_x^2 \text{ and } S_o^2 \rightarrow S_x^2 \quad (A5)$$

and for a completely "useless" estimate:

$$S_{xy}^2 \rightarrow S_x^2 \text{ and } S_o^2 \rightarrow S_x^2 + S_y^2 \quad (A6)$$

Equation (A4) indicates that as the agreement between clipped and estimated values improves, S_o^2 will more closely approach S_y^2 . Equation (A5) indicates that as the clipped plots more closely approach a constant, the variance of the estimated values around the constant clipped weight will remain and S_{xy}^2 will approach S_x^2 . Equation (A6) would provide a large penalty for a completely "useless" estimate, but this condition would seldom be expected. Under any conditions S_o^2 calculated by Equation (A3) provides no reward for having poor estimates and is more conservative than using either S_x^2 or S_y^2 .

Literature Cited

- Ahmed, J., C.D. Bonham, and W.A. Laycock. 1983. Comparison of techniques used for adjusting biomass estimates by double sampling. *J. Range Manage.* 36:217-221.
- Cochran, W.G. 1977. Sampling techniques. John Wiley and Sons.
- 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. Envir. Manage.* 8:55-72.
- Gelb, Arthur (ed.). 1974. Applied optimal estimation. MIT Press, Cambridge.
- Maybeck, P.S. 1979. Stochastic models, estimation and control. Academic Press Series on Mathematics in Science and Engineering. Vol. 141.
- 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.
- Wilm, H.G., D.F. Costello, and G.E. Klipple. 1944. Estimating forage yield by the double sampling method. *J. Amer. Soc. Agron.* 36:194-203.