SAMPLE SIZE DETERMINATION FOR ESTIMATING AMOUNTS OF AVAILABLE WHEAT FORAGE

R.W. McNew[1], G.W. Horn[2], E.K. Kocher[3] and G.J. Vogel[4]

Story in Brief

Data from six years of wheat pasture grazing studies were used to develop a model describing the sampling variation of amounts of available forage dry matter. The model was used to determine the sample size required to achieve a maximum relative error in the estimate of forage dry matter availability. The variation among samples from a pasture had a coefficient of variation that was constant across pastures, years and sampling dates. The model developed used normality and a constant coefficient of variation of .3. Based on this model, a maximum 10% relative error of the estimated pasture mean could be achieved with a 50% chance from a sample size of four. A sample size of 36 would be required to increase this chance to 95%.

(Key Words: Sampling Variability, Coefficient of Variation.)

Introduction

The evaluation of performance of grazing animals in nutrition trials frequently requires estimates of the quantity of available forage. To this end, forage samples are collected during the course of the research. It would be useful to know how much sampling effort is required. This is the standard "sample size" question that should be addressed in planning an experiment. Answering this question requires selection of a criterion to be met with the sample and knowledge of the magnitude of sampling variation.

In this report, we consider the problem of sample size determination for estimating the amount of forage dry matter (DM) of wheat pasture. Using data collected on forage samples, we attempt to identify a reasonable model describing variation in samples of forage DM availability from wheat pasture.

---

[1]Professor, Statistics [2]Professor [3]Graduate Student, Statistics
[4]Graduate Student

We then use this model, and the criterion of maximum relative error, to determine a sample size requirement.

Materials and Methods

The data set used in this report is from stocker cattle grazing studies on wheat pasture that were conducted at the Forage and Livestock Research Laboratory (USDA/ARS), El Reno, Oklahoma from 1981 to 1988 (Vogel, 1985 and 1988). Data from forage samples were available to us for each year except the 1984-85 pasture season. Eight pastures were used in each season except in 1982-83 when there were four pastures. Pasture size was approximately 10 ha. Usually, three samples were taken from each pasture on each sampling date. Sampling dates, three to five per pasture season, were chosen from different growing periods of the pasture season. Details of the data source are presented in Table 1. There were a total of 164 samples (pasture-date) available.

A sample consisted of forage in a .5 m^2 circular area, removed by hand clipping at ground level. Each sample was oven-dried, after which dry weight was measured. In this report, the units of dry matter are $g \cdot .5m^{-2}$. Multiplication by 20 converts this to $kg \cdot ha^{-1}$.

In the following analysis, we treat each pasture-date as a population from which we have a random sample of dry weights. Populations are expected to have different amounts of dry matter per unit area. The within-population variance is not assumed to be constant.

Results and Discussion

The practical objective of forage sampling is to estimate the forage dry matter per unit area in a pasture. The quality of this estimate depends on the magnitude of sampling variation and the sample size (number of forage samples observed). Assessing this quality depends on the combination of these factors and the form of the distribution of the estimate. Because of the

Table 1. Source of forage dry weight data used for analysis.

	Pasture-Season					
	1981-1982	1982-1983	1983-1984	1985-1986	1986-1987	1987-1988
No. of pastures	8	4	8	8	8	8
No. of sampling dates	4	3	5	3	4	3
Sample size	3	5	3	3	3	2,3

Central Limit theorem, the distribution of the sample mean may be approximated by the Normal distribution, even for small sample sizes if the population distribution is symmetric. Our approach was to assume normality unless the data indicated extreme skewness. When the sample size is small, it is not possible to have much opportunity to identify a distribution form. Therefore, our attempt to use the data proceeded in the following way. Using only the 155 samples of size 3 or more, the mean and standard deviation were obtained. The first value in each sample was standardized and the resulting 155 values were used to look for skewness. A histogram (Figure 1) of the distribution appeared quite symmetric and the measure of skewness was nearly 0. On this basis, we chose to retain the normality assumption.

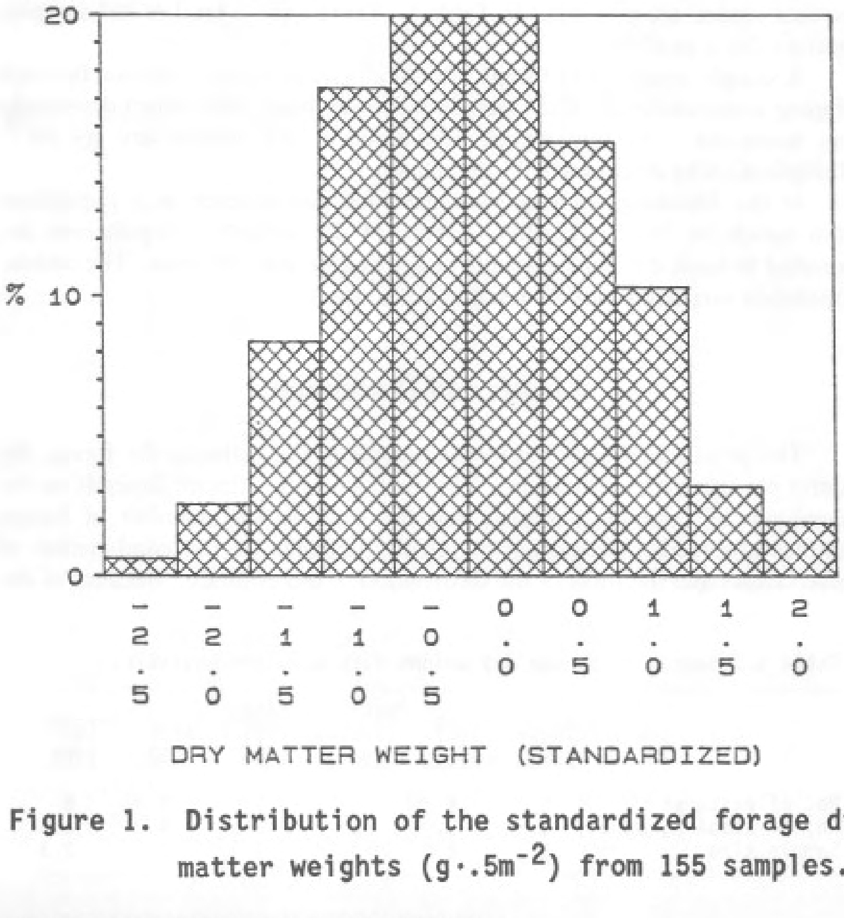

Figure 1. Distribution of the standardized forage dry matter weights ($g \cdot .5m^{-2}$) from 155 samples.

The standard deviations from the samples were observed to be quite variable. One might expect that the magnitude of the sample variation will increase as the mean level of forage dry matter increases. The plot (Figure 2) of the sample standard deviation against the sample mean supports this anticipated pattern. Furthermore, this plot suggests a linear relationship between the standard deviation and the mean, and that this linear relation extends through the origin. This pattern is consistent with a constant coefficient of variation (CV: ratio of standard deviation to mean). The straight line through the scatter plot in Figure 2 is the regression line with zero intercept, the fitting being done by ordinary least squares. The slope of this line is an estimate of the constant coefficient of variation.

Additional methods of checking for a constant CV and of estimating its value were attempted. These included other least squares fits, examination

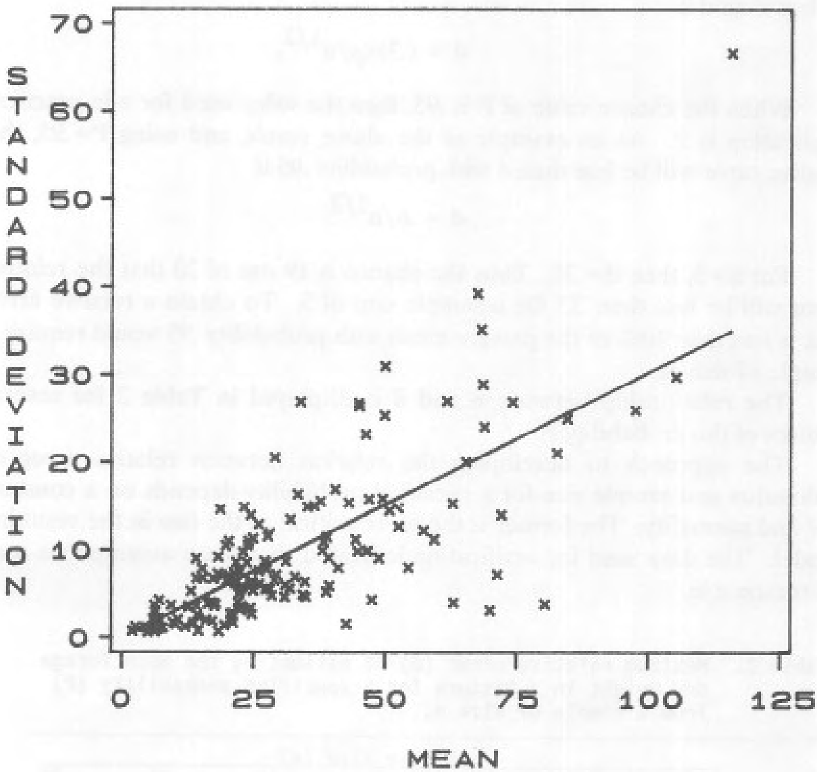

Figure 2. Plot of the standard deviation against the mean of forage dry weights ($g \cdot .5m^{-2}$) from 164 samples.

of residuals and maximum likelihood estimation. The results did not indicate a serious error in the assumption of a constant CV. The range of estimates from the different methods was .28 to .32.

The above analysis led us to the following model: The CV for sample variation is constant and the distribution of the sample mean is normal. In numerical evaluations to follow, we use CV=.30. From this model, we can obtain the relation between sample size and a desired precision by which the same mean estimates the mean dry weight ($g \cdot .5m^{-2}$) for the entire pasture. The pattern of this development is similar to that in Snedecor and Cochran (1980, pp. 441-442). The term relative error will refer to the absolute error of estimating the population mean as a fraction of the population mean. For a probability P, let z_P denote the 100(1+P)/2 percentile of the standard normal distribution. Then, for a sample of size n from a pasture, the probability is P that the relative error of an estimate of the pasture mean will be less than d if

$$d = (.3)z_P/n^{1/2}.$$

When the chosen value of P is .95, then the value used for z in practical application is 2. As an example of the above result, and using P=.95, the relative error will be less than d with probability .95 if

$$d = .6/n^{1/2}.$$

For n=3, then d=.35. Thus the chance is 19 out of 20 that the relative error will be less than .35 for a sample size of 3. To obtain a relative error that is less than 10% of the pasture mean with probability .95 would require a sample of size 36.

The relationship between n and d is displayed in Table 2 for several choices of the probability P.

The approach to developing the relation between relative error of estimation and sample size for a specified probability depends on a constant CV and normality. The former is the more critical of the two in the resulting model. The data used for verification indicated that these assumptions may be reasonable.

Table 2. Maximum relative error (d) of estimating the mean forage dry weight in a pasture for a specified probability (P) from a sample of size n.

	Sample size (n)						
P	3	4	6	9	16	25	36
.50	.12	.10	.08	.07	.05	.04	.03
.75	.18	.16	.13	.10	.08	.06	.05
.90	.29	.25	.20	.16	.12	.10	.08
.95	.35	.30	.25	.20	.15	.12	.10

The results are most useful when interest lies in relative error rather than absolute error. They permit determination of the sample size for achieving an upper limit to the relative error for a specified probability. They can also be used to evaluate the likely relative error for a chosen sample size.

If one wishes to have at most a 10% relative error in the estimate, then there is a 1 in 2 chance of achieving this when a sample size of 4 is used. With larger sample sizes of 9, 25, and 36, the chances increase to 3 in 4, 9 in 10, and 19 in 20, respectively.

Literature Cited

Snedecor, G.W. and W.G. Cochran. 1980. Statistical Methods (Seventh Edition). Iowa State University Press. Ames, IA.

Vogel, G.J. 1985. Influence of supplemental silage on forage intake and utilization and performance of stocker cattle grazing wheat pasture and bermudagrass. M.S. Thesis. Oklahoma State Univ., Stillwater.

Vogel, G.J. 1988. Kinetics of ruminal nitrogen digestion of wheat forage and high protein feedstuffs and the effects of supplemental protein on the performance of growing cattle on wheat pasture. Ph.D. Dissertation. Oklahoma State Univ., Stillwater.