*Applied Animal Behaviour Science*, 30 (1991) 17–26
Elsevier Science Publishers B.V., Amsterdam



# Sample size determinations for studying selected cattle foraging behaviors*

K.M. Havstad[1] and K.M. Olson-Rutz[2]

[1]USDA-ARS, Jornada Experimental Range, Box 30003-NMSU-Dept. 3JER, Las Cruces, NM 88003 (U.S.A.)

[2]Department of Animal and Range Sciences, Montana State University, Bozeman, MT 59717 (U.S.A.)

(Accepted for publication 6 August 1990)

ABSTRACT

Havstad, K.M. and Olson-Rutz, K.M., 1991. Sample size determinations for studying selected cattle foraging behaviors. *Appl. Anim. Behav. Sci.*, 30: 17–26.

Experimentation on livestock foraging behaviors under free-ranging conditions is frequently constrained by few experimental units (animals). Effective research designs require knowledge of expected sampling variability, but this information relative to range foraging behaviors is lacking. Sixteen 5-year-old range beef cows were used to estimate daily grazing time (DGT) and fecal output (percentage of body weight, FOBW) for 46 consecutive days during a winter grazing season. Estimates of components of variability associated with cows, days, and theoretical treatment effects were constructed from these data. Implied sample size requirements for a half-confidence-width $\leqslant 10\%$ of the mean and a Type 1 significance level of 0.05 were 7 and 9 cows for DGT and FOBW, respectively, with day as a fixed effect. For day as a random effect, implied sample sizes were 5 days and 5 animals for DGT, and 5 days and 9 animals for FOBW. Variability among animals was greater when estimating FOBW than for DGT estimates. Detecting differences between two treatments sampled over 5 days with Type 1 and Type 2 probabilities of 0.05 and 0.10, respectively, implied sample sizes of five and $>20$ cows for DGT and FOBW estimates, respectively. In general, these data indicated that $<5$ animals was insufficient and $>20$ was excessive. These sample size estimates were obtained from a relatively uniform set of animals grazing a uniform quality and quantity of forage. Estimates of N would have been greater under more variable experimental conditions. Given constraints of statistics and the logistics of range nutrition experimentation, unless sample size requirements can be satisfied it may not be appropriate to examine certain questions with current experimental methodologies.

## INTRODUCTION

Hypotheses tests of herbivore foraging traits under free-ranging conditions are frequently restricted by limited experimental units. Yet, the precision of estimated mean population characteristics partially relies on adequate sam-

*Contribution published with the approval of the Director, Montana Agricultural Experiment Station, Journal Series J2402.

0168-1591/91/$03.50 

ple numbers. Sample size requirements are dictated by population variance and the desired precision of estimation. Estimates of the former may be obtained from published information or preliminary experiments. The latter are selected by the investigator. Generally, the criteria of cost, accepted practice, and availability of experimental units dictate sampling procedures (Berndtson, 1989). Prior knowledge of sample size required to obtain a desired degree of precision can clarify requirements of experimental design and reveal potential limitations of data interpretation.

Methods of determining daily grazing time (DGT) and fecal output (percentage of body weight, FOBW) of beef cows grazing winter rangeland in Montana, U.S.A., are presented. Sensitivities of mean estimates and detectable differences of DGT and FOBW to animal numbers and variance components are identified.

## ANIMALS, MATERIALS AND METHODS

Daily grazing time and FOBW were estimated as part of an ongoing winter grazing study (Beverlin et al., 1989) at the Red Bluff Research Ranch, Norris, MT, operated by the Montana Agricultural Experiment Station. Sixteen 5-year-old pregnant Angus×Hereford or Tarentaise cows (525–575 kg) grazed a 324-ha perennial grassland range from 2 December 1986 to 28 February 1987. The study area was 65% grasses (principally dominated by *Agropyron spicatum*, *Stipa comata*, *Festuca idahoensis* and *Elymus cinerius*) and 35% forb and woody species (Ross and Hunter, 1976). Elevations range from 1400 to 1900 m with average annual precipitation ranging from 350 to 406 mm (Ross and Hunter, 1976). Dominant southwest winds remove snow and expose forage throughout winter. The study site is described further in Ross and Hunter (1976).

The cows grazed native range for 6 months before the trial. They received no winter supplements but had access to mineral salt. Each cow was fitted with a vibracorder (Stobbs, 1970) to collect DGT from 8 January to 26 February 1987. Charts were changed weekly. Grazing time within the intervals 07:01–13:00, 13:01–19:00 and 19:01–07:00 h were rounded to 15 min and recorded. Interval times were summed to estimate DGT.

Fecal output was estimated daily for each cow from 11 January to 26 February 1987 using a $Cr_2O_3$ dilution technique (Raleigh et al., 1980). All 16 cows were bolused with 10 g $Cr_2O_3$ and rectal fecal grab samples were collected daily. Daily fecal output (FO) per day was estimated from rectal grab samples based on the following equation: g FO day$^{-1}$ = [ (g $Cr_2O_3$ fed day$^{-1}$)/ ($Cr_2O_3$(%) in dry fecal sample) ]. Rectal grab samples were adjusted individually based on FO from total fecal collections collected from each cow for 4 consecutive days in both January and February. Cows were weighed twice weekly.

Means and variance components of cows, days, and error for both DGT and FOBW were estimated with the model

$$Y_{ij} = \mu + \text{Cow}_i + \text{Day}_{(j)i} + \text{Error}$$

(SAS General Linear Models procedure (Statistical Analysis Systems, 1985)). The half width of a confidence interval ($W$) and the detectable difference ($\delta$) among hypothetical treatments were calculated as a function of animal numbers for three potential research objectives.

When the research objective is to estimate animal response (DGT or FOBW) to a specific day, day becomes a fixed effect. Estimation of DGT response to daily temperature is an example of this objective. Equation (1) defines the relationship of number of animals sampled ($N$) to the precision of mean estimates ($W$) (Snedecor and Cochran, 1967).

$$W = Z_{\alpha/2} \cdot [(s_c^2 + s_e^2)/N]^{0.5} \tag{1}$$

Estimates of variance components associated with cow and error are $s_c^2$ and $s_e^2$, respectively, and $Z$ follows the standard normal distribution. Preliminary calculations use $Z$, rather than $t$, because a lack of precision in determining sums of squares diminished the importance of degrees of freedom for calculating $s$. However, $Z$ produces narrower half-confidence-widths than $t$. Snedecor and Cochran (1967) recommend using $Z$ rather than $t$ and adding one or two extra sample units for a more conservative estimate of the $N$ required.

An alternative research objective might be to estimate animal response over time; for example, while on winter range. Day is a random variable for this inference and must be included when calculating the precision of mean estimates as a function of $N$ (eqn. (2)) (Snedecor and Cochran, 1967).

$$W = Z_{\alpha/2} \cdot [(s_c^2/N) + (s_d^2/M) + (s_e^2/NM)]^{0.5} \tag{2}$$

$M$ is the number of sample days and $s_d^2$ is the variance estimate associated with day.

A third possible research objective would be to test for a response difference among treatments. Influence of different dietary supplements on DGT is an example of this situation. If treatments are applied across the same number of days to different cows, then eqn. (3) describes the difference detectable among treatments as a function of animal numbers where $N$ is sample size per treatment group. The Type 2 probability is $1-\beta$ (Snedecor and Cochran, 1967).

$$\delta^2 = 2(Z_{\alpha/2} + Z_\beta)^2 \cdot [s_c^2 + (s_e^2/M)]/N \tag{3}$$

This equation has an inherent limitation in determination of sample size. A desired width of $\delta$ cannot be guaranteed with selected probability unless a prior estimate of population variance (with known degrees of freedom) is available (Li, 1964). Estimates of variance associated with foraging behavior measurements are generally not published. However, as these estimates become available, other methods for determining sample size, such as described by Li (1964), can be substituted for eqn. (3) and provide a guaranteed $\delta$ width with known probability.

To facilitate interpretation of figures, all $s$ terms were replaced by coefficients of variation ($CV=100s$/mean) prior to calculation of $W$ and $\delta$. Both $W$ and $\delta$ are also expressed as percentages of the means. Unless otherwise stated, $\alpha=0.05$ and $\beta=0.10$.

## RESULTS

Coefficients of variation for cow ($CV_c$), day ($CV_d$), and error ($CV_e$) for DGT and FOBW were 2.5, 9.3, and 12.8, and 12.8, 5.6, and 7.4, respectively. The half-confidence-width around a mean, for a fixed day, varies with $N$ for different CVs (eqn. (1), Fig. 1). The total variation ($s_c^2+s_e^2$) determines the precision of the estimated mean for a given number of animals. The ratio of $CV_c$ to $CV_e$ is immaterial. Under these study conditions, we would need at

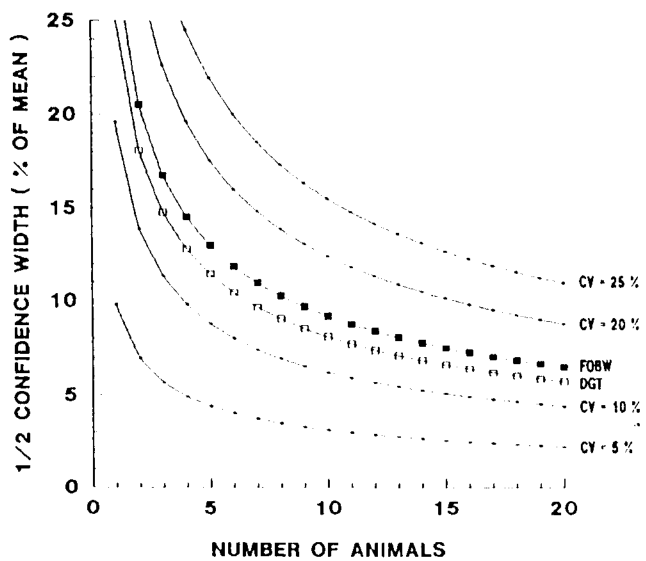


Fig. 1. One-half confidence widths (as a percentage of the mean) for estimates of daily grazing time (DGT) and fecal output (percentage of body weight, FOBW) for a specific day as affected by number of experimental units. Relationship are expressed for several different possible coefficients of total variation (CV). Day is treated as a fixed effect. Observed CVs associated with cow ($CV_c$) and error ($CV_e$) for DGT and FOBW were 2.5 and 12.8%, and 12.8 and 7.4%, respectively.

least 9 cows to estimate mean daily FOBW within 10% with 0.95 confidence. Allowing the level of confidence to reduce to 0.90 implied a reduction of the number of cows necessary for measurement of FOBW to 8 and DGT to 6 to retain a *W* of 10%.

When sampling over several days, variance among days influences animal numbers required (eqn. (2)). From our DGT data, $CV_d$ was greater than $CV_c$. Therefore, adding more days of sampling can increase the precision of the mean estimate of DGT more than adding more cows to each sampling day (Fig. 2). For example, if we used 4 cows and increased the number of days from 3 to 7, we achieve a 10% half-confidence-width. When sampling days across large environmental gradients, indiscriminately increasing number of days may not reduce either $CV_d$ or experimental unit requirements.

If $CV_c$ is greater than $CV_d$ as with our FOBW data, then increasing cow numbers improve mean precision more than sampling over more days (Fig. 3). If we sampled with 11 cows for 3 days, our FOBW estimate would have a similar precision as sampling with 7 cows for 21 days. As stated above, the span of sampling days must be increased cautiously. Nine cows sampled for 5 days would result in precision equivalent to the preceding situation. The decision to include more animals or more days depends on the relative cost of

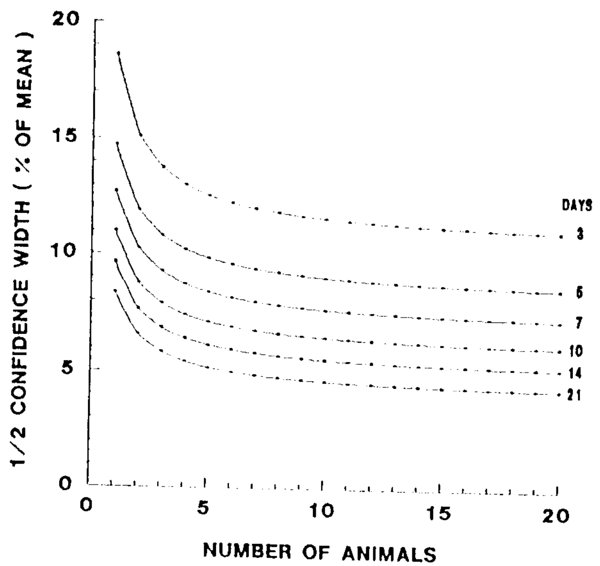


Fig. 2. One-half confidence widths (as a percentage of the mean) for estimates of daily grazing time (DGT) over a time interval as affected by number of experimental units. Day is treated as a random effect. Coefficients of variation for animal ($CV_c$), day ($CV_d$) and error ($CV_e$) were 2.5, 9.3 and 12.8%, respectively. Relationships are expressed for several different potential time spans. Type 1 significance level was 0.05.

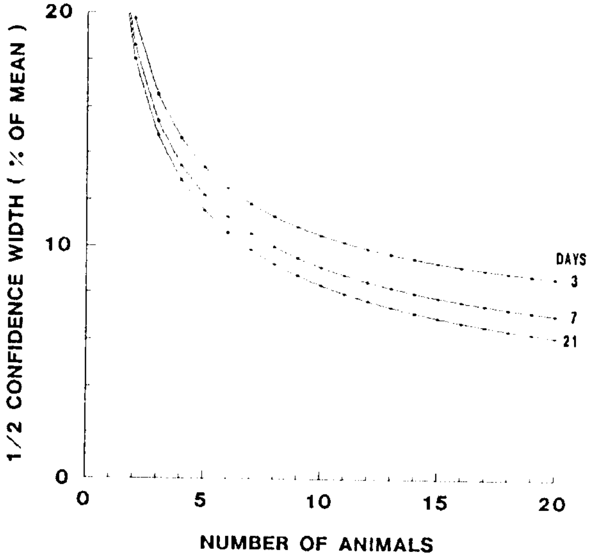


Fig. 3. One-half confidence widths (as a percentage of the mean) for estimates of daily fecal output (percentage of body weight, FOBW) over a time interval as affected by number of experimental units. Day is treated as a random effect. Coefficients of variation for animal ($CV_e$), day ($CV_d$) and error ($CV_e$) were 12.8, 5.6 and 7.4%, respectively. Relationships are expressed for several different time spans. Type 1 significance level was 0.05.

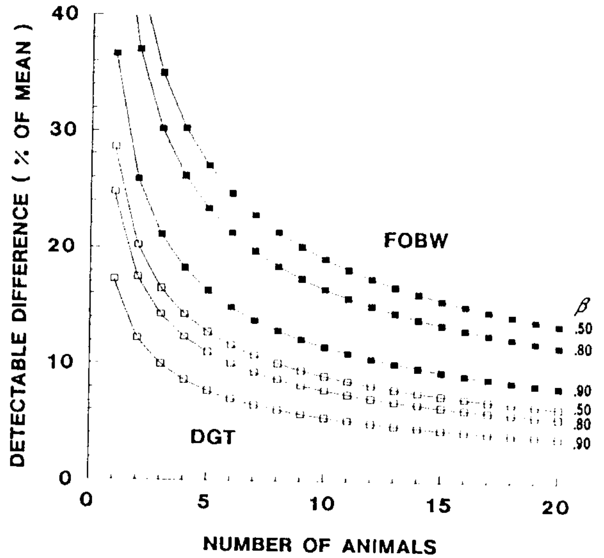


Fig. 4. Detectable differences (as a percentage of mean estimates) between two hypothetical treatments for daily grazing time (DGT) and fecal output (percentage of body weight, FOBW) over a time interval as affected by number of experimental units. Day is treated as a random effect. Relationships are expressed for a specific time interval (5 days), a Type 1 significance level of 0.05, and three Type 2 probabilities (0.10, 0.20 and 0.50).

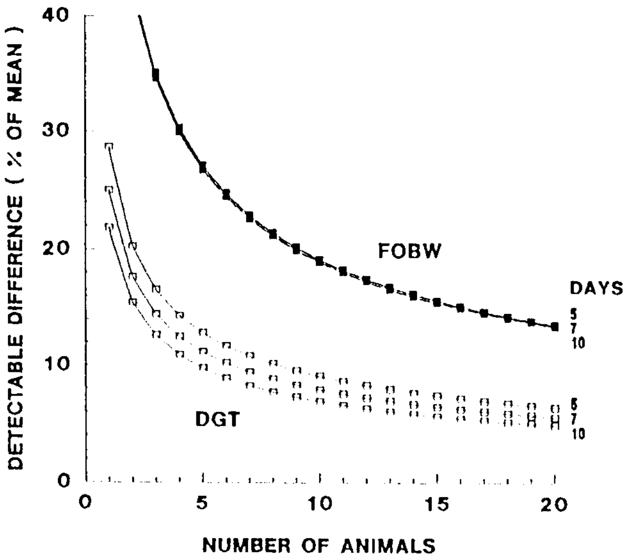


Fig. 5. Detectable differences (as a percentage of mean estimates) between two hypothetical treatments for daily grazing time (DGT) and fecal output (percentage of body weight, FOBW) over time intervals as affected by number of experimental units. Day is treated as a random effect. Relationships are expressed for a Type 1 significance level of 0.05, a Type 2 probability of 0.10, and three time intervals (5, 7 and 10 days).

expanding cow numbers compared with increasing the number of sampling days.

The balance in the experimental design produced by using the same days for both treatments allows the variability among days to be removed (eqn. (3)), assuming treatments are imposed across the same days. $CV_c$ becomes the critical term. At low $CV_c$ values, treatment differences can be detected with relatively few animals (Fig. 4). We could detect a 10% difference in DGT between two treatments using 9 cows per treatment group (e.g. supplement type) over 5 days ($\alpha=0.05, \beta=0.10$). However, the Type 2 probability level does have an effect on the potential difference detectable among treatments, especially when the $CV_c$ is high (Fig. 4). With a high $CV_c$, as noted for FOBW, the probability of detecting treatment differences decreases. The Type 2 probability level chosen affects the potential difference detectable among treatments, especially when $CV_c$ is high (FOBW, Fig. 4). Using 14 cows per treatment over 5 days would allow a 50% chance ($\beta=0.50$) of a 10% difference among treatments to be undetected. Reducing the probability of failing to detect a true treatment difference to 10% ($\beta=0.10$) would require $>20$ cows over 5 days (Fig. 4). Increasing the number of days sampled does little to reduce $\delta$ when $CV_c$ is high and is unnecessary at low $CV_c$ values (eqn. (3) and Fig. 5 for FOBW and DGT, respectively).

## DISCUSSION

Estimates of variability among animals, days, and animals by days are needed to help design experiments. Although measures of variability are published more frequently, measures of precision are not always clearly defined (Tacha et al., 1982). Also, sometimes $s_c^2$ and $s_\varepsilon^2$ (animal by day) are not adequately separated (Obioha et al., 1970; Lake and Clanton, 1972). Alternatives to published estimates of variance require pre-trial experiments or calculating estimates of variance (Gill, 1978, 1981).

Our values for $s_c^2$, $s_d^2$, and $s_\varepsilon^2$ were derived from a uniform group of cows grazing a relatively stable, dormant forage resource. Therefore, our cow, day and error components of variability should be relatively low. In comparison, Phillips and Denne (1988) reported DGT means with $CV_c$ and $CV_d$ of 24 and 15%, respectively, for 5 dairy cows in repeated 7-day trials over 4 weeks. Our estimates of animal numbers required are approximately in accordance with other recommendations (van Dyne and Meyer, 1964; Scales, 1972 (in Cordova et al., 1978)). Gary et al. (1970) suggested that > 9 cows and 6 days should be used to estimate grazing, loafing and lying time. These workers stated that 9 cows and 6 days might be adequate if cows are similar in liveweight, age and other characteristics. Increasing the number of cows per day is preferable to increasing days if the forage resource is changing rapidly (Obioha et al., 1970).

Animal variability (and consequently, *N* requirements) differs depending on the attribute investigated. Van Dyne (1969) discussed the sources and magnitude of animal variability for digestion trials. Van Dyne and Heady (1965) presented numbers of sheep or steers needed to sample specific components of animal diet on summer annual range. They reported that from 1 to 298 animals are necessary depending on the component of interest. Sheep are more variable in general than cattle (see Heaney et al. (1968) for a discussion of sheep intake trials). Fecal output measurement requires more animals than determination of diet composition. However, as available herbage increases, more animals are needed to sample diet composition whereas less are needed for fecal output (van Dyne and Meyer, 1964). Sampling cattle diets for lignin content requires more animals than for percentage nitrogen (Obioha et al., 1970). Similarly, cell wall content in steer diets varies more than dry matter disappearance (Lake and Clanton, 1972). Yet, both characteristics vary more among days than among steers, as did our estimates of variances associated with DGT.

Depending on the study objective, variance estimates are used in eqns. (1), (2), or (3) to estimate the sample sizes necessary to obtain mean estimates or detectable treatment differences with relatively high precision. If unreasonable numbers are required to obtain reasonable mean estimates, then critical review of methodology and experimental animal selection to reduce variability is required. With our data, 4 to 6 cows are sufficient to estimate a mean for either DGT or FOBW with reasonable precision only at very low values of $s$.

Equation (3), used when testing for treatment differences, not only assists sample size determination, but also the intensity of treatments to apply. For example, if the influence of forage quality on intake (related to FOBW) is to be determined, and only 10 cows are available per treatment group, then forage quality must differ enough among treatments to produce an expected difference in FOBW of 18–20% before it is likely to be detected (Fig. 4). Detecting a 5% difference would be improbable, and conducting trials with small differences in forage quality would lead to inconclusive results. Non-significant tests could be due to a lack of treatment effect, or the inability to detect an effect (Cohen, 1977; Gill, 1981; Rotenberry and Wiens, 1985).

Prior knowledge of necessary sample sizes, precision of estimated means, and detectable treatment differences can be used to design experiments to ensure that meaningful data are collected. The relationship in Figs. 1–5 should also be considered when developing research hypotheses. With the constraints of statistics and the logistics of range nutrition field work, it may not be possible to answer certain questions with current methodologies. The more heterogenous the experimental animals and more subjective the data collection methods, the less likely the results will be conclusive.

## ACKNOWLEDGMENTS

The authors appreciate the thorough manuscript reviews provided by Dean Billheimer, Richard Lund, Debbie Rossi and Steve Katchman.

## REFERENCES

Berndtson, W.E., 1989. Sampling intensities and replication requirements for detection of treatment effects on testicular function in bulls and stallions: a statistical assessment. J. Anim. Sci., 67: 213–225.

Beverlin, S.K., Havstad, K.M., Ayers, E.L. and Petersen, M.K., 1989. Forage intake responses to winter cold exposure of free-ranging beef cows. Appl. Anim. Behav. Sci., 23: 75–85.

Cohen, J., 1977. Statistical Power Analysis for the Behavioral Sciences. Academic Press, New York, 474 pp.

Cordova, M.K., Wallace, J.D. and Pieper, R.D., 1978. Forage intake by grazing livestock: a review. J. Range Manage., 31: 430–438.

Gary, L.A., Sherritt, G.W. and Hale, E.B., 1970. Behavior of Charlois cattle on pasture. J. Anim. Sci., 30: 203–206.

Gill, J.L., 1978. Design and Analysis of Experiments in the Animal and Medical Sciences, Vol. 1. Iowa State University Press, Ames, IA, 399 pp.

Gill, J.L., 1981. Evolution of statistical design and analysis of experiments. J. Dairy Sci., 64: 1494–1519.

Heaney, D.P., Pritchard, G.I. and Pigden, W.J., 1968. Variability of ad libitum forage intakes by sheep. J. Anim. Sci., 23: 159–164.
Lake, R.P. and Clanton, D.C., 1972. Sampling irrigated pasture with esophageal fistulated steers. Proc. West. Sect. Am. Soc. Anim. Sci., 23: 188–191.
Li, J.C.R., 1964. Statistical Inference I. Edwards Brothers, Ann Arbor, MI, 658 pp.
Obioha, F.C., Clanton, D.C., Rittenhouse, L.R. and Streeter, C.L., 1970. Sources of variation in chemical composition of forage ingested by esophageal fistulated cattle. J. Range Manage., 23: 133–136.
Phillips, C.J.C. and Denne, S.K.P.J., 1988. Variation in the grazing behavior of dairy cows measured by a vibrarecorder and bite count monitor. Appl. Anim. Behav. Sci., 21: 329–335.
Raleigh, R.J., Kartchner, R.J. and Rittenhouse, L.R., 1980. Chromic oxide in range nutrition studies. Oreg. Agric. Exp. Stn. Bull., 641, 41 pp.
Ross, R.L. and Hunter, H.E., 1976. Climax vegetation of Montana based on soils and climate. USDA Soil Conserv. Serv., Bozeman, MT, 61 pp.
Rotenberry, J.T. and Wiens, J.A., 1985. Statistical power analysis and community-wide patterns. Am. Nat., 125: 164–168.
Snedecor, G.W. and Cochran, W.G., 1967. Statistical Methods. Iowa State University Press, Ames, IA, 507 pp.
Statistical Analysis Systems, 1985. SAS User's Guide: Statistics. SAS Institute, Cary, NC, 956 pp.
Stobbs, T.H., 1970. Automatic measurement of grazing time by dairy cows on tropical grass and legume pastures. Trop. Grassl., 4: 237–244.
Tacha, T.C., Warde, W.D. and Burnham, K.P., 1982. Use and interpretation of statistics in wildlife journals. Wildl. Soc. Bull., 10: 355–362.
Van Dyne, G.M., 1969. Measuring quantity and quality of the diet of large herbivores. In: F.B. Golley and H.K. Buechner (Editor), A Practical Guide to the Study of the Productivity of Large Herbivores. Blackwell, Oxford, pp. 59–94.
Van Dyne, G.M. and Heavy, H.F., 1965. Dietary chemical composition of cattle and sheep grazing in common on a dry annual range. J. Range Manage., 18: 78–86.
Van Dyne, G.M. and Meyer, J.H., 1964. Forage intake by cattle and sheep on dry annual range. J. Anim. Sci., 23: 1108–1115.