

Frequency Sampling and Type II Errors

G.L. WHYSONG AND W.W. BRADY

Abstract

Probabilities of detecting frequency differences based on data obtained by random sampling were determined by computer simulation. Artificial, monotypic populations of known frequency were generated and sampled. Sample sizes of 100, 200, 500, and 1,000 plots were used to compare baseline populations of 20, 50, and 80% frequency to populations having progressively larger or smaller frequencies. Probabilities of detecting a difference in frequency from baseline populations were empirically estimated from 10,000 comparisons using a test of proportions ($P < 0.05$). Results indicated that the power of the test was substantially reduced at lower sample sizes. Equating the probability of Type I and Type II errors at 0.05 resulted in sample sizes of approximately 500 plots being needed to statistically distinguish between differences of plus or minus 10% frequency.

Key Words: trend, probability, computer simulation, sample size

The use of frequency data for evaluation of range trend has received significant interest (Hironaka 1985, West 1985). The effects of plot size on determination of vegetation frequency have been studied by several authors (Hyder et al. 1965, Kershaw and Looney 1985, Greig-Smith 1983). Hyder et al. (1965) indicated that 250 quadrats were sufficient to sample most plant species in macroplots on blue grama (*Bouteloua gracilis*) range. Fisser and Van Dyne (1966) investigated sample size using intervals from line intercept sampling of foothill bunchgrass range. Greig-Smith (1983) indicates that 100 plots should be considered a minimum when measuring vegetation frequency.

Eshelman (1985) discussed the different methods used to determine vegetation frequency within the Bureau of Land Management (BLM) and indicated that over 60% of BLM field offices were using frequency data as the primary indicator of trend. Additional interest in the use of frequency data for trend evaluation has occurred within several state agencies.

Little information exists concerning detection of vegetation differences using frequency data. Therefore, a computer simulation study on artificial vegetation populations was conducted to evaluate the sensitivity of frequency data for detecting differences using several sample sizes. The primary objective was to investigate the effect of sample size on probabilities of detecting true differences in frequency or, conversely, not detecting real differences (Type II errors).

Methods

Frequency is based on presence or absence of a plant species or vegetation group within some sampling apparatus, usually a plot. A location in computer memory can be defined to correspond to a plot location. Thousands of locations grouped together can then be used to simulate a vegetation population containing a population of possible plot locations. These populations may then be sampled, results tabulated, and statistical tests performed to evaluate sample results through computer simulation techniques.

A computer program was written to generate artificial monotypic plant populations, conduct sampling, and evaluate statistical tests of frequency estimates. Populations with known frequencies

were generated using computer memory locations randomly set to logical true or false conditions corresponding to vegetation presence or absence within sample plots. The number of plot locations within an artificial plant population was adjusted relative to the sample size being used, so that the maximum number of plots measured during any one sample was less than or equal to 5% of the population. Cochran (1963) and Scheaffer et al. (1979) indicated that no finite correction was necessary when 5% or less of the population was being sampled. Two populations of known frequency were generated during each trial. The frequency of 1 population was held constant at a predetermined baseline level (20, 50, or 80%) while the frequency of the second population was incremented between each trial. Both baseline and the population to be compared were regenerated between trials. Sampling consisted of randomly determining plot locations within each population and measuring presence or absence by evaluating the true or false condition of the corresponding memory location. Sample sizes used were 100, 200, 500, and 1,000 plots.

During a simulation, sample size was constant and the baseline and second population were sampled independently. Frequency was determined for each population and one-tailed test of proportions conducted to determine if a significant difference existed between the 2 sample estimates (Zar 1984). The alpha level was set at 0.05 for all comparisons. No correction for continuity was used. Simulations indicated that correction resulted in overly conservative test results when comparing 2 populations having the same frequency. A two-tailed test of proportions was used when comparing a baseline frequency to itself in order to check that the probability of Type I errors was maintained near the 0.05 probability level.

Probabilities of detecting significant differences in frequency between the 2 populations being compared were determined by evaluating results of statistical tests from 10,000 independently paired samples collected during each trial. Repeated trials were conducted at different known frequencies to empirically determine the power curves of the statistical tests for each baseline population. In order to avoid repeated random number sequences, the random number generator was reseeded prior to the generation of each population and after each thousand samples.

Following completion of one-tailed testing, the 50% baseline frequency was selected for further investigation of the power curve using a two-tailed test of proportions. Application of frequency measurements for detection of vegetation change in the field would be conducted on communities where actual frequencies are unknown. Two-tailed Chi-Square tests are generally recommended for testing between frequency estimates (Greig-Smith 1983, Hironaka 1985). Zar (1985) illustrated the equivalency of the two-tailed test of proportions to Chi-Square. The two-tailed test of proportions was used to evaluate the power curves of the 50% baseline frequency using sample sizes of 200 and 500 plots. This procedure closely simulated conditions under which frequency estimates between plant communities are usually tested and also allowed comparison to the corresponding one-tailed power curves previously determined.

All computer programs were written in Pascal and simulations were conducted using microcomputers located within the Division of Agriculture at Arizona State University. Due to the amount of computer time required for simulations, as many as a dozen computers were employed simultaneously.

Authors are associate professors, Division of Agriculture, Arizona State University, Tempe 85287.

The authors wish to acknowledge the contribution of Drs. R. Backhaus, J. Baxter, J. Brock, R. Chalquest, J. Diaz, H. Hatzell, K. Maddy, W. Miller, Ms. G. Patrick, Drs. M. Raccach, and J. Stutz from the Division of Agriculture. Their spirit of cooperation and willingness to allow their micro-computers to be mistreated at night and during weekends contributed significantly to the timely completion of this project. Manuscript accepted 28 May 1987.

Results and Discussion

The probability of a Type I error is determined by selection of a probability level by the investigator. When testing for differences between frequency estimates, a Type I error will occur when one concludes that there is a difference when the actual frequencies of the 2 populations are the same. The probability of a Type I error is commonly set at 0.05 since it is thought that falsely concluding that a difference exists is undesirable. A Type II error occurs when a difference in frequency actually exists but is undetected by the statistical test, resulting in a conclusion of no difference. Since the probability of a Type II error is difficult to determine and control, it has received much less attention. The widespread interest in using frequency for detecting vegetation change and evaluating range trend (Eshelman 1985, Hironaka 1985), requires that an evaluation of the probability of detecting differences be conducted.

When evaluating rangelands, is it more important to falsely conclude a vegetation change has occurred or to not detect a difference when it has occurred? Since the former can be controlled by selecting a probability level, the control of the latter becomes an important issue. The early detection of vegetation response to environmental and grazing conditions is needed for evaluation and management of rangelands.

The probability of occurrence of Type II errors can only be decreased by increasing the chance of a Type I error and/or increasing sample size. The cost of a Type I error can be high, since it may result in either unwarranted livestock reduction or overstocking. Consequently, it may be undesirable to increase the probability of this kind of error. Tanke and Bonham (1985) judged the risks associated with Type I and Type II errors of equal importance when evaluating range trend. Under these conditions, sample size remains the preferred means of exerting some control over the Type II error.

The power of a test is the probability that a statistical test will detect a difference when one actually exists and is calculated as one minus the probability of a Type II error. When determined over a range of differences a power curve may be developed for specific comparisons. Power curves obtained for the 20% frequency baseline population as compared to populations having different frequencies in increasing or decreasing order are presented in Figure 1. The curves are not symmetrical since frequency data fit the binomial distribution.

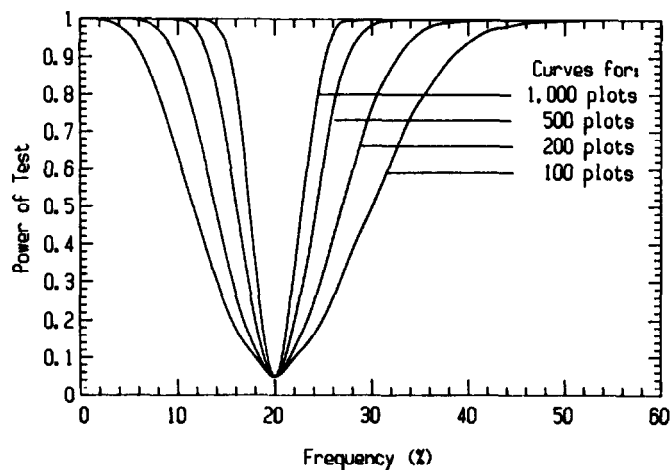


Fig. 1. Effect of sample size on the probabilities of detecting a change in frequency from the baseline frequency of 20% (one tailed).

Greig-Smith (1983) suggested that a sample size of 100 plots should be considered a minimum and more would be desirable. The results of this study indicate that appreciably more than 100 plots are necessary to avoid high probabilities of Type II errors.

Using 100 plots, a decrease in frequency of approximately 9%, from the baseline of 20%, will be detected about 50% of the time (Fig. 1). If one desired to set the probability of a Type II error equivalent to the 0.05 level used for the Type I error, differences, could only be statistically detected upon a decrease of approximately 15% or more or an increase of 21% or greater from the baseline frequency (Table 1). In other words, if the original fre-

Table 1. Upper and lower limits for each sample size and baseline frequency where the probability of a Type II error was empirically estimated at 0.05 (one-tailed).

Sample Size	Baseline Frequency (%)					
	20		50		80	
	Lower ¹	Upper ²	Lower	Upper	Lower	Upper
100	5.3	40.6	27.7	72.4	59.4	94.5
200	8.9	34.3	33.7	66.1	65.8	91.3
500	12.7	28.6	40.2	60.2	71.3	87.6
1,000	14.8	25.9	43.4	57.1	73.9	85.5

¹Frequencies at or below the indicated value result in probabilities of Type II errors of 0.05 or less.

²Frequencies at or above the indicated value result in probabilities of Type II errors of 0.05 or less.

quency was 20%, the frequency would have to decrease to 5% or less or increase to 41% or more before we could be confident of detecting that difference 95 times out of 100. The reduction in the probability of a Type II error was greater at the 500 plot sample size (Fig. 1). Proportionally less improvement occurred between 500 and 1,000 plot sample sizes than among other sample sizes. Results from the 80% baseline frequency are not presented since they are the same as those obtained from the 20% frequency with the exception that the power curves are skewed in the opposite direction.

As frequency approaches 50%, the binomial distribution approximates the symmetry of the normal distribution. Results obtained from the 50% baseline frequency are presented in Figure 2. Large

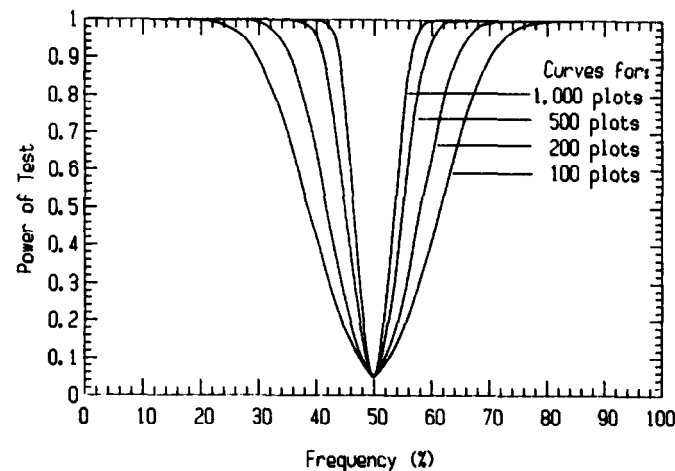


Fig. 2. Effect of sample size on the probabilities of detecting a change in frequency from the baseline frequency of 50% (one tailed).

probabilities of occurrence of Type II errors are evident at sample sizes below 500 plots unless the detection of only major differences is desired. At 500 plots a difference of 9% from the baseline frequency resulted in approximately a 10% chance of a Type II error. That is, when the original frequency was 50% a decrease to 41% or less or an increase to 59% or more would be detected 9 times out of 10 (using a sample size of 500 plots). Doubling the sample size to 1,000 plots reduced the difference from 9 to 6%.

Table 1 presents the frequency endpoints where the probability of Type II errors was computed as equal to that of a Type I error

(0.05), for detecting differences from the 3 baseline frequencies. The endpoints presented were estimated to the nearest tenth of a percent. These data illustrate the effect of sample size on the chance of occurrence of Type II errors. Although a 1,000 plot sample size for the 50% baseline frequency results in over a 300% increase in efficiency compared with 100 plots, a 7% difference from the baseline is still required to effectively reduce Type II errors to 0.05. It may be impractical to assume that the occurrence of Type II errors can be reduced to such levels when determining differences among natural vegetation communities. Instead, a decision concerning the acceptable probability of a Type II error will be required, much like that used for Type I errors. Theoretical power curves for different frequency estimates and Type I error rates may be estimated using the normal approximation procedures as presented by Zar (1984). Such procedures are essential to assess the proper sample size and the reliability of statistical tests for detecting frequency differences at a desired precision.

Frequency measurements under field conditions are conducted without knowledge of the actual vegetation frequency. The test employed to detect differences are commonly two-tailed. Figure 3 illustrates the power curves of two-tailed tests for 200 and 500 plot sample sizes conducted at the 50% baseline frequency. These

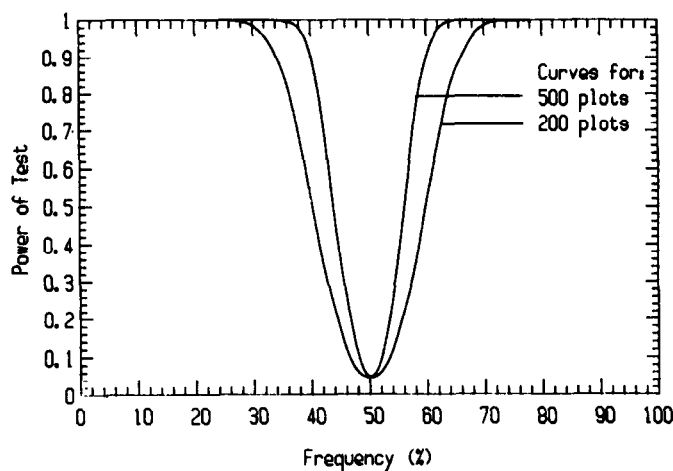


Fig. 3. Effect of sample size on the probability of detecting a change in frequency from the baseline frequency of 50% (two tailed).

curves appear similar to those for the one-tailed test (Fig. 2). However, the probabilities of Type II errors are slightly greater than those from the corresponding one-tailed tests. This is also apparent from the upper and lower endpoints where probabilities of Type II errors were estimated at 0.05 (Table 2) as compared to

Table 2. Upper and lower limits (%) for the 50% baseline frequency where the probability of a Type II error was empirically estimated at 0.05 (two-tailed).

Sample Size	Lower ¹	Upper ²
200	32.3	67.6
500	39.1	61.1

¹Frequencies at or below the indicated value result in probabilities of Type II errors of 0.05 or less.

²Frequencies at or above the indicated value result in probabilities of Type II errors of 0.05 or less.

those from one-tailed tests (Table 1). The endpoints from two-tailed tests generally increased 1% in either direction as compared to those obtained from one-tailed tests.

Conclusions

The results of this study indicate that probabilities of detecting vegetation change or differences using frequency data collected with sample sizes commonly employed today are low. The most practical way to improve chances of detection without sacrificing the probability of a Type I error is to increase sample size. Sample size should be determined based upon the range of frequency difference desired to detect and the acceptable probability of not detecting that difference. It must be recognized that detecting changes in vegetation frequency of plus or minus 10% with a high degree of success can only be obtained with sample sizes approaching 500 plots ($P < 0.05$). If it is desired to narrow this range to plus or minus 5% an excess of 1,000 plots will most likely be required.

The results obtained from this study cannot be considered applicable to all situations. The effect of probability levels other than 0.05 for Type I errors was not investigated. No attempt was made to evaluate the effects of pattern on frequency estimates, since populations were generated with a random pattern of distribution. Finally, since sampling was conducted randomly, the applicability of these results to those obtained from systematic sampling, as commonly employed, is unknown.

Despite these limitations, the importance of sample size for the detection of frequency differences is apparent. The early detection of vegetation response to grazing is necessary for timely decisions involving range condition and trend. The use of frequency data for rangeland evaluation may require more effort than other methods of vegetation measurement.

Literature Cited

- Cochran, William G. 1963. Sampling techniques. John Wiley and Sons, Inc., New York.
- Eshelman, K.R. 1985. Vegetation inventory and monitoring on the public lands. Symp. on use of frequency and density for rangeland monitoring. William C. Krueger, Chairman. Proc. 38th Annual Meeting, Soc. for Range Management. Feb. 1985. Salt Lake City. Soc. for Range Manage. 79-83.
- Fisser, H.G., and G.M. Van Dyne. 1966. Influence of number and spacing of points on accuracy and precision of basal cover measurements. J. Range Manage. 19:205-211.
- Greig-Smith, P. 1983. Quantitative plant ecology. 3rd ed. Univ. of California Press. Berkeley.
- Hironaka, M. 1985. Frequency approaches to monitor rangeland vegetation. Symp. on use of frequency and density for rangeland monitoring. William C. Krueger, Chairman. Proc., 38th Annual Meeting, Soc. for Range Manage. Feb. 1985. Salt Lake City. Soc. for Range Manage. 84-86.
- Hyder, D.N., R.E. Bement, E.E. Remmenga, and C. Terwilliger, Jr. 1965. Frequency sampling of blue grama range. J. Range Manage. 18:90-93.
- Kershaw, K., and J.H. Looney. 1985. Quantitative and dynamic plant ecology. 3rd. ed. E. Arnold Pub. Ltd., Baltimore.
- Scheaffer, Richard, L., William Mendenhall, and Lyman Ott. 1979. Elementary survey sampling. 2nd ed. Duxbury Press, Mass.
- Tanke, W.C., and C.D. Bonham. 1985. Use of power curves to monitor range trend. J. Range Manage. 38:428-431.
- West, N.E. 1985. Shortcomings of plant frequency-based methods for range condition and trend. William C. Krueger, Chairman. Proc., 38th Annual Meeting Soc. for Range Manage. Feb. 1985. Salt Lake City. Soc. for Range Manage. 87-90.
- Zar, Jerrold H. 1984. Biostatistical analysis. 2nd ed. Prentice-Hall Inc., Englewood Cliffs, N.J.