Use of Power Curves to Monitor Range Trend

WILLIAM C. TANKE AND CHARLES D. BONHAM

Abstract

Type I and Type II errors and the power of the test when testing the null hypothesis of static range trend are discussed. The consequences associated with Type I and Type II errors are judged to be similar and therefore the probability of committing a Type I or Type II error should be equal. As an example, the current range trend monitoring program for the Moose Camp Allotment on USDA Forest Service land in southwestern Montana is capable of detecting a change in range condition of one condition class 83% of the time when the probability of Type I error is set at .17 (Prob[Type I Error] = Prob[Type II Error] = .17). Doubling the sample size would increase the ability to detect a condition class change to 95% when the probability of Type I error is set at .05 (Prob[Type I Error] = Prob[Type II Error] = .05).

A goal of range management is to promote increased carrying capacity while maintaining the health and stability of the range ecosystem. On an existing grazing unit where an initial estimate of available carrying capacity is established, some method to determine the desirability of this stocking level is usually undertaken. This may be done by personal impressions from range inspections or through some formal data gathering procedure which assesses the impacts associated with a particular stocking level.

On public lands, range condition and trend measurements are the accepted method for monitoring the impacts associated with livestock grazing (Public Rangeland Improvement Act 1978, USDA Forest Service 1981), where the current species composition indicates range condition, and changes in condition over time indicate range trend.

In any range ecosystem there exists a certain amount of variability in range condition in space. This is due to the well-known concept of natural variability in biological systems (Vogl 1976). The practical importance of variability in space is that range condition will differ at different locations and therefore, a single measurement may not give a good indication of the overall condition for the management unit. One way to account for this variation is through repeated measures at different locations.

If an assumption is made concerning the distributional relationship of range condition, it then becomes possible to estimate mean range condition for a management unit and, by use of parametric statistical hypothesis testing procedures, to determine if any significant changes have occurred in mean range condition over time. In other words, statistical analyses allow the range manager to determine, with a certain degree of confidence, if range trend is upward, downward, or static.

Because of the variability inherent in range ecosystems, it is not feasible to determine mean range condition with absolute certainty since this would entail a sample size large enough to include the whole population (i.e., inferential statistics would not be required since the sample would be equal to the population). Limited resources do not allow a sample size of this magnitude and, as a consequence, errors in the determination of range condition and trend are always possible. Management decisions which are made based on the erroneous determination of range trend can have serious consequences both economically as well as environmentally. For example, if errors cause the manager to believe trend is upward when it is actually static or downward, a decision to increase stocking may be made and, as a consequence, deterioration in the system may begin or accelerate.

Statistical methodology provides the manager with a means to estimate the probability of making an error. There are 2 kinds of errors associated with the testing of the null hypothesis (H_0: static trend) against the alternate hypothesis (H_1: nonstatic trend): Type I error occurs when the null hypothesis is rejected but is actually true, and Type II errors occurs when the null hypothesis is not rejected but is actually false (Huntsberger and Leaverton 1970) (Table I).

Mention of the existence of Type II errors is surprisingly absent in recent natural resource literature. The absence of any discussion on Type II errors may mean that many of those involved in natural resource management and research are oblivious to the possibility of committing these errors and that a discussion of them may be of value. The purpose of this paper is to introduce the concepts of Type II error and power of the test, and illustrate their usefulness in evaluating the monitoring program for range trend on a specific U.S. Forest Service range allotment. While this paper focuses on the evaluation of a range condition and trend monitoring program, the concepts presented are applicable to all natural resource decision makers who utilize hypothesis testing procedures. Knowledge of Type II errors should help those who utilize hypothesis tests to better understand the benefits and limitations inherent in their use.

Statistical Errors and Power of the Test

The probability of committing a Type I error is equal to the significance level or size of the test (Mood et al. 1974). Controlling Type I errors is routine and is the basis for choosing the significance level. On the other hand, the probability of committing Type II errors is rarely considered, except indirectly, by following the accepted procedure of formulating the 2 hypotheses in such a way that Type I errors are most important (Neyman 1942, Cohen 1977). In many natural resource problems this is not an acceptable solution because both Type I and Type II errors are important. Consider the following example.
Type I error—A nonstatic trend is falsely assumed.

Reject $H_0$ when true.

Belief about range trend:
- True state of nature:
  - Management decision:
    - Consequences:
      - Short-term
      - Long-term

Type II error—A static trend is falsely assumed.

Do not reject $H_0$ when false.

Belief about range trend:
- True state of nature:
  - Management decision:
    - Consequences:
      - Short-term
      - Long-term

Table 1. Errors associated with the testing of the null hypothesis of static trend ($H_0$: Trend static).

<table>
<thead>
<tr>
<th>Hypothesis Test</th>
<th>Trend Static</th>
<th>Trend Not Static</th>
</tr>
</thead>
<tbody>
<tr>
<td>Do Not Reject $H_0$</td>
<td>$\beta_0$</td>
<td>Type I error—Erroneously change stocking</td>
</tr>
<tr>
<td>Reject $H_0$</td>
<td>$\beta_0$</td>
<td>Type II error—Erroneously maintain stocking</td>
</tr>
</tbody>
</table>

Formulate the 2 hypotheses:

$H_0$: Trend is static
$H_1$: Trend is not static

A Type I error (reject $H_0$ when actually true) occurs when the manager decides the trend is not static when in fact it is. If the implied trend is upward, the manager may erroneously decide to increase stocking and a deterioration in the ecosystem may occur. Conversely, if the implied trend is downward, the manager may erroneously decide to decrease stocking and thereby incur an economic loss. A Type II error (do not reject $H_0$ when actually false) can result in similar consequences. If the manager decides the trend is static, then no adjustments to stocking are indicated. If the trend is actually downward, this lack of action may cause the downward trend to accelerate. Conversely, if an upward trend goes undetected, economic losses may occur. These results are summarized in Table 2.

Table 2 shows that an erroneous decision may result in serious negative consequences, particularly in the long term. Furthermore, because the consequences associated with Type I and Type II errors are similar, we may judge that the probability of committing these 2 types of errors should be equal. Procedures for developing hypothesis tests with equal probability of Type I and Type II errors will be examined in this paper. Similar procedures may be followed when it is judged the probability of Type I and Type II errors should not be equal; this judgment should be based on an analysis of the consequences of Type I and Type II errors as was done here.

Accurate estimates of range condition and trend are necessary in order to effectively utilize the productive potential of the land. It is important to be able to detect any change in condition which is large enough to signal ineffective management. When testing the null hypothesis of no change in range condition ($H_0$: trend static), a Type II error (do not reject $H_0$ when $H_0$ is false) is, in essence, the inability to detect a change in range condition (nonstatic trend) when in fact a change has occurred. Type II errors associated with this hypothesis may result in incorrect decisions, and consequently, inefficient management. Conversely, the ability to detect a change in condition enables the manager to adjust stocking levels as required, and efficiently utilize the forage resource.

In statistical terminology, the probability of detecting a change in range condition (reject $H_0$ when $H_0$ false) is referred to as the power of the test. Power ($\pi$) is a probability defined as follows:

$$\pi = \text{Power} = \text{Prob}[\text{Reject } H_0 \text{ when } H_0 \text{ False}] = \text{Prob}[\text{Detecting a condition change}] = 1 - \text{Prob}[\text{Type II Error}]$$

Associated with any hypothesis test is a power curve ($\pi(\theta)$). The power curve associated with the test of the null hypothesis of static range trend gives the probability of detecting a change in range condition ($\theta$) of a given magnitude. As would be expected, as $\theta$, the magnitude of the range condition change increases, so does the probability of detecting that change, i.e., the magnitude of range trend ($\theta$) and power ($\pi$) are directly related.

Methods

Data for this study came from U.S. Forest Service range condition transects in the Moose Camp range allotment of the Deelodge National Forest. This allotment is located in southwestern Montana approximately 10 miles south of Butte, Montana. The data consist of a linear transformation of the numeric composition ratings from 6 permanent range condition transects. Each transect consists of a relocatable 100-foot linear sampling unit from which, at 1-foot intervals, the nearest plant species is recorded. The composition score, which is used by the Forest Service as a part of their overall range condition rating, is a numeric value ($X$) from 1 to 15.

Table 2. Consequences associated with various erroneous decisions resulting from Type I and Type II error when testing the null hypothesis of static trend.

<table>
<thead>
<tr>
<th>Type I Error</th>
<th>Trend up</th>
<th>Trend down</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consequences:</td>
<td>Decrease in condition</td>
<td>Decreased forage production</td>
</tr>
<tr>
<td>Implications to ranch stability</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Type II Error</th>
<th>Trend static</th>
<th>Trend up</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consequences:</td>
<td>Maintain stocking</td>
<td>Maintain stocking</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Short-term</th>
<th>Long-term</th>
</tr>
</thead>
<tbody>
<tr>
<td>Further decrease in condition</td>
<td>Decreased forage production</td>
</tr>
</tbody>
</table>
and is based on the percentage of desirable, intermediate, and undesirable species tallied for a transect (USDA Forest Service 1975). The numeric composition rating was resealed, by the formula \( Y = 100(X - 1)/14 \), to lie between 0 and 100. This resealed value \( Y \) will be referred to as the range condition for a transect. The 6 transects for which range condition values were derived are grouped into 3 range condition clusters, with each cluster containing 2 transects located in close proximity to one another. The null hypothesis of no change in range condition was tested for each of the 3 clusters, using a paired \( t \)-test, and a pooled variance estimate \( (s^2) \) (Steele and Torrie 1975). Use of a pooled variance estimate is based on the assumption that the variance associated with each of the 3 clusters is equal. The degrees of freedom \( (df) \) associated with both this variance estimate and the \( t \)-test for which this estimate is utilized is:

\[
df = 3(n - 1)
\]

where, \( n \) = the number of transects in each cluster (\( n = 2 \) for the current monitoring program).

Associated with each hypothesis test is a power curve which is a function of: the chosen significance level \( (\alpha) \), the “population” variance, the sample size \( (n) \), and the deviation of range trend from zero \( (\theta) \), i.e., the magnitude of range trend. The pooled estimate of the variance \( (s^2) \) was used for the population variance in all 3 clusters. Power curves were developed for various values of the other 3 variables \( (\alpha, n, \theta) \) using a computer program developed by the senior author. These curves were used to investigate the probability of detecting a range condition change for the current and alternate monitoring programs. A more detailed discussion of the development of the power curves used in this paper may be found in Tanke (1984). The text by Cohen (1977) provides a thorough discussion and numerous tables useful for calculating power curves for many of the tests (\( t \)-test, ANOVA, chi-square) commonly used in range science.

**Results and Discussion**

**Significance Level and Power**

The typical methodology used when applying statistical hypothesis testing techniques is to choose a small significance level \( (\alpha = \text{Prob}[\text{Type I Error}]) \) to control for Type I errors while giving little or no consideration to errors of the second kind (Lehmann 1958). This practice is objectionable whenever the consequences associated with Type II errors are important. One problem with blindly selecting a small significance level is that reducing the significance level and the probability of Type I error simultaneously increases the size of Type II error (Neyman et al. 1935; Sokal and Rohlf 1969).

Figure 1 shows a series of power curves for the current monitoring program (3 clusters with each cluster containing 2 transects). Each curve is associated with a particular significance level \( (\alpha) \), and the probability of detecting a change in condition of a given magnitude \( (\theta) \). These curves may be used to examine the relationship between the significance level \( (\alpha) \), the magnitude of range trend \( (\theta) \), and the power of the test \( (\pi) \). Because a two-tailed test was used, changes in condition associated with an upward or downward trend have identical power when their magnitude is the same. For example, for an upward or downward trend of magnitude 20 power is equal to .80 when \( \alpha = .15 \). Because it is the magnitude and not the direction of range trend that is important in determining power, only the right half of the symmetrical power curves are shown in Figure 1 and in figures to follow.

Consider, as a reference point, a change in range condition of 20. This magnitude of change is equivalent to a change of one qualitative condition class, where range condition classes are defined as follows: excellent (81-100), good (61-80), fair (41-60), poor (21-40), very poor (1-20). Assume the manager is interested in testing the null hypothesis of static trend and, following conventional procedure, has selected a significance level of 0.05. Figure 1 indicates there is less than a 50% chance this test will detect a change in condition equivalent to one condition class. Furthermore, if a 0.01 significance level had been selected, a condition class change would be detected only 15% of the time. The consequence of these low power values is: a potential loss of AUM’s if an upward trend goes undetected, or a further decrease in range condition if a downward trend goes undetected (Table 2).

This analysis clarifies why a significance level should not be selected without first considering the power of the test. Furthermore, it shows that the choice of a smaller significance level does not necessarily result in a better testing procedure. While reduction of the significance level decreases the probability of a Type I error, it also simultaneously increases the probability of a Type II error. A better method for selection of the significance level is based on the consequences of Type I and Type II errors. Previous discussion indicated that, because of the similar consequences associated with Type I and Type II errors (Table 2), the “optimal” testing procedure should have an equal probability of committing Type I and Type II errors. Therefore, for a condition change of 20 the desired significance level \( (\alpha) \) is approximately 0.17 (Fig. 1). For this significance level the probability of Type I error \( (\alpha=0.17) \) is equal to the probability of Type II error \( (\beta=1-0.83=0.17) \).

From a manager’s standpoint, a power curve may be thought of as illustrating the relative value, or importance, associated with the ability to detect a given magnitude of trend. A manager who can quantify this relationship may use this information to assist in selecting a testing procedure. For example, assume the manager feels it is:

1. Crucial to be able to detect a change in range condition of 30 and above.
2. Only 75% as important to detect a range condition change of 20.
3. Only 30% as important to detect a range condition change of 10.
4. Of no importance to detect a condition change of 5 and below.

The manager has thus defined 4 points on an “idealized” power curve:

\[
\begin{array}{c|c}
\theta & \pi \\
30 & 1.00 \\
20 & 0.75 \\
10 & 0.30 \\
5 & 0.00 \\
\end{array}
\]

Figure 1 shows this idealized curve (labeled “idealized”) superimposed on the actual power curves. The actual curve which comes
closest to matching the idealized curve is associated with a significance level of 0.10. The selection of a significance level of 0.10 represents another possible choice of $\alpha$ for the current monitoring program.

**Sample Size and Power**

When the sample size is fixed, it may not be possible to lower probabilities of Type I and Type II errors to acceptable limits. With the current monitoring program, and assuming an approach which balances the probabilities of Type I and Type II errors, the manager can expect to commit an error in 10 to 20% of the hypothesis tests performed. If this is not acceptable, increasing the sample size should be considered. An increase in the number of transects makes it possible to simultaneously lower the probability of committing both types of error.

Figure 2 shows a series of power curves for testing the null hypothesis of static trend when the significance level is equal to 0.05. Each curve is associated with a different sample size (n) equal to the number of transects in each of the 3 clusters. This graph may be used to investigate the effect of sample size (n) on power of the test ($\pi$) for different magnitudes of range trend ($\theta$). For example, consider a change in range condition of 20. There is a substantial increase in power when the sample size is increased from the present 2 transects per cluster to 3 transects per cluster. Smaller increases in power occur up to $n = 6$, thus substantial improvement in the ability to monitor range trend may be obtained through modest increases in sample size.

The idealized power curve, previously discussed, may be used to assist in selecting the sample size in the same way it was used to assist in selecting the significance level. This is illustrated in Figure 2 which shows that, for a significance level of 0.05, the actual power curve which comes closest to matching the manager's idealized curve is for a sample size of 3.

**Conclusions**

Statistical power analysis provides the range manager with a useful set of analytical tools. These tools can assist in the development of a range condition monitoring program and associated hypothesis testing procedures. They can also be used to investigate present and past monitoring programs to judge their validity and usefulness. The technique is not limited to the study of monitoring problems but can be applied to any decision-making environment where hypothesis testing is utilized.

The value of statistical power analysis is not confined to the manager. It is perhaps of more importance to the researcher who uses statistical techniques to guide and justify research. It is the results of this research, more than the results of hypothesis tests, upon which the manager typically relies when making management decisions.

Introduction of the concepts of statistical errors into natural resource publications should be beneficial to both managers and researchers alike. A working knowledge of these concepts should result in research proposals and management decisions which are both more valid and more effective.

**Literature Cited**


