Coryell A. Ohlander
6048 So. Lakeview Street
Littleton CO 80120
cohlander@polnow.net

March 27, 2000

To: Content Analysis Enterprise Team
USDA Forest Service
Bldg. 2 Suite 295
5500 W. Amelia Earhart Dr.
Salt Lake City, UT 84116 (cleanwater/wo_caet-slc@fs.fed.us)

RE: Good Science aspects in UFP
Part 1 regarding misuse of ordinal indexes

Goals "... in a unified and cost-effective manner."
Guiding principles: "...consistent and scientific approach...."
Agency Objectives: "...common science-based approach...."
"...test watershed assessment procedures...."
"...implement ... consistent with applicable legal authorities."
"...will base watershed management on good science."
"...science-based total maximum daily loads (TMDLs)."
"...sharing of scientific and technical resources;"

In pursuit of the most demanding tasks pertaining to "good science," I believe that meeting an evidentiary standard is critical to the foundation of good science. The issue comes into play for the Administrative Practices Act and, in the UFP issue, in the CWA S505 regarding citizen lawsuits. This seems like a no-brainer; however, during the July 1999 national review of T-Walk, the argument of what would meet an evidentiary standard brought to light the fact that many biologists and hydrologists would - and do - choose evaluation systems that are not supportable in Court. These systems are based on treating ordinal numbers as though they are interval or ratio data. Typically, several narrative categories are assigned a rank and - with a multi-metric concept - are added up to arrive at a grand total.

In order for information to be supportable in Court, the procedure itself must be valid. And for each class of data - nominal, ordinal, interval, and ratio - there are rules for what is appropriate analysis and what is not. When we create ordinal data and treat it as interval or ratio data, we cross the line into nonsense and obviously fail to support good science.

For UFP, this is an important issue because EPA is the spring board for numerous State biological assessments based wholly or in part on the Rapid Biological Assessment (RBP). RBP offers flagrant disregard for the rules on procedural validity. And it would be easy for UFP to adopt a system that, when the chips are down, will be defeated. Consider, for example, the habitat assessment from the 1999 RBP (page A8) with this list of factors:

Epifaunal Substrate/available cover
Embeddedness
Velocity/Depth regimes
Sediment Deposition

CAET RECEIVED
MAR 29 2000

Channel Flow Status
Channel alteration
Frequency of riffles
Bank Stability
Vegetative Protection
Riparian Vegetative Zone
Total Score _____

Each factor carries 20 points and the total score is used in comparisons with other sites or a against a reference condition. However, a score of, say, 150 can be arrived at by an almost infinite number of combinations. Or, we could have a score of 180 and find all factors are perfect (9x20) **except** all the water has been taken out by a diversion (0). Thus, the loss of water is the limiting factor and it doesn't really matter how great everything else is. The limiting factor defines the bottomline biological health. That is, there is no way to compensate no water by having great bank stability. Yet the index total suggests each factor is a substitute for every other factor - which is procedural as well as logical nonsense.

The article by Schuster and Zuuring, "Quantifying the Unquantifiable - Or, have you stopped abusing measurement Scales?" has been enclosed to highlight the issue.

We must anticipate that watershed assessments will bring proponents and opponents out of the wood work and any time a decision is made counter to their special interests, it will be challenged by either appeal or by law suit. To survive such a test, good science in UFP will need to recognize the four classes of data and to use assessment systems that are procedurally valid.

Regards

Coryell A. Ohlander

Coryell A. Ohlander
Hydrologist

enc: Schuster + Zuuring

CAET RECEIVED
MAR 29 2000

QUANTIFYING THE UNQUANTIFIABLE

Or, Have You Stopped Abusing Measurement Scales?

By Ervin G. Schuster and Hans R. Zuuring

The forestry profession finds itself in a headlong quest for quantitative information. The phenomenon is reflected in the number of quantitatively oriented courses in forestry curricula, in budget processes, and in mathematical land-management planning models. We are now quantifying what previously was thought to be unquantifiable or not needing quantification—public opinion, esthetic quality, personnel performance. The justification is that objective information is needed for sound verifiable decisions.

Unless data are properly analyzed, resulting numbers may be worthless at best, misleading, or even counterproductive. The profession has witnessed a growing array of abuses in quantification—biased sampling, ignored variability in data, and inappropriate inferences. Although some abuses come from sloppiness, others result from unawareness. The article by Stafford in the March 1985 *JOURNAL* provided a statistical basis on which data analysis and interpretation can be improved. To assist the practicing forester, we move back a notch and focus on the measurement itself. Of the abuses stemming from unawareness, the most prevalent and insidious is the failure to recognize, select, and utilize the scale of measure-

Ervin G. Schuster is research forester, USDA Forest Service, Intermountain Research Station, Forestry Sciences Laboratory, Missoula, MT. Hans R. Zuuring is associate professor, School of Forestry, and station biometrician, Montana Forest and Conservation Experiment Station, University of Montana, Missoula.

GAET RECEIVED
MAR 29 2000

"Of the abuses stemming from unawareness, the most prevalent and insidious is failure to recognize, select, and utilize the scale of measurement appropriate to the task at hand."

CAET RECEIVED
MAR 29 2000

ment appropriate to the task at hand.

Before taking measurements, foresters decide on sampling design, data analysis, and data interpretation; they also select a scale. A measurement simply assigns symbols (numbers, letters, and words) to things (objects or events) to represent facts or characteristics. Assignments are made according to rules that result in different kinds of measurements and measurement scales. Choice of scale is unavoidable and inevitable and should be conscious and explicit. The four measurement scales commonly used are nominal, ordinal, interval, and ratio.

Suppose a forester wants to measure the sensation of warmth or coldness associated with an object (temperature). The silviculturist may be interested in site temperature, the hydrologist in snow temperature, the fire specialist in flame temperature. The simplest form of measurement is made on the nominal scale.

Nominal Scale

A nominal scale consists of categories or classes into which objects are placed and counted. Nominal implies naming. The "names" of the classes are unimportant because they merely serve to label or identify each class. With a nominal scale, observations may be separated only according to categories. Although numbers could be used as labels, words or letters work just as well.

For example, each digit comprising a forest-habitat-type code could be replaced by a letter. If 0 through 9 were replaced by the letters A through J, then the code 321 (for PSME/CARU-AGSP, one of the Douglas-fir habitat types) would become DCB (rather than 321), with no loss in information. Likewise, temperature can be measured by a nominal scale consisting of the classes "good" and "bad," provided that criteria are given to define good and bad. Only one-to-one substitution of class labels is allowed, and so these classes could just as well have been labeled "1" and "18" or "pleasant" and "unpleasant," without any loss of information.

What can be said about the temperature of objects measured on a nominal scale? Not much. We are restricted to reporting the number of objects receiving the same measurement, the number of objects in each class. The silviculturist might report the temperature on 19 sites as "acceptable" and on 42 sites as "not acceptable." No conclusion can be reached about the degree of warmth or coldness from data so measured. To make conclusions regarding degree of warmth or coldness, measurements must at least be made on an ordinal scale.

Ordinal Scale

An ordinal scale is one of order or rank. It specifies the relative position of objects concerning some characteristic of interest. With an ordinal scale, observations may be arranged from smallest to largest regarding that characteristic. Like nominal scales, ordinal scales are frequently expressed in terms of clearly defined classes or categories, but ordinal scales can also be ranked by magnitude.

For example, 2 × 4 studs are graded from No.1 common to No.5 common, where No.1 is "better" than No.2 in terms of freedom from knots, crook, and wane. But these grades could just as easily be recorded A through E and still preserve the qualitative order. Similarly, an ordinal temperature scale might consist of the classes warm, warmer, and warmest; or the classes could be labeled A, B, and C, so long as the relationship C greater than B greater than A exists. This type of transformation results in no loss of information, because the order implied by the former assignment is preserved.

More can be said of ordinal-scale data, namely that the temperature of one object is not only different from that of another object, but also that it is warmer or colder. We cannot say how much temperature difference exists, however, or that this difference remains the same between classes. The sign of the difference (+ or −) is known, but neither the magnitude nor the constancy of the difference is known. These differences constitute the scale's "interval." The interval associated with the ordinal scale has been called "unequal," but probably should be called "unknown." The measurer simply does not know. To understand how much one measurement differs from another, the scale interval must exhibit a constant difference, as in either an interval or ratio scale.

Interval Scale

An interval scale is often called the "equal interval" scale. It specifies the relative positions of objects concerning characteristics of interest. The distance between these positions, the interval, is the same throughout the range of the scale. In addition to specifications identifying nominal and ordinal scales, the interval scale exhibits equality of differences or intervals. The Fahrenheit (F) and centigrade or Celsius (C) temperature scales are examples of interval scales. Because the scale's units—degrees—are represented by equal volumes of liquid expansion, a temperature difference between 32 ° and 42 ° is the same as between 202 ° and 212 °.

Interval scales can be subjected to any linear transformation of the form $Y = A + BX$. For example, when converting °C to °F we use the linear transformation

$$°F = 32 + (9/5)\,°C$$

and 18 °C becomes about 64 °F. With this conversion, the relative positions of the temperatures have remained the same even though the scale range has changed.

Interval scales are far more powerful than ordinal or nominal scales because measurements on them are additive. Because the scale interval is the same, it is possible to determine the sum of a series of measurements and calculate the arithmetic mean, the average. But some other comparisons cannot be made. For example, it is incorrect to conclude that an object with a temperature measurement of 64 °F is twice as hot as an object measuring 32 °F. This is because the interval scale has no "true" or "absolute" zero point.

Selection of the numerical zero point on an interval scale is arbitrary. In the case of temperature scales, zero might be set where water freezes (as in °C) or 32 ° below that point (as in °F). Many, perhaps most, of the measurement scales used in forestry have a meaningful, nonarbitrary zero.

Ratio Scale

A ratio scale possesses all of the interval scale's characteristics but in addition has an "absolute zero" point. Ratio scales display both equality of intervals and equality of ratios. This latter equality holds because a true zero exists. Then and only then can statements of relative comparison, such as "twice as hot," be made.

The Kelvin temperature scale, based on thermodynamics, has an absolute zero point; it is a ratio scale. If three objects were assigned Kelvin-scale values of 273 °, 291 °, and 373 °, we could conclude not only that there are 18 ° and 82 ° of equal temperature differences respectively between the objects, but that the second object is 7 percent warmer than the first and 78 percent as warm as the third.

Ratio scales can be transformed only by multiplication with a constant. Common examples of this are conversion between the measurement units in the metric and English scales. For example, a tree with a d.b.h. of 12 inches would also be 30.48 (= 12[2.54]) centimeters. The constant utilized in this example is 2.54 centimeters per inch. Clearly, ratio scales are the most powerful because the entire set of arithmetic operations can be performed on them.

Interpreting Scales

Return now to the case of measuring temperature. *Table 1* illustrates measurements taken on three objects, A, B, and C, using four different measurement scales.

Note that although all measurements are expressed as numbers, the num-

Table 1. Measurement of three objects using different scales.

Scale	Unit	Object A	Object B	Object C
Nominal	Acceptable (1), Unacceptable (2)	1	1	2
Ordinal	Warm (1), Warmer (2), Warmest (3)	1	2	3
Interval	°F	32	64	212
Ratio	°K	273	291	373

"Mathematical procedures and computers are indifferent to the origin of numbers that enter statistical computations, but that indifference should not be shared by the forester."

GAET RECEIVED
MAR 29 2000

"The profession is witnessing an array of abuses in quantification—biased sampling, ignored variability in data, and inappropriate inferences."

bers have quite different meanings. The numerical value of "2" means "unacceptable" under the nominal scale and "warmer" under the ordinal scale. Foresters do use the four scales. Examples for forestry applications are presented in the box on page 30.

A forest-inventory plot may contain measurements expressed in all four scales: dominant timber type expressed by a nominal scale; site productivity class, an ordinal scale; date of stand origin, an interval scale; and the height of a dominant tree, a ratio scale.

The problem is not the taking of measurements with different scales, but rather the manipulation, analysis, and interpretation of these measurements. One simply cannot legitimately perform all mathematical operations on measurements from the various scales. From the standpoint of arithmetic operations, only counting—not numbering—is appropriate for nominal- or ordinal-scale measurements *(table 1)*. Counting, addition, subtraction, division, and multiplication are all appropriate for the interval and ratio scales. Percentage changes are also permissible under the ratio scale, because a true zero point exists.

The appropriate statistics and significance tests also change from one measurement scale to another. A cumulative hierarchy exists. For example, the measure of central location (the center of distribution) of data measured under the nominal scale is the mode, not the median or arithmetic mean; under the ordinal scale the mode and median, not the arithmetic mean; and under either the interval or ratio scales, the arithmetic mean, mode, and median.

The appropriate measures of dispersion or variation also change from the nominal to the interval and ratio scales. Foresters commonly use the variance or its square root, the standard deviation, as a measure of variation. However, these statistics are appropriate only for interval- and ratio-scale data. Less rigorous measures of dispersion must be used with nominal- and ordinal-scale measurements. Interval and ratio scales are thought to be the more powerful of the four scales. This is because they can support more types of manipulations and analyses. Data collected under the interval or ratio scales can be converted to the ordinal or nominal scale if desired, but not vice versa. Consequently, statistics (and significance tests) appropriate for nominal and ordinal data can also be applied to interval- or ratio-scale data, but not vice versa.

Reality is somewhat more complicated, and statisticians debate the appropriateness of applying statistics suitable for interval- and ratio-scale data to ordinal-scale data. The debate focuses on how closely an ordinal scale approximates an interval or ratio scale and the extent to which statistical conclusions are sensitive to the analytical procedures used. This article reflects the traditional, and more restrictive, view on uses of measurement scales.

Abusive Practices

Practices that abuse the integrity of numbers may well originate from illusions created by the system of numbers itself. This system—cardinal numbers—is used to count virtually everything in forestry, from animal droppings to annual budgets. Although the number system is not always used in the same way, the numbers themselves always appear to be the same. The numbers 4, 5, and 6 always have the same appearance, whether used to label nominal or ordinal classes or to represent levels on an interval or ratio scale. Confusion between the appearance of numbers and the use to which they can be put may foster some abusive practices.

Ordinal means—One of the most frequent abuses involves treating ordinal numbers as if they were interval- or ratio-scale numbers. Perhaps the biggest problem is calculating arithmetic means. As indicated earlier, unless the scale is of equal interval, scale values cannot be summed, and the arithmetic mean cannot be calculated. In reality, when ordinal-scale measurements are made, they normally constitute all that is known; corresponding ratio-scale measurements do not exist. A test case is presented in the box on page 29.

Measurements of esthetic quality, public preference, personnel performance, and so on are commonly made on an ordinal scale. Equally common is the practice of calculating an "average" visual quality or employee-performance score. Ordinal scale data are insufficient to support these calculations. Use of medians and modes, not means, can solve the problem.

GAEI RECEIVED
MAR 29 2000

Transformation—From time to time foresters have occasion to estimate a linear regression expression, such as $Y = A + X$. The dependent variable Y might be stand volume, and the independent variable X might be basal area. There is no difficulty if X is measured by either the interval or ratio scales. The problem arises when X is measured on either the nominal scale (soil type 1, 2, 3, etc.) or the ordinal scale (site class 1, 2, 3, etc.).

When class designations are expressed as numbers, it is easy to use these designations incorrectly, as if they are interval- or ratio-scale values. Nominal- or ordinal-scale class designations cannot be "transformed" into interval or ratio-scale measurements. Class designations are really labels, not numbers. Nobody would consider conducting a regression analysis with X being measured by the words "low-," "medium-," and "high-" site. To replace those words by the numbers 1, 2, 3 or 2, 23, 5,000 does not fundamentally change the situation. Yet this is precisely what is being done when the computer code 321 for forest habitat type PSME/CARU-AGSP is used as a measurement on an independent variable in a regression equation.

Elements of a nominal or ordinal scale are labels, regardless of whether the labels are expressed as numbers or any other symbols. Although mathematical computations can be performed, the resulting output is meaningless. Proper use of "dummy" variables can solve that problem.

Index construction—Use of indices is increasing in forestry. In addition to traditional indices of site quality or consumer prices, indices "measuring" employee performance, site attributes, and the quality of land-management alternatives have been constructed. The topics typically addressed by indices are multidimensional. Problems arise when these dimensions are improperly combined into an index.

Index construction often involves the direct combination of measurements that may not be measured on the same scale or even in the same units. Often overlooked is the fact that the units of measure associated with a scale carry through the arithmetic operations to the results. For example, if a measurement made in feet were multiplied by one made in acres, the product carries acre-feet as its unit, a legitimate water measure.

Scale units must be compatible. If an index of recreation potential were created by combining recreation activities, persons, size of area, and site uniqueness, each unit of measure would be reflected in the index. The problem is that when some measurement units are combined, the result may carry an incomprehensible unit, such as unique-activity-person-acres.

Generally, indices based on dissimilar scales can be constructed only if the measurements being combined have been "standardized" by a transformation through which measurements become unitless. Such numbers can be combined legitimately to form an index. For example, an ecological-diversity index has been developed by summing relative species frequency, relative dominance, and relative density. Without use of relative measures to standardize each part, the resulting index would be meaningless. An index is unitless because its elements are unitless.

Another index-related problem in-

"Some disciplines have condoned calculating means of ordinal numbers for so long that individuals may no longer perceive a problem."

Test Case:

Do organizations A and B have the same average volume per acre?

Organization A	Organization B	Ordinal scale (site class)	Ratio scale (volume per acre)
..... Acres			MBF
1,000	5,000	5	5
2,000	0	4	7
6,000	2,000	3	12
2,000	0	2	20
1,000	5,000	1	32

Answer on p. 30.

CAET RECEIVED
MAR 29 2001

How Foresters Use Measurement Scales

Scale	Application
Nominal	Landforms, habitat types, personnel classes
Ordinal	Fire-danger ratings, esthetic preferences, Richter scale
Interval	pH, mean sea level, slope (degrees)
Ratio	Forb biomass, carrying capacity, tree volume

volves the weighting and mixing of scale types. The offense is committed in the process of weighting ordinal-scale measurements. Most weighting systems are inherently ratio scale. A weight of 2 is twice that of 1; 1.0 is five times that of 0.2. Because the interval between ordinal-scale measurements is unknown, the effect of weighting—emphasizing or deemphasizing—these intervals cannot be determined.

Indices commonly have several elements. An employee-performance index may use ordinal ratings on supervision, output, quantity, and quality. The result of weighting these ratings and then aggregating can be totally incomprehensible. Our point is not that weighting always aggregates problems, but rather that the effect of weighting cannot be determined without knowledge of the scale's intervals.

False Security

A range of mathematical operations can be performed on numbers *(table 2)*. The problem with inappropriate operations is that although they can be conducted, the results cannot be interpreted. It is possible to create an unlimited number of indices that have absolutely no meaning. It is mechanically possible to total a series of ordinal numbers and divide by the number of observations; but an arithmetic mean, as traditionally interpreted, has not been calculated.

It is possible to code nominal or ordinal classes as numbers and perform the manipulations normally conducted as a regression analysis; but regression coefficients will not be calculated. Pseudoindices, pseudomeans, and pseudocoefficients result from inappropriate mathematical operations.

Mathematical procedures and computers are indifferent to the origin of numbers that enter statistical computations, but that indifference should not be shared by the forester. Although procedures are blind, the forester can see the nature of numbers. Yet, individuals are not solely responsible for abuse. Some abuse has been institutionalized in standard operating procedure; for example, some disciplines have condoned calculating means of ordinal numbers for so long that individuals may no longer perceive a problem.

Measurement scales evolve over time, based on the state of the art and the changing information needs of managers. As information needs become more sophisticated and demanding, measurement scales must evolve to meet the need, knowledge permitting. But in the meantime, the state of the art may be deficient. We simply may not know how to measure esthetic quality or employee performance on a ratio scale. That is tolerable. Foresters have always had to deal with incomplete knowledge. We commit no intellectual infraction by correctly interpreting data, given the current scale of measurement.

Incorrect interpretations produce bogus information and worse. Misuse of measurement scales and performance of inappropriate mathematical operations lead to a false sense of security—the belief that more and better information exists than actually does. The incentive to push back the frontiers of ignorance is unwittingly reduced or eliminated. Although correct use of measurement scales in forestry will not by itself ensure high-quality information, it is a prerequisite. ■

Table 2. Arithmetic operations and statistical measures associated with various measurement scales.

Scale	Arithmetic operations	Statistical measures: Location	Statistical measures: Dispersion
Nominal	Count	Mode	Diversity index*
Ordinal	Count	Mode Median Deciles Percentiles Quartiles	As above, plus Minimum/maximum values Interquartile range
Interval	Count Addition Subtraction Division Multiplication	As above, plus Midrange Arithmetic mean Geometric mean Quadratic mean Harmonic mean	As above, plus Range Variance Standard deviation Coefficient of variation Standard error Absolute mean deviation
Ratio	As above, plus Percentages	As above	As above

This measure of dispersion is also known as "Information H" (Zar 1984).

Suggested Reading

CONOVER, W.J. 1971. Practical nonparametric statistics. 462 p. John Wiley & Sons, New York, NY.

KIRK, R.E. 1972. Statistical issues—a reader for behavioral sciences. 401 p. Brooks/Cole Publ. Co., Monterey, CA.

SIEGEL, S. 1956. Nonparametric statistics for the behavioral sciences. 312 p. McGraw-Hill, New York, NY.

STAFFORD, S.G. 1985. A statistics primer for foresters. J. For. 83:148–157.

STEVENS, S.S. 1968. Measurement statistics and the schemapiric view. Science 161:849–856.

ZAR, J.H. 1984. Biostatistical analysis. Ed. 2. 718 p. Prentice-Hall, Englewood Cliffs, NJ.

Answer to test, p. 29: The site-class scale (ordinal) has an unequal interval in terms of the volume-per-acre scale (ratio), although the intervals appear equal—1, 2, 3, 4, 5. If the average (arithmetic mean) site class were calculated, the conclusion would be "yes," class 3 in both cases. If the mean volume per acre could be determined from the ratio-scale information, the conclusion would be "no"; this conclusion is correct.

DATE RECEIVED