The Choice of an Appropriate Statistic: A Nonmathematical Approach

William C. Mann

Key Words: research • statistics

Statistics appear with increasing frequency in the occupational therapy literature. At the same time many therapists do not understand the reasons for selecting certain statistics for a study. This article outlines an approach to selecting statistics that can be used with many research studies. The approach is useful in evaluating the choice of statistics in a published study or a research proposal or in planning a study. The presentation is conceptual, illustrated with examples from occupational therapy literature.

William C. Mann, PhD, OTR, is Associate Professor and Director of Graduate Studies, State University of New York, Buffalo, NY 14214

Occupational therapists are generally more concerned with treatment approaches than with mathematical analyses. However, to be an informed consumer of our literature, which often examines treatment approaches, an understanding of the selection and application of statistics for research in occupational therapy is necessary. As Ottenbacher and Peterson (1985) have demonstrated, the percentage of articles in the American Journal of Occupational Therapy (AJOT) that present statistical information has increased significantly in the past 10 years. The present paper discusses guidelines for understanding why an author chose a particular statistic for the study. It also presents four tables that can serve as quick references for selecting the appropriate statistic (for those planning to conduct a study). Thus, the purpose of this paper is to present a basic outline for statistic selection; the approach is based on the measurability of variables in the study.

The American Journal of Occupational Therapy has published several articles that assist therapists in understanding our literature and in gaining an introduction to the concepts used in our research. Crocker (1977) provided a helpful approach to reading research literature. Ethridge and McSweeney (1970a, 1970b, 1971a, 1971b, 1971c) covered the basics of occupational therapy research in a series of five articles. Even earlier, Mowrer (1960) provided an overview of research methods, statistics, and theory. Hasselkus and Safrit (1976) discussed measurement in occupational therapy. A number of texts are also available which cover research and statistics in more detail (Glass & Stanley, 1970; Kerlinger, 1973; Payton, 1980).

Before any sense can be gained from reading about statistics—just as in reading about anatomy or woodworking—a basic vocabulary, unique to the subject, must be understood. Once this vocabulary is mastered, decisions about which statistic to select become simple. Already, the terms variable and measurability have been used. Definitions are needed.

A variable is any trait, quality, behavior, attribute, attitude, or type of treatment—in fact, anything that varies or changes from one person or thing to another. In any group of people, fine motor coordination will vary among people in the group; thus, fine motor coordination is a variable. In the January 1986 issue of AJOT, the following variables were studied: work simulation, back injury, affective meanings of life review, occupational therapy student learning cycles, occupational therapy student clinic performance, and chewing cycles.

In studying variables, it must be recognized that they can be classified in more than one way. First, some variables are measured while others are cate-
gorical. Measured variables are variables for which numbers can be assigned as a result of testing (e.g., intelligence described by IQ scores) or by the variable's universally accepted meaning (e.g., age described in years or months). An example from the list above will illustrate the point. Vestibular function can be measured by the Southern California Sensory Integration Tests (SCSIT) and is therefore a measured variable. For any individual tested, it is possible to arrive at a number, a measurement, that represents the individual's vestibular function. However, since the SCSIT was developed for a specifically defined group, the SCSIT scores for many individuals might not be an accurate measure of vestibular function. Age, measured in years, months, or days, is a measured variable. When we are told that someone is 30 years old, we know how much time, in the form of a number, has passed since the person was born.

While a measured variable can be represented by numbers that have meaning to some group of people, categorical variables, although often represented by numbers for coding purposes, only take on numerical values whose meaning is known to the researcher or "coder." A categorical variable separates individuals or things into groups, but the differences between the groups cannot be measured. For example, diagnosis is a categorical variable. It is not practically possible to measure, or to assign a number to, the difference between cancer, heart disease, diabetes, and hip fractures. But the differences on this variable can be explained, rather than given a single number to represent them. It should be noted that at times people do assign numbers to categorical variables for convenience in data storage and retrieval; this is called coding.

Textbooks often refine the concepts of categorical and measured variables even further (Glass & Stanley, 1970; Kerlinger, 1973). Typically four classes of variables are identified: nominal, ordinal, interval, and ratio. Nominal level variables are essentially the same as categorical variables, and ratio level variables are the same as measured variables. All ratio level variables have an absolute zero point. For example, range of motion, measured in degrees, has an absolute zero point. If an individual flexes at the elbow from a fully extended position held horizontal to the floor to a position at a right angle to the floor, it is agreed that the person has moved 90°. But if the person has severe contractures and cannot move at the elbow joint, we agree that this would be 0° of movement—absolute zero. On the other hand, temperature, measured on a Fahrenheit scale, does not have an absolute zero point. In measuring temperature, we are actually measuring the amount of heat, and 0 °F is not a total absence of heat (although it is cold); a temperature of -5 °F would not be possible if 0 °F reflected a total absence of heat. Temperature, measured on the Fahrenheit scale, is actually an interval level variable, and differs from a ratio level variable only in that the zero point is not absolute.

Ordinal level variables differ from both interval and ratio level variables in that there is no consistency in the amount of difference between any two points on the scale. With range of motion, the amount of distance between 80° and 81° is the same as between 3° and 4°. With ordinal level measurement, we know that the values of the variable have been ranked, and each adjacent value reflects more (or less if the list is reversed) of the property than the previous value, but the amount of difference between two values may not be the same as the amount of difference between two other values. For example, in measuring attitude toward work, we could ask people to state how much they like work (not at all, some, very much). It would be wrong to say that the amount of difference in liking work between "not at all" and "some" was equal to the difference between "some" and "very much." A number of occupational therapy studies use scales that reflect ordinal level measurement. Statistics have been developed specifically for ordinal level measurement, although some researchers treat ordinal level variables like interval or ratio level variables. Glass and Stanley (1970, p. 13) outline the dilemma of considering IQ scores as either ordinal or interval levels of measurement.

For this paper, ordinal level measurement will be considered as "categorical." We will also refer to nominal level variables as categorical variables, and interval and ratio level variables as measured variables. This follows the classification scheme, described by Payton (1980), in which statistics are considered "parametric" or "nonparametric."

Another way to classify variables is as either dependent or independent. In most studies of two or more variables, the researcher is interested in the relationship of one or more variables to another variable or to several other variables. The variable changes (or is hypothesized to change) in relation to changes the researcher makes in the independent variable. For example, in testing the effectiveness of sensory stimulation with geropsychiatric inpatients on several variables including reality orientation, Paire and Karney (1984) "changed" or "manipulated" the treatment variable—the independent variable with three levels: (a) sensory stimulation, including staff attention, (b) staff attention only, and (c) standard hospital treatment. In another example, Beck and Callahan (1980) looked at two levels of institutionalization, continuous and intermittent, in relation to the posture of chronic schizophrenic patients. Posture is the dependent variable, since the authors hypothesized that posture would vary with level of institution-
alization (the independent variable that they controlled or set up).

To complete the discussion of variables, let us combine the concepts of dependent and independent variables with measured and categorical variables. Each variable must be some combination of these two concepts: measured independent, categorical dependent, measured dependent, or categorical independent. Looking at the two examples above, the sensory integrative treatment variable is a categorical independent variable; it is categorical because there are only two levels, pretreatment and posttreatment. The reading variable is a measured dependent variable. In the second example, level of institutionalization is a categorical independent variable, and posture, for which graded scores were assigned, is a measured dependent variable. Readers of research literature will find it helpful to describe each variable in terms of measured versus categorical and dependent versus independent.

Descriptive Statistics

In doing research with groups of subjects, we must offer descriptions of the groups we are studying. We do this with descriptive statistics, which are numbers that summarize the large amount of data we have. Two major categorizations of descriptive statistics are measures of central tendency and measures of variability. Table 1 lists the most common measures of central tendency and of variability. In most studies, even when higher level statistics are employed, these statistics will also be used to describe the sample or to help analyze differences or similarities between two or more groups. Groups of subjects are often described collectively—a mean age or a mean IQ, for example—and standard deviations are also presented.

Measures of central tendency are single numbers used to represent and summarize a large group of scores on a measured variable and are designed to give an idea of the middle point, the most frequently occurring score, or the average of a group of scores. The mean is simply an average of all scores; add all scores and divide this sum by the number of scores. The median is the point at which half the scores are smaller and half the scores are larger. The mode is the score that occurs most frequently. If we are working with a group of scores that has a normal distribution, the mean, median, and mode will be equal. However, in taking samples, this is not always the case. The mean is the most widely used measure of central tendency and serves as the basis for many of the other higher level statistics.

Measures of variability provide an indication of how spread out a set of scores are for a measured variable. The range is simply the number of points or the distance between the lowest score and the highest score. In looking over a group of 50 Barthel Index scores, it was found that the low score was 35 and the high score was 100; thus, the range is 65 points. The mean deviation is simply the average distance of scores from the mean. One simply calculates the mean for a group of scores, determines how far each score is from the mean (positive value or absolute value), adds up the distances from the mean, and divides by the number of scores. For the above Barthel Index scores, the mean was 75. The mean alone does not indicate if the scores were close to the mean or close to the low and high scores, or something in between. The mean indicates nothing about the spread of scores. But if you know that the mean deviation is 6, then you know that the average distance of the scores from the mean is 6 points above or below the mean.

The mean deviation provides a simple explanation and example of a measure of variability. We all understand the idea of "averaging" a set of scores, such as test grades in a course. For the instructor, after giving a class a test, knowing both the average or mean and something about the spread of scores, such as the mean deviation, is useful, because not only does the instructor know what the mean score is, but it is possible to tell if the students all performed equally well (small mean deviation) or if there was much difference in the performance of students on the test (large mean deviation).

While the mean deviation is useful in illustrating the concept of a measure of variability, it is rarely used in reporting research. Instead, the variance and standard deviation are used; they essentially serve the same purpose, although they are a little more difficult to understand than the mean deviation. The reason lies in the formulae. For mean deviation, the absolute value, or positive root, of numbers must be taken. The variance relies on the fact that the square of any number, negative or positive, is a positive number. The formula for mean deviation is very similar to the formula for variance, except that in the numerator of the formula for variance squaring is used in place of taking the absolute value. While formulae are beyond the scope of this paper, the concept of variance as a measure of the spread of scores must be

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Descriptive Statistics for Use With One Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measures of Central Tendency</td>
<td>Measures of Variability</td>
</tr>
<tr>
<td>Mean: average; add all scores and divide sum by number of scores.</td>
<td>Mean Deviation: average distance of scores from the mean.</td>
</tr>
<tr>
<td>Mode: most frequently occurring score.</td>
<td>Variance: average squared distance of scores from the mean.</td>
</tr>
<tr>
<td>Median: middle point; half the scores are above the median, half below.</td>
<td>Standard Deviation: square root of variance.</td>
</tr>
<tr>
<td>Range: distance between lowest score and highest score.</td>
<td></td>
</tr>
</tbody>
</table>

October 1985, Volume 40, Number 10

Downloaded From: http://ajot.aota.org/ on 11/09/2018 Terms of Use: http://AOTA.org/terms
clear in order to proceed. Like the mean, the variance also serves as the basis for many of the higher level statistics. If you understand the concept of mean deviation, think of the variance as another statistic that provides an indication of the spread of scores around the mean.

Standard deviation is one of the most widely reported measures of variability. It is simply the square root of the variance. The standard deviation is a more easily interpretable measure of variability than the variance. If we calculate the value of the standard deviation for a set of scores, then (a) add it to the mean and (b) subtract it from the mean, those two numbers (called standard deviation above and below the mean) represent the range in which we can expect to find about two thirds of our subjects’ scores, if we have a normal distribution of scores. If we go 2 standard deviations above and below the mean, we will find about 95% of our subjects’ scores, and taking 3 standard deviations gives us a range in which we find more than 99% of our subjects’ scores. Thus, the standard deviation provides an excellent description for spread of scores: it is an excellent measure of variability.

Inferential Statistics

Inferential statistics are used in studies with independent and dependent variables. They are based on rules of probability or chance. In fact, inferential statistics were first developed by mathematicians working for gentlemen gamblers who wanted to maximize their chances of winning. In research we simply want to know the chances of getting the results (numbers) we got purely by chance. If the odds are very low (say 1 in 100), then the results must be due to something other than chance: the effect of the independent variable(s) on the dependent variable(s). Thus, we set a level of confidence, usually expressed as the probability of getting the results by chance, such as .05 (to be 95% certain the results did not occur by chance) or .01 (to be 99% certain or “confident”).

Studies that have categorical independent variables would most likely use one or more of the statistics listed in Table 2. If the study contains a categorical independent variable with only two levels, and a measured dependent variable, the t test would be used to test for significance of difference between the means for each of the two levels on the dependent variable. For example, Tickle and Yerxa (1981) studied need satisfaction (dependent variable) in relation to where older people lived (independent variable with two levels: community and institution). A mean score for need satisfaction was determined for the group of subjects who lived in the community, and this was compared to the mean score for need satisfaction of the older people living in institutions. To determine if the size of the difference between the two mean scores for need satisfaction was due to chance or a result of the effect of the independent variable (where the older people lived), a t test was employed.

Note that the t test is used only when there are two levels of the categorical independent variable. Given a study with a measured dependent and one categorical independent variable with three or more levels, the t test cannot be used; in this case the one-way analysis of variance (ANOVA) is the appropriate statistic. Taking the same example above, suppose Tickle and Yerxa had decided to study differences in need satisfaction of older people living (a) in their own homes, (b) in the homes of relatives or friends, and (c) in institutions. Now the categorical independent variable has three levels, and, to test for differences in need satisfaction among the three subgroups of subjects, a one-way ANOVA is used. Note that both differences in the means (central tendency) and in the spread of scores around each mean (variability) are considered in the formulation of t test and ANOVA. Sample size also is considered in the formulation.

If the discussion of statistics is understood to this point, then the remainder of Table 2 should be easily comprehended. If a study has more than one categorical independent variable and one measured dependent variable, then we would use an N-way ANOVA, with N equal to the number of categorical independent variables. Let’s take the above example again, but this time we are interested in differences in need satisfaction of older people (subjects) grouped not only by where they lived but also by sex and social class. We now have three independent variables, all categorical, and one dependent variable. For this analysis we would use a three-way ANOVA. We would be looking at the means and variances for need satisfaction between (a) subjects living in the community versus those living in institutions, (b) male versus female subjects, (c) lower, middle, and upper socioeconomic class subjects, and (d) interactions, such as

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Inferential Statistics for Use in Studies With Categorical Independent Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>One measured dependent variable</td>
<td>t test (with 2 levels)</td>
</tr>
<tr>
<td>More than one measured dependent variable</td>
<td>One-way MANOVA</td>
</tr>
<tr>
<td>Categorical dependent variable(s)</td>
<td>nonparametric statistics such as chi-square</td>
</tr>
</tbody>
</table>

The American Journal of Occupational Therapy

699
Groups to Be Studied in Three-Way ANOVA Example: Where Older People Lived

Figure 1

<table>
<thead>
<tr>
<th>Subjects Living in Community</th>
<th>Subjects Living in Institutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male Upper Class</td>
<td>Male Middle Class</td>
</tr>
<tr>
<td>Female Upper Class</td>
<td>Female Middle Class</td>
</tr>
</tbody>
</table>

Note that we will be looking at 12 “cell” means, as well as two means for sex, two for “where old people lived,” and three for social class, in this analysis. By doing the ANOVA, we will be able to say with some degree of certainty (say 95%) that the differences we found were not due to chance; or we will determine that the differences were more likely due to chance.

If a study has more than one measured dependent variable that can be collectively considered as contributing to some larger concept, then the study becomes multivariate. Hasselkus, Kshepakaran, and Safritz (1981) studied the relationship of handedness (categorical independent variable with two levels, right and left) to hand joint changes. The concept of “hand joint changes” was measured considering 10 separate variables, including “passive MCP lateral mobility of digits 2–5, passive MCP hyperextension of digits 2–5, passive thumb MCP range of motion, and lateral pinch strength” (p. 707). Since these “hand joint changes” variables are dependent variables and can best be considered collectively, the statistic employed in this study was a one-way multivariate analysis of variance (MANOVA).

In the example above, had there been two or more categorical independent variables (e.g., handedness, sex, occupation), then the appropriate statistic would be a three-way MANOVA (or two-way MANOVA, N-way MANOVA, depending on the number of categorical independent variables).

Continuing with this exploration of Table 2, a major change occurs when both independent and dependent variables are categorical. We can no longer rely on means or variances as the basis for our analyses. Means and variances cannot be calculated for categorical variables. Consider a mean of several ZIP codes, the number would be meaningless. What is the mean for “handedness,” or the mean for “where people live” in the examples above? Obviously, there can be no means for categorical variables, so a different approach must be taken. The statistics used with this approach are called nonparametric statistics.

One frequently used nonparametric statistic is the chi-square statistic (Glass & Stanley, 1970; Kerlinger, 1973). Chi-square considers the relative observed frequency with the expected frequency of subjects, events, or items in a cross-tabulation table. For example, if we were studying activity preferences among three different ethnic groups to determine if ethnicity was an important factor to consider in activity selection in an occupational therapy clinic, we could survey several members of each of several different ethnic groups, asking them to state their favorite activities. Since both activity and ethnicity are categorical variables, we can only compare differences in numbers of people, broken down by ethnic group, who choose each of the activities. To determine if the relative differences in choice of activity are the result of chance or are a result of differences resulting from ethnic preferences, we can use the chi-square statistic. There are several other nonparametric statistics useful when the variables meet certain criteria. These statistics are described in most basic statistics texts (see Ethridge & McSweeney, 1971a).

Table 3 shows the selection of statistics for studies with measured independent variables. The simplest case is where we have one independent and one dependent variable, both measured. Correlation analysis would be applied. In fact, correlation is one of the most widely used statistics in our literature. Katz and Mosey (1980) studied the relationship between fieldwork performance and academic course work, both of which were considered measured variables. Using correlation analysis, they calculated correlation coefficients for each course grade (independent variable) with each of the fieldwork performance ratings (dependent variable). Through this analysis they determined that some course grades did show a significant relationship to fieldwork performance, while others showed no relationship, and they were further able to determine the strength of the relationship, which in most cases was weak.

A correlation coefficient is in fact a descriptive statistic, summarizing the relationship between two variables. However, inferential statistics can be applied to the analysis to determine the probability of the results occurring just by chance.

In studies that have more than one measured independent variable and one measured dependent variable, multiple correlation—more commonly called multiple regression analysis—is used. Often,
but not exclusively, this analysis would be used to study the relationship of several variables on one other variable, possibly for the purposes of prediction. For example, if Katz and Mosey had been interested in predicting the fieldwork performance rating, they would have needed to add more variables than just course grades to the analysis. If they had added IQ scores, personality scores, and age to the analysis, they would have had four measured independent variables and one measured dependent variable. The appropriate statistic would have been multiple regression.

In the situation where we have more than one measured dependent variable and one measured independent variable, we could simply reverse the independent and dependent variables and use multiple regression. When we have more than one measured dependent variable and more than one independent variable, we would likely choose a multivariate approach with multiple regression. Taking the example above, we might wish to consider the two fieldwork performance ratings (physical disabilities and psychosocial) simultaneously in an analysis with the four independent variables (grades, IQ, personality, and age). In this case we would have a multivariate multiple regression analysis. Canonical correlation is a specialized multivariate regression approach, often used when there are large numbers of variables in the analysis.

The final situation to consider from Table 3 occurs in studies with categorical dependent variables and measured independent variables. Here we either use nonparametric techniques or, in some cases, reverse our dependent and independent variables and then use parametric statistics.

One last research situation which arises often is having a measured dependent variable(s) and having both measured and categorical independent variables. For this analysis we would likely select an analysis of covariance.

### Table 3

<table>
<thead>
<tr>
<th>Categorical dependent variables</th>
<th>More Than One Measured Independent Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>One measured dependent variable</td>
<td>Correlation</td>
</tr>
<tr>
<td>More than one measured variable</td>
<td>Reverse independent and dependent variables use multiple regression (Multivariate multiple regression/canonical correlation)</td>
</tr>
<tr>
<td>Nonparametric techniques</td>
<td>In some cases the independent and dependent variables can be reversed, permitting use of parametric statistics</td>
</tr>
</tbody>
</table>

**Conclusion**

This paper has presented an approach to selecting the appropriate statistic. If one knows whether the variables are measured or categorical, independent or dependent, then it becomes easy to select the correct statistic. There is nothing mysterious about the choice. Rules are followed, and rules can be learned. This overview is an introduction to these rules.

**References**


