DOCUMENT RESUME

ED 048 360 TM 000 438

AUTHOR Cassel, Russell N.
TITLE The Precision-Power-Gradient Theory for Teaching Basic Research Statistical Tools to Graduate Students.
PUB DATE Feb 71
NOTE 15p.; Paper presented at the Annual Meeting of the American Educational Research Association, New York, New York, February 1971

EDRS PRICE EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS Analysis of Covariance, Analysis of Variance, *Behavioral Science Research, Correlation, Discriminant Analysis, Factor Analysis, *Graduate Students, Hypothesis Testing, Research Methodology, *Research Tools, *Statistical Analysis, *Statistics

ABSTRACT

This paper relates educational and psychological statistics to certain "Research Statistical Tools" (RSTs) necessary to accomplish and understand general research in the behavioral sciences. Emphasis is placed on acquiring an effective understanding of the RSTs and to this end they are are ordered to a continuum scale in terms of individual precision, power, and robustness. The Precision-Power-Gradient theory is suggested as a means of structuring an introductory graduate course in statistics. Twelve RSTs are considered, ranging from normative comparisons (at the simplest level), through parametric/non-parametric statistics, analysis of variance and covariance, correlations, factor analysis, to canonical correlation analysis, and, briefly, other advanced techniques. Equally important is a working knowledge of the computer, which enables the RSTs to be used with greater facility and accuracy. Computer orientation should include a basic knowledge of FORTRAN programming, nature and availability of computer-based statistical libraries, the ability to utilize such libraries, with a general understanding of computer output. (Author/AE)

THE PRECISION-POWER-GRADIENT THEORY FOR TEACHING BASIC RESEARCH STATISTICAL TOOLS TO GRADUATE STUDENTS

Russell N. Cassel
University of Wisconsin-Milwaukee 53201

This paper is concerned primarily with relating educational and psychological statistics to certain "Research Statistical Tools" (RSTs) necessary to accomplish and understand general research in the behavioral sciences. It considers the RSTs as a means or vehicle to accomplish and understand research, not as ends in themselves, and seeks to present them in just such a functional setting.

Emphasis is placed on having the individual acquire an effective understanding of such RSTs as are necessary for accomplishing and understanding general research in the behavioral science areas. In order to facilitate the meaningfulness of such RSTs, they are ordered to a continuum scale in terms of their individual precision, power, and robustness.

Equally important to the RSTs is a working knowledge of the electronic high speed computer; for it is this new vehicle that enables the scientist to use the RSTs with much greater facility and accuracy than ever before. Such computer orientation should include a basic knowledge of FORTRAN programming, nature and availability of computer-based statistical libraries, the ability to utilize such libraries, with a general understanding of such computer output.

Law of Parsimony

The law of parsimony states that "When two different theories explain a phenomena equally well, the task scientist is to use the simpler of the two". This is not intended to imply that the precision, power, or robustness of a particular RST will serve as the basis for its selection or use. Often the decision as to which particular RST is to be used may be based on an empirical test of which one does the job. Recently, for example, at the University of Wisconsin-Milwaukee (UWM) a leadership experiment was conducted involving a sizeable number of ROTC students. A t test between the means (Ms) for the pre- and post tests failed to support the hypothesis. It was noted that a great number of students involved reflected small differences in support of the hypothesis, and that the Sign Test from RST III - Chi Square and Nonparametric Statistics supported the hypothesis at the 01 per cent level of confidence. More often than not, however, it is the t test that supports the hypothesis when Chi Square fails to reveal such significance of difference to be present.

Levels of Confidence

Evaluation efforts have typically concentrated on measurement of outcome variables, especially those that were specifically stated as objectives of the program. Often this is accomplished by comparing obtained differences between scores on a pretest and scores on a posttest. Generally, it resolves to a consideration of probability statistics,"Could a difference as great as the obtained one be expected to have occurred by chance?" If such a difference could not have been expected to have occurred by chance, then the theoretical hypothesis for the study is supported and accepted. This, to be sure, entails the establishment of "levels of confidence" or "alpha" levels, i.e., the probability level at which we accept the theoretical hypothesis.

Statistical Significance Level

The characteristic alpha levels used in the professional literature range from .10 to .003, and with .05 and .01 being most frequently used for accepting the theoretical hypothesis in a study. There is no logical reason why a confidence level needs to be adopted prior to the accomplishment of a finding; for an individual is evaluating the finding-not the statistic-nor his own integrity.

Practical Significance Level

Fisher addressed himself to the distinction between "statistical inference" and "scientific inference". It would seem that statistical inference is a necessary, but not a sufficient condition for scientific inference; not unlike reliability and validity in measurement. For those who quote Fisher as their authority for use of the .05 level for accepting a theoretical hypothesis, it should be remembered that Fisher maintained the rule that a statistic should be three or more times it's standard error (SE) to be statistically significant at the .01 level, and that when the critical ratio (CR) is 3.00, the confidence level is actually .003.Thus, we can say that statistical inference obtains at the .05 or .01 levels, but that practical significance obtains above .001.

Variance Vs. Invariance

Too often we are concerned with single studies, and we are inclined to accept or reject the theoretical hypothesis on the basis of such findings, only to find in subsequent application of such findings similar results do not obtain. "Variance" refers to range of differences obtained in single studies, while "invariance" refers to corresponding results for multiple studies.

Replication Vs. Cross-validation

The question as to what information should be included in the report of a research study is determined on the basis of answering the question of whether such information is essential for a replication of the study. If the ultimate goal of the task scientist is the "discovery of universal truths" then not only replication but cross-validation is essential. In replication the precise conditions are duplicated; while in cross-validation perimeters may be expanded towards more universal testing.

One- or Two-Tailed Test

The question of whether to use the one- or two-tailed test is a decision made by the task scientist; not unlike the selection of a level of confidence for accepting the theoretical hypothesis of the study. If it appears reasonable to expect findings only in one direction,as in the study for improvement of reading, for example, the one-tailed test should be used. Such decisions, however, are highly critical to 'level of confidence".

Nature of Gradient

The "Precision-Power-Gradient" theory is an expedient construct for introducing educational and psychological statistics to graduate students in a meaningful setting, and based on the characteristic critical attributes of such procedures. Such attributes, to be sure, involve precision, power, and robustness in the same configuration; not as separate entities.While they may be examined separately, they can not be used separately, but only in a total context.

Precision

The notion of "precision" as it is used in this paper entails fineness or preciseness of measurement. The real "keystone" on which scientific progress, or scientific findings rests is measurement, and the keystone of all measurement is the scale. No statistics, however precise, may be substituted for such keystones. The percentile score, for example is not additive, and no statistics can be used to compensate for that inadequacy. Three principal considerations are involved in the use of precision as it is used in this paper: (1) the preciseness of mathematical treatment afforded, (2) the yielding of diagnostic indicators, and (3) the actual generation of findings that are present in study.

Power

The concept of "power" and precision have much in common, as they are related to statistical inferences, but power as it is used in this paper refers to the extent of domain over which it makes assessment, and how it interrelates the attributes of that domain. Three separate considerations are included in the concept of power: (1) the number of variables that may be analyzed at once, (2) the degree to which variables being analyzed are integrated, and (3) the meaningfulness of obtained findings.

Robustness

This deals with the kind of conditions under which statistical procedures may be utilized, and it involves the notions of "assumptions" and "presumptions". Too often this area receives the principal attention in graduate level courses of statistics; while there is little or no agreement among the experts as to precisely what is involved, or how the task scientist treats such involvements. The term "assumption" as it is used in this paper implies that the scientist make empirical tests to determine if basic conditions are met for the use of a statistical procedure; while "presumption" implies that no empirical test is necessary.

Gradient

This, to be sure, implies a construct with the properties of an "ordinal" scale, where one end of the scale includes considerations less precise and powerful, and with the other end including those more precise and powerful.

RST Repertoire

Statistics, not unlike many other scientific areas, is a rapidly growing and expanding field of inquiry. No course, however complete, could be expected to cover all statistical procedures. The present proposal is concerned largely with the introductory graduate course in statistics. It seeks to group the basic statistical procedures into twelve "Research Statistical Tools"(RSTs) which are believed to be essential for accomplishing and understanding general research in the behavioral sciences. The notion of RST implies a research function that is presumably independently organized, but which may be accomplished by a wide variety of statistical procedures, and in a variety of ways.It is designed to relate mathematical procedures to a functional setting that is pertinent to the law of parsimony, and to the precision and power of the RST involved.

RST I - Normative Comparisons

In as much as all phenomena in nature, including educational and psychological considerations, distribute on the Gaussian normal probability curve, comparisons of obtained data with such a hypothetical probability curve tends to give such data. It serve to infer the degree to which it deviates from the universe of which it is a part. It suggests the degree to which selective sampling has been operating. At the simplest level it merely includes a personal observation of the difference between an obtained score and the mean (M) score for a corresponding population. Or it may include comparison between an obtained standard deviation (SD), and the SD for some corresponding population. At a more sophisticated level it may make comparisons between measures of central tendency, for example, to depict the nature of a frequency distribution, i.e., where such measures coincide, the distribution is bisymmetrical and normal; where such measures have disparity, the distribution is skewed with the M determining the tail and the mode the head or peak of the distribution. The degree of disparity of such measures of central tendency is an indication of the degree of skewness. This RST may be used for matrix sampling and comparisons as well; for determining the presence of kurtosis; or even depicting evidence of "test compromise"-where M increases and SD narrows.

RST II - Two Variable Parametric Hypothesis Testing

This RST is concerned largely with the t test and the F test where ten or more observations are present (preferably 25 or more), where Ms are from populations that are fairly normally distributed, where the variance in the two populations is fairly homogenous, and where the dependent variable is continuous. The t and F-tests are comparable measures, as the $(t\ test)^2$ =

F-test. They are arrived at, to be sure, by different statistical procedures. The t test always involves determining the difference between two different statistics (means, medians, correlations, proportions, etc.), and then dividing that difference by the standard error (SE) of the difference. Thus, it is a standard deviation measure divided by a SE. Since the SE is in large part a function of the number of cases, or the degrees of freedom, we can assume that the larger the population, the smaller the SE. The critical question to which the task scientist is addressed is whether or not the obtained difference is greater than chance, and if so the theoretical hypothesis is supported and retained. There is a considerable amount of evidence to suggest that if the number of cases involved is 30 or greater, that significant skewness does not tend to invalidate the results. There is an equal amount of evidence to suggest that nonparametric methods are even less applicable where the underlying assumptions for the t test are not present (normality, variance, and continuity). For example, the Whitney U test is a rather complicated function of the mean, the skew, and the kurtosis; so that a positive skew tends to decrease the obtained U, and a negative skew to increase the U.

RST III - Chi Square and Simple Nonparametric Statistics

The chief difference between the statistical procedures in RST II and RST III is that in the former we are testing one difference at a time; while in the latter many differences may be tested at once. Because chi square, the sign test,and other similar distribution-free statistics deal with more than one difference at the same time, it is placed higher on the "Precision-Power-Gradient" scale, and becomes RST III. The outcomes for this RST are always based on a difference between "expected" (on basis of normal probability) and obtained frequencies. Most human behavior is multiple determinant in nature, and where multiple differences are being assessed in terms of crude comparisons with expected frequencies, chi square is an appropriate approach. The chief criterion for determining whether a statistic is parametric or nonparametric is whether or not the populatiors involved meet the basic assumptions for parametric statistics,i.e., (1) measures are independently drawn from the population and are independent of each other; (2) measures are drawn from populations that are normally distributed; (3) populations have substantially the same variance; and (4) measures are continuous with equal intervals. For nonparametric s t atistics, including chi square, there is but one assumption, and it is treated like a presumption-not tested empirically, and it the assumption of "independence",i.e., that data is independently drawn from the population. Many of the statistical procedures, like rho, phi correlation, and others may be used either as parametric or nonparametric statistices; depending solely on the degree to which data being analyzed meets the assumptions for the parametric statistic interpretation.

RST IV - Analysis of Variance and Covariance

Like chi square analysis of variance (A of V) deals with two or more variables at the same time, and the data must meet the assumptions for parametric statistics. Unlike many of the nonparametric statistics, A of V compares each score against the M for each group, and the M for each group against the grand M. This, to be sure, is much more precise than grouping of scores into expected frequency categories, and making comparisons by groups of scores.Always, in the A of V comparisons are made between intra-variance and inter-variance; the variance within groups as opposed to the variance between groups. In the two-way A of V in addition to the main effect, one is able to assess an interactional effect of one variable against a second. In the factor A of V one is able to permit the separation of information; to study interaction effects; to assess implications of multiple variables against expected outcomes of a dependent variable, and to test several hypotheses simultaneously. In analysis of covariance one is able to control independent varisbles statistically by adjusting Ms through the use of regression analysis. Factor A of V, and which includes two-way A of V, provides some basis for inferring cause and effect relationships through the interaction effects. Thus, A of V is not only more precise, but also more powerful than most of the nonparametric techniques. There is a preponderance of evidence which suggests that the assumption of "homogeneity of variance" should be considered as a presumption, and need not be tested empirically.

RST V - Linear Correlation of Paired Variables

The correlation technique provides a means for assessing the relationship between two arrays of paired variables, and of determining the degree or magnitude to which one of the variables in the pair tends to increase or decrease with the second one. The pairs are usually labeled "X" and "Y", with the "Y" standing for the dependent one, or the one being predicted; and with the correlation of X on Y being precisely the same as for Y on X. When the rate of change between the X and Y variables is uniform, the correlation is linear, and a straight line may be drawn through the Ms for the rows and columns. The obtained coefficient of correlation is the degree to which Ms for the related rows and columns are in agreement with each other based on data being assessed. The obtained correlation coefficient is in part a function of the ranges of the bivariate distribution, with greater ranges tending toward larger r's. High positive coefficients of correlations (above .90, for example) tend to be associated with negatively skewed distributions; while high negative ones are typically associated with positively skewed distributions. By squaring the obtained r's we have the amount of variance common to both variables in terms of percentage-Kelley's coefficient of determination. Not only does the correlation provide the basis for acceptance or rejection of the theoretical hypothesis, but it also yields a regression weight which may be used for prediction of the dependent variable (usually the Y variable), when the X one is known.

<u>Correlations involving artificial dichotomies</u>. Where correlations involve artificial dichotomies for one or both variables in the pair specialized techniques are available to compute an approximation of the Pearson r. If only one of the variables is dichotomized, the r_b (biserial r) is used; while if both variables are dichotomized, the r_{tet} (tetrachoric r) is used. When the r_b or the r_{tet} are used, the obtained correlation matrix from several variables may not be used as the basis for further correlational analyses,i.e. multiple regression analysis, or factor analysis, etc.

<u>Correlations involving real or discrete dichotomies</u>. Where one variable in the pair is a real dichotomy, the r_{pb} (point biserial r) is used. Where the both variables are discrete dichotomies the r_{phi} (phi coefficient) is used. Where fewer than 30 cases are involved in the computation, the data is usually ranked for both variables, and rho is used. In all three of these cases, the obtained r is equivalent to a Pearson r, and is treated and interpreted the same way. A correlation matrix for either of these methods may be used as the basis for further correlational analysis,i.e., multiple correlation, factor analysis, etc.

<u>Partial correlations</u>. This includes both first and second order partial correlations, and in much the same manner as with analysis of covariance, we partial out the influence of one variable while the relationship between two other variables is assessed. In first order partial correlation the influence of just one variable is partialled out; while in second order, we partial out the influence of two variables.

<u>RST VI - Nonlinear Correlation</u>

Man tends to demonstrate a progressive increase in physical prowess during ealy childhood and youth development, and just as clearly a decline in such attributes with old age. Eta or curvelinear correlation may be used in assessing such correlation coefficients. If there is some question as to whether the paired variable distributes on a curved basis, eta may be used;

since if the distribution is linear, eta will yield precisely the same coefficient as a Pearson r. If, on the other hand, the distribution is curved, eta will yield the larger coefficient.Thus, it may be readily seen that eta is more precise and more powerful than the Pearson r, and that is why it is assigned a higher position on the gradient scale. Eta exists only on the positive end of the scale; not unlike the multiple correlations.

RST VII - Fisher'a Discriminant Function

With this RST a seriea of two or more variables uaed in concert form a correlation matrix which serves as the basis for discerning between two independently established populations. All variables may be included in the matrix except where experimental dependency is evident,i.e., as with part and total scores. An r_{pb} with each of the separate variables used in the concert serves as the fulcrum on which to diacern the popula ons, and yields multiple regression weights depicting the degree to which each variable in the matrix contributes to the discerning function. A multiple R_c which is corrected for shrinkage is obtained depicting the degree to which discerning ia accomplished based on the concert of variables. This RST represents the simpleat of the multiple correlational analyses.

RST VIII - Multiple Linear Correlation

With this RST the first task is the identification of a dependent variable, usually depicting the expected outcome, and then proceed to assess the degree of relationship between that single dependent variable and the two or more independent variables used in concert. RST VIII is more powerful and precise than RST VII because the dependent variable is always a serial score rather than a dichotomy, and a serial scale has greater precision than a dichotomy. Stepwise multiple correlation is also included in RST VIII, as it has the same statistical functional basis, but provides an observation in terms of the various steps involved, and depicting the major variable which contributes to the respective step.

RST IX - Factor Analysis and Inverse Factor Analysis

This is a valuable and meaningful technique for use in grouping variables into related cluster groupings, which are called factors, and of determining the nature and identification of such factors. Factor analysis is used for the grouping of variables in R-methodology into common clusters; while inverse factor analysis is used with Q-methodology for grouping of persons into similarly related groupings. While there are a number of different approaches to factor analysis, all of them deal with the same two basic questions that are typically asked by the task scientist: (1) How many independently organized factors are present in a given set of variables? (2) Are there a certain number of independently organized factors present in a given set of variables? After the factors are extracted, they may be rotated to simple structure in a number of different ways for purposes of identification. Sometimes the scientist will include variables of known identity for purposes of identifying certain factor loadings.

RST X - Multiple Discriminant Analysis

This RST deals with an extension of the single-classification of A of V, but which is designed to include two or more dependent variables; rather than the single one previously described under traditional A of V techniques, and as contained in RST IV. The problem is to determine the extent and manner in which two or more previously defined groups of subjects may be differentiated by a set of dependent variables operating together. With two groups of subjects this separation can be represented only along a single dimension, but with more than two groups the differentiation may be described in terms of multiple independent dimensions. The maximum number of these factors necessary to represent group differences will be the number of groups minus one, or the number of variables, whichever is smaller.

RST XI - Canonical Correlation Analysis

The goal of this RST is to define the primary independent dimensions which relate one set of variables to another set of variables. The technique is primarily descriptive, although the method used involves finding sets of weights which will yield two composite variables (one for each set of original variables) which will correlate maximally. The output of a canonical correlation analysis should suggest answers to questions concerning the number of ways in which the two sets of measures are related, the strengths of the relationships, and the nature of the relationship so defined. One of the two sets always deals with independent variables, just as with the multiple correlation in RST VIII, but the other set deals with multiple dependent variables. Each canonical "root" will be associated with two vectors of weights, which may be applied to the original scores on their respective sides to yield the composite canonical scores.

RST XII - Other Advanced Statistical Techniques

This RST is intended as a catch-all for other advanced statistical techniques. It would include such techniques as "Heuristic" problem solving, "Decision" making techniques, Multiple curvelinear regression techniques, and a host of other ones in various states of present development.

Use of Computer

More and more the computer is becoming the principal vehicle for the use of the RSTs, and it is essential that present graduate students be properly introduced to its use. Generally, the computer orientation should include the following:

.. Use of key punch.

.. Introduction to FORTRAN programming.

.. Comparison of interaction and batch modes of processing.

.. Use of computer RST libraries.

.. Interpretation and meaning of printouts for RSTs.

REFERENCES

1. Boneau, C.A. The effects of violations of assumptions underlying the t test. Psychological Bulletin, 1960, 57, 49-64.

2. Cassel, R.N. Basic research statistical tools (RSTs) distribute on a precision or power gradient scale. Psychology, 1970, 7(3),

3. Downie, N.M., and Heath, R.W. Basic Statistical Methods. Third Edition. New York: Harper and Row, 1970.

4. Edwards, A.L. Statistical Methods. Second Edition. New York: Holt, Rinehart and Winston, 1967.

5. Edwards, A.L. Experimental Design in Psychological Research. New York: Holt, Rinehart and Winston, 1962.

6. Filstead, W.J. Qualitative Methodology. Chicago: Markham Publishing Company, 1970.

7. Guilford, J.P. Fundamental Statistics in Psychology and Education. New York: McGraw-Hill Book Company, 1965.

8. Horst, P. Psychological Measurement and Prediction. Belmont, California: Wadsworth Publishing Company, 1966.

9. Hsu, T.C., and Feldt, L.S. The effect of limitations on the number of criterion score values on the significance level of the F-test. American Educational Research Journal, 1969,VI(4),515-527.

10. Veldman, D.J. Fortran Programming for the Behavioral Sciences. New York: Holt, Rinehart and Winston, 1967.

11. Lehman, R.S., and Bailey, D.E. Digital Computing. New York: John Wiley and Sons Inc.,1968.

12. Norton, D.W. An empirical investigation of the effect of non-normality and heterogenity upon the F-test of analysis of variance. Unpublished Dissertation for Ph.D. University of Iowa, 1952.

13. Nunnally, J.C. Psychometric Theory.

New York: McGraw-Hill Book Company, 1967.

14. Spogren, D.D. Measurement techniques in evaluation. Review of Educational Research, 1970, 40(2), 301-320.

15. Stanley, J.C. Measurement in Today's Schools. Fourth Edition. Englewood, New Jersey: Prentice-Hall, Inc.,1964.