DOCUMENT RESUME

ED 436 571 TM 030 433

AUTHOR	Peet, Martha W.
TITLE	The Importance of Variance in Statistical Analysis: Don't Throw Out the Baby with the Bathwater.
PUB DATE	1999-11-00
NOTE	17p.; Paper presented at the Annual Meeting of the Mid-South Educational Research Association (Point Clear, AL, November 17-19, 1999).
PUB TYPE	Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Ability; Analysis of Covariance; *Analysis of Variance; Classification; *Effect Size; Middle School Students; Middle Schools
IDENTIFIERS	*Variance (Statistical)

ABSTRACT

This paper analyzes what happens to the effect size of a given dataset when the variance is removed by categorization for the purpose of applying "OVA" methods (analysis of variance, analysis of covariance). The dataset is from a classic study by Holzinger and Swinefors (1939) in which more than 20 ability test were administered to 301 middle school students to discern which abilities determined overall academic performance. One-way analyses of variance (ANOVA) were run with the dichotomized variable as the independent factor to compare the reduction in effect size in the ANOVAs with the effect size of regression run with the intact continuous independent variable. The effect sizes of the ANOVAs that were run with the dichotomized variable were about half of the regression effect sizes run with the unaltered independent variable. Each of the effect sizes after dichotomization would fit into a lower category, resulting in a serious underestimation of the substantive effect. (Contains 3 tables and 16 references.) (SLD)

The Importance of Variance in Statistical Analysis: Don't Throw Out the Baby with the Bathwater

Martha W. Peet
University of North Texas

Paper presented at the annual meeting of the Mid-South Educational Research Association, Point Clear, Alabama, November 16-19, 1999.

The desire to understand variability, both between groups and individuals, motivates much of educational research. Good design standards require careful thought about the theoretical basis in the field, the selection of the appropriate research method and a well thought out plan for the analysis of the data. One result of good research design is clearly interpretable results (Prosser, 1990) . The measurement of variability plays a central role in both research design and statistical analysis. "In research, control is the control of variance. ...[The experimental method attempts] to increase the variance between groups, minimize the error variance and control the extraneous variance" (Pedhazur,1982, p.97). Hence, without the ability to account for sufficient variance, a psychometric instrument cannot be considered to yield reliable information (Thompson, 1986).

However, not all researchers recognize the importance of variance in their analysis of data, selecting data analytic methods that fail to honor the variation in their variables of interest. As one thoughtful observer noted, "Despite the recognized importance of variance, some researchers in the behavioral sciences choose techniques which discard it by categorizing variables" (Prosser, 1990, p 4). In a recent review of research methods reported in American Educational Research Journal, Educational Researcher, and Review of Educational Research, Elmore and Woehlke found that the most commonly used methods were ANOVA and ANCOVA (Elmore & Woehlke, 1996). ANOVA and the various other types of "OVA" methods (ANCOVA, MANOVA, and MANCOVA) are used by educational researchers even though these methods sometimes restrict the quality of data that can be used. The wide use of OVA methods might be due to the fact that some researchers "unconsciously and erroneously associate ANOVA with the power of experimental designs" (Thompson, 1992, p 1). This may be due in part to the fact that OVA methods were first used in agronomic experimentation, where the treatments were artificially manipulated for the experimental design. However, OVA methods were never intended to be used in non-experimental research, even though this application of OVA methods is a common practice among education researchers (Lopez, 1989). Unfortunately, analyzing data in an ANOVA format tends to create the false impression that a non-experimental design has thereby been transformed into an experimental design, or at the very least, into something closely approximating it (Pedhazur & Schmelkin, 1991). Thompson (1992) noted:

> Researchers often value the ability of experiments to provide information about causality; they know that ANOVA can be useful when independent variables are nominally scaled and dependent variables are intervally scaled; they then begin to unconsciously identify the analysis of ANOVA with design of an experiment. (p.1)

Data can be analyzed in a number of related but distinct ways. All parametric statistical methods are correlational. Analysis of variance (ANOVA) is a special case of regression analysis, and both ANOVA and regression are techniques derived from the general linear model(Cohen, 1968). Regardless of the specific correlational technique employed, correlations are maximized when patterns of systematic variance across variables are maximized. Kerlinger (1986) stresses that variance is "of the highest importance in research and in the analysis of research data"(p. 84). This paper looks at the reasons that regression is a better, more powerful tool for statistical analysis in many cases.

ANOVA and related methods

OVA methods are used to test whether there is a statistically significant difference between the means of two or more experimental groups. In other words, they are used to determine whether one or more factors have a significant effect on the dependent variable being measured. OVA methods are suited for experimental methods in which qualitative treatments are manipulated in an appropriately orthogonal relationship. Ideally, ANOVA requires the dependent variable to be measured in the form of equally spaced intervals and equal sized samples per treatment if computational simplicity is to be maintained.

Advantages of OVA methods

OVA methods have several perceived advantages (Prosser, 1990). When ANOVA was first introduced in 1925 by Sir Ronald Fisher in his seminal work, Statistical Methods for Research Workers, it was of prime importance that the new method was much quicker to compute then previously used methods. It saved time to summarize data as groups. The residual effect was that many researchers exhibited a casual readiness to discard variance in continuous predictor variables so as to create grouped data. Unfortunately these data simplifications are neither appropriate nor justifiable(Cohen, 1983). Moreover, with present computer power and the prevalence of statistical software packages the importance of calculation time is minimized . Quickness of calculations is no longer an advantage. Nelson and Zaichkowsky (1979) postulate that

"ANOVA has been applied so frequently because of historical momentum: it was, perhaps, the first multivariate solution made available to researchers" (p. 328).

Another advantage of OVA methods is ease of interpretation. The independent variables are uncorrelated in the omnibus test when a balanced design is used. Hence, OVA methods give a clear cut interpretation of results. The third perceived advantage is the ability to test interaction effects, however other techniques such as regression, can also be used for this purpose without possibility of certain problems associated with OVA methods.

Problems with OVA methods

There are two specialized conditions that make OVA methods problematic. The first requirement and main problem with OVA methods is that the independent variable(s) must be nominally scaled (Prosser, 1990). A researcher might treat a continuous independent variable as comprised of distinct categories and analyze the data using an ANOVA. For example, if a researcher wanted to use IQ as one of the independent variables in an ANOVA analysis, the IQ test results would have to be divided into a number of discrete categories, such as high, middle, low. Interval data loses a substantial amount of variance when it is categorized into lower scales. Cohen demonstrated how artificially dichotomizing a continuous variable reduces the power of statistical tests (Cohen, 1983). He calculated the cost of dividing a continuous variable in half. For a bivariate normal population divided at the mean, Pearson's r would decrease to .798 r, accounting for only ($.798^2 =$) .637 as much variance in the other variable. Cohen stated further that by dichotomizing both independent variables to create a 2X2 ANOVA, further power is lost. When r is between .2 and .5, double dichotomization at the mean is equivalent to discarding 60 % of the subjects at both the two-tailed .01 and .05 levels. He sums up by stating:

> It is no exaggeration to say that double dichotomization may result in the loss of as much as two-thirds of the proportion of variance that could be accounted for on the original variables, with a resulting loss of power equivalent to throwing away as much as two-thirds of the sample. (Cohen, 1983, p. 252)

Another way that Cohen (1983) demonstrated the danger of dichotomization is by determining the increase in sample size needed to offset dichotomizing:

> If a population r = .30 is assumed, the sample size needed for power to equal .80 for a two-tailed .05 test is 84. For "optimal" dichotomization at the mean, the resulting r of .239 [.798r=.798(.30)=.239] requires 133 cases under the same conditions, an increase in the necessary sample size of 58%. (p. 251)

Kerlinger (1986) warned against demoting intervally scaled data to nominal scale:

> Partitioning a continuous variable into a dichotomy or trichotomy throws information away…To reduce a set of values with a relatively wide range to a dichotomy is to reduce its variance and thus its possible correlation with other variables. A good rule of research data analysis, therefore, is: Do not reduce continuous variables to partitioned variables (dichotomies, trichotomies, etc.) unless compelled to do so by circumstances or the nature of the data (seriously skewed, bimodal, etc.). (p. 558)

As Pedhazur (1983) noted, "categorization leads to a loss of information, and consequently to a less sensitive analysis" (p. 453). Categorization is especially harmful to the analysis when there is a great deal of variance in the original data. In this case, all the subjects within a category are treated alike even though they may have been originally quite different in the continuous variable. Pedhazur further notes that "it is this loss of information about the differences between subjects, or the reduction in the variability of the continuous variable, that leads to a reduction in the sensitivity of the analysis, not to mention the meaningfulness of the results" (p. 454). Hence, discarding variance is not generally regarded as a good research practice (Thompson, 1988). As Kerlinger (1986) pointed out, "variance is the 'stuff' on which all analysis is based." (p. 558).

Another problem with dividing the independent variable into categories is the problem of the dividing points. Each group might have been highly variable before it was condensed into one value. Data that are quite close to the dividing point are associated with points that are quite different instead of the data on the other side of the dividing point which is more similar. Cliff (1987) summarized this :

> Such division is not infallible. Think of the persons near the borders. Some who should be highs are actually classified as lows, and vice versa. In addition, the "barely highs" are classified the same as the "very highs," even though they are different. Therefore, reducing a reliable variable to a dichotomy makes the variable more unreliable, not less. (p. 30)

Most researchers when using OVA methods use a balanced cell design. A balanced cell design has an equal number of subjects in each experimental condition. By maintaining a balanced cell design, the independent variables appear to be uncorrelated. However, maintaining a balanced design throughout an experiment is rarely achieved without problems. As the experiment proceeds, subjects are lost by attrition. To keep a balanced design, the "extra" subjects in the remaining cells are often discarded with the variance that they exhibit. Throwing out subjects to achieve balance reduces the power of the analysis: "Many [researchers] are under the erroneous impression that OVA methods offer more power and, therefore, more protection against Type II error" (Prosser, 1990, p. 9). On the contrary, as variance is discarded, reliability decreases.

OVA methods should only be used in "experimental manipulations along one or more dimensions (main effects), resulting in subgroups of observations in multifactor cells, treatment conditions"(Cohen, 1968, p. 440). There is no advantage in analyzing the data obtained with these rigid conditions (naturally occurring categorical data and balanced cell design) by another method such as regression. OVA methods would under these restrictions be the tool of choice for the analysis. OVA can be seen as a "shortcut to an analysis by the linear model which analyzes by batches and capitalizes on the fact that batches are orthogonal" (Cohen, 1968, p. 440). Unfortunately, OVA methods are widely used in the analysis of data from both experimental and non-experimental research. Prosser (1990) summarizes the case against overuse of OVA methods:

> There are ... seriously disturbing aspects about the use of OVA methods in nonexperimental research. Their misuse can compromise the integrity of an entire study. This is true mainly because of the specialized conditions required by the OVA's. (p. 5)

Linear Regression

Linear regression is another popular statistical technique used by educational researchers. Regression is a broader test than ANOVA and its various analogs because the independent variable(s) can be either categorical or continuous. Both regression and ANOVA try to ascertain if some of the variance of the dependent variable(s) can be related to variance in the independent variable(s). Regression provides more power while doing everything that OVA methods can do. Regression does not require the specialized conditions that are required for OVA analysis. It does not require preciously gathered continuous data to be categorized, throwing away variance. In regression, nominal, ordinal, or interval data are treated alike. Regression can also be used when cell frequencies are unequal and disproportionate. It can be used to study trends in the data. Regression has been shown repeatedly to be superior to OVA methods (Daniel, 1989; Thompson, 1986). Kerlinger and Pedhazur (1972) described the advantages of regression:

> It can be used equally well in experimental or non-experimental research. It can handle continuous and categorical variables. It can handle two, three, four or more independent variables...multiple regression analysis can do anything the analysis of variance does--sum of squares, mean squares, F ratios--and more. (p. 3)

Regression allows variables "to retain their highest level of scale and requires that the researcher thoughtfully examine data from several different perspectives" (Lopez, 1989, p. 12).

Method

This paper analyzes what happens to the effect size of a given dataset when the variance is removed by categorization for the purpose of applying OVA methods. The dataset is from a classic study by Holzinger and Swineford (Holzinger & Swineford, 1939), who administered over 20 ability tests to a large group (301) of middle school students to determine which abilities determined overall academic performance. Three pairs of tests with different levels of intercorrelation were chosen for demonstration purposes in this study. One pair was comprised of two tests of verbal abilities which were highly correlated ($r^2 = .538$). The tests were the Paragraph Comprehension test (Dependent variable, DV) and the Sentence Completion test (Predictor, P). The second two tests were moderately correlated ($r^2 = .432$). The tests were the Paragraph Comprehension test (P) and the General Information Verbal test (DV). The third pair had only a small correlation ($r^2 = .157$). The tests were Memory of target words (P) vs. Memory of target numbers (DV).

Results

The first thing that was verified was Cohen's cost of dichotomization. Each independent variable was divided into two groups at the mean. The Pearson r was found for the correlation between the continuous variable and the dichotomized version of itself. Cohen calculated that the correlation of a variable with a dichotomized version of itself was $r_d = .798r$, which is verified with this dataset as shown in Table 1. Each of the correlations shows about a 20 % reduction in value.

INSERT TABLE 1 ABOUT HERE

One-way ANOVAs were run with the dichotomized variable as the independent factor to compare the reduction in effect size in the ANOVAs with the effect size of regression run with the intact continuous independent variable. The effect sizes of the ANOVAs that were run with the dichotomized variable were about half of the regression effect sizes run with the unaltered independent variable as shown in Table 2. The magnitude of the original effect size does not influence the percentage of the effect size that was still accounted for after stripping out the variance in the independent variable. Each of the effect sizes after dichotomization would fit into a lower category, resulting in a serious underestimation of the substantive effect. In each of these ANOVA analyses, the null hypothesis was rejected at $p>.001$.

INSERT TABLE 2 ABOUT HERE

Next, the effect of dividing the independent variable into different numbers of groups was investigated. As Table 3 shows, as the number of groups increases more of the variance in the dependent variable is accounted for. With four groups, almost 90% of the variance that is accounted for by the continuous variable is accounted for with the categorical variable.

INSERT TABLE 3 ABOUT HERE

Discussion

"It is important for educational researchers to make informed choices in respect to methods of analysis of their data" (Lopez, 1989, p. 11). Each of the foregoing examples illustrates why variance should not be discarded in the interest of running an ANOVA. In each case, the effect size is seriously underestimated simply by categorizing the independent variable. The reason to run any statistical test is to test whether the result obtained is "real." The researcher should maintain all the variance in a dataset to increase the power to make a correct decision whether to reject the null hypothesis and/or to support the alternative hypothesis.

Beyond statistical problems with discarding variance, Siebold and McPhee (1979) explained substantively why variance should not be discarded:

> Advancement of theory and the useful application of research findings depend not only on establishing that a relationship exists among predictors and the criterion, but also upon determining the extent to which those independent variables, singly and in all possible combinations, share variance with the dependent variable. Only then can we fully know the relative importance of independent variables with regard to the dependent variable in question. (p. 355)

Likewise, Prosser (1990) noted, "Keeping all variance is vital to a sensitive, effective analysis of data (p. 15).

Table 1

Correlation between Variable and Dichotomized Variable

Variable	Correlation of variable with dichotomized variable
Sentence Completion	.820
Paragraph Comprehension	.766
Memory of Target Words	.774

Table 2

ANOVA Effect Size (dichotomized independent variable) vs.Regression Effect Size

Variables	Regression effect size	ANOVA effect size	Percentage of ANOVA effect size in regard to regression effect size
IV = Memory of Target words DV = Memory of Target numbers	.157	.08	51%
IV = Paragraph Comprehension Test DV = General Information verbal test	.432	.216	50%
IV = Sentence completion test DV = Paragraph completion test	.538	.3	56%

Table 3

Comparison of ANOVA run with the Independent Variable divided into 2, 3 and 4 levels

Variables	Regression effect size	ANOVA effect size DV = 2 groups	ANOVA effect size DV = 3 groups	ANOVA effect size DV = 4 groups
IV = Memory of Target words DV = Memory of Target numbers	.157	.08 (51%)	.109 (69%)	.138 (88%)
IV = Paragraph Comprehension Test DV = General Information verbal test	.432	.216 (50%)	.352 (81%)	.37 (85%)
IV = Sentence completion test DV = Paragraph completion test	.538	.3 (56%)	.41 (76%)	.47 (87%)

Bibliography

Cohen, J. (1968). Multiple regression as a general data-analytic system. Psychological Bulletin, 70, 426-443.

Cohen, J. (1983). The cost of dichotomization. Applied Psychological Measurement, 7, 249-253.

Daniel, L. (1989). Commonality analysis with various parametric methods. Paper presented at the annual meeting of the Southwest Educational Research Association, Houston, TX. (ERIC Document Reproduction Service No. ED 314 483)

Elmore, P.B., & Woehlke, P.L., (1996, April). Research methods employed in American Educational Research Journal, Educational Researcher, and Review of Educational Research from 1978 to 1995. Paper presented at the annual meeting of the American Educational Research Association, New York, NY. [ERIC Document Reproduction Service No. ED 397 122]

Fisher, R. A. (1925). Statistical methods for research workers. Edinburgh, England: Oliver and Boyd.

Holzinger, K.J., & Swineford, F. (1939). A study in factor analysis: The stability of a bi-factor solution (no. 48). Chicago, IL: University of Chicago. (p. 81-91)

Kerlinger, F. (1986). Foundations of behavioral research (3rd ed.). New York: Holt, Rinehart and Winston.

Kerlinger, F.N., &Pedhazur, E. (1973). Multiple regression in behavioral research New York: Holt, Rinehart and Winston.

Lopez, K. A. (1989, November). Testing interaction effects without discarding variance. Paper presented at the annual meeting of the Mid-South Educational Research Association, Little Rock, AR. [ERIC Document Reproduction Service No. ED 322 167]

Nelson, L.R., & Zaichkowsky, L.D. (1979). A case for using multiple regression instead of ANOVA in educational research. Journal of Experimental Education, 47, 324-330.

Pedhazur, E. (1982). Multiple regression in behavioral research: Explanation and prediction New York: Holt, Rinehart and Winston.

Pedhazur, E., & Pedhazur-Schmelkin, L. (1991). Measurement, design, and analysis: An integrated approach Hillsdale, NJ.: Lawrence Erlbaum Associates Publishers.

Prosser, B. (1990, January). Beware the dangers of discarding variance. Paper presented at the annual meeting of the Southwest Educational Research Association, Austin, TX. [ERIC Document Reproduction Service No. ED 314 496]

Thompson, B. (1986). ANOVA versus regression analysis of ATI designs: An empirical investigation. Educational and Psychological Measurement, 46, 917-928. [ERIC Document Reproduction Service No. EJ 348 419]

Thompson, B. (1988). Discard variance: A cardinal sin in research. Measurement and Evaluation in Counseling and Development, 21, 3-4.

Thompson, B. (1992). Interpreting regression results: Beta weights and structure coefficients are both important. Paper presented as the annual meeting of the American Educational Research Association, San Francisco, CA. [ERIC Document Reproduction Service No. ED 344 897]

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: The Importance of Variance in Statistical Analysis: Don't Throw out the Baby with the Bathwater

Author(s): Martha W. Peet

Corporate Source: U. of North Texas

Publication Date:

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 1	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2A	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2B
Level 1	Level 2A	Level 2B
☑	☐	☐
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, → please

Signature: Martha W Peet	Printed Name/Position/Title: Martha W. Peet	
Organization/Address: 2123 Carlton Carrollton, TX 75007	Telephone: 972-492-2380	FAX:
	E-Mail Address: mwp0003@unt.edu	Date: 11-17-99

(over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

University of Maryland
ERIC Clearinghouse on Assessment and Evaluation
1129 Shriver Laboratory
College Park, MD 20742
Attn: Acquisitions

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
1100 West Street, 2nd Floor
Laurel, Maryland 20707-3598

Telephone: 301-497-4080
Toll Free: 800-799-3742
FAX: 301-953-0263
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com