DOCUMENT RESUME

ED 124 571 TM 005 324

AUTHOR Williams, John D.
TITLE Should a First Course in ANOVA Be Taught Through MLR?
PUB DATE [Apr 76]
NOTE 5p.; Paper presented at the Annual Meeting of the American Educational Research Association (60th, San Francisco, California, April 19-23, 1976)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Analysis of Covariance; *Analysis of Variance; Computer Assisted Instruction; Costs; Courses; *Higher Education; Instructional Design; *Mathematical Models; *Multiple Regression Analysis; Statistics; *Teaching Methods
IDENTIFIERS MLR; *Multiple Linear Regression

ABSTRACT

Before implementing a course in the analysis of variance (ANOVA) taught through multiple linear regression, several concerns must be addressed. Adequate computer facilities that are available to students on a low-cost or cost-free basis are necessary; also students must be able to meaningfully communicate with their major advisor regarding their statistical knowledge. The second concern implies some usage of traditional ANOVA terminology. Using a standard ANOVA topic such as the analysis of covariance is generally convincing in the efficacy of the multiple regression approach. The analysis of covariance can be accomplished through two rather simple linear models. (Author)

Should A First Course in ANOVA Be Taught Through MLR?

John D. Williams
The University of North Dakota

While many adherents to the MLR approach might be oriented toward answering the question given in the title to this paper with a "yes", several points of view should be examined. Given that it makes pedagogic sense (which traditionalists may dispute) to use a linear models approach, practical considerations may temper the decision to implement MLR in the teaching of ANOVA.

First, a most important consideration is an adequate computer together with the necessary software. As most colleges and universities will have the computer and necessary software, the concern changes to accessibility to students. Are key-punch machines readily available? Are other computer services cost-free to the student? Is the software easily accessible? If the answers to any of these questions is no, then implementing a MLR approach can be extremely frustrating.

Second, it must be remembered that most students in an applied course are going to eventually work on a research project of some sort and will, of course, have to be able to communicate with their major advisor. In general, most of our non-statistician colleagues are familiar with traditional ANOVA designs and terminology. Thus, to help the student and his advisor to communicate, some degree of dependence on traditional ANOVA terminology is necessary.

Implementing a Regression Approach in a First Course in ANOVA

If the concerns just given are reasonably satisfied (computer availability, and the expectations of the major departments) then a definite _yes_ can be given as an answer to the question posed in the title to this paper. Still to be resolved, though, is the question, "Should the course, if taught by MLR, use traditional ANOVA terminology?" Also, "Should such a course be oriented toward a direct translation of ANOVA type questions to MLR solutions?"

Remembering that students will eventually have to communicate with their major advisors and also will be reading journals using a traditional ANOVA format, the most judicious choice seems to me to use ANOVA terminology, but also point out the conveniences provided by the MLR approach. Perhaps this process can be illustrated through the use of the analysis of covariance (at least as I do this on a personal basis).

First, before any discussion is made in class of the linear models involved in a solution, the students are required to read either Lindquist's (1953), Edward's (1968), or Winer's (1971) presentation on the analysis of covariance. They are specifically instructed to follow the numerical example to completion and then reflect upon their cognitive understanding of the process. Then, in the ensuing class period, the students consider the same problems, but through multiple regression. For example*, if there are three groups and one covariate, then a full model can be defined as

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + e_1, \qquad (1)$$

where

Y = the score,

X_1 = the covariate,

X_2 = 1 if a member of Group 1, 0 otherwise,

X_3 = 1 if a member of Group 2, 0 otherwise,

$b_0 - b_3$ are regression coefficients and e_1 is the error in prediction with this model.

*under the assumption of homogeniety of regression

Then a restricted model can be formed as

$$Y = b_0' + b_1X_1 + e_2. \qquad (2)$$

Equation 1, the full model, contained all the information (both covariate and group membership variables) whereas equation 2 included only the covariate. The "difference" between these models would then tell us what group differences there are that are independent of the covariate and are tested by

$$F = \frac{(R^2_{FM}-R^2_{RM})/(k-1)}{(1-R^2_{FM})/(N-k-1)}, \qquad (3)$$

where

R^2_{FM} = the multiple correlation squared value from the full model;

R^2_{RM} = the multiple correlation squared value from the restricted model;

k = the number of groups, and

N = the number of subjects.

Several students often then bring up the question, "If it's this easy, then why did you make us read these other books?" Some ask, "Why would anyone ever use the methods described in those other books, if we can get the same answers through two simple linear models?"

While I have no answer to the last posed question, I would note that those students who ask either of the last two questions have become "converts" to the use of multiple linear regression for solving ANOVA problems.

As to whether the course should be oriented toward a direct translation of ANOVA designs to MLR, the overriding concern (at least to me) regards the original data. If the data is associated with natural mutually exclusive groups (male-female, Protestant-Catholic-Jew-other) then a direct ANOVA type solution is appropriate; if the ANOVA design is arrived at through "cutting up" the subjects into groups such as high-middle-low on a continuous variable, it makes more sense to me to use the original continuous variable.

References

Edwards, A.L. Experimental design in psychological research. New York: Holt, Rinehart and Winston, 1968.

Lindquist, E.F. Design and analysis of experiments in psychology and education. Boston: Houghton-Mifflin, 1953.

Winer, B.J. Statistical principles in experimental design. New York: McGraw-Hill, 1971.