Behavior Research Methods, Instruments, & Computers
1985, 17 (3), 411-412

Data transformation maximizing homoscedasticity and within-group normality

JEFFREY LEE RASMUSSEN
Indiana University–Purdue University at Indianapolis
Indianapolis, Indiana

Two assumptions underlying the analysis of variance (ANOVA) technique are that the data be sampled from a normally distributed population and that the population within-group variances be equal. Applying ANOVA to data sets that display a clear violation of these assumptions may result in a substantial deviation of the actual Type I and Type II error rates from their nominal levels. This deviation may result in an overly liberal or overly conservative test (Bradley, 1978; Glass, Peckham, & Sanders, 1972; Levine & Dunlap, 1982, 1983).

Data transformation may be used in order to better meet the assumptions of ANOVA. Furthermore, an appropriate transformation may result in a substantial increase in the power of ANOVA (Levine & Dunlap, 1982, 1983). Theoretically, it is often beneficial to transform data to meet the assumptions of ANOVA. In practice, however, as Games (1983) pointed out, it is often difficult to pick an appropriate transformation. Some guidelines have been offered by Bartlett (1947), Mueller (1948), and Tabachnick and Fidell (1983); nevertheless, finding a workable transformation is often tedious.

Dunlap and Rasmussen (1982) and Rasmussen and Dunlap (1983) have presented FORTRAN programs that select a transformation from the Box and Cox (1964) family of power transformations to meet assumptions of regression statistics. Dunlap and Duffy (1974) presented a FORTRAN program that chooses a power transformation that minimizes within-group skew.

The present program serves as a substitute for the Dunlap and Duffy (1974) program, which picks a transformation that minimizes only skew; the present program minimizes nonnormality and heterogeneity of variance. Use of the Dunlap and Duffy program may result in extreme heterogeneity of variance or extreme kurtosis values, both of which may affect the Type I and Type II error rates. The transformation described here serves to minimize nonnormality and heterogeneity of variance.

Program Execution. The program interrogates the user with respect to the parameters of the study, such as the number of subjects, the number of groups, and so forth. It then prints out, group by group, a histogram of the untransformed data. The program then applies a series of power transformations to the data and calculates normality and homoscedasticity statistics for each transformation. Normality is calculated using a statistic by Rasmussen and Dunlap (1983); homoscedasticity is measured by Levene's (1960) F test for homogeneity of variance. The program chooses the transformation that maximizes homoscedasticity and the transformation that maximizes normality. Since these transformations are usually not identical, the program solves for the transformation that serves to maximize both by minimizing the geometric mean of the standardized indices of normality and homoscedasticity. The program prints out the three transformation values and histograms of the transformed data.

When the three values are similar, users should employ the transformation that simultaneously maximizes normality and homoscedasticity. When the values are quite dissimilar, users should take greater care in selecting the appropriate transformation. For example, a conservative choice might be to correct for heteroscedasticity and accept the nonnormality: Nonnormality usually results in a conservative test, whereas heteroscedasticity may lead to an overly liberal test.

Requirements. The program is written in single-precision FORTRAN IV and runs on a Digital Equipment Corporation DEC-2060 computer. Since the program is interactive, most of the input and output is to the computer terminal. A limited number of nonstandard FORTRAN statements make the program run conveniently on the DEC system. Users with other systems will probably need to modify assignments of logical units, READ statements with an "*" in the usual place of a format statement number (which causes free format on the DEC), and input format statements with the final symbol "$" (which causes the cursor to remain at the end of a line on the DEC system). The source program requires about 12 KB of memory. Run time depends upon the number of cases processed; for example, running the program on the DEC-2060 requires approximately 2 CPU sec for 10 cases, 4 CPU sec for 50 cases, and 16 CPU sec for 200 cases.

Program Availability. A listing of the program is available at no cost from the author.

The author's mailing address is: Department of Psychology, Indiana University—Purdue University at Indianapolis, P.O. Box 647, Indianapolis, IN 46223.

REFERENCES

Bartlett, M. S. (1947). The use of transformations. *Biometrics Bulletin*, **3**, 39-52.

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society* (Series B), **26**, 211-243.

Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology*, **31**, 144-152.

Dunlap, W. P., & Duffy, J. A. (1974). A computer program for determining optimal data transformation minimizing skew. *Behavior Research Methods & Instrumentation*, **6**, 46-48.

Dunlap, W. P., & Rasmussen, J. L. (1982). A program to determine data transformations maximizing linear regression. *Behavior Research Methods & Instrumentation*, **14**, 357-358.

Games, P. A. (1983). Curvilinear transformation of the dependent variable. *Psychological Bulletin*, **93**, 382-387.

Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of failure to meet the assumptions underlying the fixed effects analysis of variance and covariance. *Review of Educational Research*, **42**, 237-288.

LEVENE, H. (1960). Robust tests for equality of variances. In I. Olkins (Ed.), *Contributions to probability and statistics*. Stanford, CA: Stanford University Press.

LEVINE, D. W., & DUNLAP, W. P. (1982). Power of the F test with skewed data: Should one transform or not? *Psychological Bulletin*, **92**, 272-280.

LEVINE, D. W., & DUNLAP, W. P. (1983). Data transformation, power, and skew: A rejoinder to Games. *Psychological Bulletin*, **93**, 596-599.

MUELLER, C. G. (1948). Numerical transformations in the analysis of experimental data. *Psychological Bulletin*, **46**, 198-223.

RASMUSSEN, J. L., & DUNLAP, W. P. (1983). A program for simultaneously transforming two variables to maximize linear regression, normality and homoscedasticity. *Behavior Research Methods & Instrumentation*, **15**, 478-480.

TABACHNICK, B. G., & FIDELL, L. S. (1983). *Using multivariate statistics*. New York: Harper & Row.

(Revision accepted for publication February 7, 1985.)