Behavior Research Methods & Instrumentation
1975, Vol. 7 (5), 484

Nonparametric one-way ANOVA with multiple covariates

EDWARD F. GOCKA and BERNARD HANES
VA Hospital, Sepulveda, California and Dept. of Health Sciences
California State University, Northridge, California 91343

Description. Unlike the parametric analysis of variance, the parametric analysis of covariance F test is appreciably affected by non-normality, even in balanced classifications, and overall the degree of sensitivity is determined by the distribution of the concomitant variables and the structure of the design matrix (Atiqullah, 1964). In balanced conditions, the design matrix X possesses the property of quadratic homogeneity which is characterized by the projection operator $P = X(X'X)^{-1}X'$ (Rao, 1965; Winer, 1968) having its diagonal terms all equal; this property largely nullifies non-normality effects on the F test in a one-way ANOVA with equal numbers of observations per group and in many standard orthogonally cross-classified or balanced designs such as two-way ANOVAs, Latin squares, etc. (Atiqullah, 1962). In balanced classifications, therefore, the covariates alone determine the sensitivity to non-normal conditions while in nonbalanced classifications both the quadratic heterogeneity of P and the covariates will determine the sensitivity to non-normality. An equally serious consideration has been shown to be the violation of the assumption of linearity between the dependent variable Y and the covariates C (Atiqulla, 1964).

The method programmed here is based on a rationale given by Quade (1967) for a one-way design with multiple covariates which avoids, by rank transformations, the important parametric model assumptions of normality and linearity. It is assumed, however, that the population treatment group distributions for a covariate are the same for each Group K. The number of observations within each group is recommended to be at least 5, and preferably 10, since the behavior of the method is not fully known for small samples. A variance ratio to be compared with a critical value of F with K − 1 and N − K degrees of freedom is used to test the hypothesis of identical conditional distributions of each Y group vector on the respective matrix of covariates C.

As a first step, the Y and C variables are pooled without regard to distinctions between treatment groups and then are ranked within variable, with the smallest being ranked *one* and the largest *N*, where $N = N_1 + N_2 + \cdots + N_K$. "Average ranks" are used in the case of a tie. These ranked variables are corrected for the mean by subtracting (N + 1)/2 from each value to give the matrices y and c. If we have q covariates, then y is an N × 1 matrix of "corrected" ranks of responses and c is an N × q matrix of "corrected" ranks of the concomitant variates. The N × 1 matrix of fitted or predicted values $\hat{y}$ is obtained by least squares procedures as

$$\hat{y} = c(c'c)^{-1}c'y$$

from which we obtain the N × 1 matrix of residuals as

$$Z = y - \hat{y}$$

Taking P as the projection operator defined earlier and I as a conformable identity matrix with the design matrix X having a rank of K − 1, the variance ratio (VR), as given by Quade (but using different notation), is

$$VR = (N-K)(Z'PZ)/(K-1)(Z'[I-P]Z).$$

It should be noted that this VR is evaluated against an F with K − 1 and N − K df, while the usual parametric analysis of covariance is evaluated against an F with K − 1 and N − K − q df.

Program Characteristics. The program is written in FORTRAN. It handles a single variable Y in any one analysis but allows for up to 12 treatment groups (K) and up to 10 covariates (q). Cell size has a maximum set at 30 cases. A control card preceding the data serves to indicate the number of covariates, the number of treatment groups, and the number of cases in each group. The data deck requires all of the values for a subject to be punched on one card in variable format, but in order $Y, C_1, C_2, \cdots, C_K$. The output consists of the original input data, its rank conversion, group means for the original and ranked data, and the value of the variance ratio VR along with the appropriate degrees of freedom for this F-type statistic.

Hardware: The computer used was an IBM 1130 with 16K of memory and an external IBM 2310 disk storage device.

Availability. A listing of the program and user document may be obtained at no cost from Edward F. Gocka, PhD, Veterans Administration Hospital, 16111 Plummer Street, Sepulveda, California 91343.

REFERENCES

Atiqullah, M. The estimation of residual variance in quadratically balanced least squares problems and the robustness of the F test. *Biometrika*, 1962, **49**, 83-91.

Atiqullah, M. The robustness of the covariance analysis of a one-way classification. *Biometrika*, 1964, **51**, 365-372.

Quade, D. Rank analysis of covariance. *Journal of the American Statistical Association*, 1967, **62**, 1187-1200.

Rao, C. R. *Linear statistical inference and its applications*. New York: Wiley, 1965.

Winer, B. J. The error. *Psychometrika*, 1968, **33**, 391-403.