

1988

ONEWAY: A BASIC program for computing ANOVA from group summary statistics

Joseph S. Rossi

University of Rhode Island, jsrossi@uri.edu

Follow this and additional works at: https://digitalcommons.uri.edu/psy_facpubs

Terms of Use

All rights reserved under copyright.

Citation/Publisher Attribution

Rossi, J. S. ONEWAY: A BASIC program for computing ANOVA from group summary statistics. *Behavior Research Methods, Instruments, & Computers*, 20(3), 347-348. doi: 10.3758/BF03203856

Available at: <http://dx.doi.org/10.3758/BF03203856>

This Article is brought to you for free and open access by the Psychology at DigitalCommons@URI. It has been accepted for inclusion in Psychology Faculty Publications by an authorized administrator of DigitalCommons@URI. For more information, please contact digitalcommons@etal.uri.edu.

ONEWAY: A BASIC program for computing ANOVA from group summary statistics

JOSEPH S. ROSSI

University of Rhode Island, Kingston, Rhode Island

In the past few years, dozens of excellent analysis of variance (ANOVA) programs have become available to microcomputer users. The varieties of such programs fill a wide range of user needs, including factorial designs (with or without equal cell sizes), repeated measures, multiple dependent variables, mixed model designs, nested designs, inclusion of covariates, and assorted follow-up tests (e.g., Collani & Waloszek, 1983; Corrigan, Bonelli, & Borys, 1980a, 1980b; Coulombe, 1984, 1985; Galla, 1984; Hacker & Angiolillo-Bent, 1981; Wolach, 1983). However, all of these programs require the user to input raw data. Described here is a BASIC ANOVA program for which group summary data—means, standard deviations, and sample sizes—are used as input, so that data from individual subjects are not needed for the analysis.

Statistical Tests Based on Grouped Data. Many commonly used statistical tests, such as the two-sample t test and the chi-square test of independence, can be calculated directly from grouped data. Most introductory statistics textbooks provide the appropriate formulas (e.g., Howell, 1982, p. 104 [chi-square test], p. 134 [t test]).

Researchers do not usually appreciate the fact that the one-way ANOVA F test for independent groups can be computed from summary statistics, as well as from raw subject data. Gordon (1973) seems to have been the first to publish the appropriate equations. Unfortunately, there was an error in one of his equations, which was corrected in Rossi (1987b). The correct equations for the between-groups and within-groups sums of squares are the following:

$$SS_{BG} = \sum n_i M_i^2 - \frac{\sum (n_i M_i)^2}{\sum n_i} \quad (1)$$

and

$$SS_{WG} = \sum (n_i - 1) s_i^2, \quad (2)$$

where n_i is the sample size of the i th group, M_i is the mean of the i th group, and s_i is the unbiased estimator of the standard deviation of the i th group. Once the sums of squares have been obtained, the determination of the degrees of freedom, mean squares, and F ratio is straightforward. Despite its simplicity and convenience, this procedure is not contained in any of more than 100 recently published statistics textbooks that I examined.

Program development was partially supported by Grant CA27821 from the National Cancer Institute. The author's mailing address is Department of Psychology, Chafee Social Science Center, University of Rhode Island, Kingston, RI 02881-0808.

These formulas may be useful to both researchers and teachers in a variety of situations in which raw data may not be readily accessible. Perhaps the most common situation involves the analysis of already published data. For example, an investigator may wish to verify the accuracy of a published F value (see Rossi, 1987a) or may want to know the F value for a subset of treatment conditions. In many research papers, F values are not even reported, especially when the overall test is not significant. Determination of an F value in this situation may help the reader evaluate the research report with respect to such rarely reported factors as magnitude of experimental effect (e.g., omega-squared, eta-squared) and statistical power. Finally, these formulas may also be of interest to teachers of statistics and experimental methods courses, especially for classroom exercises such as those described by Rossi (1987a).

Program. To facilitate the use of these formulas, a simple interactive BASIC computer program was developed. The program asks the user how many groups are to be included in the analysis, and then prompts the user to supply the mean, the standard deviation (unbiased estimator version), and the sample size for each group. Output is provided in familiar ANOVA summary table format. Also provided are several standard measures of effect size, including Cohen's (1977) f , omega-squared, eta-squared, and eta. The p value associated with the F statistic is not provided, since this would add substantially to the length and execution speed of the program, and since p can be determined easily using commonly available tables of the F distribution. Users who wish to incorporate an appropriate algorithm into the program may choose from among several easily available sources (e.g., Collani, 1983; Cooke, Craven, & Clarke, 1982; Edgeman, 1984; Wood & Wood, 1986).

When the number of groups is equal to two, the program prints the value of the t statistic, its degrees of freedom, and the following effect size measures: Cohen's (1977) d , omega-squared, eta-squared, and eta. The program also provides a modified t value (t') for the heterogeneous variance case (Howell, 1982, p. 137), as well as Satterthwaite's (1946) approximation for degrees of freedom. It is left to the user to decide when variances are heterogeneous.

The program was developed on an IBM PC/AT microcomputer and is written in IBM BASIC version 3.10. It should run on most IBM-compatible machines with very little modification. The program uses approximately 2K of memory and contains 69 lines of code, including 2 REMarks and 19 output formatting statements. Typical execution times on the IBM PC/AT are less than 1 sec.

Program Accuracy. Program accuracy is dependent on the number of significant digits used in entering the summary statistics. Accuracy tests conducted by the author (available upon request) suggest that error rates

of less than 1% can be expected when at least three significant digits are used. Use of fewer than three significant digits is not recommended.

Availability. A listing and sample run of the program may be obtained without charge from the author. For a disk copy of the program, send a formatted IBM diskette (5.25-in. only, DOS 2.10 or later) in a returnable diskette mailer.

REFERENCES

- COHEN, J. (1977). *Statistical power analysis for the behavioral sciences* (rev. ed.). New York: Academic Press.
- COLLANI, G. VON. (1983). Computing probabilities for F, t, chi-square, and z in BASIC. *Behavior Research Methods & Instrumentation*, *15*, 543-544.
- COLLANI, G. VON, & WALOSZEK, G. (1983). UNIVARAN: A universal analysis of variance program. *Behavior Research Methods & Instrumentation*, *15*, 607.
- COOKE, D., CRAVEN, A. H., & CLARKE, G. M. (1982). *Basic statistical computing*. London: Edward Arnold.
- CORRIGAN, J. C., BONELLI, P. J., & BORYS, S. V. (1980a). BASIC programs for one-way through four-way between-subjects and within-subjects ANOVAs. *Behavior Research Methods & Instrumentation*, *12*, 468.
- CORRIGAN, J. C., BONELLI, P. J., & BORYS, S. V. (1980b). BASIC programs for two-way through four-way mixed design ANOVAs. *Behavior Research Methods & Instrumentation*, *12*, 546.
- COULOMBE, D. (1984). Two-way ANOVA with and without repeated measurements, tests of simple main effects, and multiple comparisons for microcomputers. *Behavior Research Methods, Instruments, & Computers*, *16*, 397-398.
- COULOMBE, D. (1985). Two-way multivariate analysis of variance: A BASIC program for microcomputers. *Behavior Research Methods, Instruments, & Computers*, *17*, 137-139.
- EDGEMAN, R. L. (1984). P-DIST: A microcomputer program for obtaining probabilities from univariate distributions. *American Statistician*, *38*, 321.
- GALLA, J. P. (1984). Simple analysis of covariance: A BASIC program for microcomputers. *Behavior Research Methods, Instruments, & Computers*, *16*, 564-565.
- GORDON, L. V. (1973). One-way analysis of variance using means and standard deviations. *Educational & Psychological Measurement*, *33*, 815-816.
- HACKER, M. J., & ANGIOLILLO-BENT, J. S. (1981). A BASIC package for N-way ANOVA with repeated measures, trend analysis, and user-defined contrasts. *Behavior Research Methods & Instrumentation*, *13*, 688.
- HOWELL, D. C. (1982). *Statistical methods for psychology*. Boston: Duxbury.
- ROSSI, J. S. (1987a). How often are our statistics wrong? A statistics class exercise. *Teaching of Psychology*, *14*, 98-101.
- ROSSI, J. S. (1987b). One-way ANOVA from summary statistics. *Educational & Psychological Measurement*, *47*, 37-38.
- SATTERTHWAITE, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin*, *2*, 110-114.
- WOLACH, A. H. (1983). *BASIC analysis of variance programs for microcomputers*. Monterey, CA: Brooks/Cole.
- WOOD, D. L., & WOOD, D. (1986). Precise F integration. *Behavior Research Methods, Instruments, & Computers*, *18*, 405-406.

(Revision accepted for publication February 16, 1988.)