

SMOOTH: A BASIC program for smoothing a data series

MICHAEL J STRUBE

Washington University, St. Louis, Missouri

When examining a series of data points, a researcher may find it useful to highlight the general trends. This aids in the interpretation of subsequent statistical analyses. A common procedure for highlighting the trends in a series is to "smooth" the series (Hartwig & Dearing, 1979; Tukey, 1977). Smoothing a series removes the influence of outlying cases so that the underlying pattern can emerge. The most common smoothing procedures entail either the use of running medians or the calculation of medians of successive overlapping groups of constant size throughout the series. The procedures can be quite time-consuming, particularly for a long series. The program described here was designed to calculate the smoothed version of a data series.

The program utilizes the 53R'H method (see Hartwig & Dearing, 1979). This method first requires calculation of running medians of groups of five data points. At the end points, the original data are carried over, and the points adjacent to the end points are determined by taking the median of the three data points at each end of the series. Next, running medians of groups of three data points are calculated (again, end points are carried over to the new series). This process is continued until there is no change in the series from one iteration to the next. Finally, a procedure known as "hanning" is conducted. This represents the calculation of running medians of

groups of two data points. This, of course, simply requires calculating the average of successive pairs of data points (the end points are carried over to the new series). This procedure is conducted twice. The final result of these three major steps in the smoothed series.

Program Input. The program allows the user either to enter the data directly or to enter the data from an existing sequential file on disk. Following data entry, the original series can be examined and corrections made. At the user's option, the original data series can be saved on disk.

Program Output. The user chooses whether the output is to be displayed on the screen or routed to a printer. If the output is displayed on the screen, the program prints only enough data to fill the screen and then prompts the user to continue. At the user's option, the smoothed series can be saved on a disk.

Additional Comments. The calculation of medians is made with adjustment for multiple data points in the score interval containing the median. This procedure is more precise than that of simply taking the score value of the interval containing the median (cf. Hartwig & Dearing, 1979, who use the latter).

Program Availability. The program is written in BASIC and is IBM-PC compatible. A listing of the program can be obtained from the author.

REFERENCES

- HARTWIG, F., & DEARING, B. E. (1979). *Exploratory data analysis*. Volume 16, Sage series: Quantitative applications in the social sciences. Beverly Hills, CA: Sage Publications.
TUKEY, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.

Address correspondence to Michael J Strube, Department of Psychology, Washington University, St. Louis, MO 63130.

(Manuscript accepted for publication September 9, 1986.)