Data manipulation and reduction in a 32K environment

KEVIN FLANAGAN

St. Louis University, St. Louis, Missouri 63103

Examples of data transformation of EBR binary data and subsequent statistical manipulation are shown using standard Data General software.

Psychologists have made the point that behavior is determined by environment (Skinner, 1971). At this time, I would like to report on some adaptive behaviors available to the psychologist manipulating and reducing data in a minicomputer environment.

The presenting problem was to multiply analyze data representing 114 averaged evoked brain responses (EBRs) using a particular Data General Nova 820 installation (King, 1975). The averaged EBR came from prior studies (Dinges, Note 1; Schweitzer & Tepas, 1974), and were comprised of 500 values per EBR, signifying relative potential differences (in arbitrary units) at consecutive 1-msec intervals, between electrode pairs attached to the scalps of human subjects. Furthermore, each of the averaged EBRs was already a hybrid form of scalp-recorded data, being the average of over 800 single EBRs recorded during the 500-msec period following identical stimulus events defined by the onset or cessation of auditory or visual stimuli of specified intensities. The resulting 57,000 numerical values were available as Teletype listings and punched paper tapes, the outputs of St. Louis University's APE II program (Tepas, Kress, & Klingaman, 1975) used with the Digital Equipment Corporation (DEC) LAB-8 system.

The demands of my project required the analysis of these data using various techniques suggested in the literature in order to compare their relative utilities (Flanagan, Note 2).

METHOD

The principal methodological hurdle was the problem of inputting the pre-existing data to the Data General system. The manual keying of 57,000 seven-digit numbers was a frightening and impractical prospect.

This paper is a presentation from the Data General User's Group meeting at the fifth annual meeting of the National Conference on the Use of On-Line Computers in Psychology. The research reported is a portion of the author's doctoral dissertation work at St. Louis University. The author wishes to thank David Dinges, Paula Schweitzer, and Donald Tepas for making their data available to him. He also thanks John King for making the XLAT program available and for helping make the Nova 820 system at the Illinois State University accessible.

Unfortunately, the particular field format used for the DEC paper tapes was not immediately intelligible to the Data General system, however. To put one problem in descriptive terms, for instance, the DEC tapes placed the "minus sign" to the left, while the Data General system looked for it on the right. To one unskilled in direct field manipulations, this posed a difficult problem; but the Data General system provided an indirect translation alternative which required essentially nothing more than the same BASIC language programming skills necessary to run the subsequent analyses.

Using a pre-existing Fortran routine, XLAT (King, Note 3), the data tapes were read onto disk files so that every field point was given representation as the ASCII value "1" or "0" for "punched" and "unpunched," respectively. Once this was accomplished, it was only necessary to write a BASIC program (Flanagan, Note 2) to manipulate these "data" files so that they were meaningful as EBRs again by merely inverting the rules of their original encoding.

Although the minicomputer makes this a very simple translation method for the user, the question of its possible use is an important one. There was, in fact, insufficient space to store ASCII representations of the 16,000 paper-tape field points for each of 114 EBRs at the same time (a total of over 1.8 million ASCII values), much less reserve room to generate the translations to supersede them. However, the Data General system was able to do this comfortably taking 30 EBRs at a time.

Once the ASCII files established by XLAT were created, the BASIC program translated these files and performed the additional operation of normalizing values within each EBR as Z scores, so that the line printer listings of the 30 normalized EBRs were complete within about 90 min. To evaluate the efficiency of this translation operation, consider that, on a per EBR basis, this is about one-fifth the time it would take for the original APE II listings to be completed on the LAB-8 system using a Teletype.

This indirect manipulation of field format also included a bonus advantage. Tape format itself, which needed to be dealt with in any case, was handled in the

same step. Thus, the absence of a punched leader code on the DEC tapes, the use of only six of the eight points per paper-tape line, and the use of four paper-tape lines per numerical value were all accommodated at the same time as the manipulations within the data field format.

Besides the normalization of EBR time-sampled potential values which was done along with the data translation, the normalized EBR data were subjected to four subsequent analyses. These were the derivation of an autocorrelogram (Walter, 1963) and periodogram (Box & Jenkins, 1970) for each EBR, the derivation of crosscorrelograms (Walter, 1963) between selected pairs of EBRs, and the computation of point-by-point *t* tests (Peimar, 1968) between selected pairs of EBRs averaged across subjects within stimulus conditions. These analyses were performed by BASIC programs (Flanagan, Note 2).

Each of these analyses demands iterative or even nested iterative procedures when applied to 500-point time series data. How efficiently can the minicomputer run through such number crunching? In this case, the autocorrelogram and crosscorrelogram computations nested routines iterated from 375 to 500 times within routines repeated 126 and 251 times, respectively. Yet, the autocorrelogram took about 4 min and the crosscorrelogram about 8 min to be computed and printed. The periodogram, which was produced in about 6 min, was derived from two routines repeated 30 times, one of which included two other routines iterated 500 times. The point-by-point *t* test iterated two routines 500 times, requiring only about 2 min.

Perhaps the efficiency of speed is not as surprising as the efficiency of accuracy which was also noted. The fact that the first term of the autocorrelogram represents the correlation of the 500 EBR sample points with themselves offered an opportunity to check this accuracy in the context of such iterative manipulations. This first value, computed by a routine iterated 500 times on data which have already been

manipulated in the translation and normalization procedures cited, ideally equals exactly one. The operations of multiplication, division, and extraction of square roots, as well as addition and subtraction, contributed to the determination of this value. In 55 of 114 cases, or about 48% of the time, the computed value was 1.000000. In the other 59 cases, cumulative round-off errors averaged to just under 3×10^{-6} , ranging from 1×10^{-6} in 18 cases to a maximum of 1.2×10^{-5} in one case. Such accuracy establishes minicomputers of this sort as more than adequate for this and many other applications.

REFERENCE NOTES

1. Dinges, D. *Human visual evoked cortical potentials to stimulus onset and offset*. Unpublished thesis, St. Louis University, 1973.
2. Flanagan, K. *Mathematical analyses of evoked response data*. Dissertation in preparation, St. Louis University, 1976.
3. King, J. Personal communication, January 1975.

REFERENCES

- BOX, G., & JENKINS, G. *Time series analysis: Forecasting and control*. San Francisco: Holden-Day, 1970.
- KING, J. Time-shared control systems: Promises and problems. *American Psychologist*, 1975, **30**, 226-229.
- PEIMAR, I. On the application of computing techniques to the investigation of short-time nonperiodic processes in the electroencephalogram (Discussion). In M. Livanov, & V. Rusinov, (Eds.), *Mathematical analysis of the electrical activity of the brain*. J. Barlow (trans.) Cambridge: Harvard University Press, 1968.
- SCHWEITZER, P., & TEPAS, D. Intensity effects of the auditory evoked brain responses to stimulus onset and cessation. *Perception & Psychophysics*, 1974, **16**, 396-400.
- SKINNER, B. F. *Beyond freedom and dignity*. New York: Bantam, 1971.
- TEPAS, D., KRESS, G., & KLINGAMAN, R. APE: Average potential evaluation software for the LAB-8 system. *Behavior Research Methods & Instrumentation*, 1975, **7**, 120-124.
- WALTER, D. Spectral analysis for electroencephalograms: Mathematical determination of neurophysiological relationships from records of limited duration. *Experimental Neurology*, 1963, **8**, 155-181.