

A computer-aided procedure for producing interesting assignments for students of applied statistics

ROBERT J. TRACY
De Paul University, Chicago, Illinois 60614

The GEN computer program is part of an overall plan to make statistics interesting for the reluctant student, while simultaneously pursuing important cognitive goals. Interest is aroused by using real data collected on students themselves, rather than using fictitious workbook examples or data which have been artificially generated (e.g., Salmon & Tracy, 1975).¹ The student can subsequently see his own scores being entered into various data analyses.

Prior to using the GEN program, it is recommended that the students themselves participate in a brief study selected for its interest value. An example is a study on ESP in which students "predict" for 20 tosses of a coin whether the coin will land "heads" or "tails." Students can test whether the class performed significantly higher or lower than expected by chance. Additional data can be collected, such as students' sex, year in school, major, or GPA, so that students can test additional hypotheses, e.g., whether people who have higher ESP scores tend also to have higher grade point averages. There are other studies which can be conducted in just a few minutes of class time, and when data from such studies are the basis for a variety of assignments, student interest and learning may be increased for the whole course.

Certain cognitive goals may be efficiently pursued in the context of such class exercises. Students are exposed to the data collection process, including needed controls and possible difficulties; students may gain some substantive knowledge of the phenomenon investigated; and they should also learn about statistics. Since data on many variables can be collected, students can choose which variables to examine, which hypotheses, and which statistical procedures are appropriate—decisions which are identical to those required of a researcher. The GEN computer program will generate a unique exercise for each student, thus discouraging plagiarism. The program also produces not only the final answers to problems but also intermediate summary statistics, so that student computation errors may be located and partial credit given.

Program Use. The instructor enters original data obtained from one or more groups of subjects, with each subject having a score on each variable. There is no limit on the number of subjects or variables except for that imposed by CPU storage.

The GEN program is subsequently used in one of two possible modes: print and statistics. In the *print* mode, the program generates and prints each data set for distribution to students. The program generates

each data set by randomly deleting from the original data a specific number of subjects indicated by the user, a process which generally assures the uniqueness of each generated data set.

Students are given several sequential assignments for the school term, requiring, for example, computation of the mean and standard deviation of one variable or the correlation between two variables, as the topics are covered in the course. In order to complete a homework assignment, the student selects a variable (or variables) which most interests him, indicates the variable name on his homework, and performs the required calculations based on that variable.

A teaching assistant can grade the completed assignment by using the GEN program in the *statistics* mode. The program will construct each student's data set internally (without printing it out) and in a fashion which allows the grader to do computations on any variable the student may have selected for study.

The grader specifies whether statistics will be computed on one or two variables. If selecting one variable, the grader must indicate whether or not answers will be printed for a *t* test for a single sample, a frequency distribution, a mean and standard deviation (which also causes other summary statistics to be printed), and raw scores. If the grader selects two variables, then he indicates whether results will be printed for: Pearson *r*, regression, matched-groups *t*, independent-groups *t*, means and standard deviations, and raw scores for the two variables.

Without interrupting a computer run, the grader may change the statistics being computed, branch to another student's data set, or select another variable(s) for analysis. The GEN program is terminated by depressing the **BREAK** key when the program is in operation or by simultaneously depressing the **CTRL** and **C** keys when the computer has paused to receive user input.

Computer and language. The BASIC program was designed for time-sharing use and was tested on a 2000F Hewlett-Packard computer. The program occupies less than 6,000 words of storage.

Availability. A program listing with sample output, a description of program use, and sample homework assignments may be obtained at no charge from Robert J. Tracy, Department of Psychology, 2219 North Kenmore Avenue, Chicago, Illinois 60614.

REFERENCE

- SALMON, P. G., & TRACY, R. J. Computer-generated computation exercises. *Behavior Research Methods & Instrumentation*, 1975, 7, 307.

NOTE

1. The Salmon and Tracy (1975) programs are written in BASIC. A comparable set of programs, written in FORTRAN, is Tracy, R. J., and Salmon, P. G., *Computer-generated computation exercises*. Baldwin, New York: Life Science Associates, 1974.