SEPP (SAMPLING ERROR PROGRAM PACKAGE)

Martin R. Frankel, Baruch College, CUNY
and University of Michigan

INTRODUCTION

The SEPP (Sampling Error Program Package)[1] consists of three computer programs which produce sampling errors (BRRP, PSALMS and ABSERD). BRRP (Balanced Repeated Replication Package) computes sampling errors using the method of balanced repeated replications. At present it can compute these sampling errors for ratio means, simple differences of ratio means, simple correlation coefficients, multiple regression coefficients (standardized and unstandardized), (p-1) order partial correlation coefficients, and multiple correlation coefficients.

PSALMS (Paired Selection Algorithm for Multiple Subclasses) computes sampling errors using a method based on the Taylor approximation (also known as the δ or delta method, or linearization or propogation of error, and sometimes the Keyfitz method). At present it can compute sampling errors for simple ratios, and linear combinations of ratios (e.g., differences of ratios). Unlike BRPP it can easily deal with many ratios from different subclasses of the sample in a single run. In addition, unlike BRRP it is designed to deal automatically with poststratification class weights, although this option has not yet been checked out and cannot be used at present. ABSERD (A Basic Sampling Error Program for Ratios and Differences) is a simpler and more modest program appended as a useful addition to the two major programs.

ABSERD

ABSERD (A Basic Sampling Error Program for Ratios and Differences) was written for users who do not have access to large computers and need only a limited number of sampling errors for ratio means and their differences. ABSERD is written entirely in basic FORTRAN, and can be compiled and run on machines with a minimum core size of 16K. Unlike PSALMS and BRRP, the original data cases do not have to be sorted by strata and psu within strata before entry into the program. On a single program set-up and run sampling errors may be computed for a pair of ratio means and their difference. It is possible to compute sampling errors over different subclasses of the total sample in this single run. Because of its simplicity, this program has no provisions for handling missing data (blanks) or missing data codes and values. The user must filter out cards or records that contain these missing data indications to assure that blanks are not interpreted as zeros and missing data codes are not interpreted as data values.

BRRP

BRRP (Balanced Repeated Replication Package) was designed primarily to compute sampling errors for regression coefficients and correlation coefficients (simple, partial and multiple) associated with a single multiple regression equation. As a byproduct, the program also computes sampling errors for means and differences of means for all variables that enter the regression equation.

Sampling errors are computed by the method of balanced repeated replication (also known as pseudo-replication). [1,2] The program allows the user the option of specifing the pattern of replication to be used; fully balanced, partially balanced, or unbalanced. The program can compute sampling errors using either half-samples alone, or half as well as complement-half samples.

Communication between the user and the program is accomplished by the use of standard ISR variable description. This is described in more detail in the section on PSALMS that follows.

BRRP is modularized to simplify the computation of sampling errors for different subsets of independent variables in a reduced regression equation. The program requires an IBM 360 running under OS (Operating System).

PSALMS

The overwhelming majority of estimates produced from survey samples are ratio means (including proportions) and differences between means. PSALMS was designed to efficiently and economically compute sampling errors for these most often used estimators. Unlike other sampling error programs, PSALMS can compute variances for many different means over different

---

[1] *A more complete description of the programs is given in: Kish L.; Frankel, M.R.; and Van Eck, N. SEPP: Sampling Error Program Package. Ann Arbor: Institute for Social Research, 1972. Information about obtaining the programs is available from Neal Van Eck, Institute for Social Research, University of Michigan, Ann Arbor, Michigan, 48106. Support for the writing of SEPP came from the National Center for Health Statistics (PH-43-67-767), the National Science Foundation (GS-3191X), and from the Institute for Social Research.*

subclasses of the sample in a single run. Specification, by the user, of the ratio means and differences, as well as the sample subclasses for which sampling errors are desired is done in three steps.

First, each variable which is to be used in the formation of a mean or which enters into the definition of a sample subclass is assigned a unique identification number. This identification number is associated with a particular data field by the use of an ISR T-card. For each such variable, the user may assign values which he wishes to be interpreted as "missing data". After this assignment of identification numbers these original data variables are referred to by their number preceeded by the letter 'X'.

All means, or more correctly ratios, for which sampling errors are desired are formed as the ratio of two sample sums, a numerator and a denominator. Sometimes these numerators and denominators can be formed directly by summation of original data or X-type variables. But often, particularly if we are interested in a mean for a subclass of the total sample, summation across X-type variables will not give us the needed numerators or denominators. These are created by an internal recode language L1. Using information from the original X-type variables, the user may create 0-1 (B-type) variables or more general (Y-type) variables. For example, the L1 statement B1 IF X32=1-4; creates a denominator for a mean based on the subclass of the sample having values 1 through 4 for original variable 32. The statement Y1=X56 IF B1; forms a numerator containing the value of original variable 56, for this sample subclass.

The last step of the man-machine communication with PSALMS is the specification of the ratio means and differences for which sampling errors are desired. Here the user puts together numerators and denominators, based on either original data variables (X-type) or user created variables (B- and Y-type). For example, the statement R1=Y1/B1, tells the program to compute sampling errors for the mean of variable 56, over the subclass of the sample consisting of all cases with codes 1 through 4 on variable 32.

Like BRRP, PSALMS requires an IBM 360 running under OS.

REFERENCES

1. Kish, L. and Frankel, M., "Balanced Repeated Replications for Standard Errors." Journal of the American Statistical Association, 65 (1970), 1071-1094.

2. McCarthy, P.J., "Pseudo-Replication: Half Samples," Review of the International Statistical Institute, 37 (1969), 234-264.