

SSAIS User's Guide

1. Description

The program approximates **by Monte Carlo simulation** the expected value and the sampling distribution function of the adverse impact (AI) ratio for general **multi-stage selections**. The program focuses primarily on decisions with a **small to medium sized applicant sample**. Given the characteristics of the available selection predictors and the intended number retained or selected at each of the stages, the program estimates the average value, the median and the 25th and the 75th percentile value of the AI ratio as well as the expected majority and minority group selection rates. The median and the 25th and 75th percentile values are straightforwardly computed from the empirically derived sampling distributions of the AI statistic.

2. Assumptional Basis

The calculations are based on the assumption that the predictors have a joint multivariate normal distribution with the same variance/covariance matrix but a different mean vector in the two applicant populations. Given this assumption it is, without loss of generalization, understood that the joint distribution of the predictors is standard multivariate normal in the majority applicant population.

3. Technical Aspects

At present, the basic random generator, used to perform the Monte Carlo simulations, is set to a non-repeating state. So, different runs of the program will not produce identical (but highly similar) results. The program can handle up to four-stage selection scenarios with as many as 10,000 applicants and a maximum of 500,000 replications. To obtain reasonably accurate results, at least 10,000 replications should be used, unless the total number of applicants exceeds 1,000. In that case, and depending on the actual size of the applicant sample, between 1,000 and 10,000 replications will generally be sufficient.

To execute the program, a personal computer, running under Microsoft Windows 95, 98, XP, NT or 2000, is required. Also, the computer should have at least 64 MB RAM and, preferably, a clock speed of over 350 Mhz. As the program is computationally very intensive, the actual execution of the program may require anywhere between 1 and several hundreds of CPU seconds, depending on the size of the total applicant sample and the required number of Monte Carlo replications.

4. Input

- # 1: **NREP, NSUB, NSUB1, NF, NPRED**
Five integer values, separated by a blank, with

- **NREP**: the number of replications on which the simulation results are based ($NREP \leq 500,000$).
 - **NSUB**: the number of subjects in the total applicant group ($NSUB \leq 10,000$).
 - **NSUB1**: the number of subjects in the majority applicant group ($NSUB1 \leq NSUB$).
 - **NF**: the number of selection stages ($NF \leq 4$)
 - **NPRED**: the total number of available elementary predictors ($NPRED \leq 10$)
- # 2: **NMA(I)**, with $I = 1, \dots, NF$
NF integer values, separated by a blank. The element NMA(I) indicates the number of applicants that is retained (or selected) at the end of selection stage I.
 - # 3: **DIFP(I)**, with $I = 1, \dots, NPRED$
NPRED *real* values, separated by a blank. The element DIFP(I) refers to the effect size (with respect to the majority group) of the Ith predictor. So, effect sizes will usually have a negative value (cf. Anonymous, 2002b).
 - # 4 and following: **CORIN(I,J)** with both I and J ranging from 1 to NPRED
Row by row, NPRED *real* values, separated by a blank. CORIN specifies the **full** correlation matrix between the NPRED predictors.
 - # 5 and following: **WEP(I,J)** with $I = 1, \dots, NF$ and $J = 1, \dots, NPRED$
Row by row, NPRED *real* values, separated by a blank. WEP(I,J) specifies the weight assigned to predictor J at stage I. Zero weights are allowed, but not all the elements of a row may be equal to zero.

5. Sample Input File

Important: in preparing the input file, use a simple text editor such as Notepad, Wordpad or any other standard ASCII producing editor. DO NOT USE TEXT PROCESSING PROGRAMS SUCH AS MS-WORD or WORDPERFECT. Also, when saving the input file in Notepad, use the option “All Files” in the “Save as type” box. When saving in Wordpad, use the “Text Document-MS-DOS Format” option in the “Save as type” box, and be **aware that Wordpad has the nasty habit of adding the extension .txt to the file name that you specify**. Thus, with Wordpad, if you specify the name of the input file as “MINPUT”, the file will in fact be saved as “MINPUT.TXT”; and this is the name that you have to use in the command to run the present program.

Here is a sample input file for the *ssais* program.

```

10000 200 150 3 4
120 50 20
-1.000 -0.230 -0.090 -0.330
 1.000  0.240  0.000  0.190  0.300  0.160
 0.240  1.000  0.120  0.160  0.300  0.260
 0.000  0.120  1.000  0.510  0.180  0.200
 0.190  0.160  0.510  1.000  0.280  0.250
 1.000  0.000  0.000  0.000
 2.000  0.000  1.000  2.000
 2.000  4.000  1.000  2.000

```

6. Running the Program

Suppose you copied the executable source of the program to the d:ssel directory on your machine. In that case, the input file must also be saved in the d:ssel directory. Next, to run the program, you have to open an MS-DOS Command window. The way to do this varies from one operating system (i.e., Windows 95, 98, NT a.s.o.) to the other, and you should use your local "HELP" button when in doubt about this feature.

In the MS-DOS Command window you type d:, followed by RETURN or ENTER, and your computer will return the D:\> command prompt. Next, you type cd ssel after the D:\> command prompt, again followed by RETURN or ENTER, and your computer will respond with the D:\ssel> command prompt. Now, you can execute the program by typing `ssais < minput > moutput` where "minput" is the name of the input file and "moutput" is the name of the output file. At the end of the execution, the PC will return the command prompt D:\ssel> (**Remember, however, that the execution may take quite some time; so, be patient!**). You can then inspect the output by editing the output file with either Notepad, Wordpad or any other simple editor program.

7. Sample Output

Program execution starts on 30/11/2003 at 17:51: 1

```

+++++++
+  SSAIS  +
+++++++

```

Estimation through simulation of the sampling distribution function of the adverse impact (AI) ratio for multi-stage selections.

Program written by Wilfried De Corte, Ghent University, Belgium

The program uses routines from the Slatec library (see <http://www.geocities.com/Athens/Olympus/5564>) and the Ranlib distribution (see <http://www.netlib.org/random/>)

PROBLEM SPECIFICATION

Number of selection stages: 3

Number of replications for the simulation: 10000

Total number of applicants: 200

Number of majority group applicants: 150

Number of remaining applicants at the end of stages 1 to 3
 120 50 20

Predictor weights (from predictor one to the last)

in the different stages (rows 1 to 3}

1.000 0.000 0.000 0.000

2.000 0.000 1.000 2.000

2.000 4.000 1.000 2.000

Effect sizes (i.e., standardized difference between the minority and the majority group) of the predictors

-1.000 -0.230 -0.090 -0.330

Correlation matrix of the predictors

Row 1 1.000 0.240 0.000 0.190

Row 2 0.240 1.000 0.120 0.160

Row 3 0.000 0.120 1.000 0.510

Row 4 0.190 0.160 0.510 1.000

PROGRAM OUTPUT

Effect size predictor composite of stage 1 is: -1.000000

Effect size predictor composite of stage 2 is: -0.775957

Effect size predictor composite of stage 3 is: -0.612347

	SR	AI	SRA	SRI
Expected Value	0.100	0.255	0.123	0.030

SR: hiring rate total applicant group

SRA: hiring rate majority applicant group

SRI: hiring rate minority applicant group

Detailed Results: Number selected at the stages 120 50 20

AI: Mean and 50--25--75th perc 0.255 0.158 0.158 0.333

Frequency and Cumulative frequency function of AI statistic

	Value	Freq	Cfreq	Number
	Undef	0.000		0
1	0.000	0.213	0.213	2130
2	0.158	0.343	0.556	3431
3	0.333	0.257	0.813	2571
4	0.529	0.133	0.947	1333
5	0.750	0.040	0.986	395
6	1.000	0.012	0.998	118
7	1.286	0.002	1.000	20
8	1.615	0.000	1.000	2

Value: value of the AI ratio; Undef means that
the value is not defined

Freq: Proportion of replications that resulted in the
corresponding AI ratio value

Cfreq: Cumulative proportion

Number: Number of replications that resulted in the
corresponding AI ratio value

SRA: Mean and 50--25--75th perc 0.123 0.127 0.120 0.127

SRI: Mean and 50--25--75th perc 0.030 0.020 0.020 0.040

CPU TIME IN SECONDS 12.06

8. Description of Output

The output is self-explanatory.

9. Dependencies and Acknowledgment

The present program is written in Fortran77. It was compiled to an executable code for WIN32 PCs (i.e., Windows 95/98/ME/XP or NT/2000) with the GNU Fortran G77 compiler (cf. <http://www.geocities.com/Athens/Olympus/5564/>). The program uses routines from the SLATEC program library (cf. Fong et al., 1993; <http://www.geocities.com>).

com/Athens/Olympus/5564/) as well as some code from the RANLIB package (cf. L'Ecuyer & Cote, 1991; <http://www.netlib.org/random/>).

When the user reports results obtained by the present program, reference should be made to De Corte (2003a, 2003b).

10. References

De Corte, W. (2003a). SSAIS User's Guide.

De Corte, W. (2003b). SSAIS: A program to assess adverse impact in multi-stage selection decisions. Educational and Psychological Measurement (in press).

Fong, K. W., Jefferson, T. H., Suyehiro, & Walton, L. (1993). Guide to the SLATEC common mathematical library (<http://www.netlib.org/slatec/>).

L'Ecuyer, P. & Côté (1991). Implementing a Random Number Package with Splitting Facilities, Transactions on Mathematical Software, 17, 98-111.