

SDAIZB User's Guide

1. Description

SDAIZB is identical to the SDAIZ program, except that it is no longer assumed that the applicant samples are random samples from corresponding populations in which the predictor scores follow a normal distribution. Instead, the predictor score distribution within the applicant populations is equated to the marginal test score distribution obtained from the bivariate normal distribution (that represents the joint distribution, with correlation parameter ρ , of the test scores, X , and the scores on a pre-screening variable, Z), after suitable truncation from below (at the pre-screening variable cut score value z_c with respect to the minority applicant population) of the recruitment variable Z . Thus, z_c corresponds to the retention rate with respect to the pre-screening variable: retention rate, $s_z = 1 - \Phi(z_c)$. More specifically, it is assumed that (X, Z) have the standard bivariate normal distribution, with correlation parameter ρ , in the minority applicant population, such that the marginal density of the predictor scores in the minority population can be computed as

$$f(x) = \frac{\phi(x) \left[1 - \Phi \left(\frac{z_c - \rho x}{\sqrt{1 - \rho^2}} \right) \right]}{1 - \Phi(z_c)}$$

(cf. formula 46.173 in Kotz, Balakrishnan & Johnson (2000). Continuous Multivariate Distributions, Vol. 1.) where z_c is the pre-screening cutoff value z_c with respect to the minority applicant population, and ρ is the correlation between X and Z in the two populations. Alternatively, it is assumed for the majority group applicants that $(X, Z) \sim N_2(\boldsymbol{\mu}, \rho)$, where $\boldsymbol{\mu} = (d_X, \rho * d_X)'$ and d_X is the predictor effect size. As a consequence, the marginal density of the predictor scores in the majority population, after truncation with respect to the recruitment variable Z , is

$$f(x) = \frac{\phi(x - d_X) \left[1 - \Phi \left(\frac{z_c - \rho d_X - \rho(x - d_X)}{\sqrt{1 - \rho^2}} \right) \right]}{1 - \Phi(z_c - \rho d_X)}$$

For more details see De Corte and Lievens (2005).

2. Running the Program

First, copy the executable source of the program to a directory on your machine. Next, to run the program, double click the source file. The program automatically opens a MS-DOS Command window, and asks for input information. A typical example of the resulting dialogue is as follows:

Filename to write output of the program?

```
a.o
File a.o          found - write over it (y/n) ?
y
Total number of applicants?
40
Number of applicants from the majority group?
30
Required number of new employees?
12
Effect size predictor composite?
0.4
Correlation predictor--recruitment practice
and retention rate recruitment practice?
0.5 0.4
```

So, the program first asks for the name of the file on which the output will be written, and the user responds in this case by typing `sout.out` (the name of the file can be chosen freely, however) and pressing ENTER or RETURN. The output file (`a.o`) is automatically appended to the directory where the program file was copied. When the directory already contains a file with the same name, then the program asks permission to overwrite the file.

Next, the user provides the input data, as prompted by the program (Do not forget to press ENTER or RETURN after each submission of the requested input data). **The program does not perform validity checks on these input values, and the program will abort as soon as one of the input datum is not of the appropriate kind (e.g., when typing letters instead of a number).** Observe that the first three numbers must be integers, whereas the final number should have a decimal point (cf. 0.70).

When the input is complete, the program starts and displays some output in the MS-DOS Command window (**Observe that the program takes some time to perform the computations!**). This feature was added to inform the user that the program is still running. At the end of the computations, the program asks:

```
Ready to exit (Answer with y or n)?
```

Responding with `y` will close the application, and the output will be written on the file indicated by the user (here the file `sout.out`).

3. Technical Aspects

The present program is limited to situations where the total number of applicants is at most equal to 5,000. However, because of the numerical complexity, its application

is best further limited to selections with fewer than 1,000 candidates. Even in the latter case, the computations take almost 5,000 CPU seconds on a Pentium 3, 850 Mhz machine. For smaller applications. the execution time varies between a few seconds and several minutes.

To execute the program, a PC running under MS-Windows 95, 98, NT, XP or 2000 is required.

4. Sample Output

Program execution starts on 31/ 7/2004 at 10:36:13

```
+++++++  
++ SDAIZB ++  
+++++++
```

Analytical computation of the expected value and the sampling distribution function of the adverse impact ratio for single-stage selection decisions when the applicant group is a sample from a pre-screened population.

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 a couple of algorithms from StatLib (see <http://lib.stat.cmu.edu/apstat/>)

PROBLEM SPECIFICATION

```
Total number of applicants:      40  
Number of majority group applicants:  30  
Number of minority group applicants:  10  
Required number of new employees:    12  
Effect size predictor (composite):  0.400000  
Correlation predictor--applicant pre-screening practice:  0.500  
Retention rate pre-screening practice:  0.400  
    corresponding pre-screening cutoff value:  0.253
```

PROGRAM OUTPUT

Exact Sampling Distribution Function Adverse Impact (AI).

The value -1.000 for AI means AI is not defined

N2: number selected from the minority applicant group

N1: number selected from the majority applicant group

AIvalue	Prob	CProb	N2	N1
0.000	0.0715	0.0715	0	12
0.273	0.2470	0.3185	1	11
0.600	0.3344	0.6529	2	10
1.000	0.2323	0.8852	3	9
1.500	0.0910	0.9762	4	8
2.143	0.0208	0.9970	5	7
3.000	0.0028	0.9998	6	6
4.200	0.0002	1.0000	7	5

Exact values (i.e., values derived from the exact sampling distribution) for the expected selection rate in the majority group (SRA), the expected selection rate in the minority group (SRI) and the expected adverse impact ratio (AI). Observe that the expected adverse impact is computed over only the finite values of AI!

SRA	SRI	AI
0.330	0.210	0.691

Percentile values sampling distribution AI: 5, 10, 25, 50, 75, 90 and 95th percentile

P05	P10	P25	P50	P75	P90	P95
0.000	0.273	0.273	0.600	1.000	1.500	1.500

CPU TIME IN SECONDS 126.72

5. Description of Output

The output is self-explanatory.

6. Dependencies and Acknowledgement

The present program is written in Fortran77. It was compiled to an executable code for WIN32 PCs (ie, 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 a., 1993; <http://www.geocities.com/Athens/Olympus/5564/>) and a couple of algorithms from StatLib (<http://lib.stat.cmu.edu/apstat/>).

When the user reports results obtained by the present program, reference should be made to De Corte (2004) and De Corte and Lievens (2005).

7. References

De Corte (2004). SDAIZB User's Guide.

De Corte, W., & Lievens, F. (2005). The risk of adverse impact in selections based on a test with known effect size. *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/>).