

# Chapter 11

## RIndSel: Selection Indices with R



**Gregorio Alvarado, Angela Pacheco, Sergio Pérez-Elizalde, Juan Burgueño, and Francisco M. Rodríguez**

**Abstract** RIndSel is a graphical unit interface that uses selection index theory to select individual candidates as parents for the next selection cycle. The index can be a linear combination of phenotypic values, genomic estimated breeding values, or a linear combination of phenotypic values and marker scores. Based on the restriction imposed on the expected genetic gain per trait, the index can be unrestricted, null restricted, or predetermined proportional gain indices. RIndSel is compatible with any of the following versions of Windows: XP, 7, 8, and 10. Furthermore, it can be installed on 32-bit and 64-bit computers. In the context of fixed and mixed models, RIndSel estimates the phenotypic and genetic covariance using two main experimental designs: randomized complete block design and lattice or alpha lattice design. In the following, we explain how RIndSel can be used to determine individual candidates as parents for the next cycle of improvement.

### 11.1 Background

The linear selection index theory (see Chaps. 2 to 9 for details) can be difficult to apply without the use of specific codes developed in statistical analysis system (SAS) software. At the International Maize and Wheat Improvement Center (CIMMYT, for its Spanish acronym), codes were developed in SAS software version 9.4 (SAS institute 2017) that can help to determine individuals as parents for the next selection cycle. The SAS codes can be found at the following link: <https://data.cimmyt.org/dataset.xhtml?persistentId=hdl:11529/10242>.

---

G. Alvarado · A. Pacheco · J. Burgueño (✉) · F. M. Rodríguez  
Biometrics and Statistics Unit, International Maize and Wheat Improvement Center (CIMMYT), Mexico, Mexico  
e-mail: [g.alvarado@cgiar.org](mailto:g.alvarado@cgiar.org); [J.Burgueno@cgiar.org](mailto:J.Burgueno@cgiar.org)

S. Pérez-Elizalde  
Departamento de Socioeconomía Estadística e Informática, Colegio de Postgraduados, Mexico, Mexico

Afterward, the SAS codes were translated to R language as scripts (Pacheco et al. 2017) and denoted by RIndSel (R software to analyze Selection Indices), with the objective of creating a user-friendly graphical unit interface (GUI) in JAVA. The link to download the software is: <https://data.cimmyt.org/dataset.xhtml?persistentId=hdl:11529/10854>.

## 11.2 Requirements, Installation, and Opening

RIndSel is compatible with a Windows platform, in any of the following versions: XP, 7, 8, and 10; furthermore, it can be installed on 32-bit and 64-bit computers. To install RIndSel on a computer, the user must double-click on the executable file downloaded over the link given above and then follow the instructions that appear in the installation box. Once RIndSel has been installed, it can be opened by:

1. Double-clicking on the shortcut located in the desktop.
2. Locating it in the Windows menu and clicking.
3. Locating the software via the pathway C:/RIndSel, and double-clicking on RIndSel.exe.

As we shall see, the software has been partitioned into two modules.

## 11.3 First Module: Data Reading and Helping

This module (Fig. 11.1) deploys two small boxes upper left denoted by “*Open File*” and “*Help*.” With *Open File*, the user may access a set of files where he/she can open, for example, the file of phenotypic data, which should contain information



Fig. 11.1 Module for reading data

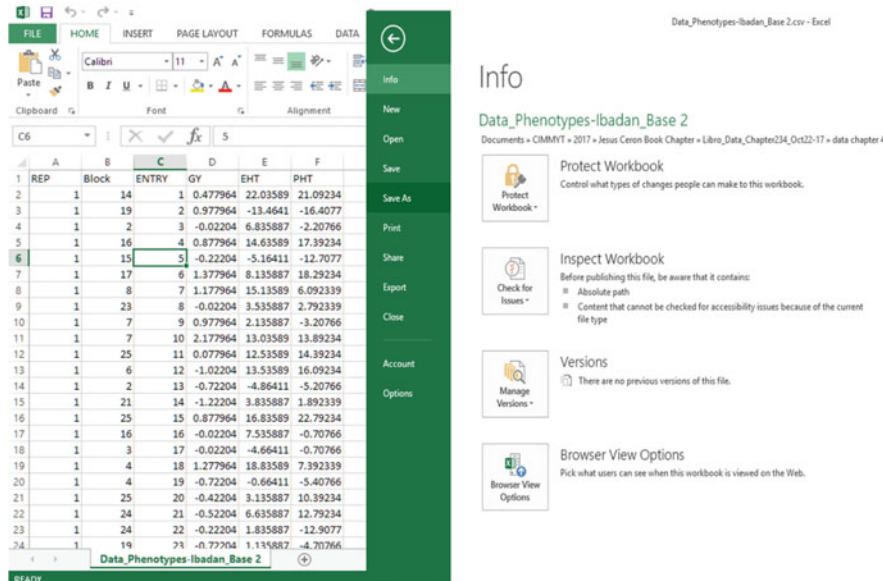


Fig. 11.2 Steps for saving a comma delimited file

associated with the experimental design. This file contains information about the field book where the experimental design variables can be identified in the first columns, whereas the remaining columns contain information about traits measured in the field; design variables and traits are connected by the plot number. Previously, the data set should have been captured in a spreadsheet using Excel or any other similar software and saved as a comma delimited file. To save the data as a comma delimited file in Excel, the following steps should be taken. In the Excel file that contains the data set (Fig. 11.2), select from the main menu: FILE → Save As → Browser View Options (look for the path where the data will be saved) → Save as type (look for CSV, comma separated values). The end of the file name should be “.csv,” indicating that the file is ready to be used.

The small box “Help” (Fig. 11.1) shows basic features such as the installation manual and software licenses. The installation manual provides a brief description of the selection indices that can be calculated and the pathway to where the software is located (Fig. 11.3). Furthermore, it shows folders related to the software features such as how the software could be used. There is also a folder called “Examples,” where the user can find data for test phenotypic selection indices, selection indices of coded score markers, and wide genome selection indices. The folders “Lib” and “Programs” contain information related to the software functioning; therefore, the authors highly recommend not modifying these folders.

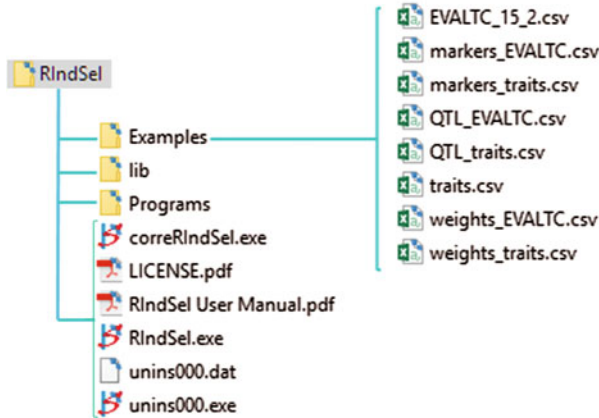


Fig. 11.3 Tree diagram of the RIndSel structure

## 11.4 Second Module: Capturing Parameters to Run

Once the data have been read (first module), RIndSel moves to the second module (Fig. 11.4), where some feedback is required:

1. To choose the selection index to calculate.
2. To select the experimental design.
3. To identify the variables of experimental design.
4. To choose the traits that will be used to calculate the selection index in the data file.

This module is structured in such a way that calculating any selection index is relatively easy. There are three other small buttons located upper left of the module: “Back,” “Analyze,” and “Help.” *Back* returns to the previous module (Fig. 11.1), *Analyze* executes and calculates the selection index, and *Help* provides the same functions as described in the previous section. In addition, there are four windows, each of which must be filled with the correct parameters. The first one is related to the indices that RIndSel is able to calculate (Fig. 11.5).

## 11.5 Selection Index

In this menu, it is necessary to define the percentage of genotypes that will be selected. By default, it is 5%, but any other percentage can be chosen. RIndSel uses the correlation matrix or the variance–covariance matrix to obtain the index; however, by default, the variance–covariance matrix is used. To work with the correlation matrix box, “Correlation” should be checked. The sign for “economic weights”

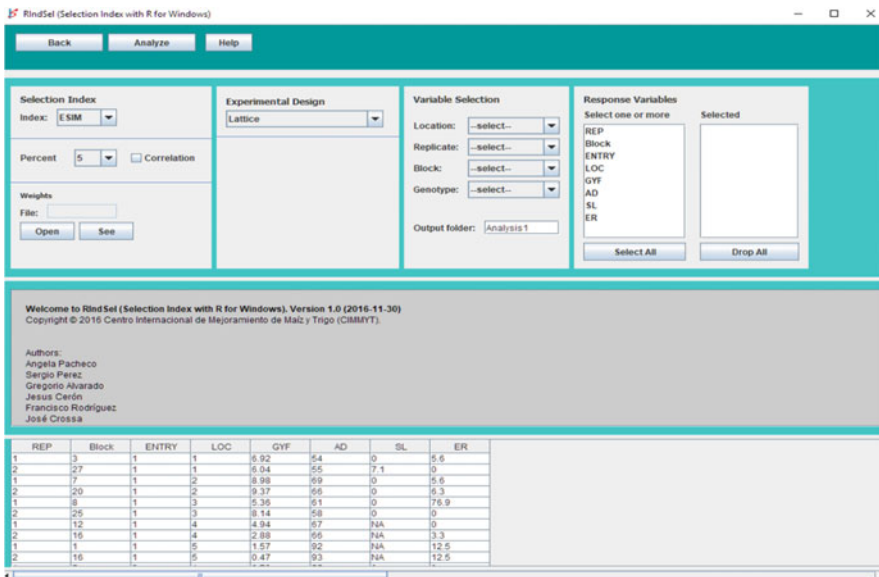


Fig. 11.4 RIndSel module of analysis

can be used to determine the behavior of the expected genetic gain of the traits. For example, with  $-1$ , the mean of the traits tends to decrease, whereas with  $1$ , it increases. It is also possible to use the trait heritability. The economic weights can be assigned by creating a comma-delimited file with the name of the trait and economic weight sign (Fig. 11.6a). Once the file has been created, it can be browsed by pressing the open button and where the \*.csv file is located (Fig. 11.6b).

To calculate the restricted linear phenotypic selection index (RLPSI or K&N, see Chap. 3 for details), it is necessary to create the same file and incorporate an additional column called “Restrictions.” This last column must be filled with the number one for those traits that remain fixed (restricted) and zeros for those traits that change (Fig. 11.7). An additional option is to ignore the “Weights” box, which means that RIndSel automatically presents an Excel file covering the options for capturing economic weights; the only requirement is that the file must be saved as a comma delimited file.

## 11.6 Experimental Design

The menu allows the user to select the field array design to be used. There are two choices:

1. Lattice or alpha-lattice
2. Random complete block designs

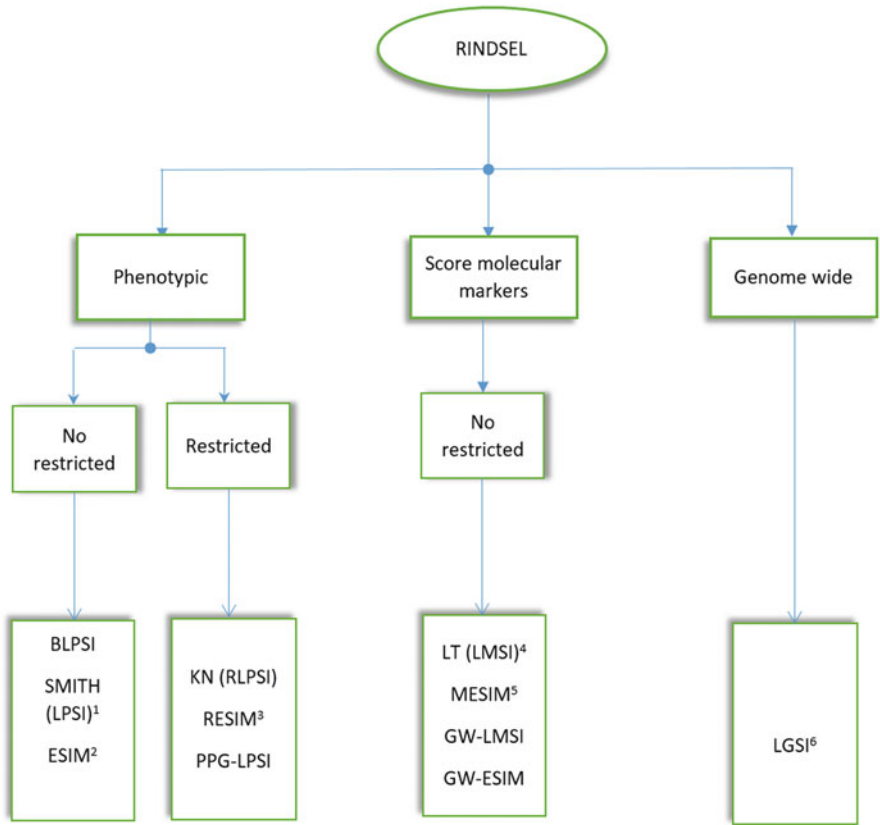


Fig. 11.5 Flow diagram of the selection indices that RIndSel is able to calculate; <sup>1</sup>Smith (1936), <sup>2,3</sup>Cerón-Rojas (2008a), <sup>4</sup>Lande R, Thompson R (1990), <sup>5</sup>Cerón-Rojas (2008b), <sup>6</sup>Cerón-Rojas (2015)

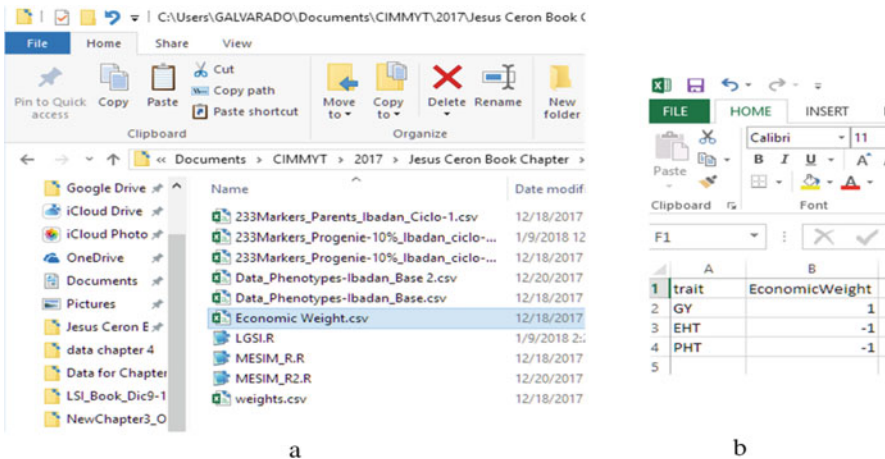


Fig. 11.6 Example of content for (a) economic weights of (b) file location

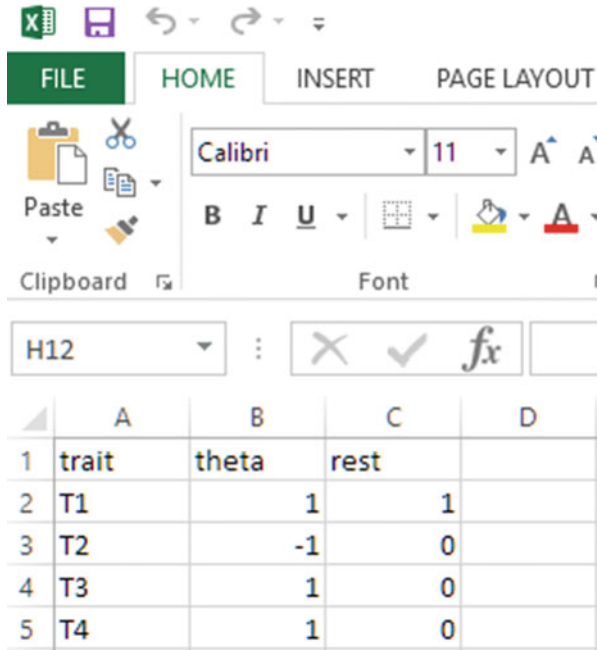


Fig. 11.7 Economic weights for restricted selection indices

### 11.7 Variable Selection

Experimental design is strongly related to the “*Variable Selection*” menu, where it is possible to identify the variables that constitute the experimental design. Thus, we can choose variables that match with the “*Location*,” replicate for random complete block design and block, provided that we have a lattice or alpha-lattice experiment.

### 11.8 Response Variables

In this menu, the user can select traits to be used to calculate the selection index. It can be activated by clicking on the trait to be selected. Figure 11.8 shows an example of how this window must be filled when a Smith phenotypic selection index is calculated.

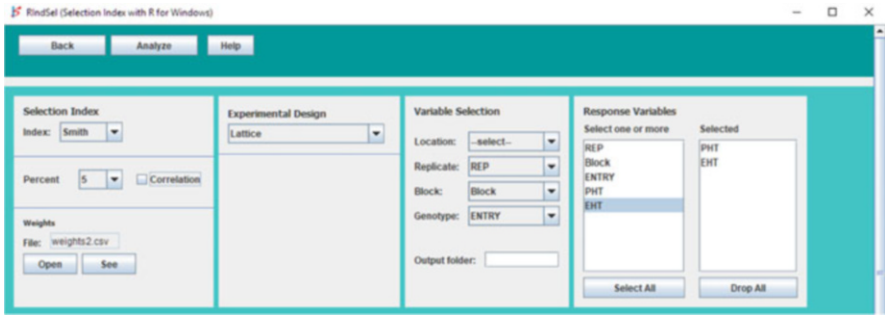


Fig. 11.8 Example of parameters that could be used to calculate a phenotypic selection index

### 11.9 Molecular Selection Indices

If the selection index to be calculated is molecular, such as the Lande and Thompson (1990) or the linear molecular selection index (Fig. 11.9, and see Table 1.1, Chap. 1, for details), two additional files are required:

1. Whole molecular markers matrix (green arrow).
2. Marker scores or estimated quantitative trait loci values (red arrow).

Marker scores can be obtained by making a regression of the phenotypic values on a codified molecular markers matrix (see Chap. 4 for details). The file can be created in Excel and must have the score with its respective marker for each trait; this file is saved with a .csv extension. An example of how these kinds of files must be generated is shown in Fig. 11.10a.

To calculate the scores in an F2 population, it is important for the molecular marker to have previously been codified as -1, 0, and 1 for genotypes aa, Aa, and AA respectively. When data come from a recombinant inbred line population, the molecular marker should be codified as -1 and 1 for homozygous genotype aa and AA respectively. In the genomic selection indices (LGSi) context (see Chap. 5 for details), it is only necessary to codify the molecular marker matrix (Fig. 11.10b), as these indices do not require a marker score.

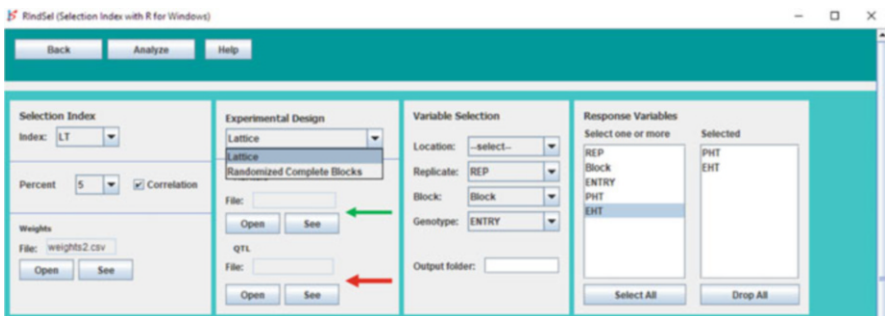
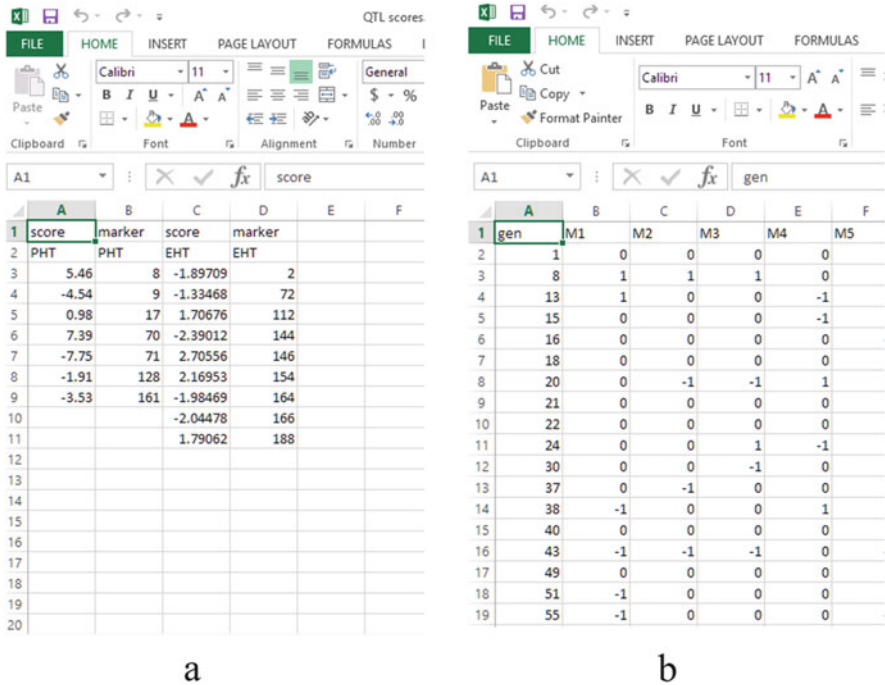


Fig. 11.9 Example of parameters that could be used to calculate a molecular selection index





**Fig. 11.10** Comma delimited files read in Excel for (a) scores of markers for traits plant height (PHT) and ear height (EHT), (b) a codified molecular marker matrix

### 11.10 How to Use RIndSel

The use of RIndSel can be illustrated with an example from the Smith linear phenotypic selection index (LPSI) (Smith 1936, see Chap. 2 for details). Figure 11.11 shows the phenotypic data (Fig. 11.11a), together with the file of economic weights (Fig. 11.11b). Three simulated traits (T1, T2, and T3) described in Chap. 2 were used. T1 and T3 are positive (economic value = 1), whereas trait T2 is negative (economic value = -1). It is important to remember that all data files must be saved in comma delimited format (\*.csv).

After the data and economic weights files have been generated, the data need to be loaded into RIndSel; thus, it is important to be able to find the pathway to where the files are located (e.g., "C://Book/datafile/C1\_PSI\_05\_Phen.csv"). Once the data file has been located, it must be uploaded, which can be done by clicking on the file, causing it to automatically begin this process. It is then possible go to the second module (Fig. 11.12) and select subsequent parameters from the menus. In this case, **Selection Index: Smith; Percent: 5; Weights: here we must look for where the economic weights are**, for example "C://Book/datafile/C1\_PSI\_05\_Phen Weights.csv." Once this file has been located, it must be selected by clicking.

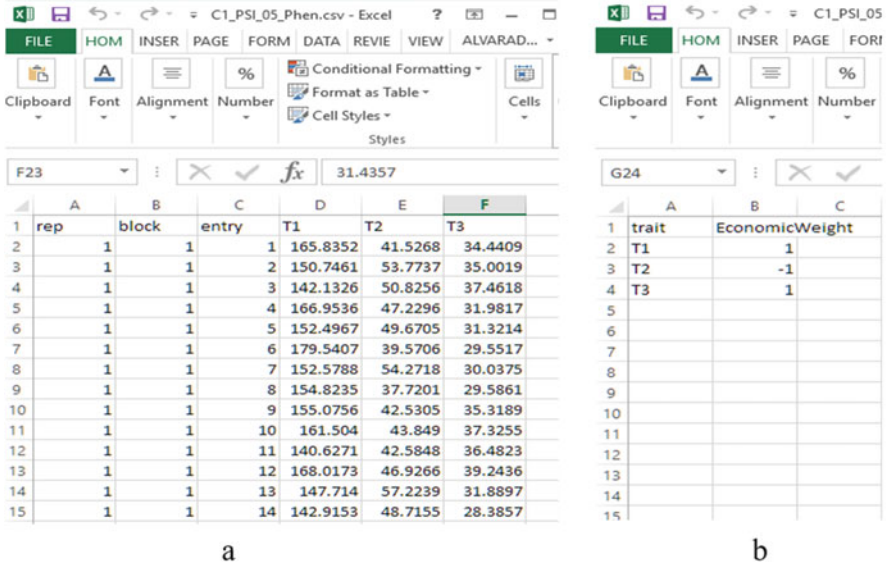


Fig. 11.11 Simulated data from Chap. 2 with (a) array in an alpha-lattice and (b) economic weights required to test the Smith linear phenotypic selection index (LPSI)

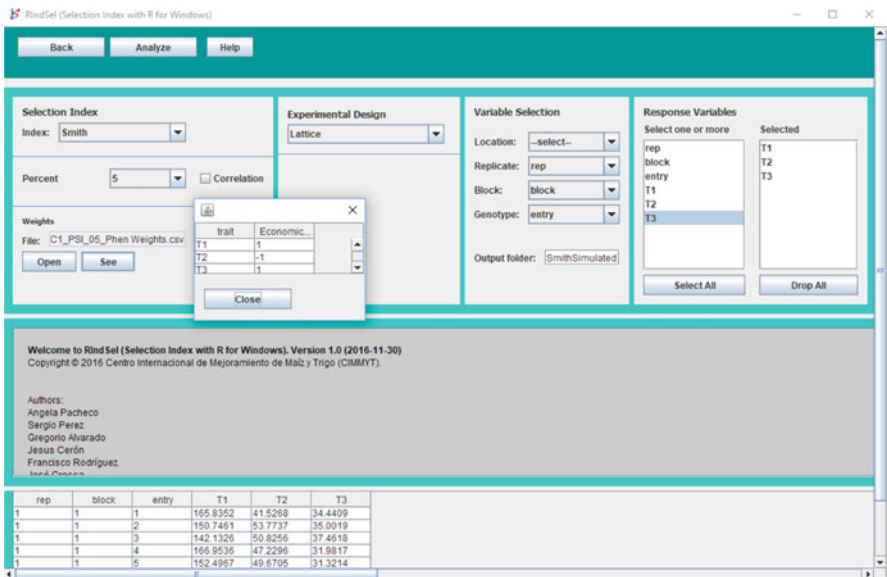


Fig. 11.12 Example of filling in a phenotypic selection index without restrictions

After the selection index windows are filled, the following menu is called: **Experimental design**, which allows the user to select the appropriate design – (for example, a lattice). To select the design variables, the user must navigate to the **Variable Selection**. In this example, the experiment has only one location, and the following should be selected: *rep* as **Replicate**, *block* as **Block** and *entry* as **Genotype**. An output name of the index must be assigned by writing its name in the **Box Output** folder, which is below the **Variable Selection** menu. For the Smith LPSI, the name chosen was *SmithSimulated*. Finally, the **Response Variables** menu should be filled by selecting the traits T1, T2, and T3.

## 11.11 RIndSel Output

This section explains the structure of the RIndSel output. First, RIndSel presents the genotypic variance–covariance matrix and the phenotypic variance–covariance matrix (Table 11.1). In addition, when the selection index involves molecular data, RIndSel presents an additional molecular variance–covariance matrix, which contains the additive variability associated with the markers (Table 11.2).

RIndSel also presents a table with the estimated values of the index parameters (Table 11.3). These estimates are the covariance of the selection index, the variance of the selection index, the net genetic merit (breeding value), the correlation between the selection index and the net genetic merit, the selection response, and the heritability of the index (see Chap. 2 for additional details).

Additional results are presented in Table 11.4, which show the ranked selected individuals; this ranking was done as a function of the estimated selection index values. Table 11.4 also presents the means of the traits of the selected individuals; the means of the traits of the total population; the selection differential (see Chap. 2),

**Table 11.1** Matrices of variance–covariance deployed by RIndSel

<i>rownames</i>	T1	T2	T3
Genetic covariance matrix			
T1	36.21	–12.93	8.35
T2	–12.93	13.04	–3.40
T3	8.35	–3.40	9.96
Phenotypic covariance matrix			
T1	62.50	–12.74	8.53
T2	–12.74	17.52	–3.38
T3	8.53	–3.38	12.31

**Table 11.2** Molecular covariance matrix

<i>rownames</i>	T1	T2	T3
T1	62.50	–12.74	8.53
T2	–12.74	17.52	–3.38
T3	8.53	–3.38	12.31

**Table 11.3** Estimated selection index parameters given by the RIndSel output

Parameter	Output
Covariance between the selection index and the breeding value	86.7185
Variance of the selection index	86.7185
Variance of the breeding value	108.5746
Correlation between the selection index and the breeding value	0.8937
Response to selection	16.3431
Heritability	0.8168

**Table 11.4** Values of the three traits for selected individuals and the values of the Smith linear phenotypic selection index, means and gains with  $k = 5\%$ 

<i>rownames</i>	T1	T2	T3	Index
Entry 353	189.68	38.16	36.13	103.97
Entry 370	178.27	34.38	37.79	103.45
Entry 480	174.84	42.72	45.12	100.66
Entry 300	177.38	39.15	40.34	100.65
Entry 273	181.18	35.94	35.14	100.52
Entry 275	167.94	36.82	42.20	99.92
Entry 148	173.37	37.07	39.62	99.86
Entry 137	185.48	46.48	42.55	99.77
Entry 351	173.79	38.38	40.52	99.68
Entry 236	182.85	37.88	34.96	99.20
Entry 217	175.13	38.48	39.16	98.84
Entry 356	171.09	39.60	41.98	98.47
Entry 167	175.39	38.73	37.73	97.17
Entry 230	169.73	37.10	38.69	96.80
Entry 243	171.90	41.53	41.45	96.29
Entry 55	170.02	36.92	37.76	96.15
Entry 68	172.56	37.18	36.70	96.13
Entry 36	175.80	38.86	36.34	95.75
Entry 164	173.61	38.37	36.42	95.14
Entry 140	170.53	42.52	41.97	95.05
Entry 146	177.40	39.64	35.50	94.89
Entry 432	174.01	40.73	38.26	94.84
Entry 378	176.62	42.69	38.47	94.44
Entry 288	172.14	39.37	37.26	94.23
Entry 386	175.77	42.89	38.81	94.13
Mean of selected individuals	175.46	39.26	38.83	
Mean of all individuals	161.88	45.19	34.39	
Selection differential	13.58	-5.92	4.44	
Expected genetic gain 5%	9.51	-5.48	4.22	

**Table 11.5** First 20 values of the entries and their corresponding selection index for all individuals when three traits are analyzed

<i>rownames</i>	T1	T2	T3	Index
Entry 1	164.46	39.63	34.66	86.81
Entry 2	144.39	50.77	34.65	63.82
Entry 3	157.48	48.04	37.90	77.52
Entry 4	167.30	47.98	30.49	74.97
Entry 5	164.11	49.89	32.03	72.85
Entry 6	166.26	40.44	29.93	81.81
Entry 7	154.59	52.22	30.31	63.22
Entry 8	160.00	42.91	31.23	77.12
Entry 9	158.51	46.32	34.52	76.25
Entry 10	163.63	45.43	35.73	81.35
Entry 11	156.16	46.75	35.58	75.62
Entry 12	171.38	41.17	35.13	89.52
Entry 13	153.17	54.18	36.23	66.79
Entry 14	149.89	52.33	31.13	61.39
Entry 15	159.63	49.01	31.72	70.96
Entry 16	160.70	42.51	32.99	79.85
Entry 17	157.07	45.49	28.40	69.68
Entry 18	167.50	41.69	36.73	88.55
Entry 19	159.17	50.60	36.25	73.93
Entry 20	161.80	46.58	37.33	80.85

and the expected genetic gain per trait. Selected individuals can be identified by the first column called “*rownames*,” as columns 2 to 4 contain the best linear and unbiased estimator for each mean trait. Finally, column 5 presents the estimated selection index values.

Comparison between means of selected individuals and all individuals is done by selection differential, where in general traits whose economic weight was 1 are positive, whereas those traits whose economic weight was  $-1$  are negative. The expected genetic gain is an inferential tool based on normal distribution that depends on the percentage of selected individuals and gives the estimated index expected genetic gain per trait.

Finally, Table 11.5 shows the best linear and unbiased estimators for all individuals accompanied by its respective selection index. In this case, only the first 20 individuals were included. This table output is important, because on some occasions, it is necessary to determine the specific behavior of a group of genotypes that may not have a good performance, even though they have shown a good general performance from previous analyses. Another possibility is that a group of individuals belongs to a specific population group; thus, it is possible to select the best individual for this population group.

## References

- Cerón-Rojas JJ, Sahagún-Castellanos J, Castillo-González F, Santacruz-Varela A, Crossa J (2008a) A restricted selection index method based on eigenanalysis. *J Agric Biol Environ Stat* 13(4):421–438
- Cerón-Rojas JJ, Sahagún-Castellanos J, Castillo-González F, Santacruz-Varela A, Benítez-Riquelme I, Crossa J (2008b) A molecular selection index method based on eigenanalysis. *Genetics* 180:547–557
- Cerón-Rojas JJ, Crossa J, Arief VN, Basford K, Rutkoski J, Jarquín D, Alvarado G, Beyene Y, Semagn K, DeLacy I (2015) A genomic selection index applied to simulated and real data. *Genes/Genomes/Genetics* 5:2155–2164
- Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* 124:743–756
- Pacheco A, Pérez S, Alvarado G, Ceron J, Rodríguez F, Crossa J, Burgueño J (2017) RIndSel: selection indices for plant breeding. hdl:11529/10854, CIMMYT Research Data & Software Repository Network, V1
- SAS Institute (2017) SAS user's guide: statistics module. Version 9.4. Ed. Cary, NC
- Smith HF (1936) A discriminant function for plant selection. In: *Papers on quantitative genetics and related topics*. Department of Genetics, North Carolina State College, Raleigh, NC, pp 466–476

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

