



Program R: applications in plant breeding

Luiz Alexandre Peternelli^{1*}

Received 2 April 2011

Accepted 21 May 2011

ABSTRACT - Nowadays the demand for so-called free, or open source software for data analysis as well as the appeal to use it is great. An public domain software that has become extremely well-known, with ever-increasing numbers of fans and even co-workers, is Environment R, or simply R. R is extremely useful for data analysis and manipulation in view of a range of tools already implemented. Also, R is not simply a statistical program, because, by its easy on using internal functions and also creating new ones, statistical procedures applied to data can also be created, manipulated, evaluated and interpreted. R contains numerous libraries (or packages), some already included in the default setting. This course will focus on the application of R in statistical analyses in plant breeding. Explanations on the use of various commands and functions will be illustrated with examples, to facilitate the interpretation and adaptation to other similar problems.

Key words: *statistical analysis, design of experiments, data analysis.*

For data analysis, software and statistical packages are of great importance, from the development and application of methods to data analysis and result interpretation. However, the purchase cost of these software packages is generally relatively high. Currently, the demand for so-called free, or open source software as well as the appeal to use it is great. A public domain software that has become extremely well-known, with ever-increasing numbers of fans and even co-workers, is Environment R, or simply R, as users call it.

The Program R is a freely available, open source code and can be changed or complemented with new procedures and functions developed by users at any time. R is extremely useful for data analysis and manipulation in view of a range of tools, e.g., parametric and non-parametric tests, linear and nonlinear modeling, time series analysis,

survival analysis, and spatial simulation and statistics, besides facilitating the drawing of various types of graphics. It can be downloaded free of charge at www.r-project.org, in pre-compiled versions for operating systems such as UNIX, Windows or Macintosh. In addition, this site provides further details about the use and a correspondence center by which professionals from different countries can contribute to the implementation of new features.

One strength of R is the ability to interact with several other programs, be they statistical or from a database. It is important to note that R is not simply a statistical program, because, by its easy on using internal functions and also creating new ones, statistical procedures applied to data can also be created, manipulated, evaluated and interpreted. The R Development Core Team (2011) classifies

¹ Universidade Federal de Viçosa, Departamento de Estatística, 36.570-000, Viçosa, MG, Brazil. *E-mail: peternelli@ufv.br

it as Environment R, due to its extensive characteristics. Here however, we discuss it as an integrated system for the execution of common statistical tasks.

In addition to the statistical procedures, R allows simple mathematical operations, manipulation of vectors and matrices and graphing. Packages or libraries are the names most often used to describe a set of functions (commands) and/or grouped data. The basic functions of R, for example, are in a library called “base”. R contains numerous libraries, some already included in the default setting. Several of them were developed by R users who, at some point considered it important to add functions that met their needs. Later, these users made these functions available in the form of a package (a library) with a certain name, so that others who need the same functions would not have to implement them again. It is this mutual collaboration that has turned R into a broad and interdisciplinary program.

The aim of this course is not merely to introduce a software or review statistical concepts and methods, but rather to provide a starting point for people who wish to start using Environment R and its statistical tools. Further

details can be found in specific books, e.g., of Peternelli and Mello (2011).

This course will focus on the application of R in statistical analyses in an attempt to address the most relevant information clearly and objectively, in several analyses. Explanations on the use of various commands and functions will be illustrated with examples, to facilitate the interpretation and adaptation to other similar problems. Initially, an introductory approach to R will show how to create, manipulate and delete the various object types and some examples involving statistical problems will be given. Reading and data entry as well as the organization of the analysis outputs will also be addressed. Finally, the use of some tools and statistical analyses will be discussed, using specific packages (eg., package ‘agricolae’ (Felipe de Mendiburu 2010) in some cases, particularly emphasizing applications in plant breeding, e.g., an analysis of the commonly used experimental designs of this area. It will also be shown how to use R to create functions for problems where no routines (or packages) are available or in cases where an automated analysis of large amounts of simulated data is desired.

Programa R: aplicações no melhoramento de plantas

RESUMO - *Atualmente a demanda por software de código aberto, chamados software livre, para a análise dos dados, é grande. Um software de domínio público que se tornou muito conhecido, com o crescente número de fãs e até mesmo colegas de trabalho, é o ambiente R, ou simplesmente R. R é extremamente útil para manipulação e análise de dados, tendo em conta uma série de ferramentas já implementadas. Além disso, R não é simplesmente um programa estatístico, porque, pela sua facilidade em usar as funções internas e também criar outras novas, procedimentos estatísticos aplicados a dados também podem ser criados, manipulados, avaliados e interpretados. R contém numerosas bibliotecas (ou pacotes), alguns já incluídos na configuração padrão. Este curso abordará sobre a aplicação de R em análises estatísticas no melhoramento de plantas. Explicações sobre o uso de vários comandos e funções serão ilustrados com exemplos, para facilitar a interpretação e adaptação a outros problemas semelhantes.*

Palavras-chave: *análises estatísticas; delineamento de experimentos, análise de dados.*

REFERENCES

Felipe de Mendiburu (2010) **Agricolae**: statistical procedures for agricultural research. R package version 1.0-9. Available at <<http://CRAN.R-project.org/package=agricolae>> Assessed in April 2011.

Peternelli LA and Mello MP (2011) **Conhecendo o R**: uma visão estatística. Editora UFV, Viçosa, 185p.

R Development Core Team (2011) **R**: a language and environment for statistical computing. R Foundation for Statistical Computing. Available at <<http://www.R-project.org/>> Assessed in April 2011.