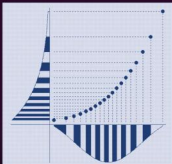


The R Series

Using R for Introductory Statistics

Second Edition



John Verzani



CRC Press
Taylor & Francis Group

A CHAPMAN & HALL BOOK

Preface

These notes are an introduction to using the statistical software package **R** for an introductory statistics course. They are meant to accompany an introductory statistics book such as Kitchens’ *Exploring Statistics*. The goals are not to show all the features of **R**, or to replace a standard textbook, but rather to be used with a textbook to illustrate the features of **R** that can be learned in a one-semester, introductory statistics course.

These notes were written to take advantage of **R** version 1.5.0 or later. For pedagogical reasons the equals sign, `=`, is used as an assignment operator and not the traditional arrow combination `<-`. This was added to **R** in version 1.4.0. If only an older version is available the reader will have to make the minor adjustment.

There are several references to data and functions in this text that need to be installed prior to their use. To install the data is easy, but the instructions vary depending on your system. For Windows users, you need to download the “zip” file, and then install from the “packages” menu. In UNIX, one uses the command `R CMD INSTALL packagename.tar.gz`. Some of the datasets are borrowed from other authors notably Kitchens. Credit is given in the help files for the datasets. This material is available as an **R** package from:

<http://www.math.csi.cuny.edu/Statistics/R/simpleR/Simple0.4.zip> for Windows users.

<http://www.math.csi.cuny.edu/Statistics/R/simpleR/Simple0.4.tar.gz> for UNIX users.

If necessary, the file can be sent in an email. As well, the individual data sets can be found online in the directory

<http://www.math.csi.cuny.edu/Statistics/R/simpleR/Simple>.

This is version 0.4 of these notes and were last generated on August 22, 2002. Before printing these notes, you should check for the most recent version available from

the CSI Math department (<http://www.math.csi.cuny.edu/Statistics/R/simpleR>).

Copyright © John Verzani (verzani@math.csi.cuny.edu), 2001-2. All rights reserved.

Contents

Introduction	1
What is R	1
A note on notation	1
Data	1
Starting R	1
Entering data with c	2
Data is a vector	3
Problems	7
Univariate Data	8
Categorical data	8
Numerical data	10
Problems	18
Bivariate Data	19
Handling bivariate categorical data	20
Handling bivariate data: categorical vs. numerical	21
Bivariate data: numerical vs. numerical	22
Linear regression.	24
Problems	31
Multivariate Data	32
Storing multivariate data in data frames	32
Accessing data in data frames	33
Manipulating data frames: stack and unstack	34
Using R ’s model formula notation	35
Ways to view multivariate data	35
The lattice package	40
Problems	40

Random Data	41
Random number generators in R– the “r” functions.	41
Problems	46
Simulations	47
The central limit theorem	47
Using <code>simple.sim</code> and functions	49
Problems	51
Exploratory Data Analysis	54
Our toolbox	54
Examples	54
Problems	58
Confidence Interval Estimation	59
Population proportion theory	59
Proportion test	61
The z -test	62
The t -test	62
Confidence interval for the median	64
Problems	65
Hypothesis Testing	66
Testing a population parameter	66
Testing a mean	67
Tests for the median	67
Problems	68
Two-sample tests	68
Two-sample tests of proportion	68
Two-sample t -tests	69
Resistant two-sample tests	71
Problems	71
Chi Square Tests	72
The chi-squared distribution	72
Chi-squared goodness of fit tests	72
Chi-squared tests of independence	74
Chi-squared tests for homogeneity	75
Problems	76
Regression Analysis	77
Simple linear regression model	77
Testing the assumptions of the model	78
Statistical inference	79
Problems	83
Multiple Linear Regression	84
The model	84
Problems	89
Analysis of Variance	89
one-way analysis of variance	89
Problems	92
Appendix: Installing R	94
Appendix: External Packages	94
Appendix: A sample R session	94
A sample session involving regression	94
t -tests	97
A simulation example	99

Appendix: What happens when R starts?	100
Appendix: Using Functions	100
The basic template	100
For loops	102
Conditional expressions	103
Appendix: Entering Data into R	103
Using c	104
using scan	104
Using scan with a file	104
Editing your data	104
Reading in tables of data	105
Fixed-width fields	105
Spreadsheet data	105
XML, urls	106
“Foreign” formats	106
Appendix: Teaching Tricks	106
Appendix: Sources of help, documentation	107