

Statistical Analyses with R
Matching Moore and McCabe (2004)

Anne Boomsma

Department of Statistics & Measurement Theory
University of Groningen

January 17, 2006

MoorR.tex

Introduction

What follows is an overview of R commands that can be used when studying and practicing material from Moore and McCabe (2004). As a reference to R commands the book of Verzani (2005) was taken; other books might have suited as well, for example Faraway (2005).

References

- Moore, D.S. & McCabe, G.P. (2004). *Introduction to the practice of statistics* (5th ed). New York: Freeman.
- Verzani, J. (2005). *Using R for introductory statistics*. Boca Raton, FL: Chapman & Hall/CRC.

The corresponding site is <http://wiener.math.csi.cuny.edu/UsingR/>. Parallel applications of R can be found at <http://cran.r-project.org/doc/contrib/Verzani-SimpleR.pdf>, under the title *SimpleR: Using R for introductory statistics*.

- Faraway, J.J. (2005). *Linear models with R*. Boca Raton, FL: Chapman & Hall/CRC.

The main part of the book can be found in PDF and postscript format at <http://www.stat.lsa.umich.edu/~faraway/book/> under the title *Practical regression and Anova in R*.

1. Looking at Data – Distributions

Table 1. Looking at Distributions.

Object	Moore & McCabe	R command	Verzani
bar graph	7	<code>barplot(·)</code> [†]	33
pie chart	7	<code>pie(·)</code>	36
stemplot	11	<code>stem(·)</code>	41
back-to-back stemplot	12		77
histograms	14	<code>hist(·)</code>	56
mean	41	<code>mean(·)</code>	43
median	42	<code>median(·)</code>	44
sorting	42	<code>sort(·)</code>	120
quantile	44	<code>quantile(·)</code>	50
five-number summary	46	<code>fivenum(·)</code>	52
summary	46	<code>summary(·)</code>	52
boxplot	46	<code>boxplot(·)</code>	64
side-by-side boxplot	46	<code>boxplot(·)</code>	77
interquartile range, IQR	48	<code>IQR(·)</code>	51
standard deviation	49	<code>sd(·)</code>	49
variance	49	<code>var(·)</code>	51
density curves	64	<code>density(·)</code>	60
normal distribution function	74	<code>pnorm(·)</code>	152
normal quantiles	78	<code>qnorm(·)</code>	152
normal quantile-quantile plot	80	<code>qqnorm(·)</code>	80

Note: † For more online information about each R command use `help(command)`,
in this case `help(barplot)`.

2. Looking at Data – Relationships

Table 2. Looking at Relationships.

Object	Moore & McCabe	R command	Verzani
scatterplot	104	plot(\cdot)	82
scatterplot smoothers	110	scatter.smooth(\cdot)	101
scatterplot smoothers	110	smooth.pline(\cdot)	101
scatterplot smoothers	110	supsmu(\cdot)	101
correlation	124	cor(\cdot)	86
least squares regression	132	lm(\cdot)	93
regression plot	133	plot(\cdot)	93
regression plot	133	res = lm(\cdot)	93
regression plot	133	abline(res)	93

2. Looking at Data – Relationships

Table 3. Extractor Functions of Linear Regression Analysis.

Object	Moore & McCabe	R command	Verzani
summary information	139	summary(\cdot)	283
diagnostic plots	153	plot(\cdot)	283
regression coefficients	137	coef(\cdot)	283
residuals	151	residuals(\cdot)	283
fitted values \hat{y}_i	137	fitted(\cdot)	283
residual sums of squares	136	deviance(\cdot)	283
performs predictions	137	predict(\cdot)	283
finds various sums of squares	139	anova(\cdot)	283
Akaike's information criterion	—	AIC(\cdot)	283

3. Producing Data

Table 4. Random Sampling from Population Distributions.

Sampling object	M & M	R command	Verzani
simple random sample (SRS)	219	<code>sample(·, replace=FALSE)</code>	146
SRS with replacement	219	<code>sample(·, replace=TRUE)</code>	146
bootstrap sample	394	<code>sample(·, replace=TRUE)</code>	176
uniform	282	<code>runif(nr[†], min=0, max=1)</code>	149
binomial	335	<code>rbinom(nr, n, prob)</code>	151*
Poisson	–	<code>rpois(nr, lambda)</code>	–
hypergeometric	–	<code>rhyper(nr, m,n, k)</code>	–
normal	69	<code>rnorm(n, mean=0, sd=1)</code>	152*
Student's <i>t</i>	450	<code>rt(nr, df)</code>	157*
Fisher's <i>F</i>	517	<code>rf(nr, df1, df2)</code>	157*
chi-square	596	<code>rchisq(nr, df)</code>	157*
exponential	364	<code>rexp(nr, rate[‡]=1)</code>	156*
lognormal	–	<code>rlnorm(nr, meanlog=0, sdlog=1)</code>	157*
geometric	358	<code>rgeom(nr, prob)</code>	–
Weibull	367	<code>rweibull(nr, shape, scale=1)</code>	–

Note: † nr: number of replications, ‡ the rate equals 1/mean, * implicitly described.

4. Probability: The Study of Randomness

The `r`-functions listed in Table 4 return random samples from a probability distribution. There are three other corresponding functions for each distribution; see Verzani (2005, p. 149) for details.

- The `d`-functions return a value of the probability mass function of the distribution: for discrete variables single probabilities, for continuous variables probability density functions.
- The `p`-functions return a value of the cumulative distribution function.
- The `q`-functions return a quantile of the probability distribution.

5. Sampling Distributions

Table 5. Topics Related to Sampling Distributions.

Topic	M & M	R command	Verzani
law of large numbers	295	–	160
binomial probabilities	338	<code>dbinom(k, n, prob)</code>	150
cumulative binomial probabilities	338	<code>pbinom(k, n, prob)</code>	150
normal approximation of binomial	343	<code>pnorm(x, mean=0, sd=1)</code>	162
binomial coefficient	349	<code>choose(n, k)</code>	151
geometric distribution	358	<code>dgeom(k, prob, ·)</code>	–
normal quantile-quantile plot	358	<code>qqnorm(·)</code>	80
central limit theorem	362	–	160
central limit theorem	362	–	168
exponential distribution	364	<code>dexp(x, rate[†]=1, ·)</code>	156*
Weibull distribution	367	<code>dweibull(x, shape, scale=1, ·)</code>	–

Note: † the rate equals 1/mean, * implicitly described.

6. Introduction to Inference

Table 6. Estimating with Confidence.

Topic	M & M	R command	Verzani
confidence intervals (CIs)	385		181
CIs for a population mean	387	<code>c(·)</code>	190
bootstrap samples	394	<code>sample(·)</code>	176
null hypothesis significance tests	400		213
<i>p</i> -values under normal law	405	<code>pnorm(z, mean=0, sd=1)</code>	216

7. Inference for Distributions

Table 7a. Inference for the Mean of a Population.

Topic	M & M	R command	Verzani
Student t quantiles	451	qt(\cdot)	157
Student t distribution function	451	pt(\cdot)	157
one-sample t confidence intervals	452	t.test(\cdot)	191
one-sample Student t test	454	t.test(\cdot)	222
matched pairs one-sample t test	459	t.test(\cdot , paired=TRUE)	241
log-transformation (base e)	466	log(x)	–
log-transformation (base 10)	466	log10(x)	–
normal quantile-quantile plot	467	qqnorm(\cdot)	80
normal quantile-quantile line	467	qqline(\cdot)	80
sign test	468	qbinom(\cdot)	207

7. Inference for Distributions

Table 7b. Comparing Two Population Means.

Topic	M & M	R command	Verzani
Student t quantiles	451	<code>qt(·)</code>	157
Student t distribution function	451	<code>pt(·)</code>	157
two-sample Student t test	488	<code>t.test(·)</code>	237
two-sample t confidence intervals	492	<code>t.test(·)</code>	240
pooled two-sample t test	499	<code>t.test(·, var.equal=TRUE)</code>	202
side-by-side boxplot	510	<code>boxplot(·)</code>	203
F test for equal variances	516	<code>var.test(·)</code>	248
Fisher's F probabilities	517	<code>pf(nr, df1, df2)</code>	157
Fisher's F quantiles	517	<code>qf(nr, df1, df2)</code>	157
power of two-sample Student t test	519	<code>power.t.test(·)</code>	—
noncentral Student t distrib. function	519	<code>pt(·, ncp=δ)</code>	157

8. Inference for Proportions

Table 8. Inferences on Proportions.

Topic	M & M	R command	Verzani
CIs [†] & hypothesis tests for proportion p	537	<code>binom.test()</code>	188
CIs & hypothesis tests for proportion p	537	<code>prop.test()</code>	184
CIs & hypothesis tests for two proportions	555	<code>binom.test()</code>	188
CIs & hypothesis tests for two proportions	555	<code>prop.test()</code>	184

Note: † CIs means confidence intervals.

9. Data Analysis of Two-Way Tables

Table 9. Data Analysis for Two-Way Tables.

Topic	M & M	R command	Verzani
make a table by combining columns	582	<code>cbind()</code>	70
make a table by combining rows	582	<code>rbind()</code>	70
marginal distributions	584	<code>margin.table()</code>	72
conditional distributions	584	<code>prop.table()</code>	74
bar graphs	586	<code>barplot()</code>	74
chi-square test	592	<code>chisq.test()</code>	258
chi-square distribution function	596	<code>pchisq()</code>	158
chi-square quantiles	596	<code>qchisq()</code>	158

10. Inference for Regression

Table 10. Simple Linear Regression.

Object	M & M	R command	Verzani
regression coefficients	639	<code>coef(·)</code>	283
residuals	639	<code>residuals(·)</code>	283
summary information	640	<code>summary(·)</code>	283
fitted values \hat{y}_i	642	<code>fitted(·)</code>	283
diagnostic plots	642	<code>plot(·)</code>	283
normal quantile-quantile plot	643	<code>qqnorm(·)</code>	80
confidence intervals (CIs) for β_0 and β_1	644	—	292*
CIs for mean response $\mu_y x$	646	<code>predict(·,int="conf")</code>	297
prediction intervals for future $Y x$	648	<code>predict(·,int="pred")</code>	295
analysis of variance	653	<code>anova(·)</code>	283
residual sums of squares	653	<code>deviance(·)</code>	283
correlation	664	<code>cor(·)</code>	86

Note: * implicitly described.

Remarks

1. The most important extractor functions for linear regression are summarized in Table 3.
2. Chapter 10 of Verzani (2005) deals with simple and multiple linear regression. More on linear models and R can be found in Faraway (2005).

Statistical Analyses with R

Copyright © 2006 by Anne Boomsma, Department of Statistics & Measurement Theory, University of Groningen, The Netherlands

Alle rechten voorbehouden. Niets in deze uitgave mag worden verveelvoudigd, opgeslagen in een geautomatiseerd gegevensbestand, en/of openbaar gemaakt, in enige vorm of op enige wijze, hetzij elektronisch, mechanisch, door fotocopie, microfilm of op enige andere manier, zonder voorafgaande schriftelijke toestemming van de auteur.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the author.