# A Handbook of Statistical Analyses Using ${\sf R}$

Brian S. Everitt and Torsten Hothorn



#### CHAPTER 16

# **Errata**

The document gives a list of typos, errors, inconsistencies etc. which have been spotted. Moreover, small numeric output differences which are due to updated packages are reported here. To get a full list of differences run R CMD check HSAUR on the source package. All issues marked with R1, R2 etc have been silently fixed in first reprint, second reprint etc.

#### **Preface**

```
Typo in name of vignette for Chapter 1, should read
R> vignette("Ch_introduction_to_R", package = "HSAUR")
and
R> edit(vignette("Ch_introduction_to_R", package = "HSAUR"))
As of version 1.0-3, only the correctly named vignette is available (R1).
```

# 16.1 Introduction to R

- $\bullet$  Type at page 11: '.' needs to be double-quoted in list.files  $(\mathbf{R1})$
- Typo at page 20 (Ex. 1.5): number of companies, not number of countries (R1).

#### 16.2 Simple Inference

Typo at page 31, code line 4: use argument varwidth = TRUE, not var.width = TRUE ( $\mathbf{R1}$ ).

## 16.3 Conditional Inference

- The names of the test statistics in the output have been changed from T to Z or chi-squared throughout the chapter (R1).
- Reference Hothorn et al. (2006a) updated (R1)

## 16.4 Analysis of Variance

Figure 4.3 had wrong line types in legend (Beef and Cereal interchanged) (**R2**).

4 ERRATA

#### 16.5 Multiple Linear Regression

• Page 74, Table 5.1: The values for cloudcover and sne had to be exchanged. The corresponding coefficients and Figures in this chapter change accordingly (R1).

- Page 83: both fitted and predict can be used to compute fitted values, the later on can be applied to new unseen data as well (R1).
- Page 87:  $\hat{y}_i$  instead of  $\hat{y}$  in the definition of the standardized residual.

# 16.6 Logistic Regression and Generalised Linear Models

- page 97: predictions are to be computed for plasma\_glm\_2, not plasma\_glm\_1 (affects Figure 6.4) (R2).
- Function myplot (page 100): the vfont argument in text has been changed to family = "HersheySerif" (the resulting plots remain the same) (R1).

#### 16.7 Density Estimation

- Page 117: typo: in instead of is
- Page 121: small numeric differences for the output of optim
- update to mclust version 3.0-0 (new names of parameters in mclust objects)

#### 16.8 Recursive Partitioning

- Page 138: the probability for glaucoma is predict(trees[[i]], newdata = GlaucomaM)[,1] and the code for converting average class probabilities in factors has to be reverted, too. Affects Figure 8.4. (which is now in sync with the interpretation).
- Page 139: small differences in predtab
- Page 140: small differences in table at bottom of this page
- Reference Hothorn et al. (2006b) updated (R1)
- Page 142, Ex. 8.1.: regression tree, not classification tree.

#### 16.9 Survival Analysis

- The name of the test statistic in the output of surv\_test has been changed to chi-squared (R1).
- Denominator s was missing from h(t) (page 147) (**R2**).

#### 16.10 Analysing Longitudinal Data I

Page 168, Figure 10.2: summary does not provide degrees of freedom and p-values in newer versions of *lme4*.

#### 16.11 Analysing Longitudinal Data II

-nothing known-

# 16.12 Meta-Analysis

- Page 202:  $\mu_i \sim \mathcal{N}(\mu, \tau^2)$ , not  $N(\mu, \tau^2)$  (**R2**).
- Page 202:  $W_i = 1/(V_i + \hat{\tau}^2)$  since  $V_i$  is the within-study variance.
- Page 207: square for selogs was missing (R2).

#### 16.13 Principal Component Analysis

-nothing known-

# 16.14 Multidimensional Scaling

In the formula for  $b_{ij}$  on page 231 the last term in the parentheses should have a plus sign not a minus sign.

#### 16.15 Cluster Analysis

- update to mclust version 3.0-0 (new plot method)
- Page 248: the likelihood needs  $|\Sigma_i|^{-1/2}$
- Page 248:  $W_j$  is a  $q \times q$  matrix
- Page 248:  $\Sigma_j = \Sigma, j = 1, ..., c$ .
- Page 248:

$$l(\vartheta, \gamma) = -\frac{1}{2} \sum_{j=1}^{c} \operatorname{trace}(\mathbf{W}_{j} \Sigma_{j}^{-1}) + n_{j} \log |\Sigma_{j}|$$

• Page 248:  $\sum_{j=1}^{c} n_j \log |\mathbf{W}_j/n_j|$ 

#### Thanks

We would like to thank the following people for pointing out errors, typos etc and for making suggestions for improvements:

- Tobias Verbeke
- Brian D. Ripley
- Yihui Xie
- Mike Cheung
- Lothar Häberle
- Radoslav Harman



# Bibliography

Hothorn, T., Hornik, K., van de Wiel, M. A., and Zeileis, A. (2006a), "A Lego system for conditional inference," *The American Statistician*, 60, 257-263, URL http://statmath.wu-wien.ac.at/~zeileis/papers/Hothorn+Hornik+VanDeWiel-2006.pdf.

Hothorn, T., Hornik, K., and Zeileis, A. (2006b), "Unbiased recursive partitioning: A conditional inference framework," *Journal of Computational and Graphical Statistics*, 15, 651–674, URL http://statmath.wu-wien.ac.at/~zeileis/papers/Hothorn+Hornik+Zeileis-2006.pdf.