

## Grundlagen der Datenanalyse mit R, 1. Auflage – Errata

Stand: 23. April 2012

### Inhaltlich relevante Korrekturen

- Abschn. 1.2.1, S. 9:  
“Wird der Anfang eines Befehlsnamens in die Konsole eingegeben und dann die Tabulator-Taste gedrückt, vervollständigt R den begonnenen Befehl. Gibt es dafür mehrere Kandidaten, zeigt R diejenigen Funktionen an, die den begonnenen Befehl vervollständigen könnten, nachdem zweimal Tab gedrückt wurde.”
- Abschn. 1.2.7, S. 15:  
“Zusätzlich sollte abweichend von der Voreinstellung `dependencies=TRUE` gesetzt werden, damit vom zu installierenden Paket ggf. benötigte andere Pakete automatisch ebenfalls mit installiert werden.”  
→ Zwingend benötigte Pakete werden bereits in der Voreinstellung `dependencies=NA` automatisch mit installiert. Durch `dependencies=TRUE` geschieht dies auch für Pakete, deren gemeinsame Verwendung empfohlen ist.
- Abschn. 1.2.7, S. 16, Fußnote 19:  
“Dafür muss eine Textdatei `Renviron.site` im Unterordner `etc/` des R-Programmordners existieren und eine Zeile der Form `R_LIBS=<Pfad>` (z. B. `R_LIBS="c:/rlibs/"`) mit dem Pfad zu den Paketen enthalten.”
- Abschn. 2.9.2, Fußnote 41, S. 81:  
“Für nicht invertierbare ( $p \times p$ )-Matrizen  $A$  ( $\text{Rang}(A) < p$ ) ermittelt die `Null()` Funktion aus dem Paket `MASS` eine Basis des Kerns von  $A^t$  (engl. Null Space bzw. Kernel).”
- Abschn. 2.9.3, S. 84:  
“> `(idxMin <- which.min(mDist))`”
- Abschn. 2.9.5, S. 89:  
“Insgesamt gilt damit  $X = u \cdot D \cdot v^t$ , wenn  $D$  die aus Singulärwerten gebildete Diagonalmatrix ist.”  
“`Xsvd$u` `%%Xsvd$d` `%% t(Xsvd$v)`”
- Abschn. 3.2.7, S. 134:  
“> `res$newIV <- factor(sample(rep(c("A", "B"), c(4, 4)), 8, replace=FALSE))`”
- Abschn. 6.3.1, S. 189:  
Spearman's  $\rho$  und Kendall's  $\tau$  stehen im Abschnitt “6.3 Maße für Zusammenhang und Übereinstimmung kategorialer Daten”, setzen jedoch stetige Variablen voraus.
- Abschn. 6.4, S. 199:  
“Die im folgenden aufgeführten Tests lassen sich mit Werten geordneter kategorialer Variablen, also etwa mit Rangdaten durchführen [...].”  
→ Die aufgeführten Tests setzen stetige Variablen voraus, damit sich Rangbindungen (mehrfach auftretende Werte) nur mit einer Wahrscheinlichkeit von 0 ergeben.

- Abschn. 6.4.1, S. 200:  
“(pLess <- 1 - pbinom(N-obs - 1, N, 0.5)) # linksseitig”  
“[1] 0.9940918 0.02069473”  
“> (pTwoSided <- 2 \* (1 - pbinom(max(obs, N-obs) - 1, N, 0.5)) min(pLess, pGreater))”
- Abschn. 7.2.1, S. 223:  
“> orgPred <- predict(model, interval="prediction confidence", level=0.95)”
- Abschn. 7.4, S. 234:  
“Die AV muss ein Objekt der Klasse factor mit zwei Stufen sein, die Auftretenswahrscheinlichkeit  $P$  bezieht sich dann auf jene Ausprägung der AV, die als erste zweite Faktorstufe definiert ist.”
- Abschn. 8.5.2, S. 274:  
“effNj <- mdPQ / apply(1/cellNs, 1, sum)”  
“effNk <- mdQP / apply(1/cellNs, 2, sum)”
- Abschn. 9.6.1, S. 331:  
“Roys Maximalwurzel: entweder der größte Eigenwert  $\theta_1$  von  $T^{-1} \cdot B$  (so definiert in  $\mathbb{R}$ ) oder der größte Eigenwert  $\lambda_1$  von  $W^{-1} \cdot B$  (so definiert in  $\mathbb{R}$ ).”
- Abschn. 11.4.1, S. 404:  
“So lässt sich mit curve() zunächst keine Dichtefunktion einer normalverteilten Variable mit Erwartungswert 100 darstellen, da an curve() nur die Funktion dnorm, nicht aber deren Argument mean übergeben werden kann.”  
→ Es ist möglich, auch vollständige Funktionsaufrufe inkl. der Argumente an curve() zu übergeben, etwa curve(dnorm(x, mean=10, sd=1)).
- Literatur, S. 413:  
“Backhaus, K., Erichson, B., Plinke, W. & Rolf, W. Weiber, R. (2008). [...]”

## Weitere Hinweise

- Abschn. 2.6.3, S. 53:  
“Den Modalwert erhält man zusammen mit seiner Auftretenshäufigkeit durch Indizieren der Tabelle mit diesem Index.”  
→ Allerdings ist der Modalwert selbst nicht das von tab[modIdx] zurückgegebene Element, sondern ist nur der zur genannten Häufigkeit gehörende Name. Er ist damit zwar in der Ausgabe abzulesen, aber nicht in weiteren Rechnungen zu verwenden. Günstiger ist es daher, den Modalwert mit unique(vec)[tabIdx] ausgeben zu lassen.
- Abschn. 2.11.4, S. 95:  
“Um relative Häufigkeiten zu ermitteln, eignet sich auch in Kreuztabellen die prop.table() Funktion. Für bedingte relative Häufigkeiten kann sie mit sweep() kombiniert werden.”  
→ Einfacher ist es, für bedingte relative Häufigkeiten die Funktion prop.table() selbst zu verwenden: sie besitzt ein zweites Argument margin=(Nummer), an das 1 für auf die Zeilen, und 2 für auf die Spalten bedingte relative Häufigkeiten übergeben werden kann.
- Abschn. 8.7.4, S. 289:  
“Auch hier stellt sich jedoch die Frage, ob Einzelvergleiche sinnvollerweise durchgeführt werden sollten [...]”  
→ Diese Einschränkung gilt nur für Kontraste, die sich auf Zellvergleiche oder auf mittlere

Erwartungswerte des Intra-Gruppen Faktors (UV2) beziehen. *A*-Kontraste lassen sich wie Kontraste im CR-*p* Design testen, nachdem zu blockweise zentrierten Daten übergegangen wurde.

- Abschn. 8.8.1, S. 294:  
“Zur Berechnung von Quadratsummen vom Typ III kann zum einen auf die `Anova()` Funktion aus dem `car` Paket zurückgegriffen werden.”  
→ Hier wird keine Interaktion berücksichtigt. Daher ist die Quadratsumme vom Typ III gleichzeitig jene vom Typ II. Aus diesem Grund ist hier anders als in Abschn. 8.5.2 (S. 273) auch kein Übergang zur Effektcodierung notwendig.
- Abschn. 9.3, S. 318:  
“Mit der Voreinstellung `"none"` für das Argument `rotation` liegt der Rechnung das Modell unkorrelierter Faktoren zugrunde.”  
→ Mittlerweile ist die Voreinstellung `rotation="varimax"`. Für die Reproduktion der Ergebnisse muss daher explizit `rotation="none"` gesetzt werden. Dies gilt insbesondere für die auf S. 320 angesprochene Gleichheit der aufgeklärten Varianz eines Faktors mit dem zugehörigen Eigenwert der geschätzten reduzierten Korrelationsmatrix.
- Abschn. 9.6.2, S. 332:  
→ Auch im multivariaten Fall ist zu beachten, daß R mit Quadratsummen vom Typ I rechnet, die bei ungleichen Zellbesetzungen (wie im gewählten Beispiel) andere Ergebnisse liefern als etwa solche vom Typ III.

## Tipp- und Druckfehler

- Abschn. 1.1.1, S. 1:  
“ehe eine Teilmenge von Beobachtungen ausgewählt und mit ihr ein statistischer [...]”
- Abschn. 1.2.1, S. 7:  
“Ein Beruf kann dabei `dabei` zur Gruppe `bc` (Blue Collar), `wc` (White Collar) [...]”
- Abschn. 2.2.3, S. 39:  
“> `IQcls[(IQ > 100) & (IQ <= 115)] <- 2`”
- Abschn. 2.3.3, S. 43:  
“[...] dies entspricht dem Binomialkoeffizienten  $\binom{n}{k}$ .”  
“[...] man im Englischen beim Binomialkoeffizienten von „*n* choose *k*“, woraus [...]”
- Abschn. 2.5.1, S. 47:  
“> `sample(1:6, 20, replace=TRUE)`”
- Abschn. 2.6.5, S. 54:  
“> `sum((age-mean(age))^2) / (length(age)-1)`”
- Abschn. 2.6.5, S. 55:  
“> `sum((age-mean(age))^2) / length(age)`”  
“> `sqrt(sum((age-mean(age))^2) / length(age))`”  
“[...] lassen sich über die Funktionen `skewness()` und `kurtosis()` aus dem [...]”
- Abschn. 2.11.1, S. 92:  
“Für diese Rechnung existiert die Funktion `prop.table()`, welche [...]”

- Abschn. 2.11.6, S. 98:  
“[...] über `plot()` direkt graphisch abbilden (Abb. 6.1 und 10.21, vgl. Abschn. 10.6.6).”
- Abschn. 2.14.1, S. 114:  
“[...] sich jedoch mit `format(x=(Date-Objekt), format="⟨Format-String⟩")` [...]”
- Abschn. 3.1.1, S. 118:  
“`> (myList2 <- list(c(1:4), matrix(1:4, 2, 2), c("Lorem", "ipsum")))`”
- Abschn. 3.1.1, S. 119:  
“`> (myList3 <- list(c(1:5), c(8:4), word="dolor"))`”
- Abschn. 3.2.11, S. 143:  
“# range der numerischen Variablen”
- Abschn. 3.2.12, S. 146:  
“`> myDf2 <- data.frame(cbind(x1, y1))`”  
“`> myDf3 <- data.frame(cbind(x2, y2))`”
- Abschn. 4, S. 151:  
“Für viele Befehle ist die Pfadangabe zu einer Datei notwendig, die es zu [...]”
- Abschn. 4.2.5.1, S. 158  
“[...] das Dateiformat `xlsx` aus Excel 2007 derzeit aber nicht unterstützt.”
- Abschn. 6.1.3, S. 178:  
“[...] die Abwandlung mit Lilliefors-Schranken sowie mit `ad.test()` den [...]”
- Abschn. 6.1.5, S. 181:  
“data: `rbind(dataFreq, freqExp)`”
- Abschn. 6.4.9, S. 210:  
“`> symmetry_test(formula=(Formel), data=(Datensatz), subset=NULL,`”
- Abschn. 6.4.10, S. 212:  
“`> categ <- c(1:3)`”
- Abschn. 7.2.3, Legende Abb. 7.3, S. 226:  
“Vorhersage”
- Abschn. 7.3, S. 229:  
“[...], d. h. alle  $f$   $p$  Prädiktoren werden mit `+` verbunden hinter die `~` geschrieben.”
- Abschn. 7.4, S. 235:  
“[...] die Deviance, also die Summe der quadrierten Residuen<sup>20</sup> und dem Wert [...]”
- Abschn. 8.3.6.1, S. 255:  
“[...] wird festgelegt, ob die  $H_1$  gerichtet oder ungerichtet (`"two.sided"`) ist.”
- Abschn. 8.5.1, S. 269:  
“[...] assoziierten einfaktoriellen Varianzanalyse (vgl. auch `?interaction` `?interaction`).”
- Abschn. 8.6.2, S. 283:  
“[...] (andere analog), `SS` für Quadratsumme und `df` für ihre Freiheitsgrade steht.”
- Abschn. 8.8.2, S. 297:  
“[...] wobei wie oben das  $b$ -Gewicht konstant sein und nur die Variation des [...]”

- Abschn. 9.3, S. 317:  
“Im Modell unkorrelierter Faktoren vereinfacht sich die Gleichung zu  $K_x = \Lambda \cdot \Lambda^t + D_e$ .”
- Abschn. 9.3, S. 319:  
“[...] nur Werte über einer bestimmten absoluten Größe angezeigt werden (Abb. 9.2).”
- Abschn. 9.3, S. 321:  
“[...] um einen besonders markanten Sprung zu identifizieren (Scree-Test).”
- Abschn. 9.4, S. 325:  
“[...] versucht nicht, die euklidischen Distanzen in exakte Übereinstimmung [...]”
- Abschn. 9.5.2, S. 328:  
“Die Ausgabe gleicht der aus der univariaten Anwendung von `anova()`, [...]”
- Abschn. 9.6.1, S. 331:  
“Für die Umrechnung von  $\theta_1$  in  $\lambda_1$  gilt [...]”
- Abschn. 10.3.1, Tab. 10.2, S. 344:  
“TRUE: gesamte Figure-Region steht zur Verfügung, NA: gesamte Device-Region [...]”
- Abschn. 10.3.1, S. 345:  
“Symbole 21–25 sind ausgefüllte Datenpunkte, deren Füllfarbe in Zeichenfunktionen [...]”
- Abschn. 10.7.1, S. 382:  
“Das Argument `f` von `lowess()` gibt ebenso wie das Argument `span` von `loess.smooth()` die Spannweite des Glätters [...]”
- Abschn. 10.8.2, S. 388, Fußnote 24:  
“[...] , die Graphiken interaktiv in Echtzeit verändern, sie etwa an geänderte Ursprungsdaten anpassen.”
- Abschn. 10.8.4, S. 393:  
“`ellCG <- t(t(RRcg) %*% rbind(cos(angles), sin(angles)))`”  
“`ellT <- t(t(RRt) %*% rbind(cos(angles), sin(angles)))`”
- Abschn. 10.9.3, S. 399:  
“Um eine mit `split.screen()` erzeugte Teilfläche abzuschließen und damit ein weiteres Hinzufügen von Elementen zu verhindern, [...]”  
“> `screen(all.screens=TRUE) # Bearbeitung abschließen`”
- Abschn. 11, S. 401:  
“[...] sondern ist gleichzeitig eine Programmiersprache, die dieselbe Syntax wie [...]”
- Abschn. 11.1.2, S. 403:  
“[...] nach dem der Inhalt von Argumenten erst mit ihrer ersten Verwendung [...]”
- Literaturverzeichnis, S. 414:  
“Friedrichsmeier, T., [...] (2009). RKWard Data Analysis Tool [...]”