

## Grundlagen der Datenanalyse mit R, 4. Auflage – Errata

Stand: 11. Juni 2020

### Inhaltlich relevante Korrekturen

- Abschn. 2.7.3, S. 72:  
“Den Modalwert erhält man zusammen mit seiner Auftretenshäufigkeit durch Indizieren der mit `unique()` gebildeten **und dann geordneten** Einzelwerte des Vektors mit diesem Index.  
> `sort(unique(vec))[modIdx]`”
- Abschn. 4.2.5, S. 176:  
“Das Argument `what` benötigt eine Angabe der Form `"logical" logical()`, `"numeric" numeric()` oder `"character" character()`, die Auskunft über den Datentyp der folgenden Werte gibt. Zeichenketten müssen durch die Angabe `"character" character()` nicht mehr in Anführungszeichen eingegeben werden,  
> `charVec <- scan(what="character" character())`”
- Abschn. 6.6.3, S. 225:  
“> `elNet <- glmnet(matScl[ , c("height", "age", "sport")],  
+ matScl[ , "weight"],  
+ lambda=lassoCVelNetCV$lambda.min, alpha=0.5)`”
- Abschn. 9.4.3, S. 367:  
“Die **Survival-Funktionen** **logarithmierten kumulativen hazards** sollten im Diagramm linear mit  $\ln t_i$  ansteigen und zudem parallel verlaufen.”
- Abschn. 10.5.4, S. 418:  
“Die diskret verteilte Teststatistik ergibt sich, indem von dieser Rangsumme ihr theoretisches Minimum  $\sum_{i=1}^{n_1} i = \frac{n_1(n_1+1)}{2}$  abgezogen wird.”
- Abschn. 13.3, S. 535:  
“`on.exit(paroptions(op))` # bei Verlassen: Option zurücksetzen”

### Weitere Hinweise

- Abschn. 1.1.4, S. 5:  
OpenAnalytics Architect ist jetzt unter folgender URL erhältlich – allerdings ist seit Längerem keine neue Version erschienen: <https://www.getarchitect.io/>  
Auf der IDE Eclipse basiert auch die aktiv entwickelte Umgebung Bio7: <https://bio7.org/>  
Microsoft hat die Entwicklung der *R Tools for Visual Studio* offenbar eingestellt.
- Abschn. 2.7.5, S. 73; Abschn. 2.10.6, S. 102:  
Für die Berechnung des Prozentrangs enthält das Paket `DescTools()` jetzt die Funktion `PercentRank()`.
- Abschn. 7.1, S. 231:  
“Um die power eines Tests zu vergrößern, wird mitunter ein höher als übliches  $\alpha$ -Niveau in der Größenordnung von 0.2 gewählt.”  
Dieser Ansatz ist sehr informell. Statistisch angemessen lassen sich Anpassungstests im Kontext

von Tests auf Äquivalenz und Nichtunterlegenheit behandeln. Sie dazu (insbesondere Kap. 9):

Wellek, S. (2010). *Statistical Hypotheses of Equivalence and Noninferiority*, 2. Aufl. Boca Raton, FL: Chapman Hall/CRC.

- Abschn. 8.2.2, S. 323f.; Abschn. 8.3.2, S. 330:  
Verschiedene Pseudo- $R^2$  Werte für ordinale oder multinomiale Regressionsmodelle, die mit der Funktion `vglm()` aus dem Paket `VGAM` angepasst wurden, lassen sich nun auch direkt mit `PseudoR2(<vglm-Modell>)` aus dem Paket `DescTools` ermitteln.
- Abschn. 8.2.3, S. 325; Abschn. 8.3.3, S. 331:  
Für ordinale oder multinomiale Regressionsmodelle, die mit `vglm()` aus dem Paket `VGAM` angepasst wurden, lassen sich nun über `confint(<vglm-Modell>, method=profile)` Konfidenzintervalle mit der Profile-Likelihood Methode berechnen.
- Abschn. 12.9.2, S. 502, Fußnote 55:  
Für die gewichtete Effektcodierung stellt das Paket `wec` die Funktion `contr.wec()` bereit. Siehe auch:  
Nieuwenhuis R, te Grotenhuis M, Pelzer B. Weighted Effect Coding for Observational Data with `wec`. *The R Journal* 2017; 9: 477–485.

## Tipp- und Druckfehler

- Abschn. 1.1.3, S. 4:  
“C:\ProgrammeR\R-<Version>”
- Abschn. 1.4.3, S. 25:  
“nur als Zeichenkette in einem anderen Objekts gespeichert ist”
- Abschn. 2.3.3, S. 45:  
“Für eine einzelne zufällige Permutation s. Abschn. 2.4.3”
- Abschn. 2.5.1, S. 51:  
“gibt eine sortierte Version des Vektors aus”
- Abschn. 2.10.6, S. 101:  
“die relative Häufigkeit `1/length(x)` für jedes Elements von `x`”
- Abschn. 2.11.3, S. 106:  
“> `apply(matNA, 1, FUN=mean, FUN=na.rm=TRUE)`”
- Abschn. 2.12.4, S. 115:  
“existieren die Funktionen `startsWith("<Z>")`, `prefix="<Muster>"`) und `endsWith("<Z>", suffix="<Muster>")`”
- Abschn. 2.12.5, S. 117:  
“sorgt u. U. schon die Suche nach einfachen Mustern für Schwierigkeiten.”
- Abschn. 3.3.9, S. 154:  
“in einen einzigen Faktor zu umzuwandeln”
- Abschn. 4.2.4, S. 169:  
“ist deshalb mitunter für die Entwickler der entsprechenden R-Funktionen nicht sicher zu ermitteln.”

- Abschn. 4.2.4, S. 172:  
“um Daten zwischen **der** Datenbank und R auszutauschen”
- Abschn. 4.4, S. 182:  
“Einfacher**er** ist es jedoch, die sehr gute Unterstützung”
- Abschn. 7.2.2, S. 237:  
“Bei**m** Verwendung einer Modellformel”
- Abschn. 8.1.6, S. 319:  
“Analog sollten Modellvergleich**e** mit einem  $F$ -Test”
- Abschn. 8.2.3, S. 326:  
“> lrtest(vglmNP, vglm**NP**) # Likelihood-Quotienten-Test”
- Abschn. 8.4.4, S. 339:  
“Separate Wald**td**-Tests der Parameter”
- Abschn. 9.5.3, S. 376:  
“Die Ausgabe ist weitgehend analog zu jener von `summary(<coxph-Objekt>)`**)**”
- Abschn. 11.1.1, S. 431:  
“Als Beispiel diene die **die** Situation eines  $t$ -Tests”
- Abschn. 11.3.1, S. 446:  
“Wähl**t** man für  $\hat{\theta}$  die Teststatistik”
- Abschn. 12.2.2, S. 472:  
“befinden sich jedoch in der Komponente `loadings` der zurückgelieferten Liste **enthalten**.”
- Abschn. 14.1.1, S. 538:  
“während die Inhalte der anderen `device`**s** unverändert bleiben”
- Abschn. 14.7, S. 590:  
“So **können** Messwerte in ihrer Ausprägung von mehr als einer numerischen Variable abhängen”
- Abschn. 14.7.4, S. 597:  
“können auf diese Weise etwa die **die** Einträge der Korrelationsmatrix”
- Abschn. 15.5.2, S. 621:  
“`theme(axis.ticks=element_blank())`**)**”
- Abschn. 15.5.3, S. 624:  
“keine Einträge zur Art der verwendeten Datenpunktsymbole**n** enthält.”