

Errata zu: Parametrische Statistik

–

Verteilungen, *maximum likelihood* und GLM in R

Carsten F. Dormann

Biometrie & Umweltsystemanalyse
Universität Freiburg

3. März 2016

1 Gemeinheiten

Titanic Der Datensatz `Titanic` im α Paket **effects** wurde in `TitanicSurvival` umbenannt.¹
Das führt zu einigen Änderungen im Buch (Vielen Dank, Severin Hauenstein!):

6.3 Seite 100 Richtig: `data(TitanicSurvival)`

6.3 Seite 102 Richtig: `plot(TitanicSurvival$..., TitanicSurvival$...)`
In drei Zeilen ändern!

12.4 Seite 221 Richtig:

```
data(TitanicSurvival)
fm.tita <- glm(survived ~ passengerClass, data=TitanicSurvival, family=binomial)
```

Daumenregel Da ist mir ein Anglizismus unterlaufen: die englische *rule of thumb* ist auf Deutsch natürlich eine „Faustregel“. Vielen Dank an Jan Hanspach! Seite 189, 237, 280 und 315.

2 Inhaltliche Fehler

3.2 Exkurs: Bedingte Wahrscheinlichkeit Seite 42 Der 2. Satz nach der untersten Formel sollte *richtig* heißen:

Wenn B eine Teilmenge von A ist, dann *muss* A eintreten, wenn B eingetreten ist:
 $P(A|B) = 1$.

Vielen Dank an Jana Jaeck fürs Finden!

3.5 S. 62, letzter Absatz Es sollte *richtig* heißen „(Abb. 3.6, hellgraue Linie)“ (nicht: dunkelgraue Linie). Vielen Dank an Jan Hanspach!

3.5.2 S. 64 Die korrekte Formel für $AICc$ ist: $AICc = AIC + \frac{2p(p+1)}{n-p-1} = -2LL + 2p + \frac{2p(p+1)}{n-p-1}$

¹Vermutlich, weil es den ganz anders aufgebauten Datensatz `Titanic` bereits in **datasets** gab und das zu Verwirrung führte.

7.1.2 S. 109 Hier haben sich ein paar Notationsunsauberkeiten eingeschlichen, sowohl in der Gleichung, als auch im nachfolgenden Text (a und b sind dort anders definiert als zuvor). Es sollte *richtig* sein:

„Die Wahrscheinlichkeitsdichte jedes beobachteten Werts y_i ist abhängig von $\lambda = ax_i + b$. Unter der Annahme, dass die Daten Poisson-verteilt sind setzen wir für einen Wert y_i einfach in Gleichung 3.10 auf Seite 50 ein:

$$L(y_i) = \frac{\lambda^{y_i}}{y_i! e^\lambda} = \frac{(ax_i + b)^{y_i}}{y_i! e^{(ax_i + b)}}$$

Nehmen wir unseren ersten Schnäpperdatenpunkt: Attraktivität = 1, Stücke = 3. Für angenommene Werte von $a = 1$ und $b = 4$ ergibt sich somit eine *likelihood* von

$$P(k = y_i | \lambda = a \cdot x_i + b) = P(k = 3 | \lambda = 1 \cdot 1 + 4) = \frac{(1 \cdot 1 + 4)^3}{3! e^{1 \cdot 1 + 4}} = \frac{125}{3 \cdot 148.4} = 0.140$$

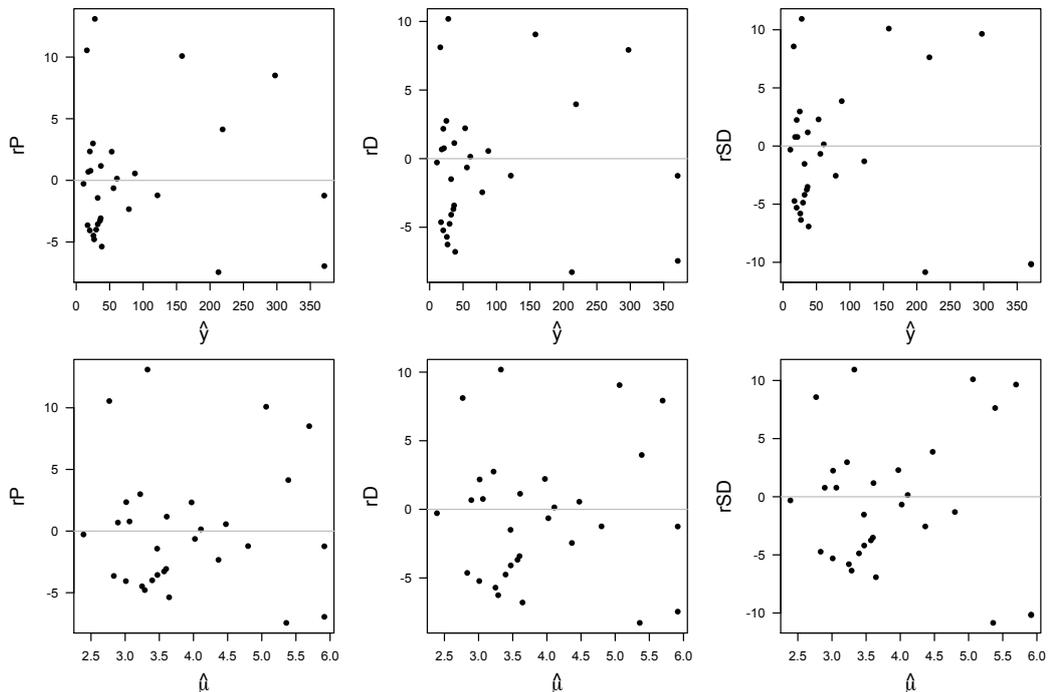
7.1.1 Seite 107 *Richtig:* $y \sim \text{Pois}(\lambda = ax + b)$

Das Zeichen λ ist inkorrekt dargestellt.

8.2.1 Seite 138 *Richtig:* „Unter dem Listeneintrag \$value erscheint der Funktionswert im Optimum, ...“.

Fälschlicherweise hatte ich dort \$counts geschrieben.

9.1.4 Seite 157 Abb. 9.3 Hier ist die Abbildung 10.2 nochmals reingerutscht, statt, wie in meiner Originalvorlage, die folgende Abbildung. Macht nicht so viel aus, sind aber einfach die falschen Daten.



10.1 Seite 167 Abb. 10.1 Rechte und linke Abbildung sind vertauscht.

10.1.1 Seite 168 *Richtig:* $88 - 31 = 57$ Einheiten (*Falsch:* $88 - 31 = 75$ Einheiten)

Vielen Dank an Patric Schlager fürs Finden!

10.1.5 Seite 180 Abb. 10.5 *Richtig:* „Daten und der Fit unseres Poisson-Modells mit quadratischem Term (fm2.2). Beachte, dass die x -Achse *links* linear, *rechts* hingegen logarithmiert dargestellt ist.“ (Abbildungsbeschriftung rechts und links waren vertauscht.)

Vielen Dank an Tina Mölter fürs Finden!

11.1.1 Seite 181, ganz unten Sollte heißen: $t = -0.843/1.65$ (Falsch: $/3.74$). Danke an Gregor Didenko!

11.1.2 Seite 190 Gleichung 11.2 sollte *richtig* heißen:

$$t = \frac{\frac{1}{n} \sum_{i=1}^n (x_{1i} - x_{2i})}{SE_{x_{1i} - x_{2i}}}$$

Leider war das $\frac{1}{n}$ verloren gegangen.

12.2.2 Seite 217 mitten im Text *Richtig*: „Abbildung 12.2 links ...“ (Falsch: rechts).

Vielen Dank an Tina Mölter fürs Finden!

11.2.1 Seite 243, mitten drin Hier ist der *output* des `tapply`-Befehls verrutscht. Er sollte eigentlich ganz oben stehen, direkt im Anschluss an den Befehl auf der Seite zuvor; etwa so:

```
> tapply(Artemisia, list(comp, nematodes), mean, na.rm = T)
```

```
      0      1
0 15.216455 14.339500
1  2.069063  2.116063
```

Danke an Gregor Didenko!

14.2.3 Seite 253 Da fehlt ein „fast“ vor „signifikant“ („Da der *exclosure*-Effekt fast signifikant ist, ist dies nicht verwunderlich.“), denn mit $p = 0.0525$ ist der Herbivorenausschluss ja gerade nicht mehr signifikant. Vielen Dank an Jan Hanspach!

14.3.2 Seite 258 Formel für die Varianz-optimale Zuordnung: im Nenner sollte als Index j stehen:

$$n_i = n \frac{A_i s_i}{\sum_{j=1}^k A_j s_j}$$

(Mit dem i ist es nicht wirklich falsch, aber verwirrend, denn nur im Nenner wird ja über alle Strata summiert, während der Wert in Zähler konstant je Stratum ist.)

14.3.2 Seite 259, Fußnote Formel für die Kosten-und-Varianz-optimale Zuordnung: Im Nenner sollte $\sqrt{\kappa_j}$ als Faktor, nicht als Divisor auftauchen, obwohl er im Zähler als Divisor auftaucht:

$$n_i = n \frac{A_i s_i / \sqrt{\kappa_i}}{\sum_{j=1}^k A_j s_j \sqrt{\kappa_j}}$$

Ursprung dieser

16.1.2 Seite 305 Abb. 16.5 Beschreibung Dreiecke sind ohne, Kreise mit Glukose! (Genau falsch herum beschriftet.)

16.1.2 Seite 306 *Richtig*:

$$\begin{aligned} & \text{Intercept} + \text{est}_{\text{glucoseNein}} + (\text{est}_{\log_{10}(\text{conc})} + \text{est}_{\log_{10}(\text{conc}):\text{glucoseNein}}) \log_{10}(\mathbf{conc}) \\ & = 37.6 - 1.1 + (-2.9 + 0.7) \log_{10}(\mathbf{conc}) \end{aligned}$$

3 Tippfehler

- 3.3.2 S. 48** „LL($(\mu = 40, \sigma = 30)|\text{Umfänge}$) = -457.97 “ (nicht andersrum) Danke an Anne-Christine Mupepele!
- 3.4.1 Seite 53, 1. Satz** „kontinuierliche Verteilung“. Danke an Gregor Didenko!
- 3.5 Seite 61, 2. Absatz von unten:** „lebendgebärend“ (nicht: legendgebärend ...). Danke an Jan Hanspach!
- 3.5 Seite 62, 3. Absatz von oben:** „Anpassungsgüte“ (nicht: Anpassunggüte ...). Danke an Jan Hanspach!
- 3.5 Seite 62, Fußnote 14:** „Die in der Vegetationskunde übliche arc-Sinus-Wurzeltransformation für Vegetationsaufnahmen“ (nicht: üblich ...) Danke an Jan Hanspach!
- 5.1 Seite 88, 2. Absatz** „einem Standardfehler“. Danke an Gregor Didenko!
- 5.1.2 Seite 89, 2. Abschnitt von unten:** „Von allen ist der χ^2 -Test der häufigste und einfachste, und er ist im nächsten Abschnitt erklärt.“ (nicht: Abschnit ...) Danke an Jan Hanspach!
- 5.2 Seite 91, 1. Satz; 15.5 Seite 290; 16.1.2 Seite. 303:** „kategorial“ (nicht: kategorisch ...) Danke an Jan Hanspach!
- 5.2 Seite 92, über Gleichung 5.3** In der allgemeinen Formulierung sollte in der rechts-unteren Zelle stehen $\frac{(b+d) \cdot (c+d)}{N}$
- 5.2 Seite 92, unter Gleichung 5.4** „diversen Büchern“. Danke an Gregor Didenko!
- 6 Seite 93, 1. Block R-Code** Es fehlt der Zeilenumbruch vor `names(gv)`. (In der Vorlage ist er vorhanden, in diesem Fall also ein Fehler des Verlegers.)
- 6.1 Seite 95, 1. Absatz** „der gleichen Syntax“. Danke an Gregor Didenko!
- 7.2 Seite 117, 2. Absatz von unten:** „haben“ (nicht: haven ...) Danke an Jan Hanspach!
- 9.1.5 Seite 159, 1. Absatz** „einen linearen Zusammenhang“. Danke an Gregor Didenko!
- 9.1.5 Seite 161, mitten drin** „ist der Effekt auch sehr gering“. Danke an Gregor Didenko!
- 11 Seite 187, 2. Absatz** „normalverteilter Antwortvariablen“. Danke an Gregor Didenko!
- 11.1.1 Seite 189 unten** „... teilen dann durch se: $t = -0.843/1.65 = -0.510$ “ (statt /3.74). Danke an einen anonymen Leser!
- 11.2 Seite 193, unteres Viertel** „vergleichen wir in in einer ANOVA den Anteil erklärbarer Varianz mit dem unerklärbaren“. Danke an Gregor Didenko!
- 11.2.2 Seite 198, Fußnote 11:** Fehlendes Leerzeichen nach R. Dank an - na wen wohl: Jan!
- 11.3.1 Seite 204, 4. Punkt:** „dort mit 334.4 ...“ (nicht: 333.4) Danke an Jan Hanspach!
- 11.3.1 Seite 204, 4. Punkt:** „... und mit 84.4 die SS_{Residuen} , sowie ...“ (nicht: SS_{Effekt}). Danke an einen anonymen Leser!
- 11.3.1 Seite 204, Mitte:** „ $17.6 - 10 = 7.6$ “ (statt $7.6 - 10$). Danke an einen anonymen Leser! (Name dem Verfasser bekannt.)

- 14 Seite 236, 1. Punkt:** „Die Bibel ...“(nicht: Bible²) Danke an Jan Hanspach!
- 15.2 Seite 271, 2. Absatz** „Wie erkennen wir ...“. Danke an Gregor Didenko!
- 15.2 Seite 274, 1. Absatz** „eine Alternative“. Danke an Gregor Didenko!
- 15.3.1 Seite 275** „sternenklarer Nachhimmel“. Dank an Anne-Christine Mupepele!
- 15.3.2 Seite 281, 1. Absatz** „multivariate normalverteilt“. Danke an Gregor Didenko!
- 15.3.2 Seite 281, 2. Absatz** „um eine weitere Variablen hinzuzufügen“. Danke an Gregor Didenko!
- 15.3.4 Seite 284, 2. Punkt** „des Quotienten“. Danke an Gregor Didenko!

²Wahrscheinlich ein Anglizismus und kein Ausdruck meiner atheistischen Grundhaltung.