

AZWS - Lösung 2

Peter von Rohr

2018-04-20

Aufgabe 1: Relativierung

Anhand des Zahlenbeispiels im Skript (Seite 21 und siehe auch Aufgabe 2) wurde die Relativierung mit der folgenden Gleichung erklärt

$$e = y - Xb$$

wobei

- e der Vektor der relativierten Beobachtungen ist
- y der Vektor der phänotypischen Beobachtungen ist
- b Vektor der Least-Squares Schätzungen für Betriebe
- X Design-Matrix, der Betriebe den Beobachtungen zuordnet

Ihre Aufgabe:

- Stellen Sie die Vektoren y , b und die Matrix X auf
- Berechnen Sie aus den gemachten Angaben die relativierten Beobachtungen

Lösung

- Der Vektor y entspricht dem Vektor der Zunahmen. Somit können schreiben

$$y = \begin{bmatrix} 1.26 \\ 1.32 \\ 1.40 \\ 1.44 \\ 1.52 \\ 1.50 \\ 1.42 \\ 1.46 \\ 1.34 \\ 1.32 \\ 1.24 \\ 1.28 \\ 1.44 \\ 1.40 \\ 1.54 \\ 1.56 \end{bmatrix}$$

- Der Vektor b entspricht den Least-Squares Lösungen der Betriebe. Diese lauten

$$b = \begin{bmatrix} 1.32 \\ 1.48 \end{bmatrix}$$

- Die Matrix X ist die Designmatrix, welche die Betriebe und die Beobachtungen verknüpft.

$$X = \begin{bmatrix} 1.00 & 0.00 \\ 1.00 & 0.00 \\ 1.00 & 0.00 \\ 1.00 & 0.00 \\ 0.00 & 1.00 \\ 0.00 & 1.00 \\ 0.00 & 1.00 \\ 0.00 & 1.00 \\ 1.00 & 0.00 \\ 1.00 & 0.00 \\ 1.00 & 0.00 \\ 1.00 & 0.00 \\ 0.00 & 1.00 \\ 0.00 & 1.00 \\ 0.00 & 1.00 \\ 0.00 & 1.00 \end{bmatrix}$$

- Der Vektor e der relativierten Beobachtungen wird nun berechnet als

$$e = y - Xb = \begin{bmatrix} 1.26 \\ 1.32 \\ 1.40 \\ 1.44 \\ 1.52 \\ 1.50 \\ 1.42 \\ 1.46 \\ 1.34 \\ 1.32 \\ 1.24 \\ 1.28 \\ 1.44 \\ 1.40 \\ 1.54 \\ 1.56 \end{bmatrix} - \begin{bmatrix} 1.00 & 0.00 \\ 1.00 & 0.00 \\ 1.00 & 0.00 \\ 1.00 & 0.00 \\ 0.00 & 1.00 \\ 0.00 & 1.00 \\ 0.00 & 1.00 \\ 0.00 & 1.00 \\ 1.00 & 0.00 \\ 1.00 & 0.00 \\ 1.00 & 0.00 \\ 1.00 & 0.00 \\ 0.00 & 1.00 \\ 0.00 & 1.00 \\ 0.00 & 1.00 \\ 0.00 & 1.00 \end{bmatrix} * \begin{bmatrix} 1.32 \\ 1.48 \end{bmatrix} = \begin{bmatrix} -0.06 \\ -0.00 \\ 0.08 \\ 0.12 \\ 0.04 \\ 0.02 \\ -0.06 \\ -0.02 \\ 0.02 \\ -0.00 \\ -0.08 \\ -0.04 \\ -0.04 \\ -0.08 \\ 0.06 \\ 0.08 \end{bmatrix}$$

- Vergleichen wir den manuell berechneten Vektor e der relativierten Zunahmen mit den Residuen aus der Anpassung der Regression mit `lm()` dann sehen wir die Übereinstimmung.

Tier	Residuen	RelZunahme
12	-0.065	-0.065
13	-0.005	-0.005
14	0.075	0.075
15	0.115	0.115
16	0.040	0.040
17	0.020	0.020
18	-0.060	-0.060
19	-0.020	-0.020
20	0.015	0.015
21	-0.005	-0.005
22	-0.085	-0.085
23	-0.045	-0.045
24	-0.040	-0.040
25	-0.080	-0.080

Tier	Residuen	RelZunahme
26	0.060	0.060
27	0.080	0.080

Aufgabe 2: Zuchtwertschätzung mit Nachkommenleistungen

Berechnen Sie die geschätzten Zuchtwerte und die Bestimmtheitsmasse für die drei Väter des Zahlenbeispiels (Seite 21) im Skript. Wir nehmen an, dass alle Nachkommen (Tiere 12-27) Halbgeschwister sind. Als Abweichung für die Relativierung der Leistungen nehmen wir wieder die Least Squares Lösungen der Betriebe an.

Das Zahlenbeispiel aus dem Skript ist hier in reduzierter Form noch einmal dargestellt

Tier	Vater	Betrieb	Zunahme
12	1	1	1.26
13	1	1	1.32
14	1	1	1.40
15	1	1	1.44
16	1	2	1.52
17	1	2	1.50
18	1	2	1.42
19	1	2	1.46
20	2	1	1.34
21	2	1	1.32
22	2	1	1.24
23	2	1	1.28
24	2	2	1.44
25	2	2	1.40
26	3	2	1.54
27	3	2	1.56

Ihre Aufgabe

- Schätzen Sie die Zuchtwerte für die drei Väter aufgrund der Nachkommenleistungen
- Vergleichen Sie die Rangierung der Väter aufgrund der geschätzten Zuchtwerte und aufgrund der Mittelwerte der Nachkommenwerte der Väter.

Lösung

Für die Berechnungen und die Umformatierung der Resultate brauchen wir zwei R-packages, welche wir gleich als erste laden.

```
suppressPackageStartupMessages( require(magrittr) )
suppressPackageStartupMessages( require(dplyr) )
```

Bei der Schätzung der Zuchtwerte gehen wir nach dem im Skript vorgestellten Prinzip vor.

- **Schritt 1:** Relativieren - dafür verwenden wir wieder die Least Squares Lösungen für die Betriebe. Diese werden hier nochmals gerechnet. Wichtig ist, dass wir die Betriebe zuerst in Faktoren verwandeln, sonst werden diese nicht als fixe Effekte, sondern als Kovariablen behandelt.

```
tbl_beef_data_reduced$Betrieb <- as.factor(tbl_beef_data_reduced$Betrieb)
lm_beef_farm <- lm(formula = Zunahme ~ 0 + Betrieb, data = tbl_beef_data_reduced)
```

Die Least-Squares Lösungen entsprechen dann:

```
coefficients(lm_beef_farm)
```

```
## Betrieb1 Betrieb2
##      1.325      1.480
```

Die um die Betriebseffekte korrigierten Zunahmen erhalten wir als Residuen von `lm_beef_farm`. Diese brauchen wir für die Zuchtwertschätzung also fügen wir den Vektor der Residuen als zusätzliche Kolonne zum Tibble (spezieller Dataframe) mit den Ausgangsdaten hinzu.

```
tbl_beef_data_reduced <- tbl_beef_data_reduced %>%
  mutate(RelativierteZunahme = residuals(lm_beef_farm))
```

- **Schritt 2:** Gewichtung - Für die Zuchtwertschätzung aufgrund von Nachkommenleistungen, wobei die Nachkommen alle Halbgeschwister sind, ist die Gewichtung für den Vater s im Skript gegeben als

$$b_s = \frac{2n_s}{n_s + k}$$

wobei $k = (4 - h^2)/h^2$ und die Erblichkeit (h^2) ist gegeben als $h^2 = 0.25$. Dabei bleibt der Wert von k für alle Väter konstant und beträgt $k = 15$. Der Term n_s entspricht der Anzahl Nachkommen für den Vater s und ist für jeden Vater unterschiedlich.

Somit empfiehlt es sich zuerst für jeden Vater die Anzahl Nachkommen zu bestimmen.

```
tbl_beef_offspring <- tbl_beef_data_reduced %>%
  group_by(Vater) %>%
  summarize(AnzahlNachkommen = n())
```

Als Zwischenresultate können wir diese in einer Tabelle zeigen

Vater	AnzahlNachkommen
1	8
2	6
3	2

Die Gewichtungsfaktoren b_s lassen sich aus k und der Anzahl Nachkommen berechnen.

```
tbl_beef_offspring <- tbl_beef_offspring %>%
  mutate(GewichtsfaktorB = 2*AnzahlNachkommen/(AnzahlNachkommen + k))
```

Dies fügt dem Tibble (spezieller Dataframe) einen neue Kolonne mit den Gewichtungsfaktoren hinzu. Das Resultat sieht, wie folgt aus

Vater	AnzahlNachkommen	GewichtsfaktorB
1	8	0.6956522
2	6	0.5714286
3	2	0.2352941

- **Schritt 3:** Zusammenstellen der Resultate - Aus den relativierten Zunahmen rechnen wir jetzt für jeden Vater noch den Durchschnitt und multiplizieren diese durchschnittlichen relativierten Leistungen mit

den Gewichtungsfaktoren.

```
# Mittlere Zunahmen
tbl_beef_offspring <- cbind(tbl_beef_offspring,
  tbl_beef_data_reduced %>%
    group_by(Vater) %>%
    summarise(MittlereZunahme = mean(Zunahme)) %>%
    select(MittlereZunahme))
# Mittlere relativierte Zunahmen
tbl_beef_offspring <- cbind(tbl_beef_offspring,
  tbl_beef_data_reduced %>%
    group_by(Vater) %>%
    summarise(MittlereRelZunahme = mean(RelativierteZunahme)) %>%
    select(MittlereRelZunahme))
```

Die mittleren relativierten Zunahmen geben wir hier wieder als Zwischenresultate in einer Tabelle an.

Vater	AnzahlNachkommen	GewichtsfaktorB	MittlereZunahme	MittlereRelZunahme
1	8	0.6956522	1.415000	0.0125
2	6	0.5714286	1.336667	-0.0400
3	2	0.2352941	1.550000	0.0700

Als letztes müssen wir jetzt noch die Gewichtungsfaktoren und die mittleren relativierten Leistungen miteinander multiplizieren.

```
tbl_beef_offspring <- tbl_beef_offspring %>%
  mutate(GeschZuchtwertNk = GewichtsfaktorB * MittlereRelZunahme)
```

Die geschätzten Zuchtwerte der drei Väter mit allen Zwischenresultaten sind in der nachfolgenden Tabelle gezeigt.

Vater	AnzahlNachkommen	GewichtsfaktorB	MittlereZunahme	MittlereRelZunahme	GeschZuchtwertNk
1	8	0.6956522	1.415000	0.0125	0.0086957
2	6	0.5714286	1.336667	-0.0400	-0.0228571
3	2	0.2352941	1.550000	0.0700	0.0164706

Die Rangierung der Väter aufgrund der beiden Kriterien lautet

Vater	RangMittelwert	RangZws
1	3	3
2	1	1
3	2	2

Aufgabe 3: Anzahl Nachkommen

Wie viele Nachkommen müsste Vater 1 haben, damit sein Zuchtwert grösser ist, als der von Vater 3?

Lösung

- Bei dieser Aufgabe setzen wir den geschätzten Zuchtwert von Vater 3 der Formel zur Berechnung von Vater 1 gleich, wobei die Anzahl Nachkommen als unbekannte Variable in der Formel enthalten ist. Aus dieser Formel berechnen wir dann die Anzahl Nachkommen.
- Der geschätzte Zuchtwert vom Vater 3 beträgt: 0.0164706.
- Der geschätzte Zuchtwert von Vater 1 wird berechnet als

$$\hat{u}_1 = b_1 * \bar{e}_1 = \frac{2n}{n+k} * \bar{e}_1 \quad (1)$$

Wobei \bar{e}_1 die mittlere relativierte Zunahme über die Nachkommen von Vater 1 ist. Diese beträgt 0.0125. Der geschätzte Zuchtwert für Vater 1 \hat{u}_1 soll nun gleich oder grösser als 0.0164706. Die gesuchte Grösse ist n und wir lösen die Gleichung 1 nach n auf.

$$(n+k) * \hat{u}_1 = 2n * \bar{e}_1$$

$$k * \hat{u}_1 = 2n * \bar{e}_1 - n * \hat{u}_1 = n * (2 * \bar{e}_1 - \hat{u}_1)$$

Somit ist

$$n = \frac{k * \hat{u}_1}{2 * \bar{e}_1 - \hat{u}_1} = \frac{15 * 0.0164706}{2 * 0.0125 - 0.0164706} = 28.9655172$$

Als Resultat haben wir gefunden, dass Vater 1 mindestens 29 Nachkommen haben muss, damit sein Zuchtwert grösser ist, als der von Vater 3.