

# Züchtungslehre - Lösung 8

*Peter von Rohr*

*2017-11-24*

## Aufgabe 1

In der Vorlesung wurden die linearen gemischten Modelle zur Voraussage von Zuchtwerten vorgestellt. Damit der Unterschied zwischen linearen Modellen mit nur fixen Effekten (Regressionsmodelle) und den linearen gemischten Modellen besser verständlich wird, wollen in dieser Übung nochmals anschauen, wie multiple lineare Regressionen an Daten angepasst werden.

Gegeben sei der folgende Datensatz.

| Tochter | Herde | Vater | Leistung |
|---------|-------|-------|----------|
| 1       | 1     | C     | 110      |
| 2       | 1     | A     | 100      |
| 3       | 2     | B     | 110      |
| 4       | 2     | A     | 100      |
| 5       | 2     | C     | 100      |
| 6       | 3     | C     | 110      |
| 7       | 3     | C     | 110      |
| 8       | 3     | A     | 100      |
| 9       | 3     | B     | 100      |

Die Zahlen in der Kolonne Leistung der oben gezeigten Tabelle sollen als beobachtete phänotypische Leistungen aufgefasst werden. Im anzupassenden Regressionsmodell sollen **Herde** und **Vater** als fixe Effekte verwendet werden.

## Ihre Aufgabe

- Stellen Sie das Regressionsmodell für die oben aufgelisteten Daten.
- Berechnen Sie die Schätzungen für die fixen Effekte der **Herde** und des **Vaters**.

## Hinweise

- Die Funktion zur Anpassung von Regressionen in R lautet `lm()`
- Die Angabe der Daten für die Funktion `lm()` muss über ein **Dataframe** gemacht werden. Dies kann entweder über die direkte Angabe der Daten in der Funktion `data.frame()` gemacht werden, oder die Daten können auch mit `read.csv2()` von der Webseite (<https://charlotte-ngs.github.io/LBGHS2017/ex/w9/mlrdata.csv>) eingelesen werden. Die direkte Angabe des Dataframes sieht wie folgt aus:
- Das Modell wird mit der speziellen Model-Syntax von R spezifiziert. Diese würde für unser Beispiel, wie folgt aussehen:

`Leistung ~ -1 + Herde + Vater`

- Die Resultate des angepassten Modells können mit der Funktion `summary()` angezeigt werden.

## Lösung

a) Das Regressionsmodell lautet

$$y = Xb + e$$

wobei der Beobachtungsvektor  $y$

$$y = \begin{bmatrix} 110 \\ 100 \\ 110 \\ 100 \\ 100 \\ 110 \\ 110 \\ 100 \\ 100 \end{bmatrix}$$

Die Inzidenzmatrix  $X$

$$X = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \end{bmatrix}$$

Der Vektor  $b$  der fixen Effekte

$$b = \begin{bmatrix} Herde_1 \\ Herde_2 \\ Herde_3 \\ Vater_A \\ Vater_B \\ Vater_C \end{bmatrix}$$

b) Angenommen, die Daten sind in einem Dataframe namens `dfMlrData` gespeichert. Dann wird das Regressionsmodell mit folgenden Anweisungen angepasst.

```
lmMlrData <- lm(Leistung ~ -1 + Herde + Vater, data = dfMlrData)
summary(lmMlrData)
```

```
##
## Call:
## lm(formula = Leistung ~ -1 + Herde + Vater, data = dfMlrData)
##
## Residuals:
##      1      2      3      4      5      6      7      8      9
##  1.2281 -1.2281  5.4386  1.0526 -6.4912  2.6316  2.6316  0.1754 -5.4386
##
## Coefficients:
```

```

##           Estimate Std. Error t value Pr(>|t|)
## Herde1  101.228      4.400  23.006 2.11e-05 ***
## Herde2   98.947      4.155  23.812 1.84e-05 ***
## Herde3   99.825      4.027  24.786 1.57e-05 ***
## VaterB    5.614      5.216   1.076   0.342
## VaterC    7.544      4.218   1.789   0.148
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.461 on 4 degrees of freedom
## Multiple R-squared:  0.9988, Adjusted R-squared:  0.9973
## F-statistic: 659.1 on 5 and 4 DF,  p-value: 6.423e-06

```

## Aufgabe 2

Anstelle des Regressionsmodells in der Aufgabe 1 verwenden wir jetzt ein gemischtes lineares Modell. In diesem Modell werden die Herden als fixe Effekte und die Väter als zufällige Effekte modelliert. In der Vorlesung werden wir diese Art von Modellen noch genauer unter dem Namen “Vatermodell” anschauen. Die Kovarianzmatrix  $G$  der zufälligen Vätereffekte setzen wir dabei zu

$$G = 0.1 * I$$

wobei  $I$  die Einheitsmatrix bezeichnet.

### Ihre Aufgabe

- Stellen Sie die Mischmodellgleichungen für die Daten in Aufgabe 1 auf.
- Berechnen Sie die Schätzer für die fixen Effekte der Herde und machen Sie eine Voraussage für die zufälligen Vätereffekte

### Lösung

- Die Mischmodellgleichungen lauten

$$\begin{bmatrix} X^T X & X^T Z \\ Z^T X & Z^T Z + G^{-1} \end{bmatrix} \begin{bmatrix} \hat{H}erde_1 \\ \hat{H}erde_2 \\ \hat{H}erde_3 \\ \hat{V}ater_A \\ \hat{V}ater_B \\ \hat{V}ater_C \end{bmatrix} = \begin{bmatrix} X^T y \\ Z^T y \end{bmatrix}$$

Abgekürzt können wir die Mischmodellgleichungen schreiben als

$$M * \hat{s} = r$$

Die einzelnen Komponenten lauten

$$X^T X = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 4 \end{bmatrix}$$

$$X^T Z = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 2 \end{bmatrix}$$

$$Z^T Z = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{bmatrix}$$

$$G^{-1} = 10 * I$$

$$X^T y = \begin{bmatrix} 210 \\ 310 \\ 420 \end{bmatrix}$$

$$Z^T y = \begin{bmatrix} 300 \\ 210 \\ 430 \end{bmatrix}$$

Stellen wir alle Komponenten zusammen erhalten wir

$$\begin{bmatrix} 2 & 0 & 0 & 1 & 0 & 1 \\ 0 & 3 & 0 & 1 & 1 & 1 \\ 0 & 0 & 4 & 1 & 1 & 2 \\ 1 & 1 & 1 & 13 & 0 & 0 \\ 0 & 1 & 1 & 0 & 12 & 0 \\ 1 & 1 & 2 & 0 & 0 & 14 \end{bmatrix} \begin{bmatrix} \hat{H}erde_1 \\ \hat{H}erde_2 \\ \hat{H}erde_3 \\ \hat{V}ater_A \\ \hat{V}ater_B \\ \hat{V}ater_C \end{bmatrix} = \begin{bmatrix} 210 \\ 310 \\ 420 \\ 300 \\ 210 \\ 430 \end{bmatrix}$$

b) Die Lösungen berechnen wir als

$$\hat{s} = M^{-1} * r$$

$$\hat{s} = \begin{bmatrix} 105.0784 \\ 103.3333 \\ 104.7854 \\ -1.0152 \\ 0.1568 \\ 0.8584 \end{bmatrix}$$