

ASMNW - Übung 1

Peter von Rohr

2018-03-02

Kontrollfrage 1

Welches Modell wurde vor der Genomischen Selektion zur Zuchtwertschätzung verwendet und welche Tiere bekamen in diesem Modell Zuchtwerte?

Kontrollfrage 2

Beim gängigen Verfahren zur genomischen Zuchtwertschätzung braucht es mehrere Schritte, wie sehen diese aus?

Kontrollfrage 3

Was bedeuten die a -Werte in den Modellen der genomischen Zuchtwertschätzung und welchem genetischen Modell werden diese entnommen?

Kontrollfrage 4

Im Paper zur Deregression (auf dem Stick oder unter: <http://gsejournal.biomedcentral.com/articles/10.1186/1297-9686-41-55>) stehen nach Gleichung (8) zwei Probleme, weshalb mit BLUP geschätzte Zuchtwerte nicht ideal sind als Beobachtungen in genomischer Zuchtwertschätzung. Fassen Sie diese zwei Probleme mit Ihren Worten kurz zusammen.

Aufgabe 1: Modellierung

In einem kleinen Beispieldatensatz sind die SNP-Genotypen für 5 Tiere gegeben. Für jedes Tier liegen Typisierungsergebnisse an 10 SNPs vor. Die Bezeichnung $(G_k G_l)_{ij}$ steht für den Genotypen für Tier i an der SNP-Position j mit den Allelen k und l . Da wir nur SNPs mit zwei Allelen betrachten können als k und l nur entweder 0 oder 1 sein. Wir nehmen an, dass das Allel 0 immer das Allel mit der gewünschten Ausprägung ist. Im Substitutionseffekt ignorieren wir alle Dominanzeffekte, d.h. alle d -Werte werden auf 0 gesetzt. In der folgenden Tabelle sind die SNP-Genotypen für alle Tiere aufgelistet, wobei die Indices i und j weggelassen wurden.

	Tier 1	Tier 2	Tier 3	Tier 4	Tier 5
SNP1	$G_0 G_0$	$G_0 G_1$	$G_0 G_1$	$G_0 G_1$	$G_0 G_0$
SNP2	$G_0 G_1$	$G_0 G_1$	$G_0 G_1$	$G_0 G_0$	$G_0 G_1$
SNP3	$G_0 G_0$	$G_0 G_1$	$G_0 G_1$	$G_0 G_1$	$G_0 G_1$
SNP4	$G_1 G_1$	$G_0 G_1$	$G_0 G_1$	$G_0 G_1$	$G_0 G_1$
SNP5	$G_0 G_1$	$G_0 G_0$	$G_0 G_1$	$G_0 G_1$	$G_0 G_1$
SNP6	$G_1 G_1$	$G_0 G_1$	$G_0 G_1$	$G_0 G_0$	$G_0 G_1$
SNP7	$G_1 G_1$	$G_0 G_1$	$G_1 G_1$	$G_0 G_1$	$G_0 G_1$
SNP8	$G_0 G_1$	$G_0 G_1$	$G_0 G_1$	$G_0 G_0$	$G_0 G_1$
SNP9	$G_0 G_1$	$G_0 G_1$	$G_0 G_0$	$G_0 G_0$	$G_0 G_1$
SNP10	$G_0 G_1$	$G_0 G_0$	$G_1 G_1$	$G_0 G_1$	$G_0 G_1$

Wir möchten aufgrund des gegebenen Datensatzes die a -Werte schätzen. Dafür verwenden wir das folgende Modell

$$\hat{g}_d = 1\mu + Ma + \epsilon_d \quad (1)$$

wobei \hat{g}_d Vektor der deregressierten BLUP-Zuchtwerte
 μ allgemeines Mittel
 a Allelsubstitutionseffekte
 M Inzidenzmatrix, welche \hat{g}_d und a verknüpft
 ϵ_d zufällige Resteffekte

Wir nehmen an, dass für jedes Tier nur ein deregressierter Zuchtwert vorliegt.

Ihr Aufgabe:

Stellen Sie den Vektor a und die Matrix M für den gezeigten Genotypendatensatz und das Modell (1) auf.

NB

Diese Aufgabe dient nur der Anschauung. Für den praktischen Einsatz wäre der Datensatz viel zu klein.

Aufgabe 2: Schätzung der a-Effekte

Wir nehmen an, dass wir aus einer unabhängigen Studie wissen, dass nur die beiden SNP 1 und 6 einen wirklichen Einfluss auf die Ausprägung des Merkmals haben. Für unser Merkmal konnten die folgenden deregressierten Zuchtwerte (y) für unsere 5 Tiere gefunden werden.

Tier	y
1	18.563077
2	8.413929
3	11.506708
4	16.673748
5	18.793892

Ihre Aufgabe

Schätzen Sie die a -Effekte der beiden SNPs 1 und 6, auf die oben gezeigten Beobachtungen mit einer multiplen linearen Regression. Verwenden Sie dazu die Funktion `lm()` in R.

Aufgabe 3: Reduktion der Varianz

Bei der BLUP-Zuchtwertschätzung haben die geschätzten Zuchtwerte im Vergleich zu den phänotypischen Werten eine reduzierte Varianz. Dies können wir an folgendem Beispiel mit R zeigen. Wir verwenden dazu einen Datensatz aus einer Übung der Züchtungslehre. Der Datensatz wird mit folgendem Befehl eingelesen:

```
dfLmm <- read.csv2(file =  
  "https://charlotte-ngs.github.io/GELASMFS2018/ex/w2/z1_w7_u5_DataLmm.csv")
```

Die Struktur der Daten können wir mit dem Befehl `str` anzeigen. Die Kolonne `y` enthält die beobachteten Daten.

```
str(dfLmm)
```

```
## 'data.frame':  240 obs. of  3 variables:
## $ ID          : int  1 1 1 1 1 1 1 1 1 1 ...
## $ FixerFactor: int -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 ...
## $ y           : num -35.6 -34 -35.4 -33.7 -37.4 ...
```

BLUP-Zuchtwerte werden mit dem Package `pedigreemm` geschätzt.

```
library(pedigreeemm)
```

```
## Loading required package: lme4
```

```
## Loading required package: Matrix
```

```
nAnzAnim <- 6
```

```
pedP1 <- pedigree(sire = as.integer(c(NA,NA,1, 1,4,5)),
                 dam  = as.integer(c(NA,NA,2,NA,3,2)),
                 label = as.character(1:nAnzAnim))
```

```
fitReml <- pedigreeemm(formula = y ~ FixerFactor + (1 | ID),
                      data = dfLmm,
                      pedigree = list(ID = pedP1))
```

Die geschätzten Zuchtwerte erhalten wir aus dem Slot `u` aus dem Resultat-Objekt `fitReml`. Der Befehl

```
fitReml@u
```

zeigt den Vektor der geschätzten Zuchtwerte. Die Funktion `var()` kann nun verwendet werden um die Varianz der Beobachtungen mit der Varianz der geschätzten Zuchtwerte zu vergleichen.