

ASMNW - Übung 4

Peter von Rohr

2017-03-17

Aufgabe 1: Einfaches Beispiel mit nur einem SNP

Wir betrachten ein einfaches Beispiel mit nur 10 Tieren, welche für einen einzigen SNP typisiert sind. Die folgende Tabelle gibt eine Übersicht über die Daten mit den SNP-Allelen und den phänotypischen Beobachtungen.

Animal	Phentype	Sn Allele1	Sn Allele2
1	2.03	1	1
2	3.54	1	2
3	3.83	1	2
4	4.87	2	2
5	3.41	1	2
6	2.34	1	1
7	2.65	1	1
8	3.76	1	2
9	3.69	1	2
10	3.69	1	2

Wir nehmen an die Tiere seien nicht verwandt miteinander. Somit können wir die Beziehung zwischen dem einen SNP und den phänotypischen Beobachtungen mit einem einfachen Regressionsmodell testen. Unser Modell lautet:

$$y = 1_n \mu + Wg + e \quad (1)$$

wobei

- y Vektor der Länge n mit phänotypischen Beobachtungen
- μ allgemeines Mittel, welches fixe Effekte repräsentiert
- 1_n Vektor der Länge n mit lauter Einsen
- g additiver Effekt des Marker-SNP
- W Inzidenzmatrix, welche die Beobachtungen zum Marker-Effekt verbindet
- e Vektor der zufälligen Resteffekte

Die Inzidenzmatrix W hat n Zeilen und so viele Kolonnen, wie SNP-Marker. Für unser Beispiel hat die Matrix W somit 1 Kolonne. Die Elemente der Matrix W zählen die Anzahl Allele mit positiver Wirkung. In diesem Beispiel sei das Allel "2".

Ihre Aufgabe

Stellen Sie das Modell aus Gleichung (1) für den gegebenen Datensatz auf und bestimmen Sie welche Modellkomponenten bekannt und welche unbekannt sind.

Aufgabe 2: Least Squares Lösungen

Da wir 10 Beobachtungen und nur einen SNP betrachten ist die Bedingung für die Schätzung der unbekannt Parameter mit Least Squares erfüllt. Somit können wir die unbekannt Parameter μ und g mit der folgenden Gleichung schätzen.

$$\begin{bmatrix} \hat{\mu} \\ \hat{g} \end{bmatrix} = \begin{bmatrix} 1_n^T 1_n & 1_n^T W \\ W^T 1_n & W^T W \end{bmatrix}^{-1} \begin{bmatrix} 1_n^T y \\ W^T y \end{bmatrix} \quad (2)$$

Berechnen Sie aufgrund der Gleichungen in (2) die Lösungen für $\hat{\mu}$ und \hat{g} .

Aufgabe 3

Überprüfen Sie die unter Aufgabe 2 erhaltenen Least Squares Lösungen für μ und g mit der Funktion `lm()` R.

Hinweise

- Lesen Sie die Daten aus der in Aufgabe 1 gezeigten Tabelle in den Dataframe namens `dfAufgabe3` ein.

```
nAnzTiere <- 10
dfAufgabe3 <- data.frame(
  Animal = c(1:nAnzTiere),
  Phentype = c(2.03, 3.54, 3.83, 4.87, 3.41, 2.34, 2.65, 3.76, 3.69, 3.69),
  SnpAllele1 = c(1, 1, 1, 2, 1, 1, 1, 1, 1, 1),
  SnpAllele2 = c(1, 2, 2, 2, 2, 1, 1, 2, 2, 2))
```

- Fügen Sie dem Dataframe eine zusätzliche Kolonne namens `Genotype` hinzu, welche die Genotypen-Codes enthält. Diese Codes entsprechen den Anzahl an "2" Allelen mit positiver Wirkung.
- Verwenden Sie die phänotypischen Beobachtungen in `dfAufgabe3$Phentype` als Zielgröße und `dfAufgabe3$Genotype` als erklärende Variable und passen Sie das Regressionsmodell mit der Funktion `lm()` an.