

ASMNW - Lösung 3

Peter von Rohr

2018-03-16

Kontrollfrage 1

1. Aus welchen Komponenten besteht ein lineares Modell?
2. Welcher Zusammenhang besteht zwischen diesen Komponenten

Kontrollfrage 2

1. Welche fixen und welche zufälligen Effekte gibt es in einer multiplen linearen Regression
2. Wie lauten Erwartungswerte und Varianzen der zufälligen Effekte

Kontrollfrage 3

Welche vier Ziele wollen wir mit einer multiplen linearen Regression erreichen?

Kontrollfrage 4

1. Was bedeutet der Ausdruck

$$\hat{\beta} = \operatorname{argmin}_{\beta} \|\mathbf{y} - \mathbf{X}\beta\|^2$$

2. Welcher Schätzer resultiert aus dem Ausdruck unter 1?

Aufgabe 1: Lineare Modelle

In der Vorlesung wurden die Eigenschaften von linearen Modellen besprochen. Die folgende Tabelle enthält vier verschiedene Modelle. Tragen Sie in die Kolonne ganz rechts ein, ob es sich jeweils um ein lineares oder nicht-lineares Modell handelt.

Nr	Model	Typ(linear/nicht-linear)
1	$y_i = \beta_1^2 + \sqrt{\beta_2}x_{i,2} + \epsilon_i$	
2	$y_i = \beta_1x_{i,1} + \beta_2x_{i,2} + \beta_3x_{i,3} + \epsilon_i$	
3	$y_i = \beta_1x_{i,1} + \beta_2 \log x_{i,2} + \beta_3 \sin(x_{i,3}) + \epsilon_i$	
4	$y_i = \beta_1^2 \log x_{i,1} + \sqrt{\beta_2}x_{i,2} + \epsilon_i$	

Aufgabe 2: Regressionsanalyse

Während der Vorlesung habe wir das Beispiel mit der Zunahme nach dem Absetzen als Zielgrösse betrachtet. Wir haben aber nirgends die Koeffizienten und die geschätzte Restvarianz berechnet. Stellen Sie für dieses kleine Beispiel das lineare Modell gemäss Vorlesungsunterlagen auf und berechnen Sie die Koeffizienten des linearen Modells und die geschätzte Restvarianz. Schätzen Sie die unbekannt Parameter einmal mit und einmal ohne Achsenabschnitt und vergleichen Sie die Resultate.

Aufgabe 3: GBLUP

Gegeben sind die Genotypen von 10 Tieren an 1000 SNP-Genorten. Die Genotypen können von der Datei `asmas_w04_u03_genotypes.txt` gelesen werden. Da wir nach dem Einlesen die Genotypen in Form einer Matrix haben wollen, werden die Inputdaten am einfachsten mit den Funktionen `matrix(scan())` eingelesen und einer Variablen zugewiesen, wie dies im folgenden Statement gezeigt wird (Die beiden `#` gehören nicht zum Einlese-Statement und deshalb müssen Sie diese weglassen).

```
## mat_data <- matrix(scan("asmas_w04_u03_genotypes.txt"), ncol=1000, byrow = TRUE)
```

Ihre Aufgabe

Aus den eingelesenen SNP-Genotypen soll die genomische Verwandtschaftsmatrix G berechnet werden. Dazu müssen Sie zuerst herausfinden, in welcher Codierung die SNP-Genotypen vorliegen. In der Vorlesung haben wir gesehen, dass wir zwischen den Codierungen $(0, 1, 2)$ und $(-1, 0, 1)$ unterscheiden können. In den Vorlesungsunterlagen wurde die Herleitung von G für die zweite Codierung gezeigt. Sie können die Codierung der eingelesenen Genotypen mit der Funktion `table()` bestimmen. Dabei geben Sie die Matrix der eingelesenen Genotypen als Argument für die Funktion an.

Damit sollten Sie alle nötigen Angaben haben, damit Sie die Matrix G berechnen können.

Zusatzaufgabe

Da wir die Matrix G anschliessend in den Mischmodellgleichungen von GBLUP verwenden wollen, muss diese Matrix positiv-definit sein, da wir ihre Inverse G^{-1} benötigen. Für die Überprüfung ob die Matrix G positiv-definit ist, können Sie ihre Eigenwerte mit der Funktion `eigen()` berechnen. Die Funktion `eigen()` gibt als Resultat eine Liste zurück, wobei die Eigenwerte als Vektor unter dem Namen `values` abgelegt sind.