

R-Blatt 3: Regressionsmodelle in R - Aufgaben

Statistical Aspects (09-202-2413)

Janne Pott

Last compiled on 20 September, 2022

Session Setup

```
rm(list = ls())
time0<-Sys.time()

source("../sourceFile.R")
setwd(pathToExercise)

knitr::opts_chunk$set(echo = TRUE)
```

Lineare Regression

Bitte laden Sie den Datensatz *SNP.RData* mittels *load()*. Die Beschreibung der Variablen finden Sie in Tabellen [1](#).

Tabelle 1: Parameterbeschreibung zum Datensatz SNP

Variable	Beschreibung	Codierung / Einheit
id	Durchlaufende ID-Nummer	NA
trait	Merkmal/Phänotyp	kontinuierlich
sex	Geschlecht	1 = Mann; 2 = Frau
SNP	Genotyp	0 = AA; 1 = AB; 2 = BB

- Untersuchen Sie die Effekte von *sex* und *SNP* auf *trait* mittels linearer Regression!
- Untersuchen Sie die multiplen Effekte und die Interaktion der Einflussvariablen *sex* und *SNP*!
- Welche von den vier Modellen ist besser geeignet, um *trait* zu beschreiben? Begründen Sie Ihre Entscheidung!
- Welches genetische Modell wird hier verwendet?
- Erstellen Sie je einen Wahrscheinlichkeits-Vektor pro Genotyp **AA**, **AB** und **BB** (1 = dieser Genotyp trifft zu, 0 = trifft nicht zu).
- Erstellen Sie nun je einen Vektor pro genetischen Modell (additiv, dominant, rezessiv). (Hinweis: Überlegen Sie zuerst, welchen der Vektoren Sie bereits haben (Teilaufgabe d))
- Untersuchen Sie die Effekte der verschiedenen Modelle auf *trait* mittels lineare Regression (univariat)! Welches Modell ist am besten geeignet? Wie könnten Sie Ihr Ergebnis testen?
- Können wir annehmen, dass der SNP auf den Autosomen liegt, oder könnte es sich auch um einen X-chromosomalen SNP handeln? Begründen Sie Ihre Antwort.

Logistische / Proportional Odds Regression

Bitte verwenden Sie den Datensatz *SNP.RData* aus der obigen Aufgabe.

- Berechnen Sie den **Median** von *trait* und nutzen Sie diesen als Cut-off, um *trait* in einen binären Phänotyp *trait2* zu zerlegen.
- Untersuchen Sie die univariaten und multivariaten Effekte von *sex* und *SNP* auf *trait2* mittels logistischer Regression! (Hinweis: Funktion *glm()* mit *family="binomial"*)
- Bestimmen Sie die **Quartile** von *trait*! Zerlegen Sie *trait* nun in einen 4-stufigen Phänotypen *trait3*, in dem Sie die Quartile als Kategorien nutzen.
- Untersuchen Sie die univariaten und multivariaten Effekte von *sex* und *SNP* auf *trait3* mittels proportional odds regression! (Hinweis: Funktion *polr()* aus dem Paket **MASS** mit *Hess=T*)
- Vergleichen Sie Ihre Ergebnisse von b) und d) mit den Ergebnissen von Aufgabe 1

Nichtlineare Regression

Bitte laden Sie den Datensatz *MichMenten.RData* mittels *load()*. Die Beschreibung der Variablen finden Sie in Tabellen 2.

Tabelle 2: Parameterbeschreibung zum Datensatz MichMenten

Variable	Beschreibung	Codierung / Einheit
cS	Substratkonzentration	in 10^{-5} mol
vA	Umsatzgeschwindigkeit, gemessen in Erwachsenen	in $\mu\text{mol}/(\text{mg Enzym}) \cdot \text{min}$
vE	Umsatzgeschwindigkeit, gemessen in Embryonen	in $\mu\text{mol}/(\text{mg Enzym}) \cdot \text{min}$

- Bestimmen Sie V_{max} und K_m für Erwachsene und Embryonen getrennt, indem Sie die Funktion *nl()* und folgende Startwerte nutzen: $V_{max} = \max(v)$ und $K_m = \frac{1}{2}\max(v)$
- Was passiert wenn man die Startwerte weglässt?
- Fassen Sie die Ergebnisse in einem Plot zusammen und interpretieren Sie diesen!

Session Information

```
sessionInfo()
message("\nTOTAL TIME : " ,round(difftime(Sys.time(),time0,units = "mins"),3)," minutes")
```