

R-Blatt 5: Genomweite (Meta-)Analysen - Aufgaben

Statistical Aspects (09-202-2413)

Janne Pott

Last compiled on 06 Oktober, 2022

Session Setup

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

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

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

In dieser Aufgabe sollen mehrere GWAS zu einem Phänotyp auf dem gleichen Datensatz durchgeführt werden.

- Datensatz: *1KG_PCA* (bekannt aus R-Blatt 4 - PCA)
- Phänotyp: *Phenotype* in Datei *PhenoFile.txt*
- Kovariablen: *sex* und *age* in Datei *CovarFile.txt* und die 10 PCs aus R-Blatt 4

Voranalyse

Bitte laden Sie die Phänotyp-, Kovariablen und PC-Dateien in R ein.

- a) Prüfen Sie die Verteilung von *Phenotype*. Muss dieser noch geeignet transformiert werden, oder ist er in etwa normalverteilt?
- b) Prüfen Sie, ob es eine Beziehung zwischen *Phenotype* und den Kovariablen gibt. Prüfen Sie zusätzlich, ob es einen Effekt der Ethnien gibt. Visualisieren Sie die Ergebnisse mit geeigneten Boxplots und Interaktionsplots (*sex* und *ethnie* auf *Phenotype*).
- c) Mergen Sie die Kovariablen mit den PCs und speichern Sie die Tabelle als *CovarFile_withPCs.txt*

GWAS

Wir werden zunächst 4 GWASs durchführen:

- 1) $Phenotype \sim SNP + sex + age$
- 2) $Phenotype \sim SNP + sex + age + PCs$
- 3) $Phenotype \sim SNP + age + PCs$ in Frauen
- 4) $Phenotype \sim SNP + age + PCs$ in Männern

- a) Erzeugen Sie eine Datei, die pro GWAS Modell die richtigen Samples enthält. Falls Sie im R-Blatt 4 nur die PCA mit 3 x 246 Individuen erzeugt haben, können Sie für Modell 1 & 2 die gleiche Datei *mySamples.txt* nutzen. Für Männer und Frauen müssen Sie diese Datei oder die *1KG_PCA.psam* Datei noch geeignet filtern und abspeichern.
- b) Führen Sie die GWAS für alle 4 Modelle durch. [PLINK2](#) Befehle können online nachgeschlagen werden. Beispielcode für Modell 1:

```
myCall1 = paste0(pathToPLINK2,
  " --pfile ",pathToData,"1KG_PCA",
  " --glm allow-no-covars hide-covar firth-fallback",
  " cols=chrom,pos,ref,alt,firth,test,nobs,machr2,alfreq,",
  "alfreqcc,alcountcc,orbeta,se,ci,tz,p",
  " --pheno ",pathToData,"PhenoFile.txt",
  " --covar ",pathToData,"CovarFile.txt",
  " --covar-variance-standardize",
  " --out ",pathToData,"Model1")
system(myCall1)
```

- c) Die Ergebnisse wurden in den Dateien *ModelX.Phenotype.glm.linear* abgelegt. Laden Sie diese Daten mit *fread()* in R ein und erzeugen Sie QQ-Plots pro Modell. Bestimmen Sie den Inflationsfaktor λ als den normierten Median der χ^2 -verteilten Zufallsvariablen der P-Werten (*qchisq()*). Interpretieren Sie das Ergebnis.

$$\lambda = \frac{\text{median}(Y_1^2, \dots, Y_n^2)}{0.456}$$

- d) Filtern Sie das Topergebnis! Welche rs-ID hat dieser SNP, welche Allelfrequenz? Gibt es Unterschiede in Männern und Frauen bezüglich der Allelfrequenz oder der Effektschätzers?

GWAMA

Nun werden wir eine Meta-Analyse einiger SNPs durchführen. Dazu verwenden wir die Daten der stratifizierten GWASs.

- a) Filtern Sie alle SNPs, die in den Modellen 2-4 einen p-Wert $< 10^{-4}$ haben, und erstellen Sie eine Top-Liste mit:
 - SNP-ID, Chromosom, Position, Effekt- und anderes Allele
 - Allelfrequenz, Fallzahl, Effektschätzer, Standardfehler und p-Wert pro Strata
- b) Führen Sie für diese Top-Liste eine Meta-Analyse pro SNP durch. Verwenden Sie dazu die Funktion *metagen* aus dem R-Paket *meta*.
- c) Interpretieren Sie das Ergebnis! Gibt es genomweit signifikante Ergebnisse? Unterscheiden sich die Ergebnisse der Meta-Analyse von denen der kombinierten Analyse (Modell 2)?

Session Information

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