

# Modul Statistische Aspekte der Analyse molekularbiologischer und genetischer Daten

## Übungsblatt 5: Imputation und GWAS

Janne Pott

WS 2021/22

Sie können Ihre Lösungen zu Aufgabe 3 als PDF in Moodle hochladen (Frist: 03.01.2022).

### Aufgabe 1: Imputation

Beschreiben Sie die drei wichtigsten Schritte einer Imputation.

### Aufgabe 2: Gemischte Modelle

Ein paar Fragen zu Mixed Models:

- Generell, was „mischt“ man bei den gemischten Modellen?
- Sie haben Labordaten von 990 Probanden, die in 33 Batches analysiert wurden (30 Probanden auf einer Batch). Jede Batch hatte eine unterschiedliche Lagerdauer auf Eis, Boxnummer, Aliquotierung, PEG-Fällung oder Assay-Datum. Wie könnten Sie mit Hilfe eines Mixed Models so adjustieren, dass kein Batch-Effekt mehr auftritt?
- Sie haben Fitness-Daten von sechs Probanden, die drei verschiedenen Ergometer je dreimal genutzt haben. Sie interessieren sich für den Effekt der Ergometer, wobei es Variationen zwischen den Probanden und innerhalb der Probanden gibt. Stellen Sie ein geeignetes Modell auf! (Hinweis: Sie wollen  $y_{ijk}$  bestimmen, für den  $i$ -ten Proband, an dem  $j$ -ten Ergometer und der  $k$ -ten Wiederholung!)
- Was ist die graphische Bedeutung eines gemischten Modells?

### Aufgabe 3: Haplotypen & Haploblöcke

Auf dem Übungsblatt 1 und 3 haben wir zwei SNPs kennengelernt, die das AB0-Blutgruppensystem beeinflussen (Blatt 1 Aufg. 3, Blatt 3 Aufg 4).

- Definieren Sie die Begriffe **Haplotyp** und **Haploblock** anhand der SNPs!
- Es gibt noch viele weitere SNPs im *AB0* Gen. Warum reichen die Genotypen dieser zwei SNPs dennoch fast immer aus, um die Blutgruppen zu bestimmen?
- Kann ein Kind Blutgruppe B haben, wenn die Mutter Blutgruppe A und der Vater Blutgruppe O hat?
- Nehmen Sie an, Sie haben mittels SNP-Array (Coverage 80%) die Genotypen der zwei SNPs von 5000 Europäern bestimmt. Was bedeutet eine Coverage von 80% und wovon hängt dies ab? Können Sie damit die Blutgruppen aller Individuen bestimmen?

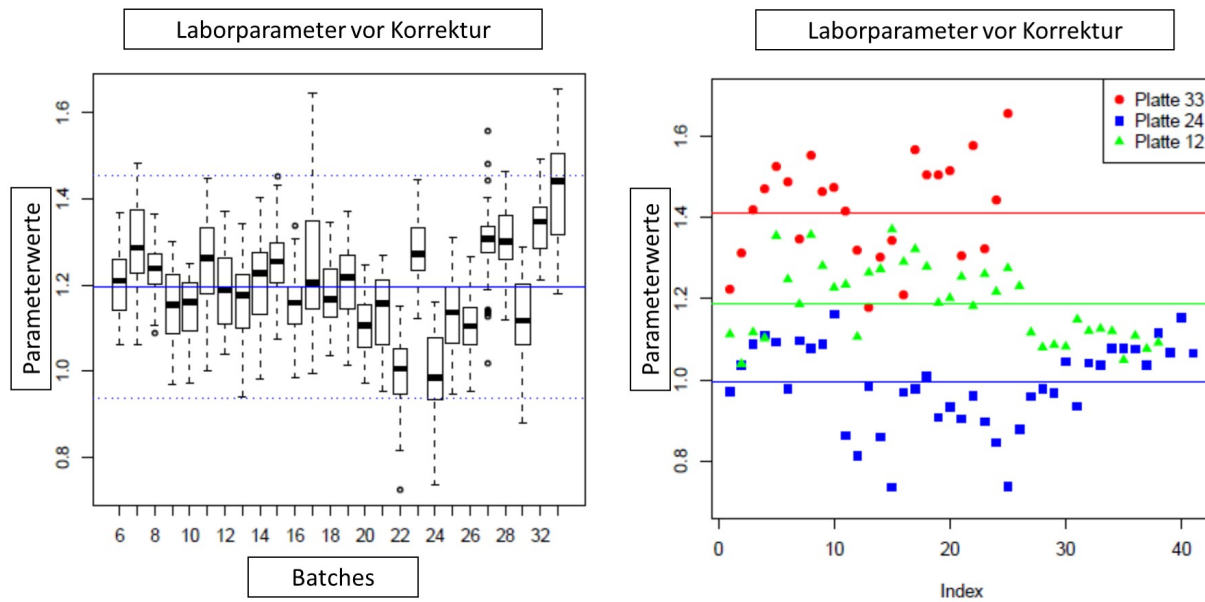


Figure 1: Beispieldaten zu Aufgabe 2: Die Streuung in den Batches ist ähnlich, aber die Mittelwerte unterscheiden sich deutlich.

## Aufgabe 4: Studiendesigns

In der Vorlesung wurde Ihnen die LIFE-Studie vorgestellt. Wie unterscheiden sich LIFE-Adult und LIFE-Heart im Studiendesign? Nennen Sie je eine Fragestellung, die sich mit der jeweiligen Studie besser beantworten lassen!