Aufgabe 1

Leiten Sie die restringierte log-Likelihood über Herausintergrieren von $\boldsymbol{\beta}$ aus der gemeinsamen Likelihood $L(\boldsymbol{\beta}, \boldsymbol{\vartheta})$ her.
Leiten Sie dazu zunächst folgende Gleichheit her:

$$(\boldsymbol{y} - \boldsymbol{X\beta})^\top \boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X\beta}) = (\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^\top \boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^\top (\boldsymbol{X}^\top \boldsymbol{V}^{-1}\boldsymbol{X})(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}),$$

vgl. Vorlesung Kapitel 2.2 *Schätzung der Kovarianzstruktur*.

Aufgabe 2

Betrachten Sie den Datensatz `sprint` zu den 100m-Läufen, den Sie aus der Vorlesung kennen.

```
# Simulation der sprint Daten #
# =========================== #

set.seed(123)

n1 = 10
n2 = n1
n = n1 + n2
M = 3
b = rep(rnorm(n1 + n2, 0, 0.5), each = M)
sex = c(rep(0, n1), rep(1, n2))
beta = c(rep(10.6, n1 * M), rep(11.9, n2 * M))
eps = rnorm(n, 0, 0.2)
times = beta + b + eps
athlete = rep(1:n, each = M)

sprint <- data.frame(athlete, sex = rep(sex, each = M), times)
```

(a) Legen Sie ein `groupedData`-Object für die Daten an (R-Paket `nlme`), das die Gruppierungsstruktur widerspiegelt. Erstellen Sie eine Graphik für die Daten.

(b) Fitten Sie das Modell mit einem festen Effekt für Geschlecht und zufälligen Effekten für die Athleten (Modell wie in der Vorlesung besprochen) einmal über ML-Schätzung und einmal über REML-Schätzung. Was ist der Unterschied zwischen REML- und ML-Schätzung? Vergleichen Sie die Ergebnisse. Lassen Sie sich jeweils die Schätzer für die festen und die zufälligen Effekte ausgeben.
Hinweis: Funktion `lme()` aus dem R-Paket `nlme` fittet lineare gemischte Modelle.

(c) Berechnen Sie nun die Schätzer der festen Effekte in R "zu Fuß", d.h. nutzen Sie die Formel für den verallgemeinerten KQ-Schätzer und berechnen Sie die Schätzer aus den entsprechenden Matrizen und Vektoren. Zur Bestimmung der Kovarianzmatrix $\boldsymbol{V}$ der Zielgröße können Sie die Ergebnisse aus dem Modellfit in (b), sowie die Formel aus dem Skript verwenden verwenden.
Hinweis: Die Funktion `getVarCov()` extrahiert den Schätzer für die Varianz-Kovarianzmatrix der zufälligen Effekte aus einem `lme`-Objekt.

(d) Erstellen Sie eine Graphik, die die Zeiten für alle Athleten darstellt. Zeichnen Sie zusätzlich die Prognosen aus den festen Effekten sowie die Prognosen aus den festen und zufälligen Effekten ein.

(e) Generieren Sie eine zusätzliche Variable `nation`, die 4 Ausprägungen hat und die Nationalität der Läufer darstellt (bereits im Code enthalten). Schätzen Sie ein adäquates Modell für die Zeiten der Läufer, das auch die Nationalität berücksichtigt.

Aufgabe 3

Betrachten Sie erneut den Datensatz `Machines`, den Sie bereits von Blatt 1 kennen. Dieser Datensatz ist im R-Paket `nlme` enthalten.

(a) Fitten Sie folgendes Modell in R:

$$y_{ijk} = \beta_j + b_i + \varepsilon_{ijk}, \quad i = 1, \ldots, 6,\ j = 1, \ldots, 3,\ k = 1, \ldots, 3$$

$$b_i \overset{\text{iid}}{\sim} N(0, \tau^2), \quad \varepsilon_{ijk} \overset{\text{iid}}{\sim} N(0, \sigma^2), \quad b_i,\ \varepsilon_{ijk} \text{ unabhängig}$$

Dabei ist y_{ijk} der Score für Arbeiter i an Maschine j in der k-ten Wiederholung. β_j sind also die festen Effekte der Maschinen und b_i die zufälligen Effekte der Arbeiter.
Lassen Sie sich die festen und die zufälligen Effekte ausgeben. Erstellen Sie einen Plot der Residuen.

(b) Fitten Sie nun ein Modell, das zusätzlich die Interaktion von Arbeiter und Maschine berücksichtigt. Lassen Sie sich die geschätzte Kovarianzmatrix der zufälligen Effekte sowie die geschätzte Varianz der Fehler ausgeben.

(c) Löschen Sie zufällig 10 Beobachtungen aus dem Datensatz und berechnen Sie die Modelle aus (a) und (b) auf den unbalancierten Daten. Was fällt Ihnen auf wenn Sie die festen Effekte dieser beiden Modelle vergleichen?

Aufgabe 4 (Selbststudium)

Betrachten Sie erneut den Datensatz `Oxide`, den Sie bereits von Blatt 1 kennen. Dieser Datensatz ist im R-Paket `nlme` enthalten.

(a) Fitten Sie folgendes Modell in R:

$$y_{ijk} = \beta_0 + b_{i0} + b_{ij} + \varepsilon_{ijk}, \quad i = 1, \ldots, 8,\ j = 1, \ldots, 3\ , k = 1, \ldots, 3$$

$$b_{i0} \overset{\text{iid}}{\sim} N(0, \tau_0^2), \quad b_{ij} \overset{\text{iid}}{\sim} N(0, \tau_1^2), \quad \varepsilon_{ijk} \overset{\text{iid}}{\sim} N(0, \sigma^2), \quad b_{i0}, b_{ij},\ \varepsilon_{ijk} \text{ unabhängig}$$

Dabei ist y_{ijk} die Dicke der Oxidschicht gemessen an der k-ten Stelle an der j-ten Scheibe in der i-ten Parzelle. Lassen Sie sich die festen und die zufälligen Effekte ausgeben.

(b) Erstellen Sie einen Plot der Residuen.

Aufgabe 5 (Selbststudium)

Betrachten Sie nun die `blutdruck` Daten.

```
# Simulation der SBD Daten #
# ======================== #

set.seed(123)

Ji <- rep(6, 30)
n <- sum(Ji)
I <- length(Ji)
ID <- rep(1:I, Ji)
gender <- 1 * (ID > 15)
dosis <- rep(c(10, 20, 40, 80, 120, 160), I)
RI <- rnorm(I, 10)
RS <- rnorm(I, sd = 0.09)
eps <- rnorm(n, sd = 2)

fun <- function(x){
  -2 * sqrt(x) - x/20
}

SBD <- 150 + fun(dosis) - 2 * gender - dosis * gender/100 +
        RI[ID] + RS[ID] * dosis + eps

blutdruck <- data.frame(ID = ID, gender = gender, dosis = dosis, SBD = SBD)
```

(a) Fitten Sie ein lineares Modell, das die Abhängigkeit in den Daten ignoriert:

$$\text{SBD}_i = \beta_0 + \beta_1 \text{gender}_i + \beta_2 \text{dosis}_i + \beta_3 \text{gender}_i \text{dosis}_i \; + \varepsilon_i, \quad i = 1, \dots, 180, \;\; \varepsilon_i \text{ unabh.}$$

(b) Fitten Sie nun ein Modell mit zufälligen Intercepts:

$$\text{SBD}_{ij} = \beta_0 + \beta_1 \text{gender}_i + \beta_2 \text{dosis}_{ij} + \beta_3 \text{gender}_i \text{dosis}_{ij} + b_i + \varepsilon_{ij}, \quad i = 1, \dots, 30, \;\; j = 1, \dots, 6$$

$$b_i \overset{\text{iid}}{\sim} N(0, \tau^2), \;\; \varepsilon_{ij} \overset{\text{iid}}{\sim} N(0, \sigma^2), \quad b_i, \;\varepsilon_{ij} \text{ unabhängig}$$

(c) Fügen Sie dem Modell aus (b) zufällige Steigungen über die Dosis hinzu. Lassen Sie sich die geschätzte Kovarianzmatrix der festen sowie die der zufälligen Effekte ausgeben.