Aufgabe 1:

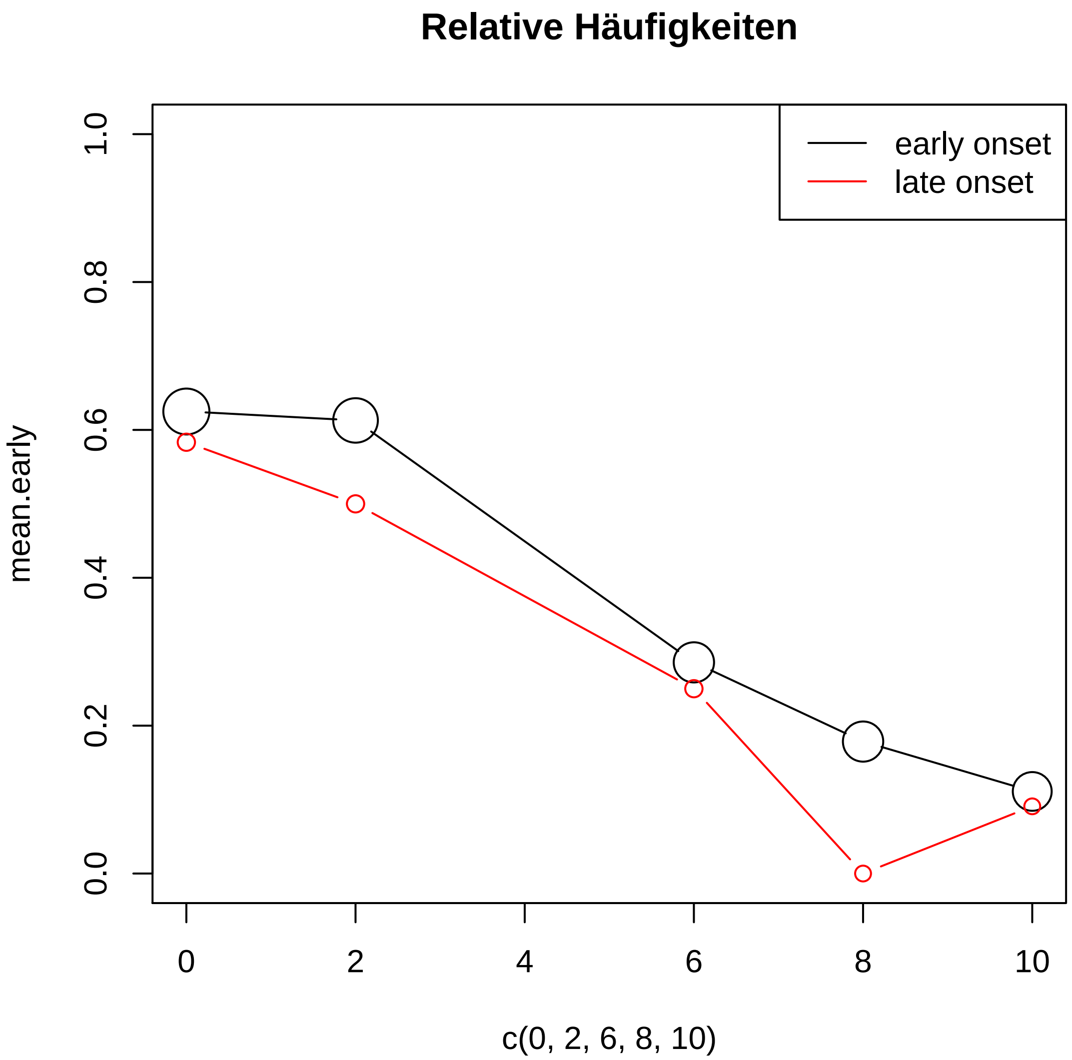
Der Datensatz `schizo.txt` enthält die Krankheitsverläufe von 44 an Schizophrenie erkrankten Frauen. Erhoben wurde das Vorhandensein von "thought disorder" (Verwirrtheit) zu Beginn des stationären Aufenthalts und nach 2, 6, 8, und 10 Monaten stationären Aufenthalts. Von Interesse ist hier die Frage ob sich der Krankheitsverlauf von Patientinnen, bei denen die Krankheit bereits in jungen Jahren (< 20) akut wurde, von dem der Patientinnen unterscheidet, bei denen die Krankheit erst später akut wurde. Benutzen Sie den folgenden Code um den Datensatz einzulesen:

```
R>url <- "http://www.statistik.lmu.de/institut/lehrstuhl/semwiso/longitudinal_ss08/"
R>schizo <- read.table(paste(url, "download/fev1.txt", sep = ""),
+     header = T)
```

(a) Berechnen Sie für die beiden Patientengruppen jeweils die relativen Häufigkeiten des Auftretens von "thought disorder" im zeitlichen Verlauf. Plotten Sie die Verläufe.

Lösung:

```
R>(mean.early <- with(subset(schizo, onset == "< 20 yrs"), tapply(unclass(disorder) -
+     1, month, mean, na.rm = T)))

        0         2         6         8        10
0.6250000 0.6129032 0.2857143 0.1785714 0.1111111

R>(n.early <- with(subset(schizo, onset == "< 20 yrs" & !is.na(disorder)),
+     tapply(disorder, month, length)))

 0  2  6  8 10
32 31 28 28 27

R>(mean.late <- with(subset(schizo, onset == "> 20 yrs"), tapply(unclass(disorder) -
+     1, month, mean, na.rm = T)))

        0         2         6         8        10
0.5833333 0.5000000 0.2500000 0.0000000 0.0909091

R>(n.late <- with(subset(schizo, onset == "> 20 yrs" & !is.na(disorder)),
+     tapply(disorder, month, length)))

 0  2  6  8 10
12 12 12 11 11

R>plot(c(0, 2, 6, 8, 10), mean.early, type = "b", ylim = c(0, 1),
+     cex = n.early/10, main = "Relative Häufigkeiten")
R>lines(c(0, 2, 6, 8, 10), mean.late, type = "b", cex = n.late/10,
+     col = 2)
R>legend("topright", col = 1:2, lwd = 1, legend = c("early onset",
+     "late onset"))
```



Größe der Punkte proportional zur Anzahl der eingehenden Beobachtungen.

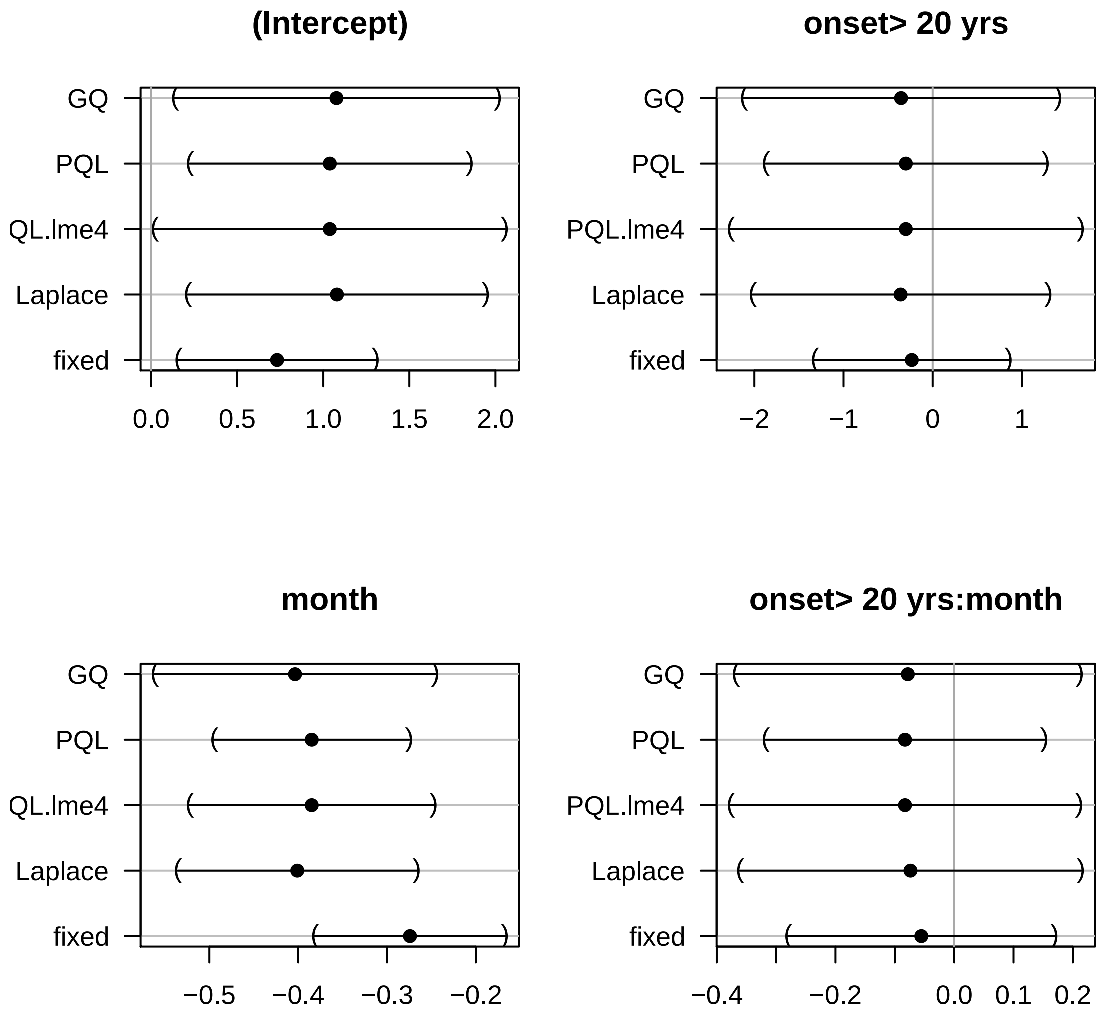
(b) Berechnen Sie ein gewöhnliches GLM und vergleichen Sie mit Schätzungen mit zusätzlichem zufälligem Personeneffekt über Laplace-Approximation, PQL und Gauss-(Hermite)-Quadratur.

Hinweis: Benutzen Sie z.B. `lmer()` aus dem `lme4`-Paket, `glmmPQL()` aus dem `MASS`-Paket und `glmmMl()` aus dem `glmmMl`-Paket.

Lösung:

```
R>library(lme4)
R>library(MASS)
R>library(glmmML)
R>glm.fixed <- glm(disorder ~ onset * month, family = binomial,
+     data = schizo)
R>glmm.laplace <- lmer(disorder ~ onset * month + (1 | subject),
+     family = binomial, data = schizo, scale = F)
R>glmm.PQL <- lmer(disorder ~ onset * month + (1 | subject), family = binomial,
+     data = schizo, method = "PQL")
```

```
R>glmm.PQL2 <- glmmPQL(disorder ~ onset * month, random = ~1 |
+     subject, family = binomial, data = schizo)

iteration 1
iteration 2
iteration 3
iteration 4
iteration 5

R>glmm.GQ5 <- glmmML(I(unclass(disorder) - 1) ~ onset * month,
+     cluster = subject, family = binomial, data = schizo, method = "ghq",
+     n.points = 5)
R>glmm.GQ30 <- glmmML(disorder ~ onset * month, cluster = subject,
+     family = binomial, data = schizo, method = "ghq", n.points = 30)
R>coef.list <- list(fixed = summary(glm.fixed)$coefficients[, 1:2],
+     Laplace = summary(glmm.laplace)@coefs[, 1:2], PQL.lme4 = summary(glmm.PQL)@coefs[,
+         1:2], PQL = summary(glmm.PQL2)$tTable[, 1:2], GQ = cbind(glmm.GQ30$coefficients,
+         glmm.GQ30$coef.sd))
R>plot.coef <- function(coef.list, se.factor = 1.96) {
+     m <- length(coef.list)
+     p <- sapply(coef.list, NROW)
+     if (min(p) == max(p))
+         p <- p[1]
+     else stop("different no. of parameters in models")
+     par(mfrow = c(ceiling(p/2), 2))
+     nam <- rownames(coef.list[[1]])
+     tab <- matrix(NA, nrow = m, ncol = 3)
+     for (i in 1:p) {
+         for (j in 1:m) {
+             tab[j, 2] <- coef.list[[j]][i, 1]
+             se <- coef.list[[j]][i, 2]
+             tab[j, c(1, 3)] <- tab[j, 2] + c(-1, 1) * se.factor *
+                 se
+         }
+         plot(tab[, 2], 1:m, main = nam[i], col = "white", xlim = range(tab),
+             yaxt = "n", xlab = "", ylab = "")
+         abline(h = 1:m, col = "grey")
+         abline(v = 0, col = "darkgrey")
+         points(tab[, 2], 1:m, pch = 19)
+         segments(tab[, 1], 1:m, tab[, 3], 1:m)
+         points(tab[, 1], 1:m, pch = "(")
+         points(tab[, 3], 1:m, pch = ")")
+         axis(2, at = 1:m, labels = names(coef.list), las = 2)
+     }
+ }
R>plot.coef(coef.list)
```



Der Plot zeigt approx. 95%-Konfidenzintervalle für die Schätzer der festen Effekte. ⇒ ohne personenspezifischen Intercept wird Variabilität unterschätzt. Unterschiede zwischen Schätzalgorithmen der verschiedenen GLMMMs im Bezug auf feste Effekte relativ gering.

```
R>deviance(glmm.PQL)

[1] 206.1479

R>deviance(glmm.laplace)

[1] 203.4931

R>deviance(glmm.GQ30)

[1] 202.3941
```

⇒ beste Modell-Anpassung (kleinste Devianz) für aufwändigste Approximation an das Integral (Gauss-Quadratur mit 30 Punkten), schlechteste für PQL.