

### Aufgabe 1:

In dieser Aufgabe sollen Sie sich mit dem Ratten-Datensatz vertraut machen. Lesen Sie sich dazu zunächst die auf der der Veranstaltungshomepage verlinkte Beschreibung des Datensatzes durch.

- (a) Lesen Sie den Datensatz in R ein. Wandeln Sie `GROUP` in eine `factor`-Variable mit entsprechenden Labels für die 3 Experimentalgruppen um.
- (b) Benutzen Sie `reshape` um den Datensatz so umzuformatieren, dass pro Tier nur noch eine Zeile im Datensatz steht. Benutzen Sie den umformatierten Datensatz `rats.wide` um sich einen Überblick über die Ausfallraten in den verschiedenen Experimentalgruppen zu verschaffen. Wie viele vollständig beobachtete Tiere gibt es in den einzelnen Experimentalgruppen, wie viele Tiere mit nur 5, 4, 3 etc. Beobachtungen?
- (c) Laden Sie das `nlme`-Paket. Legen Sie, ausgehend von den Original-Daten, einen neuen Datensatz `rats2` im `groupedData`-Format an, der nur die Tiere mit mindestens 3 Beobachtungen enthält. Plotten Sie die Verläufe der einzelnen Tiere getrennt nach Experimentalgruppen.
- (d) Fitten Sie ein lineares Modell auf dem Datensatz der kompletten Beobachtungen und plotten Sie die Verläufe der Residuen für die einzelnen Tiere.

```
R>rats$SUBJECT <- factor(rats$SUBJECT)
R>rats.complete <- rats[complete.cases(rats), ]
R>lm1 <- lm(RESPONSE ~ TIME, data = rats.complete)
R>xyplot(resid(lm1) ~ rats.complete$TIME | rats.complete$SUBJECT,
+       panel = function(x, y) {
+         panel.points(x, y)
+         panel.lines(x, y)
+         panel.lines(x = x, y = 0, col = "black")
+       })
```

Was fällt ihnen auf? Sind die Annahmen des gewöhnlichen linearen Modells erfüllt?

- (e) Benutzen Sie `lmList` um für jedes Tier ein separates lineares Modell mit zeitlichem Trend  $I(\text{TIME} - 45)$  anzupassen und betrachten sie den Residuenplot. Warum macht es Sinn die Zeit-Variable so zu transformieren? Was fällt am Residuenplot auf?

```
R>lmList1 <- lmList(RESPONSE ~ I(TIME - 45), rats2, na.action = na.omit)
R>plot(fitted(lmList1), resid(lmList1))
R>abline(h = 0, col = "grey")
R>o <- order(fitted(lmList1))
R>lines(lowess(fitted(lmList1)[o], resid(lmList1)[o]),
```

```

+     col = 2)
R>complete <- complete.cases(rats2)
R>xyplot(resid(lmList1) ~ rats2$TIME[complete] | rats2$SUBJECT[complete],
+     panel = function(x, y) {
+         panel.points(x, y)
+         panel.lines(x, y)
+         panel.lines(x = x, y = 0, col = "black")
+     })

```

- (f) Die Verläufe sind wohl nicht linear. Legen Sie eine Variable  $\log T = \log(1 + (\text{TIME} - 45)/10)$  an. Benutzen Sie wieder `lmList` um für jedes Tier ein separates lineares Modell mit zeitlichem Trend anzupassen, überprüfen Sie den Residuenplot (s.o.) und visualisieren Sie die Parameterschätzungen mit den folgenden Befehlen:

```
R>plot(intervals(lmList - Objekt))
```

Interpretieren Sie die Plots.

- (g) Untersuchen sie die subjektspezifischen  $R_{adj}^2$  und vergleichen Sie diese mit einem Modell mit quadratischem Trend in  $\log T$ .

*Hinweis:*

Betrachten sie zunächst `str(summary(lmList1[[1]]))` um herauszufinden wo  $R_{adj}^2$  abgelegt ist. Dann könnte `lapply()` weiterhelfen.

Was ist bei den gegebenen Daten problematisch an den quadratischen Fits?