Florian Wickelmaier
Forschungsmethoden und Mathematische Psychologie
Fachbereich Psychologie
Universität Tübingen

Statistik III/Advanced Statistics
SS 2026

Exercises

Exercise 25. This exercise demonstrates the application of generalized linear mixed models in the context of a longitudinal design. It considers the data of 90 patients who were hospitalized for first-episode schizophrenia at the Government General Hospital in Madras, India, during the period 1981-1982. The *Madras Longitudinal Schizophrenia Study* (Thara et al., 1994) investigates the presence or absence of (positive or negative) symptoms through monthly assessments during the first year after hospitalization.

The file `madras-data.txt` contains the following columns:

variable	descriptor
`id`	patient ID
`y`	symptom indicator (1 schizophrenic symptoms present; 0 otherwise)
`month`	months since hospitalization
`age`	age at onset (1 age less than 20 years; 0 otherwise)
`gender`	gender (1 female; 0 male)

a) The description of the data by a base model (`m1`) with random intercept leads to the following model equations:

$$\begin{aligned}
&\text{(level 1)} && g(E(y_{ij}) \mid \boldsymbol{\upsilon}_i)) &&= b_{0i} + b_{1i}\, t_{ij} \\
&\text{(level 2)} && b_{0i} &&= \beta_0 + \upsilon_{0i} \\
& && b_{1i} &&= \beta_1 \\
&\text{(2) in (1)} && g(E(y_{ij}) \mid \boldsymbol{\upsilon}_i)) &&= \beta_0 + \beta_1\, t_{ij} + \upsilon_{0i}
\end{aligned}$$

where y_{ij} denotes the j-th observation of patient i and t_{ij} denotes respective time point.

Specify the model equations for a generalized model (`m2`) with random slopes.

b) Estimate parameters of both models `m1` and `m2` using the `glmer()` function in R and compare the models by means of a likelihood ratio test. (Note: Redefine the variable `id` as a factor).

c) Consider gender as a level-2 covariate. Formulate two models (`m3` and `m4`) that generalize Model `m2`, and in which gender affects the intercept only, or both the slope and the intercept.

d) Implement both models `m3` and `m4` in R as described above and compare them with the previously considered models. Which of the model is to be preferred in the context of likelihood ratio tests?

e) Answer the following questions regarding the parameter estimates of the model selected in the previous task:

   (a) What can be inferred from the results about the baseline values and their variability across patients?

   (b) Provide an interpretation of the resulting correlation of the random effects?

   (c) Compute the expected mean odds ratio for female vs. male patients. What is the probability π_F for the presence of schizophrenic symptoms in female patients assuming $\pi_M = 0.5$ for male patients?

f) For the selected model, calculate the mean expected probability of schizophrenic symptoms separately for male and female patients. Plot the results.

Note: An illustration of the predictions of models `m1` and `m3`, for example, can be obtained with the following R code (using packages `lme4` and `lattice`).

```
pred1 <- predict(m1, type = "response")
pred3 <- predict(m3, type = "response")
xyplot(pred1 + pred3 ~ month | id, dat, type = "l",
       xlab = "Month", ylab = "Prob. of symptoms")
```

Exercise 26. Reconsider the analysis of the data of the *Madras Longitudinal Schizophrenia Study* in the preceding exercise. The decision on the successively nested models `m1`, `m2`, `m3`, and `m4` was based on likelihood ratio tests. Is this decision backed up by the results of either or both AIC and BIC?