Florian Wickelmaier
Forschungsmethoden und Mathematische Psychologie
Fachbereich Psychologie
Universität Tübingen

Statistik III/Advanced Statistics
SS 2026

Example Exam Problems

Problem 1. Consider the items

- Q2 My sexual thoughts and behaviors are causing problems in my life
- Q3 My desires to have sex have disrupted my daily life
- Q4 I sometimes fail to meet my commitments and responsibilities because of my sexual behaviors
- Q7 I feel that my sexual thoughts and feelings are stronger than I am
- Q8 I have to struggle to control my sexual thoughts and behavior
- Q9 I think about sex more than I would like to

selected from the 10 items of the *Sexual Compulsivity Scale* (SCS; Kalichman, & Rompa, 1995). Responses are on a four-category Likert scale (Not at all like me, Slightly like me, Mainly like me, Very much like me). The following analysis is based on an online study with sample size $N = 3215$. The data can be downloaded from:
`https://github.com/cran/difR/blob/master/data/SCS.txt`

a) Perform a principle component analysis on the correlation matrix of the six selected items. What are the eigenvalues? What are the eigenvectors?

b) How many components are to be extracted according to the Kaiser-Guttman criterion?

c) Compute the loadings matrix of a two-component solution. What is the proportion of variance explained by the solution in total?

d) Determine the varimax rotation matrix $\mathbf{M}$ for the two-component solution. Determine the rotated loadings matrix and the proportion of variance explained by each of the rotated components.

e) Assign items to rotated components via their maximum absolute correlation and provide an interpretation of the components.

Problem 2. A large-scale study investigates the dependence of the academic self-concept of students (y, total mean 0) on their achievements in mathematics (x, total mean 0). It is expected that beyond individual achievements also the average achievement of the respective class has an impact. The data can be downloaded from:
`https://www.mathpsy.uni-tuebingen.de/pw/fw/Umulti/bigfish.csv`

a) How can the analysis account for mathematics achievements at both individual and class level?

b) Set up a linear mixed model that takes into account both levels and includes random intercepts and slopes. Spell out the equations for each of the levels, and for the complete model.

c) Fit the model to the data with the `lme4` package. Interpret the parameter estimates.

Problem 3. In the course of a stationary psychotherapeutic treatment, for 34 patients their well-being is monitored in 6 weekly assessments by administering a questionnaire. Data are analyzed with a mixed model including correlated random intercepts and slopes. The model equation reads

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + \upsilon_{0i} + \upsilon_{1i} t_{ij} + \varepsilon_{ij},$$

where y_{ij} denotes the j-th observed value of person i with $t_{i1,\ldots,i6} = 0, \ldots, 5$. The data can be downloaded from:
`https://www.mathpsy.uni-tuebingen.de/pw/fw/Umulti/therapy.csv`

a) Which part of the equation characterizes the fixed effects and which part the random effects included in the model?

b) Describe how it may be statistically tested whether the random slopes are necessary for predicting the data. Specify the null hypothesis of this test.

c) Fit the model to the data with the `lme4` package.

d) Interpret the parameter estimates in the context of the application. In particular, answer the following questions:

   (a) What is the predicted average change of well-being per week?

   (b) What can be said about the dependence between individual intercepts and slopes? Compute the corresponding correlation.

Problem 4. Before and after a couples therapy husband (X_1) and wife (X_2) of 12 married couples independently fill in a questionnaire on the quality of the relationship. The data can be downloaded from:
`https://www.mathpsy.uni-tuebingen.de/pw/fw/Umulti/couples.csv`

a) Calculate the mean vectors $\bar{\mathbf{x}}_{pre}$ and $\bar{\mathbf{x}}_{post}$ and the covariance matrices $\mathbf{S}_{pre}$, $\mathbf{S}_{post}$ and $\mathbf{S}_d$ for the vector of differences $\mathbf{d} = \mathbf{x}_{post} - \mathbf{x}_{pre}$.

b) Use a multivariate test to check whether the results before and after the therapy differ ($\alpha = 0.05$).

c) What are the assumptions the test is based on?

d) Does a significant result of the multivariate test imply that the univariate tests on each of the single variables is significant? Provide reasons for your decision.

Problem 5. The organization committee of an international fair wants to investigate the impact of the variables "distance to place of residence" (D) and "gender" (G) on the variable "visit the fair" (Y). The variables may attain the following values

Y: visit ($y1$), no visit ($y0$);
D: <20 km, 20–100 km, >100 km;
G: male, female.

The data obtained from questioning 50 persons are:

y1	y0	D	G
1	6	<20	m
6	2	<20	f
1	7	20-100	m
6	2	20-100	f
1	9	>100	m
1	8	>100	f

They are to be analyzed with an additive logit-model with dummy coding (first category as reference).

a) Set up the design matrix $\mathbf{X}$. Notice that the 50 observations can be summarized into categories, and $\mathbf{X}$ need only have as many rows as categories.

b) Fit the model to the data and interpret the parameter estimates.

c) For the considered model, formulate the assumption that the distances 20–100 km and >100 km do not exhibit differential effects (compared to distance <20 km) as a linear hypothesis and test it ($\alpha = 0.05$, `car` package allowed).

Problem 6. Generalized linear mixed models (GLMM) offer a framework for formulating item response theories.

a) Specify the design matrices in the general model equation of a GLMM such that the Rasch model with three items as fixed effects and a single person as random effect results.

b) Extend this model by including gender as a fixed effect for the cases of a male and a female person.