Florian Wickelmaier
Forschungsmethoden und Mathematische Psychologie
Fachbereich Psychologie
Universität Tübingen

Statistik III/Advanced Statistics
SS 2026

Exercises

Exercise 23. Consider again the data of the Box (1950) experiment, and analyze the data contained in `box-weight.txt` in the long format (data frame `dat`). For hypothesis testing, assume a Type I error rate of $\alpha = 0.05$.

a) Extend the model considered in the preceding exercise, which includes group-specific as well as individual intercepts and linear time trends, by adding group-specific and individual quadratic time trends. Provide ML parameter estimates using the `lmer()` function.

b) Which of these models is to be preferred when compared by a likelihood ratio test?

c) Does the mean quadratic trend differ across groups? Formulate the null hypothesis of a likelihood ratio and run the test.

d) For each rat, plot the observed weights and the weights predicted by the preferred model as a function of time (`lattice` package allowed).

e) Interpret the parameter estimates of the fixed and random effects of the preferred model in the context of the specific application.

Exercise 24. Reconsider the data from the national survey "High School & Beyond" in U.S. public and Catholic high schools from the National Center for Education Statistics (NCES), which were already discussed in the lecture. Read the data into R from the file `hsbdataset.txt` as indicated in the lecture. Base statistical tests on $\alpha = 0.05$ and specify the corresponding null hypotheses.

a) The parameter estimates of the considered linear mixed model (Model 1) suggested a simplification, where, concerning the dependence of math achievement on `cses`, the intercept but not the slope varies across schools. Formulate the model equations for this Model 2 at both levels and derive the respective model equation.

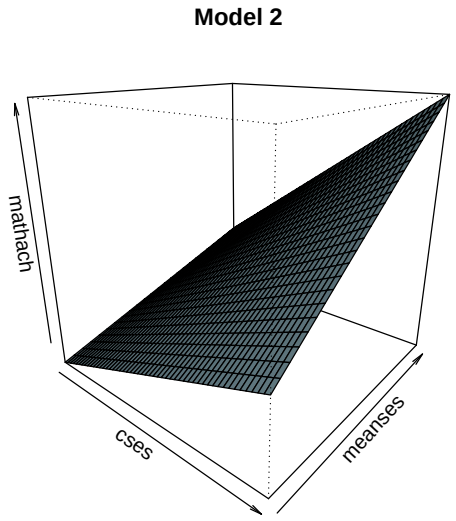
b) For Model 2, calculate the ML parameter estimates using the `lmer()` function from the `lme4` package and compare Models 1 and 2 with a likelihood ratio test.

c) Perform a likelihood ratio test with an appropriately restricted Model 3 to show that the interaction of `cses` and `meanses` contained in Model 2 is necessary for predicting the data.

d) What is the interpretation of a model prediction based on fixed effects only?

e) The following R code produces a graphical illustration of the model prediction of Model 2 for public high schools (`sector` $= 0$) based on the fixed effects.

```
> c <- seq(-2, 2, length.out = 51)
> m <- seq(-1, 1, length.out = 26)
> ndat <- expand.grid(cses = c, meanses = m)
> ndat$sector <- factor(0, levels = 0:1)
> z2 <- predict(m2, newdata = ndat, re.form = ~0) |> matrix(51)
> persp(c, m, z2, theta = 40, phi = 20, col = "lightblue",
+       ltheta = 60, shade = .9, xlab = "cses", ylab = "meanses",
+       zlab = "mathach", main = "Model 2")
```



Create a corresponding graphical illustration for the public high schools under Model 3. Comparing the illustrations, describe how the interaction effect included in Model 2 is represented graphically.