Overview

This four-lesson local teaching session consists of

1. the learn team coaching as preparation for the third central teaching block,

2. the analysis of the soy bean data. The questions below guide you through the analysis. This example is mainly meant to cover the analysis of factorial designs.

3. Individual work, e.g. on the graded hand-in exercises. The hand-in exercises cover random intercept models.

The learn team coaching may take more than one lesson this week.

1 The soy bean data

This is an adapted old exam question. The `soybean.csv` file contains data on the `yield` of soybean (in grams per plot) for 12 `geno`types, at three `loc`ations. We first ignore the effects of the three blocks for this problem.

a.) Overall, one genotype clearly produces the highest average yield. Which one, and how high is its average yield?

b.) Do you find evidence of an effect of the location, the genotype or their interaction on the average `yield`? First, give only your `R` code to answer this question, . . .

c.) . . . then name the tools you use, give p values for each effect and answer the question.

d.) A researcher claims that some genotypes yield more in one location, some in another location. Comment on this in view of c.). Justify your answer.

e.) Regardless of your result in c.), fit a model without interaction effect and then pairwisely compare all the genotypes based on this model. Give your `R` code to do all of this and report all genotypes which are significantly different from genotype `N72-3058`.

f.) According to the model from e.), how many kg of yield do we obtain for the `Centennial` genotype, grown in Plymouth, on average? Justify your answer.

g.) How high is the variance of the yield values, and what proportion of it is explained with the model from e.)? Explain your approach.

h.) Perform the required model diagnostics (assumption checks), report your `R` code and the results.

Optionally, answer questions b.) to h.) while accounting for the effect of the blocks.