**Stat 414 – Lab 1**

**Due Tuesday, Nov. 13**

**You may work together with one partner on this assignment and turn in one report with both names. It is your responsibility to make sure you contribute equally.**

|  |
| --- |
| Some software lessons we’ve learned   * In R,   + (1 | school:class) specifies random intercepts with school fixed and class random.   + (IQ | school:class) specifies random intercepts and slopes at the class level   + (IQ | school\class) specifies random intercepts and slopes at both the school and class levels (and a covariance between them)     - For no covariance between the intercepts and slopes use (1 | class) + (IQ | class)     - If classes are numbered uniquely, we don’t have to specify the nesting, so you can make different specifications for school and class, e.g., (IQ | class) + (1 | school)     - Similarly, a\*b species main effects and interaction but a:b is just the interaction * In JMP,   + Nest Random Coefficients will create random intercepts and slopes but you can delete one of them   + For more than 2 levels, nest in the higher level first and then nest in the two combined (e.g., select both class and school)   + If classes are numbered uniquely, we don’t have to specify the nesting, so you can make different specifications for school and class |

Large scale alteration (e.g., destruction) of native prairie communities has been associated with numerous problems (e.g., soil erosion, lack of biodiversity of plants, increase in atmospheric CO2). This has led to an increase in prairie reconstruction projects, but there has been a lot of variability in the success of these projects, even those using the same seed combinations and dispersal techniques in different years. A 3x2x2 factorial design was conducted to investigate the impact of soil type (remnant, cultivated, restored), sterilization (yes or no), and species (leadplant and cornflower) on the height on germinating plants. Each of the 12 treatments was replicated in 6 pots, for a total of 72 pots. Six seeds were planted in each pot. (OK, a few pots had more than six plants, probably because two of the microscopically small seeds stuck together when planted.) Measurements on each plant in each pot were taken at 13, 18, 23, and 28 days after planting. Plants that did not germinate are removed from the analysis (so we will restrict our study conclusions to plants that germinate!). Not all plants survived to the end of the 28th day.

(0) Remove the 154 plants that did not germinate by Day 28, but keep plants that had some growth but did not survive for the whole study (Document how you do so. You can use brute force, but what if you later find a typo…. Also a hint you are going to want to do some quality check with these data.) Make sure the remaining missing values are coded in a way R/JMP can understand.

(a) Identify the three-levels in this data set.

(b) Calculate the mean height for each plant for the existing measurements for that plant. (Document how you do so.) Examine boxplots of these plant heights across the soil type, separately for the coneflowers and the leadplants. Examine boxplots of the plant heights across sterilization, separately for the coneflowers and the leadplants. Summarize what you learn, including any differences in associations between species.

Convert the data to long format (document how you do so)

(c) Examine spaghetti plots of the plant heights across the measurements for each of the species. Is it reasonable to assume linear growth between Day 13 and Day 28? Does the initial height and/or rate of growth seem to differ between the species? Is there more variability in one species than the other?

(d) Examine spaghetti plots of the plant heights over time separately for the three types of soil, separately for each species. What do you learn?

(e) Examine spaghetti plots of the plant heights over time separately for the two levels of sterilization, separately for each species. What do you learn?

Split the data into two datasets, one with leadplants and one with coneflowers.

Focusing on just the leadplants

(f) “Center” the time variable by subtracting 13. Fit an “unconditional means” or random intercepts null model with no predictors but the three-level hierarchy. How many parameters are estimated? Provide an interpretation of each, including the variance components.

(You should have 413 obs, 107 plants, 32 pots. I can tell you a clever way to check this in JMP?)

(g) Include the centered time variable in the model, assuming linear growth, with random intercepts and slopes (at both levels). How much of the within-plant variability is explained by the linear changes over time? Interpret the fixed effects. Are either of the fixed effects statistically significant? How many variance/covariance parameters are there?

(h) Now add the sterilization and soil type variables. You will want to include interactions with the time variable. Is this model a significant improvement from the model in (g)?

(i) You are probably running into some boundary conditions with this model. One option is to simplify the model, e.g., removing some variance components. Modify the previous model so that the Level 3 intercepts are random but the Level 3 slopes are fixed. What is the practical interpretation of this modelling choice? Does this change the estimates of the fixed effects? How many parameters have you removed from the model? [This model should be more stable, and if you check shouldn’t be significantly worse.]

(j) Now add the Level 3 interaction between sterilization and soil types (including the interaction with time!). Compare this to the model that does not allow sterilization or soil type (or their interaction) to influence Day 13 measurements. Why do I make this suggestion? Is there a significant difference between these two models?

(k) This is your final model. Interpret it! (A brief summary of the important features, especially as the agree/disagree with your exploratory data analysis. What seems to maximize growth!)

(l) Reconsider adding the Level 3 interaction in (j), but for the model in (i) that still had random slopes at Level 3. Run this model; how many parameters are in this model?

Repeat for coneflowers

(k) Repeat the initial model in (j) for the coneflowers. Does the model also suggest that sterilization and soil type (or their interaction) don’t impact Day 13 measurements for this species? What other differences do you see in comparing the models for leadplants and coneflowers?

**Submit one report with both your names. Make sure the analysis is reproducible and your output is integrated with your discussion.**