Homework 4

Please follow the homework submission instructions provided on Piazza.

Due on Canvas before midnight on Sunday, December 3 2023.
Each part of the problems 5 points

 Note: KNNL refers to Kutner, Nachtsheim, Neter & Li (2005). *Applied Linear Statistical Models*, 5th Ed, McGraw-Hill. The dataset is available at
`https://users.stat.ufl.edu/~rrandles/sta4210/Rclassnotes/data/textdatasets/index.html`

1. *[The use of standard libraries is acceptable for this problem.]* Consider the dataset from KNNL problem 25.7. Assume that the model in KNNL Eq. 25.1 is applicable. The dataset is available from this url

   `http://users.stat.ufl.edu/~rrandles/sta4210/Rclassnotes/data/textdatasets/index.html`

   (a) Using the ANOVA decomposition, test whether the mean sodium content is the same in all brands sold in the metropolitan area; use $\alpha = 0.01$. State the null hypothesis, the alternative, the full mathematical expression for the test statistic, the decision rule, and the conclusion.

   (b) Using the ANOVA decomposition, estimate the mean sodium content across all brands. Use a 99 percent confidence interval.

   (c) Using the ANOVA decomposition, estimate and interpret the quantity $\frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma^2}$

   (d) Repeat (c) using Maximum Likelihood and the Restricted Maximum Likelihood approaches. Compare the results to the results of (c). Which method do you prefer, and why?

2. *[The use of standard libraries is acceptable for this problem.]* Library `lme4` contains the dataset `Penicillin`. Use `library(lme4)`, `data(Penicillin)` and `?Penicillin` to access the data. Perform an analysis of this dataset using `diameter` as the response, and assuming that both `plate` and `sample` are random.

   (a) Visualize the data, specify and fit the appropriate *additive* probability model. (Is it possible to specify a model with an interaction in this case?)

   (b) Use the ANOVA decomposition to test the null hypotheses of no between-plate and no between-sample variations, each at the confidence level 95%. Interpret your conclusions from the problem statement point of view.

   (c) Repeat (b) using Maximum Likelihood and the Restricted Maximum Likelihood approaches. Discuss the similarities and the differences. Which method do you prefer, and why?

   (d) Use model-based summaries to visually assess the appropriateness of the assumptions regarding the random effects. Interpret the results.