# STAA 554 Homework 4

## Contents

## Q1 Diet and blood coagulation.

The coagulation dataset comes from a study of blood coagulation times. Twenty-four animals were randomly assigned to four different diets and the samples were taken in a random order.

> ### (a) 2pts Plot the data.

> ### (b) 2pts
>
> Fit a fixed effects model and construct a prediction together with a 95% prediction interval for the response of a new animal assigned to diet D.

> ### (c) 4pts
>
> Now fit a random effects model using REML. A new animal is assigned to diet D. Predict the blood coagulation time for this animal along with a 95% prediction interval.

### (d) 4pts

A new diet is given to a new animal. Predict the blood coagulation time for this animal along with a 95% prediction interval

### (e) 4pts

A new diet is given to the first animal in the dataset. Predict the blood coagulation time for this animal with a prediction interval. You may assume that the effects of the initial diet for this animal have washed out. (Open for discussion – how to get at the animal effect for this prediction?)

## Q2 Lawnmowers and cutoff times.

Data on the cutoff times of lawnmowers may be found in the dataset lawn. Three machines were randomly selected from those produced by manufacturers A and B. Each machine was tested twice at low speed and high speed.

### (a) 2pts

Make plots of the data and comment.

### (b) 2pts

Fit a fixed effects model for the cutoff time response using just the main effects of the three predictors. Explain why not all effects can be estimated.

### (c) 2pts

Fit a mixed effects model with manufacturer and speed as main effects along with their interaction; include machine as a random effect.
- If the same machine were tested at the same speed, what would be the SD of the times observed?
- If different machines were sampled from the same manufacturer and tested at the same speed once only, what would be the SD of the times observed?

### (d) 3 pts

Test whether the interaction term of the model can be removed. If so, go on to test the two main fixed effects terms.

### (e) 2pts

Test whether there is any variation between machines. Use the model from part (d) with non-significant terms removed.

### (f) 2pts

Fit a model with speed as the only fixed effect and manufacturer as a random effect with machines also as a random effect nested within manufacturer. Compare the variability between machines with the variability between manufacturers.

### (g) 2pts

Construct bootstrap confidence intervals for the terms of the previous model. Discuss whether the variability can be ascribed solely to manufacturers or to machines. *hint: the confint() command allows a variety of method options.*

## Q3 5 pts Normality assumption of random effects

At the end of the School Performance Lab, there is a suggested way to explore the robustness of the normality assumption. Choose one of the following:
- the suggested gamma distribution

- the suggested double exponential distribution

- or a distribution of your choice

**Then:**
```
1. simulate new values for the underlying or performance of each school ("mu").
This is done in section "1. Generate true performance..."
2.  generate new data from these "mus" as in section "2. Generate observed...".
Save this data to a data set to be used in the following step.
3. Fit the empirical Bayes estimates as in section "Empirical Bayes estimates..."
```

**Include your results HERE in your submission, and ALSO on Canvas Discussion Board.**

Summarize your results on the discussion board under either the "Gamma simulation" or the "Double exponential simulation", or "Other simulation" conversation as appropriate. In one post please include:
   a. a histogram of your simulated values.
   b. a summary of the error in the random effect estimates, for example:
```
hist(coef(mod)$schools[,1] - sorted.data$mu)
sd(coef(mod)$schools[,1] - sorted.data$mu)
```
   c. Report the estimated school variance component, and the true value. (Please use standard deviation.)