# HW5 - Random Effects

Instructions: Use RMarkdown for your work. In order to receive credit for a problem, your solution must show sufficient detail so that the grader can determine how you obtained your answer.

Submit a single pdf file for your final outcome. All R code should be included, as well as all requested output. However, be sure not to include extraneous output - all code should be echoed but only include as output items requested or output that you wish to discuss in your answer. Upload your work to the Canvas course Web site.

## Problem 1 (25 pts)

```
set.seed(2)

sige = 3
siga = 4
a = 4
n = 20

alpha = rnorm(a, 0, siga)
mu = 2.5
x = runif(a*n, 0, 10)
b1 = 1.3
z = matrix(0, n*a, a)
for(i in 1:a){
  z[((i-1)*n+1):(i*n), i] = 1
}

y = mu + x * b1 +  z %*% alpha + rnorm(n*a, 0, sige)

dat = data.frame(y = y,
                 group = rep(1:a, each = n) |> as.factor(),
                 x = x)
```

### Part 1a

The above code chunk will generate a dataframe, `dat`, which contains a response, `y` (which has an overall mean of mu), and two predictors, `group` and `x`. Note that you know the truth - we have a true fixed effect $x * b1$ and a random effect *alpha* so you know the true values for these. Make a plot of the data with the response on the y-axis, the predictor x on the x-axis and the group by color or point type. Be sure your plot has all appropriate labels.

### Part 1b

Fit a linear model with all predictors as fixed effects and an intercept term. Use the *sum contrasts* option. Report your estimated coefficients. Are they all significant? Any issues with your model specification (check diagnostics)

### Part 1c

Now treat the `group` as a random effect but still include `x` as a fixed effect. Compare your predicted random effects in part c to the results you got in part b when you treated group as a fixed effect. How do they compare to the truth? (Remember we have a total of 4 random effect so report them all.)

### Part 1d

Use the Kenward-Roger approximation for an F-test to test the fixed effect. What is your conclusion?

### Part 1e

Test to see if the random effect is significant. What is your conclusion?

### Part 1f

Plot your fitted models from part b and part c on the same plot along with the observed data. Is there anything interesting?

## Question 2 (25 pts)

An experiment was conducted to optimize the manufacture of semiconductors. The semicond data (from `faraway`) has the resistance recorded on the wafer as the response. The experiment was conducted during four different time periods denoted by ET and three different wafers during each period. The position on the Wafer is a factor with levels 1 to 4. Note that while the `Wafer` variable only has three levels we don't reuse the Wafer for each time period - the variable `Grp` is a unique identifier with the form `a/b` where a is the Time (ET) and b is the wafer number.

```
data(semicond, package="faraway")
```

### Part 2a

Plot the response, resistance, for each unique wafer. Add a horizontal line for the overall mean. Do you see any patterns?

### Part 2b

Fit a fixed effects model with ET, position, and an interaction between ET and position (no other predictors). What terms are significant? Use the sum contrasts for your predictor factors (hint your intercept term should be what you plotted as your overall mean).

### Part 2c

Fit a model appropriate to the split plot design used here. (We have the same fixed effects as in part b but add in the appropriate random effect [a ET:Wafer or Grp random effect]). Report your estimated random effect variance (or standard deviation - but note which you are reporting).

### Part 2d

Test for the fixed effects. Compare the outcomes with what you obtained from the fixed effects model.

### Part 2e

Using the your final selected model from part d, test the random effects.

## Question 3 (10 pts)

Simulate a one-way ANOVA experimental design with a total of 6 levels:

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij} \alpha_i \sim N(0, \sigma_\alpha^2) \epsilon \sim N(0, \sigma^2)$$

(hint you can look at problem 1 for help). For this simulation choose $\sigma_\alpha^2$ and $\sigma^2$ so that $\sigma^2 = 5\sigma_\alpha^2$.

Create a box plot of the data. Compute your *ANOVA estimators* for $\sigma_\alpha^2$ and $\sigma^2$ and compare to your maximum likelihood estimates from the *lmer* function.