STAT 700
Homework 5 Problems
due Wed. Oct. 24

3 Problems. Show all work.
Please follow the Lab report directions off the homework web page for R Problems.
Please work in Groups!

1. Two methods, A and B, were used in determination of the latent heat of fusion of ice (Natrella, 1963). The investigators wished to find out how much the methods differed. The dataset gives the change in total heat from ice at -72 °C to water 0 °C in calories per gram of mass. It is available off the class web page:

`https://edoras.sdsu.edu/~babailey/stat700/ice.dat`

Use the Nonparametric Bootstrap Lab bootstrap function to construct a 95% percentile bootstrap interval for the difference in the population means. Use B = 1000 bootstrap replicates. It is fine to call the bootstrap function two times. Use the R `set.seed` function to set the seed. Make a histogram of the 1000 bootstrapped differences in the means. Compare your percentile interval to the one obtained by the t.test function using the var.equal=T option.

Note: You could bootstrap the difference in the medians, but let's use the means to be directly comparable to the t-test.

2. Return to enzyme kinetics. Most analyses of enzyme kinetics fit the initial velocity of the enzyme reaction as a function of the substrate concentration. In the Nonlinear Regression Lab we fit a nonlinear model to data from a biochemical experiment where the initial rate or velocity of a reaction was calculated for different concentrations of the substrate are given in the data frame `Puromycin`. We will use the "untreated" dataset (not the "treated" dataset).

It is clear from inspection of these data that velocity increases with concentration, seeming to "level off" at high concentration levels. A a standard model postulated to describe the mean relationship is the Michaelis-Menton model

$$f(x;\theta) = \frac{\theta_1 x}{x + \theta_2} \tag{1}$$

where x is concentration. Another possible model that allows for "shifting" is

$$f(x;\theta) = \theta_3 + \frac{\theta_1 x}{x + \theta_2}. \tag{2}$$

You have already fit model (1) and model (2) with nonlinear least squares.
Note that if $\theta_3 = 0$, then model (2) reduces to model (1). We can bootstrap the data to construct a percentile interval for θ_3. Use the bootstrap function from the Nonparametric Bootstrap Lab. You will have to make your own function for the theta argument.

Using 500 bootstrap replicates, bootstrap pairs to obtain the bootstrap estimates of θ_3. Make a histogram of the 500 bootstrap estimated coefficients. Use the R `set.seed` function to set the seed and use `set.seed(1)` before calling the bootstrap function. Construct a 95% percentile interval for θ_3. What is your conclusion about null the hypothesis that θ_3=0?

2. **Dyestuff Data:** (Ref: Davies, 1960) The variation of the strength of (coloring powder) of a dyestuff from one manufacturing batch to another was studied. Strength was measured by dyeing a square of cloth with a standard concentration of dyestuff and visually comparing the result with a standard. The result was numerically scored as the percentage strength of the dyestuff. Large samples were taken from six batches and from each batch six subsamples were taken. The 36 subsamples were submitted to the laboratory in a random order for testing as described above. There are two sources of variability: batch-to-batch variability and measurement error.

We return to the one-way ANOVA model,

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

where μ is the overall mean level, α_i is the random effect of the ith batch and they are iid $N(0, \sigma_\alpha^2)$ and ε_{ij} are iid $N(0, \sigma^2)$.

The file off the class web page:
`https://edoras.sdsu.edu/~babailey/stat700/dye.dat`

and you can use the read.table command with option header=T.

We will ignore the Subsample and make the Batch a factor by,

```
> dye$Batch <- as.factor(dye$Batch)
```

(a) Make strip chart of Strength by Batch. What do you notice?

(b) To test the hypothesis $H_0 : \sigma_\alpha^2 = 0$ vs $H_1 : \sigma_\alpha^2 \neq 0$, use the `lm` and `anova` function to construct the ANOVA table. Recall, the test statistic used to detect treatment effects is exactly the same as that used in the fixed effects setting.

(c) In class, it was discussed that $E(MS_E) = \sigma^2$ and $E(MS_{Tr}) = \sigma^2 + r\sigma_\alpha^2$. ($MS_E = MS_{Residuals}$). Use the Mean Squares in the ANOVA table from part (b) to show how σ^2 and σ_α^2 can be estimated from the data. Calculate these estimates.