STAT 700
Homework 6 Problems
due Wed. Oct. 26

2 Problems. Please follow the Lab report directions off the homework web page and work in HW Groups.

1. Return to Midterm Part II: We will still use the "untreated" dataset (not the "treated" dataset). To compare the nonlinear models to a linear model that is quadratic in the x-variable we consider

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + error. \tag{1}$$

(a) Fit this model with least squares using the `lm` function. Hint: There is an inhibit `I` function that will prevent R from simplifying the linear and quadratic terms. Superimpose the fitted values over the scatter plot of the data. Make diagnostic plots of the linear model fit and include the linear model summary. How well does the model fit the data?

(b) Since it is reasonable to assume the covariate concentration is fixed, we will bootstrap the residuals. Using 500 bootstrap replicates, bootstrap residuals to obtain the bootstrap estimates of the parameters $(\beta_0, \beta_1, \beta_2)$ in this model. Make a histogram of the each of the 500 bootstrap estimated coefficients.

Use the bootstrap function from the Nonparametric Bootstrap Lab and use the R `set.seed` function to set the seed and use `set.seed(6)` before calling the bootstrap function.

(c) Construct a 95% percentile interval for each of the parameters and compare your intervals with the 95% CI constructed with normality assumptions and no bootstrapping.

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 $N(0, \sigma^2)$.

To get the data, the file off the class web page:

`http://www.rohan.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) We showed in class 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.