STAT 700, Fall 2011
Homework 7 Problems
due Wednesday November 2

2 Problems. Please follow the Lab report directions off the homework web page for R Problems.

1. **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, source the file off the class web page:

```
 > dye <- read.table(‘‘http://www.rohan.sdsu.edu/~babailey/stat700/dye.dat’’,
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.

2. turn page over

2. Return (again) to housing demand data. An economist is interested in the relationship between the demand for housing (as measured by housing starts), price, and national disposabl e income.

It is available off the class web page:
`http://www.rohan.sdsu.edu/~babailey/stat700/housing.dat`

Let Y be the housing demand in appropriate units, AP be a variable representing average price, and DI be a variable representing disposable income. There are $n = 6$ observatio ns. We will consider the model,

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{\beta}' = (\beta_0\ \beta_1\ \beta_2)'$.

Assume that the ε_i are independent $N(0, \sigma^2)$ random variables.

We will use the QR decomposition to obtain the least squares estimates.

(a) Make the X matrix with a column of 1's and a column of AP and DI values so that you obtain estimates of $\boldsymbol{\beta}$.

(b) Find a Q and a R matrix such that $X = QR$. Hint: Use the help file for the R function `qr` and notice under See Also there are some functions for the reconstruction of matrices.

(c) Now use the Q and the R matrices from part (b) to obtain the least squares estimates. (They should be the same as the estimates from the `lm` function).