STAT 700
Homework 4 Problems
due Wed. Oct. 3

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

1. **Please read Lab2: Two-Way ANOVA Section**. The survival times (in hours) for animals in an experiment whose design consisted of three poisons, four treatments, and four observations per cell (Ref: Rice, 1995)
We will use data available off the class web page:

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

You can use the header information already in the file.

(a) Plot the data using strip charts and interaction plots. Describe any differences or interactions that you see.

(b) Conduct a two-way ANOVA to test the effects of the two main factors and their interactions. Give the R ANOVA table and state your conclusion. Include diagnostic plots of the residuals. Use the Bonferroni method for multiple comparisons to determine if there are significant pairwise difference among poisons and among treatments. (Make sure that a factor is a factor!)

(c) Box and Cox (1964) analyzed the reciprocals of the survival data, pointing out that the reciprocal of survival time can also be interpreted as rate of death. Conduct a two-way ANOVA, and compare your results to part (b).

2. An economist is interested in the relationship between the demand for housing (as measured by housing starts), price, and national disposable income.

It is available off the class web page:
`https://edoras.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 $n = 7$ observations. 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.

(a) In R, use matrix and vector operations to find the least estimate $\hat{\boldsymbol{\beta}}$. You will need to make the X matrix with a column of 1's and a column of AP and DI values so that you obtain estimates of $\boldsymbol{\beta}$.

Note: You can check your answer with the results from the R function `lm`.

(b) Now, we will use the QR decomposition to obtain the least squares estimates. 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. Now use the Q and the R matrices to obtain the least squares estimates. (They should be the same as the estimates from the `lm` function).

(c) Using nested models and ANOVA, test if the AP predictor can be dropped from the model. Be sure to state the null and alternative hypotheses. You should use the R function `lm` and `anova`. (See Lab 1 under the ANOVA Section.)

3. Return to previous HW: Consider the linear model from class,

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

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

$$\boldsymbol{Y} \sim N_n(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I}_n).$$

Also, assume that $\boldsymbol{X}'\boldsymbol{X}$ is invertible.

The prediction of a future observation, $Y_0 = \boldsymbol{x}_0'\boldsymbol{\beta} + \varepsilon_0$ at a given vector of independent variables $\boldsymbol{x}_0'$, is given by $\widehat{Y}_0 = \boldsymbol{x}_0'\widehat{\boldsymbol{\beta}}$.

Use the answers to the previous HW and write down the T random variable that can be used to construct a 95% CI for the mean response, $\boldsymbol{x}_0'\boldsymbol{\beta} = \mu_Y$

Then, we can set-up the probability statement: $P(-t_{df,\alpha/2} \leq T \leq t_{df,\alpha/2}) = 1 - \alpha$.

Note: You do not have to construct the CI, just derive the T random variable that can be used in the above probability statement and give the degrees of freedom!

You can look at `https://onlinecourses.science.psu.edu/stat501/node/273/` for information about a CI for the mean response.