MATH 282B – Homework 3
Due Monday, 02/15/2016, by 11:59 PM

Send your code to `math282ucsd@gmail.com`. *Follow the following format exactly. For Homework 1, in subject line write "MATH 282B (HW 1)" and nothing else in the body. There should only be one file attached, named* `hw1-lastname-firstname.R`. *Make sure your code is clean, commented and running. Keep your code simple, using packages only if really necessary. If your code does not run, include an explanation of what is going on.*

Problem 1. (Interactions with a discrete variable) Consider the `04cars` dataset. (Focus on complete cases.)

A. Fit a degree 2 polynomial explaining `mpg` (highway MPG) as a function of `hp` (horsepower). Produce a scatterplot with the fitted curve.

B. Add the factor `drive` (type of drivetrain: forward, backward, all-wheel) as an intercept. Produce a scatterplot with the points and fitted curves colored according to the type of drivetrain. Add a legend. Test whether including this new variable(s) brings a statistically significant improvement to the fit.

C. Repeat, now allowing `drive` to also interact with linear term in `hp` (in addition to the intercept).

D. Repeat, now allowing `drive` to also interact with the quadratic term in `hp` (in addition to the intercept and the linear term).

Problem 2. (Hierarchical ANOVA testing) When fitting a model on where all predictors are factors, there are different options as to how to test for the statistical significance of a particular term in the model.

A. *Sequential (aka Type I).* This is what R does when using the function `anova`. It depends on the order in which the variables are entered in the model.

B. *Hierarchical (aka Type II).* This amounts to comparing the smallest hierarchical model that contains the term with the same model with the term removed. For example, assume the variables are x_1, x_2, x_3 and that they are binary. If we want to test for the significance of $x_1x_2x_3$ we would test $y \sim x_1 + x_2 + x_3 + x_1x_2 + x_2x_3 + x_3x_1$ versus $y \sim x_1 + x_2 + x_3 + x_1x_2 + x_2x_3 + x_3x_1 + x_1x_2x_3$. If we want to test for the significance of x_2x_3 we would test $y \sim x_1 + x_2 + x_3 + x_1x_2 + x_3x_1$ versus $y \sim x_1 + x_2 + x_3 + x_1x_2 + x_3x_1 + x_2x_3$. If we want to test for the significance of x_3 we would test $y \sim x_1 + x_2 + x_1x_2$ versus $y \sim x_1 + x_2 + x_1x_2 + x_3$.

C. *Fully adjusted (aka Type III).* This amounts to comparing the full model to the model without the term. It corresponds to placing the term last and performing sequential ANOVA.

Note that a term can correspond to several variables when the factors have more than two levels. The last two types are implemented in `Anova` (recall that R is case sensitive) in the package `car`. Consider the `exmpl8.10` dataset. It is used in the following wonderful (free) book on experimental design (`http://users.stat.umn.edu/~gary/Book.html`). To read the data, use the command: `dat = read.table("exmpl8.10.txt", header=TRUE)`. Apply all three procedures to test the significance of the interaction between growth temperature and variety. (Make sure that R knows these are categorical.) Each time, provide the corresponding p-value. (This can be read from the ANOVA table and simply written as a comment in your code.)

Problem 3. (L_1 versus L_2) Let's compare least squares (or L_2) regression with least absolute (or L_1) regression in simulations. Consider a 'vanilla' linear model where $x_1, \ldots, x_n$ are iid standard normal and $y_1, \ldots, y_n$ are defined via $y_i = 3x_i + \varepsilon_i$ where $\varepsilon_1, \ldots, \varepsilon_n$ are iid with zero mean and standard deviation 0.5.

A. First, assume that the errors are normal. Repeat the following $B = 200$ times. Generate the model with $n = 100$. Fit the model by L_2 and record the estimate for the slope. Repeat with L_1. Produce a side-by-side boxplots of these slopes. Reiterate the whole thing with $n = 1000$. Offer some brief comments.

B. Next, assume that the errors are double-exponential (aka Laplace), and go over the same process.