STAT-S-670 HOMEWORK # 1

DUE WEDNESDAY 9/9/2015

Hello Class, here is the first homework assignment. If you have trouble with any of these questions don't hesitate to come and see me or my TA during office hours. For future reference, I am available most days and times and my TA's office hours are Wednesday from 3-5.

The purpose of the first two problems is to practice using R effectively. Pretend you are teaching a Calculus class and you want to give your students a graphical feeling for the gradient of a bivariate function. Let us take the bivariate function

$$z = f(x, y) = cos(x)sin(y)$$

1. Create a function in R which will produce either an image or perspective plot of the bivariate function $f(x, y)$ over a specified x and y range. Hint: You can use the function expand.grid(a, b) in R to create a grid of (x, y) values for which you can create the plot.

2. Now create some R code which would allow the user to overlay some arrows on your image plot which point in the direction of the gradient of $f(x, y)$ for a sequence of test points $\{(x_1, y_1), (x_2, y_2), \ldots (x_n, y_n)\}$ For this exercise you can use the deriv() function in R along with the expand.grid(a, b) function.

3. For the next two problems use Data from UREDA page 31, problem 2(b), 2(c).

   (a) Give stem-and-leaf displays for each of those batches of data. Choose one dataset to present your display by hand using two lines per stem (justify whether this is suggested by the rules), and the other dataset using R function using different values for "scale".

   (b) Briefly describe the distributions from your displays (shape, symmetry, outliers, etc).

4. Generate 1000 data from Gamma distributions (R: rgamma()) with two different sets of parameter values picked by you. For each of the generated datasets

   (a) Produce a plot of the exact density function using the curve() function in R.

   (b) Produce QQ plots of the generated data and interpret. Your interpretation might include symmetry, shape, heavy/light tail(s), location of the median, etc.

   (c) Produce histogram plots of the data.

   (d) Produce non-parametric density plots using the density() function for three different kernel functions. Which kernel function do you think best represents the data?