Homework 3

Due March 22, 2011

Directions Please turn-in a hard copy of your R code along with a brief write-up of the solutions (do not submit raw output). Also submit via e-mail (njc23@pitt.edu) a copy of your R code.

1. In the following data with 12 subjects, the first column is a sex indicator (1=male and 0=female) the second column is a treatment indicator (1=treatment and 0=placebo) and the remaining columns are the recorded outcome at three follow-up visits.

ID	Sex	Treatment	T.1	T.2	T.3
1	0	0	94	23	61
2	0	1	46	92	97
3	0	0	40	65	43
4	0	1	64	15	8
5	0	0	6	34	59
6	0	1	30	37	10
7	1	0	47	85	88
8	1	1	36	41	3
9	1	0	92	60	95
10	1	1	1	100	47
11	1	0	32	66	62
12	1	1	25	43	93

   (a) Enter the data into R using `scan()` and `matrix()`.

   (b) Reshape the data from wide format to long format so that the repeated measures are on separates rows.

   (c) Plot the above data using `xyplot()` in the `lattice` package. Use separate plotting symbols for treatment and plot males and females on different panels. Create a clean figure with a legend, meaningful labels and meaningful axes. (Hint: use factor variables.)

2. Answer the following questions using the function,

$$f(x, y) = \frac{-3y}{x^2 + y^2 + 1}.$$

   (a) Plot $f(x, y)$ using `filled.contour()`.

   (b) Numerically approximate the double integral,

$$\int_{-2}^{2} \int_{-2}^{0} f(x, y)\, dydx$$

   (c) Symbolically differentiate $f(x, y)$ with respect to x and y.

   (d) Find the maximum and minimum points of $f(x, y)$ using two different techniques. First using a method that does not require the gradient and a second time using a method that does require the gradient. Using your answer to part (c) which optimization function is the most convenient when using the gradient to find the optimum point?

   (e) Plot and label the maximum and minimum points from part (d) on the contour plot from part (a).

3. Posted on the course website are two Excel files `follow.up.xls` and `pred.xls` that represent data from a clinical trial with 30 subjects. The file `pred.xls` contains baseline predictor data `X1`, `X2`, and `X3`. This is information that was collected when the subject enrolled in the trial. The file `follow.up.xls` contains follow-up data `Y` collected at each follow-up visit as well as the date of the visit, `visit`. Note that not every subject had the same number of follow-up visits. Missing data is denoted by `"9999"`.

   (a) Import `follow.up.xls` and `pred.xls` into `R`. Replace missing data with `NA` and convert `visit` to an `R` Date object. (Hint: look at the arguments for `read.table()` in the documentation)

   (b) Merge the two datasets and create one dataset in long format that has 103 observations and where the data from `pred.xls` is repeated for each subject's follow-up visit.

4. Using the dataset produced in question 3, create a dataset of just the first follow-up visits for each subject and a second dataset of all of the last follow-up visits.