Statistical Computing

**Homework 4**

**Directions** Please submit a copy of your R code along with a brief write-up of the solutions.

1. Simulation of a queuing problem: a clinic has one doctor. Patients come into the clinic at random, starting at 9 a.m., according to a Poisson process with time parameter 20 minutes: that is, the time after opening at which the first patient appears follows an exponential distribution with expectation 20 minutes and then, after each patient arrives, the waiting time until the next patient is independently exponentially distributed, also with expectation 20 minutes. When a patient arrives, he or she waits until the doctor is available. The amount of time spent by the doctor with each patient is a random variable, uniformly distributed between 5 and 20 minutes. The office stops admitting new patients at 4 p.m. and closes when the last patient is through with the doctor. Simulate this process and find:

(a) How many patients came to the office?

b) How many had to wait for a doctor?

c) What was their average wait?

d) When did the office close?

2. 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 |

* 1. Enter the data into R using two methods you know.
  2. Reshape the data from wide format to long format so that the repeated measures are on separates rows.
  3. Plot the above data using xyplot() in the lattice package. Use separate plotting symbols for treatment and plot males and females on diff t panels. Create a clean fi with a legend, meaningful labels and meaningful axes. (Hint: use factor variables.)

3. Load the MASS library and take a look at the dataset called survey.

a) Find the mean pulse rate of the students. What goes wrong here?

b) Try looking at the documentation for mean() to see how to get around this.

c) The ages are recorded as fractions representing a number of months. Change that columns of the data frame so that it contains whole years (the floor() command may be useful).

d) Find the mean pulse rate for students under 20.

e) Find the mean age of students who write with their right hand.

f) What proportion of left handers do not clap with their left hand on top?

g) Using the plot() command, plot the pulse of the subjects against their age. Try subtracting 10 from the age and taking the logarithm (using the function log()), to obtain a slightly clearer picture.

4. Posted on blackboard are two Excel files follow.up.xls and pred.xls that represent data from a clinical trial with 30 subjects. The fi pred.xls contains baseline predictor data X1, X2, and X3. This is information that was collected when the subject enrolled in the trial. The fi 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".

* 1. 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)
  2. Merge the two datasets and create one dataset in long format that has 103 obser- vations and where the data from pred.xls is repeated for each subject’s follow-up visit.

(c)Using the dataset produced in part (b), 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.

5. The attitude dataset in R contains the results of a survey of clerical employees for a large financial organization. Questionnaires were completed by employees from 30 randomly selected departments. The numbers in the dataset give the percent of favorable responses to seven questions for each department. The seven questions dealt with,

# Variable Name Description

rating Overall rating (response) complaints Handling of employee complaints privileges Does not allow special privileges learning Opportunity to learn

raises Raises based on performance

critical Too critical

advance Advancement

1. Make the complaints variable a discrete variable with three levels, ”bad”, ”okay” and ”good.” Where 0-60 is ”bad”, ”61-80” is ”okay” and ”81-100” is good. For the remaining questions use this categorical complaints variable, not the continuous version.

Calculate the average overall rating for each level of complaint