

Course EPIB-613 - Introduction to Statistical Software

Assignment 2

1. Use a combination of the `rep`, `seq` and other commands to create a data frame that looks like the following:

```
      a b c d   e
1  0 0 0 0 0.01
2  0 0 0 1 0.02
3  0 0 1 0 0.03
4  0 0 1 1 0.04
5  0 1 0 0 0.05
6  0 1 0 1 0.06
7  0 1 1 0 0.07
8  0 1 1 1 0.08
9  1 0 0 0 0.09
10 1 0 0 1 0.10
11 1 0 1 0 0.11
12 1 0 1 1 0.12
13 1 1 0 0 0.13
14 1 1 0 1 0.14
15 1 1 1 0 0.15
16 1 1 1 1 0.16
```

2. Use `rbind` to create a 2 row by 16 column matrix from the data frame created in #1 above, using `a` as the first row and `e` as the last row.

3. Download the “foreign” package to your computer, and then load it into your R session. This package contains many useful functions for getting data into and out of R. Use the `read.dta` command to read the `kidney.dta` data set into R. The data set can be found on the course web page. Type the name of the data set to view its contents.

4. Use the `summary` command to summarize the `kidney.dta` data set. Note that all variables contain some missing data. Apply the `mean` and `cor` commands to determine the mean age of subjects in the data set, and also the correlation between the `creat.clear` and `creat.conc` variables. Note that you will need to use options within these commands to handle the missing data.

5. Use the appropriate R commands to create a data set that includes only complete cases from the `kidney.dta` data set. Look at your data, and verify that it has worked correctly by comparing it with the original data set with NAs.
6. Use the `write.dta` command to save the data set of complete cases as `kidneyComplete.dta`. You now have a way to take Stata data sets into R, manipulate and analyze them as needed, and save and changes back to Stata.