

Homework 2

Due February 15, 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. Create a binary operator that will concatenate two strings with a space separating the terms. Write an operator where quotations around the terms are optional. That is create an operator like `%&%` that gives the following output for quoted and/or unquoted arguments,

```
> Hello %&% World
[1] "Hello World"
> "Hello" %&% "World"
[1] "Hello World"
```

2. (a) Create an S4 class for hospital billing data. The class should resemble a dataset and have 4 slots,
 - Patient's name
 - Amount due
 - Date admitted
 - Date discharged

Include a function that will verify the components. Date admitted needs to occur before date discharged, amount due should always be positive, and every patient should have at least a first and last name.

- (b) Write an S4 show method that will print the contents of the object created in part (a)

- (c) Create a new S4 object for the following dataset

Name	Due	Admit	Discharge
Peggy Jones	1150	4/10/2008	4/11/2008
Mark Smith	9530	11/20/2006	11/25/2006
Bill Peterson	5482	8/4/2004	8/7/2004

3. (a) Write a function that calculates the geometric mean of a vector $\mathbf{x} = (x_1, \dots, x_n)$ and returns an error message if any element of \mathbf{x} is negative. The geometric mean of \mathbf{x} is,

$$\mathbf{x} = \left(\prod_{i=1}^n x_i \right)^{1/n}.$$

- (b) Apply your function to each column of the dataset given below. Some of the columns have “data entry errors” and contain negative numbers. Create a labeled output vector that gives the geometric mean for valid data and NA for columns with invalid data. (Hint use `try()`).

```
set.seed(123)
data <- matrix(rnorm(10000, mean=3), ncol=25,
               dimnames=list(NULL, paste("X", 1:25, sep=".")))

```

4. Write a function that demonstrates the central limit theorem. The function should accept at least two arguments: a function for generating random data and a sample size. Using the given random number generator and sample size simulate the sampling distribution of the sample mean. The only output the function needs to return is a histogram of the simulated distribution. In the margins of the figure give the mean and variance of the simulated sampling distribution as well as the name of the function used to generate the random data. Apply your function to these cases,
- A standard normal distribution with a sample size of 25.
 - A gamma distribution with parameters $\text{shape}=2$ and $\text{scale}=0.5$ for a sample size of 50.
5. Create the world’s most dazzling graph. Using several different plotting options create a graph that is flashy, colorful, gaudy and completely uninformative. Your graph should at a minimum include,
- Two plot regions
 - Three different plotting symbols
 - Three different line types
 - Three different colors
 - Text in the outer margin, figure margin and plot region
 - Math expression
 - Two different looking axes
 - Legend

Generate whatever data you need to get the look you want. Have fun with it!