

Homework Assignment 1  
(Due Wednesday, September 23, 2015 before class)

Please hand in a print-out of your answer and R code, and also email your R code to Zhiyuan (Jason) Xu <[xu0284@umn.edu](mailto:xu0284@umn.edu)>. Note: John Verzani's simpleR notes is available on <https://cran.r-project.org/doc/contrib/Verzani-SimpleR.pdf>

1. Problem 2.1-2.3. (p7) from John Verzani's simpleR notes. (9 points)
2. Problem 2.5. (p7-p8) from Jon John Verzani's simpleR notes. (5 points)
3. Problem 2.6. (p8) from John Verzani's simpleR notes. There is a typo in simpleR notes for this exercise. It should read: Note, we use  $X_1$  to denote the first element of  $x$  (which is 1) etc. (8 points)
4. Specify R code, using the function `rep`, to create the vector (1, 1, 1, 1, 1, 2, 2, 2, 2, 3, 3, 3, 4, 4, 5). (3 points)
5. Specify R code for pulling out the rows of the `mat` object (below) for which the fourth column is not missing and is less than 14. (4 points)  
`>mat<-matrix(c(1:13, NA, 14:15), nrow=4)`
6. Use the FAMuSS dataset (described in page 20 of Foulkes book)
  - (a) create a new data set with more than 500 rows and 4 columns that does not contain any missing values (5 points)
  - (b) Use **write.table** to save this data frame to a file. Read it in back and check the structure of the data. [Hint: use **str**] (5 points)
  - (c) Use **save** to save this data frame to a file (.RData). **load()** the file. Check the structure of the data. (5 points)
7. Describe the following terms (6 points):
  - (a) allele frequency
  - (b) heterozygous
  - (c) genotype
  - (d) haplotype
  - (e) SNP marker
  - (f) gene expression
8. Create a  $n \times m$  matrix of random numbers. Then determine how long it take to calculate the mean of each column using [Hint: use **proc.time** to track time]
  - (a) a **for** loop (4 points)
  - (b) **apply** (6 points)
9.
  - (a) Simulate a string of 10,000 characters drawn uniformly and independently from the set {A, C, G, T} [Hint: **sample**] (7 points)
  - (b) Create a frequency table of the string [Hint: **table**] (3 points)
  - (c) Write a function to create a contingency table of adjacent  $k$ -tuples. For example, with  $k=3$  and with the string "CAGACAAAAC", you would want to produce the following table: [Hint: **paste(, collapse=""**)] (20 points)
 

AAA	AAC	ACA	AGA	CAA	CAG	GAC
2	1	1	1	1	1	1
10.  $x! = 1 \times 2 \times 3 \dots \times x$ ;  $0! = 1$ .  $x$  is an integer  $\geq 0$ . Write your own function to perform the calculation. (10 points) [Do not use the function **prod** and **factorial** in R]