This is a **closed-note**, **closed-script**, and **closed-internet** exam. That is, students are not allowed to reference their course notes, any previously developed R script, or any internet site while completing the exam. Students should only use the internet to download and upload the midterm documents from the Sakai course site. At any other point during the exam, internet browsers should be closed. The only R script that should be open in R Studio is the one that accompanies the exam. Students are permitted to use the R Help Pages within R Studio during the exam. Failure to comply with these rules will result in a score of zero on the exam and being reported to the B&B MB Program for an Honor Code Violation.

* Show ALL work! Partial credit will be given.
* Please save your responses, R code, and R output in this word document. Save the file using the naming convention below, and upload to your Sakai Drop Box:
  + LastName\_FirstName\_Midterm.docx
* By typing your name below, you are agreeing to abide by the Duke Honor Code.
  + **Name:**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. Consider the R object THING printed below:

> THING

$n

[,1] [,2] [,3]

[1,] 2 5 0

[2,] 3 10 7

$s

[1] "aa" "bb" "cc" "dd" "ee"

$b

[1] TRUE FALSE TRUE FALSE FALSE

* 1. What type of object is THING? Explain your choice. (5 pts)
  2. Suppose the following R commands were submitted to the processor:

THING[[2]][1] <- "ta"; THING[[2]]

What would be returned to the console? If output, please provide the expected output. If an error, explain what the error would be. Note: If the command would execute, but may return a warning, you do NOT need to list the warning, just provide the output that would be returned to the console. (5 pts)

* 1. What type of variables are stored in the object named b in THING? If possible, state the mode of the object. If not, explain why there is not enough information to determine the mode of the object. (5 pts)

1. Suppose a wild life biologist is planning to study the different species of song-birds present at a local conservatory. The researcher collects birds by setting a small, no-harm trap that is big enough for a single bird. The researcher is able to check the trap once a day to collect specimens. Suppose there are 4 species of song-birds on the conservatory with the following distribution: 50%, 20%, 20%, and 10%, respectively. The researcher would like to estimate the number of days he will need to set traps in order to obtain 10 birds of each species. It is safe to assume that he will be able to catch a bird each day.
   1. Develop R code that will estimate the number of days the researcher will need to set traps in order to obtain 10 birds of each species. Make sure that the results of the estimation are reproducible each time the R code is submitted. Please submit R code and output below. (20 pnts)
      1. Note: The R command below can be used to simulate the random selection of species that will be present in the trap on a given day: rmultinom(1,1,prob=c(0.5,0.2,0.2,0.1)). It creates a p by 1 matrix of zeros and ones where p is the length of the ‘prob’ vector supplied to the rmultinom() function and the 1 indicates the species of bird that species present in the trap on a given day.
   2. The researcher is planning to conduct this study at several conservatories around North Carolina. The number of song-bird species and their distribution may vary across the conservatories. Using the R code developed in part (a), create a function that will estimate the number of days the researcher will have to set traps in order to obtain N birds of each species for any number of species, with any distribution, for any value of N. Call this function Trap\_Days. Make sure that the results of the estimation are reproducible each time the function is run for the same set of parameters. Demonstrate that the function can produce the same results given in part (a) and can produce an estimate for a new value of N, a new number of species, and new distribution of species (the distribution of proportions must add to 1). Please submit R code and output below. (15 pnts)
2. Suppose the data frame DATA contains weekly biomarker measures for subjects enrolled in a clinical study along with their gender.

set.seed(3214)

Week1 <- rnorm(100,300,5)

Week2 <- rnorm(100,290,5)

Week3 <- c(rnorm(88,250,8),rep(NA,12))

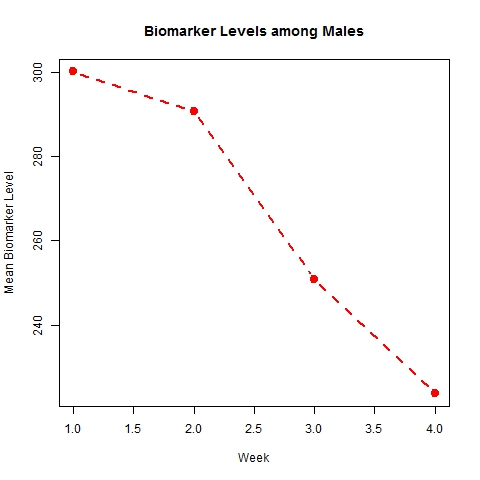
Week4 <- c(rnorm(73,225,8),rep(NA,27))

Gender <- sample(c('Female','Male'),100,replace=TRUE)

DATA <- data.frame(Week1,Week2,Week3,Week4,Gender)

Use this data frame to answer the questions below.

* 1. Create a data frame that contains the observations for all male subjects enrolled in the study. Call the data frame MALES. Please submit R code below. (5 pts)
  2. Create a 2 by 4 matrix that contains the weekly means in the first row and the weekly standard deviations in the second row for non-missing values among males. Call the matrix StatsM. Name the rows Mean and StdDev, respectively. Name the columns Week 1 … Week 4, respectively. Please submit R code and output below. (10 pnts)
  3. Using the matrix created in part (b), plot the mean biomarker level among males across the follow-up period. That is, recreate the figure given below. Please submit R code below. (25 pnts)
     1. Note: For the purposes of this problem, ignore the fact that R automatically creates a finer grid of value for the Week axis (i.e. presents options like Week 1.5). Also, there is NO NEED to get the limits of the y-axis exactly the same – you only need to ensure that no plotting data is cut off.



* 1. Using the figure created in part (c), the matrix created in part (b), and the segments() function in R, add vertical bars around each mean that represents ± one standard deviation for each week. That is, recreate the figure given below. Please submit R cope below (10 pnts)
     1. Note: For the purposes of this problem, ignore the fact that R automatically creates a finer grid of value for the Week axis (i.e. presents options like Week 1.5). Also, there is NO NEED to get the limits of the y-axis exactly the same – you only need to ensure that no plotting data is cut off.

