Introduction to Data Science

Instructor: Daniel D. Gutierrez

MODULE 4 QUIZ

Question 1

Using the `iris` data set, which of the following R code statements returns a vector of the means of the variables `Sepal.Length`, `Sepal.Width`, `Petal.Length` and `Petal.Width`?

(a) `rowMeans(iris[, 1:4])`

(b) `apply(iris[, 1:4], 1, mean)`

(c) `apply(iris, 1, mean)`

(d) `apply(iris, 2, mean)`

(e) `apply(iris[, 1:4], 2, mean)`

(f) `colMeans(iris)`

Question 2

Given the definition of the vector: `g <- c("M", "F", "F", "I", "M", "M", "F")`, which of the following R code statements will produce the list object shown below?

```
$F
[1] 2 3 7

$I
[1] 4

$M
[1] 1 5 6
```

(a) `split(3, g)`

(b) `split(1:3, g)`

(c) `split(1:7, g)`

(d) None of the above

Question 3

If you need to generate 10 random numbers based on the normal probability distribution, which of the following would you use? [HINT: remember positional vs. keyword argument passing to functions]

(a) `dnorm(10, sd=2)`

(b) `rpois(10, 2)`

(c) `rnorm(n=10, sd=2, mean=10)`

(d) `rnorm(5, mean=10)`

Question 4

If you have the vector `v <- 1:20`, which R statement would you use to take a random sample of 5 items without replacement?

(a) `sample(v, 5)`

(b) `sample(v, 10, replace=FALSE)`

(c) `sample(v, 5, replace=TRUE)`

(d) `sample(v[1:10], 5, replace=FALSE)`

Question 5

Which of the following statements is false when considering the data and time functionality in base R?

(a) Date values are represented by the `Date` class

(b) Times are stored internally as the number of seconds since January 1, 1970

(c) Time values are represented by the `POSIXlt` and `POXIXct` classes

(d) Say you have a date/time value stored in a character string: `"2015/12/8 12:04"`. You can use `as.POSIXlt()` to store the value as a variable of `Date` class