Tuan Syazana Tuan Ab Rashid

Beginner Test

1. What **class** of object is mtcars? What function did you use to find out?

“data.frame”

I used > class(mtcars)

1. Is precip defined as a **1-dimensional array** or a **vector**? How did you find out?

 It is a vector. it says “NULL” when I enter dim(precip)

1. How would you convert the **data.frame** trees into a matrix?

> data.matrix(trees, rownames.force=NA)

1. What is the name of the 14th city in the precip dataset?

 Atlanta

1. What function would you use if you wanted to combine all three data sets into a single object?

> list( )

example: jj<-list(mtcars,trees,precip)

1. Does precip consist of numeric data? How did you find out?

> is(precip, "numeric")

[1] TRUE

1. Code **four** different ways to **subscript** the **2nd row** and **7th column** of mtcars using bracket notation - i.e., 17.02.

> mtcars[2,7]

> mtcars[2,"qsec"]

> mtcars[,"qsec"][2]

1. How would you change the precipitation values of "Juneau", "Phoenix", and "Sacramento" to 23, 46, and 12 in theprecip dataset. (Hint: You will need to use **subscripts** and the <- operator).

> precip["Juneau"]<-23

> precip["Phoenix"]<-46

> precip["Sacramento"]<-12

1. Are there **any** trees in the trees dataset with more **girth** than **volume**? How did you find out?

y<-trees["Girth"]>trees["Volume"]

1. Take the sum of all elements in column **height** of the trees dataset, call this value **A**. Take the sum of all elements in row **Valiant** of the **mtcars** dataset, call this value **B**. Take the sum of the first **8 elements** of the **precip** dataset, call this value **C**. Divide **C** by **B** and add **A**. What is your final answer?

> A<-sum(trees["Height"])

> B<-sum(mtcars[6,1:11])

> C<-sum(mtcars[1:8])

> C/B+A

> 2391.59

Intermediate Test

**Section 1 Questions**

1. What does the REPLACE= argument of the sample( ) function do?

It randomly replaces elements that were true in the prior list.

1. Using as(MyMatrix,"numeric") will not convert MyMatrix to numeric data! Can you think of a property of logicals that you can use to convert the logicals to 0's and 1's other than the as( ) function?

Perform any type of mathematical operation on MyMatrix and R will convert TRUE to 1 and FALSE to 0. In this way we don’t need the as() function.

1. If you wanted to check if **all** of the elements in each row are true, how would you do this?

> apply(MyMatrix,1,all)

**Section 2 Questions**

1. How many times does the number 7 occur in MyMatrix?

> sum(MyMatrix==7)

1. How do you find the sum of each column?

> apply(MyMatrix,2,sum)

1. How do you find the product of each column?

> apply(MyMatrix,2,prod)

1. How would you change every instance of the number 10 to 12?

MyMatrix[which(MyMatrix==10)]<-12

1. How many values in MyMatrix are greater than 3 and less than 8?

> length(MyMatrix[(MyMatrix) > 3 & (MyMatrix <8)])

> 33

1. How do you change the elements of column 12 into **character data**, while keeping columns 1- 11 as numeric data??

> MyMatrix[,12]<-NA

1. Find which rows of MyMatrix have a sum >70. Make a *new* version of MyMatrix where the 13th column is a set of TRUEand FALSE values denoting which rows have a sum >70. (Hint: What type of object allows you to store both logical and numeric data at once?)

> apply(MyMatrix,1,sum)

Rows 2, 5, 6, 7 and 8.

Advanced Concepts

1. Load the iris dataset we used in the earlier tests. Write a function that takes iris as its argument, and returns three subsets of the data.frame split by the three different types of species (saved as a single object).
2. Write a function that takes iris as its argument. The function should, *for* each row, add **Sepal.Length** and**Petal.Length** *if* **Sepal.Width** is > 3.1. It should substract **Petal.Length** from **Sepal.Length** *if* **Sepal.Width** is <3.1. The answer should be returned as a vector.
3. Load the mtcars dataset we used in the earlier tests. Write a function that takes a number of cylinders as its argument. Have the function return the average miles per gallon (column **mpg**) for *all* cars with that many cylinder (column **cyl**).
4. Write a function that simulates 1,000,000 powerball drawings. A powerball drawing takes a random **sample** of 5 numbers (without replacement) from 1 through 69, plus one powerball number ranging from 1 through 26. The function should return a single object recording all of your draws.
5. Write a function that take a single set of lottery numbers (as a vector) as its **argument**. As before, write a function that simulates 1,000,000 powerball drawings. Have the function return a TRUE or FALSE value if you won any of the
6. Drawings.

Expert Test

Load in the precip dataset.

1. What is the **mean**, **median**, and **standard deviation** of precip?

> mean(precip)

[1] 34.88571

> median(precip)

[1] 36.6

> sd(precip)

[1] 13.70665

1. Is precip best visualized using a barplot( ) or hist( )? Why?

precip is best visualised using a barplot() because this gives a more accurate visual representation of the data in R. hist() doesn’t plot each unique value but rather groups the values together.

1. Generate a vector of random numbers drawn from a normal distribution with the same mean, standard deviation, and number of elements as in the precip dataset. Name this vector RandomNormal.

> length(precip)

[1] 70

> RandomNormal<-sample(precip, 70)

1. Write a function that tests, based on the means of each distribution, whether it is likely that RandomNormal and precipwere drawn from the same underlying distribution.

> ks.test(precip,RandomNormal)

Two-sample Kolmogorov-Smirnov test

data: precip and RandomNormal

D = 0, p-value = 1

alternative hypothesis: two-sided

p-value=1 >0.05 , so two vectors were drawn from identical distributions

1. Create a density( ) plot of precip and RandomNormal. Is the test you performed above (question 4) a good or bad indicator of whether the two distributions are identical? Why or why not?

> density(precip)

Call:

density.default(x = precip)

Data: precip (70 obs.); Bandwidth 'bw' = 3.848

x y

Min. :-4.544 Min. :1.666e-05

1st Qu.:16.228 1st Qu.:2.687e-03

Median :37.000 Median :9.502e-03

Mean :37.000 Mean :1.202e-02

3rd Qu.:57.772 3rd Qu.:1.649e-02

Max. :78.544 Max. :3.610e-02

> density(RandomNormal)

Call:

density.default(x = RandomNormal)

Data: RandomNormal (70 obs.); Bandwidth 'bw' = 3.848

x y

Min. :-4.544 Min. :1.666e-05

1st Qu.:16.228 1st Qu.:2.687e-03

Median :37.000 Median :9.502e-03

Mean :37.000 Mean :1.202e-02

3rd Qu.:57.772 3rd Qu.:1.649e-02

Max. :78.544 Max. :3.610e-02

Good because both tests are sensitive to locations and shape of distribution of the two samples.