



ITU

Patricia Hoffman, PhD Homework 1

Abstract

This homework uses the data from
<https://archive.ics.uci.edu/ml/machine-learning-databases/adult/adult.data>
It is described at
<https://archive.ics.uci.edu/ml/datasets/Adult>

Question 1

What does the myFun function return when it is called with the statement myFun(4)?

myFun()

```
> myNum <- 4
> myFun<- function(x){
+   myNum <- 2
+   return(myNum*yourFun(x))
+ }
> yourFun <- function(x){
+   x*myNum
+ }
```

Answer 1

Question 2

Use R to find the answer to the following set of equations:

$$3x + 2y = 12$$

$$x + y = 5$$

That is, what are the values for x and y that solve these equations.

Answer 2

Question 3

Now, write a function `solveLin()` that will accept six numbers as formal arguments: u, v, r, s along with b1 and b2. Provided an answer exists, the function should return the answer to the following set of equations:

$$ux + vy = b1$$

$$rx + sy = b2$$

That is, what are the values for x and y that solve these equations.

Check to make sure your function is correct by calling the function with the parameters given in Question 2. That is the call

`solveLin(u=3,v=2,r=1,s=1,b1=12,b2=5)`

should result in the following printed out

x = 2

y = 3

Answer 3

Question 4

Use the `adult.data` from the University of California at Irvine repository to answer this question.

4a

For this question, only consider the people making less than or equal fifty thousand dollars. What is the percent of people making fifty thousand dollars or less ($\leq 50K$) that are male? In other words, what is the fraction of males making less than equal fifty thousand dollars divided by the number of all people making less than or equal to fifty thousand dollars?

Answer 4a)

```
read.csv("C:/Dev/workspace/RProgramming/DataSets/adultData.txt")
```

4b

What is the percent of people making fifty thousand dollars or less ($\leq 50K$) that are female? (Once again the denominator should be the total number of people making less than or equal to fifty thousand dollars)

Answer 4b)

4c

Repeat questions 4a and 4b, by ignoring all the people who worked less than 40 hours per week.

Answer 4c)

4d

Only consider the people who worked at least 40 hours per week. What is the percent of people making more than fifty thousand dollars that are female? What percent of these people are male? (Here, the denominator should be the number of people who worked at least 40 hours per week and made more than fifty thousand dollars.)

Answer 4d)

4e

For this question only consider the people who worked at least 40 hours per week. Of the people with a bachelors degree, what is the percent making more than fifty thousand dollars? (i.e. (num working full time with Bachelors making more than 50 thousand) divided by (num working full time with Bachelors)) Of the people with less than 14 years of education what is the percent making more than fifty thousand dollars? Of the people with at least 16 years of education, what percent are making more than fifty thousand dollars? How many people have at least 16 years of education?

Answer 4e)

Question 5

Look at the adult.data from the University of California at Irvine repository. Please tell me an interesting fact you can tell me about this data set.

Answer 5)