READING CSV DATA IN R

By

Israel Nolazco

In this document the use of the programing language and environment R is going to use in order to perform some of the following exercises:

- Read .csv file
- Determine the number of observations in a dataset
- Calculate the mean for Age and Salaries columns
- Calculate the Standard Deviation of Age and Salaries columns

1. To read a csv file the command it is first important to change the directory. In a mac computer we can go over to MISC → CHANGE WORKING DIRECTORY. If anything, the shortcut ⌘D would also work. As seen in the image below the only file inside of that directory is the "ages-salaries.csv"

2. Much like R and many different programming languages we are going to set the name for an instance. In this case “rest_file” is the name of the instance for reading our "ages-salaries.csv" file with the command read.csv(file, header = T , sep = “,”). Technically, we did not have to use the header option since by default it is set as true and the separator is also “,”

3. To then look at the observations we can use the command “str(reference)”

4. To then calculate the mean for the Age and Salaries columns the command “mean(reference$columnName)”

5. Optionally, we could also set another instance of reference for the column. i.e “deviation_age = rest_file$age” and then we could use that reference to make our calculations. In this case we are going to find the standard deviation “sd(reference)”

```
R version 3.6.3 (2020-02-29) -- "Holding the Windsock"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin15.6.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[R.app GUI 1.70 (7735) x86_64-apple-darwin15.6.0]

[History restored from /Users/izzy/.Rapp.history]

> dir()
[1] "ages-salaries.csv"
>
> rest_file = read.csv("ages-salaries.csv", header = T , sep = ",")
>
> str(rest_file)
'data.frame':    26 obs. of  3 variables:
 $ name  : Factor w/ 26 levels "Alice","Bob",..: 1 2 3 4 5 6 7 8 9 10 ...
 $ age   : int  25 30 28 32 42 51 31 57 42 40 ...
 $ salary: int  35000 65000 70000 15000 43013 51777 100000 99999 30001 45710 ...
>
> mean(rest_file$age)
[1] 40.34615
>
> mean(reat_file$salary)
Error in mean(reat_file$salary) : object 'reat_file' not found
> mean(rest_file$salary)
[1] 70091.04
>
> deviation_age = rest_file$age
>
> deviation_salary = rest_file$salary
>
> sd(deviation_age)
[1] 12.92731
>
> sd(deviation_age)
[1] 12.92731
>
>
> sd(deviation_salary)
[1] 60344.87
>
>
> |
```