
Course: Basics of R programming language for statistical analysis

Instructor: Marina FERENT [marinaferent@gmail.com]

CHAPTER 2: CONTROL STRUCTURES AND FUNCTIONS | Statistical measures

Meeting 5: RECAP REPETITIVE STRUCTURES AND CONDITIONAL STATEMENTS |

Challenge: Median values

Exercises

PRODUCE Tasks:

1. rBasics_Meeting5.r>line123>EXERCISE POINT_1: Save locally the matrix medianValues as "campusCrimes_medianValues.csv".

[Estimated time: 5 min]

2. rBasics_Meeting5.r>lines 131-134>EXERCISE POINT_3: Add a second column to medianValues in which you store the interpretation of each median value.
e.g. medianValues[1,2]= 50% of the colleges had at most 11 990 students enrolled => the medianValues matrix should look like this:

	Median	Interpretation
enrollments	11990	50% of the colleges had at most 11 990 students enrolled
privateCollege	2	Median value does not make sense. Private college = qualitative nominal variable (private or public).*
Police	16	50% of the colleges had at most 16 police officers in the campus.
crime	187	50% of the colleges had at most 187 crimes in the campus.

*Remember from descriptive statistics: Private college = qualitative nominal variable (private or public). It does not make sense to say that 50% of the colleges were AT MOST private.

HINT: check the paste0 function -> see for e.g. rBasics_Meeting2.r -> lines 100-101
Name the new column: Interpretation

[Estimated time: 15 min]

DEBUG Tasks:

1. rBasics_Meeting5.r>lines 125-129>EXERCISE_POINT_2: Debug the following piece of code:
medianValues=matrix(NA, nrow=ncol(campusData), ncol=2)
rownames(medianValues)=colnames(campusData)
colnames(medianValues)="Median"
View(medianValues)

[Estimated time: 15 min]

REPRODUCE Tasks:

1. Find an R built-in function that computes the quartiles. Compute quartiles for <<number of enrolled students>> in <<Campus crime.csv>> dataset.

[Estimated time: 10 min]

2. Compute the quartiles (in exercise 1 above) by yourself through the usage of a conditional statement of your choice.

[Estimated time: 30 min]

3. Find an R built-in function that computes the quartiles for all the columns in a dataframe. Compute quartiles for <<Campus crime.csv>> dataset.

[Estimated time: 10 min]

4. Compute the quartiles (in exercise 3 above) by yourself through the usage of a loop and a conditional statement of your choice.

[Estimated time: 30 min]

ⁱ The data set comprises part of *campus* data from Wooldridge, Jeffrey M. (2013). *Introductory econometrics: a modern approach*. Mason, Ohio: South-Western Cengage Learning. Wooldridge Source: These data were collected by Daniel Martin, a former MSU undergraduate, for a final project. They come from the FBI Uniform Crime Reports and are for the year 1992. The original data set is available for download at:

(1) https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041 Or

(2) <https://cran.r-project.org/web/packages/wooldridge/wooldridge.pdf>