
Course: Basics of R programming language for statistical analysis

Instructor: Marina FERENT [marinaferent@gmail.com]

Meeting 7

Exercises

PRODUCE Tasks:

1. Compute the correlation matrix for all the variables in your data set (Wages.csv[i]) – R predefined function cor().

[Estimated time: 10 min]

2. Compute the correlation matrix for all the quantitative continuous variables in your data set (Wages.csv) - R predefined function cor().

[Estimated time: 10 min]

3. Using a for loop, plot the scatter plots of salary against all the other quantitative variables in your data set (Wages.csv).

[Estimated time: 20 min]

4. Provide a cross-tab analysis for variables gender and job category (Wages.csv):
   - gender = 1 male, 0 female
   - job category = 1 worker, 2 admin, 3 manager

HINT: gender = qualitative nominal variable; job category = qualitative ordinal variable => bar chart/gender OR pie chart/job category; Chi-squared + Pearson's contingency coefficient (see for example: Example_Relations between 2 qualitative variables.xlsx)

[Estimated time: 20 min]

5. Provide a cross-tab analysis for variables salary and job category (Wages.csv):
   - job category = 1 worker, 2 admin, 3 manager

HINT: salary = quantitative continuous variable; job category = qualitative ordinal variable => histogram/jobcat; ANOVA (see for example: Example_Relations between 1 quantitative and 1 qualitative variable.xlsx)

[Estimated time: 20 min]

COMMENT Tasks:

1. Comment rBasics_Meeting7_SAMY.r code.

[Estimated time: 20 min]

2. Comment rBasics_Meeting7_ALINA.r code.

[Estimated time: 20 min]

---

[i] The data set is a slightly altered version of *engin* data from *Wooldridge, Jeffrey M. (2013). Introductory econometrics: a modern approach. Mason, Ohio: South-Western Cengage Learning*. *Wooldridge Source: Thada Chaisawangwong, a former graduate student*

at MSU, obtained these data for a term project in applied econometrics. They come from the Material Requirement Planning Survey carried out in Thailand during 1998.

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

Current data set changed the definition of the gender variable and created the variable job category for instructional purposes (Manager=those employees that have higher than Q3 total experience and salary; Admin = those employees that have higher than Q2 salaries or total experience; Worker=the rest). I also dropped some of the variables of the original data set.