ELEMENTS OF STATISTICS AND PROBABILITY
R EXERCISE
Submission Deadline: Monday 3 October 2022

The data frame spain2005 contains data on 218 used flats sold in Vitoria (Spain) in 2005. The variables included are the following:

row.labels: the number of the observation

totalprice: the market total price (in euros) of the apartment including garage(s) and storage room(s)

area: the total living area of the apartment in square meters

zone: a factor indicating the neighborhood where the apartment is located with levels Z11, Z21, Z31, Z32, Z34, Z35, Z36, Z37, Z38, Z41, Z42, Z43, Z44, Z45, Z46, Z47, Z48, Z49, Z52, Z53, Z56, Z61, and Z62.

category: a factor indicating the condition of the apartment with levels 2A, 2B, 3A, 3B, 4A, 4B, and 5A. The factors are ordered so that 2A is the best and 5A is the worst.

age: age of the apartment

floor: floor on which the apartment is located

rooms: total number of rooms including bedrooms, dining room, and kitchen

out: a factor indicating the percent of the apartment exposed to the elements. The levels E100, E75, E50, and E25, correspond to complete exposure, 75% exposure, 50% exposure, and 25% exposure respectively.

conservation: is an ordered factor indicating the state of conservation of the apartment. The levels 1A, 2A, 2B, and 3A are ordered from best to worst conservation.

toilets: the number of bathrooms

garage: the number of garages

elevator: indicates the absence (0) or presence (1) of elevators.

streetcategory: an ordered factor from best to worst indicating the category of the street with levels S2, S3, S4, and S5

heating: a factor indicating the type of heating with levels 1A, 3A, 3B, and 4A which correspond to: no heating, low-standard private heating, high-standard private heating, and central heating respectively.

tras: the number of storage rooms outside of the apartment

1. Load the data from the csv file.
2. Create a table of the number of flats according to the number of garages.
3. Find the mean of totalprice according to the number of garages.
4. Create a frequency table of flats using the categories: number of garages and number of elevators.
5. Find the mean flat price (total price) for each of the cells of the table created in question (4).
6. What command will select only the flats having at least one garage?
7. Define a new file called data.c with the flats that have category="3B" and have an elevator. How many flats does the new file include?
8. Find the mean of totalprice and the mean of area using the information in (a) data.c, and (b) spain2005. Comment on your findings.
9. Create a frequency table, a piechart, and a barplot showing the number of apartments grouped by the variable out. For you, which method conveys the information best?
10. Characterize the distribution of the variable totalprice.
11. Characterize the relationship between totalprice and area.
12. Characterize the relationship between totalprice and area according to the number of toilets.
13. Find the minimum and maximum area for the intersection of apartments with one and two bathrooms.
14. Use the area values reported in (13) to create a subset of apartments that have both one and two bathrooms. By how much does an additional bathroom increase the appraised value of an apartment? Would you be willing to pay for an additional bathroom if you lived in Vitoria, Spain?
15. Create a function that given a data frame, and a vector of integers, will return the data frame sorted by the multiple columns indicated by the indices in the integer vector. Apply your function to sort the data frame created in question (14) by the market total price and total living area.

Deliverables

- A short report with comments on your findings
- Your script in an R file

Grading

Your score at the assignment will count for 20% of your final score