

R Homework for Chapter 4

The Numbeo website (www.numbeo.com) provides access to a variety of data. One table lists prices of certain items in selected cities around the world. They also report an overall cost-of-living index for each city compared to the costs of hundreds of items in New York City. For example, London at 110.69 is 10.69% more expensive than New York. In the data file `Cost_of_living_2013.txt` included are the *Cost of Living Index*, a *Rent Index*, a *Groceries Index*, a *Restaurant Price Index*, and a *Local Purchasing Power Index* that measures the ability of the average wage earner in a city to buy foods and services. All indices are measured relative to New York City, which is scored 100.

You can find the data file on Blackboard. Download it and analyze it using **R** to answer the following questions.

1. Produce a scatterplot between the *Cost of Living Index* and EACH of the other index variables. As a result, there should be 4 scatterplots in total. Examine the relationship shown in each scatterplot in terms of its *form*, *strength* and *direction*.
2. Compute the correlation coefficients for all the scatterplots obtained above.
3. Verify the conditions for EACH correlation coefficient computed above.
4. Fit a linear regression model between the *Cost of Living Index* and each of the other index variables. As a result, there should be 4 regression models in total. Interpret the resulting estimated slope in each model.
5. Based on the correlation coefficients and the regression models obtained above, which item would be the best predictor of overall cost in these cities? Which would be the worst? Explain.
6. Find the cost of living as predicted by *Groceries Index* and its residual for Beijing, China. (**Hint:** Find row index of Beijing in the dataset, and then use that index to extract the corresponding fitted value and residual from the regression result.)