Intelligent Data Analysis - Homework 1.2

Ignacio Rodriguez, Daniel Minguez

October 24th, 2018

1 Basic exploration of a dataset

1.1 Introduction

1.2 Descriptive statistics

1.2.1 Questions

a) Choose a quantitative variable and explore its distribution in terms of descriptive measures of center, dispersion, skewness and kurtosis. Is a normal model a plausible one for its distribution? If the answer is no, can you think of a transformation of the variable that improves normality. Are there any outliers?

b) Choose two quantitative variables and describe its joint bivariate distribution. Does it seem to be Normal? Are there any outliers?

c) Choose a subset of 4 or 5 quantitative variables and explore linear relationships

2 Permutation tests

2.1 Choose a subset of 4 or 5 quantitative variables and explore linear relationships

a) R matrix of pairwise correlations

b) Matrix of partial correlations

c) Coefficient of determination (function r2multv() we define in R)

d) The determinant of R (correlation matrix) as an overall measure of linear relationships.

e) An eigenanalysis of matrix R, looking for really small eigenvalues.

2.2 **Repeat the analysis deleting the values for three customers that left a tip greater than 30% of the bill. These generous customers seem to be outliers.**