Politecnico di Milano
Scuola di Ingegneria Industriale e dell'Informazione
Applied Statistics
July 3rd, 2017

First Name and Family Name:
ID Number:

Problem n.1

The file `kimono.txt` collects the value (k€) of 528 silk kimonos sold in Mitsukoshi department stores in Kyoto and Tokyo, both tailor-made and ready-to-wear.

a) Formulate an ANOVA model for the value of a kimono as a function of the factors city (Kyoto or Tokyo) and type (tailor-made, ready-to-wear). Verify the assumptions of the model.

b) Through appropriate statistical tests, propose a reduced model.

c) Provide Bonferroni intervals (global level 95%) for the differences between the mean value of kimonos belonging to the homogeneous groups identified by the model at point (b).

Problem n.2

Having picnics in parks is very common in Japan, especially for the traditional custom of Hanami (flower viewing) during the bloom of cherry blossoms. The file `bento.txt` contains the total amount [g] of rice, sashimi, vegetables and okashi (traditional sweets) in the bentō's (packed lunches) of 32 volunteer families, consumed on March 26th 2017 (for Hanami) and on May 7th 2017 (normal season). Assuming independent the composition of bentō's of different families and not independent that of the same family, answer the following question.

a) Perform a statistical test to verify if there is evidence of an impact of Hanami on the mean amount of rice, sashimi, vegetables and okashi in families bentō's. Verify the needed assumptions.

b) Provide four T^2 simultaneous confidence intervals (global confidence 95%) for the increase in the mean consumption of rice, sashimi, vegetables and okashi in correspondence of the bloom of cherry blossoms. Comment the results.

Problem n.3

The file `geisha.txt` collects data about *Geisha hunting* in Kyoto (i.e., tours finalized to spot a Geisha). The data report the duration (minutes) and starting time (in minutes after 16:00) of 130 trials (not all successful).

a) Use a hierarchical clustering method based on Euclidean distance and single linkage to identify two groups of data (i.e., successful and unsuccessful tours). Report the centers of the clusters, the size of the clusters, the cophenetic coefficient and a qualitative plot of the results.

b) Evaluate the quality of the clustering at point (a) and, in case you deem it unsatisfactory, repeat the procedure with another linkage at your choice.

c) Identify the successful tours with the smaller group found with the clustering method at point (b). Having introduced and verified the needed assumptions, provide 4 Bonferroni intervals (global level 90%) for the difference in the mean characteristics of successful and unsuccessful tours, and for the mean characteristics of a successful tour.

d) Comment the results at point (c) and suggest a successful strategy for Geisha hunting.

Problem n.4

The file `garden.txt` collects the number of carps, maple trees, cherry trees and stones, and the extension [m^2] of 156 of garden in the Kantō region of Japan. Experts believe that, to achieve an overall balance of elements, the Japanese gardens follow the model

$$E = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \beta_3 \cdot x_3 + \beta_4 \cdot x_4 + \varepsilon,$$

with E the extension of the garden, x_1, x_2, x_3, x_4 the number of carps, maple trees, cherry trees and stones respectively, and $\varepsilon \sim N(0, \sigma^2)$.

a) Estimate the 6 parameters of the model and verify the model assumptions. Evaluate the residuals of the model.

b) Perform two statistical tests to verify if

- there is statistical evidence of a dependence of the mean garden extension on the number of maple or cherry trees;
- there is statistical evidence of a dependence of the mean garden extension on lake elements (stones, carps).

c) Based on the results at point (b), comment on possible model weaknesses and, if needed, reduce the dimensionality of the regressors. Comments the analysis and interpret the results.

d) Update the estimates of the parameters using the analysis at point (c).