Politecnico di Milano
Scuola di Ingegneria Industriale e dell'Informazione
Applied Statistics
July 10th, 2018

First Name and Family Name:
ID Number:

Problem n.1

The file `Revenues.txt` contains the total monthly income of 37 beach clubs in Santa Margherita Ligure (GE), during the months of June, July, August and September 2017.

a) Perform a test of level 10% to verify if there exists any significant difference in the mean of monthly revenues along the considered period. Introduce and verify the appropriate assumptions.

b) Build 4 simultaneous confidence intervals (global level 90%) to describe the dynamic over time of the mean revenues. Comment the results.

Problem n.2

The file `Running.txt` collects the times of 80 participants to the *Portofino>Run* 2018, for the sections Santa Margherita Ligure to Paraggi and Paraggi to Portofino.

a) Use a hierarchical clustering method (Euclidean distance and single linkage) to identify two groups of participants, i.e., those who actually run and those who walked instead. Report the number of data within each cluster, the mean within the clusters and a qualitative plot of the results.

b) Evaluate the performances of the clustering method at point (a) and identify the possible issues (if any). Propose, if needed, an alternative clustering method to identify the two groups of participants (report the number of data within each cluster, the mean within the clusters and a qualitative plot of the results).

c) Provide four Bonferroni intervals (global level 95%) for the mean and the variances of the *running* times of the two sections.

Problem n.3

The files `Presales.txt` and `Sales.txt` contain the prices of flip flops and swimsuits in 45 shops of Rapallo, before and after the summer sales 2017, respectively.

a) Perform a statistical test (level 1%) to verify if the prices were discounted in mean of 20% during sales. Introduce and verify the appropriate assumptions.

b) Provide a point estimate of the mean and of the covariance matrix of the discounts on flip flops and swimsuits during the summer sales 2017.

c) Estimate an elliptical region that contains 99% of the discounts. Reports its analytical expression, its center, and the direction and length of its semi-axes.

Problem n.4

The file `Focaccia.txt` contains the amount [kg] of *focaccia* sold in a bakery in Recco during 43 days of June and July 2017. For the kilos of sold *focaccia* consider the following linear model

$$Y_g = \beta_{0,g} + \beta_{1,g} \cdot t + \epsilon,$$

with $t \in [1:61]$ the index of the day, $g = \{weekend, weekday\}$ the day of the week and $\epsilon \sim N(0, \sigma^2)$.

a) Estimate the 5 parameters of the model $\{\beta_{0,g}, \beta_{1,g}, \sigma\}$. Verify the model assumptions.

b) Perform two statistical tests – each at level 5% – to verify if

- there is a significant dependence of the mean sales on the day of the week;
- there is a significant difference between weekend and weekdays, in the *increase* of the mean sales along time.

c) Based on point (b), reduce the model and update the estimates of the parameters.

d) Perform a test of level 5% to verify if during weekends the mean amount of sold focaccia increases of 60 kg. In case, update the parameters' estimates.

e) Based on the last model, provide a point prediction of the mean sales of focaccia on the 28th July 2017 (day 61, weekday).