Politecnico di Milano
Scuola di Ingegneria Industriale e dell'Informazione
Applied Statistics
June 28th, 2018

First Name and Family Name:
ID Number:

Problem n.1

The files `Morning.txt` and `Evening.txt` contain the data about delays [min] registered in July 2017 on the flights operated by the company BeLate from Mexico City to Oaxaca and viceversa. The flights were scheduled twice per day, in the morning time (`Morning.txt`) and evening time (`Evening.txt`). Negative values indicate flights arrived earlier than scheduled. Assume the flights operated on the same day to be dependent, and the flight operated in different days to be independent.

a) Verify at level 1% if there is a significant difference in the mean delay between morning and evening flights (introduce and verify the appropriate assumptions).

b) Provide four Bonferroni intervals (global level 1%) for the mean delay in the morning flights and for the mean delay in the evening flights (introduce and verify the appropriate assumptions). Comment the results.

c) On July 28, the scheduled arrival time for the morning flight from Oaxaca to Mexico City is 10:15. Would you believe (at level 99%) to be able to take an international flight to Italy at 12:15, knowing that the gate closes at 11:45?

Problem n.2

The file `Mexican.txt` collects the prices [Mexican pesos] of 300 Mexican meals consumed in July 2017 by tourists in three areas of Mexico (near Mexico City, Cancun, and Guanajato). The dataset also reports whether the meal was based on Tacos or Fajitas.

a) Build a complete ANOVA model to characterise the price of a Mexican meal. Verify the assumptions of the model.

b) Reduce the model based on appropriate statistical tests.

c) Build Bonferroni intervals (global level 99%) for the mean differences between the price of a Mexican meal, in the groups identified at point (b).

Problema 3

The file `Precolombian.txt` collects the data about 167 Precolombian statues found in the area of Tula (Mexico) and exhibited in the National Museum of Mexico City. Some of these statues were attributed by experts to the Maya civilization, some to the Aztecs and the remaining to Toltecs. The dataset reports the dating (estimated year in which the statues were built) and the aspect ratios of the statues (ratio between their hight and width). Knowing that, out of 100 statues found in the area, on average 20 of these are attributed to Aztecs, 10 to Maya and 70 to Toltecs, answer the following questions.

a) Build a classifier for the variable civilization based on the available quantitative features. Report the mean within the groups identified by the variable civilization and a qualitative plot of the classification regions. Introduce and verify the appropriate assumptions.

b) Compute the APER of the classifier.

c) How would you classify a new statue dated 986 a.D. and with aspect ratio 1.4?

Problem n.4

The file `Hotels.txt` contains the rates for a double room in 4^* hotels at Playa del Carmen, recorded for 83 days of the year 2017. For the rates consider a linear model, accounting for the time of the year, for the seasonality (wet season or not) and for the position of the hotel with respect to the sea (seafront or not):

$$Y_g = \beta_{0,g} + \beta_{1,g} \cdot \left(1 + \cos\left(\frac{4\pi}{365}\, t\right)\right) + \epsilon,$$

with $\epsilon \sim N(0, \sigma^2)$ and g the grouping structure induced by the seasonality and by the seafront (consider a model *without interaction*).

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

b) Perform three statistical tests – each at level 1% – to verify if

- there is a significant dependence of the mean rates on the seasonality (wet season or not);
- there is a significant dependence of the mean rates on the position (seafront or not);
- there is a significant difference in the *temporal dynamic* of the mean rates along the year depending on the seasonality or on the position of the hotel.

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

d) Provide a confidence interval (level 99%) for the maximum of the mean rates of 4^* hotels at Playa del Carmen.