First Name and Family Name:
ID Number:

Problem n.1

Pigeons are rigorously monogamous animals. The criterions employed by these animals to choose their life-long partner have been widely discussed among ethologists. File `pigeons.txt` collects the weights [g] and wingspans [cm] of 134 couples of pigeons (a couple being made by a male and a female animal).

a) Perform a statistical test to verify if a significant difference exists among the characteristics of male and female animals.

b) Support the conclusions of the test at point (a) with appropriate Bonferroni intervals (global level 90%). Comment the results.

c) Ethologists believe pigeons choose their partners such that the wingspan of the male animal should be at least 20% higher than that of the female. Perform a statistical test of level 10% to verify the ethologists' hypothesis.

Problem n.2

A thermal resort in Sirmione (BS) is evaluating the effectiveness of its wellness programs. A trial based on 14 clients was made: each client tested the effects of the 11 major wellness programs offered by the resort; each test was performed independently along a year. The variation in a client well-being after a wellness program was quantified through a numerical indicator valued in $[-1, 1]$, positive values indicating that the program brought an improvement, negative otherwise. The file `wellness.txt` reports the measures collected during the trial.

a) For each wellness program, perform a one-sided permutation test aimed to test if a significant *improvement* occurs in clients' well-being. Use as test statistics the mean of the indicator, and use 5000 random permutations with random seed equal to 321 to estimate the permutational distribution. For each program, report the value of the test statistics and their corresponding p-values. Identify the effective programs using a level 1% in each test.

b) Report the wellness programs associated with a significant improvement in clients' well-being, imposing a probability at most 1% that at least one of the non-effective programs is judged as effective.

Problem n.3

The file `sansiro.txt` reports the daily weather conditions (mean temperature [°C] and precipitation [mm]) recorded on 45 rainy days in Milan, in which concerts where planned at San Siro stadium. Of the latter, 15 took place regularly, 15 were suspended, 15 cancelled.

a) Assuming that, in rainy days, on average only 3% of the representations are cancelled and 8% are suspended, build a classifier for the representations (with classes "regular", "suspend", "cancel") that minimizes the expected number of misclassifications. Introduce and (if possible) verify the appropriate assumptions; report the mean within the three groups and a qualitative plot of the regions of classification.

b) Compute the APER of the classifier.

c) How would you classify the concert which was planned for the 17th June, knowing that the mean temperature was 28°C, and the precipitation 12 mm?

d) Do you deem relevant to consider the variable *temperature* in the construction of the classifier? Answer to the question by using appropriate statistical instruments.

Problem n.4

The city of Kyoto is famous for its masterpieces in terms of kimonos. For the cost of a kimono [€], consider the following model:

$$Y = \alpha_g + \beta_g \cdot l + \gamma_g \cdot n_c + \varepsilon,$$

where l is the amount of silk [m] used for its manufacturing, n_c is the number of colors within the fabric, $g = 1, 2$ describes the quality of the manufacturing (1=prestige, 2=medium quality) and $\varepsilon \sim N(0, \sigma^2)$. Based on the data contained in `kimonos.txt` answer the following questions.

a) Estimate the 7 parameters of the model (report the estimated of $\alpha_g, \beta_g, \gamma_g$ for $g = 1, 2$ e σ^2).

b) Having verified the needed assumptions, perform a test of level 10% to verify if the factor *quality* has a significant impact on the cost of the kimono.

c) Do the number of colors have a significant impact on the mean cost of a kimono?

d) Reduce the model and update the estimates of its parameters.

e) Interpreting α_g as a fixed cost of production and the remaining terms as variable costs, provide confidence intervals of global level 90% for: the mean fixed cost for a kimono of high/medium quality, the mean total costs for a kimono which makes use of 6.5 m of silk and is of high/medium quality, made of 3 colors fabric.