Proiect de Certetare Lab4

**Language chosen:**

R has been chosen over rivaling programming languages like Python or MATLAB that also cover statistical computing and data analysis because of the numerous built-in packages to handle multivariate statistics like stats, caret, cluster, NbClust as well as the tools to display the results generated like ggplot2 or corrplot. R also offers integration with excel as well as the means of handling mixed data types like dplyr and tidyr that are present in the current scenario.

It is also worth mentioning the extensiveness of R documentation and tutorials offered by portals like rdocumentation and r-bloggers where you can find official documentation and community created guides and articles.

The main reason R has been chosen over MATLAB is the open-source aspect of R, offering state-of-the-art statistical tools without the licensing fees.

Python doesn’t fall short of functionality however R has been chosen over because of its seamless function and package integrations as well as general ease of use, leaving more time for development over package management and deployment.

**Data sets description:**

BD Oua UV 2022:

Using a spectrophotometer the light travelling through the egg shell was measured, thus obtaining the percentage of light that passes through the shell for each wavelength from 30 to 700 nm. The spectrum was split in three types of wavelengths: UV, MW and LW;

It is of interest to find out if there exist correlations between the amount of light that passes through the shell and the egg's properties: membrane, colour, habitat, incubation period and the type of species. For each species there have been done five measurements.

BD Gall Rall Cot Egg submersion:

An experiment has been done on three species of birds: Rallus aquaticus, Coturnix coturnix and Gallinula galeata. Their eggs have been split into three groups: normal hatching, submersed for two hours and for three hours. The submersion was done before incubation;

It is of interest to find the difference between species and the resistances to the embryos being inundated.

**Data sets fields:**

BD Gall Rall Cot Egg submersion:

Specie: Abbreviated name of the bird's species (nonnumeric);

> Method: Method used to conclude hatching flag. Normal(m), Two hours(2), Three hours(3) (nonnumeric);

> Hatching: Egg hatched(1) or not(0) (numeric)

BD Oua UV 2022:

Species: Abbreviated name of the bird's species;

Membrana: Presence of egg membrane;

Light\_UV: Average percentage of UV light that penetrates the egg shell (numeric);

Light\_MW: Average percentage of MW light that penetrates the egg shell (numeric);

Light\_LW: Average percentage of LW light that penetrates the egg shell (numeric);

Specia: Full name of the bird's species (nonnumeric);

Altricial/Precocial: Species group. Altricial(A), Precocial(P) (nonnumeric);

Incubation: Range of incubation time in weeks (nonnumeric);

Egg\_length: Length of the egg in mm (numeric);

Egg\_width: Width of the egg in mm (numeric);

Habitat: Dry habitat(D), Open habitat(O), Cultivated habitat (C) (nonnumeric);

Egg\_spots: Presence of egg spots (numeric);

Colors: Blue(B), White(W), Blue-green(BG) (nonnumeric);

mean\_Inc\_per: Average incubation time in weeks (numeric).

Eggshells2021\_MedianTrans10nm:

Spec: Abbreviated name of the bird's species (nonnumeric);

Membrana: Presence of egg membrane (numeric);

x-<y fields: Percentage of light that penetrates the eggshell on the (x, y) range in nm with a 10nm step from 300nm to 750nm (numeric);

100%: Sum of all the percentages on that line (numeric).

A - Z nm:

B: Abbreviated name of the bird's species (nonnumeric);

x-<y fields: Percentage of light that penetrates the egg shell on the (x, y) range in nm with a 10nm step from A nm to Z nm (numeric);

suma: Sum of all the percentages on that line (numeric);

index: The sum divided by the 100% field (numeric);

100%: Identical to the 100% field in Eggshells2021\_MedianTrans10nm (numeric).

**Source code:**

****

****

****

****

****

**Github history:**