Statistical analysis: rice phenotype

Houda Aiboud Benchekroun and Thomas Roiseux (group 8)

2022-11-19

General information

Genetic data hold important information about living beings. These information can be used to understand and even predict the ability of a species, or an individual, to protect itself from a specific disease. Using this and statistical tools, we may be able to find some predispositions for individuals of a population to have or a not specific disease, and by making predictions, to find the most suitable gene that will grant the best protection against several problems.

To explore this, we will use a data set made with genetic information of rice. These have been extracted from the UK Bio-Bank. As rice is one of the basics in food in several countries, it production must be protected to avoid any starvation in these countries. That's why studying its genetic data to protect it is quite important nowadays.

We will then try to find to understand which variables of the genetics data are useful to understand and explain the resistance of rice to certain diseases.

But, before that, we are going to list in this paper all the variables we have and make a first analysis to determine which are target and which will be the explanation for the targets.

Variables available