Study of fungal invasion of different varieties of apples by different fungal strains

Introduction

Apples are among the best-preserved fruits. However, after a while fungi strains are inevitably starting to grow. In F.G. Gregory and A.S.Horne study, the conditions of infection, the progression of the invasion and the final stage of invasion was studied for two varieties of apple : the Cox’s Orange Pippin from Burwell, Cambridgeshire and Bramley’s seeding from 6 different localities. The dataset contains the variety of apple. the fusarium strains, they day at which the apples have been infected, the temperature of storage of the apples, the apple weight, the apple radius, the fungal radial advance and the rate of advance. The aim of our study is to compare the infecting power depending on the conditions of infection (strain of fusarium, day of infection and storage temperature) and the resistance to invasion of different varieties of apples and their characteristics (weight and radius).

Explanatory analysis

To begin with, we want to get familiar with the data. The storage temperature data was embedded in the variety and days parameters. That’s why, in order to simplify the analysis, we created a variable temperature in the dataset.

Then we observe that for two apples, the fungal radial advance is bigger than the radius of the apple. This doesn’t make any physical sense. Therefore, we chose to remove them from the dataset.

Moreover the sample 17 creates a very high variance for day 54 and variety 3 for no obvious reason. So we decided to remove it from the dataset.

After explanatory data analysis, we then decided to remove three outliers from the dataset.

+ analysis of collinearities

Model selection for the study of the rate of advance

We start our model selection with a formula with all the variables at the first order. Strain + variety+days+weight+radius+temperature.

Using the F-test on the analysis of variance of this model, we can see that the variables strain and days are the most significative in the modelling of the fungal rate of advance and that the temperature could also be included in the model.

In order to select the more relevant variables to describe the fungal rate of advance, we use a stepwise method. The selection of our model is based on the

ANOVA tells us to keep the radius more than the temperature while the step wise method tells us to keep only strain, days and temperature. Do you have an idea on how to explain that properly ? We chose the stepwise model so are we talking about anova at all (as it is the name of our subject?)