ANOVA

1 Introduction

During this project, we're going to analyze the process of invasion of varieties of apples by differents strains of funghis. This was the aim of the study done in the article [1]. The method used in the experiences gave us the radial advance, therefore the rate of advance of the funghis.

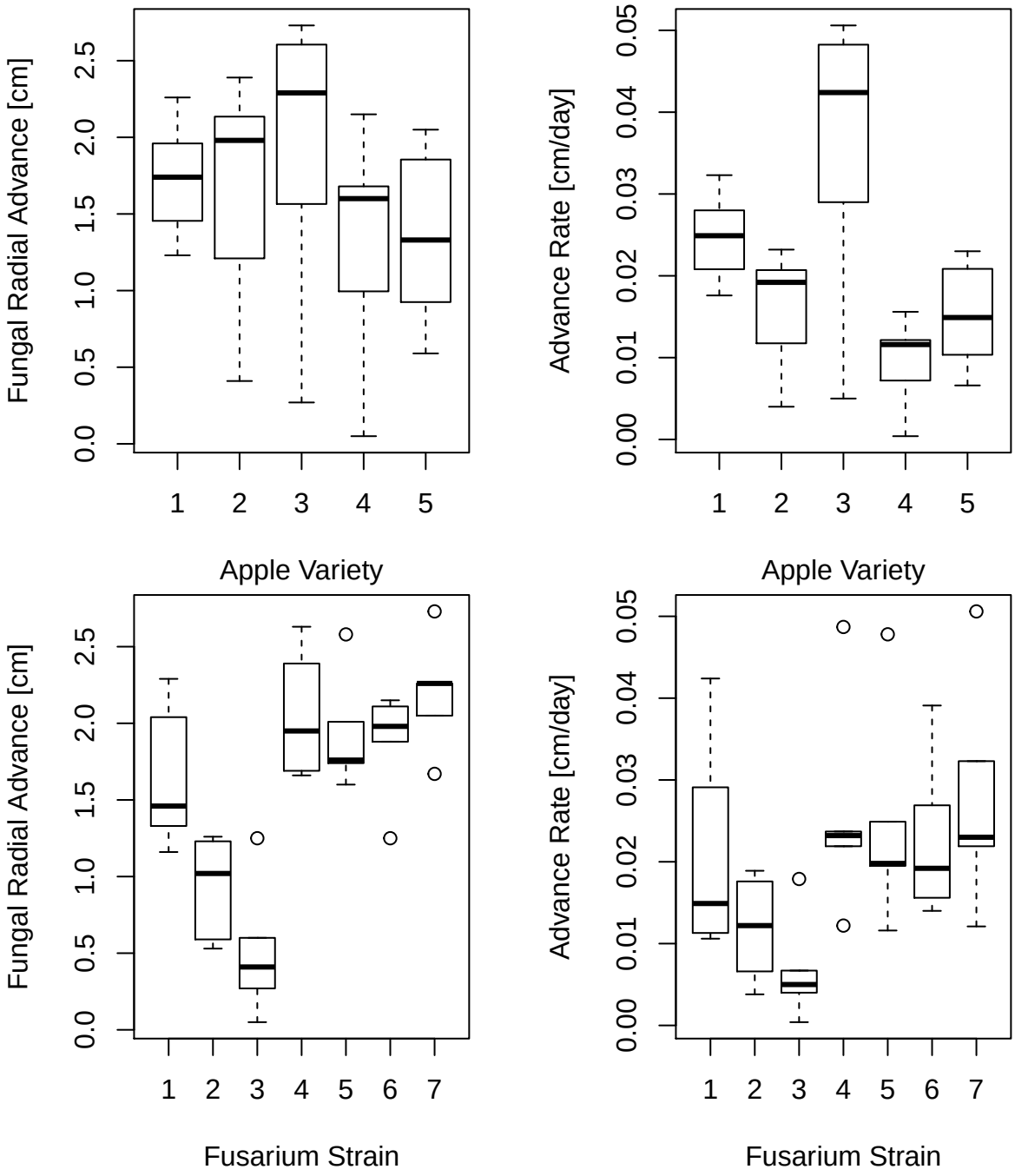
2 Exploring the data

The experience has been done for each variety of apple for different types of funghis. There's 5 varieties of apples and seven strains of funghis. But each one of these didn't have the same day of exposure and the 35 observations has been done on apples with different weight and radius. The parameters for the experience are then : variety, type of strain, number of days of exposure, weight and the radius. And the results are the radial advance and the rate of the fusarium. As first, to explore more the data, we plot the radial advance function of the variety and then the advance function of the strain. We obtain the results depicted in figure ANALYSER.

And after analyzing the correlation coefficients that the program gives us, our conclusions are :

- *radius* and *weight* are really low correlated to the rate
- the correlation between the factor change significantly with the *varieties* and the *strains*
- $advance = rate \cdot days$ so we decide to focus only on the data rate.

Now that we have an idea of the dependency of the factors with the results we try to find the right model that fits our data correctly. In order to do that we use the Analysis of Variance (ANOVA).

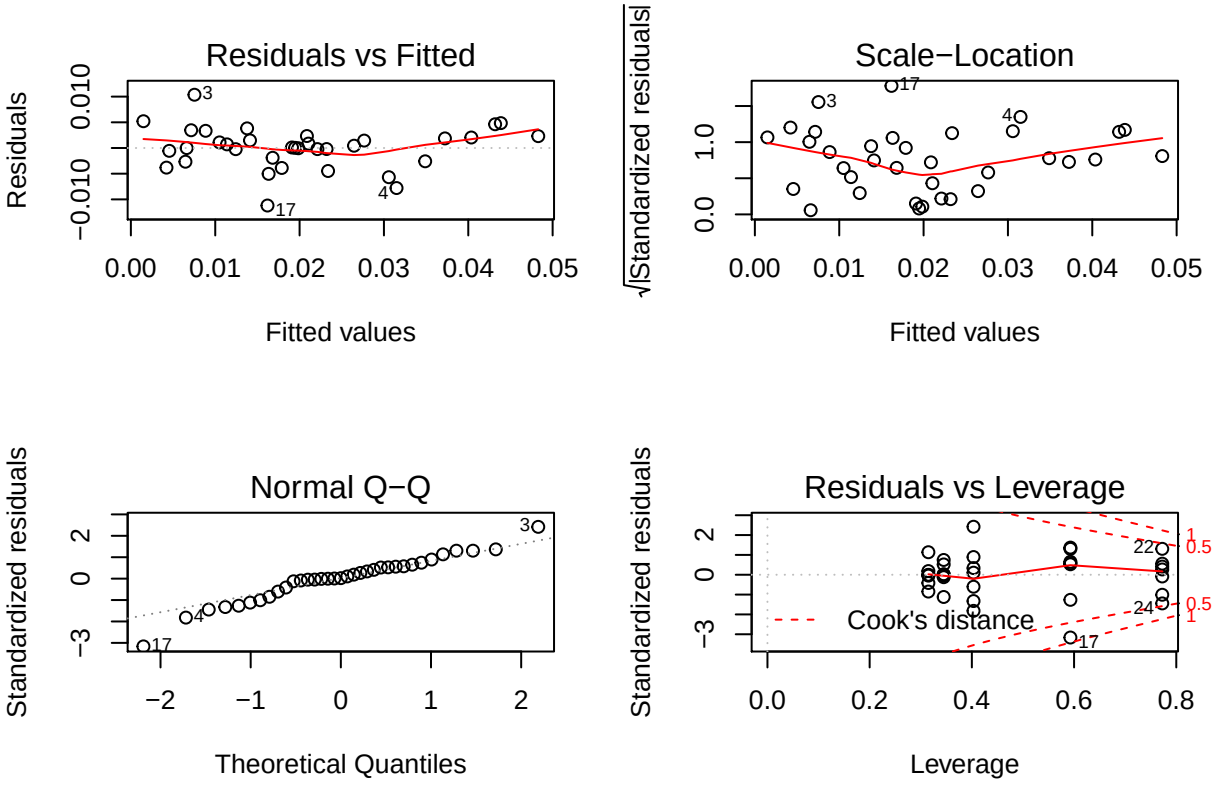


3 ANOVA

Before starting our analysis we need to verify if our data can be modeled with an ANOVA. First we verify that the variances in each varieties are equal, so that we have homoscedasticity.

Table 1: Analysis of Variance Model

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
varieties	4	0.003006	0.0007514	24.43	4.583e-07
strains	6	0.001867	0.0003112	10.12	5.942e-05
days:strains	6	0.000313	5.216e-05	1.696	0.1794
Residuals	18	0.0005536	3.076e-05	NA	NA



3.1 Model

After trying different models and analysing each time the Pr values, we decide to we choose one depending on the the value of Pr to see if we reject the NULL hypothesis or not. The one that gaves us the best the values that are inferior to 5% is the following one :

$$rate \sim varieties + strains + days + days : strains \tag{1}$$

3.2 Result-Analysis

The ANOVA table is the following one :

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
days	1	0.0025257	0.0025257	82.122	3.98e-08	***
varieties	3	0.0004799	0.0001600	5.201	0.0092	**
strains	6	0.0018670	0.0003112	10.117	5.94e-05	***
days:strains	6	0.0003130	0.0000522	1.696	0.1794	
Residuals	18	0.0005536	0.0000308			

As said before the value of Pr is inferior to 5% for the first 3 factor, we conclude that we can neglect the null hypothesis.

Now that we decide on a model, we need to analyze the quality. To do so we look at :

- Residual vs fitted values
- Normal QQ plot

METTRE les GRAPHS \ In fact we see that in the figure (Residual vs Fitted values) the behaviour is constant and around zero. For the Normal QQ plot, the behaviour is generally linear, but as we can see there are points (especially for theoretical quantities between 1 and 2) the points follows less the linear tendency. This can be explain by the fact that we consider that generally the initial weight of the apple didn't influence the rate. But the values of the weight varies a lot in the data, (the variance of the rate is around XXX) so this can be a cause of the change of behaviour for large Theoretical values.

An other really important aspect to analyse is the variance of the data. We need to make sure that the data is homocedency in order to be sure that an anova can be correctely applied. The table XXXX illustrates the result. In fact we see that the variance is generaly constant, so we can conclude the data is homocedency

References

[1] A. S. Horne Frederick Gugenheim Gregory Vernon Herbert Blackman "A quantitative study of the course of fungal invasion of the apple fruit, and its bearing on the nature of disease resistance.—Part II. The application of the statistical method to certain specific problems". **102**. *Proceedings of the Royal Society of London. Series B, Containing Papers of a Biological Character* http://doi.org/10.1098/rspb.1928.0018