UE 9.16
Advanced statistics

Kemely **JEAN BAPTISTE**

Joshua Elukut **AMAITUM**

I. MATERIAL AND METHODS

1.1. Data structure

The dataset used comes from the Lake Matapedia Experimental Forest (LMEF). The LMEF originates from a clearcut that was carried out in the mid-1930s. In 1989, a network of 23 permanent plots was established to monitor the growth of individual trees. Each fixed-radius plot covers an area of $400m_2$. Within the plots, all the trees with diameter at breast height (dbh, 1.3m in height) greater than 1cm were tagged and their dbh was measured. The plots were measured again in 1994, 2000 and 2006. The file contains the following fields.

Table 1 : *The different fields which make up the dataset*

Variable	Description	Units
plotID	unique id for each plot	-
treeID	unique id for each tree	-
dateYr	date	year
densityHa	plot-level tree density	stems ha^{-1}
basalAreaM2ha	plot-level basal area	m^2 ha^{-1}
species	species group of the tree	-
dbhCm	tree dbh	cm
BALM2Ha	basal area of all trees with dbh greater than the subject	m^2 ha^{-1}
dead	a binary variable (1= dead, 0=alive)	-
diMm	dbh increment over the growth interval	mm

1.2. Data preparation

The plot 5 were considered by our group and we decided to check the category of hypothesis about the diameter increments:

1. Diameter increments are different among the species.
2. Larger the trees have smaller diameter increments.
3. The more intense the competition, the smaller the diameter increment is.

Plot 5 contains data for 1989, 1994 and 2000. Three species make up this plot: Fir, Spruce and Others. The Fir species represented 98% percent of the whole elements. With 1.47 % for Spruce and 0.47 % for Others, we have estimated that those percentages were not significant enough to be considered in the analysis. Then, only the Fir species have been considered for a better output. This led us to discard the first hypothesis because, by deciding to work only with only Fir species, any analysis considering the discrimination among species could not have been performed. So, we worked on the last two. The following table is summary of the necessary fields:

Table 2 : *Measurement date, densityHa, basalAreaM2Ha, mean, min and max for dbh, Basal area and diameter increments*

Measurement date	densityHa	basalAreaM2Ha		dbhCm	BALM2Ha	diMm	living/dead	
1989	1925	30.65	*Mean* *Min-Max*	13.45 (3.1 - 25.5)	20.4 (0 - 30.64)	9.686 (0 - 24)	70	6
1994	1775	33.85	*Mean* *Min-Max*	14.7 (3.1 - 27.6)	22.64 (0 - 33.83)	8.76 (0 - 22)	63	7
2000	2850	36.11	*Mean* *Min-Max*	15.62 (3 - 29.3)	23.89 (0 - 35.8)	12.49 (0 - 29)	57	8

Our plot has data over three periods (1989, 1994, 2000), in order to have homogeneity among the different periods, we decided to split our plot into three data frames according to the years. After we merged our data frames by TreeID, with All=FALSE. This process has allowed observations for each tree across the three periods, while removing dead trees. while viewing our merged dataframe, we saw some null values among the necessary variables, and we removed them. Afterwards, we rebuilt our data frames according to the years (1989, 1994, 2000), but this time only with necessary variables (diMm, dbhCm, BALM2Ha).

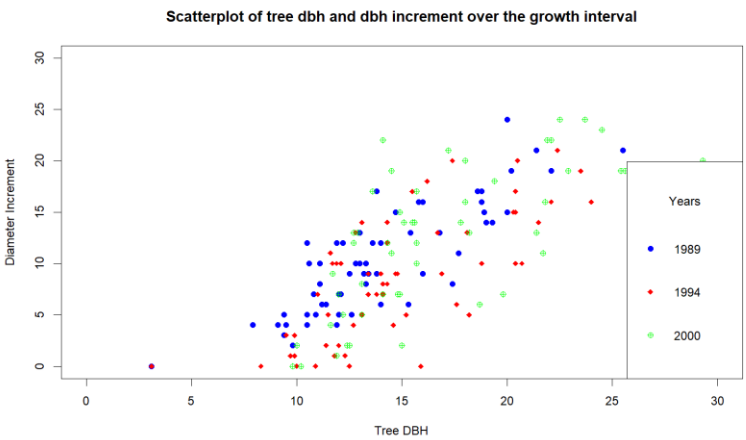
1.3.Statistical technique

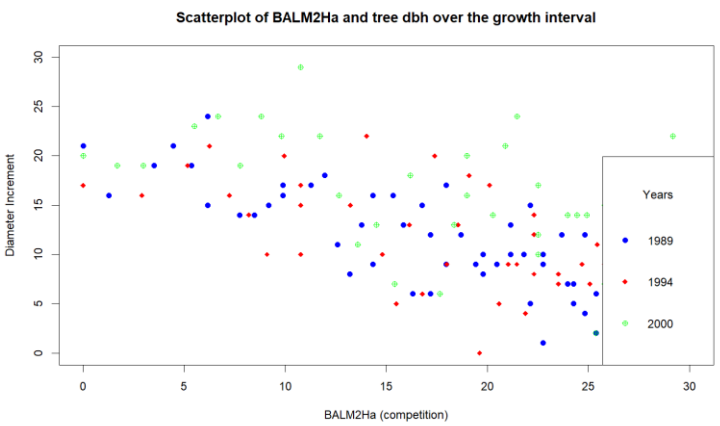
Since we had to deal only with numeric values, we decided to perform a linear regression model. The dbh increment over the growth interval (diMm) is the response variable for the hypothesis we wanted to check (2 and 3). The tree dbh (dbhCm) and the basal area of all trees with dbh greater than the subject (BALM2Ha) are the explanatory variables for the second and the third hypothesis respectively.

1. With scatter plots, we visualize our data : Plot **dbhCm** against diameter increments (**diMm)** and **BALM2Ha** against diameter increments (**diMm)** for each period to observe the general trend.
2. We made correlations analysis. We calculate correlation coefficients (Pearson's correlation) in order to measure the strength and direction of the linear relationship between **dbhCm** and diameter increments (**diMm)** and between **BALM2Ha** and diameter increments (**diMm)** for each period
3. By using linear regression models, we conducted regression analysis to model the relationship between **dbhCm** and diameter increments (**diMm)** and between **BALM2Ha** and diameter increments (**diMm)**. We looked at the coefficients in order to understand the direction and significance of the relationship.
4. We validated the findings of our linear regression models and ensured robustness, by checking assumptions and performing relevant statistical tests.
5. We Interpreted the results of our analysis in order to determine whether there is evidence to support the hypothesis.

II. RESULTS

Data visualization using a scatter plot to visually inspect the relationship between tree size and diameter increments was performed. Results show that diameter increment increases with bigger tree dbh and reduces with higher competition. Therefore, we disagree that larger trees have smaller diameter increments. Furthermore, we confirm that the higher the competition, the smaller the diameter increments.

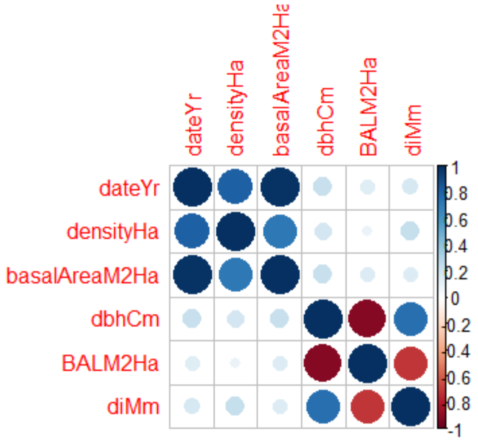




Correlation analysis was performed to quantify the strength and linear direction of the relationship between both tree dbh and competition with respect to diameter increment. Correlation between tree size and diameter increment = 0.7558055 while Correlation between BALM2Ha (competition) and diameter increment = -0.7148737. Therefore, there is a strong positive relationship between tree size and diameter increment, and a strong negative relationship between competition and diameter increment.

The correlation plot is displayed below.

Correlation plot of the variables



Regression analysis was performed to model the relationship and assess the statistical significance. Only data for fir trees was considered for this analysis. The modelling results are shared below.

1. Diameter increment = -6.13472 + 1.09179*dbhCm

2. Diameter increment = 21.38434 - 0.51923*BALM2Ha

Table 3: Model Summary

Model	R-Square	Adjusted R Square	Std. Error of the estimate
1	0.5691	0.5668	4.258
2	0.5159	0.5133	4.513

Table 4: Model Coefficients

Model		Unstandardized Coefficients		t	Sig.
		B	Std. Error		
1	(Constant)	-6.13472	1.09380	-5.609	.000
	dbhCm	1.09179	0.06984	15.632	.000
2	(Constant)	21.38434	0.85840	24.91	.000
	BALM2Ha	-0.51923	0.03698	-14.04	.000

From the results, the coefficient of tree size (dbhCm) is positive and statistically significant thus supporting the notion that larger trees have bigger diameter increments. **Therefore, we reject the stated hypothesis that larger trees have smaller diameter increments**.

The coefficient of competition (BALM2Ha) is negative and statistically significant thus supporting the notion that higher competition leads to smaller diameter increments. **Therefore, we support the stated hypothesis that the more intense the competition, the smaller the diameter increment.** In both cases, the assumptions of linear regression were met.

The analysis was also performed on a year-by-year basis for 1989, 1994, and 2000. The data is summarized in the table below. The year with the lowest average tree diameter (1989) had comparatively lower diameter increments while the year with the lowest competition (2000) had comparatively higher diameter increments which confirms both the hypothesis.

Correlation between tree size and diameter increment

Year	1989	1994	2000
Correlation	0.7772922	0.6852535	0.6751358

Correlation between basal area of larger trees and diameter increment

Year	1989	1994	2000
Correlation	-0.775161	-0.6771198	-0.6738281

Table 5: Model Coefficents

Year	Model	Unstandardized Coefficients		t	Sig.
		B	Std. Error		
1989	(Constant)	-4.0804	1.8654	-2.187	.034
	dbhCm	1.0686	0.1248	8.560	.000
	(Constant)	20.49983	1.14114	17.964	.003
	BALM2Ha	-0.49597	0.05834	-8.501	.000
1994	(Constant)	-4.0804	1.8654	-2.187	.034
	dbhCm	1.0686	0.1248	8.560	.000
	(Constant)	19.60136	1.58595	12.359	.000
	BALM2Ha	-0.45604	0.07154	-6.375	.000
2000	(Constant)	-4.9766	2.9549	-1.684	.099
	dbhCm	1.0879	0.1716	6.341	.000
	(Constant)	24.18961	1.87398	12.908	.000
	BALM2Ha	-0.49333	0.07808	-6.318	.000