METMA International Workshop on Spatio-Temporal Modelling
Santiago de Compostela
30 Jun - 2 Jul 2010

Analysis of Clinical Outcome of Patients with Lung Emphysema

I. Sousa[1], L. Rocha[1] and R. Menezes[1]

[1] *University of Minho, Department of Mathematics and Applications, isousa@math.uminho.pt, ID2704@alunos.uminho.pt, rmenezes@math.uminho.pt*

Abstract. *The aim of this work is to analysis and modelling the Clinical Outcome of Patients with Lung Emphysema. For that, we propose a regression model with independence between the observations and a longitudinal model with correlation structure.*

Keywords. *Lung Emphysema; Longitudinal Models; Regression Models*

1 Introduction

Clinical studies, very useful in understanding diseases, are often used to monitor the evolution of the health status of patients over time.
If previously regression models are used in the adjustment of models for this type of data today, the longitudinal models have been increasingly used the modelling of such data. The longitudinal models are more realistic in this context because the measurements of the same individual are not considered independent, i.e., is taken into account a correlation between the measurements of the same individual.
In this study we attempted to fit to the data a longitudinal model assuming a correlation structure between measurements and a linear regression model assuming independence between them.
The data used in the evolution of the clinical status of 48 patients with emphysema were in to two types of treatment: surgical or non-surgical. To check the status of the patient at each visit, the amount of air that can inspire the patient within one second was recorded. The greater the amount of air inspired, the better the health of the patient at that time.
The purpose of this study is to set appropriate models and compare the data taking into account two types of approaches: longitudinal models with correlation structure and linear regression models assuming independence between the measurements. Another aim is to assess which of the treatments (surgical or non-surgical) is more effective in dealing with this disease.

2 The data set

The data used relate to the amount of air breathed in a second by the 48 patients in consultations over the monitoring of the disease (see [4])). Of these 48 individuals, 35 were males and 13 were female. Data were collected between 1982 and 1999.
The age is between 42 to 74 years, and the average is about 61 years. The body mass index of patients, on average, shows that the relationship between weight and height of patients is normal and healthy (BMI values from 18.6 up to 24.9). However, there are 7 patients with underweight and 9 patients above the desired weight. These variations in weight may be associated with the disease in question.
During the study, some patients were dying. Note that the patient's death may or may not be associated with the disease in question. There is a fee of 56 % survival in this study. When comparing the survival times of patients undergoing surgical treatment with survival times of patients undergoing non-surgical treatment, it appears that surgical treatment provides survival times higher. However, performing a hypothesis test for comparison of survival times in both types of treatment it appears that there are no statistical significance differences between the two treatment.
The follow up of patients was a maximum of 2381 days (about 6 years and half). The average length of follow-ups of patients is 1153 days or approximately 3 years.
Individuals who underwent surgical treatment wait on average 225 days (about seven and a half months) for surgery.
Figure 1 shows three graphs for representation of data in the study. The first is the evolution of all patients during the follow-up time. The second graph represents the evolution of the patients who underwent non-surgical treatment over time for monitoring. The third graph represents the evolution of the patients who underwent surgical treatment during the follow-up time. The solid line represents the mid-line of evolution estimated using the function *smooth.spline* in software R (see [2]).

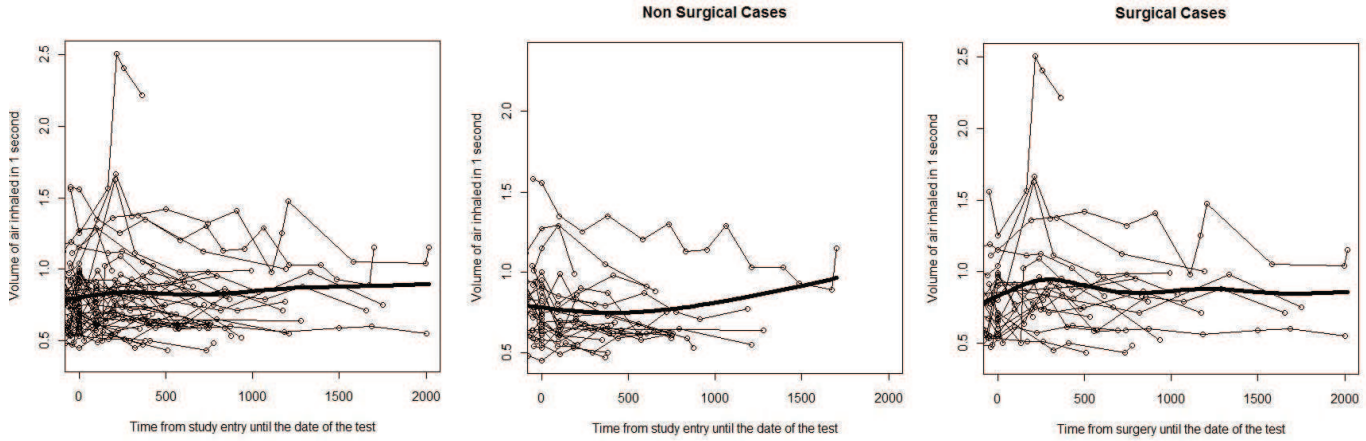

Figure 1: Evolution of patients throughout the study

3 Regression Models and Longitudinal Models

Consider the y_i amount of air inhaled by the patient during a second query i. In a linear regression model is assumed independence between observations and model equation is given by

$$Y_i = a + \beta_1 X_{i1} + \ldots + \beta_p X_{ip} + \varepsilon_i, \quad i = 1, \ldots, 340 \tag{1}$$

where p represents the number of covariates that are related to the progress of patients throughout the study and $\varepsilon \sim N(0, \sigma^2)$ homoscedasticity and are not correlated.
Models in longitudinal measures of the same individual are not considered independent, ie it is considered a correlation between measurements from the same patient. The model equation is given by

$$Y_{ij} = \beta_1 X_{ij1} + \ldots + \beta_p X_{ijp} + \varepsilon_{ij}, \quad j = 1, \ldots, 48 \quad i = 1, \ldots, n_j \tag{2}$$

where p represents the number of covariates that are related to the progress of patients throughout the study and $\varepsilon \sim MVN(0, V(t,\ alpha))$. The big difference between these two models is that ε in the longitudinal models are a combination of three sources of variation: random effects (differences between individuals), serial correlation (of the individual variation in time) and measurement errors (see Diggle (2002) [1] and Pinheiro (2000) [3]).

4 Results and Discussion

By analysis of the correlation coefficient of Pearson finds that the variables that indicate how long the patient waited for surgery and time between the start of the study and consultation are correlated with the dependent variable and should be included in the models. Table 1 shows estimates for the parameters of the models calculated.

	Intercept	β_1	β_2
Regression Model	0.8338	$-4.9 * 10^{-5}$	not significance
Longitudinal Model	0.8309	-0.0001242	0.0000754

Table 1: Estimativas dos parâmetros dos modelos.

In the regression model it appears that the time patients wait for surgery is not significant in the model, ie the type of treatment that the individual is not subject to influence developments in the clinical state of the patient. Not so in the longitudinal model. In this model, the evolution of the patient's clinical condition is related to the type of treatment that the individual is subject and the time from study entry until the query.
The comparison between the two models was performed using the criterion Log.likelihood (LogLik). The model shows the value of longitudinal LogLik lowest (38.26) while the regression model has a value of -70.9 LogLik. Figure 2 shows the residuals of the regression model. They present themselves around zero, not creating some kind of structure. In Figure 3 are represented two graphs. The first shows the variability among 48 individuals. It appears that the values are around the average. The second graph shows the fitted values versus the waste. It appears that there is no defined structure. It is expected that the longitudinal model is the most appropriate model to the data concerned.

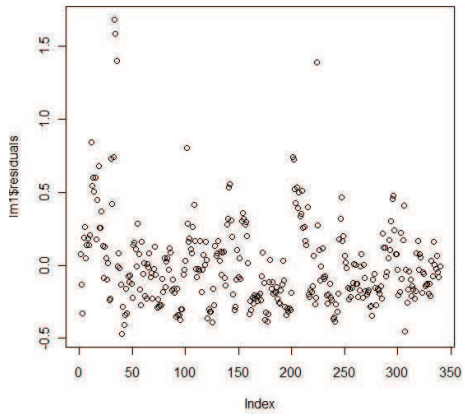

Figure 2: Waste of Regression Model

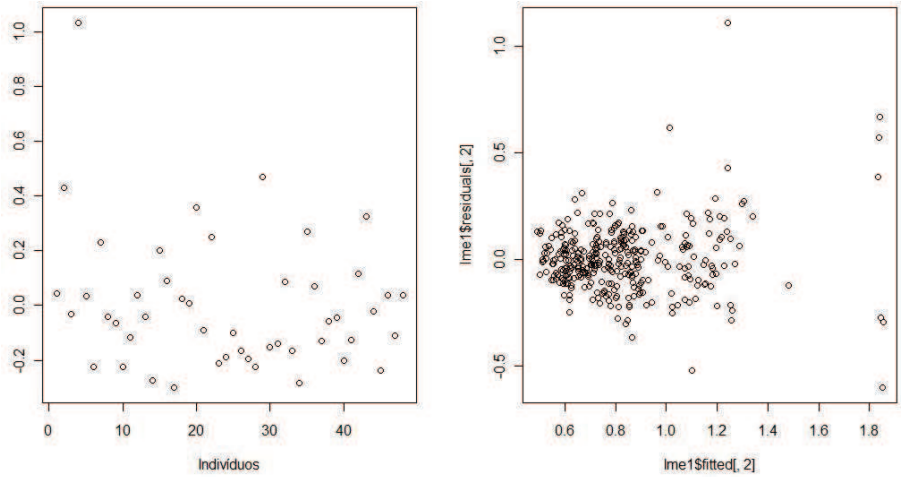

Figure 3: Waste Longitudinal Model

Acknowledgments. The author L. Rocha has a PhD scholarship by FCT SFRH/BD/61368/2009.

References

[1] Diggle, P.J., Heagerty, P., Liang K-Y. and Zeger, S.L. (2002). *Analysis of Longitudinal Data* (second edition). Oxford: Oxford University.

[2] Hatie, T.J. and Tibshirani, R.J. (1990). *Generalized additive models*. Chapman and Hall, New York.

[3] Pinheiro, J., Bates, D.(2000). *Mixed-Effects Models in S and S-PLUS* (first edition). Springer.

[4] Sousa, I. (2007) *Analysis of Repeated Measurements and time-to-event Outcomes in Longitudinal studies* PhD thesis, Lancaster, UK.