Mixed and functional analysis models for studying Heart Failure (re)-hospitalizations

POLITECNICO MILANO 1863

Castellani Tomaso, Corbo Gabriele, Fabroni Davide, Lucchini Marco

Abstract

Heart failure is a pathophysiological state in which the heart fails to supply the required amount of blood and oxygen to the body. It is a common, costly, chronic and potentially fatal condition widespread all over the world.
The main aim of this study is to provide statistically relevant methods for identifying the **potentially high-risk patients** in advance and understanding which are the main patient-level features that have an impact on the re-hospitalization. Moreover, another goal of the project is to find out if there are differences among patients due to a possible **hospital effect**. Finally, another purpose is to evaluate the impact of adherence to the **drugs** in (re)-hospitalizations. The study is based on the recordings of the hospitalizations and the medical prescriptions of 187493 patients from the Lombardy region but in this project, the main focus is centred on the people from Milano.

Dataset

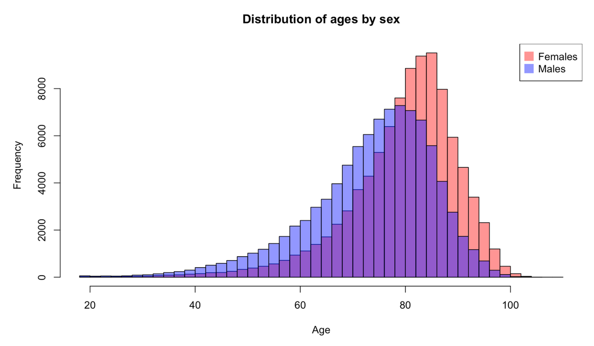
The dataset, which includes all the recordings of **hospitalizations** and **medical prescriptions**, stores information about patients, hospital and prescription levels. It involves more than 180'000 individuals among which about 30'000 reside in Milano.

- Age
- Gender
- Date of death
- ASL code
- Dates of study
- Final condition

- Date of admission
- Length of stay
- Comorbidities
- Structure ID
- Clinical interventions

- Date of purchase
- ATC code
- Length of prescription

Exploratory analysis and Clustering for comorbidities



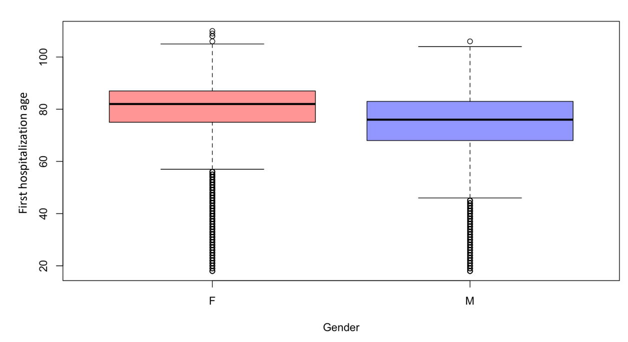


Fig. 1 and 2: We can see that there is a difference between the two genders in the age at first hospitalization distribution

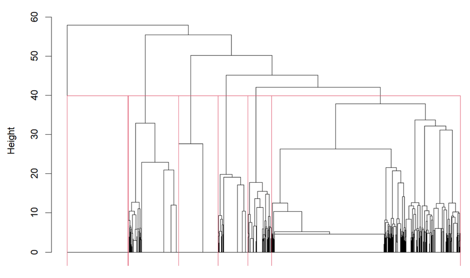
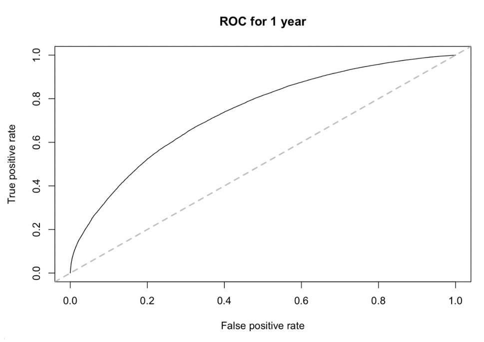


Fig.3 : Dendogram

As we can see from the first two images, there is a significant difference in the mean number of hospitalizations and age between the two genders, confirmed by a MANOVA test.
We tried to synthesize the information about the **comorbidities** at first hospitalization through a **clustering method** in order to reduce the complexity of the future models. We identified 6 clusters using the Euclidean distance and the ward linkage, getting a cophenetic coefficient of 0.6291303.
The cluster can be interpreted as follows: in the second one, there are patients mainly affected by hypertension, in the third one by tumor, in the fourth one by pulmonary diseases, in the fifth one by arrhytmia , in the sixth one by renal diseases. Finally, in the first (and biggest) cluster the is no prevalence of any comorbidity.

Life expectancy (Logistic mix. model)

Predicting a patient's life expectancy after the first heart failure hospitalization is crucial to assigning correct therapy. We decided to focus on the time frame of 1,3 and 5 years and built three **logistic models** to explain the probability of death in that time frame.
The relevant covariates are Age, Gender, membership in a cluster and the interventions performed in the hospital at the first hospitalization.
For the 1-year model the results are the following:



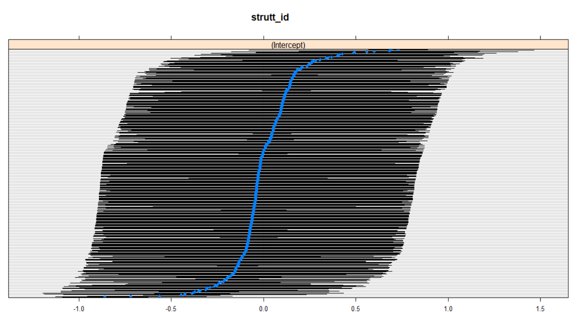
AUC = 0.74
Choosing threshold 0.6 we obtain a prediction error of 25,6% and this confusion matrix :

True \ Pred	Alive	Dead
Alive	6277	1769
Dead	1717	4161

For the 3 and 5-year models, the AUC increases to 0.7760562 and 0.8121018. By using thresholds 0.3 and 0.2 it is possible to obtain errors of 23% and 18% respectively. What is more relevant is that the error when the prediction is Alive is always bounded and less than 25%. Assuming a random effect due to the patient's hospital, we created a **mixed-effects** model for life expectancy at 1 year. We obtain a PVRE of 0.043 and classify it with a threshold of 0.6 the error is 24.4% on the dataset with the following results:

AUC = 0.77
Choosing threshold 0.6 we obtain a prediction error of 24%

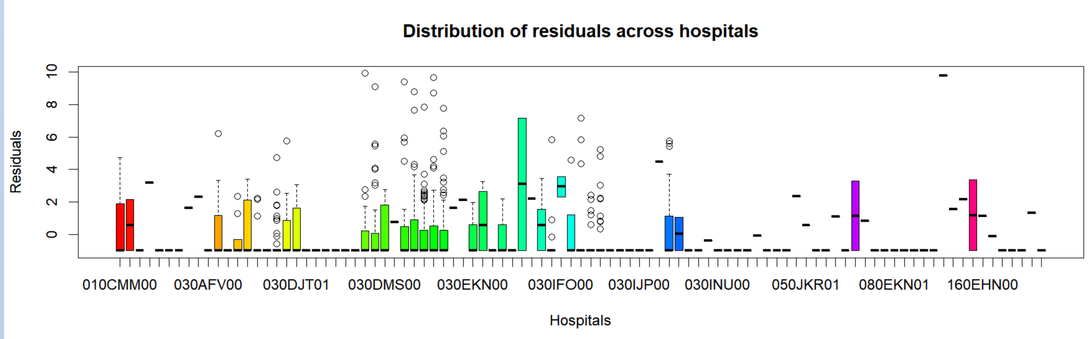
True \ Pred	Alive	Dead
Alive	6164	1882
Dead	1435	4443



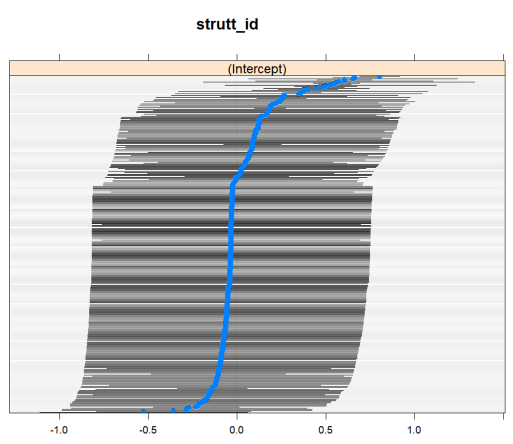
We, therefore, decided to keep the mixed effects model to explain life expectancy. Within the same hospital the factors that reduce life expectancy the most are **age** and the membership in the third group (characterized by the highest presence of **tumor** among clusters).

Number of Re-hospitalizations (Poisson mix. model)

Following the idea behind the previous part, we made three **Poisson regression** models to predict, given a patient at the beginning of the study, the **number of hospitalizations** that he will have to carry out in the next 1,3 and 5 years.
We considered for each individual his initial comorbidities (represented by the belonging to a cluster and by the number of comorbidities), age and gender, and for each model we just analysed the patients that stayed in the study for at least *n* years.
The AIC of the five models are: 13754, 8101.2 and 3134.8.



Since we noticed that the distribution of the residuals is not homoscedastic with respect to the hospital, we built a **mixed model** with the id of the structure as a random intercept. The PVRE are respectively 0.0966, 0.2194 and 0.3350.



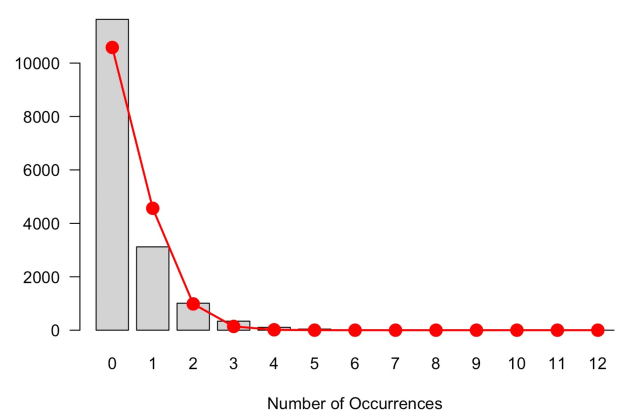
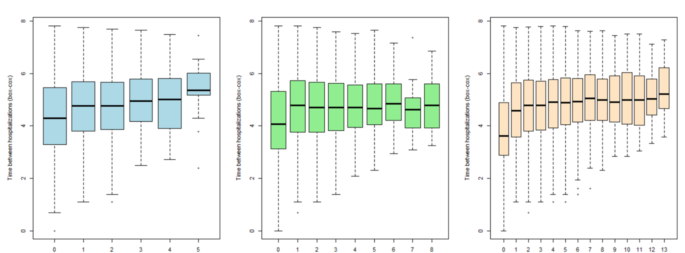


Fig. : The histogram shows that the distribution of the number of hospitalizations is almost a Poisson density function with lambda = 0.4.
The Poisson density is:

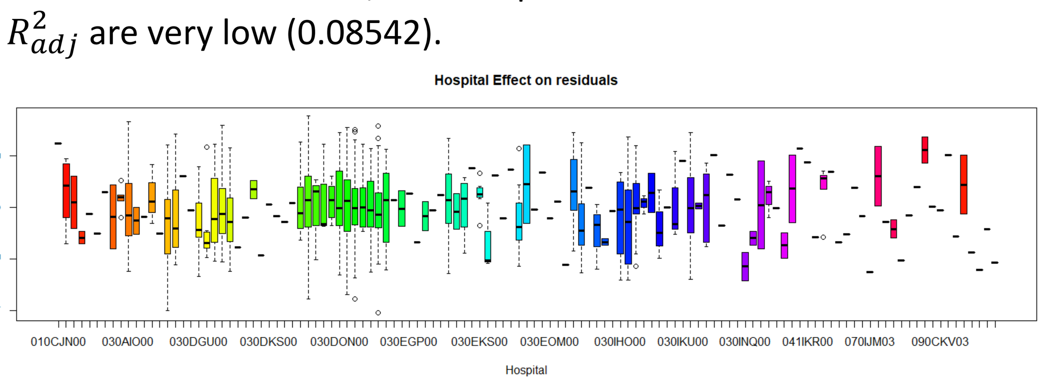
$$f(n) = \frac{\lambda^n}{n!}e^{-\lambda}$$

Time between hospitalizations (Linear mix. model)

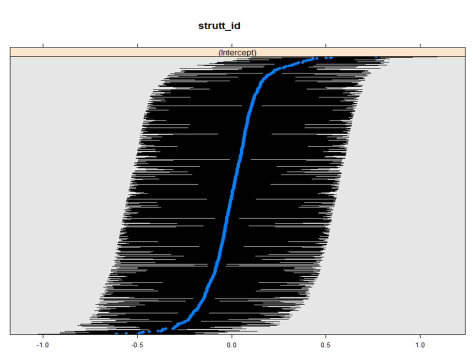
The aim of this part is to understand how the variables, such as age, gender, number of comorbidities, length of stay and in particular assumption of drugs can affect the **time between two consecutive hospitalizations** of the same patient. Moreover, this study could provide a useful tool for the doctors to decide the frequency of the checks for each patient knowing his current conditions. From *Fig. below* we can expect a **medicine effect** and we can confirm it with the following analysis.

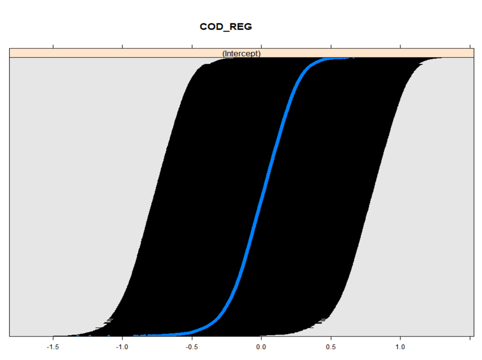


The initial model constructed is a linear one in which all the features used seem to be relevant, but the performances measured in terms of R^2_{adj} are very low (0.08542).



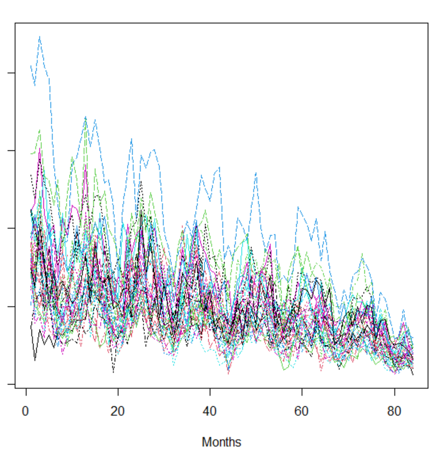
In order to try to catch the dependence of the observations (*Fig. above*) caused by the hospital or the patient, two independent **random components** on the intercept have been added. We obtain a PVRE of 0.1435431 (mainly due to the dependence on the same patient).

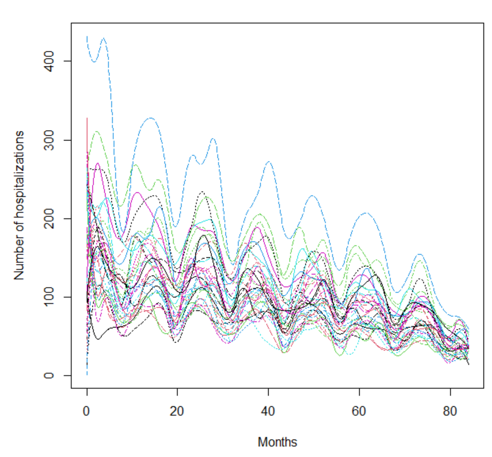




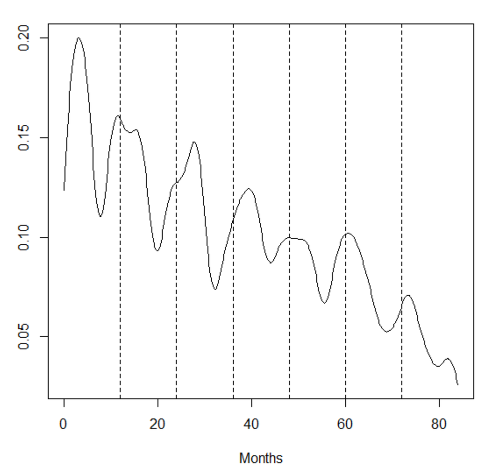
Functional Data Analysis

Finally, we selected the 33 hospitals in Lombardy with the **highest number of entries** and we studied the hospital entries in each month of the seven years of experiments. The objective is to find out if there is a **particular trend** in hospitalizations.





Just from the *Figs. on the left*, we can denote a downward trend. Performing a **Functional Principal Components Analysis**, the study has revealed that the first principal component alone explains almost 90% of the variability so we proceed with its interpretation.



The plot shows an **increment during the winter months**, suggesting a higher number of hospitalizations following the lowering of temperatures. Indeed, as reported in the article *A SWEDEHEART Nationwide Observational Study* published by *Mohammad MA, Koul S, Rylance R* on *JAMA Cardiol* in 2018, the cold and windy climate leads to a contraction of blood vessels, increasing the workload of the heart and consequentially the risk of heart failures.

Conclusions

Our models were created to help doctors to make **better decisions**. The first and the second one predict the severity of a patient at the first hospitalization and this information can be used to identify the patients who need more care. The objective of the third model was to plan in advance control hospitalizations based on the need of a patient and allow the hospitals to group them and plan days with specialized doctors. Also, this could reduce the waiting time and the stress connected to the patients. Unfortunately, this model can't be used directly since shows very low performances but suggest that adherence to drugs is relevant and further investigations can be done. Functional data analysis shows that the winter period is the worst for heart failure disease so information brochures may be distributed to patients to suggest habits to reduce the incidence of the phenomenon. **Random effects** in the models suggest that hospitals make a difference in the therapy so studying the difference between the treatments made in the most virtuous hospitals and the others may lead to an improvement in the overall quality of the service.

Extra