

Lab 2

Data

Emergency services data, again. Similar dataset as the last lab. Each observation is the stay of one patient in one emergency service.

- It has already been (mostly) cleaned.
- All variables with missing values have had their missing values imputed
- Patients with ages over 120yo have been removed
- Patients with length of stay over 4320 minutes have been removed

Variables

- Region: geographical region where the service is located
- Service: name of the service
- id: unique identifier of the patient in the service
- severity: severity of the health status of the patient (= how bad it is). Scale of 1 to 5.
- Discharge mode: where the patient went when they exit the emergency service.
 - Home: patient was sent home.
 - Hospitalised: patient was hospitalised in the same hospital as the emergency service is located in.
 - Transferred to other hospital: patient was hospitalised in an hospital different from the one the emergency service is located in.
 - Died: patient died during the stay
- arrival_mode : how the patient arrived to the emergency service
 - Ambulance
 - Ambulance + firefighters
 - Ambulance + life support (ambulances with specialised equipment for more severe cases)
 - Cops
 - Helicopter
 - Personal (you went to the emergency service by yourself)
- los: length of stay. How long you've been in the service. In minutes.
- age: time between your date of birth and your date of arrival in the service. In years.
- Diagnostic: Main diagnostic of the patient.
- entry: date and hour of entry in the service
- exit: date and hour of exit in the service
- complexity: how complex the pathology of the patient is to treat, on a scale of 1 (lower) to 10 (higher). Complexity is calculated at the level of the diagnostic, not at the individual level (i.e. each diagnostic will have the same complexity level, even if for some patients, the same diagnostic may be more or less severe). It is a score (or latent variable) built from a PCA that includes, for each diagnostic:
 - The median length of stay
 - The percentage of entries with an ambulance (alone, or + firefighters, or + life support) or helicopter
 - The percentage of patients aged <1 or >75
 - The percentage of hospitalisations/transfers to other hospitals
 - The percentage of severity 3, 4 or 5.

- cost: total cost of the stay, in €.
- sas: "Service d'Accès aux Soins". When you have a medical emergency and your general practitioner ("médecin traitant") is not available, you are supposed to call "le 15". Initially, it was le SAMU (Service d'Aide Médicale Urgente). Now, it is the SAS. They will give you medical advice, and orient you towards some alternate offers: a teleconsultation, an emergency medical consultation in a dedicated healthcare centre ("centre de soins non-programmés"), an emergency service. For serious emergencies, an ambulance/helicopter is sent. In the context of the data, this variable indicates, for people that have come to the service by themselves (arrival_mode == "Personal"), if they called the SAS and were told to come ("yes"), if they did not call the SAS ("no") or if they called the SAS but were not told to go to an emergency service ("no").

Tasks

Variable creation

- Relative number of entries

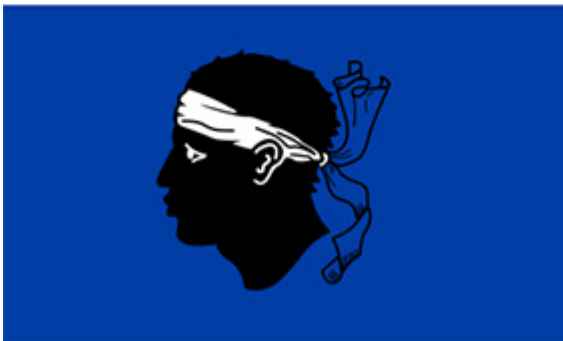
This is more of a little challenge, it may be harder that it seems

Average number of patient entries per hour during the stay of the patient divided by the average number of patient entries per hour in the service (the total number of hours will be, per service, the time between the earliest entry and the latest exit in the dataset)

$$RNE = \frac{\text{Patients/hour}_{\text{stay}}}{\text{Patients/hour}_{\text{service}}}$$

- Length of stay > 8 hours. Binary variable: takes 1 (or "yes") if los >= 480 minutes. 0 (or "no") if < 480 minutes.

The exercise



The Union Démocratique du Québécois is facing economic difficulties. Its democratically elected president, Jean-François Québécois, has announced an unexpected deficit of 5% this year, well over the expected 2%, due to poor management and planning by his Minister of Quebec Economics, Brunon Le Mar. This has sparked severe public outcry, especially given the rise in economic inequality over the last 15 years of Jean-François' presidency, the deterioration of multiple public services (including healthcare) and multiple scandals involving the president's lavish spending despite his repeated calls for austerity.

Fortunately, Brunon Le Mar has been dismissed and sent far away to reflect on his mismanagement. Unfortunately, you have been randomly chosen to replace him and Jean-François will publicly pin the whole economic collapse of the country on you if you don't fix their mess.

Management of emergency services fall under the responsibility of your ministry. This year, the total public spending for emergency services reached 435,751,480€ and is not expected to rise next year. As part of your plan to reduce public deficit to 2% next year, you aim to reduce the public spending for emergency services by at least 1.5%.

There is not much you can ethically do to reduce public spending on this matter. However, focusing on patients that enter emergency services without going through the SAS (Service d'Accès aux Soins) is one lever you can pull to reduce spending: they are known to affect costs, and length of stay (notably, stays over 8h), but you don't know by how much.

You have a meeting with le Grand Conseil des Ministres du Québécois in two weeks, and they expected results.

You anticipate they want to know:

- What are the main factors associated with length of stay and costs, and how are they associated? Which ones are the most important?
- What options do we have to reduce costs? How much can we expect to save?

What you could do:

- univariate descriptions of the data (costs, length of stay, over relevant variables that you may include in your analyses)
- bivariate descriptions (with length of stay, costs)
- a multivariable analysis (regression, or something more complicated if you can explain it)
- Counterfactual comparisons (see `marginaleffects` package)
- Some type of importance plot (you can get that with `ranger`, or use DALEX)

My recommendations/advice:

- There are multiple angles you can take to tackle this cost issue: some services may be more inefficient than others, the SAS variable, overcrowding...
- Remember that some variables can affect both length of stay and costs, and since length of stay affects costs...
- Try to answer this question: "how much money could the country save if all patients went through the Service d'Accès aux Soins (SAS)?" This will be your introduction to something called [G-computations](#).

Useful tools

Tabyl

For creating tables with categorical variables: <https://cran.r-project.org/web/packages/janitor/vignettes/tabyls.html>

Marginal Effects

Counterfactual comparisons/G-computations: <https://marginaleffects.com/>

DALEX

Model-agnostic methods for explaining the output of statistical models. <https://ema.drwhy.ai/>

This chapter for variable importance: <https://ema.drwhy.ai/featureImportance.html>