

Non-linear Regression Background

September 25, 2021

1 Overview

This work is for a project studying the judicial system in South Carolina. In that system, judges rotate across counties, instead of just hearing cases in a single county. Furthermore, defendants are allowed to "choose" (to a certain extent), which judge hears their case. The goal of the project is to study how changes to the system would affect different sentencing outcomes, and we do that through simulation.

As part of this, we are currently trying to estimate how long it takes judges to process pleas and trials. We have a dataset that includes how many days each judge was assigned to each county, how many pleas they sentenced each day in each county, and how many trials they sentenced each day in each county. There are certain days of the year in which institutional knowledge suggests that judges only worked on pleas and trials. We were hoping to focus on these days to determine how long it takes judges to process pleas and trials.

There is large demand heterogeneity amongst counties, i.e. some counties see a lot more sentences than others. We first tried to estimate the service rates by estimating the following model, where the unit of observation i corresponds to a judge-county combination, and c denotes a county: $\text{Days}_i = \alpha_c + \beta_T \text{Trials}_i + \beta_P \text{Pleas}_i + \epsilon_i$. The county fixed effects, α_c , were all positive and we interpreted these as the idleness in the different counties, or the "fixed costs" that judges have while working.

However, we were worried because the number of days that judges are assigned to different counties aren't always on the same scale. For example, given a certain county, c , there are some judges that are assigned there for 25 days, and there are others that are assigned there for only 5 days. If the estimated α_c is 4, then the model would imply that both judges "wasted" 4 days. In some cases, the model would imply that, because of the fixed costs, judges had no time to work on pleas or trials, even though the data shows that they did. We decided that this was because we were trying to use the same intercept, α_c , for all judges, regardless of how many days they were assigned to a specific county.

We thought that a better approach would be to model idleness as a proportion of total time. For example, it might be the case that judges in general spend 10% of their time doing administrative work instead of hearing cases. So, we wanted to estimate the following model so that we could estimate each judge's 'proportional idleness'.

$$\begin{aligned} u_c \text{Days}_i &= \beta_T \text{Trials}_i + \beta_P \text{Pleas}_i \\ \log(\text{Days}_i) &= \log\left(\frac{1}{u_c}\right) + \log(\beta_P \text{Pleas}_i + \beta_T \text{Trials}_i) \end{aligned}$$

I pointed out to my PI that this can equivalently be expressed as a mixed effects model, but he insisted that I try to estimate this particular model, but I haven't been able to figure out how to implement it in R. Any modeling suggestions are also very welcome.