

STA 721 Fall 2013: Final HW/Data Analysis Take-Home

November 24, 2013

Data Analysis Problem: Health plans use many tools to try to control the cost of prescription medicines. For older drugs, generic substitutes that are the equivalent to name-brand drugs are available at considerable savings. Another tool that may lower costs is restricting drugs that the physician may prescribe. For example if three similar drugs for treating the same condition are available, a health plan may require the physician to prescribe only one of them, allowing the plan to negotiate discounts based on a higher volume of sales.

The following data from 29 health plans can be used to explore the effectiveness of these two strategies in controlling drug costs. The response is COST, the average cost of the prescriptions to the plan per day (in dollars).

Potential explanatory variables are:

- RXPM: Average number of prescriptions per member per year
- GS: Percent generic substitute used by the plan
- RI: Restrictiveness Index, from 0 (no restrictions) to 100 (total restrictions on the physician)
- COPAY: Average member copay on prescriptions
- AGE: Average member age
- F: percent female members
- MM: Member months, a measure of the size of the plan
- ID: an identifier for the name of the plan

Data are on Sakai under Resources in costs.txt.

Using EDA and all methods that we have covered in this course explore the relationship of GS and RI to COST adjusting for the other variables. Summarize your results with regard to the importance of GS and RI and include quantitative summaries and appropriate uncertainty measures to summarize findings, and not just qualitative answers. Each analysis should address any outliers or influential observations. In particular, I would like you to include

1. the g-prior with $g=n$ using the full model
2. the lasso
3. ridge regression

4. Bayesian Model Averaging using the Zellner-Siow Cauchy prior
5. any other methods mentioned class (optional)

Summarize in a 2 page report with an introduction to the problem, brief methods, results, and discussion suitable for someone at the level of 601.

1. For BMA, based on the probabilities would you recommend picking one model?
2. For all, discuss sensitivity of the results to the modeling assumptions.
3. Do the different methods lead to similar conclusions or different conclusions?

Based on the method that you think provides the best solution, provide a one paragraph executive summary that would be suitable for a national insurance company that recommends which strategy would be most effective in controlling costs and provide projected savings (with uncertainty measures if possible) based on average settings for the other covariates.

Please provide details of your R code and analyses in an appendix, rather than the body of the report.