Data Mining
BANL 6625/BANL 4425
Department of Economics & Business Analytics
Spring 2019/Instructor: Rodriguez

Smoking and Mortality

The Whickham data set includes data on age, smoking, and mortality from a one-in-six survey of the electoral roll in Whickham, a mixed urban and rural district near Newcastle upon Tyne, in the UK. The survey was conducted in 1972-1974 to study heart disease and thyroid disease. A follow-up on those in the survey was conducted twenty years later.

Describe the association between smoking status and mortality in this study. Test using a chi-square test to determine whether any association is statistically significant at the 95% level of significance.

Run a binomial logit to again examine the association but this time controlling for age.

Show the results using plot_model in the package sjPlot.

Zombie GRIT

Do "grittier" employees perform better? This is, of course, empirical.

You ask all your 1000 employees to fill out a personality survey. You end up with a score depicting individual personality characteristic GRIT, which can run from 0 (no grit) to 7 (grit galore). You also assemble the respective data for Job Performance (measured 0 – 100) and data on individual Annual Salary. In order to get a sense of the effects, you may decide to visualize both these relations in scatter plots. Note that the data has NA's.

From these visualizations it would look like GRIT relates significantly and positively to both employees' performance and their yearly salary. Should you select more gritty people to improve your overall company performance? Or would this be discriminating against emotionally-stable, cool, mellow employees when it comes to salary?

Taking a closer look at the subgroups in your data, you might however find very different relationships. For instance, the positive relationship between grit and performance may only apply to humans, but not to zombies. In addition, examine the impact of employees by education level; it becomes clear that there is a relationship between grit and education level that may explain the earlier association with salary.

More educated employees receive higher salaries and within these groups, grite is actually related to lower yearly income.

Compile script into PDF, Word, or HTML format – and upload to blackboard.