Data Madness - Measuring morality

Henry Mauranen, Martin Gassner, Matteo Maggiolo

March 27, 2018

1 Introduction

We worked on the dataset Measuring morality, concerned with assessing the character of moral beliefs in individual in the United States. We will answer the following questions:

1. Can we find ways to interpret the answers of each block as one value?

2. Can we identify distinct clusters within different blocks of questions?

3. Based on these values and/or clusters, are there clear correlations between interpretable blocks?

4. Are there interpretable types of individuals?

5. Can we link the clusters to personal metrics?

2 Methods

We also applied the k-modes algorithm on a per block basis. For each block of multiple questions, we extracted the answer vector of each participant, and then we performed clustering on these vectors. In order to choose an appropriate number of clusters, for each value from 2 to 20 we performed 15 tests. Then, for each k we averaged the maximum total intra-cluster difference, and we applied the elbow method to choose the appropriate value.

After performing the clustering, we performed some experiments and discovered that some of the clusters found for one block had an almost complete correspondence with another cluster found on another block of questions. Using this information together with the clusters made on the entire set of blocks, we could identify some distinct group of people with very similar opinions.

We employed principal component analysis to obtain a linear combination proportionfor each block, weighting the influence of each question within. These weights were obtained from the sum of loadings per variable weighted by the variance explained by each principal component. We transformed the original data into a matrix only containing one value for each block and used k-means on it.

3 Results

1. Yes we can! We used PCA to obtain weights.

2. Yes we can! We performed EM-clustering and k-modes clustering on a per block basis and obtained the cetroids of each cluster. We then related the results with the questions presented in the questionnaire.

3. Sometimes. Our results show that the correlations between variables are not particularly strong. However we could find correlations between individual clusters for specific blocks. Additionally, there was often overlap between the clusters and only few variables clearly separated a cluster from the others.

4. Based on the per-block basis clustering we could find pairs of clusters correlated over multiple blocks. This established a link between blocks of questions for a specific cluster, such as: (a) a group of individuals never taking a side (b) a group of individuals with high answers to integrity and ethical values questions

   Furthermore, we could find four clusters on the transformed data with k-means, namely: (a) Materialistic pragmatists (b) Individualists (c) Average people (d) Saints

5. Corresponding to the previous 4 types: (a) rather young, educated, male, not married (b) more ethnically diverse, less tertiary educated, fewer living in detached houses (c) average, everywhere (d) most likely to be married and to live in a detached house, predominantly women