Exam 2

Isaac Haberman

April 18, 2016

Introduction

\paragraph While the fall of the Berlin Wall is the iconic moment of the Fall of Communism, the Communist Bloc had begun to fail in the 1980's, years before the fall of the Berlin Wall. The newly free countries quickly adopted democracy and capitalism as a replacement for the communist and planned economies that had previously terrorized their countries. With this change in countries, researchers were interested in the thoughts of the citizens political, historical and human rights views.

We have analyzed the anonymous researchers; dataset `partisan.csv` which is the results of a survey done twice, once in 1998 and once in 2003, in a country which we will refer to as "Partisan." The researchers theorized that demographics categorized citizens political views which then categorized their views on history and human rights. To test the theory of the researchers, we clustered the opinions of the citizens using mixture models, tested for predictability and independence between the clusters before concluding.

Attitudes Towards the Past

To begin our work, we did some exploratory data analysis of each of the variable groupings. Of the historical era variables, `monarchy` and `feudal` had the highest correlation at 0.5834756, while the lowest correlation was between `intl.stds` and `socialist` at 0.0082927. We speculated the relatively high correlation between `monarchy` and `feudal` was due to their intertwined relationship throughout history, while the low correlation between `intl.stds` and `socialist`is explained by the opposition between democracy and socialism. Interestingly, none of the correlations are negative, meaning all the variables have positive relationships. Hopefully, the mixture model will shed further light on data.

	class.1..	class.2..	class.3..
postsocialist.Pr.1.	0.0000000	0.6588080	0.2224646
postsocialist.Pr.2.	1.0000000	0.3411920	0.7775354
intl.stds.Pr.1.	0.5595043	0.6089604	0.1112323
intl.stds.Pr.2.	0.4404957	0.3910396	0.8887677
socialist.Pr.1.	0.3752744	0.7487267	0.2219665
socialist.Pr.2.	0.6247256	0.2512733	0.7780335
monarchy.Pr.1.	0.9979212	0.9962030	0.0000000
monarchy.Pr.2.	0.0020788	0.0037970	1.0000000
feudal.Pr.1.	0.9812946	1.0000000	0.0000000
feudal.Pr.2.	0.0187054	0.0000000	1.0000000

	X1	X2	X3
postsocialist.1	0.0000000	0.1714003	0.1895175
postsocialist.2	0.0000000	0.1714003	0.1895175
intl.stds.1	0.0332439	0.0208444	0.1398087
intl.stds.2	0.0332439	0.0208444	0.1398087
socialist.1	0.1599733	0.0190613	0.1447652

	X1	X2	X3
socialist.2	0.1599733	0.0190613	0.1447652
monarchy.1	0.0029958	0.0025317	0.0000000
monarchy.2	0.0029958	0.0025317	0.0000000
feudal.1	0.0087439	0.0000000	0.0000000
feudal.2	0.0087439	0.0000000	0.0000000

hist.mix.P	hist.mix.P.se
0.3754246	0.1605943
0.6175683	0.1605958
0.0070072	0.0023301

We cross-validated loglikelihood's to find the proper number of clusters for our final mixture model of attitudes towards the historical eras. The resulting number of clusters was 3. The tables above show the results of our mixture model. There are clusters, with the second cluster containing the majority of the data, followed by the first and third cluster. Above the table of classes is the table of probabilities for the different variables of the historical era.

General political Values

The second set of variables we analyzed were the political value variables. `personal.dignity` and `national.dignity` had the highest correlation at 0.1900721, while `selfdetermination` and `national.dignity` had the lowest correlation at 0.0066949. We speculated that the relatively high correlation between `personal.dignity` and `national.dignity`stems from the extension of national dignity from personal dignity. While the relatively low correlation between `selfdetermination` and `personal.dignity` is explained by national self determination not factoring into personal dignity. As with the previous table of variables, there are no negative correlations.

	class.1..	class.2..
freedom.oppression.Pr.1.	0.9936273	0.8964312
freedom.oppression.Pr.2.	0.0063727	0.1035688
personal.dignity.Pr.1.	0.8520491	0.0509277
personal.dignity.Pr.2.	0.1479509	0.9490723
selfdetermination.Pr.1.	0.9023370	0.8318168
selfdetermination.Pr.2.	0.0976630	0.1681832
national.dignity.Pr.1.	0.9430141	0.7806531
national.dignity.Pr.2.	0.0569859	0.2193469

	X1	X2
freedom.oppression.1	0.0060881	0.0331926
freedom.oppression.2	0.0060881	0.0331926
personal.dignity.1	0.1550918	0.0740752
personal.dignity.2	0.1550918	0.0740752
selfdetermination.1	0.0124002	0.0288427
selfdetermination.2	0.0124002	0.0288427
national.dignity.1	0.0128394	0.0539767
national.dignity.2	0.0128394	0.0539767

polit.mix.P	polit.mix.P.se
0.6164972	0.1343899
0.3835028	0.1343899

We cross-validated loglikelihood's to find the proper number of clusters for our final political value mixture model. Using the functions we used earlier, we chose to use 2 clusters. The tables above shows the probabilities, standard errors of the clusters and the variables.

Human Rights

The last set of variables we analyzed were the citizens thoughts on human rights. `hr.democracy` and `hr.mentioned` had the highest correlation at `r human.cor[66,3]`, while the lowest correlation was between `hr.natl.respect` and `hr.violated` at -0.0046654. We conjectured the relatively high correlation between `hr.democracy` and `hr.mentioned` could be explained by the broad terms of mentions and democracy. The relatively low correlation between `hr.natl.respect` and `hr.violated` can be explained by the apparent lack of similarities between self-determination of an individual and respect for a nation.

	class.1..	class.2..	class.3..	class.4..
hr.personal.dignity.Pr.1.	0.0193488	1	1.0000000	0.6828993
hr.personal.dignity.Pr.2.	0.9806512	0	0.0000000	0.3171007
hr.equality.Pr.1.	0.8133045	1	1.0000000	0.9380718
hr.equality.Pr.2.	0.1866955	0	0.0000000	0.0619282
hr.political.freedom.Pr.1.	0.1159095	1	1.0000000	0.8940726
hr.political.freedom.Pr.2.	0.8840905	0	0.0000000	0.1059274
hr.participation.Pr.1.	0.5139428	1	0.8853331	0.9901273
hr.participation.Pr.2.	0.4860572	0	0.1146669	0.0098727
hr.econ.freedom.Pr.1.	0.3752059	1	1.0000000	0.4629600
hr.econ.freedom.Pr.2.	0.6247941	0	0.0000000	0.5370400
hr.socioeconomics.Pr.1.	0.8558247	1	1.0000000	0.9378781
hr.socioeconomics.Pr.2.	0.1441753	0	0.0000000	0.0621219
hr.selfdetermination.Pr.1.	0.8800914	1	1.0000000	0.8839274
hr.selfdetermination.Pr.2.	0.1199086	0	0.0000000	0.1160726
hr.natl.respect.Pr.1.	0.6648906	1	1.0000000	0.7003808
hr.natl.respect.Pr.2.	0.3351094	0	0.0000000	0.2996192
hr.violated.Pr.1.	0.9780655	1	0.9527519	0.9827496
hr.violated.Pr.2.	0.0219345	0	0.0472481	0.0172504
hr.support.Pr.1.	0.7330583	1	1.0000000	0.6984378
hr.support.Pr.2.	0.2669417	0	0.0000000	0.3015622
hr.democracy.Pr.1.	0.1686029	1	1.0000000	0.2267397
hr.democracy.Pr.2.	0.8313971	0	0.0000000	0.7732603
hr.mentioned.Pr.1.	0.0000000	1	0.3765934	0.0000000
hr.mentioned.Pr.2.	1.0000000	0	0.6234066	1.0000000

	X1	X2	X3	X4
hr.personal.dignity.1	0.0157901	0	0.0000000	0.0368783
hr.personal.dignity.2	0.0157901	0	0.0000000	0.0368783
hr.equality.1	0.0302282	0	0.0000000	0.0183825
hr.equality.2	0.0302282	0	0.0000000	0.0183825
hr.political.freedom.1	0.0375124	0	0.0000000	0.0259926

	X1	X2	X3	X4
hr.political.freedom.2	0.0375124	0	0.0000000	0.0259926
hr.participation.1	0.0411348	0	0.1169681	0.0084798
hr.participation.2	0.0411348	0	0.1169681	0.0084798
hr.econ.freedom.1	0.0390004	0	0.0000000	0.0359948
hr.econ.freedom.2	0.0390004	0	0.0000000	0.0359948
hr.socioeconomics.1	0.0272581	0	0.0000000	0.0172259
hr.socioeconomics.2	0.0272581	0	0.0000000	0.0172259
hr.selfdetermination.1	0.0261767	0	0.0000000	0.0233176
hr.selfdetermination.2	0.0261767	0	0.0000000	0.0233176
hr.natl.respect.1	0.0356258	0	0.0000000	0.0307138
hr.natl.respect.2	0.0356258	0	0.0000000	0.0307138
hr.violated.1	0.0110239	0	0.0599726	0.0089434
hr.violated.2	0.0110239	0	0.0599726	0.0089434
hr.support.1	0.0338564	0	0.0000000	0.0309039
hr.support.2	0.0338564	0	0.0000000	0.0309039
hr.democracy.1	0.0289015	0	0.0000000	0.0302076
hr.democracy.2	0.0289015	0	0.0000000	0.0302076
hr.mentioned.1	0.0000000	0	0.3418672	0.0000000
hr.mentioned.2	0.0000000	0	0.3418672	0.0000000

human.mix.P	human.mix.P.se
0.1659080	0.0131116
0.6162198	0.0151222
0.0132252	0.0080312
0.2046470	0.0139257

Like the last two sets of variables, we cross-validated loglikelihood's to find the proper number of clusters for our mixture model of human rights, which resulted in 4 clusters. The tables above show the probabilities and standard errors of the clusters and variables within the model.

Conditional Independence Relations

We moved away from the mixture models and began testing for relationships between sets of variables. Testing the first parent-child relationship of the conjecture, we tested the predictive power of demographic variables on political views. To do this, we used a generalized additive model of the demographic variables predicting on the clustering of our earlier political view clustering. While the predictive power was not great at only 0.609 it was greater than predicting on the larger of the two clusters. These results showed a relationship from the conjecture and we, therefore, continued in our testing.

Demographic Predicting Membership

```
## Warning in chisq.test(polit.table1): Chi-squared approximation may be
## incorrect
```

```
## Warning in chisq.test(polit.table2): Chi-squared approximation may be
## incorrect
```

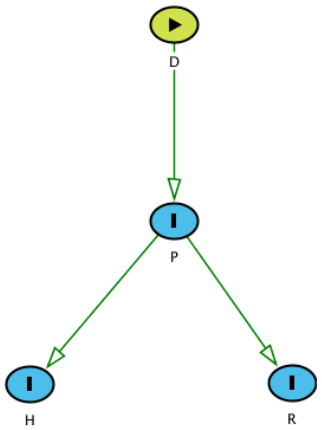
\paragraph After testing the first relationship in the researchers' model, we chose to test the final relationship, human rights, and the historical era. We created a new data frame containing the clustering of our three mixture models, split the data frame in two, based on the clustering of political values and ran chi-square tests for independence. The first cluster, the larger of the two, returned a p-value of 0.8117331, indicating dependence of variables. While the second cluster, the smaller of the two, returned a p-value of `r.pchi2$p.value`, Indicating independence of variables. Since the majority of the data is dependent, we categorized all the data as dependent. In other words, given a citizens political values, their historical era values and human rights values are associated.

Independence Given Membership

After testing the ends of the researchers' conjecture, we needed to test the middle relationship, to ensure validity of the theory. We ran a general additive model predicting historical era values from demographic values and political views. The resulting accuracy of our model is 0.715, which is greater than the accuracy of always choosing from the largest cluster, 0.702. Our model shows that given political values, one can predict historical era values using demographic values, which was the middle relationship in the researchers' theory. Using the previous section, we know a simialr relationship exists for human rights.

Directed Acyclic Graphs

Directed Acyclic Graph



Above is our graphical interpretation of the researchers conjecture of the data set. Historical and Human Rights values are caused by the a citizens general political values which are in turn caused by their pre-existing demographic variables. The graph implies that the estimated cluster memberships used in the previous sections should follow the same conditional independence relations as the actual latent variables.

Other Directed Acyclic Graphs

The graphical model above is the only graphical model which predicts this pattern of conditional independences among the observed and estimated variables from the researcher' conjecture. Assuming we follow that aforementioned conjecture, historical and human rights are children nodes of political which is a child of demographics. There is no room for other causal parents or children from the data set due to the language of the conjecture.

Missing Values

The previous sections all removed the 26 rows of missing values, under the assumption that those rows had no bearing on the theory or statistical work. By bootstrapping and testing for significance, we received a p-value of 2.3026973, which was insignificant under the alpha level of 0.05. Meaning, the rows removed had no special bearing on the data and could be removed without fear of tampering with the results.

Conclusions

We tested each relationship of the directed acyclic graph as well as the missing values of the data, set. We, therefore, felt confident in writing an assessment of the researchers' theory about the causal structure of the data. We saw from the previous sections, the dependence of political views on demographics, the dependence of political views for human rights and historical era views and the predictive ability of demographic ability on historical era views. Each of those steps showed parts of the directed acyclic graph that was the researcher's conjectures. By checking the missing values, we showed that there were not meaningful patterns obscured by missing data points. Therefore, we feel that the researchers' should be confident in their assessment of the data due to the aforementioned reasons.