

Predicting the Likelihood of Belief in Conspiracy Theories

Calix Barrus, Dallan Olsen, Laren Edwards, Elijah Cox

November 17, 2021

Abstract

The purpose of this project was to find what factors most correlate with a person believing or disbelieving in conspiracy theories. To discover this we took a dataset that scored peoples' belief of conspiracy theories based on their answers to certain questions. The dataset also included demographics and other information on the subjects which we used with a linear regression model to find which traits most correlate to a belief in conspiracy theories.

Introduction

Are there some factors that correlate with a person believing or disbelieving in conspiracy theories? Some believe that only uneducated people, or overly educated people, have such beliefs, but do these impressions hold any water? (At the very least, among those who would complete an on-line survey?) For our project we will use online survey data about belief in conspiracy theories to find what factors most correlate to one believing or disbelieving in conspiracy theories. This question has been a topic of research for several decades and has produced dozens of papers. Different studies focus on certain traits and how they correlate to conspiracy theory belief. For example, research done by Natasha Galliford and Adrian Furnham in "Individual difference factors and beliefs in medical and political conspiracy theories" attempted to find correlations between both demographic and psychological identifiers (age, race, sex, whether a person is introverted or extroverted, etc.) and belief in political and medical conspiracy theories. Using step-wise regression, they concluded that religion and age tended to be big factors in whether one is more accepting of conspiracies. Another study called, "Looking under the tinfoil hat: Clarifying the personological and psychopathological correlates of conspiracy beliefs" (Bowes et al., 2020) attempted to find correlation between abnormal personality and psychological traits and conspiracy theory belief. They used meta-regression to determine the strength of this correlation, and found that any personality trait that was correlated was usually weakly so. Our project will be using linear regression on a dataset to see if different factors will have a stronger correlation than the ones used above. The data used was taken from "Measuring belief in conspiracy theories: the generic conspiracist beliefs scale." It was a study that used the generic conspiracist beliefs scale (GCBS) score to quantify one's belief in conspiracy theories. The study also asked questions about demographic and personality traits which we will use to look for correlations. We used this dataset because it had an extensive amount of data, which we needed to get a good estimator, and because all the data was available to the public. However, it is important to keep in mind that the data came from an internet survey. Thus, it is likely that not every person took it seriously, and some people likely lied. To fix this, we cleaned the data by finding people who may not have taken the survey seriously (took little to no time to complete the survey, put the same answer for every question, entered invalid responses, etc.). We then ended with a clean dataset that we were able to analyze. As part of our analysis,

we found which traits correlated most with conspiracy theory belief. This is done by performing an L1 regularization on the data to find which traits had the strongest correlation. We then ran a linear regression on these traits to find exactly how they were correlated to conspiracy theory belief.

Data Cleaning

Data Cleaning and Feature Engineering The survey includes many important features including (but not limited to) age, education, religion, major, and vocabulary comprehension. Most were categorized and one-hot encoded, but not all. Here we go over the changes we made in the fields.

GCB The first field we created was the GCB field, or “General Conspiracy Belief”. The first 15 questions of the survey asked about level of belief in specific conspiracies on a 1-5 scale. The GCB is the average of these values for each person, so 2.5 would indicate an average level of belief of 2.5 in various conspiracies. This was used as the target “dependent” variable in our analysis. It simplifies the regression a lot to use an average of these answers rather than trying to have 15 different dependent variables.

Major The last field that we made large changes to was the “major” field. Here, survey respondents would put in their major as a string. Because this was a free response, we had hundreds of unique responses for major. Because there are so many majors (and so many major misspellings), we decided to group the responses together in larger categories. We used STEM, Humanities, Business/Law, Arts, and Other as our categories. Logically, this makes sense because we expect majors within the same sector to have similar correlations to a person’s belief in conspiracies. Because there were such varied responses, we manually assigned each major to a group. We then one-hot encoded these new variables to facilitate performance of classification.

Feature Engineering

We looked at the respondents who took the least amount of time and the most time to complete the survey, as well as the respondents who put the same answer for every question on belief in conspiracy theories. However, upon visual inspection of all of these edge cases, the demographic information seemed appropriately varied, and was completely filled out, implying that these rows should not be thrown out. We concluded that it is likely that the publishers of this survey data performed similar data cleaning and so any unfinished surveys or invalid responses would already be thrown out.

We create two new features from the data regarding vocabulary, namely `vocabulary_knowledge` (the percentage of real words the respondent knew) and `vocabulary_misclassification` (the percentage of fake words the respondent claimed to know).

Additionally, we create `y`, the feature we would like to predict, as the normalized average of responses to questions 1-15 on the survey. Those questions include, for example, “Secret organizations communicate with extraterrestrials, but keep this fact from the public” and “The spread of certain viruses and/or diseases is the result of the deliberate, concealed efforts of some organization”, and participants are asked to respond with their level of agreement from 1 to 5.

Results

The figure below shows the coefficient of determination R^2 and correlation coefficients for different features. None of our features seems to be strongly correlated with GCB, and we only account

for about 1% of the variance in beliefs with these features. Among what we did find, however, the strongest predictors of belief in conspiracy theories seem to be religion (in particular marking 'other' as one's choice of religion) and low knowledge of the given vocabulary words. The engineered features "vocabulary knowledge" and "vocabulary_misclassification" are what determines this.

Dep. Variable:	GCB	R-squared:	0.139
Model:	OLS	Adj. R-squared:	0.133
Method:	Least Squares	F-statistic:	22.29
Date:	Wed, 17 Nov 2021	Prob (F-statistic):	3.82e-68
Time:	12:56:52	Log-Likelihood:	579.98
No. Observations:	2495	AIC:	-1122.
Df Residuals:	2476	BIC:	-1011.
Df Model:	18		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
TIP12	0.0142	0.002	6.565	0.000	0.010	0.018
TIP15	0.0056	0.003	2.008	0.045	0.000	0.011
TIP16	-0.0014	0.002	-0.670	0.503	-0.006	0.003
vocabulary_misclassification	0.0706	0.017	4.051	0.000	0.036	0.105
STEM	-0.0383	0.010	-3.774	0.000	-0.058	-0.018
education_2	0.0380	0.009	4.120	0.000	0.020	0.056
education_3	0.0088	0.011	0.812	0.417	-0.012	0.030
urban_3	0.0242	0.008	2.882	0.004	0.008	0.041
gender_2	0.0219	0.008	2.707	0.007	0.006	0.038
engnat_1	0.0061	0.009	0.653	0.514	-0.012	0.025
religion_2	-0.0753	0.009	-8.266	0.000	-0.093	-0.057
religion_3	0.0873	0.028	3.119	0.002	0.032	0.142
religion_7	0.0603	0.014	4.179	0.000	0.032	0.089
religion_12	0.0996	0.013	7.947	0.000	0.075	0.124
orientation_2	-0.0120	0.011	-1.130	0.259	-0.033	0.009
orientation_5	0.0431	0.017	2.525	0.012	0.010	0.077
voted_2	0.0260	0.009	3.022	0.003	0.009	0.043
married_1	-0.0284	0.010	-2.842	0.005	-0.048	-0.009
constant	0.2304	0.013	17.307	0.000	0.204	0.257
const	0.2304	0.013	17.307	0.000	0.204	0.257

Omnibus:	85.135	Durbin-Watson:	1.925
Prob(Omnibus):	0.000	Jarque-Bera (JB):	42.486
Skew:	0.107	Prob(JB):	5.95e-10
Kurtosis:	2.398	Cond. No.	5.64e+17

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The smallest eigenvalue is 6.51e-31. This might indicate that there are strong multicollinearity problems or that the design matrix is singular.

Ethical Implications

It is important to remember that, although we have found traits with positive and negative correlation to conspiracy theory belief, our model has a very low R^2 score. That means that our model does not account for a lot of the variance in the data. Thus, just because someone may have many of the traits that correlate with conspiracy theory belief, that does not mean they do believe in conspiracy theories. Human beings are diverse and multi-faceted in their beliefs. Thus, absolute assumptions should not be made from the data above. Instead, the data above should be used as merely a soft guide to indicate that a person may believe conspiracy theories. This could be useful in anticipating a group's beliefs prior to giving them information about controversial topics, or in determining the issues most important to a group of people. But again, it should not be used to make absolute assumptions. Additionally, it is important that predictions from this model not be used to discriminate against individuals or disqualify them from anything, such as jobs or political offices. High correlations are good leads for other questions of research, such as "why is belief in conspiracy highly correlated with ____?" Understanding the causes behind the trends could help us answer questions about nature vs. nurture, psychology, and other topics in the field of study.