Investigating Predictors for American Attitudes on Justifiability of Abortion

STA 210 - Project

Bayes' Harem - Christina Wang, Kat Cottrell, David Goh, Ethan Song

Introduction and Research Question

Understanding public attitudes on divisive political issues is an important way for political leaders to mobilize voters and for lawmakers to draft laws that represent their constituents. While it may be easier to poll constituents' positions on an issue, it can be challenging to assess the complex factors that influence and predict those stances.

Abortion is one such divisive issue in the United States. Both pro-choice and pro-life groups have a history of mobilizing in states across the country in support of legislation for their respective sides (Ziegler, 2020). However, following the Supreme Court's 1973 decision in Roe v. Wade, a ruling which protected an individual's right to have an abortion before fetal viability, the issue has risen in political salience. Both the pro-choice and pro-life movements have gained national prominence, and the two major political parties have polarized around the issue, with the Democratic Party in favor of and the Republican Party against policies legalizing and increasing access to abortion (Weinberger 2022). Abortion has also increasingly become a key issue that voters consider when making their choice at the ballot box, with an increasing share of Americans identifying as "single-issue voters" regarding abortion (Brenan 2020).

In May 2022, a leaked draft opinion revealed that the US Supreme Court is prepared to overturn Roe v. Wade. Overturning Roe would dramatically change the trajectory of abortion politics in the US. Unless Congress passes a national policy, states would be able to decide whether or not to legalize abortion and gain much greater leverage in regulating access to the procedure (Weinberger 2022).

Given the potential overturning of Roe and the polarizing nature of the issue, it is important to understand how the American public feels about whether abortion should be legal or not, how accessible the procedure should be, and which factors influence these opinions. Understanding public opinion on the issue will ensure that political leaders are able to mobilize the correct constituencies, and that policy experts are able to pass policies on this issue that accurately reflect the preferences of the American people.

1

The dataset used in this analysis observes attitudes on the justifiability of abortion among respondents in the United States across six "waves" of the World Values Survey (1982-2011), which is administered every few years and collects information about people's values and beliefs worldwide, alongside basic demographic characteristics. The data are a nationally representative sample of the American population and are collected via face-to-face interviews at the respondents' homes. In this data set, the response variable, justifiability of abortion, is a numerical measure on a scale of 1 to 10 on the individual person's attitude toward whether abortion is justifiable or not. Individuals responded 1 for "abortion is never justified" and 10 for "abortion is always justified.

Our research question is as follows: "Do an individual's political ideology and their personal attitudes/preferences towards other issues, such as the importance of religion in their life and their respect for authority, among others, predict their attitude on the justifiability of abortion?" We will attempt to answer this question using an EDA-informed predictive model. Because the data are nationally representative, we can infer that all the observations are independent and the variables of interest may show a linear relationship. Given that this issue has become highly polarized by political party, we predict that liberal ideology and liberal-leaning attitudes on other issues will correlate with belief that abortion is more justified.

Exploratory Data Analysis
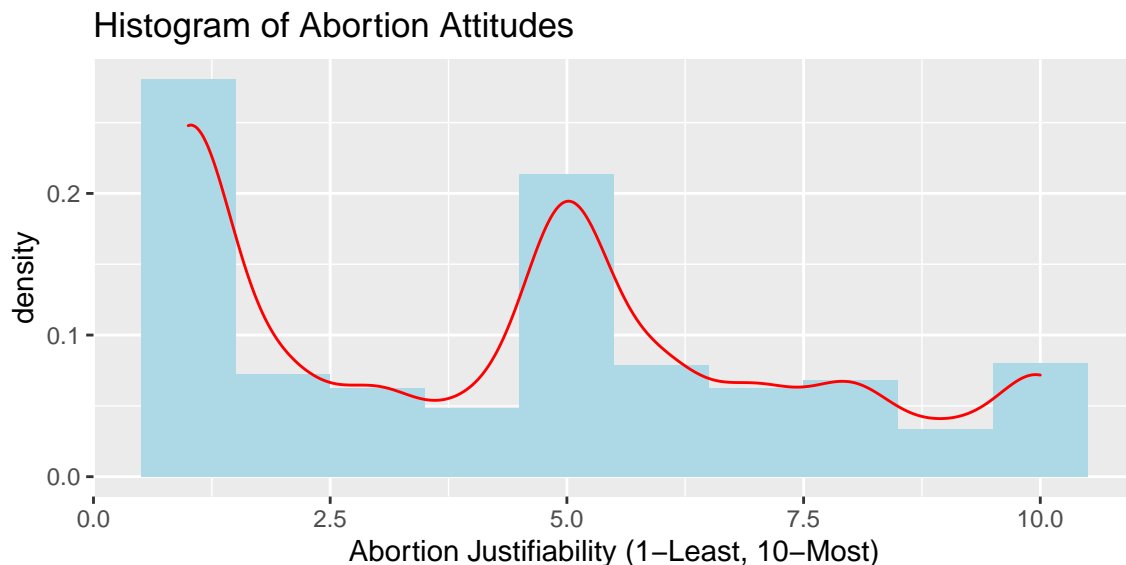
Histogram of Abortion Attitudes

Fig.1: The visual shows that the distribution of attitude toward abortion is trimodal, rather than bell–shaped. This may worsen our model's ability to fit the data.

From the above visualization of the distribution of attitudes toward the justifiability of abortion, we observe that it is not a bell shape, but rather trimodal. This may be because the

2

question's phrasing is similar to a yes/no question, but respondents were asked to give their level of agreement on a scale of 1-10. This may worsen the ability of a multiple linear regression (MLR) model to fit the data. Consequently, further research may choose to truncate the attitude on the justifiability of abortion into a categorical variable such as (Agree, Disagree, Undecided), conducting a binomial or multinomial logistic regression thereafter.

We now extend our exploratory data analysis (EDA) to the predictor variables of interest: year of survey, ideology, Child Autonomy Index, Importance of God, Respect for Authority, and National Pride. The EDA for each variable comprises a histogram and a boxplot of the response variable, grouped by value of the predictor. We also provide a correlation matrix to detect any multicollinearity between our predictor variables, which would increase the uncertainty of our model's parameters.
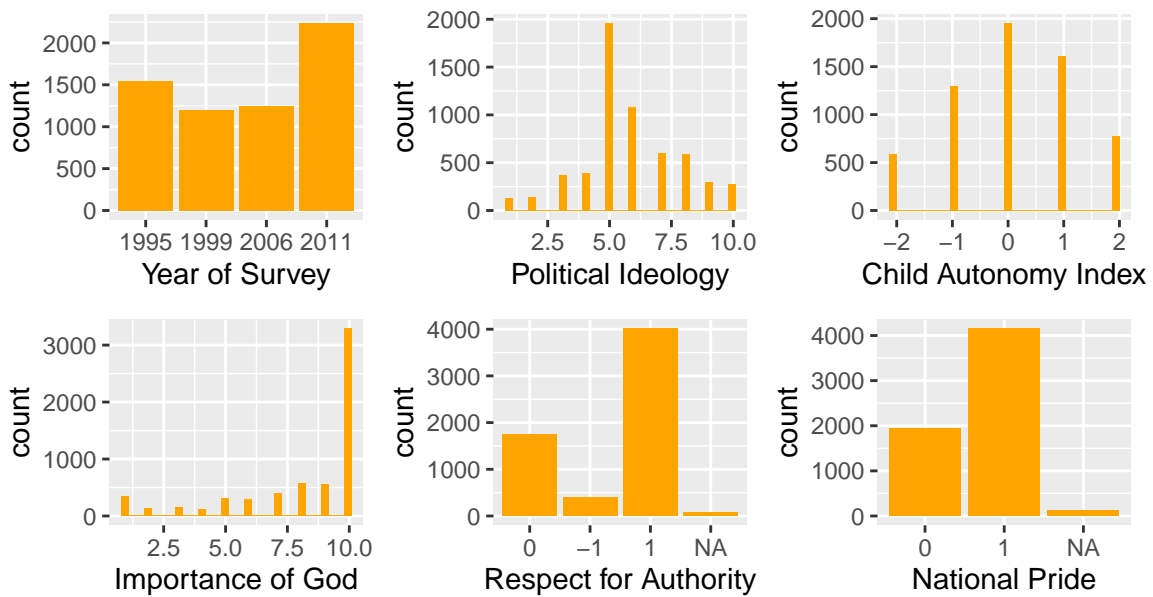
Distributions of variables of interest

Fig.2: The visuals show that while some variables have bell–shaped distributions, others are left–skewed.

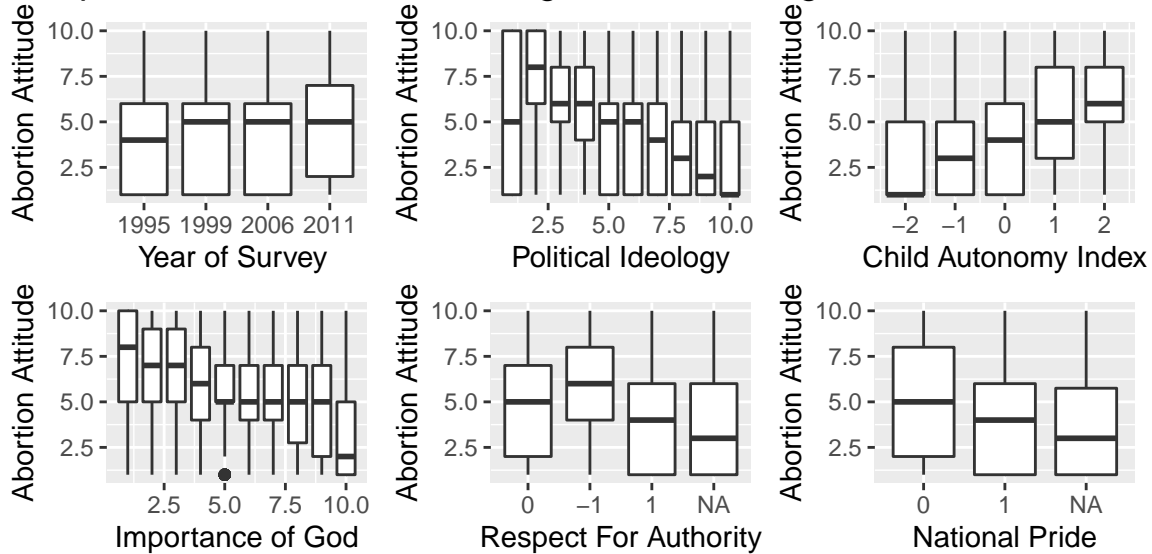## Boxplots of Abortion Attitudes against interesting variables

Fig.3: The visuals show that there may be a relationship between the response and each of these variables, making them likely predictors.

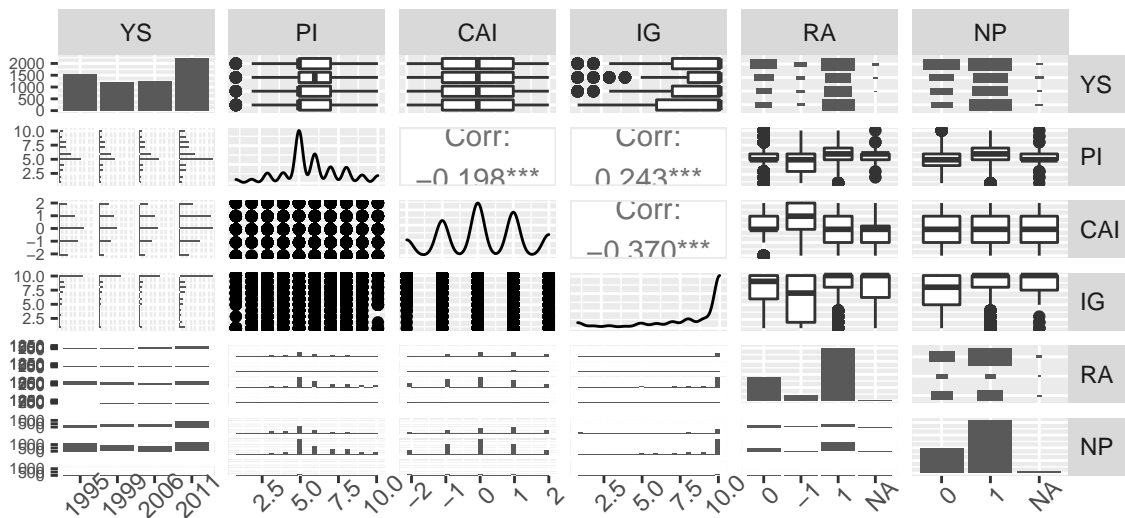## Correlation Matrix of Predictors of Interest

Fig.4: This matrix shows that the correlations between pairs of predictors are low, with the highest being 0.370.

Methodology

Given that our response variable, attitude towards the justifiability of abortion, is measured on a numeric scale from 1 to 10, and that there are multiple predictor variables being tested against the null hypothesis, we will conduct multiple linear regression (MLR) to model the effect of these predictors on the variation in abortion attitudes.

We chose to consider the predictor variables our exploratory data analysis (EDA) suggested were potentially correlated with the response variable. Because the predictor variables were discrete, our EDA was presented as a series of box plots. We observed the differences in median and quartile values of abortion attitude across different predictor values, ruling out variables with no significant influence on the outcome. This led to the following predictor variables: Year of survey, Political Ideology, Child Autonomy Index, Importance of God, Respect for Authority, and National Pride.

Many survey questions were only added from 1995 onward, so observations before 1995 have a high number of missing values. For these reasons, observations prior to 1995 were removed.

Year of survey, Respect for Authority, and National Pride were changed into factors since they are categorical variables. Year has 4 levels (one corresponding to each year), Respect for Authority has 3 levels (0 being the baseline - neutral, -1 being respect for authority bad, and 1 being respect for authority good), and National Pride is a binary (0 being the baseline - having no feelings of national pride, and 1 being having feelings of national pride) Political Ideology, Child Autonomy Index, and Importance of God had numerical sample spaces and were therefore treated as numerical variables.

Fitting the Model

We conducted a 75% - 25% data split into training and testing sets, using the random seed 206. Fitting the MLR, we obtain the following table of parameter point estimates.

Table 1: This table displays the estimate, standard error, statistic and p.value for each parameter of the model.

term	estimate	std.error	statistic	p.value
(Intercept)	6.596	0.170	38.688	0.000
ideology	-0.280	0.021	-13.354	0.000
cai	0.468	0.037	12.622	0.000
godimportant	-0.287	0.016	-17.776	0.000
year_X1999	0.330	0.120	2.743	0.006
year_X2006	0.339	0.120	2.826	0.005
year_X2011	0.521	0.107	4.889	0.000
respectauthority_X.1	0.140	0.173	0.807	0.420

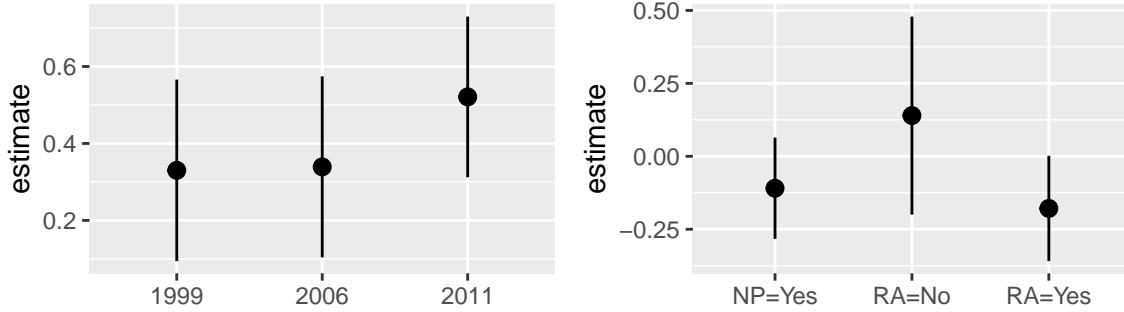| term	estimate	std.error	statistic	p.value
respectauthority_X1	-0.179	0.092	-1.941	0.052
nationalpride_X1	-0.110	0.089	-1.237	0.216

Hence, our MLR model equation is

$$y = 6.596 + 0.468x_1 + 0.33x_2 + 0.339x_3 + 0.521x_4 + 0.14x_5$$
$$- (0.28x_6 + 0.287x_7 + 0.179x_8 + 0.11x_9)$$

where x_1 = child autonomy index, x_2 = year 1999, x_3 = year 2006, x_4 = year 2011, x_5 = respect authority -1, x_6 = ideology, x_7 = God important, x_8 = respect authority 1, and x_9 = national pride.

Model Inference and Prediction

We will take a quick look at the 95% **confidence intervals** of the parameters from our fitted model, displayed in the following plots:
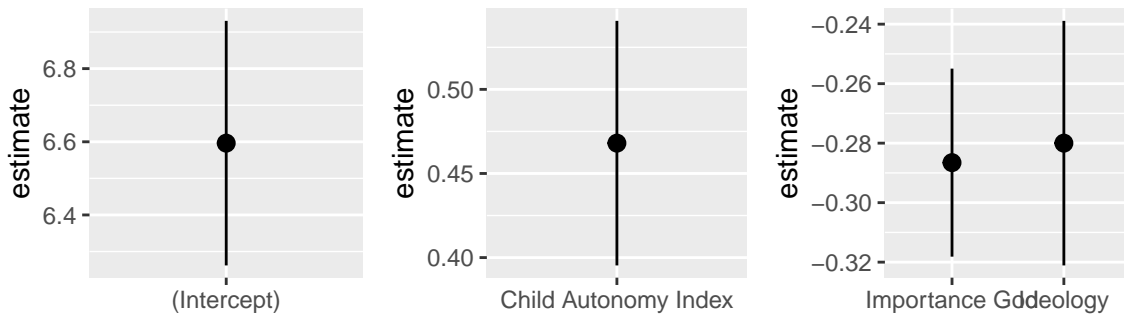

Fig.5: 95% confidence intervals of estimates for model parameters, using points and error bars

6

We then proceed to our **hypothesis test**. As stated in our introduction, given that this issue has become highly polarized by political party, we suspect that liberal political ideology and political attitudes will correlate with belief that abortion is more justified.

We define the following as liberal political ideology and attitudes: - ideology: values closer to 1 (most left-leaning) - cai: values closer to 2 (belief in child self-determination and independence) - godimportant: values closer to 1 (belief that God is not important) - respectauthority: -1 (belief that greater respect for authority is a bad thing) - nationalpride: 0 (not proud to be American)

This leads to the following hypotheses:

H_0 (null hypothesis): An individual's political ideology (ideology), their beliefs about child autonomy (cai), the importance of God in their life (godimportant), their view on whether respect for authority is good (respectauthority), and their pride in being American (national-pride) are not significant predictors of aj, and the variables' respective slope point estimates will be 0.

H_A (alternative hypothesis): An individual's political ideology (ideology), their beliefs about child autonomy (cai), the importance of God in their life (godimportant), their view on whether respect for authority is good (respectauthority), and their pride in being American (national-pride) are significant predictors of aj, and the variables' respective slope point estimates will not be 0.

Looking at the p-values in our output displayed above, we can reject the null hypothesis for the predictors political ideology, child autonomy index, and importance of God, as the p-values are very small, way below the 0.05 significance threshold. Hence, we can conclude that the data provide convincing evidence that an individual's political ideology, beliefs about child autonomy, and the importance of God in their life are significant predictors of their abortion attitudes.

However, we fail to reject the null hypothesis for the predictors respect for authority and national pride, as the p-values for each level of these predictors are above the 0.05 significance threshold. Hence, we cannot conclude that the data provide convincing evidence that an individual's view on respect for authority and their pride in being American are significant predictors of their abortion attitudes.

We now consider the strength of our model in **predicting** the attitude toward the justifiability of abortion of a person of certain characteristics.

Earlier, we guessed that liberal ideology and liberal-leaning attitudes on other social issues will correlate with belief that abortion is more justified. For a person surveyed in 2011 who identifies with a left leaning political ideology ("1"), has a high child autonomy index of 2, considers God and religion unimportant ("1"), and has little respect for authority, our our model predicts that their attitude toward the justifiability of abortion is 9.24, that is to say, very amenable to the justifiability of abortion. This supports our hypothesis based on contextual knowledge of this issue.

However, calculating the R-squared and root-mean-squared-error (RMSE) values of our model shows that our model is not very strong at predicting attitude toward the justifiability of abortion, with a R-squared of 0.247 and RMSE of 2.53.

We will discuss potential reasons for this lackluster performance in our model diagnosis and conclusion sections; however, we will first conduct model evaluation and comparison to ascertain if our chosen model is still a better fit to the data compared to some alternatives.

A quick glance at our data might seem that the models are not that strong at predicting the abortion attitude (aj) of a given observation. We will look into this further now with cross-validation, then a test in AIC and BIC statistics.

Model Evaluation

To see how well our model fits the data on the attitudes toward the justifiability of abortion, we conducted a 10-fold cross-validation with seed 206, calculating various metrics – the Akaike information criterion (AIC), Bayesian Information Criterion (BIC) and adjusted R-squared value – to evaluate the performance of our model on 10 subsets of our data.

We also compared this to two alternative models that used other combinations of predictors from the data set.
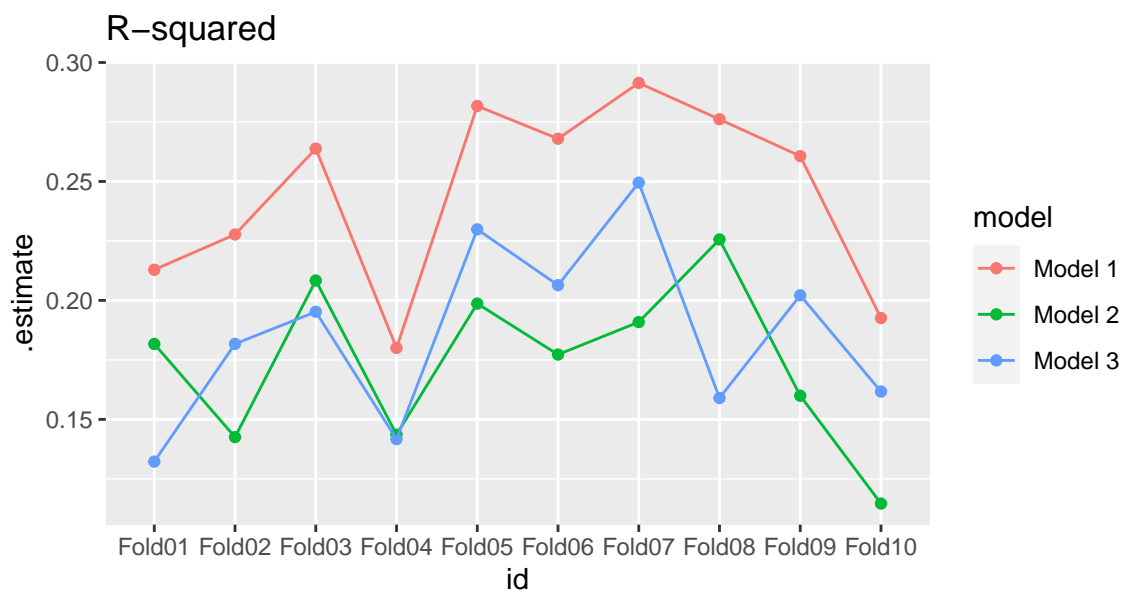

Fig.6: This plot of R–squared values demonstrates our chosen model 1 is the best of the three.

From the above visualization of R-squared values, we have evidence that our chosen model is the strongest model and best fit for the data compared to two alternatives with fewer predictors. Below we will observe the AIC, BIC, and Adjusted R-squared values for each model as well.

8

Table 2: With the lowest AIC and BIC and highest adjusted R-squared value, our chosen model is best of the three.

	AIC	BIC	Adj. R-squared
Model 1	19839.43	19909.27	0.246
Model 2	21483.39	21521.84	0.174
Model 3	20333.67	20384.53	0.186

The table above further supports that Model 1 is our best fitting model.

Model Diagnosis

Below we will check each of the following conditions for our model: Linearity, Constant Variance, Normality, and Independence.

As shown in the following plot of residuals against fitted values, the linearity condition is not met as the residuals are not scattered around the horizontal red line. The constant variance condition is also not met.
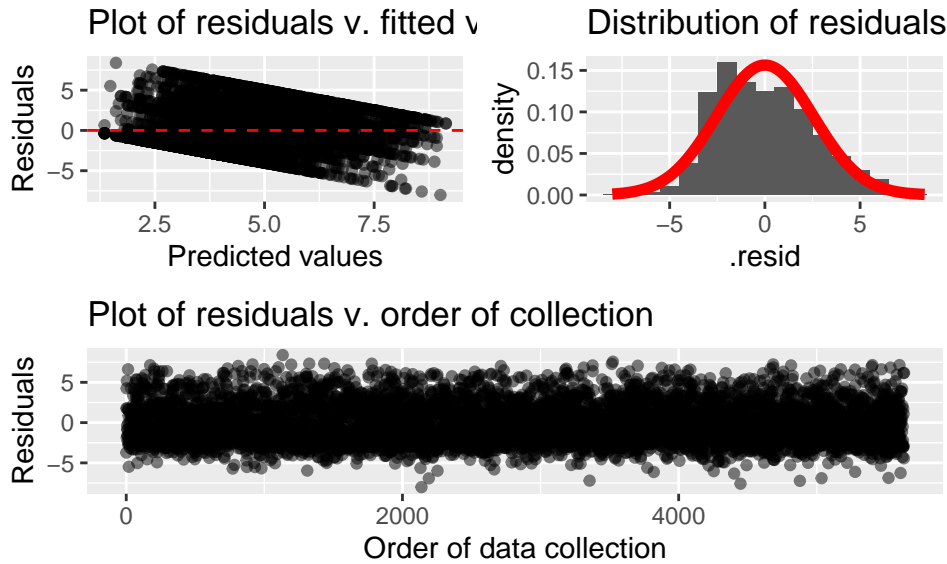

Fig.7: The linearity condition and constant variance condition is clearly not met in this plot of residuals v. fitted values.

Fig.8: The residuals appear to follow a normal distribution, which looks like a bell-shape, as illustrated in red.

Fig.9: The residuals plotted by observation order are randomly scattered, showing no structure.

9

Conclusion + Discussion

The model that we fitted to predict abortion attitudes generally confirmed our initial research question that liberal ideology and attitudes towards other issues predicts abortion attitudes. Inference on the model (through both a confidence interval and hypothesis test) showed that having liberal political ideology, low belief in the importance of God, and belief in greater child autonomy generally predict a belief that abortion is more justified. However, holding liberal attitudes on certain other issues, namely believing that respect for authority is bad and not having pride in being American, do not necessarily predict greater justifiability of abortion.

Despite the model confirming our initial research question, we have reason to question the quality of the model, and whether the model accurately fits the data on abortion attitudes. The first reason for questioning our model's quality is the low R-squared value, which indicates that the variance of our predictors may not greatly account for the variance in abortion attitudes. The second reason for questioning our model's quality is that our model does not fit the linearity and constant variance conditions. Upon closer examination of the data, we can see that the distribution of our response variable was trimodal, with peaks at the extremes and in the middle. This likely led to the data having a large number of very different observations that had different predictor values, but the same `aj` score, and may be one explanation our model's weaknesses.

Given the nature of public opinion on political issues, we are likely limited by the precision and accuracy of each observation in our dataset. Previous political science research on public opinion indicates that Americans' political beliefs are inconsistent and volatile. For example, most Americans' political beliefs have been found to be shaped by partisan and group loyalties, rather than ideology or any specific set of issue positions. Indeed, even for committed partisans, one individual's issue positions are likely to to be scattered across the ideological spectrum, and may even change over short periods of time (Achen and Bartels, 2017). This inconsistency between partisanship, ideology, and issue positions may account for the non-normal distribution of abortion attitudes in the dataset and the model's poor fit to the data. In addition, although the data are supposed to capture a nationally representative sample of the US population, we do not know if they are weighted by other factors that may represent important distinctions in the US population. This possibility of non-weighting may have influenced our model's quality as well.

Future work may build on this analysis and correct some of its weaknesses. First, the data should be weighted by more relevant factors to ensure that the sample is as representative as possible of the US population. Relevant factors may include, but are not limited to, partisan affiliation, race, religion, socioeconomic status, and geographic area. In addition, it may be helpful to consider other predictors that may influence abortion attitudes, such as partisan affiliation and race. However, it is important to keep in mind that abortion is an incredibly divisive and complex issue, so it is not an easy task to precisely pinpoint any individual's abortion attitude, no matter which predictors are used.

Data dictionary

The data dictionary can be found here.

References

Achen, C.H., & Bartels, L.M. (2017). Democracy for realists: Why elections do not produce responsive government. Princeton University Press.

Brenan, M. (2020, July 7). One in Four Americans Consider Abortion a Key Voting Issue. Gallup. https://news.gallup.com/poll/313316/one-four-americans-consider-abortion-key-voting-issue.aspx

Miller, Steven V. (2016, Jul 31). "Attitudes on the Justifiability of Abortion in the United States." GitHub. https://github.com/svmiller/wvs-usa-abortion-attitudes

Weinberger, J. (2022, May 6). How we got here: Roe v. Wade from 1973 to today. Vox. https://www.vox.com/23055389/roe-v-wade-timeline-abortion-overturn-political-polarization

Ziegler, M. (2020, October 22). Abortion politics polarized before Roe. When it's gone, the fighting won't stop. The Washington Post. https://www.washingtonpost.com/outlook/2020/10/22/roe-polarize-abortion-politics/