Gov 1372 - Hierarchies

Yanxi Fang

September 30, 2021

```
library(tidyverse)
library(stargazer)
```

Question 1: REQUIRED

Note: my survey screenshots, named `survey_1.png` *and* `survey_2.png` *(for the dominance and anti-egalitarianism sub-scales, respectively), can be found on GitHub at the following link: https://github.com/yanxifang/Gov-1372/tree/main/images.*

I chose to structure my survey by keeping the existing order of questions and by making the response categories (e.g. strongly favor, somewhat favor, etc.) relatively discrete.

I kept the existing order because when I first saw it in class, I was pleasantly surprised by the fact that the second half of the statements in each sub-scale directly contradict the corresponding first half, and I thought that such a contradiction would prompt respondents to re-evaluate their responses and ultimately provide the responses that are closest to how they actually feel. I also kept the two sub-scales separate from each other, with the goal of locking in some answers from each respondent.

I made the response categories relatively discrete (i.e. as opposed to continuous) by choosing not to present them as being on a linear continuum (such as with a slider). By making it slightly less "fun" to answer the survey, my goal was to implicitly reinforce the seriousness of the questions being asked, while also highlighting the distinctions between each category to try to get the most accurate responses possible.

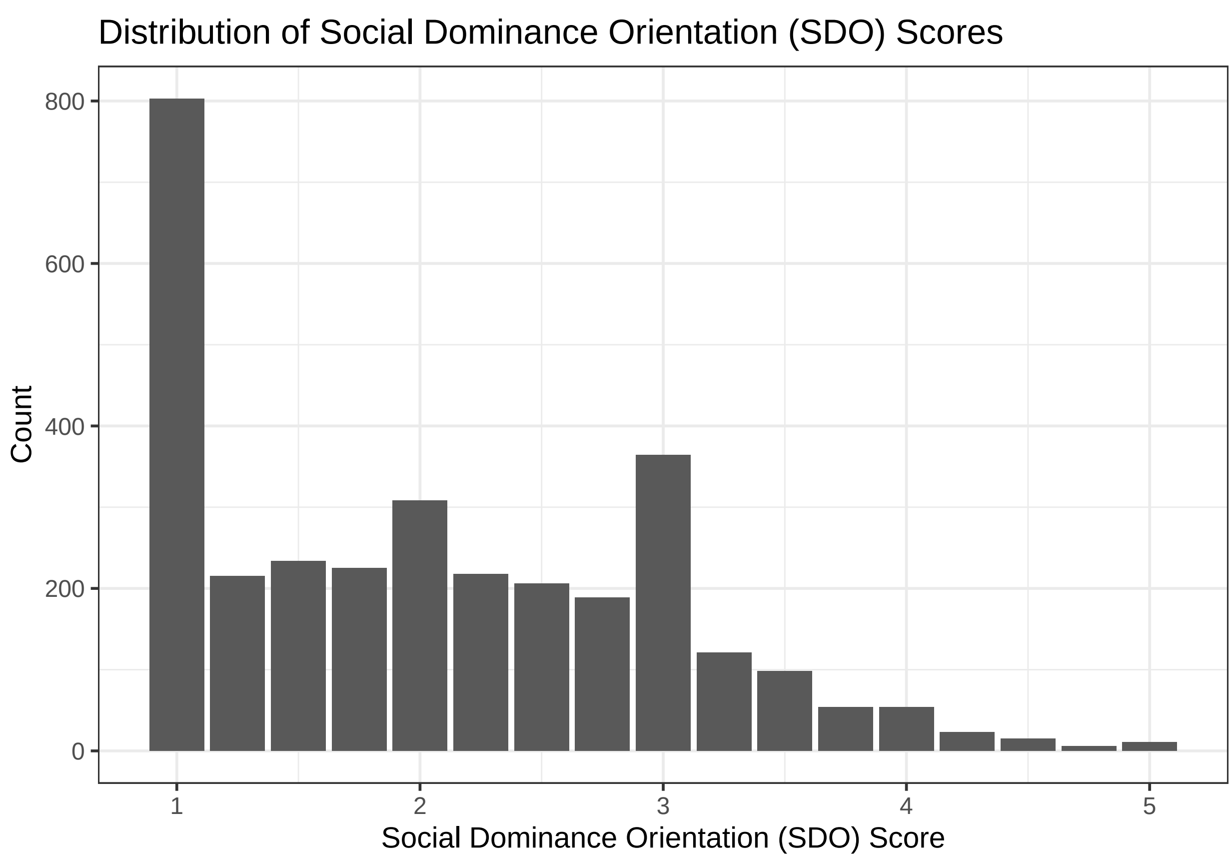
Question 2

Below, the distribution of the social dominance orientation (SDO) scores is plotted. As shown graphically, there are relatively very few respondents with SDO scores greater than 3, with a plurality (or even majority) of respondents having SDO scores lower than 2. Numerically, the mean SDO in the sample is 2.053, with a standard deviation of 0.9215.

The second plot below splits up the respondents by their party identification. The findings are consistent with the conclusions made in the Sidanius reading on SDO: the Democrats have significant more "1"s and the SDO distribution is more skewed to the right (i.e. the mode is near the left of the plot), meaning that they are less dominant on average, while the Republicans have a more normal-looking distribution, suggesting a greater average level of domination.

```r
# read in data
sdo <- read_csv("sdo_data.csv")

# create histogram
sdo %>%
  ggplot(aes(x = sdo5)) +
  geom_bar() +
  xlab("Social Dominance Orientation (SDO) Score") +
  ylab("Count") +
  ggtitle("Distribution of Social Dominance Orientation (SDO) Scores") +
  theme_bw()
```



```r
# mean and std. dev.
mean(sdo$sdo5)
```

```
## [1] 2.05256
```

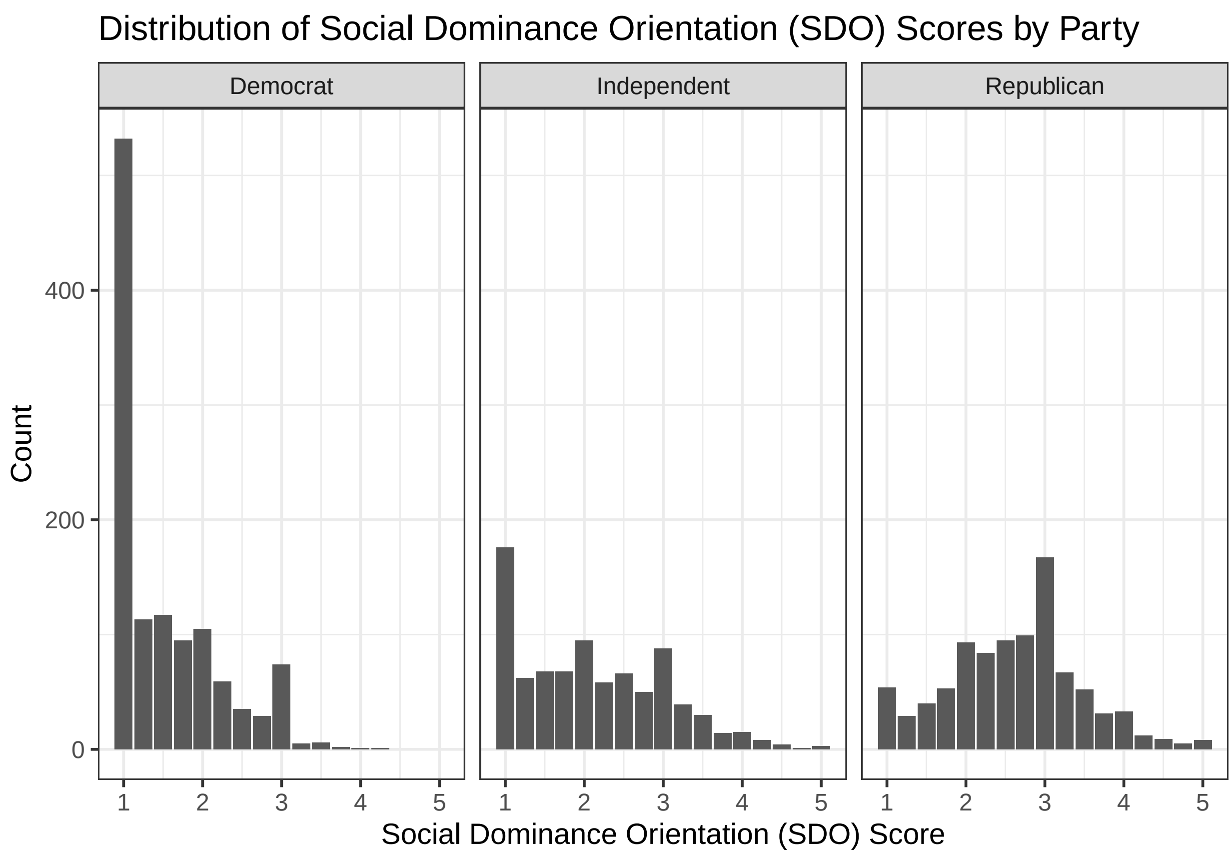
```r
sd(sdo$sdo5)
```

```
## [1] 0.9214715
```

```r
# extension: by party identification
# modify data
sdo <- sdo %>%
  mutate(party = ifelse(pid3 == 1, "Democrat",
                ifelse(pid3 == 2, "Independent", "Republican")))

# remove party identification NAs
sdo_party <- sdo %>%
  filter(is.na(party) == FALSE)

# plot by party
sdo_party %>%
  ggplot(aes(x = sdo5)) +
  geom_bar() +
  xlab("Social Dominance Orientation (SDO) Score") +
  ylab("Count") +
  ggtitle("Distribution of Social Dominance Orientation (SDO) Scores by Party") +
  facet_wrap(~party) +
  theme_bw()
```




```
# save plot
ggsave("social_dominance_byparty.png", height = 4, width = 8)
```

Question 6: Data Science Question

The `white` and `rep_vote` variables are created below, denoting white versus nonwhite respondents and Trump versus non-Trump voters, respectively. Subsequently, the following model is fitted; the results are printed on the next page.

$$\text{rep\_vote} = \beta_0 + \beta_1\text{sdo5} + \beta_2\text{female} + \beta_3\text{white} + \beta_4\text{educ} + \beta_5\text{age} + \beta_6\text{pid3} + \beta_7\text{ideo5} + \epsilon$$

The regression findings are in line with what we would expect based on social dominance theory. In the linear specification of the model, party identity and ideology are numerically the most significant explanatory variables for determining whether an average person voted for Trump in the 2016 election, followed by the 5-point SDO scale; all three have positive and statistically significant coefficients. Since the outcome variable is coded such that a `0` represents a non-Trump vote and `1` represents a Trump vote, the positive coefficients mean that on average, a person with a higher value in any of those three categories (while holding everything else equal) is more likely to vote for Trump.

While the first two variables are technically not included in SDO, they make sense because intuitively, one would expect conservative-leaning and Republican-affiliated respondents to be more likely to vote for Trump. Similarly, higher SDO scores are correlated with someone being more likely to vote for Trump. This aligns with SD theory because people with higher SDO scores are more likely to be conservatives (and thus vote for a candidate representing the Republican Party). Here, the numerical coefficient is difficult to interpret because linear regression makes the assumption that the outcome variable will be continuous (here, it is discrete – to be exact, binary), but one approach to interpretation is that on average, an increase on a person's SDO score, holding all other factors constant, is predicted to increase that person's likelihood of voting for Trump in the 2016 election.

```r
# duplicate existing dataset for manipulation
sdo_vote <- sdo

# create `white` and `rep_vote` variables
sdo_vote <- sdo_vote %>%
  mutate(white = ifelse(race == 1, 1, 0)) %>%
  mutate(rep_vote = ifelse(presvote16 == "Trump", 1,
                    ifelse(presvote16 == "Did not vote", NA, 0)))

# fit models
# note: the second model is used only in my blog post, not in this assignment
votemodel1 <- lm(rep_vote ~ sdo5 + female + white + educ + age + pid3 + ideo5,
                 data = sdo_vote)

votemodel2 <- glm(rep_vote ~ sdo5 + female + white + educ + age + pid3 + ideo5,
                 data = sdo_vote, family = "binomial")
```

```r
# print stargazer table
stargazer(votemodel1, votemodel2, header = F)
```

Table 1:

	Dependent variable:	
	rep_vote	
	OLS	*logistic*
	(1)	(2)
sdo5	0.093***	0.944***
	(0.007)	(0.095)
female	−0.002	−0.018
	(0.011)	(0.154)
white	0.043***	0.321
	(0.015)	(0.206)
educ	−0.017***	−0.240***
	(0.004)	(0.057)
age	0.001***	0.018***
	(0.0004)	(0.006)
pid3	0.264***	1.976***
	(0.010)	(0.129)
ideo5	0.111***	1.294***
	(0.007)	(0.101)
Constant	−0.646***	−10.787***
	(0.034)	(0.602)
Observations	2,749	2,749
R^2	0.681	
Adjusted R^2	0.680	
Log Likelihood		−612.019
Akaike Inf. Crit.		1,240.038
Residual Std. Error	0.281 (df = 2741)	
F Statistic	835.558*** (df = 7; 2741)	
Note:	*p<0.1; **p<0.05; ***p<0.01	