

# Personality and Worldview

Yanxi Fang

October 7, 2021

## Question 1

The first photo that I analyzed was 36449, which had a response of 3 on the political ideology scale, 4 on the BFI conscientiousness, 2 on the BFI openness, 5.5 on the TIPI conscientiousness, and 4 on the TIPI openness. After evaluating the PLSCI, my survey responses were 1 on ideology, 4 on conscientiousness, and 2.5 on openness. Here, my ideology response was off (likely because it was more influenced by an assumption about the political ideologies of Harvard students than by the actual characteristics of the room), but my conscientiousness and openness estimates were very close to the actual BFI responses. The TIPI scores are harder to compare because they were on a 10-point scale (as opposed to the 5-point scale that I scored the person with – and the 5-point BFI), but an interesting thing is that the person behind the photo had less polarized responses on TIPI than on BFI (5.5 and 4 as opposed to 4 and 2), so perhaps there is something different about how people respond to those two methods of the Big Five Inventory.

The second photo that I analyzed was 62476, which had a response of 3 on the political ideology scale, 5 on the BFI conscientiousness, 3 on the BFI openness, 6.5 on the TIPI conscientiousness, and 6 on the TIPI openness. After evaluating the PLSCI, my survey responses were 1 on ideology, 2 on conscientiousness, and 4 on openness. Here, my responses were all completely off. Again, the ideology response was more driven by my background understanding about the political ideologies of Harvard undergraduates, while my estimates of the conscientiousness and openness dimensions each went the opposite way from the center compared to the actual response. Furthermore, as described above, it appears that this person also had less polarized responses on TIPI than on BFI (a basically equivalent 6.5 and 6 on TIPI, compared to 5 and 3 on BFI).

Something interesting I noticed is that the first image 36449, where I was more accurate, was actually a stock photo, while the second image 62476 was a real photo of a Harvard dorm. Thus, one theory to explain this (assuming that I actually have some capability of judging personality from photos) is that a stock photo might be a better reflection of their true personality, since stock photos allow respondents to choose what their ideal living environment would look like (thus reflecting preferences to some degree), rather than having them show their actual living environment that they may not necessarily be happy with (e.g. with dorms, lighting and furniture setup may not be ideal).

## Question 2

After conferring with the others assigned to each of the two photos, I noticed that there were significant differences when it came to more subjective parts of the PLSCI, such as how modern, stylish, well-lit, or inviting a photographed room might be. This is completely normal because different people have different baselines (e.g. our own rooms) against which the photographs were compared. Interestingly, these differences in PLSCI did not necessarily translate into significantly different guesses about openness, conscientiousness, and ideology, since it wasn't exactly clear which specific criteria (out of the many on the PLSCI) should be used to make those guesses. Nonetheless, the differences in PLSCI point out the issues with inter-coder reliability: generally, in an experiment like this, lower reliability among different coders would imply that the scale is not necessarily a good one to use, because if coders see the questions differently, then the results would be not only less accurate, but also less convincing.

## Question 7: Data Science Question

```
# set up
library(tidyverse)
library(stargazer)

# read in data
personality <- read.csv("Oct7ClassData.csv")

# run OLS regressions
ols_personality1 <- lm(Overall ~ BFI_extraversion + BFI_agreeableness +
  BFI_conscientiousness + BFI_emot_stability +
  BFI_openness, data = personality)
ols_personality2 <- lm(Social ~ BFI_extraversion + BFI_agreeableness +
  BFI_conscientiousness + BFI_emot_stability +
  BFI_openness, data = personality)
ols_personality3 <- lm(Economic ~ BFI_extraversion + BFI_agreeableness +
  BFI_conscientiousness + BFI_emot_stability +
  BFI_openness, data = personality)
ols_nothing <- lm(Overall ~ 1, data = personality)

# alternative approach: change outcome variable to (0, 1) binary
# ... and then run a multivariate logit regression

# first, sum the social and economic scores
# this will be useful for evaluating people with `overall = 3`
# for people with `overall = 3`:
# if score is 5 or less (e.g. 2 and 3), then the person is liberal
# if score is 7 or greater (e.g. 3 and 4), then the person is conservative
# if score is 6 (e.g. 3 and 3), then the person is excluded
# ... because a middle-of-the-road score means that they are neither lib. nor cons.
# then, create 0-1 binary variable
# 0 is for liberals, 1 is for conservatives

personality <- personality %>%
  mutate(sum_socialeconomic = Social + Economic) %>%
  mutate(Overall_binary = ifelse(Overall == 1 | Overall == 2, 0,
    ifelse(Overall == 4 | Overall == 5, 1,
      ifelse(Overall == 3 & sum_socialeconomic <= 5, 0,
        ifelse(Overall == 3 & sum_socialeconomic >= 7, 1, NA)))))

# look at outputs
table(personality$Overall_binary)

##
## 0 1
## 66 5

# not so great for outcomes: very few conservatives, making it hard to be sure...
# ... that the model is actually a good fit for the data

# second try: put all `overall = 3` into conservative category
```

```

personality <- personality %>%
  mutate(Overall_binary_2 = ifelse(Overall == 1 | Overall == 2, 0, 1))
table(personality$Overall_binary_2)

##
## 0 1
## 55 26

# this approach is considerably better
# but I'll still run one regression with each of these approaches

# run logit regressions
logit_personality1 <- glm(Overall_binary ~ BFI_extraversion + BFI_agreeableness +
  BFI_conscientiousness + BFI_emot_stability +
  BFI_openness, data = personality, family = "binomial")
logit_personality2 <- glm(Overall_binary_2 ~ BFI_extraversion + BFI_agreeableness +
  BFI_conscientiousness + BFI_emot_stability +
  BFI_openness, data = personality, family = "binomial")

# also, trying OLS with the binary outcome variable
ols_personality_b1 <- lm(Overall_binary ~ BFI_extraversion + BFI_agreeableness +
  BFI_conscientiousness + BFI_emot_stability +
  BFI_openness, data = personality)
ols_personality_b2 <- lm(Overall_binary_2 ~ BFI_extraversion + BFI_agreeableness +
  BFI_conscientiousness + BFI_emot_stability +
  BFI_openness, data = personality)

```

For this question, I am hoping to see whether the trait indices (extraversion, agreeableness, conscientiousness, emotional stability, and openness), as measured by the BFI inventory, can explain the responses to the political ideology questions (overall, social, and economic liberalism/conservatism). I ran several regression models with the code above. The first three are OLS models; each uses the five trait indices as the independent variables, while the outcome variables were the overall, social, and economic ideology, respectively.

After previewing the dismal results, which showed that none of the variables had statistical significance even at the lax  $\alpha = 0.10$  level, I then ran a fourth model with no independent variables and the overall ideology as the outcome variable; here, I know that the intercept would just be the mean of the y-variable (i.e. the mean of the overall ideology across the entire dataset), so this might be a good way to compare the intercepts of the three other models.

The results are printed on the next page. They are not particularly helpful for advancing the theory that traits can explain a person's political ideology. As previously mentioned, none of the variables had statistical significance, meaning that we cannot rule out the possibility that each of the variables has a true coefficient of 0 (in other words, the possibility that none of the variables actually matter in explaining ideology). In fact, the adjusted  $R^2$  values for the three models are all negative; we would usually expect  $R^2$  to be between 0 and 1, and the negative value implies that the fitted model performs worse than a horizontal line (which would be the fourth model, with only a y-intercept) in explaining the relationship between the explanatory variables and the outcome variable.

In running these OLS models, I am making an assumption about linearity (that the true relationship between the explanatory variables and the outcome variable is going to follow a roughly linear trend), as well as an assumption that the outcome (dependent) variable is continuous (because by definition, lines necessarily encompass all values between integers). While the results suggest that the first assumption is wrong, I already knew that the second assumption is wrong, because the outcome variable is discrete, not continuous, since it has five possible categorical levels (1 through 5, on the scale of political ideology).

Thus, I chose an alternative approach. Instead of using a 1-to-5 categorical variable to represent the outcome (dependent) variable of ideology, I decided to convert that into a 0-or-1 binary variable, with 0 representing liberals and 1 representing conservatives. The idea is that having a binary outcome variable would allow me to run a logistic model, which is a different model specification and thus has the potential to offer new insights into the relationship between personality and ideology.

To operationalize this, I decided that people with 1s and 2s in the Overall category would be automatically considered liberals, while 4s and 5s would be automatically considered conservatives. People with 3s would be placed in a category based on the sum of their Social and Economic ideology scores. Since those scores were also on a 1-to-5 scale (so the mid-point for each would be a 3), I decided that a total of 5 or less on those two scores would represent a liberal, while a total of 7 or more on those two scores would represent a conservative, with 6s being removed from consideration because they would be centrists, in a sense. Unfortunately, this approach yielded only 4 conservatives (i.e. four 1s in the resulting variable), which is problematic for regression because each of the 4 observations would have such a significant impact on the regression results (in the sense that removing one of the four, or adding a fifth, would change the results by a lot).

Thus, I took an alternative approach: reasoning that most Harvard students are liberal-leaning, and may thus feel peer pressure to identify themselves further to the left than they might actually be (a phenomenon that would be in line with the theories about in-groups), I decided to code all the 3s as conservative, leaving only 1s and 2s as liberals. This produced a much more reasonable 26 conservatives.

With each of these two binary variables as the outcome variable, I ran logistic regressions, using the same five independent/explanatory variables. I also ran parallel OLS regressions for comparative purposes, even though I know that assumptions are being violated (as previously described). The results of these four regressions are printed on the page after next, in Table 2. Again, none of the explanatory variables have statistically-significant coefficients, so the interpretation is the same as before: we can't rule out the possibility that the all of the explanatory variables, the five personality measures, have a true effect of zero (i.e. no effect) on the outcome variable, a person's political ideology.

```
# print stargazer table
stargazer(ols_personality1, ols_personality2, ols_personality3, ols_nothing,
          header = F, column.sep.width = "Opt", font.size = "small",
          title = "(OLS) Regressions with Unmodified Variables")
```

```
stargazer(logit_personality1, logit_personality2,
          ols_personality_b1, ols_personality_b2,
          header = F, column.sep.width = "Opt", font.size = "small",
          title = "Regressions with Binary Outcome Variable")
```

Table 1: (OLS) Regressions with Unmodified Variables

	<i>Dependent variable:</i>			
	Overall	Social	Economic	Overall
	(1)	(2)	(3)	(4)
BFI_extraversion	0.016 (0.091)	0.046 (0.086)	0.101 (0.125)	
BFI_agreeableness	-0.185 (0.118)	-0.117 (0.112)	-0.180 (0.162)	
BFI_conscientiousness	0.063 (0.095)	0.075 (0.090)	0.043 (0.130)	
BFI_emot_stability	0.033 (0.089)	-0.018 (0.084)	-0.007 (0.121)	
BFI_openness	-0.104 (0.096)	-0.148 (0.091)	-0.015 (0.131)	
Constant	2.812*** (0.617)	2.275*** (0.581)	2.554*** (0.843)	2.198*** (0.087)
Observations	80	80	80	81
R <sup>2</sup>	0.045	0.048	0.020	0.000
Adjusted R <sup>2</sup>	-0.020	-0.016	-0.047	0.000
Residual Std. Error	0.794 (df = 74)	0.748 (df = 74)	1.084 (df = 74)	0.781 (df = 80)
F Statistic (df = 5; 74)	0.692	0.744	0.298	

*Note:*

\*p&lt;0.1; \*\*p&lt;0.05; \*\*\*p&lt;0.01

Table 2: Regressions with Binary Outcome Variable

	<i>Dependent variable:</i>			
	Overall_binary	Overall_binary_2	Overall_binary	Overall_binary_2
	<i>logistic</i>	<i>logistic</i>	<i>OLS</i>	<i>OLS</i>
	(1)	(2)	(3)	(4)
BFI_extraversion	0.090 (0.488)	0.076 (0.249)	0.011 (0.033)	0.016 (0.054)
BFI_agreeableness	-0.596 (0.637)	-0.542 (0.334)	-0.045 (0.041)	-0.113 (0.071)
BFI_conscientiousness	0.346 (0.611)	0.086 (0.268)	0.019 (0.034)	0.016 (0.057)
BFI_emot_stability	-0.605 (0.564)	0.135 (0.241)	-0.042 (0.035)	0.028 (0.053)
BFI_openness	-0.478 (0.508)	-0.393 (0.270)	-0.035 (0.034)	-0.081 (0.057)
Constant	1.095 (3.323)	1.483 (1.708)	0.366 (0.223)	0.801** (0.369)
Observations	70	80	70	80
R <sup>2</sup>			0.053	0.051
Adjusted R <sup>2</sup>			-0.021	-0.013
Log Likelihood	-16.282	-48.337		
Akaike Inf. Crit.	44.565	108.674		
Residual Std. Error			0.262 (df = 64)	0.474 (df = 74)
F Statistic			0.713 (df = 5; 64)	0.801 (df = 5; 74)

Note:

\*p&lt;0.1; \*\*p&lt;0.05; \*\*\*p&lt;0.01