Exploring the Relationship Between Religion and Trust in Science

Enzo Moraes Mescall, Charlotte Coudert, Martin Olarte

2022-12-08

Abstract

Introduction

Data

Data Cleaning

Exploratory Data Analysis

Modeling

```
library(haven)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(stringr)
library(purrr)
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --

## v ggplot2 3.3.6     v readr   2.1.2
## v tibble  3.1.8     v forcats 0.5.2
## v tidyr   1.2.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(MASS)
```

```
##
## Attaching package: 'MASS'
##
## The following object is masked from 'package:dplyr':
##
##     select
```

```r
library(lmreg)
GSS2018 <- read_sav("GSS2018.sav")
```

```r
#Create codebook with variable descriptions
n <- ncol(GSS2018)
labels_vector <- map_chr(1:n, function(x) attr(GSS2018[[x]], "label"))
variable_name <- names(GSS2018)
code_book <- data.frame(variable_name, description = labels_vector)

#Filter for religious regex
relig_codebook <- code_book[
  grep("relig|strength of affiliation|god|jew|denomination|fundamentalist|bible|muslim|belief|church|pra
       labels_vector, ignore.case = TRUE),]

relig_df <- GSS2018 %>%
  dplyr::select(relig_codebook$variable_name)


#Add response to religious df
all_vars <- code_book[
  grep("relig|strength of affiliation|god|jew|denomination|fundamentalist|bible|muslim|belief|church|pra
       labels_vector, ignore.case = TRUE),]

all_df <- GSS2018 %>%
  dplyr::select(all_vars$variable_name)
```

```r
#Filtering predictors with less than 10% NA values
least_na <- relig_df %>%
  summarise(across(everything(), ~ sum(is.na(.)))) %>%
  pivot_longer(cols = c(1:82)) %>%
  arrange(value) %>%
  filter(value < nrow(relig_df) * 0.1)

the_df <- all_df %>%
  dplyr::select(least_na$name, CONSCI)

#Fitting first ordered logit model
first_cum_logit_model <- polr(as.factor(CONSCI) ~ ., data = the_df, Hess=TRUE)

summary(first_cum_logit_model)
```

```
## Call:
## polr(formula = as.factor(CONSCI) ~ ., data = the_df, Hess = TRUE)
```

```
##
## Coefficients:
##              Value Std. Error t value
## PRAY     -0.12809    0.04723 -2.7118
## ATTEND   -0.06074    0.03127 -1.9423
## RELIG16   0.01872    0.05430  0.3448
## RELPERSN  0.03917    0.08990  0.4357
## RELIG    -0.01660    0.05179 -0.3205
## RELACTIV -0.03539    0.03311 -1.0689
## GOD       0.06206    0.05809  1.0684
## BIBLE    -0.53648    0.10556 -5.0821
## RELITEN   0.07624    0.08218  0.9278
## FUND16   -0.22966    0.10763 -2.1338
## FUND     -0.08837    0.11759 -0.7516
## POSTLIFE -0.15110    0.16783 -0.9003
##
## Intercepts:
##     Value   Std. Error t value
## 1|2 -2.1251  0.5979    -3.5543
## 2|3  0.9337  0.5989     1.5590
##
## Residual Deviance: 2075.455
## AIC: 2103.455
## (1112 observations deleted due to missingness)
```

```
ctable_first <- coef(summary(first_cum_logit_model))

#Calculate and store p values
p_first <- pnorm(abs(ctable_first[, "t value"]), lower.tail = FALSE) * 2

#Combined table
ctable_first <- cbind(ctable_first, "p_value" = p_first)

full_table_first <- data.frame(ctable_first) %>%
  mutate(significance = case_when(
    p_value < 0.1 ~ "*",
    p_value < 0.05 ~ "**",
    p_value < 0.01 ~ "***",
    TRUE ~ ""
    ))

full_table_first
```

```
##                 Value Std..Error    t.value      p_value significance
## PRAY     -0.12808811 0.04723284 -2.7118443 6.691001e-03            *
## ATTEND   -0.06073943 0.03127268 -1.9422519 5.210661e-02            *
## RELIG16   0.01872119 0.05429870  0.3447815 7.302586e-01
## RELPERSN  0.03916526 0.08990017  0.4356528 6.630887e-01
## RELIG    -0.01659847 0.05179093 -0.3204900 7.485969e-01
## RELACTIV -0.03538994 0.03310827 -1.0689154 2.851078e-01
## GOD       0.06206247 0.05809114  1.0683638 2.853564e-01
## BIBLE    -0.53647870 0.10556272 -5.0820847 3.733148e-07            *
## RELITEN   0.07624481 0.08217915  0.9277878 3.535177e-01
## FUND16   -0.22966492 0.10763116 -2.1338144 3.285797e-02            *
```

```
## FUND     -0.08837424 0.11758677 -0.7515662 4.523120e-01
## POSTLIFE -0.15110273 0.16782752 -0.9003454 3.679365e-01
## 1|2      -2.12511194 0.59790685 -3.5542526 3.790550e-04            *
## 2|3       0.93368640 0.59890560  1.5589876 1.189993e-01
```

```r
#Making each religion into it's own predictor made up of binary variables
religions <- data.frame(binaries(as.factor(all_df$RELIG)))
colnames(religions) <- c("CHRISTIAN", "CATHOLIC", "NONE", "PROTESTANT",
                         "MOSLEM/ISLAM", "ORTHODOX-CHRISTIAN", "BUDDHISM",
                         "JEWISH", "OTHER", "HINDUISM", "NA_RELIGION",
                         "INTER-NONDENOMINATIONAL", "OTHER EASTERN",
                         "NATIVE AMERICAN")

#Removing religions that would make the model rank deficient due to NAs
new_df <- the_df %>%
  cbind(religions) %>%
  dplyr::select(-c("MOSLEM/ISLAM", "ORTHODOX-CHRISTIAN", "BUDDHISM",
                   "JEWISH", "OTHER", "HINDUISM", "NA_RELIGION",
                   "INTER-NONDENOMINATIONAL", "OTHER EASTERN",
                   "NATIVE AMERICAN"))

#New model with Christian, Catholic, None, and Protestant vars
new_model <- polr(as.factor(CONSCI) ~ ., data = new_df, Hess=TRUE)

summary(new_model)
```

```
## Call:
## polr(formula = as.factor(CONSCI) ~ ., data = new_df, Hess = TRUE)
##
## Coefficients:
##                Value Std. Error t value
## PRAY        -0.12289    0.04747 -2.5887
## ATTEND      -0.05912    0.03138 -1.8838
## RELIG16      0.02441    0.05567  0.4384
## RELPERSN     0.04057    0.09009  0.4504
## RELIG       -0.08137    0.19150 -0.4249
## RELACTIV    -0.04374    0.03362 -1.3012
## GOD          0.07183    0.05866  1.2245
## BIBLE       -0.53052    0.10564 -5.0218
## RELITEN      0.01750    0.12017  0.1456
## FUND16      -0.21740    0.10947 -1.9859
## FUND        -0.10948    0.12697 -0.8622
## POSTLIFE    -0.15591    0.16803 -0.9278
## CHRISTIAN    0.35581    1.53744  0.2314
## CATHOLIC    -0.42247    0.60280 -0.7008
## NONE         0.03410    0.54563  0.0625
## PROTESTANT -0.30514    0.73071 -0.4176
##
## Intercepts:
##     Value   Std. Error t value
## 1|2 -2.5857  1.0741    -2.4073
## 2|3  0.4795  1.0732     0.4468
##
## Residual Deviance: 2073.141
```

```
## AIC: 2109.141
## (1112 observations deleted due to missingness)
```

```
ctable <- coef(summary(new_model))

#Calculate and store p values
p <- pnorm(abs(ctable[, "t value"]), lower.tail = FALSE) * 2

#Combined table
ctable <- cbind(ctable, "p_value" = p)

full_table <- data.frame(ctable) %>%
  mutate(significance = case_when(
    p_value < 0.1 ~ "*",
    p_value < 0.05 ~ "**",
    p_value < 0.01 ~ "***",
    TRUE ~ ""
    ))

#Filtering for significant predictors
signif <- full_table %>%
  filter(significance != "") %>%
  rownames()

#Default method gives profiled CIs
ci <- confint(new_model)
```

```
## Waiting for profiling to be done...
```

```
#Odds ratios
ratios_table <- data.frame(exp(cbind(OR = coef(new_model), ci)))
ratios_table %>%
  filter(row.names(.) %in% signif)
```

```
##               OR    X2.5..   X97.5..
## PRAY   0.8843580 0.8056360 0.9705379
## ATTEND 0.9425972 0.8862224 1.0023069
## BIBLE  0.5883004 0.4775784 0.7227937
## FUND16 0.8046116 0.6489952 0.9971733
```

We can also get confidence intervals for the parameter estimates. These can be obtained either by profiling the likelihood function or by using the standard errors and assuming a normal distribution. Note that profiled CIs are not symmetric (although they are usually close to symmetric). If the 95% CI does not cross 0, the parameter estimate is statistically significant.

The CIs for PRAY, BIBLE, and FUND16 do not include 0; ATTEND does. The estimates in the output are given in units of ordered logits, or ordered log odds. The coefficients from the model can be somewhat difficult to interpret because they are scaled in terms of logs. Thus, we can interpret logistic regression models by converting the coefficients into odds ratios (OR). To get the OR and confidence intervals, we just exponentiate the estimates and confidence intervals.

```r
#Relevel CONSCI to make results interpretable
relevel_df <- new_df %>%
  mutate(CONSCI_re = case_when(
    CONSCI == 1 ~ 3, #Now 3 means A GREAT DEAL of confidence in the scientific community
    CONSCI == 2 ~ 2, #2 still means ONLY SOME confidence in the scientific community
    CONSCI == 3 ~ 1, #Now 1 means HARDLY ANY confidence in the scientific community
    TRUE ~ NA_real_
  )) %>%
  dplyr::select(-CONSCI)

#New model with releveled response
new_releveled_model <- polr(as.factor(CONSCI_re) ~ ., data = relevel_df, Hess=TRUE)

summary(new_releveled_model)
```

```
## Call:
## polr(formula = as.factor(CONSCI_re) ~ ., data = relevel_df, Hess = TRUE)
##
## Coefficients:
##                Value Std. Error  t value
## PRAY         0.12289    0.04747  2.58858
## ATTEND       0.05912    0.03138  1.88380
## RELIG16     -0.02441    0.05567 -0.43843
## RELPERSN    -0.04058    0.09009 -0.45041
## RELIG        0.08144    0.19151  0.42522
## RELACTIV     0.04374    0.03362  1.30102
## GOD         -0.07183    0.05866 -1.22457
## BIBLE        0.53052    0.10564  5.02179
## RELITEN     -0.01753    0.12017 -0.14590
## FUND16       0.21740    0.10947  1.98590
## FUND         0.10944    0.12697  0.86194
## POSTLIFE     0.15594    0.16803  0.92801
## CHRISTIAN   -0.35647    1.53753 -0.23185
## CATHOLIC     0.42247    0.60281  0.70084
## NONE        -0.03409    0.54564 -0.06248
## PROTESTANT   0.30519    0.73072  0.41766
##
## Intercepts:
##     Value   Std. Error t value
## 1|2 -0.4795  1.0733    -0.4468
## 2|3  2.5857  1.0742     2.4072
##
## Residual Deviance: 2073.141
## AIC: 2109.141
## (1112 observations deleted due to missingness)
```

```r
ctable_relevel <- coef(summary(new_releveled_model))

#Calculate and store p values
p_relevel <- pnorm(abs(ctable_relevel[, "t value"]), lower.tail = FALSE) * 2

#Combined table
ctable_relevel <- cbind(ctable_relevel, "p_value" = p_relevel)
```

```r
full_table_relevel <- data.frame(ctable_relevel) %>%
  mutate(significance = case_when(
    p_value < 0.1 ~ "*",
    p_value < 0.05 ~ "**",
    p_value < 0.01 ~ "***",
    TRUE ~ ""
    ))

#Filter only significant predictors
signif_relevel <- full_table_relevel %>%
  filter(significance != "") %>%
  rownames()

#Default method gives profiled CIs
ci_relevel <- confint(new_releveled_model)
```

```
## Waiting for profiling to be done...
```

```r
#Odds ratios
ratios_table_relevel <- data.frame(exp(cbind(OR = coef(new_releveled_model), ci_relevel)))
ratios_table_relevel %>%
  filter(row.names(.) %in% signif_relevel)
```

```
##               OR    X2.5..   X97.5..
## PRAY   1.130758 1.0303565 1.241256
## ATTEND 1.060899 0.9976986 1.128385
## BIBLE  1.699812 1.3835206 2.093897
## FUND16 1.242841 1.0028347 1.540843
```

Inference

- for every one unit decrease in frequency prayed, the odds of being more confident in science increases 13%
- for every one unit increase in attendance to religious services, the odds of being more confident in science increases 6%
- for every one unit decrease in level of devotion to the Bible, the odds of being more confident in science increases 70%
- for every one unit decrease in how fundamentalist the respondent was, the odds of being more confident in science increases 24%

Conclusion