Fake News Data - Fake News Paper - Demographics (Only used in the paper + Evangelicals)

Natália Tosi

08/18/2021

```
data_code <- read_csv("fake_news_data_code.csv") %>%
  select(-1)
```

Paper Research Questions (ORIGINAL)

- RQ1. How do demographics (age, gender, political orientation and income), affect the probability to share political fake news online?
- RQ2. What is the interaction between political orientation and gender?
- RQ3. How does the perception of frequency of political fake news online affect the probability to share them?
- RQ4. How does previous online fake news sharing (unnoticed), affect the probability to share political fake news online (on purpose)?
- RQ5. How does the perception of responsibility of 1) members of the public, 2) government, politicians and elected officials and 3) Facebook and Twitter, in trying to prevent fake news affect the probability to share them online?

Paper Research Questions (ADJUSTED)

- RQ1. How do demographics (age, gender, religion, education, political orientation and income), affect the probability to share political fake news online?
- RQ2. What is the interaction between political orientation and gender?
- RQ3. How does the perception of frequency of political fake news online affect the probability to share them?
- RQ4. How does previous online fake news sharing (unnoticed), affect the probability to share political fake news online (on purpose)?

- RQ5. How does the perception of responsibility of 1) members of the public, 2) government, politicians and elected officials and 3) Facebook and Twitter, in trying to prevent fake news affect the probability to share them online?

Model Variables

Dependent Variable = Shared Fake News (`shared_fake_news_19`)

P19 - Have you ever shared a political news story online that you thought at the time was made up? (Single Answer)

1 = Yes; 0 = No

Independent Variables

1) Gender
2) Age
3) Income (try class and income)
4) Political Orientation

"Data for the demographic variables (i.e., gender, age, income, education and political orientation) were collected using standard survey measurements.

The perception of frequency of political fake news was measured by asking participants the following question: “How often do you come across news stories online that you think are almost completely made up”, on a four-point Likert scale, ranging from 1) never, 2) hardly ever, 3) sometimes and, 4) often.

Unnoticed fake news sharing was measured by asking participants the following question: “Have you ever shared a political news story online that you later found out was made up? (0) No and (1) yes.

The perception of responsibility was measured by asking participants the following question: “As you may have heard, there have recently been some instances of so called “fake news stories” circulating widely online.

How much responsibility does each of the following have in trying to prevent made up stories from gaining attention” a) members of the public, b) the government, politicians, and elected officials, c) social networking sites like Facebook and Twitter, and search sites like Google, on a four-point Likert scale, ranging from: 1) no responsibility at all, 2) not much responsibility, 3) a fair amount of responsibility, 4) a great deal of responsibility."

Data

- P19 = shared_fake_news
- frequency_fake_news
- P21 = unnoticed_share
- public
- gov and politicians (create one single)
- social media

```
data_fake_news <- data_code %>%
  select(sex, age_full, age_60, pol_orientation, P19, frequency_fake_news, P21,
         resp_population, resp_gov, resp_politicians, resp_press,
         resp_social_media, severity_fake_news, shared_fake_news_19, education,
         income, class, religion, evaluation, approval, shared_fake_news_19,
         race_adj, sex_men, region_North, region_Northeast, `region_Center-West`,
         region_Southeast, region_South, religion_Catholic, religion_Evangelicals,
         `religion_Other religion`, `religion_No religion`, `age_60_16-24 age`,
         `age_60_25-34 age`, `age_60_35-44 age`, `age_60_45-59 age`,
          `age_60_60 or more`, evaluation_Unsure, P21_No, P21_Unsure, P21_Yes,
         capital_metrop, pol_orientation_right, pol_orientation_center,
         pol_orientation_left, pol_orientation_none, race_is_white,
         education_high, income_low, class_ab, class_c, class_de, has_religion) %>%
  rename(unnoticed_share = P21_Yes) %>%
  mutate(resp_gov_politicians = if_else(((resp_gov == 1) | (resp_politicians == 1)),
                                        1, 0),
         sex_women = if_else(sex_men == 0, 1, 0))
```

Regressions - All categories

Dependent variable: `shared_fake_news_19`

Independent variables:

1) Gender
2) Age
3) Income (class)
4) Education
5) Political Orientation

Regression 1 - Share fake news with demographics (sex, age, income, education, political orientation)

```r
reg_1 <- glm(shared_fake_news_19 ~ sex_men + age_full + class_ab +
               class_c + pol_orientation_right + pol_orientation_center +
               pol_orientation_left + education_high + religion_Evangelicals,
               family = binomial(link = 'logit'),
               data = data_fake_news)

#Left out: women, class DE, no political orientation
```

Regression 2 - How does the perception of frequency of political fake news online affect the probability to share them?

```r
reg_2 <- glm(shared_fake_news_19 ~ frequency_fake_news + sex_men + age_full + class_ab +
               class_c + pol_orientation_right + pol_orientation_center +
               pol_orientation_left + education_high + religion_Evangelicals,
               family = binomial(link = 'logit'),
               data = data_fake_news)
```

Regression 3 - How does previous online fake news sharing (unnoticed), affect the probability to share political fake news online (on purpose)?

```r
reg_3 <- glm(shared_fake_news_19 ~ unnoticed_share + frequency_fake_news +
               sex_men + age_full + class_ab + class_c + pol_orientation_right +
               pol_orientation_center + pol_orientation_left + education_high +
               religion_Evangelicals,
               family = binomial(link = 'logit'),
               data = data_fake_news)
```

Regression 4 - How does the perception of responsibility of 1) members of the public, 2) government, politicians and elected officials and 3) Facebook and Twitter, in trying to prevent fake news affect the probability to share them online?

```r
reg_4 <- glm(shared_fake_news_19 ~ resp_population + resp_gov_politicians +
               resp_social_media + unnoticed_share + frequency_fake_news +
               sex_men + age_full + class_ab + class_c + pol_orientation_right +
               pol_orientation_center + pol_orientation_left + education_high +
               religion_Evangelicals,
               family = binomial(link = 'logit'),
               data = data_fake_news)
```

Regression 5 - What is the interaction between political orientation and gender?

```
reg_5 <- glm(shared_fake_news_19 ~ resp_population + resp_gov_politicians +
               resp_social_media + unnoticed_share + frequency_fake_news +
               sex_men + age_full + class_ab + class_c + pol_orientation_right +
               pol_orientation_center + pol_orientation_left + education_high +
               sex_men*pol_orientation_right + sex_men*pol_orientation_center +
               sex_men*pol_orientation_left + religion_Evangelicals,
               family = binomial(link = 'logit'),
               data = data_fake_news)
```

```
library(DescTools)
library(mfx)

reg_1_or <- logitor(shared_fake_news_19 ~ sex_men + age_full + class_ab +
               class_c + pol_orientation_right + pol_orientation_center +
               pol_orientation_left + education_high + religion_Evangelicals,
               data = data_fake_news)


reg_2_or <- logitor(shared_fake_news_19 ~ frequency_fake_news + sex_men + age_full + class_ab +
               class_c + pol_orientation_right + pol_orientation_center +
               pol_orientation_left + education_high + religion_Evangelicals,
               data = data_fake_news)

reg_3_or <- logitor(shared_fake_news_19 ~ unnoticed_share + frequency_fake_news +
               sex_men + age_full + class_ab + class_c + pol_orientation_right +
               pol_orientation_center + pol_orientation_left + education_high +
               religion_Evangelicals,
               data = data_fake_news)

reg_4_or <- logitor(shared_fake_news_19 ~ resp_population + resp_gov_politicians +
               resp_social_media + unnoticed_share + frequency_fake_news +
               sex_men + age_full + class_ab + class_c + pol_orientation_right +
               pol_orientation_center + pol_orientation_left + education_high +
               religion_Evangelicals,
               data = data_fake_news)

reg_5_or <- logitor(shared_fake_news_19 ~ resp_population + resp_gov_politicians +
               resp_social_media + unnoticed_share + frequency_fake_news +
               sex_men + age_full + class_ab + class_c + pol_orientation_right +
               pol_orientation_center + pol_orientation_left + education_high +
               sex_men*pol_orientation_right + sex_men*pol_orientation_center +
               sex_men*pol_orientation_left + religion_Evangelicals,
               data = data_fake_news)
```

% Error: Unrecognized object type. % Error: Unrecognized object type. % Error: Unrecognized object type. % Error: Unrecognized object type. % Error: Unrecognized object type.

Table 1: Likelihood of Sharing Fake News - Logit Models (Estimates)

	Dependent variable:				
	Shared FN				
	Model 1	Model 2	Model 3	Model 4	Model 5
Resp. Pop				−0.159	−0.164
				(0.191)	(0.191)
Resp. Gov				0.233	0.233
				(0.220)	(0.220)
Resp. SM				0.064	0.060
				(0.183)	(0.183)
Unnoticed FN			2.080***	2.075***	2.077***
			(0.125)	(0.125)	(0.126)
FN Freq		0.603***	0.339**	0.330**	0.336**
		(0.145)	(0.157)	(0.158)	(0.158)
Gender(male)	−0.143	−0.135	−0.136	−0.141	0.134
	(0.116)	(0.117)	(0.128)	(0.128)	(0.198)
Age	0.005	0.006	0.005	0.005	0.005
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Class AB	0.231	0.183	0.031	0.025	0.017
	(0.177)	(0.178)	(0.196)	(0.196)	(0.197)
Class C	0.400**	0.359**	0.188	0.186	0.179
	(0.156)	(0.157)	(0.172)	(0.172)	(0.172)
Politics (R)	0.365**	0.342**	0.289*	0.287*	0.378
	(0.149)	(0.150)	(0.164)	(0.165)	(0.247)
Politics (C)	0.544***	0.503***	0.338	0.343	0.504
	(0.192)	(0.193)	(0.214)	(0.214)	(0.332)
Politics (L)	0.345**	0.311**	0.242	0.240	0.512**
	(0.145)	(0.145)	(0.160)	(0.160)	(0.202)
Educated	0.046	0.029	0.074	0.078	0.079
	(0.126)	(0.126)	(0.139)	(0.140)	(0.140)
Evangelical	0.276**	0.278**	0.339**	0.353**	0.336**
	(0.122)	(0.123)	(0.135)	(0.137)	(0.138)
Gender(male)*Pol(R)					−0.258
					(0.333)
Gender(male)*Pol(C)					−0.369
					(0.437)
Gender(male)*Pol(L)					−0.718**
					(0.330)
Constant	−2.060***	−2.499***	−2.901***	−3.016***	−3.108***
	(0.237)	(0.262)	(0.284)	(0.319)	(0.324)
Nagelkerke R^2	0.021	0.036	0.247	0.248	0.252
Cox and Snell R^2	0.013	0.023	0.159	0.16	0.162
Observations	1,934	1,934	1,934	1,934	1,934
Log Likelihood	−987.360	−977.967	−832.598	−831.698	−829.274
Akaike Inf. Crit.	1,994.720	1,977.934	1,689.197	1,693.396	1,694.548

Note: *p<0.1; **p<0.05; ***p<0.01

Variance Inflation Factor (VIF)

"For a given predictor (p), multicollinearity can assessed by computing a score called the variance inflation factor (or VIF), which measures how much the variance of a regression coefficient is inflated due to multicollinearity in the model.

The smallest possible value of VIF is one (absence of multicollinearity). As a rule of thumb, a VIF value that exceeds 5 or 10 indicates a problematic amount of collinearity (James et al. 2014)."

```r
model_1 <- as.data.frame(car::vif(reg_1)) %>%
  rownames_to_column("variables")

model_2 <- as.data.frame(car::vif(reg_2)) %>%
  rownames_to_column("variables")

model_3 <- as.data.frame(car::vif(reg_3)) %>%
  rownames_to_column("variables")

model_4 <- as.data.frame(car::vif(reg_4)) %>%
  rownames_to_column("variables")

model_5 <- as.data.frame(car::vif(reg_5)) %>%
  rownames_to_column("variables")

vif_test <- model_1 %>%
  left_join(model_2, by = c("variables" = "variables")) %>%
  left_join(model_3, by = c("variables" = "variables")) %>%
  left_join(model_4, by = c("variables" = "variables")) %>%
  left_join(model_5, by = c("variables" = "variables"))

names(vif_test) <- c("variables", "model_1", "model_2", "model_3",
                     "model_4", "model_5")

vif_test %>%
  kable(caption = "Variance Inflation Factor (VIF) per variable and model",
      align = "c")
```

Table 2: Variance Inflation Factor (VIF) per variable and model

variables	model_1	model_2	model_3	model_4	model_5
sex_men	1.068161	1.071654	1.069363	1.072113	2.542111
age_full	1.059130	1.060514	1.060088	1.064317	1.076270
class_ab	2.122597	2.132611	2.125674	2.132818	2.138996
class_c	1.933145	1.944887	1.924123	1.925463	1.931028
pol_orientation_right	1.280661	1.285474	1.268561	1.276673	2.855346
pol_orientation_center	1.170760	1.173993	1.164312	1.165669	2.793771
pol_orientation_left	1.229419	1.232917	1.228927	1.229279	1.932797
education_high	1.208696	1.214429	1.224700	1.226145	1.229491
religion_Evangelicals	1.054773	1.053607	1.056399	1.082931	1.087877