Fake News Data - Correlation

Natália Tosi

08/18/2021

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

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 (class)
4) Education
5) Political Orientation
6) Religion (Evangelicals)

Data

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

```r
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))
```

Correlation Matrix

```r
variables <- c("sex_men", "age_full", "class_ab", "class_c", "pol_orientation_right",
               "pol_orientation_center", "pol_orientation_left", "education_high",
               "religion_Evangelicals", "frequency_fake_news", "unnoticed_share",
               "resp_population", "resp_gov_politicians", "resp_social_media")



variables_rename <- c("Gender(male)", "Age", "Class AB", "Class C", "Politics (R)",
                      "Politics (C)", "Politics (L)", "Educated", "Evangelical",
                      "FN Freq", "Unnoticed FN", "Resp. Pop", "Resp. Gov",
                      "Resp. SM")
```

```r
correlation_matrix <- function(df,
                               type = "pearson",
                               digits = 3,
                               decimal.mark = ".",
                               use = "all",
                               show_significance = TRUE,
                               replace_diagonal = FALSE,
                               replacement = ""){

  # check arguments
  stopifnot({
    is.numeric(digits)
    digits >= 0
    use %in% c("all", "upper", "lower")
    is.logical(replace_diagonal)
    is.logical(show_significance)
```

```
    is.character(replacement)
  })
  # we need the Hmisc package for this
  require(Hmisc)

  # retain only numeric and boolean columns
  isNumericOrBoolean = vapply(df, function(x) is.numeric(x) | is.logical(x), logical(1))
  if (sum(!isNumericOrBoolean) > 0) {
    cat('Dropping non-numeric/-boolean column(s):', paste(names(isNumericOrBoolean)[!isNumericOrBoolean]
  }
  df = df[isNumericOrBoolean]

  # transform input data frame to matrix
  x <- as.matrix(df)

  # run correlation analysis using Hmisc package
  correlation_matrix <- Hmisc::rcorr(x, type = )
  R <- correlation_matrix$r # Matrix of correlation coeficients
  p <- correlation_matrix$P # Matrix of p-value

  # transform correlations to specific character format
  Rformatted = formatC(R, format = 'f', digits = digits, decimal.mark = decimal.mark)

  # if there are any negative numbers, we want to put a space before the positives to align all
  if (sum(R < 0) > 0) {
    Rformatted = ifelse(R > 0, paste0(' ', Rformatted), Rformatted)
  }

  # add significance levels if desired
  if (show_significance) {
    # define notions for significance levels; spacing is important.
    stars <- ifelse(is.na(p), "   ", ifelse(p < .001, "***", ifelse(p < .01, "** ", ifelse(p < .05, "*
    Rformatted = paste0(Rformatted, stars)
  }
  # build a new matrix that includes the formatted correlations and their significance stars
  Rnew <- matrix(Rformatted, ncol = ncol(x))
  rownames(Rnew) <- colnames(x)
  colnames(Rnew) <- paste(colnames(x), "", sep =" ")

  # replace undesired values
  if (use == 'upper') {
    Rnew[lower.tri(Rnew, diag = replace_diagonal)] <- replacement
  } else if (use == 'lower') {
    Rnew[upper.tri(Rnew, diag = replace_diagonal)] <- replacement
  } else if (replace_diagonal) {
    diag(Rnew) <- replacement
  }

  return(Rnew)
}
```

```
data_fake_news <- data_fake_news[, variables]
```

```r
colnames(data_fake_news) <- variables_rename

correlation <- correlation_matrix(data_fake_news)

correlation_1 <- correlation[,c(1:7)]

correlation_2 <- correlation[,c(8:14)]
```

Table 1: Correlations between independent variables

	Gender(male)	Age	Class AB	Class C	Politics (R)	Politics (C)	Politics (L)
Gender(male)	1.000	-0.002	0.030	-0.021	0.187***	0.116***	-0.044
Age	-0.002	1.000	0.062**	-0.049*	0.101***	0.021	-0.136***
Class AB	0.030	0.062**	1.000	-0.619***	0.025	0.034	0.008
Class C	-0.021	-0.049*	-0.619***	1.000	-0.034	-0.030	0.018
Politics (R)	0.187***	0.101***	0.025	-0.034	1.000	-0.172***	-0.287***
Politics (C)	0.116***	0.021	0.034	-0.030	-0.172***	1.000	-0.177***
Politics (L)	-0.044	-0.136***	0.008	0.018	-0.287***	-0.177***	1.000
Educated	0.005	-0.140***	0.283***	0.039	-0.029	-0.033	0.065**
Evangelical	-0.016	-0.004	-0.135***	-0.027	0.078***	-0.004	-0.050*
FN Freq	-0.013	-0.046*	0.043	0.036	0.009	0.034	0.054*
Unnoticed FN	0.002	0.014	0.015	0.050*	0.018	0.053*	0.026
Resp. Pop	0.036	-0.051*	0.083***	-0.008	-0.026	-0.011	0.051*
Resp. Gov	0.042	-0.038	0.077***	-0.003	0.002	-0.013	0.045*
Resp. SM	0.014	-0.049*	0.049*	0.002	-0.045*	-0.002	0.040

Table 2: Correlations between independent variables

	Educated	Evangelical	FN Freq	Unnoticed FN	Resp. Pop	Resp. Gov	Resp. SM
Gender(male)	0.005	-0.016	-0.013	0.002	0.036	0.042	0.014
Age	-0.140***	-0.004	-0.046*	0.014	-0.051*	-0.038	-0.049*
Class AB	0.283***	-0.135***	0.043	0.015	0.083***	0.077***	0.049*
Class C	0.039	-0.027	0.036	0.050*	-0.008	-0.003	0.002
Politics (R)	-0.029	0.078***	0.009	0.018	-0.026	0.002	-0.045*
Politics (C)	-0.033	-0.004	0.034	0.053*	-0.011	-0.013	-0.002
Politics (L)	0.065**	-0.050*	0.054*	0.026	0.051*	0.045*	0.040
Educated	1.000	-0.139***	0.064**	0.010	0.061**	0.049*	0.027
Evangelical	-0.139***	1.000	-0.017	-0.010	-0.151***	-0.151***	-0.083***
FN Freq	0.064**	-0.017	1.000	0.133***	0.087***	0.076***	0.096***
Unnoticed FN	0.010	-0.010	0.133***	1.000	0.022	0.050*	0.030
Resp. Pop	0.061**	-0.151***	0.087***	0.022	1.000	0.624***	0.533***
Resp. Gov	0.049*	-0.151***	0.076***	0.050*	0.624***	1.000	0.587***
Resp. SM	0.027	-0.083***	0.096***	0.030	0.533***	0.587***	1.000

```
kable(correlation_1, caption = "Correlations between independent variables",
      booktabs = T, linesep = "") %>%
  kable_styling(latex_options = "striped", font_size = 8) %>%
  row_spec(0, bold = TRUE)
```

```
kable(correlation_2, caption = "Correlations between independent variables",
      booktabs = T, linesep = "") %>%
  kable_styling(latex_options = "striped", font_size = 8) %>%
  row_spec(0, bold = TRUE)
```