

Athanasia Meimeteas
Harrison High School
Behavioral Sciences
2019-2020

Project Title:

Predicting the Relationship between Marriage and Discussion of Women's Issues in
Congressional Representatives: A Multiplicative Interaction Model Approach

Rationale:

Recent research exploring women's views on gender equality has focused on race, education, socioeconomic status, age, and political party. Little research has been done, however, on marriage, and how it affects views on gender inequality. Of the current research on marriage, the impact marriage has on an individual's or a family's financial stability has been explored. The U.S. economy encourages marriage by offering certain tax benefits and program entitlements for married couples (Murphy, 2019). The Urban Institute report (2019) found that among other things, marital status altered eligibility for entitlement programs, tax rates, and the availability of social safety nets.

Marriage rates, however, are predicted to drop below 70 percent in the near future. Baby boomers have a marriage rate of 91 percent, late boomers have a marriage rate of 87 percent marriage, and Gen Xers have a marriage rate of 82 percent (Murphy, 2019). According to Pew Research (2019), up to 25 percent of millennials may never get married. Therefore, the growing rate of unmarried Americans may have ramifications the U.S. economy is unprepared to handle (Murphy, 2019). This research, however, did not address any social factors influenced by marital status.

One analysis of marriage's effect on social factors determined that gender linked fate in females — one's identification with a group and the perception that one's life chances are tied to

the success of that group (Dawson, 1994) — weakens after marriage. This may be a result of women institutionalizing their relationship with men and feeling less connected to other women (Stout, Kretschmer, & Ruppaner, 2017). Marriage can lead to a female's investment in family linked fate rather than gender linked fate as they begin to have children and run a household. The findings of Stout's study suggests there are political ramifications of marriage for which we may be unaware.

To predict the relationship between marriage and a social issue, this study will create a mathematical model. In this study, analysis will be performed to determine whether married or unmarried Congressional representatives were more likely to discuss women's issues — as defined by discussion of sexual harassment, sexual assault, pay inequality, the recognition of women, human trafficking, women's health issues, gender disparities in profession, and international women's issues (Stout, Kretschmer, Bartula-Henkle, McLean, 2019). A mathematical model was created to predict the likelihood of representatives to discuss women's issues based on their marital status.

The model will be created based on data retrieved from members of the 115th Congress, which occurred during the unraveling of Harvey Weinstein thus heightening sensitivity and awareness to the MeToo movement across the board. A data set of over 400 U.S. Congress representatives' Tweets was used in this study. Twitter will be selected as the data source as it provides a frequent source of unfiltered representative communication. Other forms of outreach such as press releases and speeches are both rehearsed and less frequent. A sample of Congressional representatives will be selected as it provides a targeted group of individuals who are intended to reflect the views of the public.

Question:

To what extent can a mathematical model predict the relationship between current marital status and congressional representatives' likelihood to discuss women's issues?

Hypotheses:

The relationship between gender and frequency to discuss women's issues will be contingent upon current marital status. If congressmen are married, they will be more likely to discuss women's issues. If congresswomen are married, they will be less likely to discuss women's issues.

Culling the Data:

A web scraper will be used to collect tweets from as many members of Congress as possible from the 115th Congress. The use of data from just one Congress will ensure consistency as few representatives quit or were removed from office during this time.

Both hand coding and computer assisted content coding analysis will be used to assign each tweet with a score of 1 if it mentions gender issues or 0 if it did not, creating a dummy variable. Computer assisted content coding analysis will be used through the program Rtexttools. A sample of hand coded training set documents will be used for the computer assisted content coding to code the remaining tweets with six different algorithms. The documents will then be coded based on what the majority of the algorithms agreed was the predicted code and provided information about where disagreement occurred among the algorithms.

Choosing the Variables:

Overfitting will be considered. This will be accounted for by considering the ratio of observations and independent variables so there are not too many terms or the number of observations present.

Multicollinearity will also be considered. It is unlikely in the given variables as they are unrelated demographics; the independent variables are correlated with the dependent variable, but not with other independent variables.

The dependent variable will be “overall female appeal”, or the frequency to tweet about women’s issues. The independent variable for the regression will be marriage. In these models, gender was hand sorted in order to assess male and female representatives separately. In the multiplicative interaction, the dependent variable will remain overall female appeal. The independent variable will be gender and the conditional variable was marriage. The control variables that will be selected in order to isolate the independent variable are:

1. If the representative is white
2. The representative’s age
3. If the representative is a democrat
4. The partisan voting index of the representative’s district
5. The percent of the representative’s district that was female
6. The total tweets of the representative

Mathematical Models:

A linear regression and multiplicative interaction model will be used to model the relationship between marriage and overall female appeal. The linear regression will reflect a simplistic x-y relationship in which men and women are grouped together. In order to combat this, the OLS regression will be run again once with just men and once with just women.

In order to reflect the conditional hypothesis, an interaction is the ideal statistical analysis test. It will be possible to make predictions based on the multiplicative interaction model. In creating all of the models, assumptions and justifications will be made in order to validate the models:

1. Assumption: All representatives use Twitter with equal purpose.

Justification: Representatives will all tweet about their social and political views.

2. Assumption: All marriages can be considered as equal.

Justification: There is no way to differentiate the dynamics within each marriage studied. Thus, the assumption must be made that they are all the same.

3. Assumption: Regardless of gender roles in a household, the likelihood of representatives to speak out is the same.

Justification: The roles of each partner in a household cannot be assessed differently for each representative evaluated.

4. Assumption: Why or why not representatives are married will not affect their likelihood to speak out.

Justification: The reason that each representative is or is not married is not possible to determine. The assumption must be made that the reason behind the choice to marry or not to marry will not have any bearing on their likelihood to speak out.

5. Assumption: Representatives' activity on Twitter can be generalized to their discussion in real life.

Justification: There is no way to differentiate between the way representatives tweet and their day to day conversations in real life.

6. Assumption: All representatives are binary and gender is listed correctly.

Justification: Every representative collected in the dataset recognizes as either male (0) or female (1).

7. Assumption: The use of Twitter is representative of discussion in real life.

Justification: There has been no research to support that twitter activity is not representative of the way representatives discuss similar topics in real life.

Testing the Model:

The actual value for overall female appeal will be pulled from the dataset and compared with the predicted value for overall female appeal obtained through the interaction model. An unpaired t-test with equal variances will be run. In this test, no significance will signify that the results are similar to the actual value and predicted value. Subsequently, a scatterplot will be created in order to visually represent how closely related the two variables are.

The score of an individual representative in the 116th Congress was also calculated and compared with their coded tweets in order to determine the model's accuracy.

References:

- Brambor, T., Clark, W., & Golder, M. (2006). Understanding interaction models: improving empirical analyses. *Political Analysis*, 14(1), 63-82. doi:10.1093/pan/mpi014
- Bolzendahl, C. I., & Meyers, D. J. (2004). Feminist attitudes and support for gender equality: Opinion change in women and men, 1974-1998. *Social Forces*, 83(2), 759-789. Retrieved from JSTOR database.
- Clair, Robin Patric, et al. "# MeToo, sexual harassment: an article, a forum, and a dream for the future." *Journal of Applied Communication Research* (2019): 1-19.
- Collingwood, Loren, and John Wilkerson. "Tradeoffs in accuracy and efficiency in supervised learning methods." *Journal of Information Technology & Politics* 9.3 (2012): 298-318.
- Dawson, Michael. 1994. *Behind the Mule: Race and Class in African-American Politics*. Princeton: Princeton University Press.
- Druckman, James N., Martin J. Kifer, and Michael Parkin. "Campaign communications in US congressional elections." *American Political Science Review* 103.3 (2009): 343-366.
- Jurka, Timothy P., et al. "RTextTools: Automatic text classification via supervised learning." *R package version 1.9* (2012).
- Kearney MW (2019). "rtweet: Collecting and analyzing Twitter data." *Journal of Open Source Software*, 4(42), 1829. doi: [10.21105/joss.01829](https://doi.org/10.21105/joss.01829), R package version 0.7.0, <https://joss.theoj.org/papers/10.21105/joss.01829>.
- Murphy, Meg. "NowUKnow: Why Millennials Refuse to Get Married." *PreparedU View | Bentley University*, 3 Jan. 2019, www.bentley.edu/impact/articles/nowuknow-why-millennials-refuse-get-married.

Pietryka, Matthew T. "The roles of district and national opinion in 2010 congressional campaign agendas." *American Politics Research* 40.5 (2012): 805-843.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
URL <http://www.R-project.org/>.

Schneider, A., Hommel, G., & Blettner, M. (2010). Linear regression analysis: part 14 of a series on evaluation of scientific publications. *Deutsches Arzteblatt international*, 107(44), 776–782.
doi:10.3238/arztebl.2010.0776

Sides, John. "The origins of campaign agendas." *British Journal of Political Science* 36.3 (2006): 407-436.

Stout, C.T. (2019, November 5). [Personal interview].

Stout, C.T., Kretschmer, K.N., & Ruppanner, L. (2017). *Gender linked fate, race/ethnicity, and the marriage gap in American politics*.

Stout, C.T., Kretschmer, K., Burtula-Henkle, M., & McLean, L. (n.d.). *The Weinstein effect: The role of MeToo movement in altering the communication strategies of members of Congress*.