FINAL REPORT

Would Gender and Race Influence the CV Responses Rate?

Veronika Nuretdinova, Yubo Zhang

Abstract

We plan to test whether employers in Data Science industry treat employers unequally by race and gender. We created fictitious CV which would fit a wide range of data scientist position. Then, sent out this CV under 4 names: White male, White female, African-American male, African-American female to 40 job positions, respectively. Then, we measured the response rate within a month to test our hypothesis that women's application for the data scientist position would receive fewer responses than men's. Also CV with African-American names (female or male) sounding names would get lower response rates than White sounding names (female or male). we have observed the difference in response rate between men and women: our results indicated that men, regardless of race, received higher response rate than women. However, despite almost twice difference in rates, considering our sample size, it was not statistically significant at 5% level, our p-value is 7.5%. We have not observed a statistically significant difference in response rate between races.

Introduction

In 2016, the World Economic Forum released the Gender Gap Report and the report indicates that in the past 10 years, the global gender gap across education and economic opportunity and politics has closed by 4%, while the economic gap has closed by 3%. United States was ranked 45 among all the 144 countries [1]. Based on their ranking on United States, on average, women are more likely to work part-time, be employed in low-paid jobs and not take on management positions [2]. Another survey by Pew Research Center in 2017 also showed that women in the U.S, especially younger women (under fifty years) are substantially more likely than men to say gender discrimination is a major problem in the technology industry [3].There is evidence that gender inequalities in the workplace stem, at least in part, from the discrimination directed against women.

Several studies have documented personal discrimination against women in the tech industry. In May 2014, Google posted on its official blog that only 30 percent of its employees globally were women. In 2017, New York Times reported that many of the largest technology companies hire only 30 % female employee [4]. As for the technical and research positions, the findings are consistent and even worse than the general trend. Many media including Forbes, Time, New York Times reported that 11% of Silicon Valley executives and about 20% of software developers are women. At Google, only about 18% of technical employees are women. On Forbes' 2015 Top Tech Investors list, of 100 investors, only 5% are women. Women in technology earn less than men, with men earning up to 61% more than women. It is said in the valley that "Bias against women in tech is pervasive" [5,6,7].

Besides gender bias that exist in tech industries, based on the Pew Research Center, survey results also asked about perceptions of discrimination against two other groups underrepresented in the tech industry: African-American and Hispanics. A majority of African-American (64%) say discrimination against African-American and Hispanics is a major problem in the tech industry, and half of Hispanics agree, compare to 21% of White Americans say this is a major problem [3].

In spite of consistent evidence that higher sexism is related to greater bias toward women who want to climb the Corporate America ladder, little is known regarding the underlying processes linking sexism to discrimination. This question remains an important one, especially because the persistence of gender discrimination contradicts the anti-discrimination rules required by laws. In this study, we would like to investigate if gender and race play any important role in Data Science related jobs.

Experimental Design

We have used a famous paper in racial inequality in recruitment "Are Emily and Greg More Employable Than Lakisha and Jamal?" by Bertrand and Mullainathan (2004) as a reference for our experiment design. Our hypothesis is that female applicants for the data scientist position would receive fewer responses than male applicants. African-American sounding applicants would receive fewer responses than White-sounding applicants.

Compare to the paper by Bertrand and Mullainathan, our experiment was simpler in terms of number of treatments and number of observations.

1) Creating fictitious CV: we have studied public Linkedin profiles of the real data scientists. We also used standard job requirement as reference for our skill, experience, education and so on, this includes: Master Degree in Mathematics, Master Degree in the Data Science, 6 years of experience in statistical modelling, data science, working with distributed datasets and so on. Also, out candidate worked in retail and media industries, which cover a wide range of employers. We have also created one Cover Letter with general text to fit into the cover letter requirement. The CV and the Cover Letter are attached in the Appendix 1.

2) Choosing candidates identities and organizing the accounts: for our purpose, we needed to choose a very commonly used White male and female names, and Black female and male names. Similar with the Bertrand and Mullainathan paper, we have reviewed top ranked popular White and Black names for men and women, as well as top ranked popular surnames among different races based on Social Security Administration (https://www.ssa.gov/oact/babynames/decades/century.html). We came up with: Thomas Brown, Sarah Brown, Rasheed Jackson, Jazmine Jackson.
Then, we have also created 4 Gmail accounts, phone numbers and candidates accounts on Indeed.com (www.indeed.com). Each of us managed two accounts: Veronika - two male accounts (Thomas Brown and Rasheed Jackson), Yubo - female accounts (Sarah Brown and Jazmine Jackson).

3) Selecting and applying jobs : we were seeking for Data Scientist job applications at indeed.com and other popular job-seeking websites. No geographic factor was considered during job selection. Since our candidate had 6 year of experience, we chose job offers with at least 3 year of experience requirements. During the job application reviewing process, we have noticed that there's some threshold at 5-year. For example, the majority of employees are required to have at least either 3 years or 5 years of previous working experience . Concerned that this might influence the response rate, we have introduced blocking, sending out 20 CVs to job posts which required only 3 years of professional experience, and 20 CVs to jobs with 5 years plus of experience requirement. We would like to test if different level of work experience would influence response rate.

   To make our companies and job offers homogenous, we avoided:
   - Jobs that require certain specific expertise (e.g. deep experience in NLP).
   - Job ads from top-notch companies like Google, Apple, Amazon where we expected the competition would be much higher than at other companies.
   - Job application that requires LinkedIn page.

- Job application that requires quantitative test during initial application.

During the applying process, each candidate was selecting the corresponding gender and racial background questions toward the end of the online application.

In the first week we sent out the CVs to 40 jobs ads for each candidate, 160 in total. We were keeping track of the companies and job offers in one spreadsheet to make sure that we're sending out four CVs of to similar jobs across different companies.

Spreadsheet with job applications:
https://docs.google.com/spreadsheets/d/1UWjKq4ZGM4exuzWRdkIPI3_WdTMsw8QwpYmoAW4ZBlU

4) Measuring the response rate: we measured the response rate through emails and phone communications. Employers sent either invitations for the interview, called for a interview schedule, or a negative response that the candidate doesn't go further in the recruitment process.

Experimental Design Drawbacks

1. ***Low statistical power***

Our major concern from the beginning was a small number of CVs we could send throughout this experiment.

If we consider, only one parameter at a time (gender or race), we have 80 observations per parameter. Calculation of statistical power even for one sided test, on the assumption that women get less responses, at the assumed proportions of 10% and 20% can provide only 55% statistical power. And 42% for two-sided test.
So, even at twice the difference in proportions, the chance of detection at our number of observations is 55%.

```
 power.prop.test(n = 80, p1 = 0.1, p2 = 0.2, sig.level = 0.05, power = NULL,
                 alternative = c("one.sided"),   strict = FALSE)
##
##        Two-sample comparison of proportions power calculation
```

```
          n = 80
         p1 = 0.1
         p2 = 0.2
  sig.level = 0.05
      power = 0.4244147
alternative = two.sided
NOTE: n is number in *each* group
```

To obtain the satisfactory 0.8 statistical power, we would have to send out 150 CVs per group candidate, or 300 in total. And for two-sided test we would need 200 CV per applicant, or 400 in total.

Besides our own limitations within this experiment, which would make it difficult for us to send out 800 CVs within 1-2 week period, there is also the ethical concern. Since the employers HR would have to spend time on fake CVs, send out the responses, and in case of positive responses, also send out the reminders. We decided not to take this approach and accept 55% statistical power for this experiment.

```
power.prop.test(p1=.1,p2=.2,power=.8, alternative = c("one.sided"))

Two-sample comparison of proportions power calculation

          n = 156.6054
         p1 = 0.1
         p2 = 0.2
  sig.level = 0.05
      power = 0.8
alternative = one.sided
```

```
power.prop.test(p1=.1,p2=.2,power=.8, alternative = c("two.sided"))

Two-sample comparison of proportions power calculation

          n = 198.9634
         p1 = 0.1
         p2 = 0.2
  sig.level = 0.05
      power = 0.8
alternative = two.sided
```

2. Unknown non-compliance and attrition rate

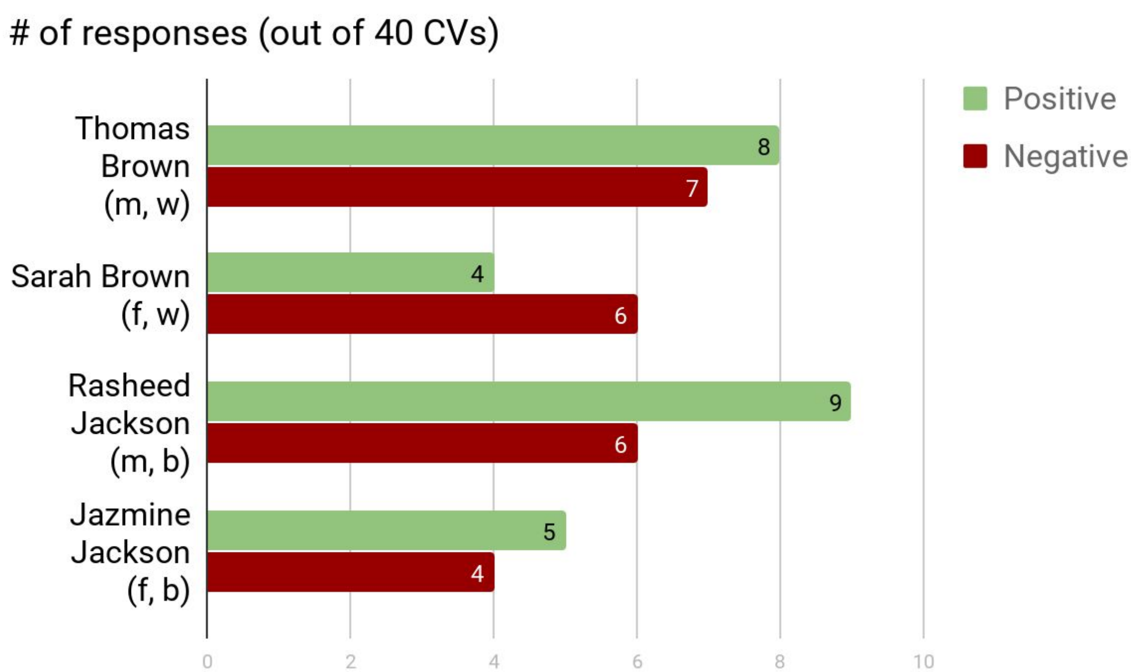
In this experiment the only type of information we could received, is whether the candidate got the response, or not.

However, we found no way to learn about:
- Non-compliance, i.e. that the employer did not review the CV at all
- Attrition, i.e. that the vacancy published is no longer available

For this reason, we've made a simple assumption, that no response = negative response.

Result

In the below chart there are numbers of negative and positive response we obtained within 4 weeks after sending out the applications. Below is the analysis of the results:



Comparison between male and female applicants

- Female positive responses: 9 out of 80
- Male positive responses: 17 out of 80

Despite almost twice difference in the response rate, our result is not statistically significant at 5% level. Our p-value in regression is 7.5%

```
Call:
lm(formula = invited ~ female, data = cv)

Residuals:
    Min      1Q  Median      3Q     Max 
-0.2152 -0.2152 -0.1111 -0.1111  0.8889 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.21519    0.04135   5.204 5.98e-07 ***
female      -0.10408    0.05812  -1.791   0.0752 .  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3675 on 158 degrees of freedom
Multiple R-squared:  0.0199,	Adjusted R-squared:  0.01369 
F-statistic: 3.207 on 1 and 158 DF,  p-value: 0.07522
```

Experimenting with the sample size

It’s interesting though, that if we had same proportions, and our sample size were 2 times smaller, the results would be statistically significant at with p-value of 1.15%.

```
Call:
lm(formula = invited ~ female, data = cv_double)

Residuals:
    Min      1Q  Median      3Q     Max 
-0.2152 -0.2152 -0.1111 -0.1111  0.8889 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.21519    0.02915   7.383 1.37e-12 ***
female      -0.10408    0.04096  -2.541   0.0115 *  
```

Comparison between Black and White applicants

We found almost no difference in the number of responses between black and white candidates. Black candidates received 13 positive responses out of 80 applications, and 12 positive responses for white candidates.

The result is not statistically significant with p-value 72%

```
Call:
lm(formula = invited ~ black, data = cv)

Residuals:
    Min      1Q  Median      3Q     Max 
-0.1728 -0.1728 -0.1519 -0.1519  0.8481 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.15190    0.04175   3.638 0.000371 ***
black        0.02094    0.05868   0.357 0.721663
```

Comparison taking into account two factors: gender and race

If we take two factors in account through interactive variable, none of the coefficients, including gender, race and gender*race is statistically significant, with very high p-values.

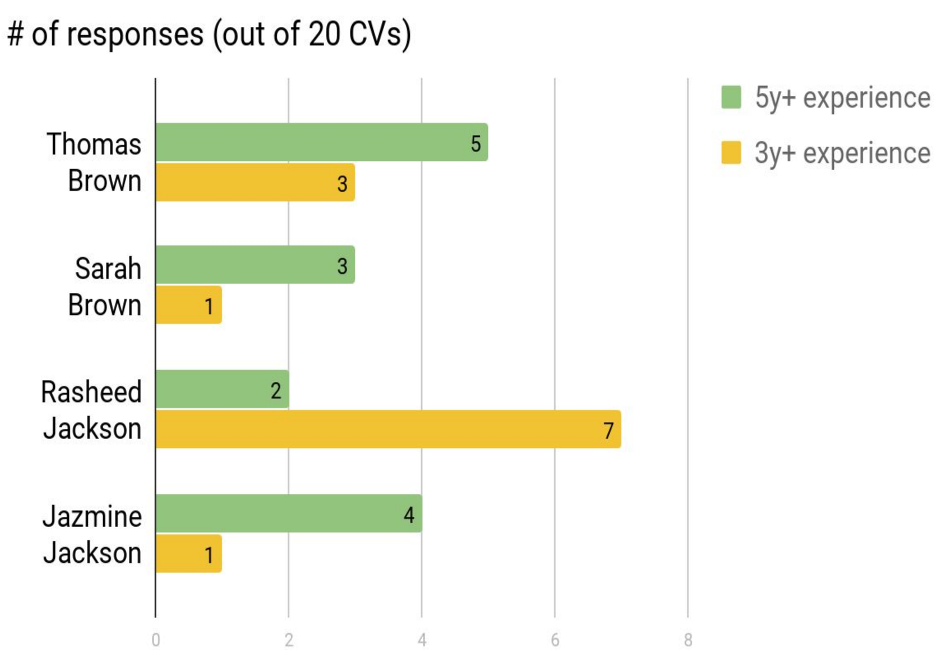
```
Call:
lm(formula = invited ~ black * female, data = cv)

Residuals:
    Min      1Q  Median      3Q     Max 
-0.2250 -0.2051 -0.1220 -0.1000  0.9000 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)   0.205128   0.059203   3.465 0.000685 ***
black         0.019872   0.083200   0.239 0.811541    
female       -0.105128   0.083200  -1.264 0.208275    
black:female  0.002079   0.116934   0.018 0.985835
```

Comparison on years of experience requested

We were also interested whether the response rate for different candidates would be different based on the job demands (years of experience). Our hypothesis was that women would also get less responses for jobs with 5+ year experience requested.

We were surprised to find the opposite of our hypothesis. In fact, all the difference in the positive response rates was created by better response rates to male candidates in 3+ years positions. Positive response rate for 5+ positions was quite similar for two genders at nearly 17%.

While for 3+ positions, men had 25% response rate, and women 5%.



The difference in 3+ year positions response rate in favor of male candidates is statistically significant with p-value of 1.13%

```
Call:
lm(formula = invited ~ five_year * female, data = cv)

Residuals:
    Min      1Q  Median      3Q     Max 
-0.2564 -0.1750 -0.1707 -0.0500  0.9500 

Coefficients:
               Estimate Std. Error t value Pr(>|t|)    
(Intercept)     0.25641    0.05863   4.373 2.23e-05 ***
five_year      -0.08141    0.08240  -0.988   0.3247    
female         -0.20641    0.08240  -2.505   0.0133 *  
five_year:female  0.20214    0.11581   1.745   0.0829 .  
```

Conclusion

- We have observed almost twice difference in the response rate between male and female candidates for general Data Science positions. However, the difference is not statistically significant with the p-value of 7.5%. We suspect that this is due to not enough sample size statistical power (only 55%).
- We have observed almost no difference in response rate between black and white races. The p-value is 72%
- We found statistically significant difference between men and women in response rate for lower level positions, those with 3+ years of experience required. Female applicants had 5% response rate, while men had 25% response rate. The p-value is 1.13% No difference observed for higher-level Data Scientist positions.

If we run a detailed professional experiment, these are the things we would do differently:

- We would increase the sample size to 200 samples per category, instead of 80, we would have 400 samples in total in order to achieve satisfactory 0.8 level of statistical power
- We would try several different names and CV types for each gender, similar with the Bertrand & Mullainathan 2004 study. This step will be included to exclude the

possibility that something particular about our CV that could create difference in response rates between men and women

- We would reproduce the experiment with higher vs lower experience demand jobs at larger sample, to investigate further whether women face more discrimination at more junior positions, and whether black men are favoured at junior positions

References

1. The Global Gender Gap Report: 2016. Geneva: World Economic Forum; 2016. (retrieved from http://reports.weforum.org/global-gender-gap-report-2016/economies/?doing_wp_cron=1533371121.3302431106567382812500#economy=ALB)

2. Arulampalam, W., Booth, A. L., & Bryan, M. L. (2007). Is There a Glass Ceiling over Europe? Exploring the Gender Pay Gap across the Wage Distribution. ILR Review, 60(2), 163-186. doi:10.1177/001979390706000201

3.Parker, K., & Funk, C. (2017, October 10). Women are more concerned than men about gender discrimination in tech industry. Retrieved from http://www.pewresearch.org/fact-tank/2017/10/10/women-are-more-concerned-than-men-about-gender-discrimination-in-tech-industry/

4. Intel Allocates $300 Million for Workplace Diversity. (2017, December 21). Retrieved from https://www.nytimes.com/2015/01/07/technology/intel-budgets-300-million-for-diversity.html?_r=1

5.Kokalitcheva, K. (2015, June 01). Google's Workplace Diversity Still has a Long Way to Go. Retrieved from http://time.com/3904408/google-workplace-diversity/

6. The New York Times Company. (n.d.). Retrieved from https://www.nytimes.com/roomfordebate/2014/10/29/reversing-gender-bias-in-the-tech-industry

7. Ong, T. (2017, August 09). In response to anti-diversity memo, YouTube CEO says sexism in tech is 'pervasive'. Retrieved from https://www.theverge.com/2017/8/9/16118018/youtube-ceo-women-tech-google-diversity-susan-wojcicki

8. Bertrand, M., & Mullainathan, S. (2003). Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination.

Spreadsheet with job applications tracking:
https://docs.google.com/spreadsheets/d/1UWjKq4ZGM4exuzWRdkIPI3_WdTMsw8QwpYmoAW4ZBlU

Appendix: Candidate CV

Jazmine M. Jackson

jazmine900jackson@gmail.com
+1(706)247-1619
Virginia, USA(ready to relocate)

Profile

- Data scientist and quantitative researcher with 6 year experience, applying predictive analytics to address complex business issues, including new product development, customers retention, market and promotion strategy

- Extensive expertise in Econometric, Statistics, Exploratory data analysis, Predictive modelling, Machine Learning, Natural Language Processing

Education

2015-2017: University of California, Berkeley
Master's degree in Information and Data Sciences

2007 - 2012: Texas A&M University
Bachelor's degree in Mathematical Science

Professional Experience

2015-Current: Associate Data Scientist at TEGNA (formerly Gannett), VA

- Created a product recommender using Bayesian Profile Regression
- Used predictive analytics to forecast churn, Applied natural language processing techniques to match disparate records
- Presented these analysis in visual, interactive format for a broad audience

2014-2015: Data Scientist at LogicNow, Cambridge, MA

- Mined large back-end SQL databases to analyze customer behavior and advise upper management on data-derived business process improvements and cost-saving adjustments, optimized and scale large amounts of data processing using Hadoop
- Develop predictive models to detect and prevent customer churn and determine the lifetime value of customers

2012-2014: Data Analyst at RetailMeNot, Inc. Austin, Texas Area

- Developed predictive models and extract insights from statistical analysis of complex problems.
- Designed a predictive model which determines the quality of a coupon and the propensity of it being clicked. Model being widely used across the company as RMN Index and for merchant coaching.
- Part of the team that designed the first machine learning algorithm for RMN.
- Developed interactive dashboards, reports and templates for ad-hoc analysis and reporting

Skills

- Programming languages: R, Python, C ,SQL , BASH shell scripting , Javascript, HTML & CSS
- Data Visualization: Tableau, Python (Bokeh, Matplotlib, Networkx) , Javascript (D3.js, DC.js, crossfilter.js)
- Distributed computing and datasets: Hadoop & HDFS , NoSQL, Amazon AWS, Google Cloud
- Deep Learning: Tensorflow, Keras, Theano, Lasagne

Certifications/Licenses

- Johns Hopkins University (9 courses+Capstone): Data Science Coursera Certification (Practical Machine Learning, Regression Models, Statistical Inference, Exploratory Data Analysis)

- DeepLearning.ai : Deep Learning Specialization (5 Courses)