

# Race and Gender on Resume Callback

Jiayi Zhou

2023-10-18

## Overview

This dataset contains experimental data from a study aimed at understanding the factors that influence resume selection for a callback. The experimental data is derived from a study that sought to investigate the impact of race and gender on job application callback rates. The study monitored job postings in Boston and Chicago for several months during 2001 and 2002 and used this data to construct a set of test cases. The research question is: How do race and gender influence job application callback rates? This is an inference problem. The dataset comprises 4,870 observations and 30 variables, including gender, race, callback receipt, job type, and job industry.

## Data Cleaning

Two variables, the indicator for whether the employer is a federal contractor and the type of company, contain missing values that cannot be easily computed or retrieved. Due to a lack of information, no observations are removed, and no values are computed to fill in.

Factorization has been applied to the following variables: the indicator for whether the employer is an Equal Opportunity Employer, the indicator for whether communication skills are required, the indicator for whether some level of education is required, the indicator for whether computer skills are required, the indicator for whether organizational skills are required, the indicator for whether there was a callback from the job posting for the person listed on this resume, the indicator for whether the resume listed a college degree, the indicator for whether the resume listed that the candidate has been awarded some honors, the indicator for whether the resume listed working while in school, the indicator for whether computer skills were listed on the resume, the indicator for whether any special skills were listed on the resume, the indicator for whether volunteering was listed on the resume, the indicator for whether military experience was listed on the resume, the indicator for whether there were gaps in the person's employment history, and the indicator for whether the resume lists an email address, based on the variable explanations on the website. No other assumption is made during the data cleaning process, and no observations are removed.

## Modeling

### Logistic Regression

Based on the question, the outcome variable is whether applicants receive a callback. Logistic regression is used because the outcome variable has two categories: receiving a callback (coded as 1) or not receiving a callback (coded as 0).

### Predictor Variables Selection

We selected predictor variables based on whether they act as confounding variables that are associated with both race, gender, and the outcome variable. Based on that consideration, the predictor variables included in the models are as follows: city where the job was located, industry of the job, type of role, inferred race associated with the first name on the resume, inferred gender associated with the first name on the resume, resume quality, years of experience listed on the resume, and all the variables that we factorized in the above section.

Total 21 variables are included, with 20 categorical variables and 1 continuous variable.

The variables that were not included are as follows with reasons:

1. Unique ID associated with the advertisement, due to the large number of categories.
2. Indicator for whether the employer is a federal contractor, due to missing values.
3. Type of company, as it contains unknown values that cannot be easily imputed.
4. Indicator for whether any job requirements are listed, as it is covered by other skills such as communication and computer skills.
5. Amount of experience required, as it has values that cannot be easily converted, such as 'some' and empty values.
6. Level of education required, as some values are not listed.
7. The first name used on the resume, due to the multitude of categories.
8. Years of college education listed on the resume, as this information is covered by the presence of a college degree.

### Model and Summary Output Table

In our model, we initially assessed Cook's distance for influential points, leading us to exclude three observations due to their influence on the results. This step became necessary because the significance level of the job industry in transportation and communication shifted from being significant to becoming insignificant. We then refitted the model and reevaluated Cook's distance. The removal of new observations did not alter the significance level, suggesting that they were not influential points in our analysis. Consequently, we proceeded with the model, which included 4,867 observations.

For logistic regression assumptions, the sole continuous variable, years of experience, exhibited a linear relationship and constant variance with the predicted value. Thus no modifications

were made to the variables. Additionally, all predictor variables had VIF scores below 5, indicating that multi-collinearity was not violated.

Based on these findings, the summary output table is presented below.

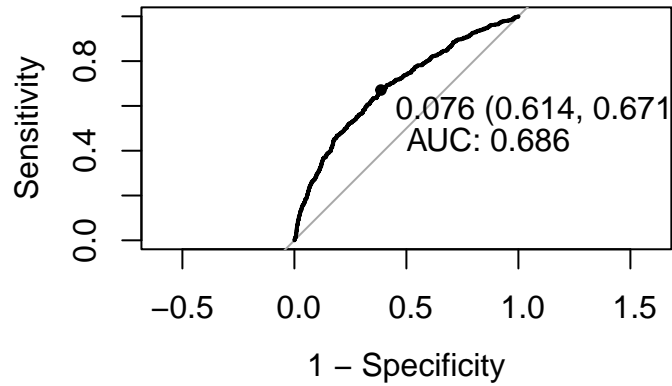
Table 1: Receive Callback

Predictors	received_callback			
	Odds Ratios	std. Error	CI	p
Intercept	0.07	0.03	0.03 – 0.16	<b>&lt;0.001</b>
Chicago	0.75	0.10	0.58 – 0.96	<b>0.025</b>
Insurance	0.87	0.19	0.55 – 1.32	0.519
Manufacture	0.69	0.17	0.41 – 1.11	0.146
Other	1.10	0.15	0.84 – 1.44	0.495
Transportation	1.57	0.45	0.87 – 2.70	0.113
Retail	0.81	0.14	0.57 – 1.13	0.217
Manager	0.61	0.15	0.38 – 0.97	<b>0.037</b>
Sales	0.77	0.18	0.49 – 1.21	0.261
Sales rep	0.62	0.15	0.38 – 1.00	<b>0.050</b>
Secretary	0.84	0.14	0.61 – 1.18	0.307
Supervisor	0.71	0.18	0.43 – 1.16	0.183
Equal opp	1.36	0.17	1.06 – 1.73	<b>0.015</b>
Req communication	1.17	0.20	0.83 – 1.64	0.353
Req education	0.58	0.13	0.37 – 0.88	<b>0.015</b>
Req computer	0.73	0.10	0.55 – 0.96	<b>0.024</b>
Req organization	0.57	0.16	0.33 – 0.96	<b>0.043</b>
Race [white]	1.56	0.17	1.26 – 1.94	<b>&lt;0.001</b>
Gender [male]	0.97	0.15	0.71 – 1.32	0.860
College degree	1.25	0.17	0.97 – 1.63	0.094
Honors	1.54	0.31	1.03 – 2.26	<b>0.030</b>
Worked during school	1.12	0.15	0.86 – 1.45	0.412
Computer skills	0.74	0.11	0.55 – 1.01	0.052
Special skills	2.07	0.25	1.63 – 2.62	<b>&lt;0.001</b>
Volunteer	0.84	0.15	0.60 – 1.19	0.316
Military	0.92	0.21	0.59 – 1.41	0.722
Employment holes	1.49	0.21	1.13 – 1.97	<b>0.004</b>
Email	1.11	0.28	0.68 – 1.82	0.688
Low resume quality	0.77	0.23	0.43 – 1.38	0.385
Years experience	1.03	0.01	1.01 – 1.05	<b>0.013</b>
Observations	4867			
R <sup>2</sup> Tjur	0.040			

## Model Assessment

Since our task is inference, we used the ROC curve to find the best cutoff point and a confusion matrix to evaluate the model. Based on figure 1, the best cutoff point is 0.076. Based on the confusion matrix, the accuracy is 0.6217, and the kappa score is 0.0986. Accuracy score is good, but kappa is low. The significant gap between these two metrics indicates that our data is imbalanced, with far fewer classifications for receiving a callback. Given this imbalance, we focus on the F1 score, which is a weighted average of precision and recall. The F1 score is 0.21892, indicating that our model doesn't perform as well.

**Fig 1: ROC Curve**



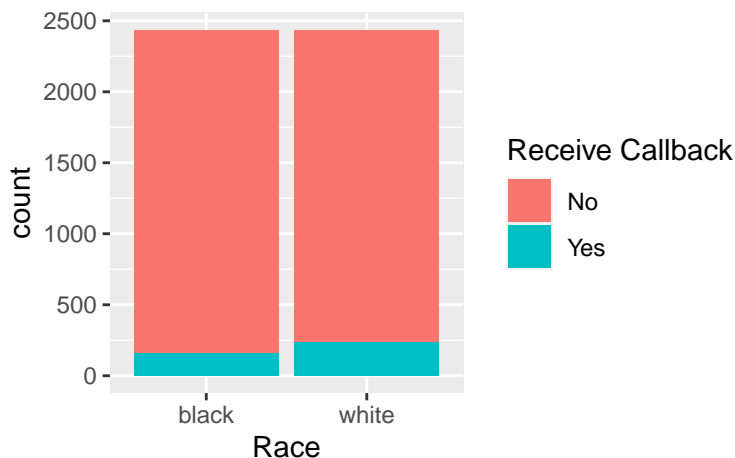
## Results

In relation to the research question, our aim is to determine the influence of race and gender on job application callback rates. Based on the summary table presented earlier, with respect to gender, while controlling for all other variables, we have found that the odds of male candidates receiving callbacks are 0.97 times higher than those of female candidates. However, based on the p-value and the significance level ( $\alpha = 0.05$ ), we do not reach a conclusive relationship between gender and applicants' callback reception. The confidence interval ranges from 0.71 to 1.32, indicating that we can be 95% confident that the odds of male candidates receiving callbacks are between 0.71 and 1.32 times higher than those of female candidates.

Regarding race, while controlling for other variables, we observed that the odds of white candidates receiving callbacks are 1.557 times higher than those of black candidates. This leads us to conclude that there is a significant linear relationship between race and callback reception in the population at the  $\alpha = 0.05$  level. The confidence interval for this comparison

falls within the range of 1.26 to 1.94, implying that we can be 95% confident that the odds of white candidates receiving callbacks are between 1.26 and 1.94 times higher than those of black candidates. Based on figure 2, we can also see that white candidates receive more callback compare to black candidates.

**Fig 2: Race and Receive Callback**



Additionally, other variables such as job location in Chicago, job type (e.g., manager or sales representative), the employer being an Equal Opportunity Employer, job-required education, computer skills, and organizational skills, applicants with a college degree, honors degree, computer skills, special skills, employment gaps, and years of experience also have an effect on job application callback rates based on p-values.

## Future Work

The validity of this analysis depends on the accuracy of the assumptions made and the representativeness of the dataset. One of the strengths of this analysis is its inclusion of numerous confounding variables related to both race, gender, and whether applicants receive a callback. However, it also has several limitations.

Firstly, we are dealing with a highly imbalanced dataset where the number of observations not receiving callbacks significantly outweighs the number of observations receiving callbacks. In the future, it would be beneficial to use a more balanced dataset for analysis.

Secondly, race and gender are inferred based on first names, but the accuracy of this inference is uncertain as we lack verified data. This introduces potential biases into the analysis.

Additionally, there are two variables with missing values: the indicator for whether the employer is a federal contractor and the type of company. If we can provide more detailed values for these variables and include them in the model, it might lead to improvements in the model's performance.