Research Question

Does race and education background affect the probability of the applicant from getting a callback from employers?

Hypothesis

"Race" has a significant effect on the probability of the applicant from getting the callback from the employers.

The years in college has a significant effect on the probability of the applicant get the callback from the employers.

The college degree listen on resume has a significant effect on the probability of the applicant to get the callback from the employers.

Our variables of interest in this hypothesis are "receivedcallback", "race", "collegeyear", "collegedegree" and "honors".

Data Cleaning

Description, type of variables and their respective levels in the data set prior to data cleaning process.

Variables	Type	Level	Description
cityjob	qualitative	2	Working city of this job
industryjob	qualitative	6	The industry of this job
typejob	qualitative	6	Role of this job
fedcontractjob	qualitative	2	whether employerthis job is a federal contractor or not
equaloppjob	qualitative	2	whether the employer is an Equal Opportunity Employer or not
ownershipjob	qualitative	4	Type of company
reqanyjob	qualitative	2	whether the job requirement is listed or not
reqcommunicationjob	qualitative	2	whether communication skills are required or not
reqeducationjob	qualitative	2	whether a certain level of education is required or not
reqminexperiencejob	quantitative	NA	Amount of minimum experience required

reqcomputerjob	qualitative	2	whether computer skills are required or not
reqorganizationjob	qualitative	2	whether organization skills are required or not
reqschooljob	qualitative	4	Required level of education
receivedcallback	qualitative	2	whether the applicant get the callback from employers or not
race	qualitative	2	Race of applicants deduced from their first names
gender	qualitative	2	Gender of applicants deduced from their first names
collegeyear	quantitative	NA	Years of college education listed on the resume
collegedegree	qualitative	2	whether the college degree was listed or not
honors	qualitative	2	whether honours awarded were listed or not
workedduringschool	qualitative	2	whether the applicant work while in school or not
yearsexperience	quantitative	NA	Years of experience listed on the resume
computerskills	qualitative	2	whether computer skills were listed or not
specialskills	qualitative	2	whether the special skills were listed or not
volunteer	qualitative	2	whether the volunteer experience was listed or not
military	qualitative	2	whether the military experience was listed or not
employmentholes	qualitative	2	whether there were employment gaps or not
hasemailaddress	qualitative	2	whether an email address was listed or not
resumequality	qualitative	2	The quality of resume classified as low or high

Column	Initial Level	Replaced Level
reqminexperiencejob	"0.5","1","10","2","3","4","5","6","7","8","some"	"0.5","1","10","2","3","4","5","6","7","8","0.25"
ownershipjob	"nonprofit","private","public","unknown"	"nonprofit","private","public",NA

For the level “some” in column “reqminexperiencejob”, we change it as 0.25. It is because we know that some experience must be larger than 0, and may not be larger than 0.5. So, we set it equals 0.25

For the level “unknown” in column “ownershipjob”, we change it to an empty value. It is because “unknown” in this column provides meaningless and ambiguous information, it could be "nonprofit", "private" or "public". To remove the ambiguity, we set it equals to an empty value.

For the level “NA” in column “fedcontractjob”, it is a string data type. We change it as an empty value.

Column “reqschooljob” has 2559 “none_listed” rows, it means that the level of education required is not listed, and is hidden. We can assume that it does not necessarily mean any education requirement. To remove the ambiguity, we planned to set it as an empty value. However, 2559 empty values are too many when the sample size is only 2870, nearly 90%. So, we decide to remove this column.

Columns “yearsexperience” and “reqminexperiencejob” should be continuous data, we set these two columns as numeric.

Column “collegeyear” should be an ordinal data, we set these columns as ordinal factors.

For the column “reqminexperiencejob”, the empty value is set as 0. It is because the whole data set does not contain any job at entry-level, it does not make sense that all jobs require more than 0 working experience.

For the “fedcontractjob” and “ownershipjob” columns, we use library “mics” to estimate the missing value based on other variables.

Methodology

To predict the binary response, logistics regression was applied. If we use the Ordinary Least Square (OLS) method in binary response, the following problem will exist:

1. <u>The response variable cannot be limited between 1 and 0</u>
   If we use OLS

$$E(Y_i) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$$

$$\pi_i = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$$

$$\pi_i = \boldsymbol{X_i}^T \boldsymbol{\beta}$$

As $0 \leq \pi_i \leq 1$, the OLS in binary response cannot allow a large value of X, otherwise π_i will be out of range.

2. Violate normality assumption in OLS

   Since $\begin{cases} Y_i = 1 \\ Y_i = 0 \end{cases}$, ε_i only take on two values.

$$\varepsilon_i = 1 - \boldsymbol{X_i}^T\boldsymbol{\beta}, \text{ if } Y_i = 1$$

$$\varepsilon_i = -\boldsymbol{X_i}^T\boldsymbol{\beta}, \text{ if } Y_i = 0.$$

   Error term ε_i is not normally distributed. The normality assumption is violated.

3. Violate homoscedasticity assumption in OLS

$$Var(\varepsilon_i) = Var(Y_i) = E[Y_i - E(Y_i)]^2$$

$$= E\left[{Y_i}^2 - 2Y_iE(Y_i) + E(Y_i)^2\right]$$

$$= E({Y_i}^2) - 2E(Y_i)E(Y_i) + E(Y_i)^2$$

$$= E\left({Y_i}^2\right) - \mathrm{E}(Y_i)^2$$

$$= \pi_i - {\pi_i}^2$$

$$= \pi_i(1 - \pi_i)$$

   The variance of error terms is not constant, it changes in different data points (i). The assumption of homoscedasticity is violated.

Based on the above issues, the OLS method cannot be applied in Binary response.

Under logistics regression:

$$\pi_i = \frac{exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)}{1 + exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)}$$

$$\pi_i = \frac{exp(\boldsymbol{X_i}^T\boldsymbol{\beta})}{1 + exp(\boldsymbol{X_i}^T\boldsymbol{\beta})}$$

$\boldsymbol{exp}(\boldsymbol{X_i}^T\boldsymbol{\beta})$ must be larger or equal to 0. In this equation, the denominator must be larger than numerator by 1, so it must be $0 \leq \pi_i \leq 1$. We can see that the value of X will not be limited.

To develop the logistics regression model, the chi-square test and a t-test with an f-test were implemented. The variables which are significantly and statistically dependent on the response variables were selected to run the AIC stepwise process. Finally, we will add the variables of interest stated before in the AIC suggested model.

By using this process, the explanation power of the interested variables will be better, as the bias from the lurking effect will be minimized by considering other variables at the same time.

Descriptive Analysis

In this part, we want to find out whether the response variable (receivedcallback) is dependent on other variables or not. We implemented the F-test and T-test for the quantitative predictors and Chi-square test for the qualitative predictors one by one.

For the continuous variables, F-test was implemented to check the equal variance between two groups of continuous data. Then, we ran the t-test to test the dependence.

F-test

Null Hypothesis: The variance of continuous data under “receivedcallback” = 0 is equal to the variance of continuous data under receivedcallback = 1
Alternative Hypothesis: The variance of continuous data under “receivedcallback” = 0 is not equal to the variance of continuous data under receivedcallback = 1

If the null hypothesis is rejected, a non-pooled T-test will be used, otherwise a pooled T-test.

T-test

Null Hypothesis: The continuous variables are independent to the “receivedcallback”
Alternative Hypothesis: The continuous variables is dependent on the “receivedcallback”

For categorical data, we ran the Chi-square test to test the dependence.

Chi-square test

Null Hypothesis: The categorical variables are independent to the “receivedcallback”
Alternative Hypothesis: The categorical variables are dependent on the “receivedcallback”

Finally, there are 9 predictors lower than 0.05 p-value in the tests, which are dependent on the response variables “receivedcallback”. They are:

"Typejob", "ownershipjob", "reqeducationjob", "reqminexperiencejob", "reqorganizationjob", "honors", "workedduringschool", "yearsexperience", "specialskills", and "employmentholes".

These 9 variables are considered as an input of stepwise processes to create a logistics regression.

Model

Purposeful Selection

The selection starts with 9 predictors which are statistically significant in the Chi-square test. The chosen alpha is 0.01 instead of 0.05. At step 1, we compared the null model to models having sole explanatory variables through the difference in deviances between the null model and each model. Based on the likelihood-ratio statistics, the result shows that the specialskills has the largest deviance difference and lowest AIC. Therefore, the specialskills is chosen as the first explanatory variable after step 1. At step 2, the deviance difference of the model including specialskills and typejob is statistically significant and has the lowest AIC, thereby choosing typejob as second predictors. At step 3, we choose honors as the third predictors due to the significant in deviance difference. However, because all models in step 4 are not significant, we stop the process of purposeful selection at 3 explanatory variables, including specialskills, typejob and honors.

Model	Explanatory Variables	Deviance	Df	AIC	Model Compared	Deviance Difference
Step 1						
1	None	1376.8	2319			
2	typejob	1349.6	2314	1361.6	(2) - (1)	27.2 (df=5)
3	reqeducationjob	1372.2	2318	1376.2	(3) - (1)	4.6 (df=1)
4	reqminexperiencejob	1369	2318	1373	(4) - (1)	7.8 (df=1)
5	reqorganizationjob	1373.4	2318	1377.4	(5) - (1)	3.4 (df=1)
6	honors	1360.2	2318	1364.2	(6) - (1)	16.6 (df=1)
7	workedduringschool	1372.3	2318	1376.3	(7) - (1)	4.5 (df=1)
8	yearsexperience	1373.4	2318	1377.4	(8) - (1)	3.4 (df=1)
9	specialskills	1343.3	2318	1347.3	(9) - (1)	33.5 (df=1)
10	employmentholes	1361.8	2318	1365.8	(10) - (1)	15 (df=1)
Step 2						
11	specialskills + typejob	1322.4	2313	1336.4	(11) - (9)	20.9 (df=5)
12	specialskills + reqeducationjob	1339.4	2317	1345.4	(12) - (9)	3.9 (df=1)
13	specialskills + reqminexperiencejob	1336.9	2317	1342.9	(13) - (9)	6.4 (df=1)
14	specialskills + reqorganizationjob	1341	2317	1347	(14) - (9)	2.3 (df=1)
15	specialskills + honors	1332.1	2317	1338.1	(15) - (9)	11.2 (df=1)
16	specialskills + workedduringschool	1341.9	2317	1347.9	(16) - (9)	1.4 (df=1)
17	specialskills + yearsexperience	1339.4	2317	1345.4	(17) - (9)	3.9 (df=1)
18	specialskills + employmentholes	1336.1	2317	1342.1	(18) - (9)	7.2 (df=1)
Step 3						
19	specialskills + typejob + reqeducationjob	1319.6	2312	1335.6	(19) - (11)	2.8 (df=1)

20	specialskills + typejob + reqminexperiencejob	1319.2	2312	1335.2	(20) - (11)	3.2 (df=1)
21	specialskills + typejob + reqorganizationjob	1320.8	2312	1336.8	(21) - (11)	1.6 (df=1)
22	specialskills + typejob + honors	1312.2	2312	1328.2	(22) - (11)	10.2 (df=1)
23	specialskills + typejob + workedduringschool	1319.5	2312	1335.5	(23) - (11)	2.9 (df=1)
24	specialskills + typejob + yearsexperience	1316.8	2312	1332.8	(24) - (11)	5.6 (df=1)
25	specialskills + typejob + employmentholes	1317	2312	1333	(25) - (11)	5.4 (df=1)
Step 4						
26	specialskills + typejob + honors + reqeducationjob	1309.4	2311	1327.4	(27) - (22)	2.8 (df=1)
27	specialskills + typejob + honors + reqminexperiencejob	1309.5	2311	1327.5	(28) - (22)	2.7 (df=1)
28	specialskills + typejob + honors + reqorganizationjob	1310.6	2311	1328.6	(29) - (22)	1.6 (df=1)
29	specialskills + typejob + honors + workedduringschool	1309.1	2311	1327.1	(31) - (22)	3.1 (df=1)
30	specialskills + typejob + honors + yearsexperience	1309.3	2311	1327.3	(32) - (22)	2.9 (df=1)
31	specialskills + typejob + honors + employmentholes	1308.4	2311	1326.4	(34) - (22)	3.8 (df=1)

StepAIC

Coefficients – Before StepAIC					
	Estimate	Std. Error	Z-values	Pr(>|z|)	
(Intercept)	-2.28262	0.33495	-6.815	9.44E-12	***
typejobmanager	-0.76636	0.30433	-2.518	0.01179	*
typejobretail_sales	-0.4368	0.27627	-1.581	0.11386	
typejobsales_rep	-0.42053	0.29728	-1.415	0.15718	
typejobsecretary	-0.39955	0.23673	-1.688	0.09145	.
typejobsupervisor	-1.57853	0.49979	-3.158	0.00159	**
reqeducationjob	-0.37332	0.29573	-1.262	0.20682	
reqminexperiencejob	-0.09685	0.06861	-1.412	0.15806	
reqorganizationjob	-0.50388	0.40326	-1.25	0.21147	
honors	0.71736	0.26929	2.664	0.00772	**
workedduringschool	-0.15541	0.18033	-0.862	0.38879	
yearsexperience	0.02306	0.01441	1.6	0.10962	
specialskills	0.68185	0.16333	4.175	2.99E-05	***
employmentholes	0.19272	0.19314	0.998	0.31835	

Based on the Chi-square test, 9 explanatory variables are chosen to run StepAIC. After the StepAIC, the suggested model comes up with 6 variables. However, only 3 variables, including typejob, honors and

specialskills, are statistically significant. To answer the research question and explore the hypothesis, the final model will include 3 mentioned variables, educational variables (college years and a college degree) and race. Therefore, explanatory variables for the final model are typejob, honors, specialskills, collegeyear, collegedegree and race.

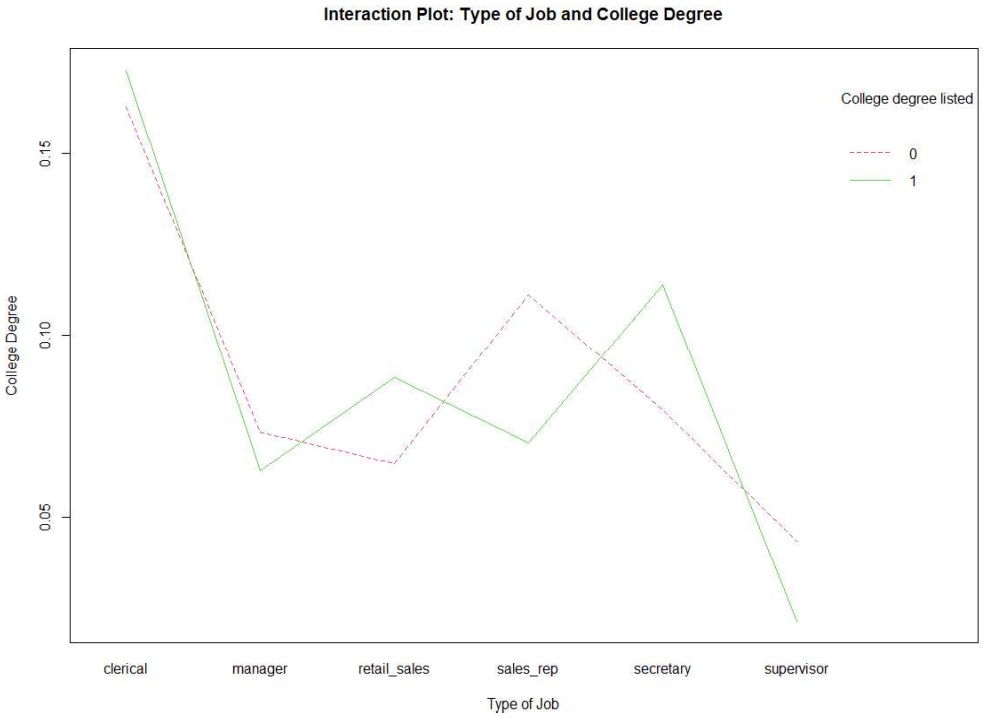
Coefficients – After StepAIC					
	Estimate	Std. Error	Z-values	Pr(>|z|)	
(Intercept)	-2.32866	0.65	-3.583	0.00034	***
typejobmanager	-1.01079	0.291	-3.474	0.000514	***
typejobretail_sales	-0.64239	0.26777	-2.399	0.016438	*
typejobsales_rep	-0.65683	0.29535	-2.224	0.026154	*
typejobsecretary	-0.53454	0.23243	-2.3	0.021459	*
typejobsupervisor	-1.71044	0.49511	-3.455	0.000551	***
honors	0.85023	0.25592	3.322	0.000893	***
specialskills	0.79636	0.16087	4.95	7.41E-07	***
collegeyear	-0.05536	0.21601	-0.256	0.797725	
collegedegree	0.3745	0.3531	1.061	0.288878	
racewhite	0.30673	0.15121	2.029	0.042506	*

Equation based on stepAIC output:

$$logitP[(y = 1)] = -2.32866 - 1.01079 * typejobmanager - 0.64239 * typejobretail\_sales - 0.65683 * typejobsales\_rep - 0.53454 * typejobsecretary - 1.71044 * typejobsupervisor + 0.85023 * honors + 0.79636 * specialskills - 0.05536 * collegeyear + 0.3745 * collegedegree + 0.30673 * racewhite$$

Baseline level = typejob clerical, black race

Adding 1 interaction term



This plot shows the interaction between type of job and whether a college degree was listed on the resume. As the lines intersected at some points, the interaction between these two variables in determining if applicants received a callback existed. Hence, the significance of the interaction term was checked using the likelihood ratio test. The residual deviances from models with and without interaction terms were summarized in Table 1.

Ho: Reduced model is appropriate
Ha: Full model is appropriate

	Model without the interaction term	Model with the interaction term (typejob and collegedegree)
Residual deviance	1305.2	1303
Degrees of freedom (df)	2309	2304
Differences between 2 models, and p-value	Residual deviance differences = 2.2, df differences = 5, p-value = 0.817	

Table 1. Summary of residual deviances and degrees of freedom for models with and without the interaction term.

The p-value from the likelihood ratio test is 0.817, greater than the significance level of 0.05. Therefore, we fail to reject Ho and conclude that the reduced model without the interaction term is appropriate and that the interaction between job type and a college degree is not statistically significant.

Interpretation of Coefficients Based on AIC (stepAIC) Output:

For each level of other predictors, the odds of people whose job type is manager is 0.3639 times the odds of people having other job types.

For each level of other predictors, the odds of people whose job type is retail sales is 0.5260 times the odds of people having other job types.

For each level of other predictors, the odds of people whose job type is sales representatives is 0.5185 times the odds of people having other job types.

For each level of other predictors, the odds of people whose job type is secretary is 0.5859 times the odds of people having other job types.

For each level of other predictors, the odds of people whose job type is supervisor is 0.1808 times the odds of people having other job types.

For each level of other predictors, the odds of people whose resume listed that the candidate has been awarded some honors is 2.3402 times the odds of people whose resume did not.

For each level of other predictors, the odds of people whose resume listed special skills is 2.2175 times the odds of people whose resume is not listed.

For each level of other predictors, one unit increase in year experience has a multiplicative effect of 0.9461 on the odds that Y=1.

For each level of other predictors, the odds of people whose resume listed a college degree is 1.4543 times the odds of people whose resume did not.

For each level of other predictors, the odds of people whose race is white is 1.3590 times the odds of people who are not.

95% Confidence Interval of Odds Ratio Based on AIC (stepAIC) Output

	2.5 %	97.5 %
(Intercept)	0.02725175	0.3483043
typejobmanager	0.20574071	0.6437468
typejobretail_sales	0.31123257	0.8890746
typejobsales_rep	0.29062937	0.9250041
typejobsecretary	0.37154104	0.9240505
typejobsupervisor	0.06850543	0.477098
honors	1.41713226	3.864505
specialskills	1.61778654	3.0393815
collegeyear	0.61956481	1.4448613
collegedegree	0.72791277	2.9053891
racewhite	1.01042141	1.8277455

Final Model

The final model used will be based from the stepAIC as typejob and specialskills were significant for applicants in receiving call back. Based on the equation below, for applicants applying for managerial role, honors, college degree and special skills were listed on resume, went to college for four years and are perceived as black from their first name would have an estimated probability of 0.2144 of receiving a call back from employers.

Equation based on stepAIC output:

$$logitP[(y = 1)] = -2.32866 - 1.01079 * typejobmanager - 0.64239 * typejobretail\_sales - 0.65683 * typejobsales\_rep - 0.53454 * typejobsecretary - 1.71044 * typejobsupervisor + 0.85023 * honors + 0.79636 * specialskills - 0.05536 * collegeyear + 0.3745 * collegedegree + 0.30673 * racewhite$$

Baseline level = typejob clerical, black race

Likelihood-Ratio Tests for Individual Explanatory Variable

Ho: $\beta = 0$	P-Value	Interpretation
typejob	0.001	We reject Ho. We have strong statistical evidence to conclude that the job type significantly increases the odds of receiving a callback, keeping all other variables constant.
honors	0.002	We reject Ho. We have strong statistical evidence to conclude that listing relevant honors awarded on the resume significantly increases the odds of receiving a callback, keeping all other variables constant.
specialskills	0.000	We reject Ho. We have strong statistical evidence to conclude that listing any special skills on the resume significantly increases the odds of receiving a callback, keeping all other variables constant.
race	0.041	We reject Ho. We have strong statistical evidence to conclude that if the applicant's first name is inferred as being of the white race, the odds significantly increase of receiving a callback, keeping all other variables constant.
collegeyear	0.800	We fail to reject Ho. We do not have strong statistical evidence to conclude that listing the number of college years on the resume significantly increases the odds of receiving a callback.
collegedegree	0.293	We fail to reject Ho. We do not have strong statistical evidence to conclude that listing a college degree on the resume significantly increases the odds of receiving a callback.

Wald Test for Individual Coefficients

Ho: $\beta = 0$	P-Value	Interpretation
typejobmanager	0.001	We reject Ho. We have strong statistical evidence to conclude that applying for a managerial job type significantly increases the odds of receiving a callback, keeping all other variables constant.
typejobretail_sales	0.016	We reject Ho. We have strong statistical evidence to conclude that applying for a retail sales job type significantly increases the odds of receiving a callback, keeping all other variables constant.
typejobsales_rep	0.026	We reject Ho. We have strong statistical evidence to conclude that applying for a sales representative job type significantly increases the odds of receiving a callback, keeping all other variables constant.
typejobsecretary	0.021	We reject Ho. We have strong statistical evidence to conclude that applying for a secretarial job type significantly increases the odds of receiving a callback, keeping all other variables constant.
typejobsupervisor	0.001	We reject Ho. We have strong statistical evidence to conclude that applying for a supervisor job type significantly increases the odds of receiving a callback, keeping all other variables constant.
honors	0.001	We reject Ho. We have strong statistical evidence to conclude that listing relevant honors awarded on the resume significantly increases the odds of receiving a callback, keeping all other variables constant.
specialskills	0.000	We reject Ho. We have strong statistical evidence to conclude that listing any special skills on the resume significantly increases the odds of receiving a callback, keeping all other variables constant.
collegeyear	0.798	We fail to reject Ho. We do not have strong statistical evidence to conclude that listing the number of college years on the resume significantly increases the odds of receiving a callback.
collegedegree	0.289	We fail to reject Ho. We do not have strong statistical evidence to conclude that listing a college degree on the resume significantly increases the odds of receiving a callback.
racewhite	0.043	We reject Ho. We have strong statistical evidence to conclude that if the applicant's first name is inferred as being of the white race, the odds significantly increase of receiving a callback, keeping all other variables constant.

Profile Likelihood Confidence Interval

Variable	Profile Likelihood Interval	Interpretation
typejobmanager	$(e^{-1.59}, e^{-0.44}) = (0.21, 0.64)$	We infer that if the job type is managerial, the odds of receiving a callback are between 0.21 and 0.64.
typejobretail_sales	$(e^{-1.17}, e^{-0.11}) = (0.31, 0.89)$	We infer that if the job type is retail sales, the odds of receiving a callback are between 0.31 and 0.89.
typejobsales_rep	$(e^{-1.24}, e^{-0.08}) = (0.28, 0.92)$	We infer that if the job type is sales representative, the odds of receiving a callback are between 0.28 and 0.92.
typejobsecretary	$(e^{-0.98}, e^{-0.07}) = (0.38, 0.93)$	We infer that if the job type is secretarial, the odds of receiving a callback are between 0.38 and 0.93.
typejobsupervisor	$(e^{-2.81}, e^{-0.82}) = (0.06, 0.44)$	We infer that if the job type is supervisor, the odds of receiving a callback are between 0.06 and 0.44.
honors	$(e^{0.33}, e^{1.34}) = (1.39, 3.92)$	We infer that if the applicant has his/her honors listed in the resume, the odds of receiving a callback are between 1.39 and 3.92.
specialskills	$(e^{0.48}, e^{1.11}) = (1.61, 3.03)$	We infer that if the applicant has his/her special skills listed in the resume, the odds of receiving a callback are between 1.61 and 3.03.
collegeyear	$(e^{-0.45}, e^{0.41}) = (0.64, 1.51)$	We infer that if the applicant has his/her number of years spent in college listed in the resume, the odds of receiving a callback are between 0.64 and 1.51.
collegedegree	$(e^{-0.33}, e^{1.06}) = (0.72, 2.89)$	We infer that if the applicant has his/her college degree listed in the resume, the odds of receiving a callback are between 0.72 and 2.89.
racewhite	$(e^{0.01}, e^{0.61}) = (1.01, 1.84)$	We infer that if the applicant's first name is inferred as being of the white race, the odds of receiving a callback are between 1.01 and 1.84.

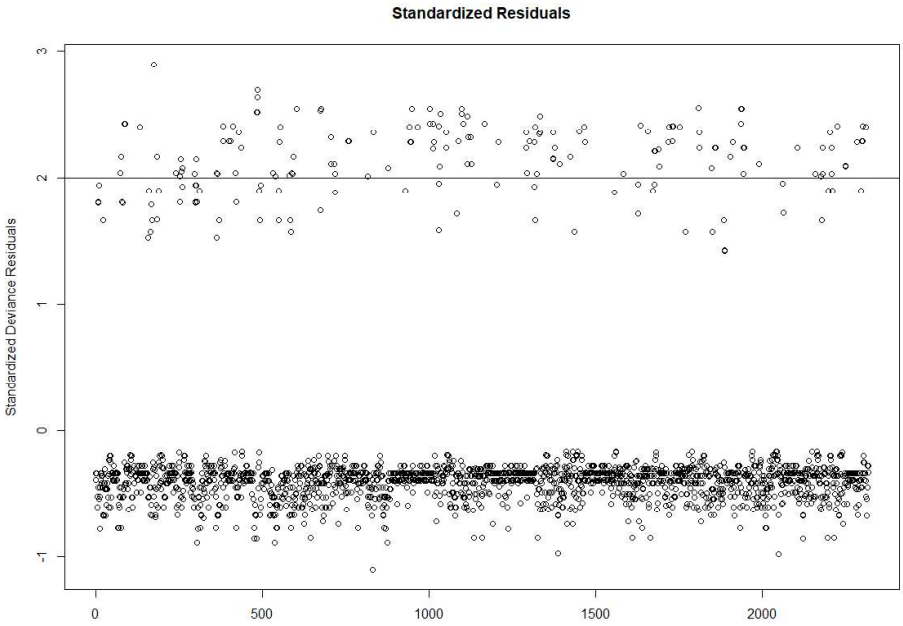
Goodness of Fit of Final Model

To determine if the current model is appropriate and is of a good fit, the Hosmer-Lemeshow test was utilized as the data is not grouped data.

Ho: Reduced model is appropriate
Ha: Full model is appropriate

The p-value from the Hosmer-Lemeshow test is 0.8651, greater than the significance level 0.05. Therefore, we fail to reject Ho. We can conclude that there is no strong evidence of a lack of fit for the current model and it is appropriate.

Residuals



The standardized deviance residual for ungrouped data was used to compare observations to the model fit in the figure above. 142 observations had values that were greater than two.

Classification Table

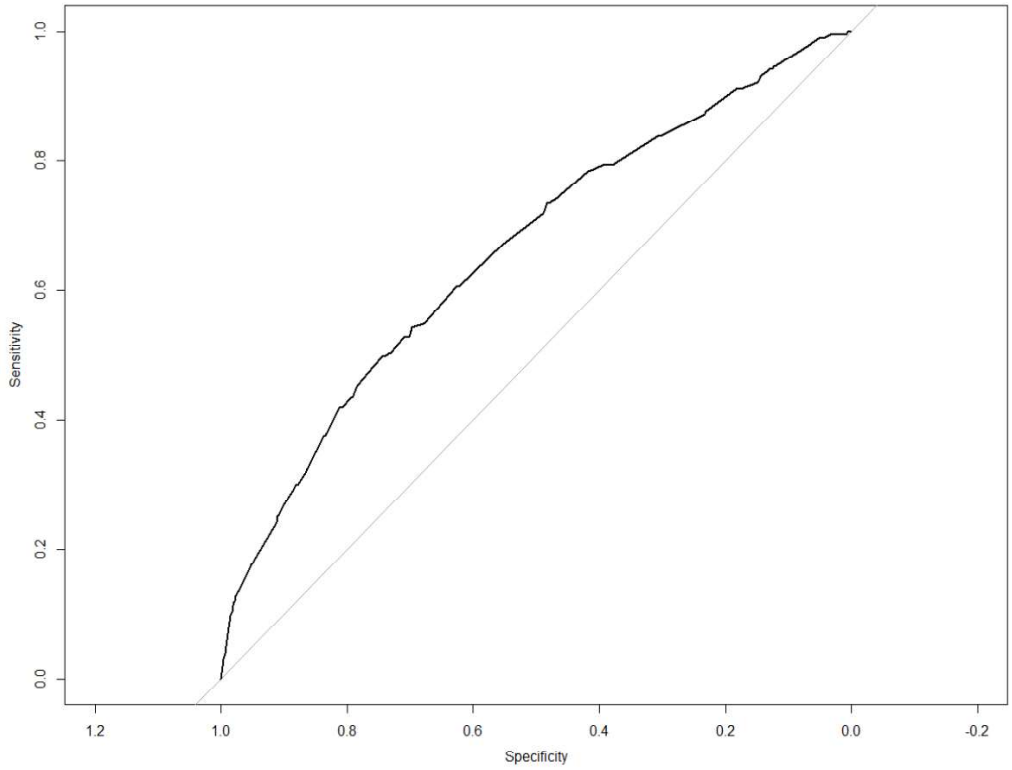
Confusion matrix also known as a classification table (Table 2) showed how well the model was able to predict the outcomes of applicants getting a callback from employers considering applicants' education and perceived race based on the final model.

	Predicted	
Actual	0	1
0	356	21
1	152	21

Table 2. Classification table of the actual and predicted number of applicants receiving a callback from employers.

The sensitivity- proportions of applicants correctly predicted to receive a callback from employers when they did receive a callback- is 0.5. The specificity- proportions of applicants correctly predicted to not receive a callback from employers when they did not receive a callback- is 0.7. Overall, the accuracy of the model in predicting the outcomes is 68.5%.

Receiver Operating Characteristic (ROC) Plot



The concordance index with predictors typejob, honors, specialskills, collegeyear, collegedegree, and race is 0.6599. In statistics, we utilize the concordance index to compare the predictive power of different models; however, overall, the ROC plot shows that our recommended model does well at predicting if an applicant will receive a callback as the ROC curve is above the plot's main diagonal line.

Conclusion

The final model was able to correctly predict the outcomes of applicants receiving callback from employers 68% of the time when considering their race and educational background. The model can be improved if the research was conducted to determine important factors which affect the outcome of receiving callback from employers.