Workplace Discrimination based on Sex, Race and Color in Canada

Xiaoke Zeng (1004005266)

Dec 21, 2020

Abstract

Discrimination in the workplace has always been a concern all over the world. This report focuses on sex and race discrimination. It utilizes Binary Logistic Regression Model to predict the probability of being treated unfairly or being discriminated at the workplace from people's sex, race, level of education, occupation, and hours worked per week. Such probability turns out to be different between discrimination based on sex and based on race for people in the same education level, occupation, and hours worked.

Keywords

Workplace Discrimination, Sex, Race, Color, Visible Minority, Education Level, Working Hours, Occupation

Introduction

Discrimination is defined as an individual or groups being treated unfairly based on certain characteristics, including religion, ethnicity, sex, age, family status and sexual orientation. Even though efforts have been made to eliminate discrimination by implementing various equity policies, it sometimes happens subtly. Typically, women of the visible minority are put at the most disadvantaged stage. This study aims to figure out the hardest-hit area of sex and race discrimination to raise awareness of inequality.

There are different workplace discrimination scenarios, including being ignored by others, setting up promotion barriers, being given too much work, etc. This study will focus on the characteristics of sex, race, and colour as the main reasons for discrimination. It will provide valuable insight into the current workplace discrimination through Canada's analysis, where people are well diversified in race. Connected with Canada's employment equity policies, the government brought up a term of "visible minority", referring to "persons, other than Aboriginal people, who are non-Caucasian in race or non-white in colour". In this study, race and colour will be simplified into whether they belong to a "visible minority".

One data set will be used to analyze the probability of being treated unfairly or being discriminated against based on race or sex. In the Methodology section, there presents data and the model that was used to perform the analysis. Results of the probability analysis are provided in the Results section. Inferences of this data, as well as forward-looking conclusions, are presented in the Conclusion section.

Methodology

#Data

The dataset for analysis is real observational data (Table 1). It includes information about respondents' personal income before-tax, current employment status, highest education level, hours worked per week, occupation for the past year, sex, whether they are visible minority, whether they have been treated unfairly based on sex, and whether they have been treated unfairly based on race. The last two are the response variables. Sex and visible minority indicators are the main focus of this study. Current employment status is used to filter out respondents who are only self-employed or not being employed.

Even though sex and visible minority indicator are the main focus, other significant factors that affect an individual's working ability and how hardworking they are, such as highest education level, hours worked per week, and occupation, are also included in the model as control variables. Controlling these variables at the same level, the probability of being treated unfairly is not supposed to be different based on sex or race. To get a more precise estimation, only respondents that provided valid information to all indicators are included in the sample frame. In other words, all responses of "N/A", "Valid Skip", "Don't know", and "Refusal" are dropped.

#Model A binary logistic regression model is used to predict the probability of being treated unfairly or being discriminated. Two regressions are made for characteristics of sex and whether they are visible minority respectively. The model follows the following format: Ln(P/(1-P))= m0 + m1 sex/minority indicator + m2 highschool + m3 trade + m4 college + m5 university + m6 bachelor + m7 above + m8 workhour16_29 + m9 workhour30_40 + m10 hourover_40 + m11 manage + m12 bus + m13 science + m14 health + m15 education + m16 art + m17sales + m18transport + m19res. To illustrate, for example, if an individual has a bachelor's degree, the input of "bachelor" is 1, while "highschool", "trade", "college", "university", "above" are 0. A positive coefficient of binary predictors makes the probability more likely and vice versa.

Some assumptions are applied here: 1.The dependent variable is binary. 2.Observations are independent of each other, which was satisfied by the nature of this survey. 3.There is little multicollinearity among the independent variables. 4.The assumptions of a large sample size are satisfied as there are over 9000 observations.

```
##
## Call:  glm(formula = df7$unfs ~ df7$mino + df7$highsch + df7$trade +
##     df7$college + df7$uni + df7$bach + df7$above + df7$manage +
##     df7$bus + df7$sci + df7$health + df7$edu + df7$art + df7$sales +
##     df7$trans + df7$res + df7$f16_29 + df7$f30_40 + df7$over_40)
##
## Coefficients:
## (Intercept)     df7$mino  df7$highsch    df7$trade  df7$college      df7$uni
##    0.004114    -0.008430     0.004935     0.006515     0.011027     0.014375
##    df7$bach    df7$above   df7$manage      df7$bus      df7$sci   df7$health
##    0.023377     0.037236    -0.011989     0.009499     0.016456     0.002141
##     df7$edu      df7$art    df7$sales    df7$trans      df7$res   df7$f16_29
##    0.005391     0.011964     0.013507    -0.008016    -0.010024    -0.005425
##  df7$f30_40  df7$over_40
##    0.001137     0.010291
##
## Degrees of Freedom: 8429 Total (i.e. Null);  8410 Residual
## Null Deviance:       201.9
## Residual Deviance: 200.2     AIC: -7566
```

```
##
## Call:  glm(formula = df7$unfr ~ df7$sex + df7$highsch + df7$trade +
##     df7$college + df7$uni + df7$bach + df7$above + df7$manage +
##     df7$bus + df7$sci + df7$health + df7$edu + df7$art + df7$sales +
##     df7$trans + df7$res + df7$f16_29 + df7$f30_40 + df7$over_40)
##
## Coefficients:
## (Intercept)      df7$sex  df7$highsch    df7$trade  df7$college      df7$uni
##    0.005430     0.002770     0.002622     0.018241     0.006722     0.016685
```

```
##    df7$bach    df7$above   df7$manage      df7$bus      df7$sci   df7$health
##    0.015638     0.014867    -0.023304    -0.017517    -0.012455    -0.003412
##     df7$edu      df7$art    df7$sales    df7$trans      df7$res   df7$f16_29
##   -0.015388    -0.026726    -0.006004    -0.013123    -0.021961     0.013024
##  df7$f30_40  df7$over_40
##    0.010061     0.007020
##
## Degrees of Freedom: 8429 Total (i.e. Null);  8410 Residual
## Null Deviance:       110.5
## Residual Deviance: 109.9     AIC: -12620
```

Results

Probability of being treated unfairly based on sex: Ln(Psex/(1-Psex)) = 0.004114 - 0.008430mino + 0.004935highschool + 0.006515trade + 0.011027 college + 0.014375university + 0.023377bachelor + 0.037236 above -0.005425 workhour16_29 + 0.001137 workhour30_40 + 0.010291 hourover_40 - 0.011989manage + 0.009499bus + 0.016456 science - 0.002141health + 0.005391education + 0.011964art + 0.013507sales -0.008016 transport -0.010024 res

From the result, the visible minority are less likely to face gender discrimination at the workplace. The higher the level of education, the more likely they are being treated unfairly based on sex. The longer hours people work, the more likely they believe they are being treated unfairly. As for occupation, it is the least likely to face discrimination at the workplace at a management position, while the probability is highest for natural and applied sciences and related occupations.

Probability of being treated unfairly based on race: Ln(Prace/(1-Prace)) = 0.005430 + 0.002770sex + 0.002622highschool + 0.018241trade + 0.006722college + 0.016685university + 0.015638bachelor + 0.014867above + 0.013024workhour16_29 + 0.010061workhour30_40 + 0.007020hourover_40 - 0.023304manage -0.017517bus – 0.012455science - 0.003412 health -0.015388education -0.026726art – -0.006004 sales -0.013123transport -0.021961res

From the result, men are more likely to face workplace discrimination based on race than women. People who possess a university education level are the most likely to face such unfair treatment. The fewer hours people work, the more likely they believe they are being treated unfairly. As for occupation, it is the least likely to face discrimination at the workplace having an art-related occupation, while the probability is highest for people having manufacturing and utility occupations.

Discussion

#Summary

To do the analysis, I first filtered the data by multiple filters and dropped the uncategorized response. Next, a binary logistic regression model is chosen. Assumptions are tested to satisfy the use of the model. Finally, I ran two logistic regression on the probability of being treated unfairly or being discriminated against at the workplace based on sex and whether they are visible minorities, respectively. The indicator of sex is excluded for the Psex prediction, and the visible minority indicator is excluded from the Prace prediction to satisfy independence assumptions..

#Conclusions

How race and visible minority interact together is a bit surprising, which contradicts common sense that female minorities are put at the least advantaged stage. The female non-visible minority are more likely to face gender discrimination, while the visible male minority are more likely to face racial discrimination. For gender discrimination, it makes sense that females don't feel being valued equivalent to their knowledge, i.e. education level, and how hardworking they are, i.e. working hours. Men are typically assumed to be better in science, and therefore women tend not to be thought highly of. As long as people reached the management position level, the probability of discrimination is significantly lower for both race and gender discrimination. As for race discrimination in different occupations, art is boundless regardless of race and colour. The highest probability for manufacturing jobs is a bit hard to explain. Considering the differences between manufacturing and sales occupations are minimal, it could be biased. Sales positions typically require strong communication and interpersonal skills, where the use of language is of vital importance. Therefore, a higher race discrimination probability for sales discrimination could be a result of language concern.

Weaknesses & Next Steps

The first potential weakness is that, even though significant independent variables have been included, some other important variables in the model may still be excluded. The second weakness rises from data dropped. Suppose the group of people who provided some invalid answers is similar to those who provided all valid answers. In that case, the results should not be significantly affected by this. The probability prediction can be biased otherwise. Lastly, the analysis is only based on information collected in 2015. Data from other years can help improve the accuracy of the probability prediction. The next step for this study is to compare the changes in workplace discrimination based on sex and race as time goes. It helps to visualize the achievements of the implementation of relevant anti-discrimination policies.

References

[1]R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

[2] Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686