

Categorical Data Analysis, Prelim 2

STSCI 4110 ---- April 9, 2019

Cover Sheet

Name _____

Net ID _____

Academic integrity is expected of all students of Cornell University at all times, whether in the presence or absence of members of the faculty.

Understanding this, I declare I shall not give, use, or receive unauthorized aid in this examination.

Signatures of Students _____

Notes:

- > This cover sheet should be the first page of your completed report.
- > You should submit both paper and electronic copies of your work. The electronic copy may be checked for plagiarism. You should submit your electronic version via Blackboard. There will be an "Assignment" link, where it can be uploaded.
- > Your electronic version should be two files: 1) Your written report with graphs and explanations of your work. 2) R code for the analysis. You do NOT need to include all the R code that you ran, just the R code that generated the analyses described in your report.
- > The report should be the work of you and your partner(s), written in your own words. Do not discuss the exam with other classmates or students, professors or other people. You may ask me questions. It should be no longer than 12 pages long.
- > The deadline for submission of your exam shall be Wednesday April 17 at 4 pm. You should hand it in at 1198 Comstock Hall.

Analysis of Immigrant Labor Certification

In the current labor market, companies in the US sometimes look to hire immigrants for certain job positions. One step in this process is the application for permanent immigration labor certification. This is a critical stage on the path to granting employment-based green cards.

The US government is required to publish records of all requests for immigrant labor certification made by companies who request to sponsor a job applicant. Applications are reviewed by Department of Labor agents, who either approve or reject labor certification.

Note that current US laws forbid discrimination on the basis of nationality. You will have the opportunity to investigate whether discrimination is prevalent or not based on the records in the dataset for this Prelim 2 exam.

This dataset contains information on labor applications made during two time periods: Oct-Dec 2015 (Obama administration) and Oct-Dec 2018 (Trump administration).

A total of 42,204 labor certification applications were made during these two time periods. A random sub-sample was taken, and data from 4027 applications are provided in the file, "Labor_Cert_FY2016_19_V1.xlsx". (Note that the original data files, which are publicly available, have many more variables, and 10 times as many observations for this time period.)

The objective of this assignment is to analyze and determine which factors might be predictive of successful immigrant labor certification. You should determine what factors are most important and how they affect immigrant labor certification. This information would be very useful for companies, lawyers and potential employees, all of whom want to know which factors or combination of factors are most predictive of the highest chances of successful immigrant labor certification.

Your report should include the fitting and evaluation of a logistic model and a description of the steps you took to fit this model. Include plots, as appropriate. If

there are situations during the analysis when you would want to discuss a next step with a lawyer or immigrant expert for advice, mention this in your report. Make any such decisions, as best you can. You should also include a final summary of your findings (in plain English).

Your report should address these three main questions:

- 1) Which factors are predictive of successful immigrant labor certification?
- 2) Is their evidence of discrimination on the basis of nationality? If so, what were the patterns of discrimination?
- 3) What differences, if any, are there between immigrant labor certification under the Obama versus Trump administrations?

You do not have to include a description of every single thing you did or every step that you took; you should outline the steps that you took, with relevant intermediate results. Your report should have more detail than you would find in a typical published paper, but less than a step-by-step description.

You should use the document called *Take Home Project Guidelines* to help guide you in the writing of your report.

If you have questions, you may email me directly at ms429@cornell.edu. If the questions are relevant to everyone, I will post them on Piazza with answers. Likewise, if I have clarifications or corrections, I will post them on Piazza.

Good luck and enjoy!

Melissa Smith
(ms429)