Introduction

Let's use logistic regression to investigate the hiring practice and potential sex discrimination in a particular company. Again, remember that:

- Every word matters, both in the problem description and your answers.
- **You must interpret the findings in the same way that was done in the class and lecture videos.** In other words, you can use what were explained in the class as interpretation templates and ***modify*** them accordingly for the present assignment.
- **You will lose points if you use x's and y's in your interpretations.** Suppose that you are interpreting your findings before a judge our jury. They are not familiar with x's and y's.
- You can log in to the SAS DataWorks on the remote server from home (or from any computer in the College of Business labs) in the same manner that we did in the class and lecture videos.

Investigation

An equal-rights group claims that females are less likely to be hired than males with the same background, experience, and other qualifications in a particular firm. You are called to investigate such allegations of sex discrimination in the hiring practices of that firm. The data for this investigation is called discrim.dat (which can be found in the Code_Data folder). Followings are the columns in that data file respectively (from left to right; you must use the same sequence to input them in your SAS code):

$$HiringStatus = \begin{cases} 1 & if\ hired \\ 0 & if\ not \end{cases}$$ **Important Note: *HiringStatus* is the dependent variable for this problem.**

$$Education = Years\ of\ higher\ education$$

$$Experience = Years\ of\ experience$$

$$Gender = \begin{cases} 1 & if\ male\ applicant \\ 0 & if\ female\ applicant \end{cases}$$

Modify the SAS logistic regression code that was explained in the lecture to conduct a logistic regression analysis for the present discrimination problem. Change the SAS code to fit the **following logistic regression model**:

$$\pi^* = \beta_0 + \beta_1 * Education + \beta_2 * Experience + \beta_3 * Gender$$

In other words, you must repeat the same process demonstrated in the lecture, but this time with the discrim.dat data file. You should also change your INFILE line, use new variable names in your INPUT line (e.g., INPUT HiringStatus Education Experience Gender;), and change your MODEL line accordingly (e.g., MODEL HiringStatus= Education Experience Gender / ctable pprob=.3 to .7 by .1 ;).

***As a point of reference**, SAS's estimate for the Gender coefficient in the Analysis of Maximum Likelihood Estimates Table should be equal to 5.6037.*

After running your code, use its output to answer the following questions.

Note: *For this assignment, it's a good strategy to read through all the questions once first, and then start answering them.*

Questions

1. See the problem statement and equation in page 1 and define π for this specific problem. Your answer should start with: "π in this problem is the probability of (or that) _____."
2. Define "odds" for this specific problem. You can either write the equation for the odds ratio, or answer like: "odds in this problem are the probability of (or that) _____ **divided** by the probability of (or that) _____."
3. What does π^* in the equation in page 1 represent?
4. What is the Global Null Hypothesis for this specific problem (express that with **both** Betas and Chi-square)? In your SAS printout, use the Likelihood Ratio row in the Global Null Hypothesis section and reason whether this is a statistically useful model or not (use $\alpha = 0.05$ to reason).
5. Is *Years of Higher Education* a statistically useful factor in the above model? Why? Use $\alpha = 0.05$ to reason.
6. Is *Years of Higher Experience* a statistically useful factor in the above model? Why? Use $\alpha = 0.05$ to reason.
7. Interpret β_2 (the impact of Experience on the odds of Getting Hired) in layman's terms.
8. What is the Null Hypothesis corresponding to *Gender* in this problem (express that with **both** Betas and Chi-square)?
9. Using the SAS findings, can you reject the Null Hypothesis in Question #8? In other words, does *Gender* statistically matters in getting hired by this particular firm? Why? Use $\alpha = 0.05$ to reason.
10. *Gender* is a categorical (binary) variable. Suppose that Gender is equal to zero (female applicant). In that case, when Gender is increased by one, it basically becomes equal to one (male applicant). In other words, gender is changing from Female to Male. Try to you use this logic to interpret β_3 (the impact of Gender on getting hired) in layman's terms. What does this interpretation tell you about discrimination against female applicants in the hiring practices of this particular firm?

    Important Point: Note that in your SAS output, $\beta_{Gender} = 5.6037$, which makes e^{β} and accordingly the resulting change percentage fairly large numbers. At the same time, remember that in this specific case you can**NOT** modify the unit change to, for example, "for every 0.1 unit change in gender…" to make the corresponding change percentage smaller. In other words, you can only change gender by one unit, e.g., from 0 (Female) to 1 (Male). Therefore, in this specific case it is perfectly fine to have a large number for the percentage of change in the odds in your interpretation.