Homework 5: Sexism at Berkeley

36-313, Fall 2021

Due at 6 pm on Thursday, 7 October 2021

In the US, discrimination in employment on the basis of sex was outlawed by the Civil Rights Act of 1964[1]. The famous "Title IX" of the Education Amendment Act[2] was passed in 1972. Before these laws, sex discrimination in education was, for the most part[3], entirely legal, and undeniably common. Whether, after the passage of these laws, schools actually compiled with them and stopped discriminating on the basis of sex was, initially at least, very much an open question.

This brings us to this week's data set, a famous[4] one about sex discrimination in admission to graduate programs at the University of California at Berkeley, a.k.a. "Cal"[5], for the fall of 1973. It's built into R, as `UCBAdmissions`, where it takes the unusual form of a 3-dimensional array. The first dimension ("rows") is "admitted" or "rejected". The second dimension ("columns") is "male" or "female". The third dimension ("layers") is a department, labeled by a letter from "A" to "F"[6]. The entries in the array are the number of applicants to that department, of that sex, by admissions status. Thus if we look at the last layer

```
data(UCBAdmissions)  # Load the data set (only needs to be done once)
UCBAdmissions[, , "F"]
```

```
##          Gender
## Admit     Male Female
##   Admitted  22     24
##   Rejected 351    317
```

we see the counts of applicants to department "F", and the department's admissions decisions about them.

1. *One department*
   a. (2) What proportion of all applicants were accepted by department F? What proportion of male applicants? What proportion of female applicants?
   b. (1) What is the difference in male and female admissions rates?
   c. (4) Using the bootstrap, give a 95% confidence interval for the difference in admissions rates. Does it contain 0? What can you conclude from the confidence interval? (Here and elsewhere, if you can't figure out how to make the bootstrap work, but can do something else to get a confidence interval, there will be partial credit.)
   d. (1) What is the ratio of male to female admissions rates?
   e. (4) Using the bootstrap, give a 95% confidence interval for the ratio. Does it contain 1? What can you conclude from this confidence interval?

---

[1] The original language of the law mentioned discrimination based on race, color, religion and national origin, but not sex. It's sometimes claimed that sex was added as an amendment by an opponent, in a failed attempt to make the proposed law unpalatable, but this is disputed, and it may have actually arisen, in part, from the Republican Party's traditional (!) support of equal rights for women (Freeman 1991). (The 1960s and 1970s saw a huge re-arrangement of which political positions were supported by which parties, and can make what happened before then seem very strange to those of us who grew up afterwards.)

[2] "No person in the United States shall, based on sex, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any education program or activity receiving Federal financial assistance." (Incidentally, the term "title" here is legal jargon; acts of Congress, and similar legal documents, are divided into parts called "titles".)

[3] There were some state and local laws against, in some parts of the country, but nothing at the national level.

[4] Among statisticans, anyway.

[5] *Disclaimer*: I went to Cal for undergrad. *Anti-disclaimer*: This data is literally from before I was born.

[6] I have been unable to discover the names of the departments.

    f. (4) Are the interval for the difference and the interval for the ratio suggesting similar or divergent stories about admissions to department F?

2. *All departments*
    a. (4) Make a table or figure showing the overall, the male and the female admissions rates for all six departments. (This will display 18 numbers.)
    b. (4) Make a table or figure showing the difference in admissions rates for each department, plus 95% confidence intervals.
    c. (4) Make a table or figure for the ratios of admissions rates by department, plus 95% CIs.
    d. (5) Over-all, do males or females have higher admissions rates? Are there exceptions to the general pattern? Justify your answers by referring to the results in Q2a, Q2b and Q2c.
    e. (5) How (if at all) could the pattern you found in Q2d be explained by discrimination by Cal?
    f. (5) How could the pattern you found in Q2d be explained *without* discrimination by Cal?

3. *The university as a whole*
    a. (2) Add up across departments to produce a 2×2 table showing the total number of applicants, broken down by sex and by admissions decision.
    b. (2) What's the difference in over-all admissions rates, and a 95% CI?
    c. (2) What's the ratio of over-all admissions rates, and a 95% CI?
    d. (4) At the university level, do males or females have higher admissions rates? Justify your answer by referring to the results in Q3a, Q3b and Q3c.
    e. (5) How could the pattern you found in Q3d be explained by discrimination by Cal?
    f. (5) How could the pattern you found in Q3d be explained without discrimination by Cal?

4. *Reconciling two analyses*
    a. (5) If you've done everything right, the natural or obvious inferences from Q2d and Q3d should point to exactly opposite conclusions about sex discrimination at Cal in 1973. Explain why the two interpretations of the same data, by basically the same methods, are in conflict.
    b. (2) For each department, calculate what proportion of the applicants are female (whether or not they're admitted). Provide your answer in the form of a table or figure.
    c. (4) Make a plot showing the admissions rate on the vertical axis and proportion of female applicants on the horizontal axis. For each department, make marks for the overall, male and female admissions rates. (There should thus be 18 points on this plot.)
    d. (5) Explain how the plot you made in Q4c resolves the apparent contradiction you found in Q4a.

5. *Conclusions* As we discuss in lecture, American law distinguishes **disparate treatment** from **disparate impact**. An organization engages in disparate treatment when it *directly* considers a protected characteristic (in this case, sex) in how it treats individuals (in this case, decides whether or not to admit applicants to graduate schools). An organization's policies or ways of doing business have disparate impact when, even if they don't directly involve protected characteristics, their effect is to "impact" people in different categories differently.
    a. (5) Do any of your analyses suggest Cal was engaging in disparate treatment by sex?
    b. (5) Do any of your analyses suggest Cal's admissions policies had disparate impact by sex?

6. *Timing* (1) How long, roughly, did you spend on this assignment?

Presentation rubric (10): The text is laid out cleanly, with clear divisions between problems and sub-problems. The writing itself is well-organized, free of grammatical and other mechanical errors, and easy to follow. Plots are carefully labeled, with informative and legible titles, axis labels, and (if called for) sub-titles and legends; they are placed near the text of the corresponding problem. All quantitative and mathematical claims are supported by appropriate derivations, included in the text, or calculations in code. Numerical results are reported to appropriate precision.

References

Freeman, Jo. 1991. "How 'Sex' Got into Title VII: Persistent Opportunism as a Maker of Public Policy." *Law and Inequality: A Journal of Theory and Practice* 9:163–84. https://www.jofreeman.com/lawandpolicy/titlevii.htm.