



# The Hidden Bias: Simpson's Paradox and 1973 UC Berkeley Gender Discrimination Case

By Dora/Xiyuan Lin

Math 1700 Ideas in Mathematics

## *UC Berkeley Lawsuit, 1973*

In 1973, the University of California, Berkeley (UC Berkeley) faced legal threats and was accused of gender discrimination in their graduate school admissions program. At first glance, the numbers appeared against their favor: 44% of the 8,442 male applicants were admitted that fall, whereas only 35% of the 4,321 female applicants were admitted (Schneps and Colmez, 2013). This seemed to demonstrate systemic bias and gender discrimination in higher education, restricting women's opportunities at one of the most prestigious schools in the country.

	All		Men		Women	
	Applicants	Admitted	Applicants	Admitted	Applicants	Admitted
<b>Total</b>	12,763	41%	8,442	44%	4,321	35%

*Figure 1: General graduate student admissions data from the 1973 Berkeley Discrimination Case (Wikipedia, 2019).*

In response, UC Berkeley launched an investigation into the data of applicants and accepted students within specific departments. When investigators broke down the admissions data by each graduate studies department, they discovered that most departments showed no unexpected gender gap in admitted students, with some departments even showing a preference for female candidates (Schneps and Colmez, 2013). However, as seen in Figure 1, more women overall were rejected from the graduate departments.

Department	All		Men		Women	
	Applicants	Admitted	Applicants	Admitted	Applicants	Admitted
<b>A</b>	933	64%	<b>825</b>	62%	108	<b>82%</b>
<b>B</b>	585	63%	<b>560</b>	63%	25	<b>68%</b>
<b>C</b>	918	35%	325	<b>37%</b>	<b>593</b>	34%
<b>D</b>	792	34%	417	33%	375	<b>35%</b>
<b>E</b>	584	25%	191	<b>28%</b>	<b>393</b>	24%
<b>F</b>	714	6%	<b>373</b>	6%	341	<b>7%</b>
<b>Total</b>	<b>4526</b>	<b>39%</b>	<b>2691</b>	<b>45%</b>	<b>1835</b>	<b>30%</b>

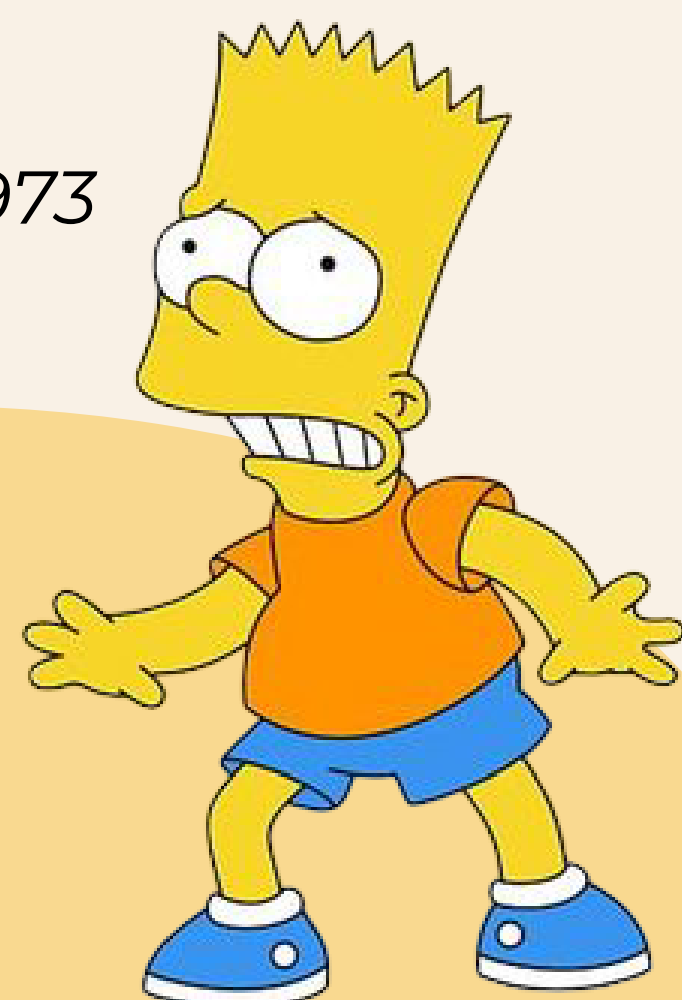
Legend:

- greater percentage of successful applicants than the other gender
- greater number of applicants than the other gender

**bold** - the two 'most applied for' departments for each gender

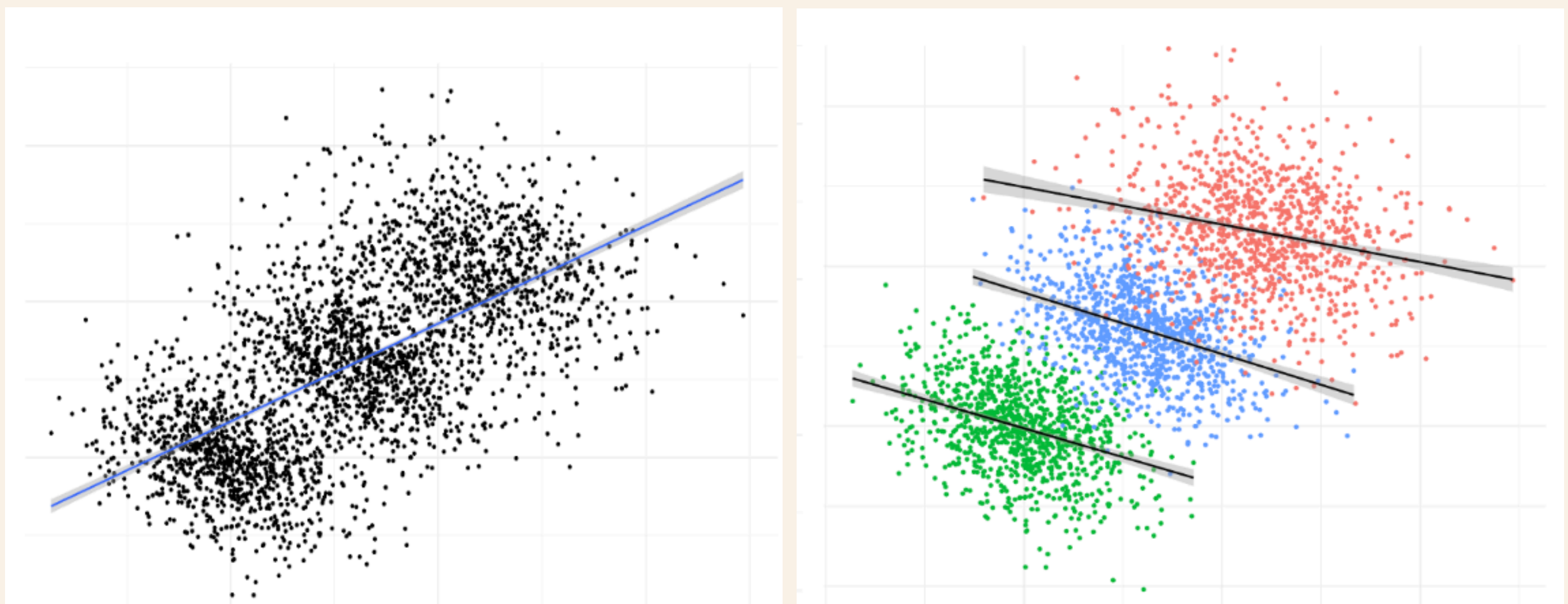
Figure 2: Admissions data from each of the six largest departments at the University of California, Berkeley, in 1973 (Wikipedia, 2019).

How is this mathematically possible?



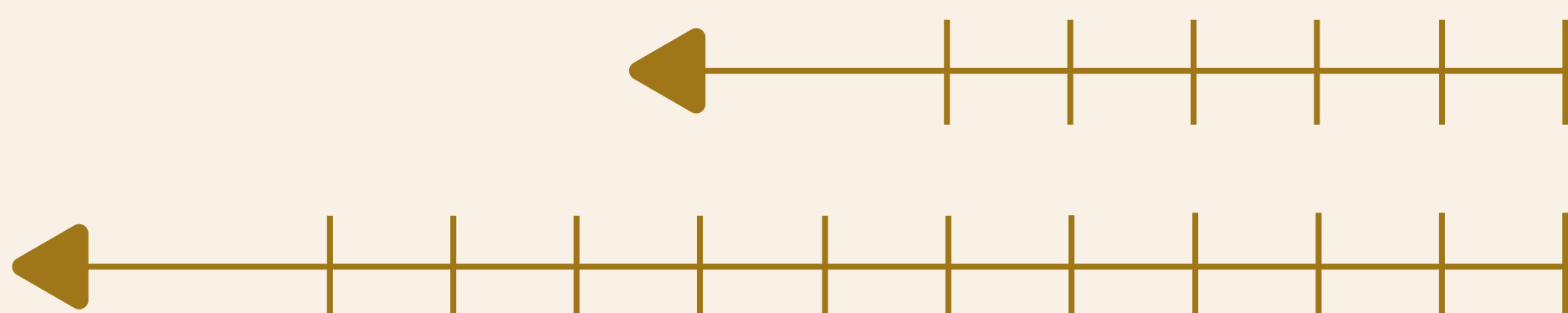
## Simpson's Paradox

This case is a paragon of Simpson's paradox, a phenomenon in which sub-populations within a large data set are reversed or disappear when combined with other groups (Hershbein, 2015). Named after British statistician Edward Simpson, Simpson's paradox is used to illustrate how statistics can be misinterpreted and is particularly common when significant differences occur between subgroups of a larger data set (Mida.so).



*Figure 3: An example of Simpson's paradox in which the pooled data demonstrate a positive correlation even though individual regions reflect a negative correlation (Laken, 2017).*

Often encountered in social science and medical science statistics, the paradox largely stems from confounding variables, or outside influences that affect both the independent and dependent variables in the experiment (Wikipedia, 2019).

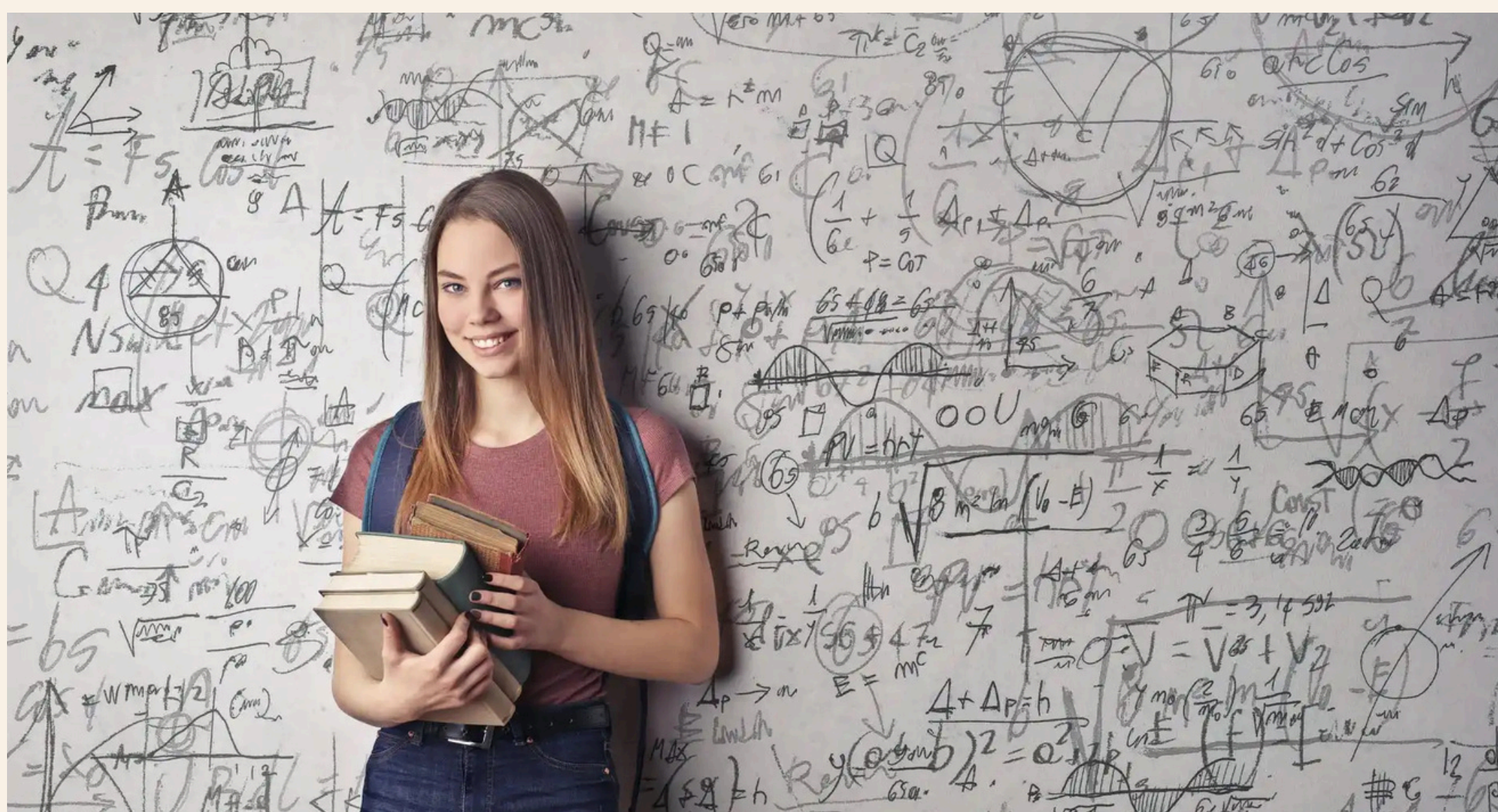


## *Simpson's Paradox at UC Berkeley*

In the UC Berkeley case, investigators found that the discrepancy stemmed from a larger proportion of female candidates applying to departments with lower acceptance rates.

Presumably, the bias lies in the factors that drew women to different fields of study than men, not directly admissions practices

(Schneps and Colmez, 2013). Here, the confounding variable is the department: under equitable admissions criteria, the more selective departments rejected more applicants of both genders. The aggregated data skewed in favor of male applicants as they were more likely than women to apply to natural science departments, which had higher acceptance rates than social sciences departments, which attracted more female applicants (Reinhart, 2016). Ultimately, despite overall data showing the discrepancy between acceptance rates (44% for men and 35% for women, as shown in Figure 1) of students, the admissions data clusters generated showed **no discrimination in the UC Berkeley case.**



## Mitigating Simpson's Paradox

It is critical to recognize the true issues and confounding variables behind cases of Simpson's paradox. While it is difficult to fully eliminate the paradox, below are some strategies for avoiding the paradox:

### Stratified Sampling

Stratified sampling involves dividing a population into groups (strata) based on shared characteristics, such as age, gender, or income, and then randomly sampling from each group to ensure fair representation of the whole population, thus allowing researchers to avoid misleading results in the overall trend (Mida.so and TGM Research).

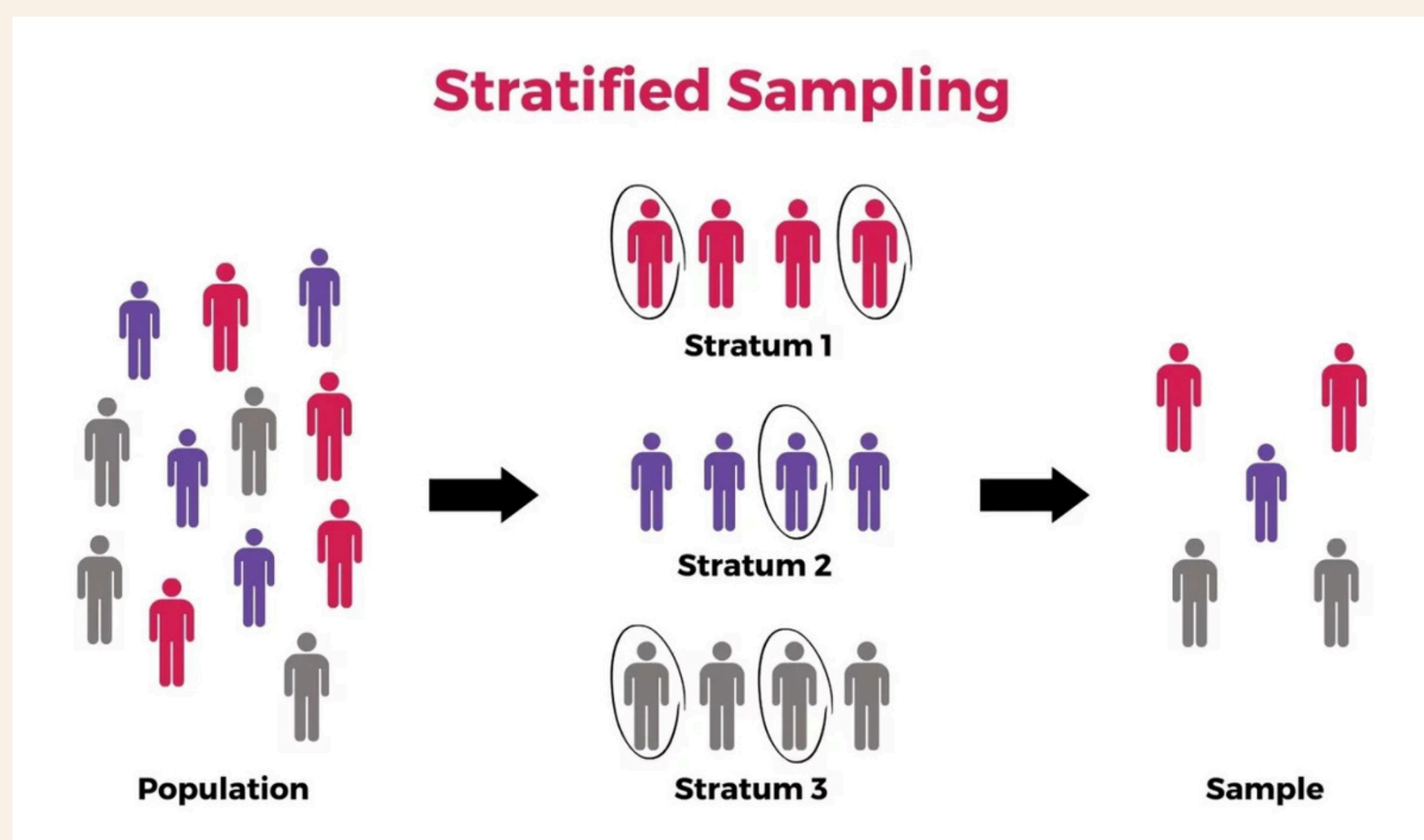


Figure 4: Visualization of the stratified sampling process (TGM Research).

### Randomization in Experimental Design

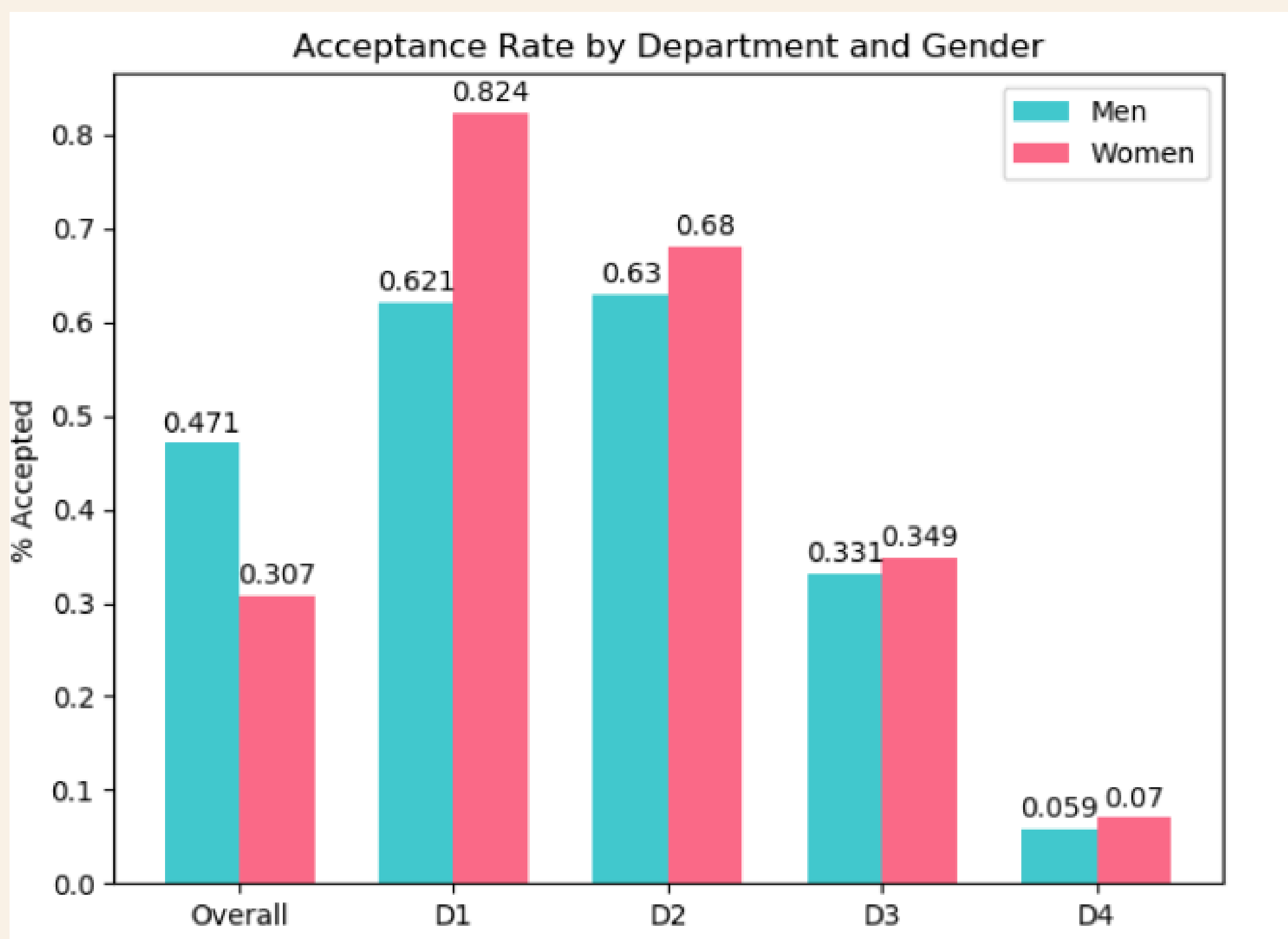
Randomization balances confounding variables and ensures that the observed effects can be attributed to the actual correlation between the dependent and independent variables (Mida.so).

## Multiple Regression Analysis

Regression analysis is a statistical method that explores relationships between a dependent variable and one or more independent variables. It can be utilized to assess the strength of the relationship between variables and control the confounding variables (Mida.so).

## Data Visualization

Through scatter plots, box plots, bar graphs, and other mechanisms of visualizing data, researchers can visually see reversals or differences in sub-groups and complete data sets (Mida.so).



*Figure 5: Bar graph of admissions acceptance rate by each department in the UC Berkeley case (Maicke, 2021).*

As shown above, there does not appear to be gender discrimination in each subgroup, yet the overall data visually highlights a reversal of this trend.

## Conclusion

Having always been interested in expanding access to higher education as a first-generation college student, I was immediately drawn to how math was applied in the legal realm and the implications behind the UC Berkeley case. Statistical errors like Simpson's paradox can be leveraged to push the political agenda forward through data manipulation and hinder progress for marginalized groups, as resources to address such issues could be diverted from the root causes of discrepancies (Setosa.io). For example, in the UC Berkeley case, the misinterpretation could have resulted in unjustified backlash for the university or even policy changes that divert resources dedicated to advancing women in STEM fields.



Meanwhile, the issue largely lies in social norms and the general push for women to pursue the social sciences and limited representation as well as opportunities for women to explore STEM fields compared to men, ultimately drawing female applicants to more competitive departments like English (Reinhart, 2016). As reflected in the admissions data, gender discrimination lies not in the admissions practices but in the stereotypes and gender expectations for paths of graduate study for women prior to their graduate school applications (Schneps and Colmez, 2013). As an aspiring lawyer with a passion for gender equity, the confounding variables in Simpson's paradox made me think more deeply about such structural forces beyond admissions practices against women and ways to unpick such phenomena. Ultimately, actively looking for reversals in trends when aggregating or disaggregating one's data helps us recognize Simpson's paradox and critically analyze the confounding variables to find solutions for issues spanning legal, policy, medical, and many other sectors.

## Sources

Hershbein, Brad. "When Average Isn't Good Enough: Simpson's Paradox in Education and Earnings." Brookings, July 29, 2015. <https://www.brookings.edu/articles/when-average-isnt-good-enough-simpsons-paradox-in-education-and-earnings/>.

Laken, Paul van der. "Simpson's Paradox: Two HR Examples with R Code." paulvanderlaken.com, September 27, 2017. <https://paulvanderlaken.com/2017/09/27/simpsons-paradox-two-hr-examples-with-r-code/>.

Maike Elisa. "Gender Bias in Admission Statistics? - Cantor's Paradise." Medium. Cantor's Paradise, June 28, 2021. <https://www.cantorsparadise.com/gender-bias-in-admission-statistics-eaabca650810>.

Mida.so. "What Is Simpson's Paradox in A/B Testing?" <https://www.mida.so/ab-testing-terms/simpsons-paradox>.

Reinhart, Alex. "Simpson's Paradox and Statistical Urban Legends: Gender Bias at Berkeley." www.refsmmat.com, May 8, 2016. <https://www.refsmmat.com/posts/2016-05-08-simpsons-paradox-berkeley.html>.

Schneps, Leila, and Coralie Colmez. Math on Trial : How Numbers Get Used and Abused in the Courtroom. New York: Basic Books, 2013.

Setosa.io. "Simpson's Paradox," n.d. <https://setosa.io/simpsons/>.

TGM Research. "An Overview of Stratified Sampling in Market Research," n.d. <https://tgmresearch.com/stratified-sampling.html>.

Wikipedia Contributors. "Simpson's Paradox." Wikipedia. Wikimedia Foundation, April 25, 2019. [https://en.wikipedia.org/wiki/Simpson%27s\\_paradox](https://en.wikipedia.org/wiki/Simpson%27s_paradox).

