Three-way tables

Wednesday, February 9, 2022 11:21 AM

Simpson's Paradox

ASSOCIATION IN THREE-WAY TABLES

SECTION 2.7

1

For some of my research in this area:

Bandyoapdhyay, P.S., Nelson, D., Greenwood, M. *et al.* The logic of Simpson's paradox. *Synthese* **181**, 185–208 (2011). https://doi.org/10.1007/s11229-010-9797-0

Example: Graduate Admissions Discrimination

In 1973, the University of California, Berkeley was charged with having discriminated against women in their graduate admissions process (Bickel et al., 1975).

	Admitted	Denied	Total
Men	1195	1486	2681
Women	559	1276	1835
Total	1754	2762	4516

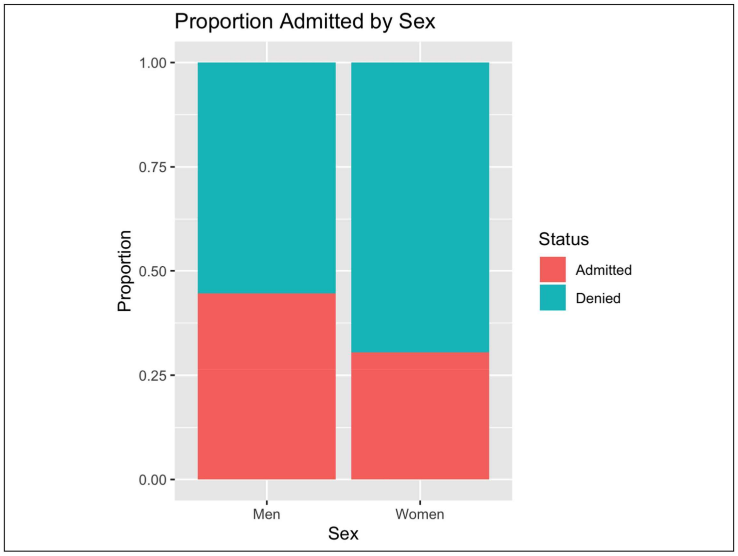
The proportion of men admitted is 1195/2681 = .446.

The proportion of women admitted is 559/1835 = .305.

.446/.305 = 1.46 → Men are ***46% more likely*** to be admitted than women!

Is this enough evidence to say UC Berkeley is discriminating against women?

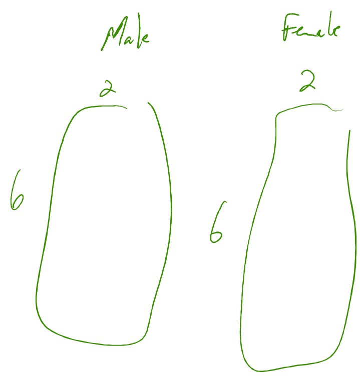
2



Graduate Admissions Discrimination?

What if we isolate the data by program, and see if a particular program (or programs) is responsible for the mistreatment of women applicants?

	Men			Women		
	Admitted	Denied	TOTAL	Admitted	Denied	TOTAL
Program A	511	314	825	89	19	108
Program B	352	208	560	17	8	25
Program C	120	205	325	202	391	593
Program D	137	270	407	132	243	375
Program E	53	138	191	95	298	393
Program F	22	351	373	24	317	341
TOTAL	1195	1486	2681	559	1276	1835



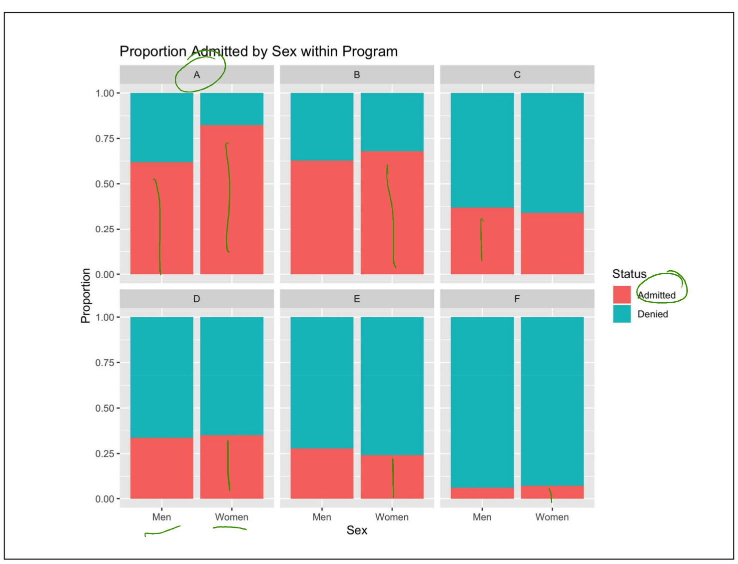
Program A

Admit Not

Male

Female

Program B



Graduate Admissions Discrimination?

	Overall Proportion Admitted	Proportion of Men Admitted	Proportion of Women Admitted	Proportion of Men Applicants	Proportion of Women Applicants
Program A	.64	.61	.82	.31	.06
Program B	.63	.62	.68	.21	.01
Program C	.35	.36	.34	.12	.32
Program D	.34	.33	.35	.15	.20
Program E	.25	.27	.24	.07	.21
Program F	.06	.05	.07	.14	.19

Each program highlighted in pink (B, D, F) accepted more women than men. So why was the overall percent of women accepted 15% less than that of men?

Example: Graduate Admissions Discrimination

So what happened?

Programs A and B had the highest acceptance rates (accepting 64% and 63% of all applicants, respectively). A total of 52% (31+21) of all men applicants applied to these programs, but only 7% (6+1) of all women applicants applied to these programs.

Programs E and F had the lowest acceptance rates (accepting 25% and only 6% of all applicants, respectively). A total of 40% (21+19) of all women applicants applied to these programs, but only 21% (7+14) of all men applicants applied to these programs.

Example: Graduate Admissions Discrimination

In summary, women tended to apply to the more competitive programs, and thus had an overall lower rate of acceptance than men, even though their acceptance rate was often higher than mens' in the individual programs.

The type of program was a **confounding variable**.

- A confounding variable is a variable that both ***affects the response variable*** and also is ***related to the explanatory variable***. How?
- The effect of a confounding variable on the response variable cannot be separated from the effect of the explanatory variable.

Simpson's Paradox: The Missing Third Variable

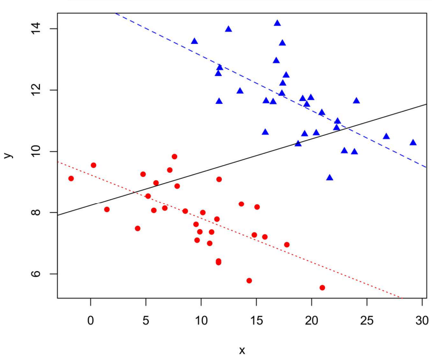
Dictionary definition of *paradox*: "a statement or proposition that, despite sound (or apparently sound) reasoning from acceptable premises, leads to a conclusion that seems senseless, logically unacceptable, or self-contradictory."

Definition: **Simpson's Paradox**

- The direction of a relationship or difference between two variables is *reversed* when studied within subgroups compared to the direction of the relationship within the whole group.

9

Example: Simpson's Paradox



Regression lines within each group both have negative slopes, so the inherent relationship between x and y is a negative linear association.

BUT regression line ignoring groups has a positive slope!

10

What did we just learn?

In an observational study, watch out for confounding variables (either known or unknown) that may explain or partially explain the observed relationship between *x* and *y*.

Within one level of the confounding variable, the effect of *x* on *y* may be entirely different from the effect of *x* on *y* when the confounding variable is ignored.

If possible, "control" for the confounding variable by assessing the relationship between *x* and *y* holding the confounding variable fixed.

11