CMPT 321
FALL 2017

Interpreting query results. Simpson's paradox

Lecture 04.04

by Marina Barsky

Example of Simpson's paradox

- Lisa and Bart are programmers, and they fix bugs for two weeks

	Week 1	Week 2	Both weeks
Lisa	60/100	1/10	**61/110**
Bart	**9/10**	**30/100**	39/110

Who is more productive: Lisa or Bart?

Explanation of Simpson's paradox

	Week 1	Week 2	Both weeks
Lisa	60/100	1/10	**61/110**
Bart	**9/10**	**30/100**	39/110

If we consider productivity for each week, we notice that the samples are of a very different size

The work should be judged from an equal sample size, which is achieved when the numbers of bugs each fixed are added together

Explanation of Simpson's paradox

	Week 1	Week 2	Both weeks
Lisa	60/100	1/10	**61/110**
Bart	**9/10**	**30/100**	39/110

Simple algebra of fractions shows that even though

$a1/A > b1/B$
$c1/C > d1/D$

$(a1+c1)/(A+C)$ can be smaller than $(b1+d1)/(B+D)$!

This may happen when the sample sizes A, B, C, D are skewed (Note, that we are not adding two inequalities, but adding the absolute numbers)

Simpson’s paradox in real life

- Two examples:
  - Gender bias
  - Medical treatment

Example 1: Berkeley gender bias case

Admitted to graduate school at University of California, Berkeley (1973)

	Admitted	Not admitted	Total
Men (**44**%)	3,714	4,727	8,441
Women (35%)	1,512	2,808	4,320

Conclusion: bias against women applicants?

Example 1: Berkeley gender bias case

Stratified by the departments

	Men		Women	
Dept.	Total	Admitted	Total	Admitted
A	825	62%	108	**82%**
B	560	63%	25	**68%**
C	325	**37%**	593	34%
D	417	33%	375	**35%**
E	191	**28%**	393	24%
F	272	6%	341	**7%**

In most departments,
the bias is towards women!

Example 2: Kidney stone treatment

Success rates of 2 treatments for kidney stones

Treatments	Success	Not success	Total
A*(78%)	273	77	350
B**(**83**%)	289	61	350

Conclusion: treatment B is better?

*Open procedures (surgery)

** Percutaneous nefrolithotomy (removal through a small opening)

Example 2: Kidney stone treatment

Stratified by stone sizes

	Treatment A	Treatment B
Small stones	**93% (81/87)**	87%(234/270)
Large stones	**73%(192/263)**	69%(55/80)
Both	78%(273/350)	**83% (289/350)**

Treatment A is better for both small and large stones,
But treatment B is more effective if we add both groups together

Implications in decision making

- Which data should we consult when choosing an action: the aggregated or stratified?

- Kidney stones: if you know the size of the stone, choose treatment A, if you don't – treatment B?

Implications in decision making

- Which data should we consult when choosing an action: the aggregated or stratified?

- The common sense: the treatment which is preferred under both conditions should be preferred when the condition is unknown

Implications in decision making

- Which data should we consult when choosing an action: the aggregated or stratified?

- If we always choose to use the stratified data, we can partition strata further, into groups by eye color, age, gender, race … These arbitrary hierarchies can produce opposite results, and lead to wrong choices

Implications in decision making

- Which data should we consult when choosing an action: the aggregated or stratified?

- Conclusion: data should be consulted with care and the understanding of the underlying story about the data is required for making correct decisions

From: Judea Pearl. *Causality: Models, Reasoning, and Inference*