LSU E. J. OURSO COLLEGE OF BUSINESS

Dr. Chun's Numb3rs & L∅gic

Simpson's Paradox

Young H. Chun, Ph.D.

Professor of Decision Science &
Cherie H. Flores Endowed Chair in MBA Studies

Simpson's Paradox

Large corporations and companies are divided into several divisions, subdivisions, departments, and so on. Except for high-level managerial positions, employment decisions usually take place at the departmental or divisional level. Analyzing aggregate employment data in such companies can give rise to a curious phenomenon known as *Simpson's paradox*.

An instructive and surprising example of Simpson's paradox occurred at the University of California at Berkeley in the 1970s. Examination of applicant data for a 1973 quarter revealed that the overall rate of admission for female applicants to the graduate school was substantially less than the rate of admission for male applicants.

Which departments at Berkeley were responsible for this imbalance?

Applicants to UC Berkeley		Number applied	Number admitted	Percent admitted
Combined	Male	110	91	82.7%
	Female	110	19	17.3%

* Hypothesis

$H_0: \pi_f \geq \pi_m$

$H_1: \pi_f < \pi_m$

* Test statistic:

$$z^* = \frac{(\bar{p}_1 - \bar{p}_2)}{\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = 9.7$$

* One-sided **p-value**:

$P[\ Z > 9.7\ ] =$ 0.00

Applicants to UC Berkeley		Number applied	Number admitted	Percent admitted
Department of Mathematics	Male	100	90	90%
	Female	10	9	90%
Department of English	Male	10	1	10%
	Female	100	10	10%
Combined	Male	110	91	82.7%
	Female	110	19	17.3%

*** Department of Mathematics**

- Acceptance rate for male = 90/100 = 90%
- Acceptance rate for female = 9/10 = 90%

*** Department of English**

- Acceptance rate for male = 1/10 = 10%
- Acceptance rate for female = 10/100 = 10%

*** No discrimination at all !**

Ask Marilyn, Parade Magazine (4/28/1996), p. 6.

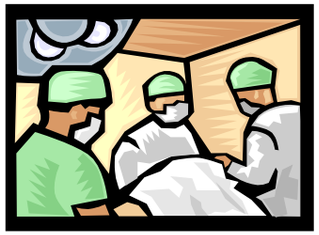
Applicants to a company		Number applied	Number admitted	Percent admitted
White-collar position	Male	200	30	15%
	Female	200	40	20%
Blue-collar position	Male	400	300	75%
	Female	100	85	85%
Combined	Male	600	330	55.00%
	Female	300	125	**41.67%**

- **White-collar position**
  - Discrimination against male applicants
- **Blue-collar position**
  - Discrimination against male applicants
- **Overall statistic?**
  - Discrimination against female applicants

Ask Marilyn, Parade Magazine (4/28/1996), p. 6.

Patients		Sample size	Number cured	Percentage cured
Study 1	A	100	36	36%
	B	1,000	450	45%
Study 2	A	1,000	600	60%
	B	100	65	65%
Total	**A**	1,100	636	**57.82%**
	B	1,100	515	46.82%

- **Study 1**
  - Treatment B is more effective than A.
- **Study 2**
  - Treatment B is more effective than A.
- **Overall statistic?**
  - Treatment A is more effective than B.



Batting Average of Two Baseball Players

A real-life example involves the batting average of two baseball players, Derek Jeter and David Justice, during the baseball years 1995 and 1996:

Player		Number of at-bats	Number of hits	Batting average
1995	Derek Jeter	48	12	.250
	David Justice	411	104	**.253**
1996	Derek Jeter	582	183	.314
	David Justice	140	45	**.321**
Total	Derek Jeter	630	195	**.310**
	David Justice	551	149	.270

In both 1995 and 1996, Justice had a higher batting average (in bold type) than Jeter did.

However, when the two baseball seasons are combined, Jeter shows a higher batting average than Justice.

The moral of the story is…

- "You can prove anything with statistics!"
- "There are lies, damn lies, and statistics!"

The aim of this course is to enhance your ability to draw insight from data and to assess more precisely the reliability of your inferences.

Because statistics are frequently "misused to give an air of legitimacy to weak claims," the course also aims to make you better consumers of statistical analyses conducted by others.

Movie Trivia

Oakland A's general manager Billy Beane's successful attempt
to assemble a baseball team on a lean budget
by employing computer-generated analysis to acquire new players.

Moneyball (2011)

Oakland A's general manager Billy Beane's successful attempt
to assemble a baseball team on a lean budget
by employing computer-generated analysis to acquire new players.