

Math 1710  
Class 16

V1u

Last Time

Contingency  
Tables - Ch. 3

Simpson's  
Paradox

# Math 1710 Class 16

More on Ch. 3,4, CLT  
Dr. Back

Oct. 2, 2009

# Sampling Distribution of a Statistic

Math 1710  
Class 16

V1u

Last Time

Contingency  
Tables - Ch. 3

Simpson's  
Paradox

Given  $n$  observations  $x_1, x_2, \dots, x_n$ , a statistic is any number determined by these observations.

# Sampling Distribution of a Statistic

Math 1710  
Class 16

V1u

Last Time

Contingency  
Tables - Ch. 3

Simpson's  
Paradox

Given  $n$  observations  $x_1, x_2, \dots, x_n$ , a statistic is any number determined by these observations.

For example the mean  $\bar{x}$  or standard deviation  $s$ .

# Sampling Distribution of a Statistic

Math 1710  
Class 16

V1u

Last Time

Contingency  
Tables - Ch. 3

Simpson's  
Paradox

## “Sampling Distribution of a Statistic”

A probability model describing the chance of different values of the statistic showing up.

# Sampling Distribution of a Statistic

Math 1710  
Class 16

V1u

Last Time

Contingency  
Tables - Ch. 3

Simpson's  
Paradox

“Sampling Distribution of a Statistic”

Two Types in the Lab:

# Sampling Distribution of a Statistic

Math 1710  
Class 16

V1u

Last Time

Contingency  
Tables - Ch. 3

Simpson's  
Paradox

## “Sampling Distribution of a Statistic”

Two Types in the Lab:

Samp. Dist. of  $\hat{p}$ ,

the proportion of infected ticks in a sample.

# Sampling Distribution of a Statistic

Math 1710  
Class 16

V1u

Last Time

Contingency  
Tables - Ch. 3

Simpson's  
Paradox

## “Sampling Distribution of a Statistic”

### Two Types in the Lab:

Samp. Dist. of  $\hat{p}$ ,

the proportion of infected ticks in a sample.

Samp. Dist. of  $\bar{x}$ ,

the average hospital costs in a sample.

# Graduation Plans

Math 1710  
Class 16

V1u

Last Time

Contingency  
Tables - Ch. 3

Simpson's  
Paradox

## What Graduates Did

	1959	1970	1980
<b>Continuing Education</b>	197	388	320
<b>Employed</b>	103	137	98
<b>In the Military</b>	20	18	18
<b>Other</b>	13	58	45

# Graduation Plans

Math 1710  
Class 16

V1u

Last Time

Contingency  
Tables - Ch. 3

Simpson's  
Paradox

## What Graduates Did

	1959	1970	1980
<b>Continuing Education</b>	197	388	320
<b>Employed</b>	103	137	98
<b>In the Military</b>	20	18	18
<b>Other</b>	13	58	45

This table is giving frequency data on all combinations of the two categorical variables:

**Year** (of graduation)

**WhatDid** (after graduation)

# Graduation Plans

Math 1710  
Class 16

V1u

Last Time

Contingency  
Tables - Ch. 3

Simpson's  
Paradox

## What Graduates Did

	1959	1970	1980
<b>Continuing Education</b>	197	388	320
<b>Employed</b>	103	137	98
<b>In the Military</b>	20	18	18
<b>Other</b>	13	58	45

This table is giving frequency data on all combinations of the two categorical variables:

**Year** (of graduation)

**WhatDid** (after graduation)

Many people use the term *contingency table* for this situation.

# Graduation Plans

Math 1710  
Class 16

V1u

Last Time

Contingency  
Tables - Ch. 3

Simpson's  
Paradox

## What Graduates Did

	1959	1970	1980
<b>Continuing Education</b>	197	388	320
<b>Employed</b>	103	137	98
<b>In the Military</b>	20	18	18
<b>Other</b>	13	58	45

First form the row and column sums:

# Graduation Plans

Math 1710  
Class 16

V1u

Last Time

Contingency  
Tables - Ch. 3

Simpson's  
Paradox

First form the row and column sums:

	1959	1970	1980	
<b>Continuing Education</b>	197	388	320	905
<b>Employed</b>	103	137	98	338
<b>In the Military</b>	20	18	18	56
<b>Other</b>	13	58	45	116
	333	601	481	1415

# Graduation Plans

Math 1710  
Class 16

V1u

Last Time

Contingency  
Tables - Ch. 3

Simpson's  
Paradox

	1959	1970	1980	
<b>Continuing Education</b>	197	388	320	905
<b>Employed</b>	103	137	98	338
<b>In the Military</b>	20	18	18	56
<b>Other</b>	13	58	45	116
	333	601	481	1415

What does the last column represent?

$$\begin{array}{r} 905 \\ 338 \\ 56 \\ 116 \\ \hline 1415 \end{array}$$

# Graduation Plans

Math 1710  
Class 16

V1u

Last Time

Contingency  
Tables - Ch. 3

Simpson's  
Paradox

	1959	1970	1980	
<b>Continuing Education</b>	197	388	320	905
<b>Employed</b>	103	137	98	338
<b>In the Military</b>	20	18	18	56
<b>Other</b>	13	58	45	116
	333	601	481	1415

What does the last column represent?

905

338

56

116

---

1415

It just gives frequency data on the single categorical variable  
*WhatDid.*

# Graduation Plans

Math 1710  
Class 16

V1u

Last Time

Contingency  
Tables - Ch. 3

Simpson's  
Paradox

	1959	1970	1980	
<b>Continuing Education</b>	197	388	320	905
<b>Employed</b>	103	137	98	338
<b>In the Military</b>	20	18	18	56
<b>Other</b>	13	58	45	116
	333	601	481	1415

Some people consider this last column to be the marginal distribution of *WhatDid*.

# Graduation Plans

Math 1710  
Class 16

V1u

Last Time

Contingency  
Tables - Ch. 3

Simpson's  
Paradox

	1959	1970	1980	
<b>Continuing Education</b>	197	388	320	905
<b>Employed</b>	103	137	98	338
<b>In the Military</b>	20	18	18	56
<b>Other</b>	13	58	45	116
	333	601	481	1415

Some people consider this last column to be the marginal distribution of *WhatDid*.

It is more informative to include the *relative* frequencies as well.

# Graduation Plans

Math 1710  
Class 16

V1u

Last Time

Contingency  
Tables - Ch. 3

Simpson's  
Paradox

	1959	1970	1980	
<b>Continuing Education</b>	197	388	320	905
<b>Employed</b>	103	137	98	338
<b>In the Military</b>	20	18	18	56
<b>Other</b>	13	58	45	116
	333	601	481	1415

Our textbook considers the marginal dist of WhatDid to be:

WhatDid	Freq.	Rel. Freq.
<b>Continuing Education</b>	905	64%
<b>Employed</b>	338	24%
<b>In the Military</b>	56	4%
<b>Other</b>	116	8%
	1415	100%

# Graduation Plans

Math 1710  
Class 16

V1u

Last Time

Contingency  
Tables - Ch. 3

Simpson's  
Paradox

	1959	1970	1980	
<b>Continuing Education</b>	197	388	320	905
<b>Employed</b>	103	137	98	338
<b>In the Military</b>	20	18	18	56
<b>Other</b>	13	58	45	116
	333	601	481	1415

Similarly the marginal distribution of Year is:

Year	Freq.	Rel. Freq.
1959	333	24%
1970	601	42%
1980	481	34%
	1415	100%

# Graduation Plans

Math 1710  
Class 16

V1u

Last Time

Contingency  
Tables - Ch. 3

Simpson's  
Paradox

	1959	1970	1980	
<b>Continuing Education</b>	197	388	320	905
<b>Employed</b>	103	137	98	338
<b>In the Military</b>	20	18	18	56
<b>Other</b>	13	58	45	116
	333	601	481	1415

(a) What % of these graduates joined the military?

# Graduation Plans

Math 1710  
Class 16

V1u

Last Time

Contingency  
Tables - Ch. 3

Simpson's  
Paradox

	1959	1970	1980	
<b>Continuing Education</b>	197	388	320	905
<b>Employed</b>	103	137	98	338
<b>In the Military</b>	20	18	18	56
<b>Other</b>	13	58	45	116
	333	601	481	1415

(a) What % of these graduates joined the military?

This is answered by an entry in the marginal distribution of WhatDid.

# Graduation Plans

Math 1710  
Class 16

V1u

Last Time

Contingency  
Tables - Ch. 3

Simpson's  
Paradox

	1959	1970	1980	
<b>Continuing Education</b>	197	388	320	905
<b>Employed</b>	103	137	98	338
<b>In the Military</b>	20	18	18	56
<b>Other</b>	13	58	45	116
	333	601	481	1415

(a) What % of these graduates joined the military?

This is answered by an entry in the marginal distribution of WhatDid.

$$\frac{56}{1415} = 4\%$$

# Graduation Plans

Math 1710  
Class 16

V1u

Last Time

Contingency  
Tables - Ch. 3

Simpson's  
Paradox

	1959	1970	1980	
<b>Continuing Education</b>	197	388	320	905
<b>Employed</b>	103	137	98	338
<b>In the Military</b>	20	18	18	56
<b>Other</b>	13	58	45	116
	333	601	481	1415

(b) What % of these students graduated in 1970?

# Graduation Plans

Math 1710  
Class 16

V1u

Last Time

Contingency  
Tables - Ch. 3

Simpson's  
Paradox

	1959	1970	1980	
<b>Continuing Education</b>	197	388	320	905
<b>Employed</b>	103	137	98	338
<b>In the Military</b>	20	18	18	56
<b>Other</b>	13	58	45	116
	333	601	481	1415

(b) What % of these students graduated in 1970?  
Based on the marginal distribution of Year.

# Graduation Plans

Math 1710  
Class 16

V1u

Last Time

Contingency  
Tables - Ch. 3

Simpson's  
Paradox

	1959	1970	1980	
<b>Continuing Education</b>	197	388	320	905
<b>Employed</b>	103	137	98	338
<b>In the Military</b>	20	18	18	56
<b>Other</b>	13	58	45	116
	333	601	481	1415

(b) What % of these students graduated in 1970?  
Based on the marginal distribution of Year.

$$\frac{601}{1415} = 42.5\%$$

# Graduation Plans

Math 1710  
Class 16

V1u

Last Time

Contingency  
Tables - Ch. 3

Simpson's  
Paradox

	1959	1970	1980	
<b>Continuing Education</b>	197	388	320	905
<b>Employed</b>	103	137	98	338
<b>In the Military</b>	20	18	18	56
<b>Other</b>	13	58	45	116
	333	601	481	1415

(c) What % of the 1970 graduates joined the military?

# Graduation Plans

Math 1710  
Class 16

V1u

Last Time

Contingency  
Tables - Ch. 3

Simpson's  
Paradox

	1959	1970	1980	
<b>Continuing Education</b>	197	388	320	905
<b>Employed</b>	103	137	98	338
<b>In the Military</b>	20	18	18	56
<b>Other</b>	13	58	45	116
	333	601	481	1415

(c) What % of the 1970 graduates joined the military?

The relevant concept here is called the **conditional distribution** of WhatDid given that Year is 1970.

# Graduation Plans

Math 1710  
Class 16

V1u

Last Time

Contingency  
Tables - Ch. 3

Simpson's  
Paradox

	1959	1970	1980	
<b>Continuing Education</b>	197	388	320	905
<b>Employed</b>	103	137	98	338
<b>In the Military</b>	20	18	18	56
<b>Other</b>	13	58	45	116
	333	601	481	1415

(c) What % of the 1970 graduates joined the military?

The relevant concept here is called the **conditional distribution** of WhatDid given that Year is 1970.

Some would just point to the 1970 column and say it is the conditional distribution of WhatDid among 1970 graduates.

# Graduation Plans

Math 1710  
Class 16

V1u

Last Time

Contingency  
Tables - Ch. 3

Simpson's  
Paradox

	1959	1970	1980	
<b>Continuing Education</b>	197	388	320	905
<b>Employed</b>	103	137	98	338
<b>In the Military</b>	20	18	18	56
<b>Other</b>	13	58	45	116
	333	601	481	1415

(c) What % of the 1970 graduates joined the military?

As before, relative frequencies help, so our textbook would report the conditional distribution of WhatDid among 1970 graduates as follows:

# Graduation Plans

Math 1710  
Class 16

V1u

Last Time

Contingency  
Tables - Ch. 3

Simpson's  
Paradox

	1959	1970	1980	
<b>Continuing Education</b>	197	388	320	905
<b>Employed</b>	103	137	98	338
<b>In the Military</b>	20	18	18	56
<b>Other</b>	13	58	45	116
	333	601	481	1415

(c) What % of the 1970 graduates joined the military?

Conditional Distribution of WhatDid for Year=1970

WhatDid	Freq.	Rel. Freq.
<b>Continuing Education</b>	388	65%
<b>Employed</b>	137	23%
<b>In the Military</b>	18	3%
<b>Other</b>	58	10%
	601	100%

# Graduation Plans

Math 1710  
Class 16

V1u

Last Time

Contingency  
Tables - Ch. 3

Simpson's  
Paradox

	1959	1970	1980	
<b>Continuing Education</b>	197	388	320	905
<b>Employed</b>	103	137	98	338
<b>In the Military</b>	20	18	18	56
<b>Other</b>	13	58	45	116
	333	601	481	1415

(d) Of the students in these surveys who joined the military, what percent graduated in 1970?

# Graduation Plans

Math 1710  
Class 16

V1u

Last Time

Contingency  
Tables - Ch. 3

Simpson's  
Paradox

	1959	1970	1980	
<b>Continuing Education</b>	197	388	320	905
<b>Employed</b>	103	137	98	338
<b>In the Military</b>	20	18	18	56
<b>Other</b>	13	58	45	116
	333	601	481	1415

(d) Of the students in these surveys who joined the military, what percent graduated in 1970?

Based on the conditional distribution of Year when WhatDid = Military. (the Military row)

# Graduation Plans

Math 1710  
Class 16

V1u

Last Time

Contingency  
Tables - Ch. 3

Simpson's  
Paradox

	1959	1970	1980	
<b>Continuing Education</b>	197	388	320	905
<b>Employed</b>	103	137	98	338
<b>In the Military</b>	20	18	18	56
<b>Other</b>	13	58	45	116
	333	601	481	1415

(d) Of the students in these surveys who joined the military, what percent graduated in 1970?

Based on the conditional distribution of Year when WhatDid = Military. (the Military row)

$$\frac{18}{56} = 32\%$$

# Graduation Plans

Math 1710  
Class 16

V1u

Last Time

Contingency  
Tables - Ch. 3

Simpson's  
Paradox

	1959	1970	1980	
<b>Continuing Education</b>	197	388	320	905
<b>Employed</b>	103	137	98	338
<b>In the Military</b>	20	18	18	56
<b>Other</b>	13	58	45	116
	333	601	481	1415

(e) What is the marginal distribution of postgraduation activities?

# Graduation Plans

Math 1710  
Class 16

V1u

Last Time

Contingency  
Tables - Ch. 3

Simpson's  
Paradox

	1959	1970	1980	
<b>Continuing Education</b>	197	388	320	905
<b>Employed</b>	103	137	98	338
<b>In the Military</b>	20	18	18	56
<b>Other</b>	13	58	45	116
	333	601	481	1415

(e) What is the marginal distribution of postgraduation activities?

Discussed when we began.

# Graduation Plans

Math 1710  
Class 16

V1u

Last Time

Contingency  
Tables - Ch. 3

Simpson's  
Paradox

	1959	1970	1980	
<b>Continuing Education</b>	197	388	320	905
<b>Employed</b>	103	137	98	338
<b>In the Military</b>	20	18	18	56
<b>Other</b>	13	58	45	116
	333	601	481	1415

(e) What is the marginal distribution of postgraduation activities?

WhatDid	Freq.	Rel. Freq.
<b>Continuing Education</b>	905	64%
<b>Employed</b>	338	24%
<b>In the Military</b>	56	4%
<b>Other</b>	116	8%
	1415	100%

# Graduation Plans

Math 1710  
Class 16

V1u

Last Time

Contingency  
Tables - Ch. 3

Simpson's  
Paradox

	1959	1970	1980	
<b>Continuing Education</b>	197	388	320	905
<b>Employed</b>	103	137	98	338
<b>In the Military</b>	20	18	18	56
<b>Other</b>	13	58	45	116
	333	601	481	1415

(f) What is the conditional distribution of postgraduation activities when Year=1959?

# Graduation Plans

Math 1710  
Class 16

V1u

Last Time

Contingency  
Tables - Ch. 3

Simpson's  
Paradox

	1959	1970	1980	
<b>Continuing Education</b>	197	388	320	905
<b>Employed</b>	103	137	98	338
<b>In the Military</b>	20	18	18	56
<b>Other</b>	13	58	45	116
	333	601	481	1415

(f) What is the conditional distribution of postgraduation activities when Year=1959?

WhatDid	Freq.	Rel. Freq.
<b>Continuing Education</b>	197	59%
<b>Employed</b>	103	31%
<b>In the Military</b>	20	6%
<b>Other</b>	13	4%
	601	100%

# Graduation Plans

Math 1710  
Class 16

V1u

Last Time

Contingency  
Tables - Ch. 3

Simpson's  
Paradox

	1959	1970	1980	
<b>Continuing Education</b>	197	388	320	905
<b>Employed</b>	103	137	98	338
<b>In the Military</b>	20	18	18	56
<b>Other</b>	13	58	45	116
	333	601	481	1415

(f) Any evidence that plans have changed over the 21 year period? Brief Description? Appropriate graph?

# Graduation Plans

Math 1710  
Class 16

V1u

Last Time

Contingency  
Tables - Ch. 3

Simpson's  
Paradox

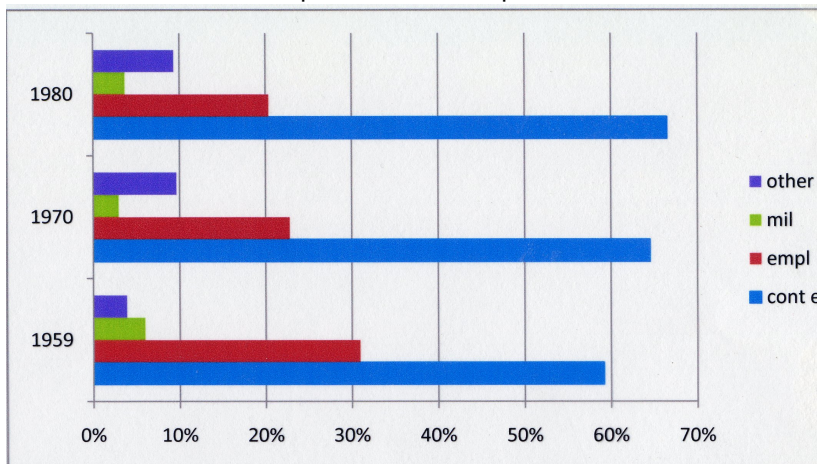
(f) Any evidence that plans have changed over the 21 year period? Brief Description? Appropriate graph?

<b>WhatDid</b>	1959	1970	1980
<b>Continuing Education</b>	59%	65%	67%
<b>Employed</b>	31%	23%	20%
<b>In the Military</b>	6%	3%	4%
<b>Other</b>	4%	10%	9%

# Graduation Plans

(f) Any evidence that plans have changed over the 21 year period? Brief Description? Appropriate graph?

Comparative Bar Graphs



Math 1710  
Class 16

V1u

Last Time

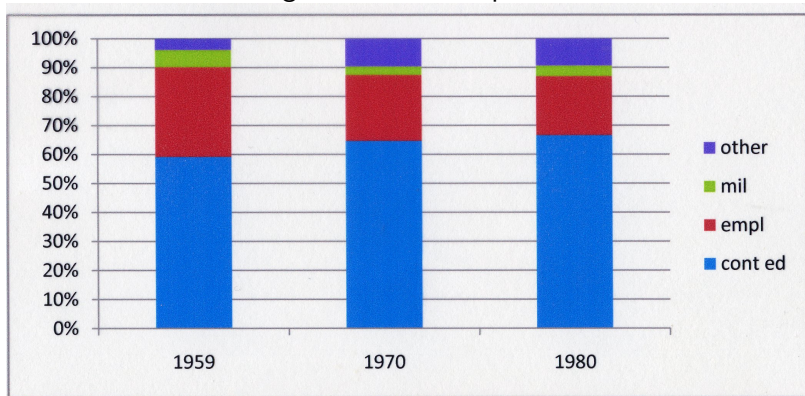
Contingency  
Tables - Ch. 3

Simpson's  
Paradox

# Graduation Plans

(f) Any evidence that plans have changed over the 21 year period? Brief Description? Appropriate graph?

## Segmented Bar Graphs



Math 1710  
Class 16

V1u

Last Time

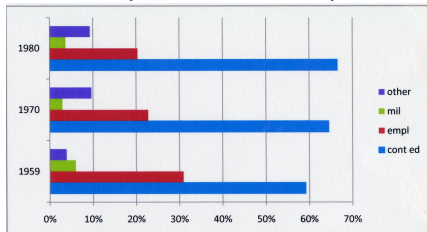
Contingency  
Tables - Ch. 3

Simpson's  
Paradox

# Graduation Plans

(f) Any evidence that plans have changed over the 21 year period? Brief Description? Appropriate graph?

## Comparative Bar Graphs



The employment percentage has gone down over the years.  
The education percentage has gone up.  
Throughout the period, education has been the first choice and employment the second.

# Two Car Rental Fleets

Math 1710  
Class 16

V1u

Last Time

Contingency  
Tables - Ch. 3

Simpson's  
Paradox

## Fleet H:

- $\frac{2}{3}$  small cars getting 3 gal/100 miles
- $\frac{1}{3}$  large cars getting 6 gal/100 miles
- Fleet Average: 4 gal/100 miles

## Fleet A:

- $\frac{1}{3}$  small cars getting 2.5 gal/100 miles
- $\frac{2}{3}$  large cars getting 5.5 gal/100 miles
- Fleet Average: 4.5 gal/100 miles

# Two Car Rental Fleets

Math 1710  
Class 16

V1u

Last Time

Contingency  
Tables - Ch. 3

Simpson's  
Paradox

## Fleet H:

- $2/3$  small cars getting 3 gal/100 miles
- $1/3$  large cars getting 6 gal/100 miles
- Fleet Average: 4 gal/100 miles

## Fleet A:

- $1/3$  small cars getting 2.5 gal/100 miles
- $2/3$  large cars getting 5.5 gal/100 miles
- Fleet Average: 4.5 gal/100 miles

## Who Gives You Better Mileage When You Rent?

- On A Small Car?
- On A Big Car?
- Overall?

# 9th Inning of World Series Game 7

Math 1710  
Class 16

V1u

Last Time

Contingency  
Tables - Ch. 3

Simpson's  
Paradox

# Simpson's Paradox

Math 1710  
Class 16

V1u

Last Time

Contingency  
Tables - Ch. 3

Simpson's  
Paradox

Don't compare combined statistics obtained by combining several disparate groups if the representation of the groups is varying.

# Simpson's Paradox

Math 1710  
Class 16

V1u

Last Time

Contingency  
Tables - Ch. 3

Simpson's  
Paradox

Don't compare combined statistics obtained by combining several disparate groups if the representation of the groups is varying.

This is an important issue in comparing success rates among hospitals.

# Berkeley 1973 Sexual Discrimination Case

Math 1710  
Class 16

V1u

Last Time

Contingency  
Tables - Ch. 3

Simpson's  
Paradox

Are women being discriminated against in grad admissions?

# Berkeley 1973 Sexual Discrimination Case

Math 1710  
Class 16

V1u

Last Time

Contingency  
Tables - Ch. 3

Simpson's  
Paradox

Are women being discriminated against in grad admissions?

	Overall	
	<b>Applicants</b>	<b>% admitted</b>
Men	8442	44%
Women	4321	35%

# Berkeley 1973 Sexual Discrimination Case

Math 1710  
Class 16

V1u

Last Time

Contingency  
Tables - Ch. 3

Simpson's  
Paradox

Are women being discriminated against in grad admissions?  
But women have a better chance in most departments ...

Major	Men Applicants	Men % admit	Women Applicants	Women % admit
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%

# Berkeley 1973 Sexual Discrimination Case

Math 1710  
Class 16

V1u

Last Time

Contingency  
Tables - Ch. 3

Simpson's  
Paradox

Are women being discriminated against in grad admissions?  
But women have a better chance in most departments ...

Major	Men Applicants	Men % admit	Women Applicants	Women % admit
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%

Women tended to apply to more competitive departments.

# Berkeley 1973 Sexual Discrimination Case

Math 1710  
Class 16

V1u

Last Time

Contingency  
Tables - Ch. 3

Simpson's  
Paradox

The above numbers are cited in Wikipedia.

An article about this is:

["Sex Bias in Graduate Admissions: Data from Berkeley"](#)

[P. J. Bickel, et. al., Science 7 February 1975](#)

# Berkeley 1973 Sexual Discrimination Case

Math 1710  
Class 16

V1u

Last Time

Contingency  
Tables - Ch. 3

Simpson's  
Paradox

## “Sex Bias in Graduate Admissions: Data from Berkeley” P. J. Bickel, et. al., Science 7 February 1975

“Examination of aggregate data on graduate admissions to the University of California, Berkeley, for fall 1973 shows a clear but misleading pattern of bias against female applicants. Examination of the disaggregated data reveals few decision-making units that show statistically significant departures from expected frequencies of female admissions, and about as many units appear to favor women as to favor men. . . . The graduate departments that are easier to enter tend to be those that require more mathematics in the undergraduate preparatory curriculum. . . . Women are shunted by their socialization and education toward fields of graduate study that are generally more crowded, less productive of completed degrees, and less well funded, and that frequently offer poorer professional employment prospects.”