431 Class 20

Thomas E. Love

2017-11-07

Today's Agenda

- Dealing with Larger two-way contingency tables (Notes 32)
    - Building a $J \times K$ Table
    - χ^2 Tests of Independence
- Dealing with an additional categorical variable (Notes 33)
    - The Cochran-Mantel-Haenszel Test
    - The Woolf test to check assumptions
    - Aggregation and Simpson's Paradox
- Review: Inference about Rates/Proportions (Notes 28-31)

Today's R Setup

```
library(magrittr); library(forcats);
library(vcd); library(tidyverse)

surd1 <- read.csv("data/surveyday1_2017.csv") %>% tbl_df

source("Love-boost.R")
```

Titanic Example

Consider the following data on survival of the Titanic.

- Among the females, 308 survived and 143 died.
- Of 851 males on board, 142 survived.

What conclusions can you draw about the relationship of sex to survival?

What are the comparisons we're making here?

Pr(survive | female) is the probability of survival, if you're female.

We want to compare Pr(survive | female) to Pr(survive | male).

How would we build such a comparison?

Starting The 2x2 Table

- Among the females, 308 survived and 143 died.
- Of 851 males on board, 142 survived.

Titanic	Survive	Die	Total
Female	308	143	??
Male	142	??	851
Total	??	??	??

The Complete 2x2 Table

Titanic	Survive	Die	Total
Female	**308**	**143**	*451*
Male	**142**	*709*	**851**
Total	*450*	*852*	*1,302*

Now, how do I get this into R?

Two-by-Two table

```
twobytwo(308, 143, 142, 709,
         "Female", "Male", "Survive", "Die")
```

```
2 by 2 table analysis:
------------------------------------------------------
Outcome   : Survive
Comparing : Female vs. Male

       Survive Die    P(Survive) 95% conf. interval
Female     308 143       0.6829    0.6385   0.7242
Male       142 709       0.1669    0.1433   0.1934

                                   95% conf. interval
             Relative Risk:  4.0928    3.4780   4.8162
         Sample Odds Ratio: 10.7541    8.2261  14.0588
Conditional MLE Odds Ratio: 10.7284    8.1546  14.1887
    Probability difference:  0.5161    0.4644   0.5636
```

Complete 2x2 Result: Conclusions?

```
2 by 2 table analysis:
------------------------------------------------------
Outcome   : Survive
Comparing : Female vs. Male

       Survive Die    P(Survive) 95% conf. interval
Female     308 143        0.6829    0.6385   0.7242
Male       142 709        0.1669    0.1433   0.1934

                                   95% conf. interval
             Relative Risk:  4.0928    3.4780   4.8162
         Sample Odds Ratio: 10.7541    8.2261  14.0588
Conditional MLE Odds Ratio: 10.7284    8.1546  14.1887
    Probability difference:  0.5161    0.4644   0.5636

             Exact P-value: 0
        Asymptotic P-value: 0
------------------------------------------------------
```

A 2x3 Table: Comparing Response to Active vs. Placebo

The table below, specifies the number of patients who show *complete*, *partial*, or *no response* after treatment with either **active** medication or a **placebo**.

Group	None	Partial	Complete
Active	16	26	29
Placebo	24	26	18

Is there a statistically significant association here? That is to say, is there a statistically significant difference between the treatment groups in the distribution of responses?

Getting the Table into R

To answer this, we'll have to get the data from this contingency table into a matrix in R. Here's one approach. . .

```
T1 <- matrix(c(16,26,29,24,26,18),
             ncol=3, nrow=2, byrow=TRUE)
rownames(T1) <- c("Active", "Placebo")
colnames(T1) <- c("None", "Partial", "Complete")
T1
```

```
        None Partial Complete
Active    16      26       29
Placebo   24      26       18
```

Getting the Chi-Square Test Results

H_0: rows and columns are independent vs. H_A: rows and columns are associated

```
chisq.test(T1)
```

```

	Pearson's Chi-squared test

data:  T1
X-squared = 4.1116, df = 2, p-value = 0.128
```

Chi-Square Assumptions

We assume that the expected frequency, under the null hypothesized model of independence, will be **at least 5** in each cell. If that is not the case, then the χ^2 test is likely to give unreliable results.

How do we calculate expected frequencies for a cell?

$$\text{Expected Frequency} = \frac{\text{Row total} \times \text{Column total}}{\text{Grand Total}}$$

This assumes that the independence model holds - that the probability of being in a particular column is exactly the same regardless of what row we're looking at.

Calculating Expected Frequencies for the 2x3 Table

```r
addmargins(T1)
```

```
        None Partial Complete Sum
Active    16      26       29  71
Placebo   24      26       18  68
Sum       40      52       47 139
```

- What is the expected frequency for the (Active, None) cell?

Calculating Expected Frequencies for the 2x3 Table

```r
addmargins(T1)
```

```
        None Partial Complete Sum
Active    16      26       29  71
Placebo   24      26       18  68
Sum       40      52       47 139
```

- What is the expected frequency for the (Active, None) cell?
- For (Active, None), row total = 71, column total = 40, grand total = 139, so expected frequency is

$$\frac{71 \times 40}{139} = 20.43$$

Expected Frequencies for the whole 2x3 table

—	None	Partial	Complete
Active	20.4	26.6	24.0
Placebo	19.6	25.4	23.0

- All of these values exceed 5, so our χ^2 results should be reasonable.
- We could have run a **Fisher's exact test**, too...

Fisher's Exact Test Results

H_0: rows and columns are independent vs. H_A: rows and columns are associated

```
fisher.test(T1)
```

```
	Fisher's Exact Test for Count Data

data:  T1
p-value = 0.1346
alternative hypothesis: two.sided
```

Working with Survey Data

431 Day 1 Survey (1 = Strongly Disagree, 5 = Strongly Agree)

9. I prefer to learn from lectures than to learn from activities.
10. I prefer to work on projects alone than in a team.

```
sur1 <- surd1 %>%
  select(student, english, lecture, alone) %>%
  mutate(lecture = factor(lecture),
         alone = factor(alone))

sur1
```

```
# A tibble: 203 x 4
   student english lecture  alone
     <int>  <fctr>  <fctr> <fctr>
 1  201701       y       2      2
 2  201702       y       3      4
 3  201703       y       3      1
```

5x5 Contingency Table

```r
sur1 %$% table(lecture, alone) %>% addmargins
```

```
       alone
lecture   1   2   3   4   5 Sum
    1     4   5   4   1   0  14
    2     6  14  21  10   2  53
    3     7  22  26  26   2  83
    4     1   7  14  14   4  40
    5     2   1   6   0   3  12
    Sum  20  49  71  51  11 202
```

Chi-Square testing

H_0: No association of rows and columns

H_A: Rows and columns are associated

```
sur1 %$% table(lecture, alone) %>% chisq.test
```

```
Warning in chisq.test(.): Chi-squared approximation may
be incorrect
```

```
    Pearson's Chi-squared test

data:  .
X-squared = 33.409, df = 16, p-value = 0.00652
```

Collapse Some Categories with `fct_recode` in `forcats`

9. I prefer to learn from lectures than to learn from activities. (1 = SD, 5 = SA)
10. I prefer to work on projects alone than in a team. (1 = SD, 5 = SA)

```
levels(sur1$lecture); levels(sur1$alone)
```

```
[1] "1" "2" "3" "4" "5"
```

```
[1] "1" "2" "3" "4" "5"
```

Collapse Categories of Lecture and Alone

```
sur2 <- sur1 %>%
  mutate(lec2 = fct_recode(lecture,
                         "Activities" = "1",
                         "Activities" = "2",
                         "Neutral" = "3",
                         "Lectures" = "4",
                         "Lectures" = "5"),
       alone2 = fct_recode(alone,
                          "Team" = "1",
                          "Team" = "2",
                          "Neutral" = "3",
                          "Alone" = "4",
                          "Alone" = "5"))
```

Result of Collapsing Categories

```
sur2 %$% table(lec2, alone2) %>% addmargins
```

```
              alone2
lec2           Team Neutral Alone Sum
  Activities     29      25    13  67
  Neutral        29      26    28  83
  Lectures       11      20    21  52
  Sum            69      71    62 202
```

Collapsed Contingency Table's Chi-Square test

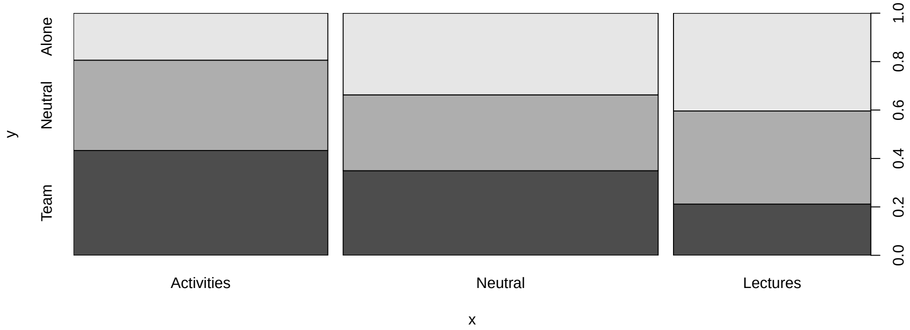
```r
sur2 %$% table(lec2, alone2) %>% chisq.test
```

```
	Pearson's Chi-squared test

data:  .
X-squared = 9.4435, df = 4, p-value = 0.05092
```

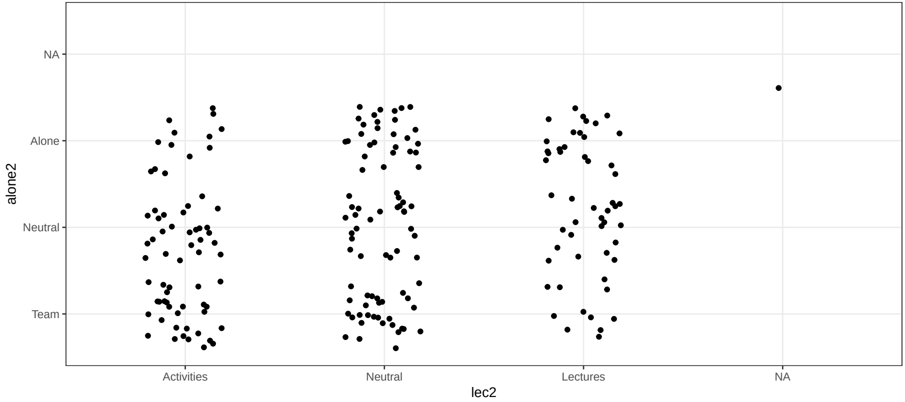
Default plot for Categorical Data (Mosaic plot)

```
plot(sur2$lec2, sur2$alone2)
```



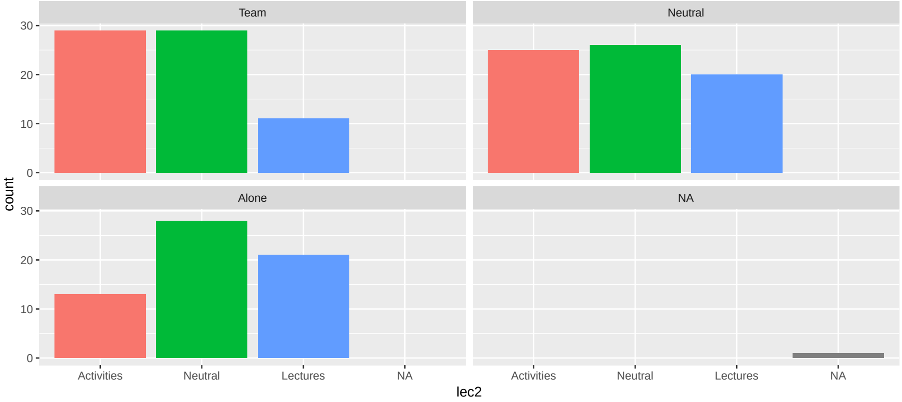
Plotting Categorical Data (Jitter)

```
ggplot(sur2, aes(x = lec2, y = alone2)) +
  geom_jitter(width = 0.2) + theme_bw()
```



Plotting Categorical Data (Bars)

```
ggplot(sur2, aes(x = lec2, fill = lec2)) +
  geom_bar() + guides(fill = FALSE) +
  facet_wrap(~ alone2)
```



Airplane Etiquette Example

https://fivethirtyeight.com/features/airplane-etiquette-recline-seat/

FiveThirtyEight asked 1040 people (not all responded) several questions, including:

1. `In general, is it rude to knowingly bring unruly children on a plane?`
2. `Is it rude to recline your seat on a plane?`
3. `Do you ever recline your seat when you fly?`
4. `Do you have any children under 18?`

The Data

The `flying` data set within the `fivethirtyeight` package provides the data here.

```
fly <- fivethirtyeight::flying %>%
  select(id = respondent_id, recline_frequency,
         recline_rude, unruly_child,
         have_kids = children_under_18) %>%
  mutate(have_kids = factor(have_kids)) %>%
  filter(complete.cases(.))
```

The `fly` data

```r
summary(select(fly, unruly_child, have_kids,
               recline_rude, recline_frequency))
```

```
  unruly_child have_kids     recline_rude
 No      :146   FALSE:657   No      :498
 Somewhat:348   TRUE :188   Somewhat:279
 Very    :351               Very    : 68

          recline_frequency
 Never              :166
 Once in a while    :254
 About half the time:116
 Usually            :175
 Always             :134
```

Airplane Etiquette Exercises for Practice

1. Estimate a 90% confidence interval for the proportion of people answering either "Somewhat" or "Very" to the question of whether it is rude to knowingly bring an unruly child on a plane. What is the margin of error?
2. Does the proportion of people who feel it is "Somewhat" or "Very" rude to knowingly bring an unruly child on a plane show a significant association with whether or not they themselves have children under 18 years of age?
3. Given the actual data, what can you conclude about the true proportion of people who feel it is rude to recline your seat on a plane?
4. Is there a significant association between how often you recline (`recline_frequency`) and your feelings about how rude (`recline_rude`) it is to do so?

Airplane Etiquette Exercises for Practice

Suppose we wish to estimate the power a study wil have to estimate the difference in proportion of people who feel that waking someone up to go for a walk is very or somewhat rude, comparing taller people to shorter people. Suppose we propose a new study, where we will collect data from 1200 tall and 1200 short people, and we look to declare as important any observed difference where one group is at 73% or more, while the other is at 70% or less.

5. Using a 10% significance level, what power will we have?
6. To obtain at least 80% power, how big a sample would we need?

One Type of Three-Way Contingency Table

We'll talk about three-way and larger contingency tables more in 432, but for now, let us focus on the situation where a 2x2 table is repeated over multiple strata (categories in a third variable.)

Duggal et al (2010) did a meta-analysis[1] of 5 placebo-controlled studies (AFREGS, ARBITER2, CLAS1, FATS and HATS) of niacin and heart disease, where the primary outcome was the need to do a coronary artery revascularization procedure.

- My Source: http://www.biostathandbook.com/cmh.html

[1] Duggal JK et al. 2010. Effect of niacin therapy on cardiovascular outcomes in patients with coronary artery disease. J Cardiovasc Pharmacology & Therapeutics 15: 158-166.

Niacin Meta-Analysis Example

For example, the FATS study had these results:

FATS	Revascularization	No Revasc.
Niacin	2	46
Placebo	11	41

- Pr(revascularization | Niacin) $= \frac{2}{2+46} = 0.042$
- Odds(revascularization | Niacin) $= \frac{2}{46} = 0.043$
- Pr(revascularization | Placebo) $= \frac{11}{11+41} = 0.212$
- Odds(revascularization | Placebo) $= \frac{11}{41} = 0.268$

and so the Odds Ratio $= \frac{2*41}{11*46} = 0.16$.

But, actually, we have data like this for each of five studies!

Building the Meta-Analysis Table

```
study <- c(rep("FATS", 4), rep("AFREGS", 4),
           rep("ARBITER2", 4), rep("HATS", 4),
           rep("CLAS1", 4))
treat <- c(rep(c("Niacin", "Niacin",
                 "Placebo", "Placebo"),5))
outcome <- c(rep(c("Revasc.", "No Rev."), 10))
counts <- c(2, 46, 11, 41, 4, 67, 12, 60, 1, 86,
            4, 76, 1, 37, 6, 32, 2, 92, 1, 93)
meta <- data.frame(study, treat, outcome, counts) %>% tbl_df
meta$treat <- fct_relevel(meta$treat, "Niacin")
meta$outcome <- fct_relevel(meta$outcome, "Revasc.")
meta.tab <- xtabs(counts ~ treat + outcome + study,
                  data = meta)
```

Five Studies in the Meta-Analysis

```r
ftable(meta.tab)
```

```
                study AFREGS ARBITER2 CLAS1 FATS HATS
treat   outcome
Niacin  Revasc.            4        1     2    2    1
        No Rev.           67       86    92   46   37
Placebo Revasc.           12        4     1   11    6
        No Rev.           60       76    93   41   32
```

The three variables we are studying are:

- `treat` (2 levels: Niacin/Placebo),
- `outcome` (2 levels: Revascularization or No Revascularization) across
- `study` (5 levels: AFREGS, ARBITER2, CLAS1, FATS, HATS)

Cochran-Mantel-Haenszel Test

The Cochran-Mantel-Haenszel test is designed to test whether the rate of revascularization is the same across the two levels of the treatment (i.e. Niacin or Placebo).

- We *could* do this by simply adding up the results across the five studies, but that wouldn't be wise, because the studies used different populations and looked for revascularization after different lengths of time.
- But we can account for the differences between studies to some extent by adjusting for study as a stratifying variable in a CMH test.
- The big assumption we'll have to make, though, is that the odds ratio for revascularization given Niacin instead of Placebo does not change across the studies. Is this reasonable in our case?

Looking at the Study-Specific Odds Ratios

We'll calculate the odds ratios, comparing revascularization odds with niacin vs. placebo, within each separate study.

Study	Rev N	Rev P	NoRev N	NoRev P	Odds Ratio
AFREGS	4	67	12	60	$\frac{4*60}{67*12} = 0.3$
ARBITER2	1	86	4	76	0.22
CLAS1	2	92	1	93	2.02
FATS	2	46	11	41	0.16
HATS	1	37	6	32	0.14

The table shows patient counts for the categories in each of the respective two-by-two tables (Rev N = Revascularization and Niacin, NoRev P = No Revascularization and Placebo, etc.)

Can we assume a Common Odds Ratio?

The Woolf test checks a key assumption for the Cochran-Mantel-Haenszel test. The Woolf test assesses the null hypothesis of a common odds ratio across the five studies.

```
woolf_test(meta.tab)
```

```
    Woolf-test on Homogeneity of Odds Ratios (no
    3-Way assoc.)

data:  meta.tab
X-squared = 3.4512, df = 4, p-value = 0.4853
```

Our conclusion from the Woolf test is that we are able to retain the null hypothesis of homogeneous odds ratios. So it's not crazy to fit a test that requires that all of the odds ratios be the same in the population.

Running the Cochran-Mantel-Haenszel test

So, we can use the Cochran-Mantel-Haenszel test to make inferences about the population odds ratio (for revascularization given niacin rather than placebo) accounting for the five studies. What can we conclude?

```r
mantelhaen.test(meta.tab, conf.level = .90)
```

```
    Mantel-Haenszel chi-squared test with
    continuity correction

data:  meta.tab
Mantel-Haenszel X-squared = 12.746, df = 1,
p-value = 0.0003568
alternative hypothesis: true common odds ratio is not equal to
90 percent confidence interval:
 0.1468942 0.4968686
sample estimates:
```

Complete CMH output

```r
mantelhaen.test(meta.tab, conf.level = .90)
```

```
Mantel-Haenszel chi-squared test with continuity correction

data:  meta.tab
Mantel-Haenszel
X-squared = 12.746, df = 1, p-value = 0.0003568

alt. hypothesis: true common odds ratio is not equal to 1

90 percent confidence interval: 0.1468942 0.4968686
sample estimates: common odds ratio 0.2701612
```

What can we conclude in this case?

The UC Berkeley Student Admissions Example

The UCBAdmissions data set contains aggregate data on applicants to graduate school at Berkeley for the six largest departments in 1973, classified by whether the applicant was admitted, and their sex.

```r
ftable(UCBAdmissions)
```

```
                Dept   A   B   C   D   E   F
Admit    Gender
Admitted Male        512 353 120 138  53  22
         Female       89  17 202 131  94  24
Rejected Male        313 207 205 279 138 351
         Female       19   8 391 244 299 317
```

Do the data show evidence of sex bias in admission practices?

Summarizing Department D

In Department D, we have

Department D	Males	Females
Admitted	138	131
Not Admitted	279	244
Applicants	417	375

Pr(Admitted if Male) $= \frac{138}{138+279} = 0.331$

Odds(Admitted if Male) $= \frac{138}{279} = 0.49$

Pr(Admitted if Female) $= \frac{131}{131+244} = 0.349$

Odds(Admitted if Female) $= \frac{131}{244} = 0.54$

Odds Ratio (Admit for Male vs Female) $= \frac{138*244}{131*279} = 0.92$

Can we use the Cochran-Mantel-Haenszel test?

Are the odds ratios similar across departments?

Department	A	B	C	D	E	F
Admitted Males	512	353	120	138	53	22
Male Applicants	825	560	325	417	191	373
Admitted Females	89	17	202	131	94	24
Female Applicants	108	25	593	375	393	341
Pr(Admit if Male)	0.62	0.63	0.37	0.33	0.28	0.06
Pr(Admit if Female)	0.82	0.68	0.34	0.35	0.24	0.07
Odds(Admit if Male)	1.64	1.71	0.59	0.49	0.38	0.06
Odds(Admit if Female)	4.68	2.12	0.52	0.54	0.31	0.08
Odds Ratio	0.35	0.8	1.13	0.92	1.22	0.83

Does it make sense to use a Cochran-Mantel-Haenszel test?

A Cochran-Mantel-Haenszel test describes a single combined odds ratio accounting for department. This assumes that the population odds ratio for admission by sex is identical for each of the six strata (departments).

- Does that seem reasonable?
- Or is there a three-way interaction here, where the odds ratios for admission by sex differ significantly across departments?

Department	A	B	C	D	E	F
Odds Ratio	0.35	0.8	1.13	0.92	1.22	0.83

How can we test this?

Woolf Test for Interaction in UCB Admissions

- H_0: There is no three-way interaction.
  - Odds ratios are homogenous, and we may proceed with the CMH test.)
- H_A: There is a meaningful three-way interaction.
  - CMH test is inappropriate because there are significantly different odds ratios across the departments.

```
woolf_test(UCBAdmissions)
```

```

    Woolf-test on Homogeneity of Odds Ratios (no
    3-Way assoc.)

data:  UCBAdmissions
X-squared = 17.902, df = 5, p-value = 0.003072
```

What's Going On Here? (1/3)

Department	A	B	C	D	E	F
Pr(Admit if Male)	0.62	0.63	0.37	0.33	0.28	0.06
Pr(Admit if Female)	0.82	0.68	0.34	0.35	0.24	0.07

What's Going On Here? (2/3)

Department	A	B	C	D	E	F
Pr(Admit if Male)	0.62	0.63	0.37	0.33	0.28	0.06
Pr(Admit if Female)	0.82	0.68	0.34	0.35	0.24	0.07
Pr (Admitted, regardless of sex)	0.64	0.63	0.35	0.34	0.25	0.06

What's Going On Here? (3/3)

Department	A	B	C	D	E	F
Pr (Admitted, regardless of sex)	0.64	0.63	0.35	0.34	0.25	0.06
% of Applicants who are Female	11.6	4.3	64.6	47.3	67.3	47.8

- The apparent association between admission and sex stems from differences in the tendency of males and female to apply to the individual departments.

What's Going On Here? (3/3)

Department	A	B	C	D	E	F
Pr (Admitted, regardless of sex)	0.64	0.63	0.35	0.34	0.25	0.06
% of Applicants who are Female	11.6	4.3	64.6	47.3	67.3	47.8

- The apparent association between admission and sex stems from differences in the tendency of males and female to apply to the individual departments.
- Females used to apply more to departments with lower admission rates.

What's Going On Here? (3/3)

Department	A	B	C	D	E	F
Pr (Admitted, regardless of sex)	0.64	0.63	0.35	0.34	0.25	0.06
% of Applicants who are Female	11.6	4.3	64.6	47.3	67.3	47.8

- The apparent association between admission and sex stems from differences in the tendency of males and female to apply to the individual departments.
- Females used to apply more to departments with lower admission rates.
- This is a common example related to what is called Simpson's Paradox.

Simpson's Paradox

Simpson's Paradox refers to a change in the direction of a relationship between two variables, depending on whether or not a third variable is controlled for.

- The Berkeley admissions result isn't quite a full reversal, but controlling for department has a big impact, so it's nearly a Simpson's paradox.

Kidney Stone Treatment Example

Suppose we compare the success rates of two treatments for kidney stones.

- 350 patients received Treatment A (all open surgical procedures) and 273 (78%) had a successful result.
- 350 patients received Treatment B (percutaneous nephrolithotomy - less invasive) and 289 (83%) had a successful result.

Which approach would you choose?

- Sources: https://en.wikipedia.org/wiki/Simpson%27s_paradox and Charig CR et al. (1986) PMID 3083922.

Kidney Stones with additional categorization of small/large stones

But what if we categorized the kidney stones as either small or large?

Disaggregated	Treatment A	Treatment B
Small Stones	81/87 (93%)	234/270 (87%)
Large Stones	192/263 (73%)	55/80 (69%)

Aggregated	Treatment A	Treatment B
All Subjects	273/350 (78%)	289/350 (83%)

Now, which treatment looks better?

Simpson's Paradox and US Wages

From 2000 to 2013, the median US wage rose 0.9%, adjusted for inflation.

But, over the same period, the median wage for

- high school dropouts (down 7.9%)
- high school graduates with no college education (down 4.7%)
- people with some college education (down 7.6%), and
- people with Bachelor's or higher degrees (down 1.2%)

have all decreased.

- Within **every** educational subgroup, the median wage is lower in 2013 than in 2000. How can this happen?

http://economix.blogs.nytimes.com/2013/05/01/can-every-group-be-worse-than-average-yes/

http://blog.revolutionanalytics.com/2013/07/a-great-example-of-simpsons-paradox.html

Coming Up

- On Statistical Significance and The p Value (Section 34)
- Type S and Type M errors (Section 35)

Deliverables

- Project Task C due **Wednesday** at noon.
- Assignment 5 due **Thursday** at noon.
- Quiz 2 will be yours on Thursday by 5 PM.
  - It's now due Tuesday Nov 14 at **8 AM**.
  - I won't ask about Cochran-Mantel-Haenszel or Woolf's test on the Quiz.