431 Class 20

thomaselove.github.io/431

2020-11-05

Today's Agenda

- Comparing Proportions using Independent Samples
  - Working with 2x2 Tables
  - Working with more general two-way tables
  - Cochran Conditions and Checking Assumptions
- Stratifying a 2x2 Table by a Third Categorical Variable
  - The Cochran-Mantel-Haenszel Test for a Three-Way Table
  - The Woolf test to check assumptions

Today's Setup and Data

```
knitr::opts_chunk$set(comment = NA)
options(dplyr.summarise.inform = FALSE)

library(Epi) # new today - for twoby2
library(vcd) # new today - for Woolf test
library(janitor)
library(knitr)
library(magrittr)
library(mosaic) # not usually something I load
library(broom)
library(tidyverse)

theme_set(theme_bw())

dm431 <- readRDS("data/dm431_2020.Rds")
source("data/Love-boost.R")
```

Comparing Two Proportions (A 2×2 table)

Using twobytwo from the Love-boost.R script

–	A1c < 8	A1c >= 8	*Total*
Never	22	12	*34*
Current	20	13	*33*
Total	*42*	*25*	*67*

Code we need is:

```
twobytwo(22, 12, 20, 13,  # note order of counts
      "Never", "Current", # names of the rows
      "A1c<8", "A1c>=8",  # names of the columns
      conf.level = 0.90)  # default is 95% confidence
```

Complete Output shown on the next slide

```
2 by 2 table analysis:
------------------------------------------------------
Outcome   : A1c<8
Comparing : Never vs. Current

        A1c<8 A1c>=8    P(A1c<8) 90% conf. interval
Never      22     12      0.6471    0.5040   0.7679
Current    20     13      0.6061    0.4613   0.7343

                                   90% conf. interval
             Relative Risk: 1.0676    0.7823   1.4571
         Sample Odds Ratio: 1.1917    0.5187   2.7377
Conditional MLE Odds Ratio: 1.1885    0.4625   3.0712
    Probability difference: 0.0410   -0.1486   0.2271

             Exact P-value: 0.8032
        Asymptotic P-value: 0.7288
------------------------------------------------------
```

Walking through the `twobytwo` Output

```
2 by 2 table analysis:
------------------------------------------------------
Outcome   : A1c<8
Comparing : Never vs. Current

        A1c<8 A1c>=8    P(A1c<8) 90% conf. interval
Never      22     12      0.6471    0.5040   0.7679
Current    20     13      0.6061    0.4613   0.7343
```

These are 90% confidence intervals for Pr(A1c$<$8) conditional on the exposure, and while we've seen five other methods for making this estimate, we use a sixth method here.

The computational details are shown on the next two slides.

90% CI for Pr(A1c < 8 | Never) (`twobytwo`)

```
        A1c<8 A1c>=8    P(A1c<8) 90% conf. interval
Never      22     12      0.6471    0.5040   0.7679
```

This result is computed using the Normal approximation for log(odds), and I'll show the entire calculation on the next slide.

Key Facts:

- odds = prob / (1 - prob) lets you convert from probability to odds
- prob = odds / (1 + odds) lets you convert from odds to probability
- standard error of the log(odds) formula is

$$se_{log(odds)} = \sqrt{\frac{1}{np(1-p)}}$$

- 90% confidence interval (two-sided) requires $Z_{\alpha/2} = Z_{0.05} = 1.645$

Calculation for 90% CI for x = 22, n = 34

```
n <- 22 + 12; prob <- 22/(22+12)
odds <- prob / (1 - prob)
logodds <- log(odds)
se_logodds <- sqrt(1 / (n * prob * (1 - prob)))
ci_logodds <- c(logodds - 1.645*se_logodds,
                logodds + 1.645*se_logodds)
  # that is the 90% CI on the log(odds) scale
  # so we exponentiate to get CI on odds scale
ci_odds <- exp(ci_logodds) # ci on odds scale
  # then convert odds to probability scale
ci_prob <- ci_odds / (1 + ci_odds) # ci on prob scale
ci_prob
```

```
[1] 0.5039485 0.7678975
```

Returning to the twobytwo output

```
2 by 2 table analysis:
------------------------------------------------------
Outcome   : A1c<8
Comparing : Never vs. Current

                             90% conf. interval
           Relative Risk: 1.0676    0.7823   1.4571
       Sample Odds Ratio: 1.1917    0.5187   2.7377
Conditional MLE Odds Ratio: 1.1885  0.4625   3.0712
    Probability difference: 0.0410 -0.1486   0.2271
```

We get confidence intervals for four different measures comparing A1c<8 rates for Never to Current, but we'll only use three in 431.

- Relative Risk
- Odds Ratio (we'll use the sample version - the cross-product version)
- Probability Difference

Relative Risk

```
Outcome    : A1c<8
Comparing : Never vs. Current

        A1c<8 A1c>=8    P(A1c<8) 90% conf. interval
Never      22      12      0.6471    0.5040   0.7679
Current    20      13      0.6061    0.4613   0.7343

                                   90% conf. interval
              Relative Risk: 1.0676    0.7823   1.4571
```

$$RR = \frac{0.6471}{0.6061} = 1.0676$$

- What does RR = 1 imply about the probabilities we are comparing?

Relative Risk

```
Outcome    : A1c<8
Comparing : Never vs. Current

        A1c<8 A1c>=8    P(A1c<8) 90% conf. interval
Never      22      12      0.6471    0.5040   0.7679
Current    20      13      0.6061    0.4613   0.7343

                                     90% conf. interval
              Relative Risk: 1.0676    0.7823   1.4571
```

$$RR = \frac{0.6471}{0.6061} = 1.0676$$

- What does RR = 1 imply about the probabilities we are comparing?
- How about RR > 1?

Relative Risk

```
Outcome   : A1c<8
Comparing : Never vs. Current

        A1c<8 A1c>=8    P(A1c<8) 90% conf. interval
Never      22     12      0.6471    0.5040   0.7679
Current    20     13      0.6061    0.4613   0.7343

                                    90% conf. interval
             Relative Risk: 1.0676    0.7823   1.4571
```

$$RR = \frac{0.6471}{0.6061} = 1.0676$$

- What does RR = 1 imply about the probabilities we are comparing?
- How about RR > 1?
- What about RR < 1?

Odds Ratio (Sample Odds Ratio)

```
Outcome    : A1c<8
Comparing : Never vs. Current

        A1c<8 A1c>=8    P(A1c<8) 90% conf. interval
Never      22     12      0.6471    0.5040   0.7679
Current    20     13      0.6061    0.4613   0.7343

                                     90% conf. interval
         Sample Odds Ratio: 1.1917    0.5187   2.7377
```

$$OR = \frac{22 \times 13}{12 \times 20} = 1.1917$$

- What does OR = 1 imply about the probabilities being compared?

Odds Ratio (Sample Odds Ratio)

```
Outcome   : A1c<8
Comparing : Never vs. Current

        A1c<8 A1c>=8    P(A1c<8) 90% conf. interval
Never      22     12      0.6471    0.5040   0.7679
Current    20     13      0.6061    0.4613   0.7343

                                   90% conf. interval
         Sample Odds Ratio: 1.1917    0.5187   2.7377
```

$$OR = \frac{22 \times 13}{12 \times 20} = 1.1917$$

- What does OR $=$ 1 imply about the probabilities being compared?
- How about OR $>$ 1?

Odds Ratio (Sample Odds Ratio)

```
Outcome   : A1c<8
Comparing : Never vs. Current

         A1c<8 A1c>=8    P(A1c<8) 90% conf. interval
Never       22     12      0.6471    0.5040   0.7679
Current     20     13      0.6061    0.4613   0.7343

                                       90% conf. interval
         Sample Odds Ratio: 1.1917    0.5187   2.7377
```

$$OR = \frac{22 \times 13}{12 \times 20} = 1.1917$$

- What does OR $=$ 1 imply about the probabilities being compared?
- How about OR $>$ 1?
- What about OR $<$ 1?

Probability Difference (also called Risk Difference)

```
Outcome   : A1c<8
Comparing : Never vs. Current

        A1c<8 A1c>=8    P(A1c<8) 90% conf. interval
Never      22     12      0.6471    0.5040   0.7679
Current    20     13      0.6061    0.4613   0.7343

                                    90% conf. interval
    Probability difference: 0.0410   -0.1486   0.2271
```

$$\Delta = 0.6471 - 0.6061 = 0.0410$$

- What will the probability difference be if the probabilities are the same?

Probability Difference (also called Risk Difference)

```
Outcome   : A1c<8
Comparing : Never vs. Current

        A1c<8 A1c>=8    P(A1c<8) 90% conf. interval
Never      22     12      0.6471    0.5040   0.7679
Current    20     13      0.6061    0.4613   0.7343

                                    90% conf. interval
    Probability difference: 0.0410   -0.1486   0.2271
```

$$\Delta = 0.6471 - 0.6061 = 0.0410$$

- What will the probability difference be if the probabilities are the same?
- What does a positive risk difference imply?

Probability Difference (also called Risk Difference)

```
Outcome   : A1c<8
Comparing : Never vs. Current

        A1c<8 A1c>=8    P(A1c<8) 90% conf. interval
Never      22     12      0.6471    0.5040   0.7679
Current    20     13      0.6061    0.4613   0.7343

                                       90% conf. interval
   Probability difference: 0.0410   -0.1486   0.2271
```

$$\Delta = 0.6471 - 0.6061 = 0.0410$$

- What will the probability difference be if the probabilities are the same?
- What does a positive risk difference imply?
- How about a negative risk difference?

Hypothesis Testing?

At the bottom of the `twobytwo` output, we have two p values...

```
Exact P-value: 0.8032        Asymptotic P-value: 0.7288
```

The `Exact P-value` comes from Fisher's exact test, and is technically exact only if we treat the row and column totals as being fixed. The `Asymptotic P-value` comes from a Pearson χ^2 test. These test:

- H_0: Pr(A1c $<$ 8 $|$ Never) $=$ Pr(A1c $<$ 8 $|$ Current) vs.
- H_A: Pr(A1c $<$ 8 $|$ Never) $\neq$ Pr(A1c $<$ 8 $|$ Current).

We usually state this as:

- H_0: rows and columns of the table are *independent* (Pr(A1c$<$8) is the same regardless of which row you're in) vs.
- H_A: the rows and columns of the table are *associated*.

Bayesian Augmentation in a 2x2 Table?

Original command:

```
twobytwo(22, 12, 20, 13,
      "Never", "Current",
      "A1c<8", "A1c>=8", conf.level = 0.90)
```

Bayesian augmentation approach: Add two successes and add two failures in each row…

```
twobytwo(22+2, 12+2, 20+2, 13+2,
      "Never", "Current",
      "A1c<8", "A1c>=8", conf.level = 0.90)
```

Output shown on next slide.

```
2 by 2 table analysis:
------------------------------------------------------
Outcome   : A1c<8
Comparing : Never vs. Current

        A1c<8 A1c>=8    P(A1c<8) 90% conf. interval
Never      24     14      0.6316    0.4965   0.7488
Current    22     15      0.5946    0.4582   0.7178

                                   90% conf. interval
             Relative Risk: 1.0622    0.7851   1.4371
         Sample Odds Ratio: 1.1688    0.5355   2.5513
Conditional MLE Odds Ratio: 1.1664    0.4837   2.8243
    Probability difference: 0.0370   -0.1437   0.2148

             Exact P-value: 0.8147
        Asymptotic P-value: 0.7424
------------------------------------------------------
```

Visit the extra slides / recording

Extra Example 1: Statin use in Medicaid vs. Uninsured within the `dm431` study.

Measuring Association using Categorical Variables with Chi-Squared Tests: Working with Two-Way Tables

A Two-Way Table with 4 rows and 3 columns

```
dm431 %>% tabyl(insurance, tobacco)
```

```
  insurance Current Former Never
 Commercial      35     60    69
   Medicaid      33     33    34
   Medicare      17     58    48
  Uninsured      11     13    20
```

Is tobacco use status associated with insurance type?

- Two factors here (insurance, tobacco) so we call it a two-way table
- This is a 4×3 table, with 4 rows and 3 columns.

Is tobacco use associated with insurance type?

H_0: Tobacco status is independent of insurance type

```
dm431 %>% tabyl(insurance, tobacco) %>%
  adorn_totals(where = "row") %>%
  adorn_percentages() %>% adorn_pct_formatting() %>%
  adorn_ns(position = "front") %>% adorn_title()
```

```
             tobacco
  insurance  Current       Former        Never
 Commercial 35 (21.3%)  60 (36.6%)  69 (42.1%)
   Medicaid 33 (33.0%)  33 (33.0%)  34 (34.0%)
   Medicare 17 (13.8%)  58 (47.2%)  48 (39.0%)
  Uninsured 11 (25.0%)  13 (29.5%)  20 (45.5%)
      Total 96 (22.3%) 164 (38.1%) 171 (39.7%)
```

Independence model: tobacco rates are 22.3%, 38.1%, 39.7% in each row

Independence Model for Insurance and Tobacco

Table shows observed counts + (expected under independence model)

- expected count = (row total) × (col total) / (grand total)
- for example, Medicaid/Current expected count is 100*96/431 = 22.3

	Current	Former	Never	Total
Commercial	35 (36.5)	60 (62.4)	69 (65.1)	164
Medicaid	33 (22.3)	33 (38.1)	34 (39.7)	100
Medicare	17 (27.4)	58 (46.8)	48 (48.8)	123
Uninsured	11 (9.8)	13 (16.7)	20 (17.5)	44
Total	96	164	171	431

Since all of these expected counts exceed 5, the Pearson χ^2 test should provide a reasonably accurate approximate p value for H_0: rows and columns are independent.

Chi-Square Assumptions

- We assume that the expected count, under the null hypothesized model of independence, will be **at least 5** (and ideally at least 10) in each cell.
- If that is not the case, then the χ^2 test is likely to give unreliable results.

The Cochran Conditions: R's warning approach

The *Cochran conditions* require us to have no cells with zero and at least 80% of the cells in our table with expected counts of 5 or higher. That's what R uses to warn you of trouble.

- Don't meet the standards? Consider collapsing categories.

dm431 association of tobacco with insurance

```r
tab43 <- dm431 %$% table(insurance, tobacco)
tab43
```

```
            tobacco
insurance     Current Former Never
  Commercial       35     60    69
  Medicaid         33     33    34
  Medicare         17     58    48
  Uninsured        11     13    20
```

```r
chisq.test(tab43)
```

```

    Pearson's Chi-squared test

data:  tab43
X-squared = 15.033, df = 6, p-value = 0.02
```

Can obtain observed + expected counts

```r
chisq.test(tab43)$expected
```

```
            tobacco
insurance      Current   Former    Never
  Commercial 36.529002 62.40371 65.06729
  Medicaid   22.273782 38.05104 39.67517
  Medicare   27.396752 46.80278 48.80046
  Uninsured   9.800464 16.74246 17.45708
```

```r
chisq.test(tab43)$observed
```

```
            tobacco
insurance    Current Former Never
  Commercial      35     60    69
  Medicaid        33     33    34
  Medicare        17     58    48
  Uninsured       11     13    20
```

Can obtain residuals

```
chisq.test(tab43)$residuals
```

```
             tobacco
insurance        Current     Former      Never
  Commercial -0.2529818 -0.3042827  0.4875409
  Medicaid    2.2727394 -0.8188378 -0.9009896
  Medicare   -1.9863151  1.6367193 -0.1145855
  Uninsured   0.3831686 -0.9146343  0.6086218
```

These are Pearson residuals - think of them as standardized, like Z scores.

- Values greater than 2 or less than -2 indicate especially poorly fit cells.
- Cells with positive residuals have higher observed counts than the independence model would predict.
- Cells with negative residuals have lower observed counts than independence predicts.

Augmented test using `mosaic::xchisq.test()`

(see full output on next slide)

```
xchisq.test(tobacco ~ insurance, data = dm431)
```

Note that flipping the rows and columns here changes the table, but not the conclusions of the χ^2 test.

```
    Pearson's Chi-squared test
X-squared = 15.033, df = 6, p-value = 0.02

   35        33        17        11
(36.53)   (22.27)   (27.40)   ( 9.80)
[0.064]   [5.165]   [3.945]   [0.147]
<-0.25>   < 2.27>   <-1.99>   < 0.38>
                                        key:
   60        33        58        13      observed
(62.40)   (38.05)   (46.80)   (16.74)   (expected)
[0.093]   [0.670]   [2.679]   [0.837]   [X-square contribution]
<-0.30>   <-0.82>   < 1.64>   <-0.91>   <Pearson residual>

   69        34        48        20
(65.07)   (39.68)   (48.80)   (17.46)
[0.238]   [0.812]   [0.013]   [0.370]
< 0.49>   <-0.90>   <-0.11>   < 0.61>
```

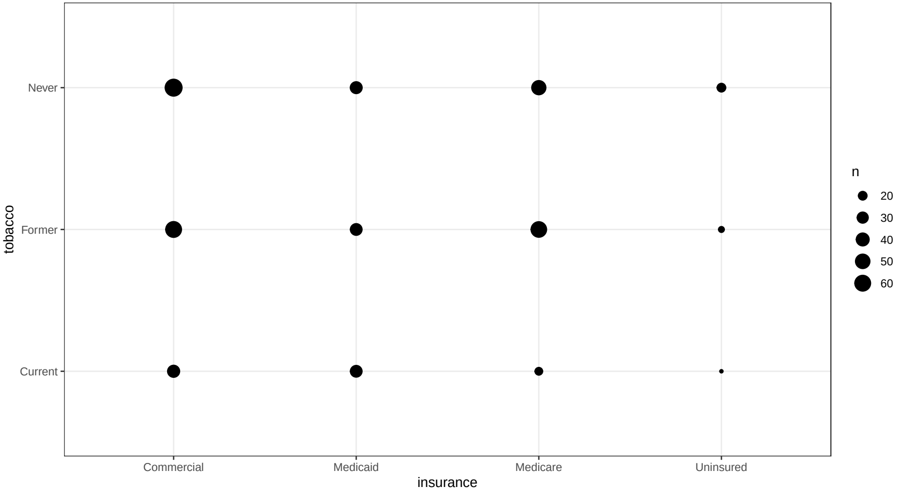
Visualizing the Association

We have cell counts for a 4×3 table here. How to visualize?

1. Use a tabyl (or table).
2. Consider the use of `geom_count`?
3. Consider a `mosaicplot`? The `mosaicplot` is a feature of the `graphics` package in base R, and has nothing to do with the `mosaic` package.
4. Consider an `assocplot`?

Using geom_count

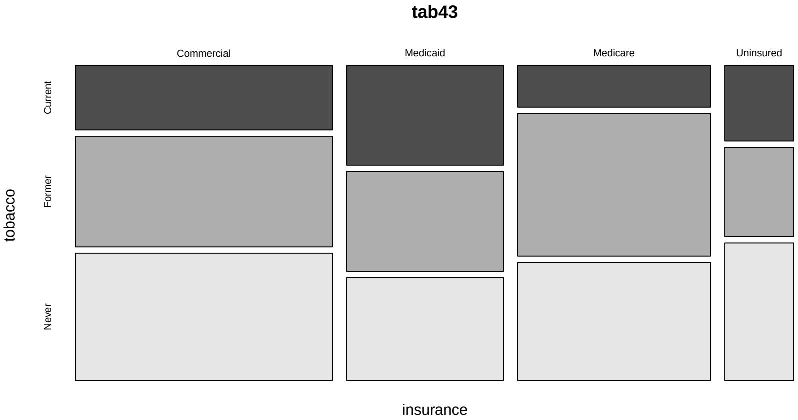
```
ggplot(dm431, aes(x = insurance, y = tobacco)) +
  geom_count()
```



Using `mosaicplot` to show what's in the table

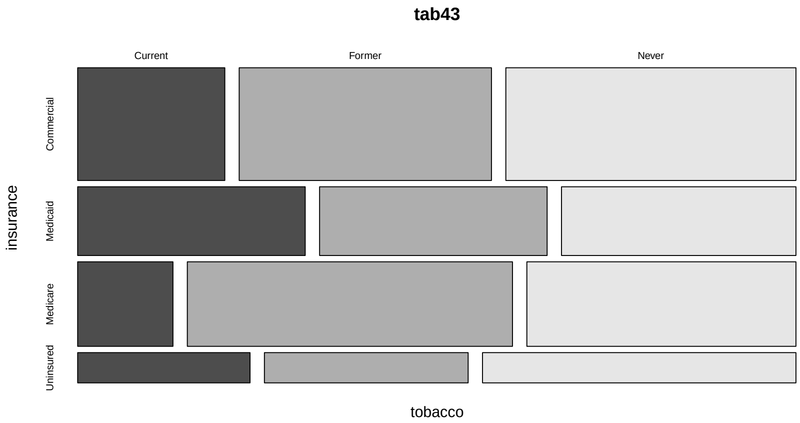
- Area of each box corresponds to its observed count

```
mosaicplot(tab43, color = TRUE)
```



Flipping the coordinates of the mosaicplot

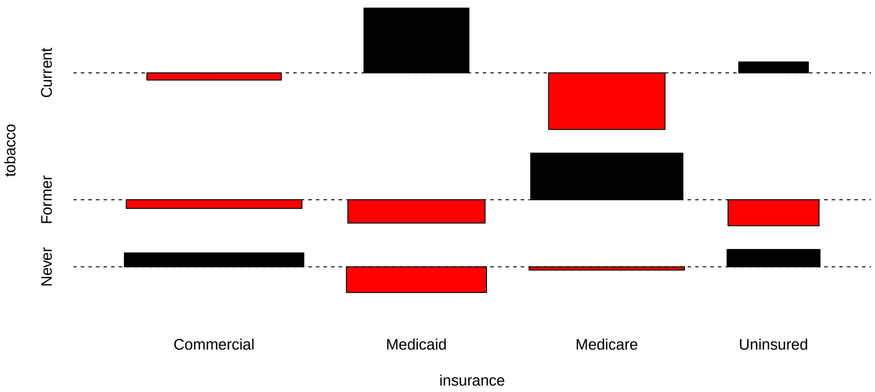
```r
mosaicplot(tab43, color = TRUE, dir = c("h", "v"))
```



assocplot to show deviations from independence

- Area of each box is proportional to (observed - expected count)

```
assocplot(tab43)
```



Visit the extra slides / recording

Extra Example 2. Statin Use by Practice in the `dm431` study

Cochran-Mantel-Haensel Test for a three-way table (technically a set of 2×2 tables)

Kidney Stone Treatment Example

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

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

Kidney Stones	Successful Outcome	Bad Outcome
A (open surgery)	273 (78%)	77 (22%)
B (less invasive)	289 (83%)	61 (17%)

Which approach would you choose?

- Sources: Wikipedia and Charig CR et al. (1986) PMID 3083922.

Kidney Stones, twobytwo results

```
twobytwo(273, 77, 289, 61, "A", "B", "Success", "Bad")
```

```
2 by 2 table analysis:
------------------------------------------------------
Outcome   : Success 
Comparing : A vs. B 

  Success Bad    P(Success) 95% conf. interval
A     273  77        0.7800    0.7336   0.8203
B     289  61        0.8257    0.7823   0.8620

                                   95% conf. interval
         Relative Risk:  0.9446    0.8776   1.0168
     Sample Odds Ratio:  0.7483    0.5146   1.0883
Probability difference: -0.0457   -0.1045   0.0133

Exact P-value: 0.154    Asymptotic P-value: 0.1292
------------------------------------------------------
```

Kidney Stones: A Third Variable

But this comparison may be misleading.

Some kidney stones are small, and some are large.

- Open Surgery (A) used in 87 small stones, and 263 large ones.
- Less Invasive (B) used in 270 small stones, and 80 large ones.

Could that bias our results?

- Should we account for this difference in "size mix"?

Kidney Stone results stratified by stone size

- For small stones, the odds ratio for a successful outcome comparing A to B is 2.08 (95% CI 0.84, 5.11)

Small Stones	Successful Outcome	Bad Outcome
A (open surgery)	81 (93%)	6 (7%)
B (less invasive)	234 (87%)	36 (13%)

- For large stones, that odds ratio is 1.23 (95% CI 0.71, 2.12)

Large Stones	Successful Outcome	Bad Outcome
A (open surgery)	192 (73%)	71 (27%)
B (less invasive)	55 (69%)	25 (31%)

Aggregated Data: % with Successful Outcome

- 78% of Treatment A subjects, 83% of Treatment B

What We Have Here is a Three-Way Table

- rows: which treatment was received (A or B)
- columns: was the outcome Successful or Bad?
- *strata* or *layers*: was the stone Small or Large?

```
Size  Treatment  Good  Bad  Total  % Good
----- ---------  ----  ---  -----  ------
Small     A        81    6     87      93
Small     B       234   36    270      87
Large     A       192   71    263      73
Large     B        55   25     80      69
```

We'll talk about three-way and larger contingency tables more in 432, but in 431, we focus on the situation where a 2x2 table is repeated over multiple strata (categories in a third variable.)

Cochran-Mantel-Haenszel Test

The Cochran-Mantel-Haenszel test is designed to test whether the rate of a successful (Good) outcome is the same across the two levels of the treatment (i.e. A or B.)

- We *could* do this by simply adding up the results across the stone sizes, but that wouldn't be wise, because the stone size is likely to be related to the outcome and the choice of procedure.
- But we can account for the differences between stone sizes to some extent by adjusting for stone size as a stratifying variable in a CMH test.
- The big assumption we'll have to make, though, is that the odds ratio for a good outcome for treatment A versus treatment B is the same for small stones and large stones. Is this reasonable here? We'll use a Woolf test to decide.

But first, let's get the data into a useful form.

Building the Three-Way Table

```
stone <- c(rep("Small", 4), rep("Large", 4))
treat <- c(rep(c("A", "A", "B", "B"), 2))
result <- c(rep(c("Good", "Bad"), 4))
counts <- c(81, 6, 234, 36, 192, 71, 55, 25)

kidney_dat <- tibble(stone, treat, result, counts) %>%
  mutate(result = fct_relevel(factor(result), "Good"))
```

What do we have so far?

```
kidney_dat
```

```
# A tibble: 8 x 4
  stone treat result counts
  <chr> <chr> <fct>   <dbl>
1 Small A     Good       81
2 Small A     Bad         6
3 Small B     Good      234
4 Small B     Bad        36
5 Large A     Good      192
6 Large A     Bad        71
7 Large B     Good       55
8 Large B     Bad        25
```

The Three-Way Table

```r
big.tab <- xtabs(counts ~ treat + result + stone,
                 data = kidney_dat)
big.tab
```

```
, , stone = Large

     result
treat Good Bad
    A  192  71
    B   55  25

, , stone = Small

     result
treat Good Bad
    A   81   6
    B  234  36
```

Three-Way Table as a “Flat” Table

```r
ftable(big.tab)
```

```
             stone Large Small
treat result                  
A     Good           192    81
      Bad             71     6
B     Good            55   234
      Bad             25    36
```

Can we assume a Common Odds Ratio?

- Recall the sample odds ratio in small stones was 2.08 and in large stones was 1.23

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 two stone sizes.

```
woolf_test(big.tab)
```

```

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

data:  big.tab
X-squared = 0.95353, df = 1, p-value = 0.3288
```

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

We then use the Cochran-Mantel-Haenszel test to make inferences about the population odds ratio. Here we define this odds ratio as the odds of a Good result for those on treat A divided by the odds of a Good result for those on treat B, while adjusting for the size of the stone. We know this is the odds ratio of interest by looking at the flattened table, and observing which `treat` and `result` are shown first.

```
ftable(big.tab)
```

```
             stone Large Small
treat result
A     Good           192    81
      Bad             71     6
B     Good            55   234
      Bad             25    36
```

- Sample odds ratio in small stones was 2.08 and in large stones was 1.23

Complete CMH output (Edited to fit on the screen)

We'll use a 90% confidence interval.

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

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

data:  big.tab
Mantel-Haenszel X-squared = 2.0913, df = 1, p-value = 0.1481

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

90 percent confidence interval: 0.9856691 2.1238011
sample estimates: common odds ratio 1.446847
```

What can we conclude in this case?

Visit the extra slides / recording

Extra Example 3: Auto Accidents in Alberta, Canada

Extra Example 4: Admissions to Departments at the University of California at Berkeley and Simpson's Paradox

Extra Example 5: A Meta-Analysis of Niacin and Heart Disease

What's Next?

Power and sample size considerations are coming in Class 21.

- I'll skip paired comparisons of proportions in 431 this year.