431 Class 19

Thomas E. Love

2017-11-02

The trouble with baseball...

It breaks your heart. It is designed to break your heart. The game begins in the spring, when everything else begins again, and it blossoms in the summer, filling the afternoons and evenings, and then as soon as the chill rains come, it stops and leaves you to face the fall alone. You count on it, rely on it to buffer the passage of time, to keep the memory of sunshine and high skies alive, and then just when the days are all twilight, when you need it most, it stops. Today, [November 2], a [Thursday] of rain and broken branches and leaf-clogged drains and slick streets, it stopped, and summer was gone.

"The Green Fields of the Mind" (A. Bartlet Giamatti)

Today's Agenda

- Overly brief discussion of Assignment 4
- An EDA Approach for Quantitative Variables
    - for One Sample (or Paired Differences)
    - for 2+ Independent Samples
- Inference about Rates/Proportions
- In a single sample
- In a 2x2 table
- Power considerations for comparing two proportions

Today's R Setup

```
library(pwr); library(forcats); library(tidyverse)

source("Love-boost.R")
dm192 <- read.csv("data/dm192.csv") %>% tbl_df
```

Assignment 4

Assignment 4 didn't go as well as we'd hoped

Questions 9-11: Many people had trouble with the `zocazo` example.

- Question 9 asked you to build a 99% confidence interval for the difference of two means, assuming a Normal distribution for each group of women.
- Question 10 asked you to do a sample size calculation, and then multiply the sample size by the cost per measurement to get a total amount of money required.
- Question 11 asked you to redo the sample size calculation, in light of a revised budget and desired confidence level.

It's worthwhile to spend some time understanding how `power.t.test` works in either the paired samples or (as in this case) independent samples setting. Remember to think about n, δ, `sd`, α (or confidence) and β (or power.)

Assignment 4 troubles

You also need to be able to parse something like this properly. . .

> *Suppose that in our new study, we assume a minimum clinically important effect 20% as large as was seen in the previous study.*

And you should definitely become intimately familiar with the set of questions I ask every time I am comparing population means/centers, in particular, people stumbled on:

- whether samples are paired or independent,
- whether a sample can be trusted to be sufficiently representative, especially if it's not a random sample,
- what visualization(s) we need to address assumptions and select an inference method, and
- what to do if we use a significance level other than 5%.

How to move forward?

Course Notes: Sections **27** (2 examples comparing means) and **36** (more general review of Part B) can help.

Two functions to help build Exploratory Data Analysis for Quantitative Variables

Exploratory Data Analysis for Quantitative Variables

Last year, Mustafa Ascha and I wrote two functions to build plots of quantitative data. They need some improvement, but they do some useful things.

- `eda.1sam` runs exploratory data analyses for a single sample (or paired differences)
- `eda.2sam` runs exploratory data analyses for two or more independent samples

which are part of the `Love-boost.R` script.

dm192 data, as a demo

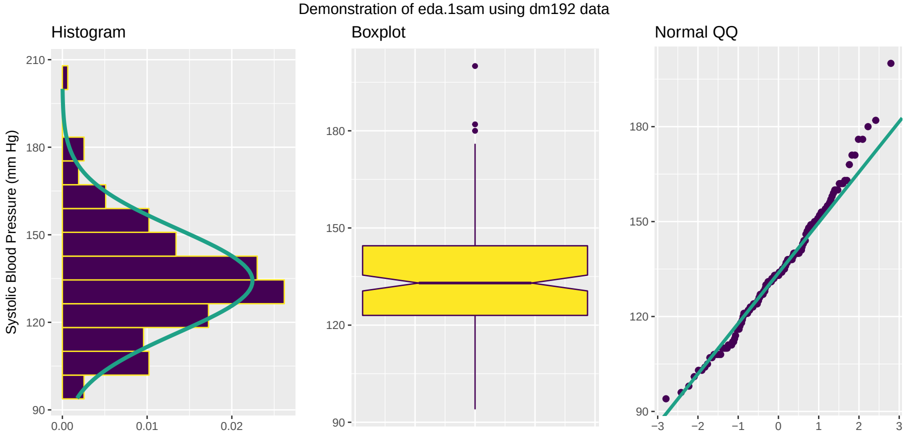
Suppose we want to look at a single sample - the sbp data from the dm192 data frame.

- Inputs: dataframe, variable, x.title, ov.title
- Output: three-panel plot (histogram, boxplot, Normal q-q plot)
- Required packages: tidyverse, pander, gridExtra
- We suggest you run it with message = FALSE

The function is called eda.1sam, and calling it looks like:

```
eda.1sam(dataframe = dm192, variable = dm192$sbp,
         x.title = "Systolic Blood Pressure (mm Hg)",
         ov.title = "Demonstration of eda.1sam
                     using dm192 data")
```

Results on the next slide. . .


Demonstration of eda.1sam using dm192 data
Histogram
Boxplot
Normal QQ
Systolic Blood Pressure (mm Hg)
210
180
150
120
90
0.00
0.01
0.02
−3
−2
−1
0
1
2
3

eda.2sam function demonstration

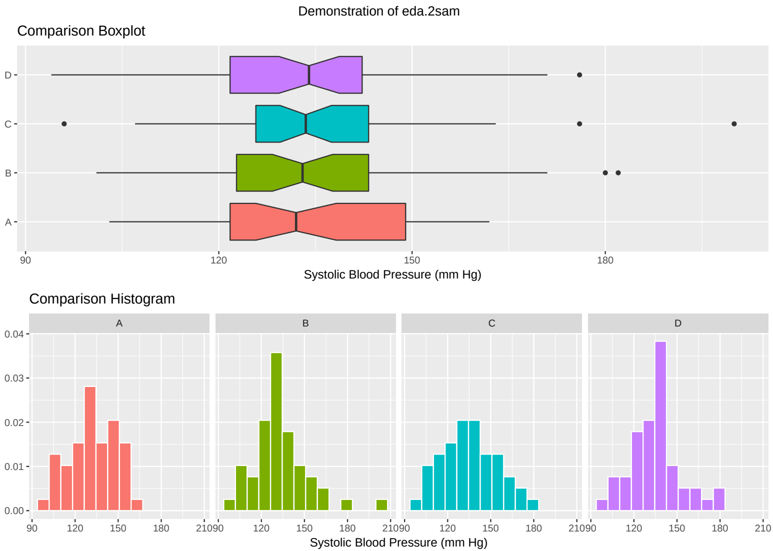
Suppose we want to look at two or more independent samples, here, we'll compare the sbp data from the dm192 data frame across the levels of the four practices (A, B, C, and D).

- Inputs: outcome, group, y.title, ov.title
- Output: comparison boxplot and faceted histograms
- Required packages: tidyverse, gridExtra, mosaic
- We suggest you run it with message = FALSE

The function is called eda.2sam, and the command would be:

```
eda.2sam(outcome = dm192$sbp, group = dm192$practice,
         y.title = "Systolic Blood Pressure (mm Hg)",
         ov.title = "Demonstration of eda.2sam")
```

- Results on the next slide.


Demonstration of eda.2sam
Comparison Boxplot
D
C
B
A
90
120
150
180
Systolic Blood Pressure (mm Hg)
Comparison Histogram
A
B
C
D
0.04
0.03
0.02
0.01
0.00
90
120
150
180
210
Systolic Blood Pressure (mm Hg)

Moving on from Means to Proportions / Rates

Moving on from Means to Proportions

We've focused on creating statistical inferences about a population mean, or difference between means, where we care about a quantitative outcome. Now, we'll tackle **categorical** outcomes.

1. We'll start by estimating a confidence interval around an unknown population proportion, or rate, which we'll symbolize with π, on the basis of a random sample of n observations from a sample which yields a proportion of $\hat{p}$, which is sometimes, unfortunately, symbolized as p. Note that this $\hat{p}$ is the sample proportion - not a p value.
2. Then we'll look at comparing proportions π_1 and π_2 - comparisons across two populations, based on samples of size n_1 and n_2.

Safer with More Guns?

A July 5-9, 2016 McClatchy-Marist poll of 1,053 registered U.S. voters nationwide asked **Do you think Americans are safer with more guns or fewer guns?** Results:

–	More Guns	Fewer Guns	Number is about right	Unsure
%	45	46	3	5

1. What can we conclude from this poll about the true percentage of registered U.S voters who would answer "More Guns"?
2. The poll lists a "margin of error" of 3 percentage points. What does this mean?

- My source: http://www.pollingreport.com/guns.htm
- Note that "Number is about right" was a voluntary (not pre-specified) response.

A Confidence Interval for a Proportion

A 100(1-α)% confidence interval for the population proportion π can be created by using the standard normal distribution, the sample proportion, $\hat{p}$, and the standard error of a sample proportion, which is defined as the square root of $\hat{p}$ multiplied by $(1 - \hat{p})$ divided by the sample size, n.

Specifically, our confidence interval is $\hat{p} \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

where $Z_{\alpha/2} =$ the value from a standard Normal distribution cutting off the top $\alpha/2$ of the distribution, obtained in R by substituting the desired $\alpha/2$ value into: `qnorm(alpha/2, lower.tail=FALSE)`.

- *Note*: This interval is reasonably accurate so long as $n\hat{p}$ and $n(1 - \hat{p})$ are each at least 5.

Estimating π in the "More Guns" Example

- We'll build a 95% confidence interval for the true population proportion, so $\alpha = 0.05$
- We have n = 1,053 subjects who responded
- Sample proportion saying "more guns" is $\hat{p} = 0.45$; we'll assume that $(1053)(0.45) = 474$ actually said this.

The standard error of that sample proportion will be

$$SE(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.45(1-0.45)}{1053}} = 0.015$$

Confidence Interval for π in "More Guns"

Our 95% confidence interval for the true population proportion, π, of voters who would choose "more guns" is

$\hat{p} \pm Z_{.025}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$,

or $0.45 \pm 1.96(0.015) = 0.45 \pm 0.029$, or (0.421, 0.479)

I simply recalled from our prior work that $Z_{0.025} = 1.96$, but we can verify this:

```
qnorm(0.025, lower.tail=FALSE)
```

```
[1] 1.959964
```

Likely Accuracy of this Confidence Interval?

Since $n\hat{p} = (1053)(0.45) = 474$ and $n(1 - \hat{p}) = (1053)(1 - 0.45) = 579$ are substantially greater than 5, the CI should be reasonably accurate.

1. What can we conclude from this poll about the true percentage of registered U.S voters who would answer "More Guns"?
   - Our best point estimate of the true population proportion who would say "more guns" is 0.45.
   - We are 95% confident that the true population proportion is between 0.421 and 0.479.
2. The poll lists a "margin of error" of 3 percentage points. What does this mean?
   - Note that our 95% confidence interval for π can also be expressed as 0.45 ± 0.029.

Happily, we don't have to do these calculations by hand ever again.

R Methods to get a CI for a Population Proportion

I am aware of at least three different procedures for estimating a confidence interval for a population proportion using R. All have minor weaknesses: none is importantly different from the others in many practical situations.

1. The `prop.test` approach (also called the Wald test)

```
prop.test(x = 474, n = 1053)
```

2. The `binom.test` approach (Clopper and Pearson "exact" test)

```
binom.test(x = 474, n = 1053)
```

3. Building a confidence interval via a `SAIFS` procedure

```
saifs.ci(x = 474, n = 1053)
```

The prop.test approach (Wald test)

The `prop.test` function estimates a confidence interval for π:

```r
prop.test(x = 474, n = 1053)
```

```

    1-sample proportions test with continuity
    correction

data:  474 out of 1053, null probability 0.5
X-squared = 10.272, df = 1, p-value = 0.001351
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
 0.4198583 0.4807948
sample estimates:
        p
0.4501425
```

binom.test (Clopper-Pearson "exact" test)

```r
binom.test(x = 474, n = 1053)
```

```
    Exact binomial test

data:  474 and 1053
number of successes = 474, number of trials =
1053, p-value = 0.00134
alternative hypothesis: true probability of success is not equ
95 percent confidence interval:
 0.4197937 0.4807706
sample estimates:
probability of success
             0.4501425
```

Estimating a Rate More Accurately

Suppose you have some data involving n independent tries, with x successes. The most natural estimate of the "success rate" in the data is x / n.

But, strangely enough, it turns out this isn't an entirely satisfying estimator. Alan Agresti provides substantial motivation for the $(x + 1)/(n + 2)$ estimate as an alternative[1]. This is sometimes called a *Bayesian augmentation*.

[1] This note comes largely from a May 15 2007 entry in Andrew Gelman's blog at http://andrewgelman.com/2007/05/15

Use (x + 1)/(n + 2) rather than x/n

- The big problem with x / n is that it estimates p = 0 or p = 1 when x = 0 or x = n.
- It's also tricky to compute confidence intervals at these extremes, since the usual standard error for a proportion, $\sqrt{np(1-p)}$, gives zero, which isn't quite right.
- (x + 1)/(n + 2) is much cleaner, especially when you build a confidence interval for the rate.
- The only place where (x + 1)/(n + 2) will go wrong (as in the SAIFS approach) is if n is small and the true probability is very close to 0 or 1.
    - For example, if n = 10, and p is 1 in a million, then x will almost certainly be zero, and an estimate of 1/12 is much worse than the simple 0/10.
    - However, how big a deal is this? If p might be 1 in a million, are you going to estimate it with an experiment using n = 10?

Practical Impact of Bayesian Augmentation

It is likely that the augmented `(x + 1) / (n + 2)` version yields more accurate estimates for the odds ratio or relative risk or probability difference, but the two sets of estimates (with and without the augmentation) will be generally comparable, so long as...

1. the sample size in each exposure group is more than, say, 30 subjects, and/or
2. the sample probability of the outcome is between 0.1 and 0.9 in each exposure group.

Bayesian Augmentation: Add a Success and a Failure

You'll get slightly better results if you use $\frac{x+1}{n+2}$ rather than $\frac{x}{n}$ as your point estimate, and to fuel your confidence interval using either the `binom.test` or `prop.test` approach.

- The results will be better in the sense that they'll be slightly more likely to meet the nominal coverage probability of the confidence intervals.
- This won't make a meaningful difference if $\frac{x}{n}$ is near 0.5, or if the sample size n is large. Why?

Suppose you want to find a confidence interval when you have 2 successes in 10 trials. I'm suggesting that instead of `binom.test(x = 2, n = 10)` you might want to try `binom.test(x = 3, n = 12)`

SAIFS confidence interval procedure

SAIFS = single augmentation with an imaginary failure or success[2]

- Uses a function I built in R for you (Part of `Love-boost.R`)

```
saifs.ci(x = 474, n = 1053)
```

```
Sample Proportion            0.025            0.975
            0.450            0.420            0.481
```

`saifs.ci` already builds in a Bayesian augmentation, so we don't need to do that here.

[2] see Notes Part B for more details.

Results for "More Guns" Rate (x = 474, n = 1053)

Method	95% CI for π
`prop.test`	0.420, 0.481
`binom.test`	0.420, 0.481
`saifs.ci`	0.420, 0.481

Our "by hand" result, based on the Normal distribution, with no continuity correction, was (0.421, 0.479).

So in this case, it really doesn't matter which one you choose. With a smaller sample, we may not come to the same conclusion about the relative merits of these different approaches.

Assumptions behind Inferences about π

We are making the following assumptions, when using these inferential approaches:

1. There are n identical trials.
2. There are exactly two possible outcomes (which may be designated as success and failure) for each trial.
3. The true probability of success, π, remains constant across trials.
4. Each trial is independent of all of the other trials.

Accuracy of these Inferences about a Proportion

We'd like to see that both $n\hat{p}$ = observed successes and $n(1 - \hat{p})$ = observed failures exceed 5.

- If not, then the intervals may be both incorrect (in the sense of being shifted away from the true value of π), and also less efficient (wider) than necessary.

None of these approaches is always best

When we have a sample size below 100, or the sample proportion of success is either below 0.10 or above 0.90, caution is warranted[3], although in many cases, the various methods give similar responses.

95% CI Approach	Wald	Clopper-Pearson	SAIFS
X = 10, n = 30	0.179, 0.529	0.173, 0.528	0.148, 0.534
X = 10, n = 50	0.105, 0.341	0.1, 0.337	0.083, 0.333
X = 90, n = 100	0.82, 0.948	0.824, 0.951	0.829, 0.96
X = 95, n = 100	0.882, 0.981	0.887, 0.984	0.894, 0.994

[3] These are great times for the Bayesian augmentation, for ‘prop.test‘ or ‘binom.test‘

Hypothesis Testing About a Population Proportion

To perform a hypothesis test about a population proportion, we'll usually use the `prop.test` or `binom.test` approaches in R[4].

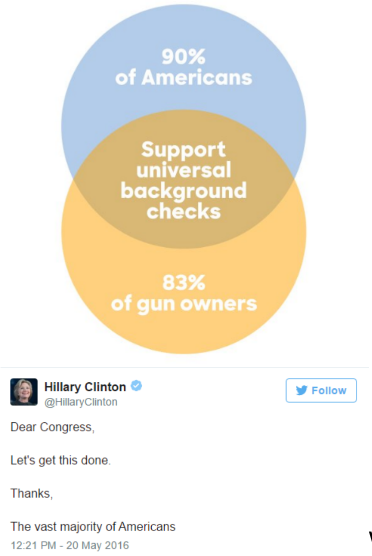
- The null hypothesis is that the population proportion is equal to some pre-specified value. Often, this is taken to be 0.5, but it can be any value, called π_0, that is between 0 and 1.
- The alternative hypothesis may be one-sided or two-sided. If it is two-sided, it will be that the population proportion is not equal to the value π_0 specified by the null hypothesis.
- In the two-sided case, we have $H_0 : \pi = \pi_0$ and $H_A : \pi \neq \pi_0$
- In the one-sided "greater than" case, we have $H_0 : \pi \leq \pi_0$ and $H_A : \pi > \pi_0$

But, as usual, the focus is usually on the confidence intervals...

[4] Bayesian augmentation is helpful here, too.

Comparing Proportions

A Troublesome Tweet (2016-05-20)



What is wrong with this picture?

Comparing Two Proportions

Quinnipiac U. poll December 16-20, 2015 of 1,140 registered U.S. voters

- **Would you support or oppose a law requiring background checks on people buying guns at gun shows or online?**
- **Do you personally own a gun or does someone else in your household own a gun?**

Reported summaries of that poll get me to the following table:

–	Support Law	Oppose Law	*Total*
No Gun	542	24	*566*
Gun Household	440	73	*513*
Total	*982*	*97*	*1,079*

- Links to sources: fivethirtyeight and pollingreport
- Source Images on next two slides

Polling Topline Images, from fivethirtyeight.com

67. Would you support or oppose a law requiring background checks on people buying guns at gun shows or online?

							COLLEGE DEG	
	Tot	Rep	Dem	Ind	Men	Wom	Yes	No
Support	89%	87%	95%	86%	84%	94%	92%	88%
Oppose	9	12	5	12	14	5	7	10
DK/NA	2	1	-	2	2	1	1	2

	AGE IN YRS..............				Gun	DENSITY............		
	18-34	35-49	50-64	65+	HsHld	Urban	Suburb	Rural
Support	98%	87%	87%	90%	84%	89%	91%	88%
Oppose	2	12	12	8	14	9	9	11
DK/NA	-	1	1	2	2	2	1	2

Polling Topline Images, from pollingreport.com

Quinnipiac University. Dec. 16-20, 2015. N=1,140 registered voters nationwide. Margin of error ą 2.9.

"Would you support or oppose a law requiring background checks on people buying guns at gun shows or online?"

	Support	Oppose	Unsure/ No answer
	%	%	%
12/16-20/15	89	9	2
Republicans	87	12	1
Democrats	95	5	-
Independents	86	12	2
4/25-29/13	83	13	3

"Do you personally own a gun or does someone else in your household own a gun?"

	Yes: Personally	Yes: Someone else	Yes: Both	No gun	Unsure/ No answer
	%	%	%	%	%
12/16-20/15	25	12	9	50	4
Republicans	37	12	15	29	7
Democrats	13	11	1	74	1
Independents	28	11	12	46	4

2 x 2 Table of Guns and Support, Prob. Difference

–	Support	Oppose	*Total*
No Gun in HH	542	24	*566*
Gun Household	440	73	*513*
Total	*982*	*97*	*1,079*

- Of those living in a no gun household, $542/566 = 95.8\%$ support universal background checks.
- Of those living in a gun household, $440/513 = 85.8\%$ support universal background checks.
- So the sample shows a difference of 10 percentage points, or a difference of 0.10 in proportions

Can we build a confidence interval for the population difference in those two proportions?

2 x 2 Table of Guns and Support, Relative Risk

–	Support	Oppose	*Total*
No Gun in HH	542	24	*566*
Gun Household	440	73	*513*
Total	*982*	*97*	*1,079*

- $\Pr(\text{support} \mid \text{no gun in HH}) = 542/566 = 0.958$
- $\Pr(\text{support} \mid \text{gun in HH}) = 440/513 = 0.858$
- The ratio of those two probabilities (risks) is $.958/.858 = 1.12$

Can we build a confidence interval for the relative risk of support in the population given no gun as compared to gun?

2 x 2 Table of Guns and Support, Odds Ratio

–	Support	Oppose	*Total*
No Gun in HH	542	24	*566*
Gun Household	440	73	*513*
Total	*982*	*97*	*1,079*

- Odds = Probability / (1 - Probability)
- Odds of Support if No Gun in HH $= \frac{542/566}{1-(542/566)} = 22.583333$
- Odds of Support if Gun in HH $= \frac{440/513}{1-(440/513)} = 6.027397$
- Ratio of these two Odds are 3.75

In a 2x2 table, odds ratio = cross-product ratio.

- Here, the cross-product estimate $= \frac{542*73}{440*24} = 3.75$

Can we build a confidence interval for the odds ratio for support in the population given no gun as compared to gun?

2x2 Table Results in R

```r
twobytwo(542, 24, 440, 73,
      "No Gun in HH", "Gun Household", "Support", "Oppose")
```

```
2 by 2 table analysis:
------------------------------------------------------
Outcome   : Support
Comparing : No Gun in HH vs. Gun Household

              Support Oppose    P(Support) 95% conf.
No Gun in HH      542     24        0.9576    0.9375
Gun Household     440     73        0.8577    0.8247
              interval
No Gun in HH    0.9714
Gun Household   0.8853

                                     95% conf. interval
             Relative Risk: 1.1165    1.0735   1.1612
```

Full Output

```
2 by 2 table analysis:
------------------------------------------------------
Outcome   : Support
Comparing : No Gun in HH vs. Gun Household

              Support Oppose  P(Support)  95% conf. int.
No Gun in HH      542     24      0.9576   0.9375   0.9714
Gun Household     440     73      0.8577   0.8247   0.8853

                                   95% conf. interval
             Relative Risk: 1.1165    1.0735   1.1612
         Sample Odds Ratio: 3.7468    2.3230   6.0431
Conditional MLE Odds Ratio: 3.7424    2.2867   6.3174
    Probability difference: 0.0999    0.0659   0.1355

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

Bayesian Augmentation in a 2x2 Table?

Original command:

```
twobytwo(542, 24, 440, 73,
      "No Gun in HH", "Gun Household", "Support", "Oppose")
```

Bayesian augmentation approach (add a success and add a failure in each row):

```
twobytwo(542+1, 24+1, 440+1, 73+1,
      "No Gun in HH", "Gun Household", "Support", "Oppose")
```

Full Output with Bayesian augmentation

```
2 by 2 table analysis:
------------------------------------------------------
Outcome   : Support
Comparing : No Gun in HH vs. Gun Household

             Support Oppose  P(Support) 95% conf. int.
No Gun in HH     543     25      0.9560 0.9357  0.9701
Gun Household    441     74      0.8563 0.8233  0.8840

                                   95% conf. interval
             Relative Risk: 1.1164    1.0731   1.1614
         Sample Odds Ratio: 3.6446    2.2768   5.8342
Conditional MLE Odds Ratio: 3.6405    2.2413   6.0875
    Probability difference: 0.0997    0.0655   0.1355

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

Using a data frame, rather than a 2x2 table

For example, in the dm192 data, suppose we want to know whether statin prescriptions are more common among male patients than female patients. So, we want a two-way table with "Male", "Statin" in the top left.

```
dm192$sex.f <- factor(dm192$sex, levels = c("male", "female"))
dm192$statin.f <- factor(dm192$statin, levels = c(1,0))
table(dm192$sex.f, dm192$statin.f)
```

```
          1  0
  male   73 21
  female 74 24
```

Running `twoby2` against a data set

The `twoby2` function from the `Epi` package can operate with the table we've generated.

```
twoby2(dm192$sex.f, dm192$statin.f)
```

```
2 by 2 table analysis:
------------------------------------------------------
Outcome   : 1
Comparing : male vs. female

        1  0    P(1) 95% conf. interval
male   73 21  0.7766    0.6815   0.8496
female 74 24  0.7551    0.6605   0.8301

                                   95% conf. interval
             Relative Risk: 1.0285    0.8795   1.2026
         Sample Odds Ratio: 1.1274    0.5775   2.2010
```

Full Output

```
2 by 2 table analysis:
------------------------------------------------------
Outcome   : 1
Comparing : male vs. female

        1  0    P(1) 95% conf. interval
male   73 21  0.7766    0.6815   0.8496
female 74 24  0.7551    0.6605   0.8301

                                    95% conf. interval
             Relative Risk: 1.0285    0.8795   1.2026
         Sample Odds Ratio: 1.1274    0.5775   2.2010
Conditional MLE Odds Ratio: 1.1267    0.5473   2.3330
    Probability difference: 0.0215   -0.0985   0.1399

Exact P-value: 0.7368       Asymptotic P-value: 0.7253
------------------------------------------------------
```

Bayesian Augmentation

```
twobytwo(73+1, 21+1, 74+1, 24+1,
         "male", "female", "statin", "no statin")
```

```
       statin no statin    P(statin) 95% conf. interval
male       74        22       0.7708    0.6764   0.8441
female     75        25       0.7500    0.6561   0.8251

                                    95% conf. interval
             Relative Risk: 1.0278    0.8783   1.2027
         Sample Odds Ratio: 1.1212    0.5814   2.1624
    Probability difference: 0.0208   -0.0988   0.1389

Exact P-value: 0.7414       Asymptotic P-value: 0.7328
```

Power and Sample Size When Comparing Proportions

Relation of α and β to Error Types

Recall the meanings of α and β:

- α is the probability of rejecting H_0 when H_0 is true.
  - So $1 - \alpha$, the confidence level, is the probability of retaining H_0 when that's the right thing to do.
- β is the probability of retaining H_0 when H_A is true.
  - So $1 - \beta$, the power, is the probability of rejecting H_0 when that's the right thing to do.

–	H_A is True	H_0 is True
Test Rejects H_0	Correct Decision (1 - β)	Type I Error (α)
Test Retains H_0	Type II Error (β)	Correct Decision (1 - α)

Tuberculosis Prevalence Among IV Drug Users

Here, we investigate factors affecting tuberculosis prevalence among intravenous drug users.

Among 97 individuals who admit to sharing needles, 24 (24.7%) had a positive tuberculin skin test result; among 161 drug users who deny sharing needles, 28 (17.4%) had a positive test result.

To start, we'll test the null hypothesis that the proportions of intravenous drug users who have a positive tuberculin skin test result are identical for those who share needles and those who do not.

Two-by-Two Table Command (with Bayesian Augmentation)

```
twobytwo(24+1, 73+1, 28+1, 133+1,
         "Sharing", "Not Sharing",
         "TB test+", "TB test-")
```

Two-by-Two Table Result

```
Outcome    : TB test+
Comparing : Sharing vs. Not Sharing

            TB test+ TB test- P(TB test+) 95% conf. int.
Sharing           25       74      0.2525  0.1767 0.3471
Not Sharing       29      134      0.1779  0.1265 0.2443

                                       95% conf. interval
             Relative Risk: 1.4194    0.8844   2.2779
         Sample Odds Ratio: 1.5610    0.8520   2.8603
Conditional MLE Odds Ratio: 1.5582    0.8105   2.9844
    Probability difference: 0.0746   -0.0254   0.1814

Exact P-value: 0.1588      Asymptotic P-value: 0.1495
```

What conclusions should we draw?

Designing a New TB Study

Now, suppose we wanted to design a new study with as many non-sharers as needle-sharers participating, and suppose that we wanted to detect any difference in the proportion of positive skin test results between the two groups that was identical to the data presented above or larger with at least 90% power, using a two-sided test and $\alpha = .05$.

What sample size would be required to accomplish these aims?

How `power.prop.test` works

`power.prop.test` works much like the `power.t.test` we saw for means.

Again, we specify 4 of the following 5 elements of the comparison, and R calculates the fifth.

- The sample size (interpreted as the # in each group, so half the total sample size)
- The true probability in group 1
- The true probability in group 2
- The significance level (α)
- The power (1 - β)

The big weakness with the `power.prop.test` tool is that it doesn't allow you to work with unbalanced designs.

Using power.prop.test for Balanced Designs

Want to find the sample size for a two-sample comparison of proportions using a balanced design

- we will use a two-sided test, with $\alpha = .05$, and power $= .90$,
- we estimate that the non-sharers will have a .174 proportion of positive tests,
- and we will try to detect a difference between this group and the needle sharers, who we estimate will have a proportion of .247

R Command to find the required sample size

```
power.prop.test(p1 = .174, p2  = .247,
                sig.level = 0.05, power = 0.90)
```

Results: power.prop.test for Balanced Designs

```r
power.prop.test(p1 = .174, p2  = .247,
                sig.level = 0.05, power = 0.90)
```

```
Two-sample comparison of proportions power calculation
n = 653.2876
p1 = 0.174, p2 = 0.247
sig.level = 0.05, power = 0.9, alternative = two.sided
NOTE: n is number in *each* group
```

So, we'd need at least 654 non-sharing subjects, and 654 more who share needles to accomplish the aims of the study.

Another Scenario

Suppose we can get 400 sharing and 400 non-sharing subjects. How much power would we have to detect a difference in the proportion of positive skin test results between the two groups that was identical to the data above or larger, using a *one-sided* test, with $\alpha = .10$?

```r
power.prop.test(n=400, p1=.174, p2=.247, sig.level = 0.10,
                alternative="one.sided")
```

```
Two-sample comparison of proportions power calculation
n = 400, p1 = 0.174, p2 = 0.247
sig.level = 0.1, power = 0.8954262
alternative = one.sided
NOTE: n is number in *each* group```
```

We would have just under 90% power to detect such an effect.

Using the pwr package to assess sample size for Unbalanced Designs

The `pwr.2p2n.test` function in the `pwr` package can help assess the power of a test to determine a particular effect size using an unbalanced design, where n_1 is not equal to n_2.

As before, we specify four of the following five elements of the comparison, and R calculates the fifth.

- `n1` = The sample size in group 1
- `n2` = The sample size in group 2
- `sig.level` = The significance level (α)
- `power` = The power (1 - β)
- `h` = the effect size h, which can be calculated separately in R based on the two proportions being compared: p_1 and p_2.

Calculating the Effect Size `h`

To calculate the effect size for a given set of proportions, just use `ES.h(p1, p2)` which is available in the `pwr` package.

For instance, in our comparison, we have the following effect size.

```
ES.h(p1 = .174, p2 = .247)
```

```
[1] -0.1796783
```

Using pwr.2p2n.test in R

Suppose we can have 700 samples in group 1 (the not sharing group) but only half that many in group 2 (the group of users who share needles).

How much power would we have to detect this same difference (p1 = .174, p2 = .247) with a 5% significance level in a two-sided test?

R Command to find the resulting power

```
pwr.2p2n.test(h = ES.h(p1 = .174, p2 = .247),
              n1 = 700, n2 = 350, sig.level = 0.05)
```

Results of using pwr.2p2n.test

```
pwr.2p2n.test(h = ES.h(p1 = .174, p2 = .247),
              n1 = 700, n2 = 350, sig.level = 0.05)
```

```
difference of proportion power calculation
for binomial distribution (arcsine transformation)

h = 0.1796783, n1 = 700, n2 = 350
sig.level = 0.05, power = 0.7836768
alternative = two.sided
NOTE: different sample sizes
```

We will have about 78% power under these circumstances.

Comparison to Balanced Design

How does this compare to the results with a balanced design using only 1000 drug users in total, i.e. with 500 patients in each group?

```
pwr.2p2n.test(h = ES.h(p1 = .174, p2 = .247),
              n1 = 500, n2 = 500, sig.level = 0.05)
```

which yields a power estimate of 0.811. Or we could instead have used…

```
power.prop.test(p1 = .174, p2 = .247, sig.level = 0.05,
                n = 500)
```

which yields an estimated power of 0.809.

Each approach uses approximations, and slightly different ones, so it's not surprising that the answers are similar, but not identical.

Coming Up

- Testing for Independence (Notes Section 32)
- Three-Way Contingency Tables (Notes Section 33)
- On Statistical Significance and The p Value (Section 34)
- Type S and Type M errors (Section 35)