

Chi-squared tests: Miscellaneous extra useful stuff

Research Methods for Human Inquiry
Andrew Perfors

Some odds and ends

- Measuring effect size (with Cramer's V)
- What are the assumptions of the tests, and what do we do when they are violated?
 - The Fisher exact test
 - The McNemar test

Effect size

- **Effect size** measures how "big" the difference between the data and the null hypothesis predictions actually were.
- Cramer's V is a popular (and simple) measure that can be applied to chi-square tests:

$$V = \sqrt{\frac{X^2}{N(k - 1)}}$$

Cramer's V

χ^2 is the goodness of fit statistic itself; it measures “strength of evidence” against the null, so it’s a good place to start

$$V = \sqrt{\frac{\chi^2}{N(k-1)}}$$

N is the sample size; since large samples let you detect even very small violations of the null, we want to correct for this when calculating effect size.

k is either the number of rows or the number of columns in the observed frequency data (often the smaller one). main thing here is it ensures V usually lies between 0 and 1

Interpreting Cramer's V

Value of Cramer's V	Rough meaning
0 to 0.1	negligible
0.1 to 0.3	weak association
0.3 to 0.5	moderate association
0.5 to 1	high association

Calculating Cramer's V directly

- Use the `CramerV()` function in the `DescTools` package (note you may need to install this)
- It uses the exact same arguments as the `chisq.test()` function, so you input exactly the same information....

```
> library( DescTools )  
> CramerV (x = boxesTable)  
[1] 0.04175778
```

- Need a separate function in a separate package (which you also need to install) for Goodness of Fit tests

```
> library(rcompanion)  
> cramerVFit(x=votingTable,p=ed)  
[1] 0.1941
```

What are the assumptions of the test?

Why assumptions matter

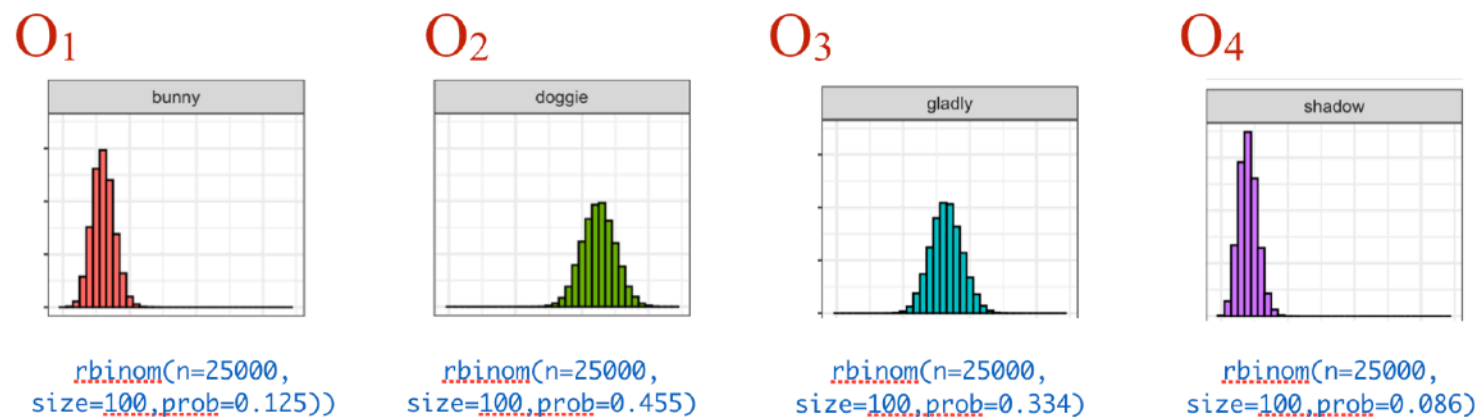
- All statistical tests rely on assumptions
- Assumptions are necessary to allow inference
- However, if the assumptions are wrong, you could end up making mistakes

- Both chi-square tests make two assumptions
 - "Large" expected frequencies
 - Independence of the data

1) "Large" expected frequencies

- On Day 1 I mentioned that the sampling distribution is only "approximately" chi-square
- This approximation breaks down if any of the expected frequencies are "too small"

The basic reason is that it only becomes chi-square if we can presume that there are enough observations for the underlying binomial distributions to be "normal" (but don't worry about this reason for the exam, you won't be assessed on it)



- Rule of thumb: you need the expected frequencies to be greater than about 5, or thereabouts

1) Solution: The Fisher Exact Test

- A analog of the chi-square test of association
 - It doesn't require large expected frequencies
 - In fact, it works best for small frequencies.
 - But it assumes that row and column totals are fixed
- It works by calculating the exact probability of obtaining a particular contingency table (i.e., cross-tabulation)
 - The p-value is calculated by summing over all contingency tables that are "more extreme" than the observed one.
 - The definition of "more extreme" is tricky, but basically means "more uneven"

Example: Bunny and Gladly fight over bows

Bunny and Gladly both really like red and really dislike blue. They also love to wear bows in their hair. There is a box with lots of bows. They are fighting over it, so Flopsy volunteers to closer her eyes and grab the bows for them.



Was Flopsy being mean to Gladly, or could this distribution have been random?

few observations
row & column totals set to 4

In R

```
> dbows  
  
# A tibble: 8 x 2  
  colour person  
  <chr>   <chr>  
1 red     bunny  
2 red     bunny  
3 red     bunny  
4 red     gladly  
5 blue    bunny  
6 blue    gladly  
7 blue    gladly  
8 blue    gladly
```

the raw
data



the contingency table



```
> bowTable <- table(dbows$person,dbows$colour)  
> bowTable  
  
      blue red  
bunny    1  3  
gladly    3  1
```

If we try to do a chi-square test, R gets
a bit panicky...

```
> chisq.test(x=bowTable)
```

Chi-squared approximation may be incorrect

Pearson's Chi-squared test with Yates' continuity correction

data: bowTable

X-squared = 0.5, df = 1, p-value = 0.4795

If we try to do a chi-square test, R gets
a bit panicky...

```
> chisq.test(x=dbows$colour,y=dbows$person)
```

Pearson's Chi-squared test with Yates' continuity correction

data: dbows\$colour and dbows\$person

X-squared = 0.5, df = 1, p-value = 0.4795

Warning message:

In chisq.test(x = dbows\$colour, y = dbows\$person) :

Chi-squared approximation may be incorrect

But this situation is perfectly suited to
the Fisher exact test...

```
> fisher.test(x=bowTable)
```



Note that you need to input the
contingency table (i.e. `bowTable`) for
this function

But this situation is perfectly suited to
the Fisher exact test...

```
> fisher.test(x=bowTable)
```

Fisher's Exact Test for Count Data

```
data: bowTable  
p-value = 0.4857  
alternative hypothesis: true odds ratio is not equal to 1  
95 percent confidence interval:  
 0.001607888 4.722931239  
sample estimates:  
odds ratio  
 0.156047
```

The output here has a bunch of stuff you
don't care about at this point

Main thing to note is the p-value

```
> fisher.test(x=bowTable)
```

Fisher's Exact Test for Count Data

```
data: bowTable
```

```
p-value = 0.4857
```

```
alternative hypothesis: true odds ratio is not equal to 1
```

```
95 percent confidence interval:
```

```
0.001607888 4.722931239
```

```
sample estimates:
```

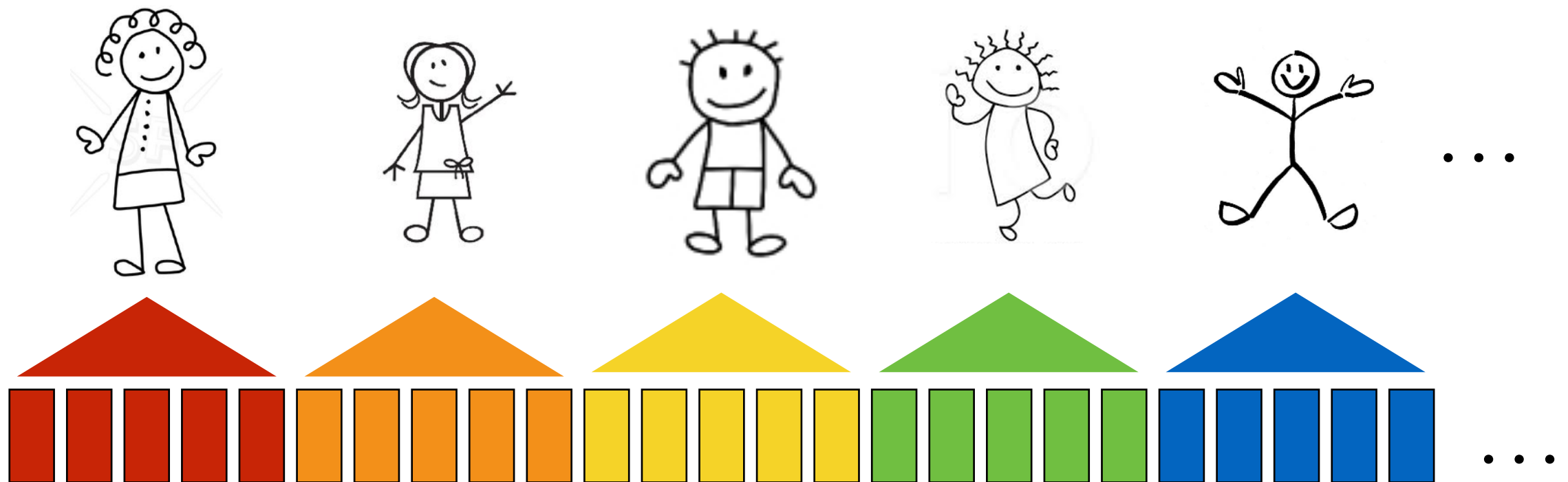
```
odds ratio
```

```
0.156047
```

So we can't reject the null hypothesis (which is that Flopsy's choices from the box were the same for Bunny and Gladly). Which means Flopsy probably was not trying to be extra mean to Gladly.

Assumption #2: Independence of the data

- This assumption is a little vague
- In practice, it means that you can't have any "special relationship" among some of your observations
 - e.g., if you had 20 people each cast 5 ballots, this would violate the assumption of independence because the 5 ballots cast by the same person have a "special" relationship to one another



Assumption #2: Independence of the data

- This assumption is a little vague
- In practice, it means that you can't have any "special relationship" among some of your observations
 - e.g., if you had 20 people each cast 5 ballots, this would violate the assumption of independence because the 5 ballots cast by the same person have a "special" relationship to one another
- Solution: McNemar test

McNemar: a limited solution to a standard problem

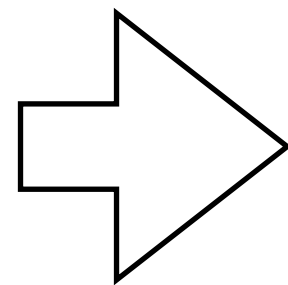
- The problem
 - What do you do when you have multiple observations from each person? (e.g., pre-test and post-test)
 - You can't do a chi-square test because this violates the independence assumption pretty severely
- A limited solution
 - Suppose you have a binary outcome measure (e.g., yes or no), and you measure it twice (e.g., pre- and post-).
 - You want to know if pre- is different from post-

An example

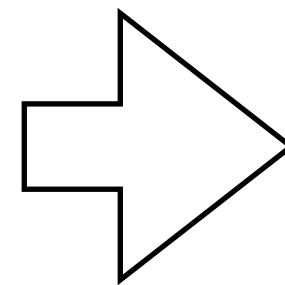
Flopsy argues that wanting to attack is fear-based reasoning, and suggests that if everybody does some meditation and deep breathing they might change their mind

100 people

“Do you want to
attack the
Others?” Y/N



<meditation>



**same 100
people**

“Do you want to
attack the
Others?” Y/N



The raw data

```
> d3
  before after
1    yes   yes
2    yes   yes
3    yes   yes
4    yes   yes
5    yes   yes
6    yes   yes
```

```
BLAH BLAH BLAH
```

```
99      no    no
100     no    no
```

We have a tibble with two variables: `before` and `after`

cross-tabulation

```
> medTable <- table(d3$before, d3$after)
> medTable
```

	no	yes
no	22	5
yes	17	56

McNemar's test

	Before = "yes"	Before = "no"
After = "yes"	56	5
After = "no"	17	22

It turns out (see textbook if you care why) that the diagnostic test in this situation is to test if these two cells have equal probability

McNemar's test is a chi-squared test that focuses on just these two squares

Doing it in R

```
> mcnemar.test(x=medTable)
```

```
McNemar's Chi-squared test with continuity correction
```

```
data: medTable
```

```
McNemar's chi-squared = 5.5, df = 1, p-value = 0.01902
```

Again, notice that you're giving it the contingency table!

So some of this may be fear based, and if people meditated and calmed themselves they might choose differently!

Exercises are in `w6day2exercises.Rmd`