Two-sample Categorical data: Measuring association

Patrick Breheny

November 3

Introduction

Today we will continue to discuss the analysis of contingency tables, dealing with two subjects:

- The variety of different study designs that lead to contingency tables
- Measuring association in contingency tables

Study designs that can be analyzed with χ^2-tests

- One reason that χ^2-tests are so popular is that they can be used to analyze a wide variety of study designs
- In addition to controlled experiments, they are widely used in epidemiology, where investigators must conduct observational studies
- Broadly speaking, observational studies in epidemiology fall into three categories: *prospective studies*, *retrospective studies*, and *cross-sectional studies*
- χ^2-tests and Fisher's exact test can be used to analyze all of these studies

Prospective studies

- We have said that the double-blind, randomized controlled trial is the gold standard of biomedical research
- When this is not possible (or ethical), the prospective study (also called a *cohort study*) is the next best thing
- In a prospective study, investigators collect a sample, classify individuals in some way, and then wait to see if the individuals develop a condition
- The classification is usually based on exposure to a risk factor such as smoking or obesity

Risk factors for breast cancer

- For example, the CDC tracked 6,168 women in the hopes of finding risk factors that led to breast cancer
- One risk factor they looked at was the age at which the woman gave birth to her first child:

	Cancer	
	No	Yes
Before age 25	4475	65
25 or older	1597	31

Risk factors for breast cancer (cont'd)

- Performing a χ^2-test on the data, we obtain $p = .19$
- Thus, the evidence from this study is rather unconvincing as far as whether the risk of developing breast cancer depends on the age at which a woman gives birth to her first child
- In other words, the data is consistent with the null hypothesis that the two groups have the same risk of developing breast cancer

Retrospective studies

- Not all researchers have the resources to follow thousands of people for decades to see if they develop a rare disease
- Instead, they often try the more feasible approach of collecting a sample of people with the condition of interest, a second sample of people without the condition of interest, and then ask them if they were exposed to a risk factor in the past
- For example, a much cheaper way to conduct the study of breast cancer risk factors would be to find 50 women with breast cancer, 50 women without breast cancer, and ask them when they had their first child
- This approach is called a *retrospective*, or *case-control* study

Fluoride poisoning in Alaska

- In 1992, an outbreak of illness occurred in an Alaskan community
- The CDC suspected fluoride poisoning from one of the town's water supplies

	Case	Control
Drank from supply	33	4
Didn't drink from supply	5	46

Fluoride poisoning in Alaska

- Testing whether this could be due to chance, the χ^2-test gives us $p = 6 \times 10^{-13}$
- The observed association was certainly not due to chance
- But the association still may be due to factors besides fluoride poisoning

Recall bias

- For example, people who got sick may think much harder about what they ate and drank than people who didn't
- This is called *recall bias*, and it is an important source of bias in retrospective studies
- The extent to which recall bias is a concern certainly depends on the study:
  - In the breast cancer example, it would not be much of a concern, since giving birth to a child is a major life event and a woman would know how old she was when it happened
  - On the other hand, if the risk factor was something like diet or exercise, recall bias would be a huge concern, as people are notoriously unreliable at recalling these things
- Furthermore, because researchers must gather separate samples of cases and controls, these studies are more prone to sampling biases than prospective studies

Electromagnetic field example

- For example, retrospective studies have been performed investigating links between childhood leukemia and exposure to electromagnetic fields (EMF)
- Families with low socioeconomic status are more likely to live near electromagnetic fields
- Families with low socioeconomic status are also less likely to participate in studies as controls
- Socioeconomic status does not affect the participation of cases, however (cases are usually eager to participate)
- This results in an observed association between EMF and leukemia potentially arising entirely due to bias

Cross-sectional studies

- The weakest type of observational study is the cross-sectional study
- In a cross-sectional study, the investigator simply gathers a single sample and cross-classifies them depending on whether they have the risk factor or not and whether they have the disease or not
- Cross-sectional studies are the easiest to carry out, but are subject to all sorts of hidden biases

Circulatory disease and respiratory disease

- For example, one study surveyed 257 hospitalized individuals and determined whether each individual suffered from a disease of the respiratory system, a disease of the circulatory system, or both
- Their results:

		Respiratory Disease	
		Yes	No
Circulatory	Yes	7	29
Disease	No	13	208

- Could this association be due to chance?
- Not likely; $\chi^2 = 7.9$, so $p = .005$

Circulatory disease and respiratory disease (cont'd)

- Okay, so it's probably not due to chance
- But does that mean that you are more likely to get a respiratory disease if you have a circulatory disease?
- The same study surveyed nonhospitalized individuals as well:

		Respiratory Disease	
		Yes	No
Circulatory	Yes	15	142
Disease	No	189	2181

Circulatory disease and respiratory disease (cont'd)

- The evidence in favor of an association is now nonexistent: $p = .48$
- What's going on?
- The issue isn't a poor sampling design: both samples were gathered carefully and are representative of their respective populations
- Instead, the issue is that cross-sectional studies are very susceptible to selection bias

Selection bias in cross-sectional studies

- In this example, the bias was that if a patient has both a circulatory disease and a respiratory disease, then he or she is much more likely to be hospitalized and to be included in the cross-sectional study
- There are many other examples:
    - Suppose we obtained a cross-sectional sample of factory workers to see if they had developed asthma at a higher rate than non-factory workers
    - Workers who developed asthma from working in the factory may be more likely to quit their job, and less likely to be included in our sample
    - Suppose we notice an association between milk drinking and peptic ulcers
    - Is it because milk drinking causes ulcers, or because ulcer sufferers like to drink milk in order to relieve their symptoms?

Hypothesis tests and confidence intervals

- We have discussed methods for testing the null hypothesis, which for all of the aforementioned study designs, can be loosely described as saying that there is "no association" between treatment/exposure and the outcome, or more formally as the hypothesis that the two events are independent
- However, we also need to be able to measure *how dependent* the two events are, and to place confidence intervals on effect sizes – otherwise, we have no way of assessing the practical and clinical significance of the association
- However, there are several different ways of measuring the association/dependence between two events

Difference in proportions

- One way of measuring the strength of an association for categorical data is to look at the difference in proportions
- For example, in Lister's experiment, 46% of the patients who received the conventional surgery died, but only 15% of the patients who received the sterile surgery died
- The difference in these percentages is 31%

Difference in proportions and rare events

- However, differences in proportions are not informative for rare events
- For example, in a rather famous study that made front-page headlines in the *New York Times*, 0.9% of subjects taking aspirin suffered heart attacks, compared to 1.7% of placebo subjects
- The difference in proportions, 0.8%, doesn't sound front-page-of-the-*New–York–Times*–worthy

The relative risk

- Instead, for proportions, we often describe the strength of an association using ratios
- When we said that the probability of suffering a heart attack was twice as large ($1.7/0.9 = 1.9$) for the placebo group as for the aspirin group, this is much more attention-grabbing
- Similarly for Lister's experiment: the risk of dying from surgery is three times lower ($46/15 = 3.1$) if sterile technique is used
- This ratio is called the *relative risk*, and it is usually more informative than the difference of proportions

Relative risks are asymmetric

- The relative risk is a good measure of the strength of an association, but it too has some shortcomings
- One is that it's asymmetric
- For example, the relative risk of dying is $46/15 = 3.1$ times greater with the nonsterile surgery, but the relative risk of living is only $85/54 = 1.57$ times greater with the sterile surgery

Relative risks and retrospective studies

- Another shortcoming is that it doesn't work with retrospective studies
- For example, consider the results of a classic case-control study of the relationship between smoking and lung cancer published in 1950:

	Cases	Controls
Smoker	688	650
Nonsmoker	21	59

- Is the probability of developing lung cancer given that a person smoked $688/(688 + 650) = 51\%$?

Relative risks and retrospective studies (cont'd)

- Absolutely not; this isn't even remotely accurate
- By design, this study included 709 people with lung cancer and 709 without; the fact that about 50% of smokers had lung cancer doesn't mean anything
- For retrospective and cross-sectional studies, then, we cannot calculate risk, let alone relative risk
- This would require an estimate of the probability of developing a disease given that an individual was exposed to a risk factor, which we can only get from a prospective study
- Instead, retrospective studies give us the probability of being exposed to a risk factor given that you have developed the disease

Odds

- A slightly different measure of association, the *odds ratio*, gets around both of these shortcomings
- Instead of taking the ratio of the probabilities, the odds ratio is a ratio of the odds of developing the disease given risk factor exposure to the odds given a lack of exposure
- The *odds* of an event is the ratio of the number of times the event occurs to the number of times the event fails to occur: $\text{odds} = \pi/(1-\pi)$
    - For example, if the probability of an event is 50%, then the odds are 1; in speech, people usually say that "the odds are 1 to 1"
    - If the probability of an event is 75%, then the odds are 3; "the odds are 3 to 1"

The symmetry of the odds ratio

- As advertised, the odds ratio possesses the symmetry that the relative risk does not
- For example, in Lister's experiment the odds of dying were $6/34 = .176$ for the sterile group and $16/19 = .842$ for the control group
- The relative odds of dying with the control surgery is therefore $.842/.176 = 4.77$
- On the other hand, the odds of surviving were $34/6 = 5.67$ for the sterile group and $19/16 = 1.19$ for the control group
- The relative odds of surviving with the sterile surgery is therefore $5.67/1.19 = 4.77$

An easier formula for the odds ratio

- Summarizing this reasoning into a formula, if our table looks like

a	b
c	d

then

$$\widehat{OR} = \frac{ad}{bc}$$

- Because of this formula, the odds ratio was originally called the "cross-product ratio"

There are two odds ratios

- Keep in mind that there are two odds ratios, depending on how we ordered the rows and columns of the table, and that they will be reciprocals of one another
- When calculating and interpreting odds ratios, be sure you know which group has the higher odds of developing the disease
- In Lister's experiment, the odds ratio for surviving with the sterile surgery was 4.77, but the odds ratio for surviving with the control surgery was $1/4.77 = 0.210$
- NOTE: When writing about an odds ratio less than 1, it is customary to write, for example, that "the sterile procedure reduced the odds of death by 79%"

Odds ratios and retrospective studies

- The symmetry of the odds ratio works wonders when it comes to retrospective studies
- So, in our case-control study of lung cancer and smoking, the odds ratio for smoking given lung cancer is

$$\widehat{OR} = \frac{688 \cdot 59}{21 \cdot 650} = 2.97$$

- However, this is also the odds ratio for lung cancer given smoking
- This is a pretty amazing result: we have managed to obtain a prospective measure of association from a retrospective study!
- Hence the popularity of the odds ratio: it can be used for any study design (prospective, retrospective, cross-sectional) that results in a 2x2 contingency table

Interpretation of odds ratios

- We will focus on odds ratios in this course, although it is worth noting that some researchers consider relative risks more easy to interpret, and prefer reporting them when possible
- Indeed, odds ratios are always larger in magnitude (i.e., further away from 1) than relative risks, something to keep in mind when interpreting clinical significance
- For example, we saw that for the Lister study, the relative risk was 3.1, while the odds ratio was 4.8
- An even more extreme example is the Nexium trial (recall that the healing rates in the two groups were 93% vs. 89%), where the relative risk is 1.04 but the odds ratio is 1.6

The χ^2-test and measures of association

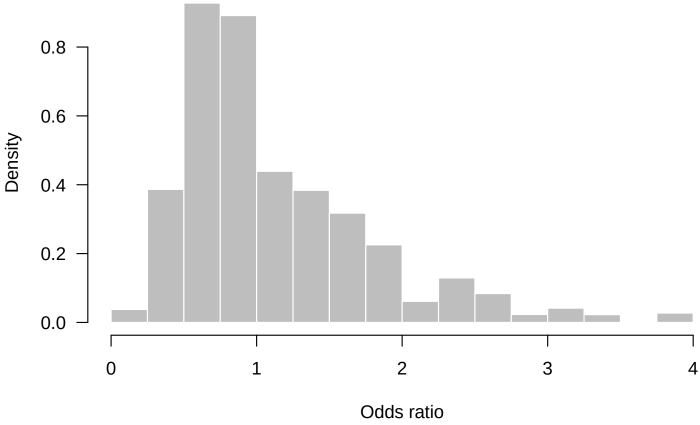
- Note that when the difference between two proportions equals 0, the relative risk equals 1 and the odds ratio equals 1
- Furthermore, when relative risk of disease given exposure equals 1, the relative risk of exposure given disease equals 1
- Indeed, all of these statements are equivalent to saying that exposure and disease are independent
- Thus, any one of these may be thought of as the null hypothesis of the χ^2-test or Fisher's Exact Test

The odds ratio and the central limit theorem

- So far in this class, we've calculated confidence intervals for quantities that involve sums and differences: averages, percentages, differences between averages
- In these cases, the central limit theorem ensures an approximately normal sampling distribution
- But we have no such guarantee for *ratios*; indeed, the sampling distribution for ratios tends to be rather skewed

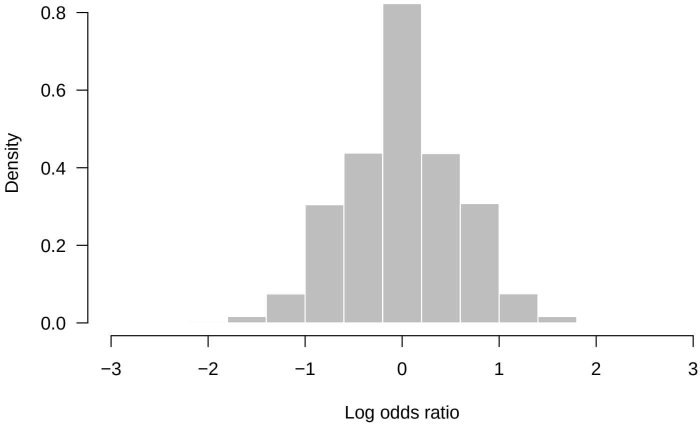
Simulation: The sampling distribution of the odds ratio

$n = 25$ per group, $\pi_1 = \pi_2 = 0.5$



The log transform

A natural solution is to consider instead the logarithm of the odds ratio, which turns the ratio into a difference:



Confidence intervals for the log odds ratio

- The log of the odds ratio is quite normal-looking and amenable to finding confidence intervals for using the central limit theorem/normal distribution
- Thus, the standard procedure for constructing approximate confidence intervals for the odds ratio actually constructs confidence intervals for the log of the odds ratio
- Getting a confidence interval for the odds ratio itself then requires an extra step of converting the confidence interval back to the odds ratio scale

Approximate distribution of the log odds ratio

- Before we consider the log of the odds ratio, let us first consider just the log odds from a single sample
- **Theorem:** Suppose $X \sim \text{Binom}(n, \pi)$, and let a and b denote the expected number of successes and failures. Then

$$\log \frac{\hat{\pi}}{1-\hat{\pi}} \dot{\sim} \text{N}\left(\log \frac{\pi}{1-\pi}, \frac{1}{a} + \frac{1}{b}\right)$$

- **Theorem:** Suppose we have data from two independent binomial samples, with expected contingency table entries a, b, c, and d, with OR denoting the true odds ratio and $\widehat{\text{OR}}$ the estimated odds ratio. Then

$$\log \widehat{\text{OR}} \dot{\sim} \text{N}\left(\log \text{OR}, \frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}\right)$$

Note: Throughout, $\log$ refers to the natural log

Confidence intervals for the odds ratio: example

With this (approximate) pivotal relationship, calculating confidence intervals for odds ratios is fairly straightforward, replacing expected counts with the observed ones; using the Lister experiment as an example:

#1 Estimate the standard error of the log odds ratio:

$$\begin{aligned} \text{SE} &= \sqrt{\frac{1}{34} + \frac{1}{6} + \frac{1}{19} + \frac{1}{16}} \\ &= 0.56 \end{aligned}$$

#2 As usual, 1.96 contains the middle 95% of the normal distribution

#3 Recall that the sample odds ratio was 4.77, so the log of the sample odds ratio is 1.56

#4 The 95% confidence interval for the log odds ratio is therefore

$$(1.56 - 1.96(0.56), 1.56 + 1.96(0.56)) = (0.47, 2.66)$$

Confidence intervals for the odds ratio: example (cont'd)

#5 The 95% confidence interval for the odds ratio is therefore

$$(e^{0.47}, e^{2.66}) = (1.60, 14.2)$$

- Note that the confidence interval doesn't include 1; this agrees with our test of significance
- Note also that this confidence interval is asymmetric (its right half is much longer than its left half) – this is as it should be, and impossible to achieve without the log transform
- Now we have an idea of the possible clinical significance of sterile technique: it may be lowering the odds of surgical death by a factor of about 1.6, or by a factor of 14, with a factor of around 5 being the most likely

Reverse example

Note also the symmetry that arises if we had decided to calculate a confidence interval for the other odds ratio (the relative odds of dying on the sterile surgery):

- $\widehat{\mathrm{OR}} = 0.21$
- $\log(\widehat{\mathrm{OR}}) = -1.56$
- 95% CI for $\log(\widehat{\mathrm{OR}})$: (-2.66, -0.47)
- 95% CI for $\widehat{\mathrm{OR}}$: (0.07, 0.62)

Summary

- In conclusion, prospective studies are the most trustworthy observational study, but like any observational study, they are subject to confounding
- Retrospective studies are often much more feasible, but potentially subject to recall bias and unrepresentative sampling
- Cross sectional studies provide a quick snapshot of an association, but need to be interpreted with care

Summary (cont'd)

- There are three natural ways to measure the association present in a 2×2 table:
  - Difference of proportions
  - Relative risk
  - Odds ratio
- One big advantage of the odds ratio is that it works equally well for both prospective and retrospective studies, unlike the other two
- Be able to construct confidence intervals for odds ratios