Lab 12: Odds ratio and relative risk

STAT218

This lab has two main objectives:

1. learn to estimate odds ratios and relative risk in R using epitools
2. distinguish which measure of association is most appropriate based on desired interpretation and/or study design

In addition, you’ll get some practice combining inference of association in two-way tables with inference for an appropriate measure of association and interpreting results correctly.

We’ll use the datasets from class, along with a few additional examples.

library(tidyverse)  
library(epitools)  
load('data/smoking.RData')  
load('data/asthma.RData')  
load('data/chd.RData')  
load('data/outbreak.RData')  
malaria <- openintro::malaria

### Odds ratios

It is worth underscoring at the outset that odds ratios can be estimated for any study design; they are not reserved strictly for case-control studies.

That said, the smoking dataset is an example of a situation in which we can *only* estimate odds ratios, because of the case-control study design: 86 lung cancer patients and 86 controls were sampled separately, and the smoking status of each participant was recorded.

Notice below how adding names when constructing the contingency table adds labels to the row and column dimensions in the output.

# construct table with dimension names group, smoking  
smoking.tbl <- table(group = smoking$group, smoking = smoking$smoking)  
smoking.tbl

smoking  
group Smokers NonSmokers  
 cancer 83 3  
 control 72 14

It is good practice to check whether expected counts are sufficiently large to use the test; if they aren’t, then we’ll know to use the Fisher’s exact -value later.

# check whether expected counts are at least ten   
chisq.test(smoking.tbl)$expected

smoking  
group Smokers NonSmokers  
 cancer 77.5 8.5  
 control 77.5 8.5

Last time, we bent the rules a little and went ahead with the test anyway because the counts were close enough to the cutoff. So this time, it’ll be interesting to see how much results differ using the exact inference method.

To compute odds ratios *and* inference for association simultaneously, use the oddsratio(...) function:

oddsratio(smoking.tbl, rev = 'both', conf.level = 0.95, method = 'wald', correction = T)

$data  
 smoking  
group NonSmokers Smokers Total  
 control 14 72 86  
 cancer 3 83 86  
 Total 17 155 172  
  
$measure  
 odds ratio with 95% C.I.  
group estimate lower upper  
 control 1.00000 NA NA  
 cancer 5.37963 1.486376 19.47045  
  
$p.value  
 two-sided  
group midp.exact fisher.exact chi.square  
 control NA NA NA  
 cancer 0.005116319 0.008822805 0.01062183  
  
$correction  
[1] TRUE  
  
attr(,"method")  
[1] "Unconditional MLE & normal approximation (Wald) CI"

This command has several moving parts:

* rev rearranges the table by **rev**ersing rows, columns, or both
* conf.level determines the confidence level of the resulting interval
* method determines how the interval is computed (the version we learned in class is a Wald approximation)
* correction determines whether a continuity correction is applied (always set to T)

The most important argument is rev: the contingency table must be arranged so that the outcome of interest is in the second position in the rows/columns.

* if the order of columns were opposite what is shown above, we would get instead an estimate of the odds of cancer among nonsmokers compared with smokers
* if the order of rows were opposite what is shown above, we would get instead an estimate of the odds of *not* having cancer among smokers compared with nonsmokers
* if both were opposite what is shown above, we would get instead an estimate of the odds of *not* having cancer among nonsmokers compared with smokers

It may be a good idea to double-check that the odds ratio you obtained is in fact the right one by doing the calculations by hand and checking the point estimate.

|  |
| --- |
| Your turn 1 |
| Compute the odds ratio using direct arithmetic to double-check that the output above gives you the odds ratio you intend. |

In this example, the exact -value is fairly close to the test; they would produce different conclusions at the 1% significance level, but not otherwise. Either would be appropriate to use here, but if you prefer to stick to the rule of thumb strictly, use Fisher’s exact test:

The data provide evidence that smoking is associated with lung cancer (Fisher’s exact test, *p* = 0.0088). With 95% confidence, the odds of lung cancer are estimated to be between 1.49 and 19.47 times higher among smokers compared with nonsmokers, with a point estimate of 5.38.

|  |
| --- |
| Your turn 2 |
| Using the outbreak data, which comprise data from a case-control study in which 30 cases and 60 controls were sampled and each subject’s exposure to raspberries was assessed, perform inference on association using the odds ratio:   1. Check assumptions for the chi square test 2. Use oddsratio(...) to perform calculations for the inference with the odds ratio 3. Interpret relevant outputs in context   # construct contingency table  # check assumptions for chi square test  # perform inference with odds ratio  # double-check your point estimate to verify data were arranged in correct orientation |

### Relative risk

The implementation for inference with relative risk is identical to that for inference with odds ratios, but a bit more care is required to orient the contingency table correctly. To obtain the correct relative risk, **it is essential to put the groups in rows and outcomes in columns**. This orientation did not matter for the odds ratio implementation.

# put groups in rows and outcome in columns  
asthma.tbl <- table(sex = asthma$sex, asthma = asthma$asthma)  
asthma.tbl

asthma  
sex asthma no asthma  
 female 49 781  
 male 30 769

# check assumptions for chi square test  
chisq.test(asthma.tbl)$expected

asthma  
sex asthma no asthma  
 female 40.25169 789.7483  
 male 38.74831 760.2517

Here, all expected counts are above 10, so when we interpret the output of riskratio(...), we can use the *p*-value from the test.

riskratio(asthma.tbl, rev = 'both', conf.level = 0.90, method = 'wald', correction = T)

$data  
 asthma  
sex no asthma asthma Total  
 male 769 30 799  
 female 781 49 830  
 Total 1550 79 1629  
  
$measure  
 risk ratio with 90% C.I.  
sex estimate lower upper  
 male 1.000000 NA NA  
 female 1.572329 1.083353 2.282007  
  
$p.value  
 two-sided  
sex midp.exact fisher.exact chi.square  
 male NA NA NA  
 female 0.04412095 0.04961711 0.05703135  
  
$correction  
[1] TRUE  
  
attr(,"method")  
[1] "Unconditional MLE & normal approximation (Wald) CI"

Here, the consequences of reversing the order of rows/columns are as follows:

* if rows were arranged opposite of what is shown above, we would obtain the relative risk of asthma among men compared with women (sensible but different)
* if columns were arranged opposite of what is shown above, we would obtain the relative risk of *not* having asthma among women compared with men (not sensible)
* if both rows and columns were arranged opposite of what is shown above, we would obtain the relative risk of *not* having asthma amond men compared with women (not sensible)

In addition, the consequence of orienting the table opposite is:

* if columns and rows were swapped, but the order were the same as shown above, we would obtain an estimate of the relative risk of being a woman among asthmatics compared with non-asthmatics (not sensible)

You can double check that the risk ratio computed is the one intended by direct arithmetic:

# compute proportions  
asthma.tbl |> prop.table(margin = 1)

asthma  
sex asthma no asthma  
 female 0.05903614 0.94096386  
 male 0.03754693 0.96245307

# risk ratio by hand  
0.05903614/0.03754693

[1] 1.572329

Since assumptions for the test were met, we can combine the inference from this test with the interval estimate for the relative risk:

The data provide evidence at the 10% significance level of an association between asthma and sex ( = 3.62 on 1 degree of freedom, *p* = 0.057). With 90% confidence, the risk of asthma is estimated to be betwen 1.08 and 2.28 times greater for women than for men, with a point estimate of 1.57.

|  |
| --- |
| Your turn 3 |
| The chd data contain observations on the incidence of coronary heart disease from a cohort study of 3000 smokers and 5000 nonsmokers.   1. Construct the contingency table with the groups you wish to compare shown in the row dimension and the outcome of interest in the column dimension. 2. Check assumptions for the test. 3. Perform inference on association with relative risk at the 1% significance level. 4. Double-check the relative risk point estimate to make sure you obtained the comparison you intended.   # construct contingency table in proper orientation for inference with relative risk  # check assumptions for chi square test  # carry out inference at 1% significance level  # double check point estimate by manual calculation |

### Practice problems

1. [L8] For the diabetes\_meds dataset from last time comparing rates of cardiovascular problems between two diabetes medications among ~200K medicare beneficiaries, determine an appropriate measure of association to add to the inference you performed previously. Provide a narrative interpretation of the result (both the test and estimates) following the style introduced in class.
2. [L8] The Learning Early About Peanut allergy (LEAP) study recruited 530 children with risk factors for developing peanut allergies and randomly allocated peanut exposure and peanut avoidance regimens to each participant. At 5 years of age, an oral food challenge (OFC) test was administered to determine whether participants had developed allergies. Data are contained in the leap dataset.
   1. Construct the contingency table for this data.
   2. Check the assumptions for the test of association.
   3. Test for association at the 1% significance level and provide point and interval estimates for an appropriate measure of association.
3. [L8] Researchers studying the link between prenatal vitamin use and autism surveyed the mothers of a random sample of children aged 24 - 60 months with autism and seperately surveyed the mothers of a random sample of children with typical development. The vitamin dataset contains observations of whether mothers in each group did or did not use prenatal vitamins during the three months before pregnancy (periconceptional period).
   1. Which proportions are possible to estimate? Based on the study design, is it possible to estimate the relative risk of autism?
   2. Construct the contingency table and check assumptions for the test.
   3. Test for an association between taking prenatal vitamins and autism at the 1% significance level; include an appropriate meausre of association and provide point and interval estimates.
   4. Interpret the result in context following the narrative style introduced in class.