

Tutorial: Contingency Tables

A well-known statistician once said, “A PhD student could write an entire dissertation on two-by-two tables only.” Continuing our health disparities research, we now consider the odds ratio and the Pearson Chi-square test.

Exercises

1. Using data from the 500 respondents of the CHIS survey, construct a 2x2 table comparing poverty level versus past doctor visit. Display the row frequencies and the expected cell counts.

```
. tabulate poverty doctor, row expected
```

2. Construct the odds ratio and corresponding 95% confidence interval (CI) for the visiting the doctor in the past 12 months for those above and below the poverty line.

```
. gen nopov = 1-pov
```

```
. cs doctor nopov, or woolf
```

Notice that the Woolf option is used, denoting that we want standard errors calculated using the formula presented in class.

3. Conduct a Pearson's chi-square test to examine the association between poverty and prior doctor visit.

```
. cs doctor poverty, or woolf
```

OR use tabulate...

```
. tabulate poverty doctor, expected chi2
```

Note that tabulate extends nicely to R x C tables...

```
. tabulate racecat doctor, expected chi2
```

- What are the null and alternative hypotheses?

Null: no association between above/below poverty line and whether an individual visited the doctor in the past 12 months.

Alternative: there is an association.

Null: $OR = 1$

Alternative: $OR \neq 1$

- Are the expected cell counts sufficiently large?

All expected cell counts are greater than 5.

- What is the value and distribution of the test statistic under the null hypothesis?

$X = 5.1 \sim \chi_1^2$

- What is the p-value?

$p = 0.024$.

- Do you reject the null hypothesis? What is your conclusion?

We reject the null hypothesis and conclude that there is evidence in the data that the odds of visiting the doctor in the past 12 months are higher in those who are above the poverty line.

4. For the above and below poverty groups, compare the following:

- 95% CI for the odds ratio
- 95% CI for the difference in two proportions (from the previous tutorial)
- the p-value from the two-sample proportion test (from the previous tutorial)
- the p-value from the Pearson Chi-square test

(a) Do you get the same general conclusion with each test?

- Difference in proportions between above and below poverty groups:

12.0 with 95% CI (0.2, 24.0)

- Odds ratio for above and below groups

2.0 with 95% CI (1.1, 3.5)

- Pearson Chi-square

$p = 0.024$

- Two sample proportion test

$p = 0.024$

(b) Which test do you find most useful?

It's always good to show both a p-value and a confidence interval. (Note that you can't show a confidence interval for the risk difference if you have a case-control study!)