Inference for Two-Way Tables

Chapter 8, Lab 2

OpenIntro Biostatistics

Topics

- The χ^2 test for independence
- Measures of association in two-by-two tables

This lab generalizes inference for binomial proportions to the setting of two-way contingency tables. Hypothesis testing in a two-way table assesses whether the two variables of interest are associated; this approach can be applied to settings with two or more groups and for responses that have two or more categories. Measures of association in two-by-two tables are also discussed.

The material in this lab corresponds to Sections 8.3 and 8.5 in *OpenIntro Biostatistics*.

Introduction

The χ^2 test of independence

In the χ^2 test of independence, the observed number of cell counts are compared to the number of **expected** cell counts, where the expected counts are calculated under the null hypothesis.

- H_0: the row and column variables are not associated
- H_A: the row and column variables are associated

The expected count for the i^{th} row and j^{th} column is

$$E_{i,j} = \frac{(\text{row } i \text{ total}) \times (\text{column } j \text{ total})}{n},$$

where n is the total number of observations.

The χ^2 **test statistic** is calculated as

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c} \frac{(E_{i,j} - O_{i,j})^2}{E_{i,j}},$$

and is approximately distributed χ^2 with degrees of freedom $(r-1)(c-1)$, where r is the number of rows and c is the number of columns. $O_{i,j}$ represents the observed count in row i, column j.

For each cell in a table, the **residual** equals

$$\frac{O_{i,j} - E_{i,j}}{\sqrt{E_{i,j}}}.$$

Residuals with a large magnitude contribute the most to the χ^2 statistic. If a residual is positive, the observed value is greater than the expected value; if a residual is negative, the observed value is less than the expected.

Measures of association in two-by-two tables

Unit 1 introduced the **relative risk (RR)**, a measure of the risk of a certain event occurring in one group relative to the risk of the event occurring in another group, as a numerical summary for two-by-two (2×2) table. The relative risk can also be thought of as a measure of association for 2×2 tables.

Consider the following hypothetical two-by-two table. The relative risk of Outcome A can be calculated by using either Group 1 or Group 2 as the reference group:

	Outcome A	Outcome B	Sum
Group 1	a	b	$a+b$
Group 2	c	d	$c+d$
Sum	$a+c$	$b+d$	$a+b+c+d=n$

Table 1: A hypothetical two-by-two table of outcome by group.

$$RR_{\text{A, comparing Group 1 to Group 2}} = \frac{a/(a+b)}{c/(c+d)}$$

$$RR_{\text{A, comparing Group 2 to Group 1}} = \frac{c/(c+d)}{a/(a+b)}$$

The relative risk should only be calculated for tables where the proportions $a/(a+b)$ and $c/(c+d)$ represent the incidence of Outcome A within the populations from which Groups 1 and 2 are sampled.

The **odds ratio (OR)** is a measure of association that remains valid even when it is not possible to estimate incidence of an outcome from the sample data. The **odds** of Outcome A in Group 1 are a/b, while the odds of Outcome A in Group 2 are c/d.

$$OR_{\text{A, comparing Group 1 to Group 2}} = \frac{a/b}{c/d} = \frac{ad}{bc}$$

$$OR_{\text{A, comparing Group 2 to Group 1}} = \frac{c/d}{a/b} = \frac{bc}{ad}$$

The χ^2 test of independence

Measures of association: the odds ratio