BIOS 6244 Analysis of Categorical Data
October 3, 2005 Lecture

Comparing Two Binomial Proportions (Section 2.2, pp. 19-22)

Difference of Proportions (Sometimes called the risk difference.)

Suppose we wish to test the hypotheses H_0: $\pi_1 = \pi_2$ vs. H_a: $\pi_1 \neq \pi_2$ and find a 95% CI $(\pi_1 - \pi_2)$

where π_1 = probability of "success" in Population 1
π_2 = probability of "success" in Population 2.

We can express this in the form of a 2x2 table, where rows denote the population and columns denote "success" or "failure," i.e.,

		success	failure	
population	1	π_{11}		π_{1+}
	2	π_{21}		π_{2+}

Then,

π_1 = Pr(success | Population 1) = π_{11} / π_{1+} (estimated by $p_1 = \frac{n_{11}}{n_{1+}}$) and

π_2 = Pr(success | Population 2) = π_{21} / π_{2+} (estimated by $p_2 = \frac{n_{21}}{n_{2+}}$).

Note that we can also re-express the hypotheses we wish to test as:

H_0: $\pi_1 - \pi_2 = 0$ vs. H_a: $\pi_1 - \pi_2 \neq 0$.

Usual Approximate Method for Finding CI $(\pi_1 - \pi_2)$:

We use $p_1 - p_2$ to estimate $\pi_1 - \pi_2$ and then use the asymptotic distribution of $p_1 - p_2$ to find a CI$(\pi_1 - \pi_2)$.

Following our general method for finding a CI for any parameter θ, an approximate 95% CI(π_1 - π_2) is given by:

$$(p_1 - p_2) \pm 1.96\ SE(p_1 - p_2),$$

where

$$SE(p_1 - p_2) = \sqrt{\frac{p_1(1-p_1)}{N_1} + \frac{p_2(1-p_2)}{N_2}}$$

$$N_1 = n_{1+}$$

and

$$N_2 = n_{2+}.$$

For our example,

$$p_1 = \frac{37}{77} = .481$$

$$p_2 = \frac{212}{350} = .606\,.$$

$$SE(p_1 - p_2) = .063$$

and an approximate 95% CI(π_1 - π_2) is given by (-.248, -.002).

Usual Approximate Method for testing H_0: $\pi_1 - \pi_2 = 0$ vs. H_a: $\pi_1 - \pi_2 \neq 0$.

Under the null hypothesis, the best estimate of $SE(p_1 - p_2)$ is

$$SE(p_1 - p_2) = \sqrt{\frac{\bar{p}\,(1-\bar{p})}{N_1} + \frac{\bar{p}\,(1-\bar{p})}{N_2}} = \sqrt{\bar{p}\,(1-\bar{p})\left[\frac{1}{N_1} + \frac{1}{N_2}\right]}$$

where $\bar{p}$ is the "pooled" estimate of the true common proportion π assuming that H_0: $\pi_1 = \pi_2 = \pi$ is true:

$$\bar{p} = \frac{n_{11} + n_{21}}{N} = \frac{n_{+1}}{N} = \frac{total\ \#\ of\ "successes"}{sample\ size}\,.$$

So an approximate test of H_0: $\pi_1 - \pi_2 = 0$ can be performed by calculating

$$z = \frac{p_1 - p_2}{\sqrt{\bar{p}\,(1-\bar{p})\left(\frac{1}{N_1}+\frac{1}{N_2}\right)}}$$

and using the standard normal distribution to calculate a p-value.

Note: It can be shown that this approximate test is equivalent to the Pearson χ^2 test for independence. (See pp. 27-31 of our text.)

For our sample data,

$$\bar{p} = \frac{37+212}{427} = .583138$$

$$SE(p_1 - p_2) = \sqrt{.583138(1-.583138)\left[\frac{1}{77}+\frac{1}{350}\right]} = .062061$$

$$\Rightarrow \quad z = \frac{.480519 - .605714}{.062061} = -2.017$$

$$\Rightarrow \text{p-value} \approx 2\Pr(Z \le -2.017) = .044 \text{ (by SimCalc)}$$

Consistency of Confidence Interval and Hypothesis Test Results

As we saw in the case of testing a single proportion, we could run into trouble since the formulas for an approximate CI(π_1 - π_2) and for the test statistic for the approximate test of H_0: $\pi_1 - \pi_2 = 0$ contain different expressions for $SE(p_1 - p_2)$, thereby creating the potential for the conclusion from the confidence interval to be inconsistent with the conclusion from the hypothesis test. For our example, the two conclusions *do* agree:

approximate 95% CI(π_1 - π_2) = (-.248, -.002) $\Rightarrow$ Reject H_0 since $0 \notin$ CI(π_1 - π_2)

approximate p-value = .044 $\Rightarrow$ Reject H_0 since p-value < .05.

We will discuss an exact method for testing H_0: $\pi_1 = \pi_2$ in Section 2.6, pp. 39-41.

See also the example in our textbook: Section 2.2.2, pp. 20-21.

Relative Risk (pp. 21-22)

In traditional statistical methods courses, techniques for comparing 2 population proportions tend to focus on the risk difference, i.e., for hypothesis testing, if we are interested in the null hypothesis H_0: $\pi_1 = \pi_2$, we convert this to H_0: $\pi_1 - \pi_2 = 0$, and we find a CI ($\pi_1 - \pi_2$).

However, in biostatistics courses, it is more common to consider the *ratio* of the 2 proportions, i.e., we convert H_0: $\pi_1 = \pi_2$ to $H_0 : \frac{\pi_1}{\pi_2} = 1$ and find a 95% CI $\left(\frac{\pi_1}{\pi_2}\right)$.

The parameter $\theta = \frac{\pi_1}{\pi_2}$ is called the *population relative risk*, and is usually denoted by *RR*.

It is particularly helpful to consider the relative risk instead of the risk difference when the proportions π_1 and π_2 are closer to .5 than when they are closer to 0 or 1. For example, suppose that, in a clinical trial, π_1 and π_2 denote the proportion of study subjects who have adverse reactions to a newly developed drug and to the standard drug, respectively. Suppose that $\pi_1 = .010$ and $\pi_2 = .001$ for adverse reaction A, yielding a risk difference of .009. Suppose that $\pi_1 = .410$ and $\pi_2 = .401$ for adverse reaction B; this also yields a risk difference of .009. Does this mean that the disparity in adverse reactions for the two drugs should be treated as equivalent for A and B? Note that the relative risks are quite different: $RR = \frac{.010}{.001} = 10$ for adverse reaction A, and $RR = \frac{.410}{.401} = 1.02$ for adverse reaction B. This means that subjects taking the new drug are 10 times more likely to have adverse reaction A than are subjects taking the standard drug. However, for adverse reaction B, the risk is essentially the same for the 2 drugs. Therefore, adverse reaction A would be considered to be a potential problem for the new drug, whereas adverse reaction B would not. In this example, the relative risk provides a better description of the *clinical significance* (as opposed to statistical significance) of the relative importance of the two adverse reactions.

Note that the *null value* is different depending on whether the risk difference or the relative risk is used: for the risk difference, the null value is 0, whereas for the relative risk, the null value is 1.

Suppose now that we have samples of size N_1 from Population 1 and size N_2 from Population 2. Denote the sample proportions of "success" by p_1 & p_2, respectively. Then the *sample relative risk* is given by

$$\widehat{RR} = \frac{p_1}{p_2}.$$

The sampling distribution of $\frac{p_1}{p_2}$ can be highly skewed, and deriving a formula for

$\text{CI}\left(\frac{\pi_1}{\pi_2}\right)$ is difficult. The most common approximate method is based on a log transformation of $\frac{p_1}{p_2}$ (Exercise 2.12, pp. 47-48).

Example, cont.

For the data in our example, $\widehat{RR} = \frac{p_1}{p_2} = \frac{.481}{.606} = .79$ and an approximate 95% $\text{CI}\left(\frac{\pi_1}{\pi_2}\right)$ = (.62, 1.02). Note that this indicates that H_0: $\pi_1 = \pi_2$ should not be rejected since $1 \in \text{CI}\left(\frac{\pi_1}{\pi_2}\right)$.

Therefore, we reach different conclusions depending on whether we examine a $\text{CI}(\pi_1 - \pi_2)$ or a $\text{CI}\left(\frac{\pi_1}{\pi_2}\right)$. In the former case, we rejected H_0: $\pi_1 = \pi_2$ and in the latter, we failed to reject H_0: $\pi_1 = \pi_2$. The choice of whether to test H_0: $\pi_1 - \pi_2 = 0$ or $H_0 : \frac{\pi_1}{\pi_2} = 1$ depends on the purpose of the test. If one is comparing two exposures or treatments in terms of a particular outcome (e.g., occurrence of a tumor), then $RR = \frac{\pi_1}{\pi_2}$ is the appropriate parameter to use. If one simply wants to compare two groups in terms of some attribute (e.g., compare Biostatistics and Epidemiology graduate students in terms of gender), then the difference of proportions $\pi_1 - \pi_2$ would be the appropriate parameter.

Public Health Implications (pp. 21-22)

Our textbook illustrates the difference between using the risk difference vs. using the relative risk with data from the Physician's Aspirin Study. (See course website for a link to this study.) This is a very famous clinical trial that examined the beneficial effects of taking low-dose aspirin to prevent adverse cardiac events.

In Table 2.3, p. 20, study data related to 5-year risk of MI (heart attack) are presented. A 95% $\text{CI}(\pi_1 - \pi_2)$ = (.005, .011), indicating an increased risk of at most 1.1% among those not taking aspirin.

On the other hand, a 95% $\text{CI}\left(\frac{\pi_1}{\pi_2}\right)$ = (1.43, 2.30), indicating that the risk of MI is increased by at least 43% if you do not take aspirin and possibly by as much as 130%. Findings such as these can have tremendous public health impact in terms of their potential effects on physician behavior and health outcomes, i.e., what will be the effect if

physicians routinely recommend that their patients take aspirin in order to prevent MI's? The Physician's Aspirin Study also found a possible *increase* in the risk of a stroke among those study subjects taking aspirin, so the potential impact of this would have to be considered as well.