Relative Risks and Odds Ratios

Statistics 149

Spring 2006

Relative Risk

Using $\pi_1 - \pi_2$ as a measure of how different two groups are isn't always a good measure. For example, a difference of 0.05 means something different for the cases

π_1	0.5	0.1	0.051
π_2	0.45	0.05	0.01

Another measure that people find useful, particularly for small π_1 and π_2, is the relative risk

$$RR = \frac{\pi_2}{\pi_1}$$

This often matches better with the way people think about small proportions.

Example: Incidence of Rhabdomyolysis and Lipid-Lowering Drugs

(JAMA, December 1, 2004 – Vol 292, No. 21, pages 2585-2590)

Drug	n	Rhabdomyolysis Cases	$\hat{p}$
Cerivastatin (Baycol)	12695	10	0.000788
Atorvastatin (Lipitor)	130865	8	0.000061

$$\widehat{RR} = \frac{0.000788}{0.000061} = 12.89$$

$$\hat{\pi}_1 - \hat{\pi}_2 = 0.000788 - 0.000061 = 0.000727$$

While the absolute difference between the two probabilities is small, it's not really the correct scale to be describing the difference.

The ratio of almost 13 times is a better description of the increased risk of problems with Cerivastatin.

Note that Cerivastatin was voluntarily removed from the market by Bayer in August 2001 for due to reports of fatal cases of Rhabdomyolysis (a severe muscle reaction to the drug).

Confidence Interval for RR

The sampling distribution of the usual estimate of RR,

$$\widehat{RR} = \frac{\hat{\pi}_2}{\hat{\pi}_1}$$

usually is not well approximated by a normal distribution. So

$$\widehat{RR} \pm z^*_{\alpha/2} SE(\widehat{RR})$$

will not work well as a confidence for RR.

However the sampling distribution of

$$\log \widehat{RR}$$

is approximately normally distributed. So we can base a confidence interval on this quantity instead.

The standard error of $\log \widehat{RR}$, as an estimate of $\log RR$, is

$$SE(\log \widehat{RR}) = \sqrt{\frac{1}{n_1 \hat{\pi}_1} - \frac{1}{n_1} + \frac{1}{n_2 \hat{\pi}_2} - \frac{1}{n_2}}$$

$$= \sqrt{\frac{1}{S_1} - \frac{1}{n_1} + \frac{1}{S_2} - \frac{1}{n_2}}$$

giving a 100(1-α)% CI for RR of

$$CI(\log(RR)) = \log(\widehat{RR}) \pm z^*_{\alpha/2} \sqrt{\frac{1}{n_1 \hat{\pi}_1} - \frac{1}{n_1} + \frac{1}{n_2 \hat{\pi}_2} - \frac{1}{n_2}} = (L, U)$$

From this we can get the confidence interval for RR of

$$CI(RR) = (e^L, e^U)$$

So for the lipid drug example

$$\hat{\pi}_1 = \frac{10}{12695} = 0.000788 \quad \hat{\pi}_2 = \frac{8}{130865} = 0.000061$$

$$\widehat{RR} = \frac{0.000788}{0.000061} = 12.89 \quad \log RR = \log 12.89 = 2.556$$

$$SE(\widehat{RR}) = \sqrt{\frac{1}{10} - \frac{1}{12695} + \frac{1}{8} - \frac{1}{130865}} = 0.4742$$

A 95% CI for $\log RR$ is

$$2.556 \pm 1.96 \times 0.4742 = (1.627, 3.486)$$

Finally, a 95% CI for RR is

$$(e^{1.627}, e^{3.486}) = (5.086, 32.643)$$

Note that that this interval is not symmetric around RR. This is reasonable in this case as RR must have a skewed distribution. For statistics with skewed distributions, usually confidence intervals based on them with not be symmetric about the observed statistic.

Confidence Intervals and Transformations

Question: Is the transformation trick done earlier to get the CI for RR valid?

Answer: Yes.

Suppose we have a valid confidence interval procedure for a parameter θ and we want to get a confidence interval procedure for $f(\theta)$, where $f(\cdot)$ is a strictly monotonic function (i.e. increasing or decreasing). For simplicity lets assume that $f(\cdot)$ is an increasing function.

Now suppose that the true parameter value for a particular problem is θ_0 and lets consider all data sets that include this value in the interval (L, U), i.e,

$$L \leq \theta_0 \leq U$$

Since f is an increasing function,

$$f(L) \leq f(\theta_0) \leq f(U)$$

i.e. $f(\theta_0)$ is in the interval $(f(L), f(U))$.

So if the procedure that generates intervals (L, U) for θ_0 has confidence level $(1 - \alpha)$, the intervals $(f(L), f(U))$ for $f(\theta_0)$ must have the same confidence level.

The proposed interval for RR satisfies this as $\exp$ is an increasing function.

So how well this interval works depends on the approximately normality of $\widehat{RR}$, which isn't that good.

Switching Failures and Successes

In the examples discussed so far, we could have looked at the failure rates (the φs) instead of the success rates (the πs). Lets do that for a new example

Example: Infant mortality in New York City in 1974

Looked at one year death rates

Birthweight	Dead	Alive	Total
$\leq$ 2500 grams	530	4340	4870
$>$ 2500 grams	333	32637	32970

What is the relative risk of death?

$$\leq 2500 \text{ grams} \quad : \quad \hat{\pi}_1 = \frac{530}{4870} = 0.1088$$

$$> 2500 \text{ grams} \quad : \quad \hat{\pi}_2 = \frac{333}{32970} = 0.0101$$

$$\widehat{RR}_{death} = \frac{0.1088}{0.0101} = 10.78$$

So the risk of death for low birthweight babies is almost 11 times higher than the risk for normal birthweight babies.

What if we look at the chance of being alive

$$\leq 2500 \text{ grams} \quad : \quad \hat{\varphi}_1 = \frac{4340}{4870} = 0.8912 = 1 - \hat{\pi}_1$$

$$> 2500 \text{ grams} \quad : \quad \hat{\varphi}_2 = \frac{32337}{32970} = 0.9899 = 1 - \hat{\pi}_2$$

$$\widehat{RR}_{alive} = \frac{0.8912}{0.9899} = 0.9086$$

Note that the RR based on the φ's in not a simple function of the RR based on the π's.

$$RR_s = \frac{\pi_2}{\pi_1} \qquad RR_f = \frac{\varphi_2}{\varphi_1} = \frac{1 - \pi_2}{1 - \pi_1}$$

i.e. no $g(\cdot)$ where $g(RR_s) = RR_f$.

In fact, it can be shown that if $RR_s = c$, then

- $c < 1$: $RR_f \in (1, \infty)$

- $c > 1$: $RR_f \in (0, 1)$

The difference in proportions works much nicer. Its easy to show that

$$\varphi_1 - \varphi_2 = (1 - \pi_1) - (1 - \pi_2) = -(\pi_1 - \pi_2)$$

In addition, all the inference procedures discussed last time transform the same way

- $SE(\hat{\varphi}_1 - \hat{\varphi}_2) = SE(\hat{\pi}_1 - \hat{\pi}_2)$

- $CI(\varphi_1 - \varphi_2) = -CI(\pi_1 - \pi_2)$

- $z(\varphi_1 - \varphi_2) = -z(\pi_1 - \pi_2)$

Because of this problem with the relative risk plus the poor distributional results, the relative risk isn't the most popular measure

Odds and Odds Ratio

Another measure to describe probability that is commonly used is the odds of an event

$$\omega = \frac{\pi}{1 - \pi}$$

Odds have the following properties

1. $\omega \in (0, \infty)$

2. $\pi = \frac{1}{2} \Longleftrightarrow \omega = 1$

3. If the odds of a success are ω, then the odds of a failure are $\frac{1}{\omega}$

4. If the odds of a success are ω, the probability of a success is

$$\pi = \frac{\omega}{1 + \omega}$$

Since there is a 1-1 relationship between odds and probabilities, instead of making statements about probabilities, we can make statements about odds.

Consider the situation where two populations of interest have success probabilities π_1 and π_2. The odds ratio is defined as

$$\phi = \frac{\omega_2}{\omega_1} = \frac{\frac{\pi_2}{1-\pi_2}}{\frac{\pi_1}{1-\pi_1}} = \frac{\pi_2}{\pi_1} \times \frac{1-\pi_1}{1-\pi_2}$$

The odds ratio acts like relative risk, e.g.

$$RR = 10 \Leftrightarrow \pi_2 = 10\pi_1$$

vs

$$\phi = 10 \Leftrightarrow \omega_2 = 10\omega_1$$

In addition, if π_1 and π_2 are small, then

$$\phi \approx RR$$

This is one reason why people look at the odds ratio. Its similar to RR and the distributional properties of its estimator are nicer.

There is another motivation for looking at odds. Consider the binomial density function

$$
\begin{aligned}
P[S = k] &= \binom{n}{k} \pi^k (1 - \pi)^{n-k} \\
&= \binom{n}{k} (1 - \pi)^n \left(\frac{\pi}{1 - \pi} \right)^k \\
&= \binom{n}{k} \left(\frac{1}{1 + \omega} \right)^n \omega^k
\end{aligned}
$$

So the odds is a natural parameter of the binomial distribution. Actually $\log \omega$ is the canonical parameter of the binomial (which we will talk about when we get to talking about the exponential family).

One important relationship between odds and probabilities is

$\pi_2 - \pi_1$	$= 0$	< 0	> 0
$\phi = \omega_2/\omega_1$	$= 1$	< 1	> 1

(You're asked to justify this in question 3 of the assignment.)

So we can write the hypotheses as

$$H_0 : \pi_2 - \pi_1 = 0 \quad \text{vs} \quad H_A : \pi_2 - \pi_1 \neq 0$$

or as

$$H_0 : \phi = 1 \quad \text{vs} \quad H_A : \phi \neq 1$$

Similarly we get for the one-sided hypotheses

$$H_0 : \pi_2 - \pi_1 = 0 \text{ vs } H_A : \pi_2 - \pi_1 < 0 \quad \Longleftrightarrow \quad H_0 : \phi = 1 \text{ vs } H_A : \phi < 1$$

$$H_0 : \pi_2 - \pi_1 = 0 \text{ vs } H_A : \pi_2 - \pi_1 > 0 \quad \Longleftrightarrow \quad H_0 : \phi = 1 \text{ vs } H_A : \phi > 1$$

There are four other reasons why the odds ratio is a useful measure for comparing population probabilities

1. In practice, the odds ratio tends to remain more nearly constant over levels of confounding variables.

   Note that this is an empirical result and will not hold in some examples.

2. The odds ratio is the only parameter that can be used to compare two groups of binary responses from retrospective studies.

3. The comparison of odds extends nicely to regression analysis (e.g. Logistic regression)

4. It really doesn't make a difference whether we count successes or failures

$$\phi_f = \frac{\varpi_2}{\varpi_1}$$

$$= \frac{\varphi_2}{\varphi_1} \times \frac{1 - \varphi_1}{1 - \varphi_2}$$

$$= \frac{1 - \pi_2}{1 - \pi_1} \times \frac{\pi_1}{\pi_2}$$

$$= \frac{\omega_1}{\omega_2} = \frac{1}{\phi_s}$$

where $\varpi_i = \frac{1}{\omega_i}$ is the odds of failure in group i.

Will we use this is a slightly different form sometimes

$$\log \phi_f = \log \varpi_2 - \log \varpi_1 = -\log \phi_s = -(\log \omega_2 - \log \omega_1)$$

Inference on the Odds Ratio

The usual estimate of the odds is

$$\hat{\omega} = \frac{\hat{\pi}}{1 - \hat{\pi}} = \frac{S}{n - S} = \frac{S}{F}$$

This gives an estimate of the odds ratio of

	Success	Failure
Group 2	S_2	F_2
Group 1	S_1	F_1

$$\hat{\phi} = \frac{\hat{\omega}_2}{\hat{\omega}_1}$$

$$= \frac{S_2(n_1 - S_1)}{S_1(n_2 - S_2)}$$

$$= \frac{S_2 \times F_1}{S_1 \times F_2}$$

So for the NY infant death example

Birthweight	Dead	Alive	Total
$\leq$ 2500 grams	530	4340	4870
> 2500 grams	333	32637	32970

$$\hat{\phi} = \frac{530 \times 4340}{333 \times 32637} = 11.967$$

So the odds of death in for low birth weight babies is almost 12 times the odds for normal birth weight babies.

For comparison, recall that $RR_{death} = 10.78$.

For the lipid drugs example

$$\hat{\phi} = 12.895 \qquad \widehat{RR} = 12.885$$

Drug	Rhabdomyolysis	n
Baycol	8	12695
Lipitor	10	130865

As with the sampling distribution of RR, $\hat{\phi}$ is not well approximated by a normal, unless n_1 and n_2 are extremely large.

However the sampling distribution of $\log \hat{\phi}$ is better behaved and has the following properties

1.

$$E[\log \hat{\phi}] \approx \log \phi = \mu$$

2.

$$\text{Var}(\log \hat{\phi}) \approx \frac{1}{n_1 \pi_1 (1 - \pi_1)} + \frac{1}{n_2 \pi_2 (1 - \pi_2)} = \sigma^2$$

3. If n_1 and n_2 are large, $\log \hat{\phi} \overset{approx.}{\sim} N(\mu, \sigma^2)$

   As noted in the text, there aren't good rules for what large n means, but usually if things are ok for $\hat{\pi}$s, they should be ok here.

These statements can be justified by Taylor series methods (e.g. the variance statement can be derived by the delta rule).

Confidence Interval for ϕ

Similarly to the RR case, we will start by getting a confidence interval for $\log \phi$.

To do this, we need to get the standard error. We can do this by plugging the estimated πs into the variance formula and taking the square root, giving

$$SE(\log \hat{\phi}) = \sqrt{\frac{1}{n_1 \hat{\pi}_1 (1 - \hat{\pi}_1)} + \frac{1}{n_2 \hat{\pi}_2 (1 - \hat{\pi}_2)}}$$

$$= \sqrt{\frac{1}{n_1 \hat{\pi}_1} + \frac{1}{n_1 (1 - \hat{\pi}_1)} + \frac{1}{n_2 \hat{\pi}_2} + \frac{1}{n_2 (1 - \hat{\pi}_2)}}$$

$$= \sqrt{\frac{1}{S_1} + \frac{1}{F_1} + \frac{1}{S_2} + \frac{1}{F_2}}$$

Then the confidence interval for $\log \phi$ is

$$CI(\log \phi) = \log \hat{\phi} \pm z^*_{\alpha/2} SE(\log \hat{\phi}) = (L, U)$$

Following the same approach as for RR, a confidence interval for ϕ is

$$
\begin{aligned}
CI(\phi) &= (e^L, e^U) \\
&= \left(e^{\log \hat{\phi} - z^*_{\alpha/2} SE(\log \hat{\phi})}, e^{\log \hat{\phi} + z^*_{\alpha/2} SE(\log \hat{\phi})} \right) \\
&= \hat{\phi} \times \left(e^{-z^*_{\alpha/2} SE(\log \hat{\phi})}, e^{z^*_{\alpha/2} SE(\log \hat{\phi})} \right)
\end{aligned}
$$

Again this is not a symmetric interval around $\hat{\phi}$.

To exhibit construction of this interval, lets look at the lipid drug example

$$\hat{\phi} = 12.895 \qquad \log\hat{\phi} = 2.557$$

$$SE(\log\hat{\phi}) = \sqrt{\frac{1}{10} + \frac{1}{12685} + \frac{1}{8} + \frac{1}{130857}} = 0.474$$

$$CI(\log\phi) = 2.557 \pm 1.96 \times 0.474$$
$$= 2.557 \pm 0.930$$
$$= (1.627, 3.487)$$

$$CI(\phi) = (e^{1.627}, e^{3.487}) = (5.088, 32.678)$$

So their is strong evidence that $\phi > 1$, consistent with Baycol leading to more problems with Rhabdomyolysis.

Testing based on ϕ

We can examine hypotheses like

$$H_0 : \phi = 1 \quad \text{vs} \quad H_A : \phi \neq 1$$

or equivalently

$$H_0 : \log \phi = 0 \quad \text{vs} \quad H_A : \log \phi \neq 0$$

with tests based on $\log \hat{\phi}$. To do this we need to calculate the standard error of this under the null hypothesis. The usual estimate is

$$SE_0(\log \hat{\phi}) = \sqrt{\frac{1}{n_1 \hat{\pi}_c (1 - \hat{\pi}_c)} + \frac{1}{n_2 \hat{\pi}_c (1 - \hat{\pi}_c)}}$$

where

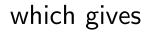$$\hat{\pi}_c = \frac{S_1 + S_2}{n_1 + n_2}$$

Then the test statistic is

$$z = \frac{\log \hat{\phi}}{SE_0(\log \hat{\phi})}$$

which is approximately distributed $N(0,1)$ under H_0.

For the lipid drug example

$$\hat{\pi}_c = \frac{10 + 8}{12695 + 130865} = 0.000125$$

which gives

$$SE_0(\log \hat{\phi})$$

$$= \sqrt{\frac{1}{12695 \times 0.000125(1 - 0.000125)} + \frac{1}{130865 \times 0.000125(1 - 0.000125)}}$$

$$= 0.830$$

$$z = \frac{2.557}{0.830} = 3.080 \qquad p\text{-value} = 0.002$$

Again supporting that Baycor has more averse events.

Prospective vs Retrospective Studies

In the examples looked at so far we are looking at the relationship between 2 binary variables. In each case their is a response/predictor relation of interest

Example	Response	Predictor
Strokes	Stroke (Yes/No)	Treatment (Aspirin/Placebo)
Lipid	Rhabdomyolysis (Yes/No)	Drug (Baycol/Lipitor)
Infant Mortality	Death (Yes/No)	Birth weight ($>$ / $<$ 2500g)

In each of these cases, the sampling is prospective (at least effectively). Subjects are assigned/observed for the predictor variable and then the response variable is observed.

In some cases, this sampling method is not feasible. One example in the text looks at the relationship between smoking and cancer.

Another example would be the relationship between genetics and breast cancer. The mutations in the gene BRCA1 have been shown to increase the risk of breast cancer in women. Lets think of a prospective study to examine the risk for both forms (wild type/mutant) of the gene.

Sample n young women and classify them by wild type (n_1) and mutant (n_2). Observe them over a period of time (say 30 years) and count the number who are diagnosed with cancer. This allows us to estimate

$$\pi_1 = P[\text{Cancer}|\text{Wild Type}]$$

$$\pi_2 = P[\text{Cancer}|\text{Mutant}]$$

and the odds ratio

$$\phi_\pi = \frac{\pi_2}{\pi_1} \times \frac{1 - \pi_1}{1 - \pi_2}$$

Not particularly feasible for a quick answer.

Another approach to studying this relationship is a retrospective case-control study. The form of the study is

- Sample m_1 subjects without breast cancer. Count number with mutant allele.

- Sample m_2 subjects with breast cancer (often matched for important covariates - age, smoking status, etc). Count number with mutant allele.

With this design we can estimate

$$p_1 = P[\text{Mutant}|\text{Cancer}]$$
$$p_2 = P[\text{Mutant}|\text{No Cancer}]$$

and

$$\phi_p = \frac{p_2}{p_1} \times \frac{1 - p_1}{1 - p_2}$$

the information on the other set of conditionals.

With this design it is not usually possible to estimate π_1 and π_2

However we can estimate ϕ_π. It can be shown that

$$\phi_\pi = \phi_p$$

So while we can't get π_1 and π_2 (and equivalently ω_1 and ω_2), at least we can get estimates of the relationship

$$\omega_2 = \phi_\pi \omega_1$$

In fact, ϕ_π is the only parameter from the prospective study than can be estimated in a retrospective study.