MATH 523: MEASURES OF EFFECT FROM 2×2 TABLES

For binary response data $Y = 0, 1$, with a single factor predictor $X = 0, 1$ representing exposure or treatment status ($X = 1$ implying treated or exposed, say), a 2×2 table can be constructed for a sample of size n. Depending on the study design, we might assume binomial sampling in the columns (*prospective*) or rows (*retrospective*).

		X		
		0	1	Total
Y	0	y_{00}	y_{01}	$n_{0.}$
	1	y_{10}	y_{11}	$n_{1.}$
	Total	$m_{.0}$	$m_{.1}$	n

Possible measure of effect that may be of interest in different contexts include:

- *Rate of Incidence* or *Risk* in Exposed and Unexposed Groups
$$\pi_1 = \Pr[Y = 1|X = 1] \qquad \pi_0 = \Pr[Y = 1|X = 0]$$
- *Odds on Incidence* in Exposed and Unexposed Groups
$$\omega_1 = \frac{\pi_1}{1 - \pi_1} \qquad \omega_0 = \frac{\pi_0}{1 - \pi_0}$$
- *Risk Difference*
$$RD = \pi_1 - \pi_0$$
- *Risk Ratio* or *Relative Risk*
$$RR = \frac{\pi_1}{\pi_0}$$
- *Odds ratio* (OR)
$$\psi = \frac{\pi_1/\pi_0}{(1 - \pi_1)/(1 - \pi_0)} = \frac{\pi_1/(1 - \pi_1)}{\pi_0/(1 - \pi_0)} = \frac{\pi_1(1 - \pi_0)}{\pi_0(1 - \pi_1)}$$
- *Excess Relative Risk*
$$ERR = \frac{\pi_1 - \pi_0}{\pi_0}$$
- *Attributable Risk*
$$AR = \frac{\pi_1 - \pi_0}{\pi_1}$$

All of these quantities are estimable from *prospective* studies, as the parameters π_0, π_1 are estimable; we have that
$$\widehat{\pi}_0 = \frac{y_{10}}{m_{.0}} \qquad \widehat{\pi}_1 = \frac{y_{11}}{m_{.1}}$$
Retrospective (case-control) studies, however, are only able to give information about
$$\varpi_1 = \Pr[X = 1|Y = 1] \qquad \varpi_0 = \Pr[X = 1|Y = 0]$$
and ψ, and functions of these parameters.

Under the *rare disease/exposure hypothesis*, however, where the *prevalence* $\phi = \Pr[Y = 1]$, and hence π_0, π_1 are **small**, case-control studies can be used to learn (approximately) about the risk ratio/relative risk π_1/π_0; this follows as, in that case, for $j = 0, 1$,
$$\frac{\pi_j}{1 - \pi_j} = \pi_j(1 - \pi_j)^{-1} = \pi_j + \pi_j^2 + \pi_j^3 + ... \approx \pi_j$$
so that
$$\frac{\pi_1/(1 - \pi_1)}{\pi_0/(1 - \pi_0)} \approx \frac{\pi_1}{\pi_0}.$$

By invariance properties of ML estimates, and the *Delta Method*, we can obtain approximate sampling distributions of these effect measures. We have for $j = 0, 1$ that

$$\widehat{\pi}_j = \frac{Y_{1j}}{m_{.j}} \dot{\sim} \mathcal{N}(\pi_j, \pi_j(1-\pi_j)/m_{.j})$$

The estimated standard error of $\widehat{\pi}_j$ is therefore

$$\sqrt{\frac{\widehat{\pi}_j(1-\widehat{\pi}_j)}{m_{.j}}}$$

The Delta method says that for statistic T_n, if

$$\sqrt{n}(T_n - \theta) \xrightarrow{d} Z \sim \mathcal{N}(0, \sigma^2)$$

then if function $g(t)$ has (non-zero) derivative $\dot{g}(t)$

$$\sqrt{n}(g(T_n) - g(\theta)) \xrightarrow{d} Z \sim \mathcal{N}(0, \{\dot{g}(\theta)\}^2\sigma^2)$$

Using the Delta method with $g(t) = \log t$, where $\dot{g}(t) = 1/t$, we have that the estimated variance of $\log \widehat{\pi}_j$ is

$$\frac{(1-\widehat{\pi}_j)}{\widehat{\pi}_j m_{.j}} = \frac{m_{.j} - y_{1j}}{m_{.j} y_{1j}} = \frac{1}{y_{1j}} - \frac{1}{m_{.j}}$$

Therefore the estimated variance of the estimator of the log relative risk, $\log(\widehat{\pi}_1/\widehat{\pi}_0)$ is

$$\left(\frac{1}{y_{10}} - \frac{1}{m_{.0}}\right) + \left(\frac{1}{y_{11}} - \frac{1}{m_{.1}}\right)$$

Similarly, with $g(t) = \log(t/(1-t))$, so that

$$\dot{g}(t) = \frac{1}{t(1-t)}$$

we have that the estimated variance of $\log(\widehat{\pi}_j/(1-\widehat{\pi}_j))$ is

$$\frac{1}{\widehat{\pi}_j(1-\widehat{\pi}_j)m_{.j}} = \frac{m_{.j}}{y_{0j}y_{1j}} = \frac{1}{y_{0j}} + \frac{1}{y_{1j}}$$

as $m_{.j} = y_{0j} + y_{1j}$. Hence the estimated variance of the estimator of the log odds-ratio

$$\log\left(\frac{\widehat{\pi}_1/(1-\widehat{\pi}_1)}{\widehat{\pi}_0/(1-\widehat{\pi}_0)}\right)$$

is

$$\frac{1}{y_{00}} + \frac{1}{y_{10}} + \frac{1}{y_{01}} + \frac{1}{y_{11}}$$