Repeated Measures Studies

Matched pairs from a single population

1. A group of subjects who are observed under two conditions (e.g., two visits such as pre- and post-treatment) for the same response variable (e.g., disorder present or absent). Subjects can be individuals with pairs of observations or aggregates of subjects such as litters, teeth in the same person, students in the same classroom, etc.
2. Case-control studies in epidemiology with explicit matching of case and control subjects on the basis of having similar background for selection and similar status for relevant characteristics.

Dichotomous response variables from a matched pairs study can be summarized with the 2 × 2 table

		Condition 2		
		Yes	No	Total
Condition 1	Yes	n_{11}	n_{12}	n_{1+}
	No	n_{21}	n_{22}	n_{2+}
	Total	n_{+1}	n_{+2}	n

The n subjects are assumed to be a simple random sample from a large population and so the $\{n_{jk}\}$ have the multinomial distribution with the probability parameters

π_{jk} = Pr{Condition 1 Response = j, Condition 2 Response = k}.

Thus, the marginal probabilities
$\pi_{j+} = (\pi_{j1} + \pi_{j2})$ = Pr{Condition 1 Response = j} and
$\pi_{+k} = (\pi_{1k} + \pi_{2k})$ = Pr{Condition 2 Response = k}
describe the distributions for the two conditions.

The $\{\pi_{jk}\}$ have the structure

		Condition 2		
		Yes	No	Total
Condition 1	Yes	π_{11}	π_{12}	π_{1+}
	No	π_{21}	π_{22}	π_{2+}
	Total	π_{+1}	π_{+2}	1

The hypothesis of equal prevalence for "yes" under conditions 1 and 2 is $H_0 : \pi_{1+} = \pi_{+1}$

$$\pi_{1+} = \pi_{+1} \Leftrightarrow \pi_{11} + \pi_{12} = \pi_{11} + \pi_{21} \Leftrightarrow \pi_{12} = \pi_{21}$$

$$\Leftrightarrow \frac{\pi_{12}}{(\pi_{12} + \pi_{21})} = \frac{1}{2}$$

The count n_{12} given the total $(n_{12} + n_{21})$ has the binomial distribution with sample size $(n_{12} + n_{21})$ and probability parameter $\theta = \pi_{12}/(\pi_{12} + \pi_{21})$.

The hypothesis of equal prevalence is H_0: $\theta = 1/2$.

An exact test for H_0 can be based on the binomial distribution.

For a two-sided test, determine

$$p = 2 \times \min\left\{ \sum_{g=0}^{n_{12}} \binom{n_{12}+n_{21}}{g} \left(\tfrac{1}{2}\right)^{(n_{12}+n_{21})}, \sum_{g=n_{12}}^{(n_{12}+n_{21})} \binom{n_{12}+n_{21}}{g} \left(\tfrac{1}{2}\right)^{(n_{12}+n_{21})}, 0.5 \right\}$$

A one-sided test would be based on the *p*-value in the corresponding tail. This method is called exact McNemar test.

An approximate test is based on noting that

$$E\{n_{12} \,|\, (n_{12}+n_{21}), H_0\} = (n_{12}+n_{21})/2 = m_{12}$$

$$Var\left\{n_{12} \,|\, (n_{12}+n_{21}), H_0\right\} = (n_{12}+n_{21})/4 = v_{12}$$

$$Q_M = \frac{\left\{\left|n_{12} - \frac{n_{12}+n_{21}}{2}\right| - 0.5\right\}^2}{(n_{12}+n_{21})/4} = \frac{\left\{\left|n_{12} - n_{21}\right| - 1\right\}^2}{(n_{12}+n_{21})}$$

is approximately chi-square with $\text{d.f.} = 1$. Usage of Q_M is reasonable when $(n_{12}+n_{21}) \geq 20$.

The data can also be represented as a set of n (2 × 2) tables for condition versus response. Let $h = 1,2,\ldots,n$ index subjects. The hth table is

		Response		
		Yes	No	Total
Condition	1	n_{h11}	n_{h12}	1
	2	n_{h21}	n_{h22}	1
	Total	n_{h+1}	n_{h+2}	2

$$\sum_{h=1}^{n} n_{h11} n_{h21} = n_{11} \qquad \sum_{h=1}^{n} n_{h12} n_{h22} = n_{22}$$

$$\sum_{h=1}^{n} n_{h11} n_{h22} = n_{12} \qquad \sum_{h=1}^{n} n_{h12} n_{h21} = n_{21}$$

The Mantel-Haenszel estimator of the common odds ratio is

$$\hat{\psi} = \sum_{h=1}^{n} (n_{h11} n_{h22} / n_{h++}) / \sum_{h=1}^{n} (n_{h12} n_{h21} / n_{h++}) = (n_{12} / n_{21})$$

The odds ratio pertains to the logistic model

$$\pi_{h11} = \exp\left(\mu_h + \beta\right) / \left\{1 + \exp\left(\mu_h + \beta\right)\right\} \text{ with } \psi = \exp(\beta)$$
$$\pi_{h21} = \exp\left(\mu_h\right) / \left\{1 + \exp\left(\mu_h\right)\right\}$$

The distribution of n_{12} given $(n_{12} + n_{21})$ and the $\{n_{h+1}\}$ is binomial with sample size $(n_{12} + n_{21})$ and probability parameter $\theta = \pi_{12}/(\pi_{12} + \pi_{21}) = \psi / (\psi + 1)$.

The Mantel-Haenszel test for no association between condition and response after adjustment for subject is

$$Q_{MH} = \frac{\left\{ \left| \sum_{h=1}^{n} (n_{h11} - 0.5 n_{h+1}) \right| - 0.5 \right\}^2}{\left\{ \sum_{h=1}^{n} n_{h+1} n_{h+2} / 4 \right\}} = \frac{\left\{ \left| n_{12} - n_{21} \right| - 1 \right\}^2}{(n_{12} + n_{21})} = Q_M$$

One way to form a confidence interval for ψ is to transform an interval for θ:

If θ_{L} and θ_{U} are the limits for θ, then $\theta_{\mathrm{L}} \le \theta \le \theta_U$ is transformed to

$$\psi_{\mathrm{L}} = \theta_L / (1 - \theta_L) \le \psi \le \theta_U / (1 - \theta_U) = \psi_U$$

since $\psi = \theta / (1 - \theta)$.

An approximate interval for θ is

$$\theta = f \pm \left\{ z_{\alpha/2} \sqrt{\frac{f(1-f)}{(n_{12}+n_{21}-1)} + \frac{1}{2(n_{12}+n_{21})}} \right\}$$

where $f = n_{12} / \left(n_{12} + n_{21} \right)$.

Another approximate interval is directed at log ψ. It is

$$\psi = \exp\left\{ \log_e \left(n_{12} / n_{21} \right) \pm z_{\alpha/2} \sqrt{\frac{1}{n_{12}} + \frac{1}{n_{21}}} \right\}.$$

The interval can be refined somewhat by adding $\frac{1}{2}$ to n_{12} and n_{21}.

A third type of approximate interval is based on

$$\frac{\left\{\left| f-\theta\right| - \frac{1}{2n_0}\right\}^2}{\theta(1-\theta)/n_0} \le z_{\alpha/2}^2$$

where $n_0 = (n_{12} + n_{21})$ and $f = n_{12} / (n_{12} + n_{21})$ and the solution of the resulting quadratic equations. (See Fleiss [1981, p.14].)

$$\theta_L = \frac{\left\{2n_0 f + z_{\alpha/2}^2 - 1\right\} - z_{\alpha/2}\sqrt{z_{\alpha/2}^2 - (2+1/n_0) + 4f(n_0(1-f)+1)}}{2\left(n_0 + z_{\alpha/2}^2\right)}$$

$$\theta_U = \frac{\left\{2n_0 f + z_{\alpha/2}^2 + 1\right\} + z_{\alpha/2}\sqrt{z_{\alpha/2}^2 + (2-1/n_0) + 4f(n_0(1-f)-1)}}{2\left(n_0 + z_{\alpha/2}^2\right)}$$

An exact interval is determined from the smallest θ_U and largest θ_L such that:

$$\sum_{g=0}^{n_{12}} \theta_U^g (1-\theta_U)^{n_{12}+n_{21}-g} \binom{n_{12}+n_{21}}{g} \leq \tfrac{\alpha}{2}$$

$$\sum_{g=n_{12}}^{n_{12}+n_{21}} \theta_L^g (1-\theta_L)^{n_{12}+n_{21}-g} \binom{n_{12}+n_{21}}{g} \leq \tfrac{\alpha}{2}$$

M	$n_{12,\mathrm{M}}$	$n_{21,\mathrm{M}}$	$\rightarrow$	ψ_{M}
F	$n_{12,\mathrm{F}}$	$n_{21,\mathrm{F}}$	$\rightarrow$	ψ_{F}

$$H_0 : \psi_M = \psi_F$$

$$\frac{n_{12,M}}{n_{21,M}} \Big/ \frac{n_{12,F}}{n_{21,F}} = \frac{n_{12,M}\, n_{21,F}}{n_{21,M}\, n_{12,F}}$$

$$\text{Fisher's Test } p \approx 1$$

Suppose H_0 is supported, estimate ψ_* for both groups:

$$\psi_* \triangleq \frac{n_{12,M} + n_{12,F}}{n_{21,M} + n_{21,F}}$$

For case-control studies with more than one case or control in each matched set, the association between condition and response with adjustment for subject can be assessed with the Mantel-Haenszel test. Use PROC FREQ for subject × condition × response.

For situations with *m* conditions, the association between condition and response with adjustment for subject can be assessed with the extended Mantel-Haenszel test. Use PROC FREQ for subject × condition × response. Such analysis applies to both dichotomous and ordinal data. (Chapter 6)

For situations with ordinal data, the association between condition and response with adjustment for subject can be assessed with the extended Mantel-Haenszel test. Use PROC FREQ for subject × condition × response. (Chapter 6)