Handout on $2 \times 2 \times K$ tables via Estimating Equations

Eric Slud 10/23/2026

The topic covered in class is count data n_{ijk} for $i = 1, 2,\ j = 1, 2,\ k = 1, \ldots, K$ cross-classified from a total of $N = \sum_{i,j,k} n_{ijk}$ subjects broken down into $n_{\cdot\cdot k}$ subjects in each of $k = 1, \ldots, K$ separate student centers. Think of $i = 1, 2$ as respectively indexing treatment A or B and $j = 1, 2$ as respectively indicating 'Y', a positive or successful response, versus 'N', a negative response. The data are assumed to have been collected in such a way that *either* $n_{i\cdot k}$ is fixed and nonrandom for $i = 1, 2$ or else that counts of subjects under treatment A or B are independent Poisson. (Here, and throughout this writeup, subscript dots are used to indicate summation over that subscript index.) In either case we assume n_{i1k} are independent Binomial$(n_{i\cdot k}, p_{i1k}/p_{i\cdot k})$ count random variables across i, k, conditionally given $\{n_{i\cdot k},\ \ i = 1, 2,\ \ k = 1, \ldots, K\}$. In either the Poisson or nonrandom $n_{i\cdot k}$ cases, denote $p_{ijk} = E(n_{ijk}\,|N)/N$ (for $N \geq 1$) for all i, j, k, and note that in the case of nonrandom numbers $n_{i\cdot k}$ of subjects on treatment indexed i in the k'th table, $n_{i\cdot k} = p_{i\cdot k} n_{\cdot\cdot k}$ with certainty. Assume none of the probabilities p_{ijk} is 0, and then by definition $\sum_{i,j,k} p_{ijk} = 1$.

In the context of Mantel-Haenszel statistics, we are interested in A vs. B success-odds-ratios

$$\theta_k \ \equiv\ \Big(\frac{p_{11k}}{p_{1\cdot k}} \Big/ \frac{p_{12k}}{p_{1\cdot k}}\Big) \Big/ \Big(\frac{p_{21k}}{p_{2\cdot k}} \Big/ \frac{p_{22k}}{p_{2\cdot k}}\Big) \ \equiv\ \frac{p_{11k}\, p_{22k}}{p_{12k}\, p_{21k}}$$

Our statistical assumption will be one of

(A1) $H_0 :\ \theta_k = 1$ for all $k = 1, \ldots, K$ **or**

(A2) For some $\theta \in (0, \infty)$, $\theta_k = \theta$ for all $k = 1, \ldots, K$.

Now define the notations (valid for $n_{\cdot\cdot k} > 1$)

$$\mu_k \ =\ \frac{n_{1\cdot k}\, n_{\cdot 1k}}{n_{\cdot\cdot k}}\ , \quad R_k \ =\ \frac{n_{11k}\, n_{22k}}{n_{\cdot\cdot k}}\ , \quad S_k \ =\ \frac{n_{12k}\, n_{21k}}{n_{\cdot\cdot k}}\ , \quad v_k \ =\ \frac{R_k\, S_k}{n_{\cdot\cdot k} - 1}$$

We showed in class that conditionally given $n_{i\cdot k}$, $n_{\cdot jk}$ for $i, j = 1, 2$, under H_0

$$n_{11k} \sim \text{Hypergeom}(n_{\cdot\cdot k},\, n_{1\cdot k},\, n_{\cdot 1k})\ , \quad E(n_{11k}) \ =\ \mu_k\ , \quad \text{Var}(n_{11_k}) \ =\ v_k \qquad (1)$$

and also that under the constant odds-ratio model (A2), the equation $\sum_{k=1}^{K} (R_k - \theta S_k) \ = 0$ is a valid Estimating Equation, the solution of which is the

Mantel-Haenszel estimator

$$\hat{\theta} \;=\; \hat{\theta}^{MH} \;=\; \sum_{k=1}^{K} R_k \,/\, \sum_{k=1}^{K} S_k \equiv R \,/\, S \tag{2}$$

Next, assuming (A2) and also that conditionally given $n_{i\cdot k}$ for $i = 1, 2$, $n_{i1k} \sim$ Binom$(n_{i\cdot k},\, p_{i1k}/p_{i\cdot k})$ are independent across $i = 1, 2$ and each k, we evaluate the variance of $R_k - \theta S_k$ and the expectation of S_k. Indeed, using standard formulas for binomial first and second moments along with independence of n_{1jk} and $n_{2j'k}$ for all j, j', k, we find (with all expectations below still conditional on $n_{1\cdot k}, n_{2\cdot k}$)

$$E(n_{ijk}) \;=\; n_{i\cdot k}\,\frac{p_{ijk}}{p_{i\cdot k}}\,, \qquad E(n_{ijk}^2) \;=\; \{n_{i\cdot k}\,\frac{p_{ijk}}{p_{i\cdot k}}\}^2 \;+\; n_{i\cdot k}\,\frac{p_{ijk}}{p_{i\cdot k}}\,(1 - \frac{p_{ijk}}{p_{i\cdot k}})$$

and

$$E(S_k) \;=\; \frac{n_{1\cdot k}\, n_{2\cdot k}}{n_{\cdot\cdot k}\, p_{1\cdot k}\, p_{2\cdot k}}\; p_{12k}\, p_{21k} \tag{3}$$

$$E(R_k^2) \;=\; \Big(\frac{n_{1\cdot k}\, n_{2\cdot k}\, p_{11k}\, p_{22k}}{n_{\cdot\cdot k}\, p_{1\cdot k}\, p_{2\cdot k}}\Big)^2 (1 \;+\; \frac{p_{12k}}{n_{1\cdot k}\, p_{11k}}) \cdot (1 \;+\; \frac{p_{21k}}{n_{2\cdot k}\, p_{22k}}) \tag{4}$$

$$E(S_k^2) \;=\; \Big(\frac{n_{1\cdot k}\, n_{2\cdot k}\, p_{12k}\, p_{21k}}{n_{\cdot\cdot k}\, p_{1\cdot k}\, p_{2\cdot k}}\Big)^2 (1 \;+\; \frac{p_{11k}}{n_{1\cdot k}\, p_{12k}}) \cdot (1 \;+\; \frac{p_{22k}}{n_{2\cdot k}\, p_{21k}}) \tag{5}$$

$$E(R_k\, S_k) \;=\; (n_{\cdot\cdot k})^{-2}\, E(n_{11k}\, n_{22k}\, (n_{1\cdot k} - n_{11k})\, (n_{2\cdot k} - n_{22k}))$$

$$= \Big(\frac{n_{1\cdot k}\, n_{2\cdot k}}{n_{\cdot\cdot k}\, p_{1\cdot k}\, p_{2\cdot k}}\Big)^2 \{p_{11k} p_{1\cdot k} - p_{11k}^2 \;-\; \frac{p_{11k}\, p_{12k}}{n_{1\cdot k}}\} \cdot \{p_{22k} p_{2\cdot k} - p_{22k}^2 \;-\; \frac{p_{22k}\, p_{21k}}{n_{2\cdot k}}\}$$

$$= \Big(\frac{n_{1\cdot k}\, n_{2\cdot k}}{n_{\cdot\cdot k}\, p_{1\cdot k}\, p_{2\cdot k}}\Big)^2 p_{11k}\, p_{12k}\, p_{22k}\, p_{21k}\, \{1 - \frac{1}{n_{1\cdot k}}\} \cdot \{1 - \frac{1}{n_{2\cdot k}}\} \tag{6}$$

Putting together equations (4)–(6), we find

$$\begin{aligned} E(R_k \;-\; \theta\, S_k)^2 \;=\; \Big(\frac{n_{1\cdot k}\, n_{2\cdot k}}{n_{\cdot\cdot k}\, p_{1\cdot k}\, p_{2\cdot k}}\Big)^2 \Big[& p_{11k}\, p_{22k}\, (p_{11k} + \frac{p_{12k}}{n_{1\cdot k}})(p_{22k} + \frac{p_{21k}}{n_{2\cdot k}}) \\ & - 2\theta\, p_{11k}\, p_{12k}\, p_{21k}\, p_{22k}(1 - n_{1\cdot k}^{-1})(1 - n_{2\cdot k}^{-1}) \\ & + \theta^2\, p_{12k}\, p_{21k}\, (p_{12k} + \frac{p_{11k}}{n_{1\cdot k}})(p_{21k} + \frac{p_{22k}}{n_{2\cdot k}})\Big] \end{aligned} \tag{7}$$

From Estimating Equation Theory, under general regularity conditions (including cases in which all n_{ijk} are large and go to infinity but K may be bounded, but also

including cases where the individual $n_{\cdot\cdot k}$ table totals remain bounded but the number K of tables goes to ∞), the estimator $\hat{\theta}$ in (2) is $\sqrt{N}$-consistent, which means that the sandwich asymptotic variance

$$\text{a.var}(\hat{\theta})) \;=\; E((R \,-\, \theta\, S)^2) \,/\, (E(S))^2 \tag{8}$$

should be of order $1/N$. Since $E(S))$ is of order N, this implies that $E((R-\theta S)^2)$ should also be of order N, and this implies that the top-order terms in (7) — the ones in which the $1/n_{i\cdot k}$ terms in curly brackets $\{\cdot\}$ are all dropped – should cancel away. And a little algebra using $\theta = p_{11k}p_{22k}/(p_{12k}p_{21k})$ shows that this is indeed true. Therefore the expression (7) becomes

$$E(R_k - \theta\, S_k)^2 \;=\; \Big(\frac{n_{1\cdot k}\, n_{2\cdot k}}{n_{\cdot\cdot k}\, p_{1\cdot k}\, p_{2\cdot k}}\Big)^2 \Big[p_{11k}\, p_{22k}\, (p_{22k}\frac{p_{12k}}{n_{1\cdot k}} + p_{11k}\frac{p_{21k}}{n_{2\cdot k}}) + 2(p_{11k}\, p_{22k})^2\, \frac{n_{\cdot\cdot k}}{n_{1\cdot k}\, n_{2\cdot k}}$$

$$+\; \frac{(p_{11k}\, p_{22k})^2}{p_{12k}\, p_{21k}}\, (p_{21k}\frac{p_{11k}}{n_{1\cdot k}} \;+\; p_{12k}\frac{p_{22k}}{n_{2\cdot k}}) \;+\; (\theta-1)^2\, \frac{p_{11k}\, p_{12k}\, p_{21k}\, p_{22k}}{n_{1\cdot k}\, n_{2\cdot k}}\Big] \tag{9}$$

The goal of this development is to get a valid estimate of the asymptotic variance of $\hat{\theta}$ valid for cases with large K and small $n_{\cdot\cdot k}$ as well as bounded K and large n_{ijk}. This was done in a series of papers around a 1986 Biometrics paper of J. Robins, N. Breslow and S. Greenland that is cited in Agresti's (2013) text as the main source of a variance estimate for $\hat{\theta}^{MH}$. The terms in that variance estimate are all of order $1/N$ in the case where K is bounded and all n_{ijk} are of order N, and therefore those formulas must have discarded the final $O(1)$ term in (9), which is

$$\frac{n_{1\cdot k}\, n_{2\cdot k}}{(n_{\cdot\cdot k}\, p_{1\cdot k}\, p_{2\cdot k})^2}\, (\theta-1)^2\, p_{11k}\, p_{12k}\, p_{21k}\, p_{22k} \tag{10}$$

Under conditions where the large-sample theory for consistency and asymptotic normality for $\hat{\theta}^{MH}$ works (with either large K or large counts $n_{\cdot\cdot k}$ or both), a logical asymptotic variance estimator is $\hat{V} = S^{-2} \sum_{k=1}^{K} (R_k \,-\, \hat{\theta}\, S_k)^2$.

Homework 3 Problem 2 is to check the theoretical formula (9) summed over k for $E(R-\theta S)^2$ by simulation, with and without this last term (10), and to show that the two answers (with and without this term) can be meaningfully different; and to check the consistency of the estimator $\hat{V}$ for the variance of $\hat{\theta}^{MH}$.