Lecture 3

25

3. Exact Inference for Various Designs

a. As with approximate analysis,
   i. case–control approach is mathematically equivalent to the stratified cohort approach
   ii. conditionality principal justifies treating the unstratified cohort design as a stratified cohort design.

b. Cohort inference is generated from distribution of $X_{00} \sim \text{Bin}(\pi_0, X_{0+}), X_{10} \sim \text{Bin}(\pi_1, X_{1+})$.

<table>
<thead>
<tr>
<th>Healthy</th>
<th>Diseased</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unexposed</td>
<td>$X_{00}$</td>
</tr>
<tr>
<td>Exposed</td>
<td>$X_{10}$</td>
</tr>
</tbody>
</table>

   i. $\pi_0 = P\{\text{Healthy|Unexposed}\}$
   ii. $\pi_1 = P\{\text{Healthy|Exposed}\}$

c. $\text{Var}_{\theta=1}[X_{00}|X_{0+}, X_{+}]= \frac{X_{+1}X_{0+}X_{01}X_{1+}}{X_{+}(X_{+}+1)}$

d. $P\{X_{00}, X_{10}|X_{0+}, X_{1+}\} = \frac{(X_{0+})!}{(X_{00})!(X_{+0})!(X_{1+})} \pi_0^{X_{00}}(1-\pi_0)^{X_{01}}\pi_1^{X_{10}}(1-\pi_1)^{X_{11}}$

e. After rewriting $\pi_1$ in terms of $\pi_0$ and $\theta$, distribution of $T$ still depends on $\pi_0$
   i. $\pi_1 = \pi_0 \theta / (1-\pi_0 + \pi_0 \theta)$
   ii. Then conditional table probabilities are

   \[
P\{X_{00}, X_{10}|X_{0+}, X_{1+}\} = \frac{(X_{0+})!}{(X_{00})!(X_{+0})!(X_{1+})} \pi_0^{X_{00}}(1-\pi_0)^{X_{01}}\left(\frac{1-\pi_0}{1-\pi_0 + \pi_0 \theta}\right)^{X_{1+}}
   \times \frac{(X_{+1})!}{(X_{+0})!(X_{1+})!} \pi_0^{X_{+0}}(1-\pi_0)^{X_{10}}\theta^{X_{11}}
   \]

   $\pi_0 \text{ contributes a constant factor to all tables with same } X_{0+}, X_{1+}$

d. Condition also on $X_{0+}$ and $X_{1+}$ as well as $X_{0+}$ and $X_{+1}$.

   a. removes dependence on $\pi$
   b. Distribution is called hypergeometric

c. If $\theta \neq 1$ called noncentral hypergeometric

d. Indicate by $|X_{1+}, X_{+k}$ conditional on $X_{0+}$ and $X_{1+}$ and $X_{1+}$ and $X_{+1}$.

   e. cuts number of tables to be examined.
       i. Both a blessing and a curse.

   f. Normal approx. works poorly unless $E_{kj} \geq 5\forall j,k$. See Figure 7.

   i. Could have continuity correction described earlier.
       a. Choice of cc and variance give 4 possible tests

g. Testing $\theta$

   i. One-sided, conditional on all margins:
       a. $H_0: \theta = \theta_0$ vs $H_A: \theta > \theta_0$
       b. Use $T = \theta - X_{00}X_{11}/(X_{01}X_{10})$ or equivalently $X_{00}$
       c. $p$-value is sum of probabilities for table with upper left corner at least observed

   ii. For two–sided test
       a. order tables according to null probability
       b. Implies something other than doubling smaller 1-sided $p$-value
       c. Result is called Fisher’s Exact Test
Lecture 3

**Asymmetry intentional**

i. Assign each of the categories a score

0 4 1 3 2 2 3 1 4 0
0 5 0 5 0 5 0 5 0 5

ii. Two-sided p-value: Sum of all table probabilities as small or smaller than the table we observe:

$$0.159 + 0.003 + 0.040 = 0.207$$

iii. Smaller than double one-sided p-value, since it avoids adding 3 2 (0.317)

v. p-value dominated by probability of observed table.

vi. Cf. p-value not conditioned on column totals: 30 tables:

$$\begin{array}{cccccccc}
0 & 4 & 1 & 3 & 2 & 2 & 3 & 1 \\
0 & 5 & 0 & 5 & 0 & 5 & 0 & 5 \\
0 & 4 & 1 & 3 & 2 & 2 & 3 & 1 \\
0 & 4 & 1 & 3 & 2 & 2 & 3 & 1 \\
2 & 3 & 2 & 3 & 2 & 3 & 2 & 3 \\
2 & 3 & 2 & 3 & 2 & 3 & 2 & 3 \\
0 & 4 & 1 & 3 & 2 & 2 & 3 & 1 \\
4 & 1 & 4 & 1 & 1 & 4 & 1 & 1 \\
\end{array}$$

vii. probabilities depend on common null $\pi$.

b. Alternatively, one can use

$$\text{Ridit scores} u_k = \frac{\sum_{i<k} K_{ij} + (K_{k+1} + 1)}{2} / \sum_{i=1}^{K} X_{ij}$$

i. Gives Mann–Whitney–Wilcoxon test

ii. Test statistic has interpretation as estimated probability that a random individual from one group has a higher score than random individual from the other

2. Calculate $T_j = \sum_{k=0}^{K-1} u_k (X_{kj} - E_{kj})$

a. $E[T_j] = 0$

3. To get variance, need conditional covariances of table entries.

a. Note $\text{Var} [X_{kj}] = X_{kj} (X_{++} - X_{kj}) X_{++} / (X_{++} - 1)$

b. For two entries in the same column:

i. $\text{Var} [X_{kj} + X_{ij}] = (X_{k+} + X_{i+}) X_{00} X_{++} (X_{++} - X_{kj} - X_{ij}) / (X_{++} - 1)$

C. Estimation of Odds Ratios:

1. Substitute cell proportions for cell probabilities

2. Called Mantel–Haenszel test

3. When $J = 2$ or $K = 2$

a. this is the same as the previous example, with any second dimension scores

b. Called Cochran–Armitage test

4. Multiple of correlation between row and column scores (1 for column $j$, and 0 for all other columns)

5. Assign each of the columns a score $v_k$

6. Calculate $T = \sum_{j=0}^{J-1} T_j = \sum_{k=0}^{K-1} \sum_{j=0}^{J-1} v_k (X_{kj} - E_{kj})$

7. Multiple of correlation between row and column scores
8. Squaring and rescaling makes it $\approx \chi_1^2$
   a. $T$ gives test of $H_0$ independence vs. $H_A$: higher rows have column probabilities putting more weight on higher columns than low rows
      i. Since $\sum_j T_j = 0$, the $T_j$ are not independent.
      ii. Properly rescaled, $\sum_j T_j^2/c_j \sim \chi_{J-1}^2$