III. Association in tables

10. Conditional methods?
   a. Same test statistic
   b. Distribution in cells is given by sequence of binomials
   c. Null hypothesis
   d. Need standard error
   i. If independent, variance would be $\sum_{k=1}^{K} u_{k}^{2} E_{k}$
   ii. Here $\operatorname{Var} \left[ \sum_{k=1}^{K} u_{k} X_{k} \right] = \sum_{k} u_{k}^{2} X_{+} \pi_{k}(1-\pi_{k}) - \sum_{k \neq j} u_{k} u_{j} X_{+} \pi_{k} \pi_{j} = X_{+} \left[ \sum_{k} u_{k}^{2} \pi_{k} - \sum_{k \neq j} u_{k} u_{j} \pi_{k} \pi_{j} \right] = \sum_{k} u_{k}^{2} E_{k} - \left( \sum_{k} u_{k} E_{k} \right)^{2} / X_{+}$
   e. Sampling distribution approximately standard normal

9. Why not CI?
   a. CI can give test when we have one parameter to test
   b. Here we need $K-1$ parameters
   c. CI becomes confidence region: more complicated.

Agr:

6. Variances and Covariances
   a. $\operatorname{Var} [X_{k} | X_{+}] = X_{+} \pi_{k}(1-\pi_{k})$ for $\pi_{k}$ equal to the proportion of person years in group $k$
   b. $\operatorname{Var} [X_{k} + X_{j} | X_{+}] = X_{+} (\pi_{k} + \pi_{j})(1 - \pi_{k} - \pi_{j})$
   c. $\operatorname{Cov} [X_{k}, X_{j} | X_{+}] = -X_{+} \pi_{k} \pi_{j}$

7. Proof that Pearson’s statistic has $\chi_{K-1}^{2}$ distribution:
   a. Let
      i. $Y$ be the $K-1$ by 1 matrix $(X_{1}, \ldots, X_{K-1})$
      (not including $X_{0}$)
      ii. $\nu$ be the $K-1$ by 1 matrix $(\pi_{1}, \ldots, \pi_{K-1})$
      iii. $1$ be the $K-1$ by 1 matrix whose entries are all 1.
   b. Hence $\operatorname{Var}[Y] = X_{+}(\operatorname{diag}(\nu) - \nu \nu^{T})$
   c. Then $V = \operatorname{Var}[Y]^{-1}$, as seen by noting that
   
   $\operatorname{Var}[Y] = (\operatorname{diag}(\nu) - \nu \nu^{T}) \operatorname{diag}(1/\nu) + \nu \nu^{T} \operatorname{diag}(1/\nu) - \nu \nu^{T}$
   
   
   $= \operatorname{diag}(1/\nu) + \nu \nu^{T} \operatorname{diag}(1/\nu) - \nu \nu^{T}$
   
   $\nu \nu^{T}$
   
   $= I + \nu \nu^{T} - \nu \nu^{T} (1 - \pi_{0}) / \pi_{0} = I$

   using
   
   i. $\operatorname{diag}(\nu) 1 = \nu$
   
   ii. $\nu^{T} (1/\nu) = 1$
   
   iii. $\nu^{T} 1 = 1 - \pi_{0}$

   d. Recall from 580 or 582 that
   
   $T = (Y - X_{+} \nu)^{T} V (Y - X_{+} \nu) \sim \chi_{K-1}^{2}$

   e. Now I show that $T$ is Pearson’s statistic:

   \[ T = (Y - X_{+} \nu)^{T} V (Y - X_{+} \nu) \]

   \[ = (Y - X_{+} \nu)^{T} (\operatorname{diag}(1/\nu) + 11^{T} / \pi_{0}) (Y - X_{+} \nu) / X_{+} \]

   \[ = (Y - X_{+} \nu)^{T} \operatorname{diag}(1/\nu) (Y - X_{+} \nu) + (Y - X_{+} \nu)^{T} 11^{T} (Y - X_{+} \nu) / (X_{+} \pi_{0}) \]

   i. First term is $\sum_{j=1}^{K} (Y_{j} - X_{+} \pi_{j}) / (X_{+} \pi_{j})$
   
   ii. $11^{T} (Y - X_{+} \nu) / (X_{+} \pi_{0}) = \sum_{j=1}^{K-1} (Y_{j} - X_{+} \pi_{j}) = -(X_{0} - X_{+} \pi_{0})$

   iii. Hence second term is $(X_{0} - X_{+} \pi_{0})^{2} / (X_{+} \pi_{0})$.

8. Ordered exposures:
   a. Give scores $u_{k}$ to groups (often 1, 2, . . .)
   b. Add up $\sum_{k=1}^{K} u_{k} X_{k} - E_{k}$
   c. Conditional expectation zero.
   d. Need standard error
   i. If independent, variance would be $\sum_{k=1}^{K} u_{k}^{2} E_{k}$
   ii. Here $\operatorname{Var} \left[ \sum_{k=1}^{K} u_{k} X_{k} \right] = \sum_{k} u_{k}^{2} X_{+} \pi_{k}(1-\pi_{k}) - \sum_{k \neq j} u_{k} u_{j} X_{+} \pi_{k} \pi_{j} = X_{+} \left[ \sum_{k} u_{k}^{2} \pi_{k} - \sum_{k \neq j} u_{k} u_{j} \pi_{k} \pi_{j} \right] = \sum_{k} u_{k}^{2} E_{k} - \left( \sum_{k} u_{k} E_{k} \right)^{2} / X_{+}$
   e. Sampling distribution approximately standard normal

   3. $J$ multinomials each with $K$ categories
   a. Get from full multinomial by conditioning on column totals $X_{+}$
   b. Probabilities $\psi_{k,j} = \pi_{k,j} / \pi_{j}$
   c. Interest parameters are now $\theta_{k,j} = \psi_{k,j} \psi_{0,j} / (\psi_{k,0} \psi_{0,j})$
   i. Quantities $\psi_{k,j} / \psi_{0,j}$ are odds
   ii. Quantity $\theta_{11}$ is odds ratio
   d. With $\psi_{0,j}$ fixed, increasing $\theta_{k,j}$ is equivalent to pushing $\pi_{k,j}$ near 1.
   i. Taking logs,
   
   $\log(\psi_{11} / \psi_{01}) = \log(\psi_{10} / \psi_{00}) + \log(\theta)$. 

   ii. Function $\log(\theta) = \log(x / (1-x)) = \log(x) - \log(1-x)$ is called logistic function.

   e. Factorization of joint probability into marginal and conditional probabilities:

   $x^{+} \prod_{k,j} \pi_{k,j}^{x_{k,j}} / x_{k,j}! = \left[ x_{+}^{+} \prod_{j} \pi_{j}^{x_{j}} / x_{j}! \right] \left[ \prod_{k} \theta_{k,j}^{x_{k,j}} / x_{k,j}! \right]

   A: 2.3.4

   f. Special Case: $J = K = 2$, $\psi_{k,j}$ small
   i. Interpretation: Rows represent exposure to a carcinogen, columns represent healthy and sick.
   ii. Alternate comparator $\psi_{11} / \psi_{01}$ relative risk
   iii. Approximately odds ratio if $\psi_{10}$ and $\psi_{00} \approx 1$

   iv. Multiplicative relationship between $\psi_{11}$ and $\psi_{01}$ fails when one (and hence both) are not really small.
Lecture 2

5. Distribution conditional on both row and column totals:

- i. Testing for
- ii. Exposure proportions proportional to those in the overall population
- iii. Odds ratios are the same
- iv. Natural probabilities for our results are conditional on disease status total
- v. Odds ratios are the same
- vi. Conditioning on row and column margins makes analysis of case-control study exactly the same as a cohort study

6. Case–Control Study

- i. Individuals are distributed into disease groups (columns of table) and exposure groups (rows of table)
- ii. According to Poisson model conditional.

B. Testing association in tables

1. Table entries

<table>
<thead>
<tr>
<th>Exp. cat.</th>
<th>Contr.</th>
<th>Cases</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>X_{00}</td>
<td>X_{01}</td>
<td>X_{0+}</td>
</tr>
<tr>
<td>1</td>
<td>X_{10}</td>
<td>X_{11}</td>
<td>X_{1+}</td>
</tr>
<tr>
<td></td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>K - 1</td>
<td>X_{K-10}</td>
<td>X_{K-11}</td>
<td>X_{K-1+}</td>
</tr>
<tr>
<td>K</td>
<td>X_{K0}</td>
<td>X_{K1}</td>
<td>X_{K+}</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2. $J = K = 2$

- i. Stratified cohort study (i.e., condition on row totals).

A: 2.4

v. Example

- Do a cohort study and find $\pi_1 = .020$, $\pi_0 = .005$.
- Hence $q = 4$.
- Analyze cohort with some other high–risk factor and get $\pi_0 = .3$.
- Hence $\pi_1 = 1.2$; invalid.

4. $K$ multinomials each with $J$ categories:

- a. Condition on row totals $X_{k+}$
- b. like previous.
- c. Conditional probabilities $\kappa_{kj} = \lambda_{kj}/\lambda_{k+}$
- d. Note that $\kappa_{kj}\kappa_{00}/(\kappa_{0j}\kappa_{k0}) = \lambda_{kj}/\lambda_{0j}/\lambda_{k0}$
- e. Odds ratios are the same.

- ii. Use same statistic as before
- iii. Solution is $\psi_{00} = \psi_{10}/(\psi_{10}\psi_{01})$
- iv. Then $P[X_{1+} = 0|X_{+j} = 1] = \psi_{00}/\psi_{10}$
- v. $P[X_{1+} = 1|X_{+j} = 1] = \psi_{01}/\psi_{00} \psi_{11}$
- vi. Then $P[X_{1+} + 2|X_{+j} = 1] = \psi_{11}/\psi_{00}$

6. Case–Control Study

- i. Individuals are distributed into disease groups (columns of table) and exposure groups (rows of table)
- ii. According to Poisson model conditional.

b. Condition on disease status totals

- i. Choose a number in each disease status
- ii. $v$ is inverse of variance of $X_{00} - E_{00}$
- iii. $X_{j+k} - E_{j+k}/E_{j+k}$
- iv. For $v = (X_{j+k} - X_{j+k}/X_{++})^{-1}$
- v. $v = (X_{j+k} - X_{j+k}/X_{++})^{-1}$
- vi. Working backwards through the above calculations

A: 2.3.5–2.3.6