Mantel-Haenszel の方法

長島 健悟

城西大学 薬学部

2008 年 6 月 12 日

Mantel & Haenszel (1959) について

- Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J. Nat. Cancer Inst.* 1959; **22**(4): 719–748.

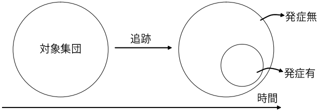
1. 前向き・後ろ向き研究 (データの特徴), 相対リスク (リスク比とオッズ比), 交絡因子のバランス (データの収集方法, マッチング, コントロールの選択など)
2. 2×2 表の検定と推定
3. 各層で共通なオッズ比が 1 かどうかの検定
    - $2 \times 2 \times k$ 表の場合の Mantel-Haenszel 検定
    - 1:1 マッチングの場合の検定
    - $2 \times j \times k$ 表の場合 (多水準の曝露因子に対する拡張)
4. 各層で共通なオッズ比の推定
    - $2 \times 2 \times k$ 表における, Mantel-Haenszel 推定量 R の提案と, その他の推定量 R_1, R_2, R_3, R_4 の比較
    - 1:1 マッチングの場合の推定

本発表での取り扱い

1. 前向き・後ろ向き研究 (データの特徴), 相対リスク (リスク比とオッズ比), 交絡因子のバランス (データの収集方法, マッチング, コントロールの選択など)
    - 論文全体に渡っているが, 特に前半 pp.719–730
    - カバーできる範囲で
2. 2×2 表の検定と推定
3. 各層で共通なオッズ比が 1 か否かの検定
4. 各層で共通なオッズ比の推定
    - 後半 pp.730–746
    - $2 \times 2 \times k$ 表, $i \times j \times k$ 表の話

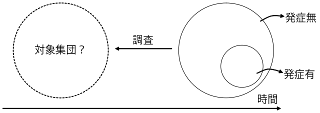
前向き研究

- 利点
    - 原因 → 結果, リスク比, リスク差が計算できる, 結果の解釈が容易
    - 複数の結果因子を調べることができる
    - 曝露因子の測定バイアスが少ない場合が多い
- 欠点
    - 追跡する時間が長い, コストが高い, 後向き研究より時間がかかる
    - 結果因子の発生確率が小さい場合に適さない



後向き研究

- 利点
  - コストが低い, 研究期間が比較的短い, 個人レベルでの研究が可能なことも
  - 結果因子の発生確率が小さい場合に適している
- 欠点
  - 原因 ← 結果, リスク比, リスク差が計算できない
  - 曝露因子の測定バイアスが問題



Mantel & Haenzel (1959)

- 事例: John Snow のコレラの研究, 喫煙と肺がんの関連についての研究
  - 役割と限界について議論する
  - 仮説を生み出す過程で後向き研究は使える
  - 結果因子の比較が可能な状況を作る必要がある
  - 交絡因子は曝露因子間で均一であることが必要
- 後向き研究と前向き研究のどちらかしか実施できないときは, 同じ結論を得たい
- 発症確率が非常に小さいとき, 倫理的にランダム化試験が行えない場合など

2×2 表

表: コホート研究

セル確率	発症あり	発症なし	合計
曝露あり	q_1	$1 - q_1$	1
曝露なし	q_2	$1 - q_2$	1

確率変数	発症あり	発症なし	合計
曝露あり	Y_{11}		m_{1+}
曝露なし	Y_{21}		m_{2+}
合計			m_{++}

$$\begin{cases} Y_{11} \sim B(m_{1+}, q_1) \\ Y_{21} \sim B(m_{2+}, q_2) \end{cases} \tag{1}$$

2×2 表

表: ケース・コントロール研究

セル確率	曝露あり	曝露なし	合計
発症あり	p_1	$1 - p_1$	1
発症なし	p_2	$1 - p_2$	1

確率変数	曝露あり	曝露なし	合計
発症あり	X_{11}		n_{1+}
発症なし	X_{21}		n_{2+}
合計			n_{++}

$$\begin{cases} X_{11} \sim B(n_{1+}, p_1) \\ X_{21} \sim B(n_{2+}, p_2) \end{cases} \tag{2}$$

リスク比とオッズ比

- 発症リスク比 ϕ (risk ratio)
- ケース・コントロール研究では曝露リスク比しか計算できない

$$\phi = \frac{q_1}{q_2} \tag{3}$$

- 発症オッズ比 ψ (odds ratio)
- 発症確率が小さいとき, リスク比に近似

$$\psi = \frac{q_1/(1-q_1)}{q_2/(1-q_2)} \left(= \frac{p_1/(1-p_1)}{p_2/(1-p_2)} \right) \tag{4}$$

- $q_1 = 0.06,\ q_2 = 0.03$ のときは

$$\phi = \frac{0.06}{0.03} = 2,\ \psi = \frac{0.06/0.94}{0.03/0.97} = 2.06 \tag{5}$$

曝露オッズ比と発症オッズ比 [8]

- 発症オッズ比 ψ_D, Case-Control 研究の曝露オッズ比 ψ_E
- 疾患 D, 非疾患 $\bar{D}$, 曝露 E, 非曝露 $\bar{E}$
- ベイズの定理より (7) 式

$$\psi_D = \frac{\Pr(D \mid E)[1 - \Pr(D \mid \bar{E})]}{\Pr(D \mid \bar{E})[1 - \Pr(D \mid E)]},\ \ \psi_E = \frac{\Pr(E \mid D)[1 - \Pr(E \mid \bar{D})]}{\Pr(E \mid \bar{D})[1 - \Pr(E \mid D)]} \tag{6}$$

$$\begin{aligned}\psi_D &= \frac{\frac{\Pr(E|D)\Pr(D)}{\Pr(E|D)\Pr(D)+\Pr(E|\bar{D})\Pr(\bar{D})} \frac{\Pr(\bar{E}|\bar{D})\Pr(\bar{D})}{\Pr(\bar{E}|D)\Pr(D)+\Pr(\bar{E}|\bar{D})\Pr(\bar{D})}}{\frac{\Pr(\bar{E}|D)\Pr(D)}{\Pr(\bar{E}|D)\Pr(D)+\Pr(\bar{E}|\bar{D})\Pr(\bar{D})} \frac{\Pr(E|D)\Pr(D)}{\Pr(\bar{E}|D)\Pr(D)+\Pr(E|\bar{D})\Pr(\bar{D})}} \\ &= \frac{\Pr(E \mid D)[1 - \Pr(E \mid \bar{D})]}{\Pr(E \mid \bar{D})[1 - \Pr(E \mid D)]} = \psi_E\end{aligned} \tag{7}$$

- 発症オッズ比 ψ_D と曝露オッズ比 ψ_E は等しい
- 以降ケース・コントロール研究の話に集中します

二項分布モデル

- 二項分布の確率関数の積を, ψ と p_2 で表わして変形する
- 後半は指数型分布属の形, $X_{+1} = X_{11} + X_{21}$ が $p_2/(1-p_2)$ の十分統計量

$$\begin{aligned}\Pr(X_{11} = x_{11},\ X_{21} = x_{21}) &= \binom{n_{1+}}{x_{11}} \left(\frac{\psi p_2}{1 - p_2(1-\psi)}\right)^{x_{11}} \left(\frac{1-p_2}{1-p_2(1-\psi)}\right)^{n_{1+}-x_{11}} \times \\ &\quad \binom{n_{2+}}{x_{21}} p_2^{x_{21}} (1-p_2)^{n_{2+}-x_{21}} \\ &= \left[1 + \exp\left\{\log\psi + \log\left(\frac{p_2}{1-p_2}\right)\right\}\right]^{-n_{1+}} \times \\ &\quad \left[1 + \exp\left\{\log\left(\frac{p_2}{1-p_2}\right)\right\}\right]^{-n_{2+}} \times \\ &\quad \binom{n_{1+}}{x_{11}}\binom{n_{2+}}{x_{21}} \exp\left\{x_{11}\log\psi + x_{+1}\log\left(\frac{p_2}{1-p_2}\right)\right\}\end{aligned} \tag{8}$$

二項分布モデルの最尤推定量と独立性の検定

- オッズ比の最尤推定量 $\hat{\psi}$

$$\hat{\psi} = \frac{X_{11}/(n_{1+} - X_{11})}{X_{21}/(n_{2+} - X_{21})} \tag{9}$$

- 独立性の検定
- 連続修正なし: $c = 0$, あり: $c = 1/2$ (以降同様の記号を使います)

$$\begin{cases} H_0: & \psi = 1 \\ H_1: & \psi \neq 1 \end{cases} \tag{10}$$

$$X^2 = \frac{(|X_{11}(n_{2+} - X_{21}) - X_{21}(n_{1+} - X_{11})| - c)^2}{n_{1+}n_{2+}X_{+1}(n_{++} - X_{+1})} \sim \chi^2(1) \tag{11}$$

非心超幾何分布モデル (条件付き推測)

表: 非心超幾何分布モデル

確率変数	発症あり	発症なし	合計
曝露あり	X_{11}		n_{1+}
曝露なし			n_{2+}
合計	$X_{+1} = n_{+1}$		n_{++}

- 局外パラメータ $p_2/(1-p_2)$ の十分統計量 $X_{+1} = X_{11} + X_{21}$ の条件付き分布に基づいて推測する
- この場合, 条件付き分布は興味のあるパラメータ ψ にしか依存しない
- 指数型分布属の重要な性質

条件付き分布と標本空間

$$\Pr(X_{11} = x_{11} \mid X_{+1} = n_{+1}) = \frac{\Pr(X_{11} = x_{11}, X_{+1} = n_{+1})}{\Pr(X_{+1} = n_{+1})} = \frac{\binom{n_{1+}}{x_{11}}\binom{n_{2+}}{n_{+1} - x_{11}}\psi^{x_{11}}}{\sum_{u \in \Omega_X}\binom{n_{1+}}{u}\binom{n_{2+}}{n_{+1} - u}\psi^{u}} \tag{12}$$

- 集合 Ω_X は十分統計量 $X_{+1} = n_{+1}$ を与えたもとでの標本空間

$$\Omega_X = \{u \in \mathbb{Z}_+ \mid \max(0,\ n_{+1} - n_{2+}) \le u \le \min(n_{1+},\ n_{+1})\} \tag{13}$$

- 無条件分布の場合の標本空間は Ω_S

$$\Omega_S = \{(u,\ v) \in \mathbb{Z}_+^2 \mid 0 \le u \le n_{1+},\ 0 \le v \le n_{2+}\} \tag{14}$$

非心超幾何分布モデルの条件付き最尤推定量

- オッズ比の条件付き最尤推定量 $\hat{\psi_C}$ は (15) 式を満たす解
- (15) 式はスコア関数=0 とおいたもの

$$x_{11} = \mathrm{E}[X_{11} \mid X_{+1} = n_{+1},\ \hat{\psi_C}] \tag{15}$$

$$\mathrm{E}[X_{11} \mid X_{+1} = n_{+1},\ \hat{\psi_C}] = \frac{\sum_{u \in \Omega_X} u\binom{n_{1+}}{u}\binom{n_{2+}}{n_{+1} - u}\hat{\psi_C}^{u}}{\sum_{u \in \Omega_X}\binom{n_{1+}}{u}\binom{n_{2+}}{n_{+1} - u}\hat{\psi_C}^{u}} \tag{16}$$

非心超幾何分布モデルの独立性の検定

- 独立性の検定 (Fisher's exact test), 仮説は (10) 式と同じ
- 帰無仮説のもとで, X_{11} は (12) 式の超幾何分布に従う
- 両側 P 値の計算には流儀がある, 何の統計量に基づいているかによって異なる

$$\Pr(X_{11} = x_{11} \mid X_{+1} = n_{+1}, \psi = 1) = \frac{\binom{n_{1+}}{x_{11}}\binom{n_{2+}}{n_{+1} - x_{11}}}{\binom{n_{++}}{n_{+1}}} \tag{17}$$

$$\begin{aligned} &(a)\ \ P\text{–Value} = 2\Pr(X_{11} \ge x_{11}) \\ &(b)\ \ P\text{–Value} = \Pr(Z_F \le z_F), \quad Z_F = \Pr(X_{11}),\ z_F = \Pr(x_{11}) \\ &(c)\ \ P\text{–Value} = \Pr(|X_{11} - \mathrm{E}_0[X_{11}]| \ge |x_{11} - \mathrm{E}_0[X_{11}]|) \end{aligned} \tag{18}$$

層別解析 (Stratified analysis)

- 後向き試験データにおける 2×2 表の解析
  - すばらしいデータ?
  - 2×2 表では明らかに交絡が存在する場合は, 層別解析やマッチング
- 層別解析 (Stratified analysis)
  層には分けるが最終的には全体の効果以外は見えない
- サブグループ解析 (Subgroup analysis)
  グループそれぞれの効果を見たい
- 多変量解析 (Multivariate analysis)
  さらに仮定をおいた数理モデルを使って, 他の因子の効果も見る

$2 \times 2 \times k$ 表

- k 個の 2×2 表
- 結果因子や曝露因子以外の因子で, さらに分けられた分割表

表: $2 \times 2 \times k$ 表

セル確率	曝露あり	曝露なし	合計
発症あり	p_{1k}	$1 - p_{1k}$	1
発症なし	p_{2k}	$1 - p_{2k}$	1

確率変数	曝露あり	曝露なし	合計
発症あり	X_{11k}		n_{1+k}
発症なし	X_{21k}		n_{2+k}
合計			n_{++k}

積二項分布モデル

- 各層が独立であることを仮定して, 二項分布の積で表現したモデル
- $X_{+1k} = X_{11k} + X_{21k}$ が $p_{2k}/(1 - p_{2k})$ の十分統計量

$$
\begin{aligned}
&\Pr(X_{11k} = x_{11k},\ X_{21k} = x_{21k}) = \\
&\quad \prod_k \left[1 + \exp\left\{\log\psi_k + \log\left(\frac{p_{2k}}{1 - p_{2k}}\right)\right\}\right]^{-n_{1+k}} \times \\
&\quad \left[1 + \exp\left\{\log\left(\frac{p_{2k}}{1 - p_{2k}}\right)\right\}\right]^{-n_{2+k}} \times \\
&\quad \binom{n_{1+k}}{x_{11k}}\binom{n_{2+k}}{x_{21k}} \exp\left\{\sum_k x_{11k}\log\psi_k + \sum_k x_{+1k}\log\left(\frac{p_{2k}}{1 - p_{2k}}\right)\right\}
\end{aligned} \tag{19}
$$

共通オッズ比 (summary relative risk/common odds ratio)

- 各層のオッズ比が等しいと仮定する
- 曝露因子と交絡因子の間に交互作用が存在しない状況

$$\psi_1 = \psi_2 = \cdots = \psi_k = \psi \tag{20}$$

- オッズ比共通の仮定をおいた積二項分布モデル

$$
\begin{aligned}
&\Pr(X_{11k} = x_{11k},\ X_{21k} = x_{21k}) = \\
&\quad \prod_k \left[1 + \exp\left\{\log\psi + \log\left(\frac{p_{2k}}{1 - p_{2k}}\right)\right\}\right]^{-n_{1+k}} \times \\
&\quad \left[1 + \exp\left\{\log\left(\frac{p_{2k}}{1 - p_{2k}}\right)\right\}\right]^{-n_{2+k}} \times \\
&\quad \binom{n_{1+k}}{x_{11k}}\binom{n_{2+k}}{x_{21k}} \exp\left\{\left(\sum_k x_{11k}\right)\log\psi + \sum_k x_{+1k}\log\left(\frac{p_{2k}}{1 - p_{2k}}\right)\right\}
\end{aligned} \tag{21}
$$

積非心超幾何分布モデル

確率変数	発症あり	発症なし	合計
曝露あり	X_{11k}		n_{1+k}
曝露なし			n_{2+k}
合計	$X_{+1k} = n_{+1k}$		n_{++k}

- オッズ比共通の仮定をおいて, 局外パラメータ $p_{2k}/(1-p_{2k})$ の十分統計量 $X_{+1k} = X_{11k} + X_{21k}$ の条件付き分布に基づいて推測する
- Ω_{X_k} は層 k の十分統計量を与えたもとでの標本空間

$$\Pr(X_{11k} = x_{11k} \mid X_{+1k} = n_{+1k}) = \prod_k \frac{\binom{n_{1+k}}{x_{11k}}\binom{n_{2+k}}{n_{+1k}-x_{11k}}\psi^{x_{11k}}}{\sum_{u_k \in \Omega_{X_k}} \binom{n_{1+k}}{x_{11k}}\binom{n_{2+k}}{n_{+1k}-x_{11k}}\psi^{u_k}} \tag{22}$$

$$\Omega_{X_k} = \{u_k \in \mathbb{Z}_+ \mid \max(0,\ n_{+1k} - n_{2+k}) \le u_k \le \min(n_{1+k},\ n_{+1k})\} \tag{23}$$

積非心超幾何分布モデルの条件付き最尤推定量

- 共通オッズ比の条件付き最尤推定量 $\hat{\psi_C}$ は (24) 式を満たす解
- (24) 式はスコア関数=0 とおいたもの

$$\sum_k x_{11k} = \sum_k \mathrm{E}[X_{11k} \mid X_{+1k} = n_{+1k},\ \hat{\psi_C}] \tag{24}$$

$$\mathrm{E}[X_{11k} \mid X_{+1k} = n_{+1k},\ \hat{\psi_C}] = \frac{\sum_{u_k \in \Omega_{X_k}} u_k \binom{n_{1+k}}{u_k}\binom{n_{2+k}}{n_{+1k}-u_k}\hat{\psi_C}^{u_k}}{\sum_{u_k \in \Omega_{X_k}} \binom{n_{1+k}}{u_k}\binom{n_{2+k}}{n_{+1k}-u_k}\hat{\psi_C}^{u}} \tag{25}$$

5 つの共通オッズ比の推定量 I

- Mantel-Haenszel (1959) の提案のうちの一つ
- 提案する共通オッズ比 R と, Haenszel, et al. (1954) の R_1, Wynder, et al. (1954) の R_2 のほかに, R_3 と R_4 を示している

$$\begin{array}{ll} A_k = X_{11k} & B_k = n_{1+k} - X_{11k} \\ C_k = n_{+1k} - X_{11k} & D_k = n_{2+k} - n_{+1k} - X_{11k} \end{array} \tag{26}$$

$$\begin{aligned} R &= \frac{\sum A_k D_k / n_{+1k}}{\sum B_k C_k / n_{+2k}} \\ R_1 &= \frac{\sum_k A_k \sum_k D_k}{\sum_k B_k \sum_k C_k} \Bigg/ \frac{\sum_k \mathrm{E}[A_k] \sum_k \mathrm{E}[D_k]}{\sum_k \mathrm{E}[B_k] \sum_k \mathrm{E}[C_k]} \\ R_2 &= \frac{\sum_k A_k \sum_k (n_{1+k} \frac{D_k}{n_{2+k}})}{\sum_k B_k \sum_k (n_{1+k} \frac{C_k}{n_{2+k}})} \end{aligned} \tag{27}$$

5 つの共通オッズ比の推定量 II

$$\begin{aligned} R_3 &= \frac{\sum_k (n_{2+k} \frac{A_k}{n_{1+k}}) \sum_k D_k}{\sum_k (n_{2+k} \frac{B_k}{n_{1+k}}) \sum_k C_k} \\ R_4 &= \frac{\sum_k (n_{++k} \frac{A_k}{n_{1+k}}) \sum_k (n_{++k} \frac{D_k}{n_{2+k}})}{\sum_k (n_{++k} \frac{B_k}{n_{1+k}}) \sum_k (n_{++k} \frac{C_k}{n_{2+k}})} \end{aligned} \tag{28}$$

- R, R_1 は $\sum X_{11k} = \sum \mathrm{E}[X_{11k} \mid n_{+1k}, \psi = 1]$ の時 1 になる共通オッズ比の推定量
- R_1 の分子は, 層を無視して 1 つの 2×2 分割表にしたときのオッズ比の推定量. 分母は, 各層の H_0 のもとでの期待値の和から計算するので 1 に近い値. R_1 は $\psi = 1$ の方向にバイアスがある
- R_4 は, 標準化リスクを用いたもの. R_2, R_3 は, 標準化に近い補正をしたもの. 共通オッズ比というよりは標準化に近い? R_2 は n_{2+k} が, R_3 は n_{1+k} が, R_4 はどちらかが 0 の場合定義できない.

共通オッズ比に対する独立性の検定

- Mantel-Haenszel (1959) の提案のうちの一つ
- 積非心超幾何分布モデルの $\psi_C = 1$ のもとでの条件付きスコア検定にも一致

$$\begin{cases} H_0: & \psi_1 = \psi_2 = \ldots = \psi_k = \psi = 1 \\ H_1: & \psi_1 = \psi_2 = \ldots = \psi_k = \psi \end{cases} \tag{29}$$

$$\chi^2_{MH} = \frac{(|\sum_k X_{11k} - \sum_k \mathrm{E}[X_{11k}]| - c)^2}{\sum_k \mathrm{V}[X_{11k}]} \sim \chi^2(1) \tag{30}$$

$$\begin{aligned} \mathrm{E}[X_{11k}] &= \frac{n_{1+k}n_{+1k}}{n_{++k}} \\ \mathrm{V}[X_{11k}] &= \frac{n_{1+k}n_{2+k}n_{+1k}n_{+2k}}{n_{++k}^2(n_{++k}-1)} \end{aligned} \tag{31}$$

Cochran-Mantel-Haenszel 検定

- Cochran (1954; pp.443–446)[4] も似たような検定だから?
- 積二項分布モデルを仮定し, 各層のリスク差の重み付き平均にもとづく統計量
- Mantel-Haenszel オッズ比 R とは関係ない
- 分散の推定量が少し違う (二項分布ベースなので $np(1-p)$ の形)

$$\mathrm{V}^C[X_{11k}] = \frac{n_{1+k}n_{2+k}n_{+1k}n_{+2k}}{n_{++k}^3} \tag{32}$$

- Cochran (1954) では他にも
  - カテゴリの併合について
  - 分割表の χ^2 検定の使い分けの基準 (正確検定, 連続性の補正, 通常の検定)
  - Cochran-Armitage trend test を提案

1:1 マッチング

- Case と Control が 2 人 1 組という条件が付く
- 多項分布モデル, $M(z_{++}, \pi_{11}, \pi_{12}, \pi_{21})$

表: 1:1 マッチング

セル確率		Control 曝露	Control 非曝露	合計
Case	曝露	π_{11}	π_{12}	p_1
	非曝露	π_{21}	π_{22}	$1-p_1$
合計		p_2	$1-p_2$	1

確率変数		Control 曝露	Control 非曝露	合計
Case	曝露	Z_{11}	Z_{12}	
	非曝露	Z_{21}	Z_{22}	
合計				z_{++}

1:1 マッチングの Mantel-Haenszel 推定量

- 元々の分割表は ⋯
- $2 \times 2 \times k$ $(= Z_{11} + Z_{12} + Z_{21} + Z_{22})$ 表
- 層の数が無限に増加する極限モデル, 無条件の最尤推定量は ψ^2 に収束 [3, 7]

	曝露	非曝露
Case	1	0
Control	1	0

$\times Z_{11}$

	曝露	非曝露
Case	1	0
Control	0	1

$\times Z_{12}$

	曝露	非曝露
Case	0	1
Control	1	0

$\times Z_{21}$

	曝露	非曝露
Case	0	1
Control	0	1

$\times Z_{22}$

1:1 マッチングの Mantel-Haenszel 推定量

- 共通オッズ比の Mantel-Haenszel 推定量 R は

$$R = \frac{\sum X_{11k}(n_{2+k} - n_{+1k} + X_{11k})/n_{+1k}}{\sum(n_{1+k} - X_{11k})(n_{+1k} - X_{11k})/n_{+2k}} = \frac{Z_{12}}{Z_{21}} \tag{33}$$

- 共通オッズ比の条件付き最尤推定量は (34) 式を満たす解
- 1:1 マッチングの時のみ $R = \hat{\psi_C}$

$$\sum_k x_{11k} = \sum_k \frac{\displaystyle\sum_{u\in\Omega_X} u\binom{n_{1+}}{u}\binom{n_{2+}}{n_{+1}-u}\hat{\psi_C}^{u}}{\displaystyle\sum_{u\in\Omega_X}\binom{n_{1+}}{u}\binom{n_{2+}}{n_{+1}-u}\hat{\psi_C}^{u}} = \sum_k \frac{\displaystyle\sum_{u\in\Omega_X} u\hat{\psi_C}^{u}}{\displaystyle\sum_{u\in\Omega_X}\hat{\psi_C}^{u}} \tag{34}$$

$$Z_{11} + Z_{12} + 0 + 0 = Z_{11} + Z_{12}\frac{\hat{\psi_C}}{1+\hat{\psi_C}} + Z_{21}\frac{\hat{\psi_C}}{1+\hat{\psi_C}} + 0$$

$$\hat{\psi_C} = \frac{Z_{12}}{Z_{21}} \tag{35}$$

1:1 マッチングの Mantel-Haenszel 検定

- Mantel-Haenszel 検定は

$$\begin{aligned}\chi^2_{MH} &= \frac{\left(\sum_k X_{11k} - \sum_k \dfrac{n_{1+k}n_{+1k}}{n_{++k}}\right)^2}{\sum_k \dfrac{n_{1+k}n_{2+k}n_{+1k}n_{+2k}}{n^2_{++k}(n_{++k}-1)}} \\ &= \frac{(Z_{11} + Z_{12} + 0 + 0 - \frac{2Z_{11}+Z_{12}+Z_{21}+0}{2})^2}{\frac{0+Z_{12}+Z_{21}+0}{4}} = \frac{(Z_{12}-Z_{21})^2}{Z_{12}+Z_{21}}\end{aligned} \tag{36}$$

McNemar 検定 [6]

- 対応のある 2×2 分割表に対する周辺同等性 (marginal homogenity) の検定
- 2×2 表なので, d に対する推測に帰着する

$$\begin{cases} H_0: & p_1 = p_2 \ \ (d = \pi_{12} - \pi_{21} = 0) \\ H_1: & p_1 \neq p_2 \ \ (d \neq 0) \end{cases} \tag{37}$$

$$X^2_{Mc} = \frac{\hat{d}^2}{\hat{\mathrm{V}}[\hat{d} \mid H_0]} = \frac{(\mid Z_{12} - Z_{21} \mid - c)^2}{Z_{12} + Z_{21}} \tag{38}$$

$$\hat{d} = \hat{\pi}_{12} - \hat{\pi}_{21} = \frac{Z_{12} - Z_{21}}{z_{++}}$$

$$\mathrm{V}[\hat{d}] = \frac{\mathrm{V}[Z_{12}] + \mathrm{V}[Z_{21}] - 2\mathrm{Cov}[Z_{12},\ Z_{21}]}{z^2_{++}} \tag{39}$$

$$\hat{\mathrm{V}}[\hat{d} \mid H_0] = \frac{\hat{\pi}_{12} + \hat{\pi}_{21}}{z_{++}} = \frac{Z_{12} + Z_{21}}{z^2_{++}}$$

$i\times j\times k$ 表

セル確率	曝露因子				合計
結果因子	B_1	B_2	$\cdots$	B_j	
A_1	π_{11k}	π_{12k}	$\cdots$	π_{1jk}	1
A_2	π_{21k}	π_{22k}	$\cdots$	π_{2jk}	1
$\vdots$	$\vdots$	$\vdots$	$\ldots$	$\vdots$	$\vdots$
A_i	π_{i1k}	π_{i2k}	$\cdots$	π_{ijk}	1

確率変数	曝露因子	B_1	B_2	$\cdots$	B_j	合計
結果因子	スコア	c_{1k}	c_{2k}	$\cdots$	c_{jk}	
A_1	r_{1k}	X_{11k}	X_{12k}	$\cdots$	X_{1jk}	n_{1+k}
A_2	r_{2k}	X_{21k}	X_{22k}	$\cdots$	X_{2jk}	n_{2+k}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ldots$	$\vdots$	$\vdots$
A_i	r_{ik}	X_{i1k}	X_{i2k}	$\cdots$	X_{ijk}	n_{i+k}
合計						n_{++k}

積多項分布モデル

$$\Pr(X_{ijk} = x_{ijk}) = \prod_i \prod_k \frac{n_{i+k}!}{\prod_j x_{ijk}!} \prod_j p_{ijk}^{x_{ijk}} \tag{40}$$

$$\psi_{ijk} = \frac{p_{ijk} p_{11k}}{p_{i1k} p_{1jk}} \tag{41}$$

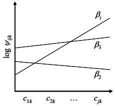
$$\Pr(X_{ijk} = x_{ijk}) = c(\boldsymbol{\psi}, \boldsymbol{\theta}) h(\boldsymbol{x}) \exp\left(\sum_{i=2} \sum_{j=2} \sum_k x_{ijk} \log \psi_{ijk} + \sum_{j=2} \sum_k x_{+jk} \theta_{jk} \right) \tag{42}$$

$$\theta_{jk} = \log\left(\frac{p_{1jk}}{p_{11k}}\right),\ c(\boldsymbol{\psi}, \boldsymbol{\theta}) = \prod_i \prod_k \left(1 + \sum_{j=2} \exp(\log \psi_{ijk}) \exp(\theta_{jk})\right)^{-n_{i+k}}$$

$$h(\boldsymbol{x}) = \prod_i \prod_k \frac{n_{i+k}!}{\prod_j x_{ijk}!} \tag{43}$$

カテゴリに順序がある場合 [8]

- 行のスコアを $(r_{1k} \le r_{2k} \le \cdots \le r_{ik})$
- 列のスコアを $(c_{1k} \le c_{2k} \le \cdots \le c_{jk})$
- 片方にスコアを割り付けて, $\log \psi_{ijk} = (c_{jk} - c_{1k})\beta_i$ の仮定をおいた検定 → 一般化拡張 Mantel 検定
- 両方にスコアを割り付けて, $\log \psi_{ijk} = (r_{ik} - r_{1k})(c_{jk} - c_{1k})\beta$ の仮定をおいた検定 → 拡張相関検定
- 行の数が 2 で, スコアが層に無関係 $c_{jk} = c_j$ の時 → 一般化 Mantel 検定 [5]



カテゴリに順序がある場合のモデル

$$\Pr(X_{ijk} = x_{ijk}) = c(\boldsymbol{\psi}, \boldsymbol{\theta}) h(\boldsymbol{x}) \exp\left(\sum_{i=2} \sum_{j=2} \sum_k x_{ijk} (c_{jk} - c_{1k}) \beta_i + \sum_{j=2} \sum_k x_{+jk} \theta_{jk} \right) \tag{44}$$

$$c(\boldsymbol{\beta}, \boldsymbol{\theta}) = \prod_i \prod_k \left(1 + \sum_j \exp[\log\{(c_{jk} - c_{1k})\beta_i\}] \exp(\theta_{jk})\right)^{-n_{i+k}} \tag{45}$$

- 興味のあるパラメータ β_i に対する十分統計量 $W_i = \sum_j \sum_k x_{ijk}(c_{jk} - c_{1k})$
- 局外パラメータ θ_{jk} に対する十分統計量 X_{+jk} の条件付き分布に基づいて考えると

$$\Pr(X_{ijk} = x_{ijk} \mid X_{+jk} = n_{+jk}) = \frac{h(\boldsymbol{x}) \exp(\sum_{i=2} W_i \beta_i)}{\sum_{\boldsymbol{u} \in \Omega_{X_{jk}}} h(\boldsymbol{u}) \exp(\sum_{i=2} W_i \beta_i)} \tag{46}$$

一般化 Mantel 検定

- 列の数が 2 で, スコアが層に無関係 $(c_{jk} = c_j)$ で, $c_1 = 0$ の時, $W = \sum_j \sum_k X_{2jk} c_j$
- 列が 2 つなのでパラメータが 1 つになる, 以下の仮説に対する検定

$$\begin{cases} H_0: & \beta = 0 \\ H_1: & \beta \neq 0 \end{cases} \tag{47}$$

$$\Pr(X_{ijk} = x_{ijk} \mid X_{+jk} = n_{+jk};\ \beta = 0) = \prod_k \frac{\prod_j \binom{n_{+jk}}{x_{2jk}}}{\binom{n_{++k}}{n_{2+k}}} \tag{48}$$

$$\mathrm{E}[W \mid X_{+jk} = n_{+jk};\ \beta = 0] = \textstyle\sum_k n_{2+k} \sum_j \frac{c_j n_{+jk}}{n_{++k}}$$

$$\mathrm{V}[W \mid X_{+jk} = n_{+jk};\ \beta = 0] = \textstyle\sum_k \frac{n_{1+k} n_{2+k}}{n_{++k}(n_{++k} - 1)} \{n_{++k} \sum_j c_j^2 n_{+jk} - (\sum_j c_j n_{+jk})^2\} \tag{49}$$

$$\chi^2_{EMH} = \frac{(W - \mathrm{E}[W \mid X_{+jk} = n_{+jk};\ \beta = 0])^2}{\mathrm{V}[W \mid X_{+jk} = n_{+jk};\ \beta = 0]} \sim \chi^2(1) \tag{50}$$

参考文献

[1] Agresti A. A survey of exact inference for contingency tables. *Statistical Science* 1992; **7**(1): 131–177.

[2] Agresti A. *Categorical data analysis. 2nd edition.* New York: John Wiley & Sons 2002.

[3] Breslow N. Odds ratio estimators when the data are sparse. *Biometrika* 1981; 68(1): 73–84.

[4] Cochran WG. Some methods for strengthening the common χ^2 tests. *Biometrics* 1954; **10**(4): 417–451.

[5] Mantel N. Chi-square tests with one degree of freedom; extensions of the Mantel-Haenszel procedure. *Journal of the American Statistical Association* 1963; **58**(303): 690–700.

[6] McNemar Q. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika* 1947; **12**(2): 153–157.

[7] 佐藤俊哉, 高木廣文, 柳川堯, 柳本武美. Mantel-Haenszel の方法による複数の 2 × 2 表の要約. 統計数理 1998; **46**(1): 153–177.

[8] 柳川堯. 離散多変量データの解析. 共立出版 1986.

超幾何分布の期待値

- $\mathrm{E}[X_{11} \mid n_{+1},\ \psi = 1]$ 等は省略して $\mathrm{E}[X_{11}]$ と表わす

$$\mathrm{E}[X_{11}] = \sum_{u\in\Omega_X} u \Pr[u] = \frac{\displaystyle\sum_{u\in\Omega_X} u\binom{n_{1+}}{u}\binom{n_{2+}}{n_{+1}-u}}{\displaystyle\sum_{u\in\Omega_X}\binom{n_{1+}}{u}\binom{n_{2+}}{n_{+1}-u}}$$

$$= \frac{1}{\binom{n_{++}}{n_{+1}}} \sum_{u\in\Omega_X} \frac{n_{1+}(n_{1+}-1)!}{(u-1)!\,(n_{1+}-u)!}\binom{n_{2+}}{n_{+1}-u} = \frac{n_{1+}\binom{n_{++}-1}{n_{+1}-1}}{\binom{n_{++}}{n_{+1}}} = \frac{n_{1+}n_{+1}}{n_{++}} \tag{51}$$

$\sum_{u=0}^{t}\binom{n}{u}\binom{m}{t-u} = \binom{n+m}{t-u+u}$

超幾何分布の分散

$$\begin{aligned}
\mathrm{V}[X_{11}] &= \mathrm{E}[X_{11}^2] - \{\mathrm{E}[X_{11}]\}^2 \\
&= \sum_{u\in\Omega_X} u(u-1)\Pr[u] + \sum_{u\in\Omega_X} u\Pr[u] - \{\mathrm{E}[X_{11}]\}^2 \\
&= \frac{n_{1+}(n_{1+}-1)n_{+1}(n_{+1}-1)}{n_{++}(n_{++}-1)} + \frac{n_{1+}n_{+1}}{n_{++}} - \frac{n_{1+}^2 n_{+1}^2}{n_{++}^2} \qquad (52) \\
&= \frac{n_{1+}(n_{++}-n_{1+})n_{+1}(n_{++}-n_{+1})}{n_{++}^2(n_{++}-1)}
\end{aligned}$$

- 積非心超幾何分布モデルは, 帰無仮説のもとで k 個の独立な超幾何分布の積になる