Comparison of confidence intervals for the difference between independent proportions

Masato Iiduka, Chikuma Hamada
Graduate school of Engineering, Tokyo University of Science
信頼区間の構成の現状（1）

- 2群の割合の差の信頼区間
  - 医薬統計では両側95%信頼区間
- 多くの医薬統計の教科書はWald型の信頼区間を使用[1][2]

\[
p_1 - p_2 \pm z_{\alpha/2} \sqrt{\frac{p_1 (1 - p_1)}{n_1} + \frac{p_2 (1 - p_2)}{n_2}}
\]

Wald信頼区間

比率(リスク)差の信頼区間

<table>
<thead>
<tr>
<th>列 1 (res = 0)</th>
<th>比率差 = -0.5333</th>
</tr>
</thead>
</table>

タイプレンジ 95% 信頼限界

Wald 0.7413 -0.3254

信頼区間の構成の現状（2）

- SAS出力結果⇒デフォルトはWaldを出力

```sas
PROC FREQ DATA = DATA;
  TABLES EXPOSURE*RESPONSE/RISKDIFF;
  WEIGHT COUNT;
RUN;
```

SAS出力結果

<table>
<thead>
<tr>
<th></th>
<th>リスク</th>
<th>漸近標準誤差</th>
<th>(漸近) 95% 信頼区間</th>
<th>(正確) 95% 信頼限界</th>
</tr>
</thead>
<tbody>
<tr>
<td>行 1</td>
<td>0.1333</td>
<td>0.0621</td>
<td>0.0117</td>
<td>0.2550</td>
</tr>
<tr>
<td>行 2</td>
<td>0.6667</td>
<td>0.0861</td>
<td>0.4980</td>
<td>0.8354</td>
</tr>
<tr>
<td>合計</td>
<td>0.4000</td>
<td>0.0632</td>
<td>0.2760</td>
<td>0.5200</td>
</tr>
<tr>
<td>差</td>
<td>-0.5333</td>
<td>0.1061</td>
<td>-0.7413</td>
<td>-0.3254</td>
</tr>
</tbody>
</table>
2×2分割表

<table>
<thead>
<tr>
<th></th>
<th>有効</th>
<th>無効</th>
<th>計</th>
<th>有効割合</th>
<th>母数</th>
</tr>
</thead>
<tbody>
<tr>
<td>薬剤群</td>
<td>(n_{11})</td>
<td>(n_{12})</td>
<td>(n_{1})</td>
<td>(p_1 = \frac{n_{11}}{n_{1}})</td>
<td>(\pi_1)</td>
</tr>
<tr>
<td>対照群</td>
<td>(n_{21})</td>
<td>(n_{22})</td>
<td>(n_{2})</td>
<td>(p_2 = \frac{n_{21}}{n_{2}})</td>
<td>(\pi_2)</td>
</tr>
<tr>
<td>計</td>
<td>(n_{1})</td>
<td>(n_{2})</td>
<td>(n)</td>
<td>(p = \frac{n_{1}}{n})</td>
<td></td>
</tr>
</tbody>
</table>

2群の有効割合の差の真値: \(\Delta = \pi_1 - \pi_2\)
2群の有効割合の差: \(\hat{\Delta} = d = p_1 - p_2\)

2群の割合の差の信頼区間[3] [4]

- Farrington-Manning信頼区間
- Hauck-Anderson信頼区間
- Newcombeスコア信頼区間
- Newcombeスコア(連続修正)信頼区間
- Wald信頼区間
- Wald(連続修正)信頼区間
- 正確な検定に基づく信頼区間
- 正確な検定に基づく信頼区間(FMスコア)
信頼区間の構成手法(1)

□Farrington-Manning信頼区間

\[ p_1 - p_2 \pm \frac{z_{\alpha/2}}{2} \sqrt{\frac{p(1-p)}{n_1} + \frac{p(1-p)}{n_2}} \]

\[ z_{\alpha/2} : \text{標準正規分布の} 100(1 - \frac{\alpha}{2})\% \text{点} \]

□Hauck-Anderson信頼区間

\[ p_1 - p_2 \pm \left( cc + \frac{z_{\alpha/2}}{2} \sqrt{\frac{p_1(1-p_1)}{n_1-1} + \frac{p_2(1-p_2)}{n_2-1}} \right) \]

連続修正 \(cc = \frac{1}{2 \text{min}(n_1, n_2)}\)

信頼区間の構成手法(2)

□Wald信頼区間

\[ p_1 - p_2 \pm z_{\alpha/2} \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \]

□Wald(修正)信頼区間

\[ p_1 - p_2 \pm \left( cc + z_{\alpha/2} \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \right) \]

連続修正 \(cc = \frac{1}{2n_1} + \frac{1}{2n_2}\)
Newcombeスコア信頼区間
(下限) = \( p_1 - p_2 - \sqrt{(p_1 - L_1)^2 + (p_2 - p_2)^2} \)
(上限) = \( p_1 - p_2 + \sqrt{(U_1 - p_1)^2 + (p_2 - L_2)^2} \)

\( L_1, U_1: p_1 \)のWilson score信頼区間
\( L_2, U_2: p_2 \)のWilson score信頼区間

Newcombeスコア(修正)信頼区間
(下限) = \( p_1 - p_2 - \sqrt{(p_1 - L_1')^2 + (U_2' - p_2)^2} \)
(上限) = \( p_1 - p_2 + \sqrt{(U_1' - p_1)^2 + (p_2 - L_2')^2} \)

\( L_1', U_1': p_1 \)のWilson score(連続修正)信頼区間
\( L_2', U_2': p_2 \)のWilson score(連続修正)信頼区間

正確な検定に基づく信頼区間\([6\ 7]\)
(下限) = \( \sup (\Delta: P_U(\Delta) > \frac{\alpha}{2}) \)
(上限) = \( \inf (\Delta: P_L(\Delta) > \frac{\alpha}{2}) \)

\( 2 \times 2 \)表の同時確率：\( f(n_{11}, n_{21}, \Delta, \pi_2) \)
\( = \binom{n_1}{n_{11}} (\Delta + \pi_2)^{n_{11}}(1 - \Delta - \pi_2)^{n_1 - n_{11}} \times \binom{n_2}{n_{21}}(1 - \pi_2)^{n_2 - n_{21}} \)

\( P_U(\Delta) = \sup_{\pi_2} \left( \sum_{\Delta \geq t_0} f(n_{11}, n_{21}, \Delta, \pi_2) \right) \)
\( P_L(\Delta) = \sup_{\pi_2} \left( \sum_{\Delta \leq t_0} f(n_{11}, n_{21}, \Delta, \pi_2) \right) \)

\( \Delta: 2 \times 2 \)table(\( n_1, n_2 \)) T(a):table内のある値aのときの割合の差
\( t_0: \)観測された割合の差
降圧薬の2×2分割表の実例[8]

臨床試験の結果を示す2×2分割表

<table>
<thead>
<tr>
<th></th>
<th>改善あり</th>
<th>改善なし</th>
<th>合計</th>
</tr>
</thead>
<tbody>
<tr>
<td>新薬</td>
<td>47</td>
<td>10</td>
<td>57</td>
</tr>
<tr>
<td>標準薬</td>
<td>26</td>
<td>18</td>
<td>44</td>
</tr>
<tr>
<td>合計</td>
<td>73</td>
<td>28</td>
<td>101</td>
</tr>
</tbody>
</table>

新薬改善割合: \[ p_1 = \frac{47}{57} = 0.8246 \]
標準薬改善割合: \[ p_2 = \frac{26}{44} = 0.5909 \]
改善割合の差: \[ d = 0.8246 - 0.5909 = 0.2337 \]

FREQプロシージャで8つの信頼区間を出力

```sas
PROC FREQ DATA = DATA;
   TABLES EXPOSURE*RESPONSE/
      RISKDIFF(CL=(WALD NEWCOMBE HA FM EXACT));
   WEIGHT COUNT;
   EXACT RISKDIFF;
RUN;
```

RISKDIFF(CORRECT CL=(WALD NEWCOMBE));
とすることで連続修正を加味して出力

EXACT RISKDIFF(FMSCORE);
とすることで正確な検定に基づく信頼区間（FMスコア）を出力
### 8種類の信頼区間の比較

<table>
<thead>
<tr>
<th>信頼区間タイプ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Farrington-Manning信頼区間</td>
</tr>
<tr>
<td>Hanek-Anderson信頼区間</td>
</tr>
<tr>
<td>Wald信頼区間</td>
</tr>
<tr>
<td>Wald（連続修正）信頼区間</td>
</tr>
<tr>
<td>Newcombeスコア信頼区間</td>
</tr>
<tr>
<td>Newcombeスコア（連続修正）信頼区間</td>
</tr>
<tr>
<td>正確な検定に基づく信頼区間</td>
</tr>
<tr>
<td>正確な検定に基づく信頼区間（FMスコア）</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>割合の差</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
</tr>
<tr>
<td>0.2</td>
</tr>
<tr>
<td>0.3</td>
</tr>
<tr>
<td>0.4</td>
</tr>
</tbody>
</table>

### 本研究の目的

- 症例数や割合の差の真値を変えて、8種類の手法について信頼区間の被覆確率を評価し、適切な方法を検討
- 一般的に使用されているWald信頼区間が、適切であるための条件を検討
信頼区間の被覆確率

被覆確率（Coverage Probability）: \( C(\Delta) \)

信頼区間が真値を含む確率

\[
C(\pi_1, \pi_2) = \sum_{n_{11}=0}^{n_1} \sum_{n_{21}=0}^{n_2} I(n_{11}, n_{21}, \pi_1, \pi_2) \times \frac{n_1}{n_{11}} \pi_1^{n_{11}} (1 - \pi_1)^{n_1 - n_{11}} \cdot \frac{n_2}{n_{21}} \pi_2^{n_{21}} (1 - \pi_2)^{n_2 - n_{21}}
\]

\[
I(n_{11}, n_{12}, \pi_1, \pi_2) = \begin{cases} 1, & \text{信頼区間が真値を含む} \\ 0, & \text{信頼区間が真値を含まない} \end{cases}
\]

信頼水準95%の信頼区間の場合、被覆確率が95%に近い信頼区間が望ましい

信頼区間の評価基準

2種類の評価基準で適切な信頼区間を検討

- 被覆確率が95%に最も近い信頼区間
- 被覆確率が保守的になる信頼区間
被覆確率の計算の設定

<table>
<thead>
<tr>
<th>$n_1$</th>
<th>$n_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>20</td>
<td>20設定1</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>70</td>
<td>70設定2</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>10</td>
<td>20設定3</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>90</td>
<td>110</td>
</tr>
</tbody>
</table>

各群の症例数が等しい場合

各群の症例数が異なる場合

信頼区間: 両側95%水準

\[ \pi_1 = 0.5, 0.7, 0.9 \]

\[ \pi_2 = 0.01, 0.02, \ldots, 0.99 \]

$n_{1.}, n_{2.}=20$とした場合の
被覆確率(設定1)

Farrington-Manning  Hauck-Anderson  Newcombeスコア  Newcombeスコア(修正)

Wald  Wald(修正)  正確  正確(FMスコア)

黒: $\pi_1 = 0.9$
灰色: $\pi_1 = 0.5$
被覆確率(設定2)

2013/7/3

被覆確率(％)

\( n_1 \cdot n_2 = 70 \)とした場合の

Farrington-Manning Hauck-Anderson Newcombeスコア Newcombeスコア(修正)

症例数別

Wald信頼区間の被覆確率

\( n_1 \cdot n_2 \)を変更した場合の被覆確率

黒: \( \pi_1 = 0.9 \)
灰色: \( \pi_1 = 0.5 \)
症例数別
Wald(連続修正)信頼区間の被覆確率

\[ \pi_1 = 0.9 \]

\[ \pi_1 = 0.5 \]

n_1 = n_2 = 20
n_1 = n_2 = 30
n_1 = n_2 = 40

n_1 = n_2 = 50
n_1 = n_2 = 60
n_1 = n_2 = 70

n_1 = 10, n_2 = 50とした場合の
被覆確率(設定3)

Farrington-Manning
Hauck-Anderson
Newcombeスコア
Newcombeスコア(修正)

Wald
Wald(修正)
正確
正確(FMスコア)

\[ \pi_1 = 0.9 \]

\[ \pi_1 = 0.5 \]
まとめ

- 被覆確率が95%に近い信頼区間
  - Newcombeスコア信頼区間
- 保守的である信頼区間
  - 正確な検定に基づく信頼区間（FMスコア）
  - Newcombeスコア（連続修正）信頼区間
- 症例数が70以上で等しく、真値が0や1付近でないときのみWald信頼区間の被覆確率は保たれる
- 各群の症例数が異なると全症例数が多くとも手法によっては被覆確率が保たれない

今後の課題

- 信頼区間の幅に重点を置いて構成方法を評価
- $2 \times 2 \times k$表による信頼区間の構成方法へ拡張
参考文献
