



Wilcoxon-Mann-Whitney オッズ(WMWodds)と Wilcoxon 順位和検定について

On Wilcoxon-Mann-Whitney Odds (WMWodds) and Wilcoxon Rank Sum Test

武田薬品工業株式会社 日本開発センター
クリニカルデータサイエンス部 統計グループ

工藤 健太郎、舟尾 暢男

武田薬品工業株式会社



はじめに

- ▶ 2 群間の比較において Wilcoxon 順位和検定を適用した場合、本検定に対応する推定方法としては Hodges-Lehmann 推定量が有名であるが、解析を行う状況によっては本推定方法が適切でない場合がある
- ▶ 本発表では、Hodges-Lehmann 推定量の復習と推定が悪くなる例を紹介した後、Wilcoxon-Mann-Whitney オッズ (WMWodds) による推定方法を紹介する。次に、いくつかのシミュレーションを行い、Hodges-Lehmann による推定との比較を行う



メニュー

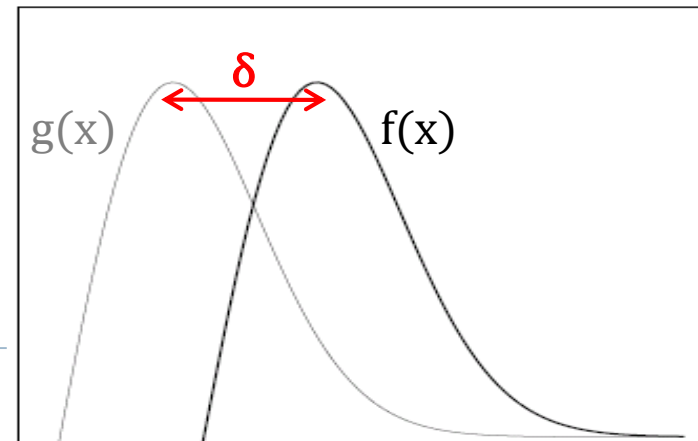
1. Wilcoxon 順位和検定と Hodges-Lehmann 型の推定方法
2. Wilcoxon-Mann-Whitney オッズ (WMWodds)
3. WMWodds と ROC 曲線下面積 (AUROC) との関係
4. 手法の比較①：Wilcoxon 順位和検定の結果との対応
5. 手法の比較②：両側 95% 信頼区間の被覆確率
6. まとめ



Wilcoxon 順位和検定

- ▶ 投与群が 2 群ある臨床試験を考える
 - ▶ 群 1（被験薬群）の応答変数： Y_{1i} ($i = 1, \dots, n_1$)
 - ▶ 群 2（対照薬群）の応答変数： Y_{2i} ($i = 1, \dots, n_2$)
- ▶ 各投与群の母集団分布の確率密度関数をそれぞれ $f(x)$ 、 $g(x)$ と表す
- ▶ Wilcoxon 順位和検定（又は、Mann-Whitney の U 検定）は、2 つの分布の形状は同じだが位置がある定数 δ だけずれている、すなわち $g(x) = f(x + \delta)$ を仮定し、以下の仮説について検定を行う

$$\begin{cases} \text{帰無仮説 } H_0: \delta = 0 \\ \text{対立仮説 } H_1: \delta \neq 0 \end{cases}$$





Hodges-Lehmann 型の推定方法

- ▶ Wilcoxon 検定に対する推定方法としては、Hodges-Lehmann 型の推定方法が一般的に用いられる

- ▶ $n_1 n_2$ 個の $d_{ij} = Y_{1i} - Y_{2j}$ を生成する

- ▶ d_{ij} を小さい順に並べたものを $d_{(1)} \leq d_{(2)} \leq \dots \leq d_{(n_1 n_2)}$ とし、

$$c_\alpha^* = \frac{n_1 n_2}{2} - Z_{1-\alpha/2} \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}} \text{ とする}$$

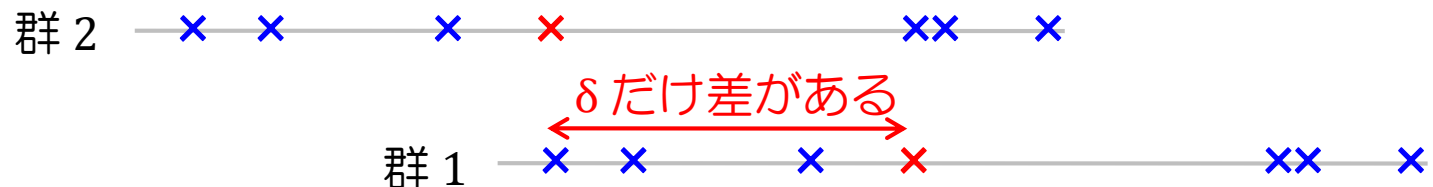
- ▶ δ の点推定値：生成した d_{ij} の中央値

- ▶ δ の両側 95% 信頼区間：

- ▶ 両側 95% 信頼区間の下限 = $d_{(c_\alpha)}$

- ▶ 両側 95% 信頼区間の上限 = $d_{(n_1 n_2 + 1 - c_\alpha)}$

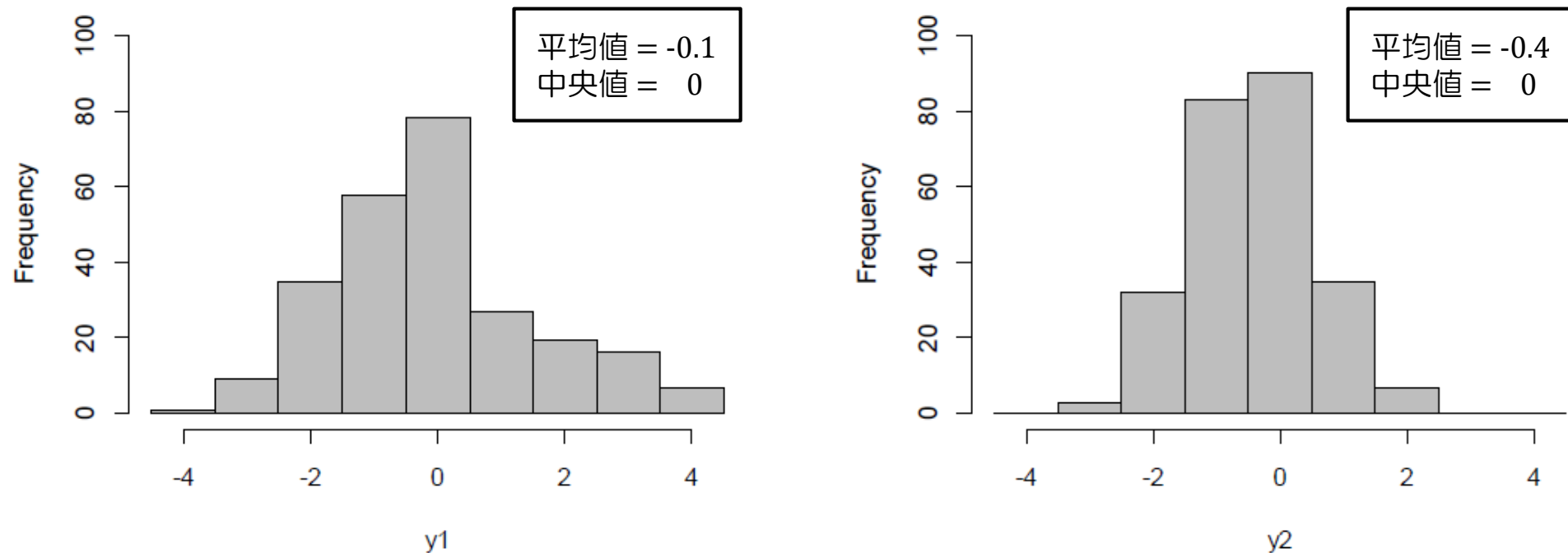
※ c_α : c_α^* 以下の値のうち最も大きい整数値





例 2.1

- ▶ 各投与群の応答変数の分布と要約統計量が下記である場合を考える
- ▶ すなわち、標本数が大きく、応答変数の実現値は整数であり、分布の中央値付近の密度が両群とも高く、分布の裾が一方の分布だけ広がる場合を考える





例 2.1

- ▶ `npar1way` プロシジャ + `hl` オプションにて Wilcoxon 順位和検定及び Hodges-Lehmann 型による δ の推定が実行できる（以下は結果の抜粋）

Wilcoxon の順位和検定 (2 標本)

正規近似

Z 1.9855
両側 Pr > |Z| 0.0471

Hodges-Lehmann 推定値
Location Shift 0.0000

95% 信頼限界
0.0000 0.0000

区間の中間点
0.0000

漸近標準誤差
0.0000

- ▶ Wilcoxon 順位和検定の結果は $p=0.0471$ となり有意差がみられるが、 δ の点推定値及び両側 95% 信頼区間の推定結果は全て 0 となり、解釈が難しい結果となる
- ▶ 特に、両側 95% 信頼区間が $[0, 0]$ となるのは問題である。



メニュー

1. Wilcoxon 順位和検定と Hodges-Lehmann 型の推定方法
2. Wilcoxon-Mann-Whitney オッズ (WMWodds)
3. WMWodds と ROC 曲線下面積 (AUROC) との関係
4. 手法の比較①：Wilcoxon 順位和検定の結果との対応
5. 手法の比較②：両側 95% 信頼区間の被覆確率
6. まとめ



Wilcoxon-Mann-Whitney オッズ (WMWodds)

- ▶ 群 1 のデータ y_{11}, \dots, y_{1n_1} と群 2 のデータ y_{21}, \dots, y_{2n_2} を小さい順に並べ「 y_{1i} 未満の群 2 のデータ数」を u_i としたとき、
 - ▶ $U = \sum_{i=1}^{n_1} u_i$ が Mann-Whitney の U 統計量となる
 - ▶ 同じ値のデータ（タイデータ）がある場合は「 y_{1i} 未満の群 2 のデータ数」に「 y_{1i} と同じ値である群 2 のデータ数の 1/2」を加えて u_i とする
 - ▶ $\#(\text{条件式})$ を「条件式を満たす要素の数」とし、以下のように略記する

$$U = \#(Y_1 > Y_2) + \frac{1}{2} \times \#(Y_1 = Y_2)$$

- ▶ Mann-Whitney の U 統計量と Wilcoxon 順位和検定の検定統計量 W との間には以下の関係が成り立ち、本質的に 2 つの統計量は同等である

$$W = U + \frac{n_1(n_1 + 1)}{2}$$



Wilcoxon-Mann-Whitney オッズ (WMWodds)

- ▶ Wilcoxon 検定と、以下の仮説に関する検定とは同等である

$$\begin{cases} \text{帰無仮説 } H_0: \pi = 0.5 \\ \text{対立仮説 } H_1: \pi \neq 0.5 \end{cases}, \quad \pi = P(Y_1 > Y_2) + \frac{1}{2}P(Y_1 = Y_2) = \frac{U}{n_1 n_2}$$

- ▶ O'Brien (2006) は、上記の π に関するオッズを定義し、これを WMWodds とした

$$\text{WMWodds} = \frac{\pi}{1 - \pi}$$

- ▶ 例えば、「WMWodds=2」は以下の様に解釈出来る
「群 1 からランダムに選んだ標本は、群 2 からランダムに選んだ標本よりも確率的に大きく、その度合いはオッズの尺度で 2 である。」



Wilcoxon-Mann-Whitney オッズ (WMWodds)

- ▶ O'Brien (2006) では WMWodds の点推定値とその両側 95% 信頼区間の計算方法と SAS マクロが提案されているが、WMWodds の計算のための SAS マクロのリンクが切れており、現在は入手できない
- ▶ そこで本発表では、O'Brien (2006) で紹介されている方法※よりも簡便に計算を行う目的で、Hanley (1982)、Mason (2002) 及び Acion (2006) 等で紹介されている「Mann-Whitney の U 統計量と、ROC 曲線の曲線下面積（以下、AUROC）との関係」を用いて、WMWodds の点推定値とその両側 95% 信頼区間の計算を行う



メニュー

1. Wilcoxon 順位和検定と Hodges-Lehmann 型の推定方法
2. Wilcoxon-Mann-Whitney オッズ (WMWodds)
3. WMWodds と ROC 曲線下面積 (AUROC) との関係
4. 手法の比較①：Wilcoxon 順位和検定の結果との対応
5. 手法の比較②：両側 95% 信頼区間の被覆確率
6. まとめ



WMWodds と ROC 曲線下面積（AUROC）との関係

- ▶ いま想定している臨床試験について、投与群（2 群）を従属変数、応答変数を独立変数とした ROC 曲線を考えると、Mann-Whitney の U 統計量と AUROC との間には以下の関係が成り立つ※

$$\text{AUROC} = \frac{U}{n_1 n_2} (= \pi)$$

- ▶ よって、WMWodds $(= \frac{\pi}{1-\pi})$ と AUROC との間には以下の関係が成り立つ

$$\text{WMWodds} = \frac{\text{AUROC}}{1 - \text{AUROC}}$$

- ▶ また、WMWodds に関する両側 95% 信頼区間は以下により計算できる

$$\left[\exp \left\{ \log \left(\frac{\pi}{1-\pi} \right) - z_{0.975} \frac{\sigma}{\pi(1-\pi)} \right\}, \exp \left\{ \log \left(\frac{\pi}{1-\pi} \right) + z_{0.975} \frac{\sigma}{\pi(1-\pi)} \right\} \right]$$

- ▶ 筆者が作成した SAS マクロ「%WMWodds」にて計算できる



SAS マクロ 「%WMWodds」

```
%macro WMWodds(_DATASET, _GROUP, _Y) ;
ods listing close ;

proc npar1way data=&_DATASET. correct=no ;
  class &_GROUP. ;
  var &_Y. ;
  ods output WilcoxonScores=WS(keep=Class SumOfScores) ;
run ;

proc transpose data=WS out=FLAG prefix=WSCORE_ ;
  var SumOfScores ;
  id CLASS ;
run ;

data _NULL_ ;
  set FLAG ;
  if (WSCORE_1 >= WSCORE_2) then call symput("_FLAG", 1) ;
  else call symput("_FLAG", 2) ;
run ;

proc logistic data=&_DATASET. ;
  class &_GROUP. ;
  model &_GROUP. =&_Y. ;
  roc ;
  ods output ROCAssociation=AUROC ;
run ;

ods listing ;
```

```
data WMWODDS ;
  set AUROC ;
  if (&_FLAG=1) then do ;
    WMWodds=Area/(1-Area) ;
    SE      =StdErr/(1-Area)**2 ;
    LnWMW   =log(Area/(1-Area)) ;
    LnSE     =StdErr/(Area*(1-Area)) ;
  end ;
  else do ;
    WMWodds=(1-Area)/Area ;
    SE      =StdErr/Area**2 ;
    LnWMW   =log((1-Area)/Area) ;
    LnSE     =StdErr/(Area*(1-Area)) ;
  end ;

  z          =quantile("NORMAL", 0.975) ;
  LowerCI     =WMWodds - z*SE ;
  UpperCI     =WMWodds + z*SE ;
  LowerCI_exp=exp(LnWMW - z*LnSE) ;
  UpperCI_exp=exp(LnWMW + z*LnSE) ;

  if _n_=1 ;
  keep Area StdErr WMWodds SE LowerCI UpperCI
      LowerCI_exp UpperCI_exp ;
run ;

title "&_DATASET." ;
proc print noobs ;
run ;
%mend ;
```



計算例

- ▶ タイがない場合とある場合の2つのデータに対して計算例を紹介する

表5.1 タイのないデータ

投与群	Y	順位	u_i
2	11	1	
1	12	2	1
2	13	3	
1	14	4	2
2	15	5	
1	16	6	3
1	17	7	3

表5.2 タイのあるデータ

投与群	Y	順位	u_i
2	11	1	
1	12	2	1
2	<u>13</u>	<u>3.5</u>	
1	<u>13</u>	<u>3.5</u>	<u>1.5</u>
2	15	5	
1	16	6	3
1	17	7	3

$$U_{5.1} = 1 + 2 + 3 + 3 = 9, \quad \pi_{5.1} = \frac{U_{5.1}}{3 \times 4} = 0.7500, \quad \text{WMWodds}_{5.1} = \frac{0.7500}{1 - 0.7500} = 3.0000$$

$$U_{5.2} = 1 + 1.5 + 3 + 3 = 8.5, \quad \pi_{5.2} = \frac{U_{5.2}}{3 \times 4} = 0.7083, \quad \text{WMWodds}_{5.2} = \frac{0.7083}{1 - 0.7083} = 2.4286$$




計算例

- ▶ U 統計量、AUROC とその標準誤差を算出するプログラムを紹介する

プログラム 5.1 (タイのないデータ)

```
data SAMPLEDATA2 ;  
  input TREAT Y ;  
  cards ;  
    2 11  
    1 12  
    2 13  
    1 14  
    2 15  
    1 16  
    1 17  
  ;  
run ;  
  
proc npar1way data=SAMPLEDATA2 correct=no ;  
  class TREAT ;  
  var Y ;  
run ;  
  
proc logistic data=SAMPLEDATA2 ;  
  class TREAT ;  
  model TREAT=Y ;  
  roc ;  
run ;
```



ROC	面積	標準誤差
モデル	0.7500	0.2152
ROC1	0.5000	0

プログラム 5.2 (タイのあるデータ)

```
data SAMPLEDATA3 ;  
  input TREAT Y ;  
  cards ;  
    2 11  
    1 12  
    2 13  
    1 13  
    2 15  
    1 16  
    1 17  
  ;  
run ;  
  
proc npar1way data=SAMPLEDATA3 correct=no ;  
  class TREAT ;  
  var Y ;  
run ;  
  
proc logistic data=SAMPLEDATA3 ;  
  class TREAT ;  
  model TREAT=Y ;  
  roc ;  
run ;
```



ROC	面積	標準誤差
モデル	0.7083	0.2282
ROC1	0.5000	0



計算例

- ▶ 表 5.1 と表 5.2 のデータに対して SAS マクロ「%WMWodds」を適用し、それぞれの WMWodds を計算する

%WMWodds (SAMPLEDATA2, TREAT, Y) ; *--- 表5.1 ;

SAMPLEDATA2							
Area	StdErr	WMWodds	SE	LowerCI	UpperCI	LowerCI exp	UpperCI exp
0.7500	0.2152	3	3.44265	-3.74747	9.74747	0.31646	28.4392

%WMWodds (SAMPLEDATA3, TREAT, Y) ; *--- 表5.2 ;

SAMPLEDATA3							
Area	StdErr	WMWodds	SE	LowerCI	UpperCI	LowerCI exp	UpperCI exp
0.7083	0.2282	2.42857	2.68272	-2.82947	7.68661	0.27866	21.1656

- ✓ Area : AUROC の点推定値
- ✓ StdErr : AUROC の標準誤差
- ✓ WMWodds : WMWodds の点推定値
- ✓ SE : WMWodds の標準誤差
- ✓ LowerCI_exp : WMWodds の
両側95%信頼区間の下限
- ✓ UpperCI_exp : WMWodds の
両側95%信頼区間の上限



メニュー

1. Wilcoxon 順位和検定と Hodges-Lehmann 型の推定方法
2. Wilcoxon-Mann-Whitney オッズ (WMWodds)
3. WMWodds と ROC 曲線下面積 (AUROC) との関係
4. 手法の比較①：Wilcoxon 順位和検定の結果との対応
5. 手法の比較②：両側 95% 信頼区間の被覆確率
6. まとめ



手法の比較①：Wilcoxon 順位和検定の結果との対応

- ▶ 各群の応答変数 Y_{1i} 及び Y_{2i} に対して同じ確率分布を仮定し、 $\delta = 1$ だけずらしたシミュレーションデータを用いて、Hodges-Lehmann型による δ の両側 95% 信頼区間と WMWodds の両側 95% 信頼区間について、Wilcoxon 順位和検定の結果とどれだけ対応が取れているかを調査する
- ▶ 各群の例数：250 例
- ▶ 確率分布：正規分布、指数分布、ポアソン分布及び負の二項分布
- ▶ 場面：それぞれの確率分布について、Wilcoxon 順位和検定の p 値が
0.05 をわずかに下回る（有意差あり）場合と
0.05 をわずかに上回る（有意差あり）場合の 2 パターンを用意



手法の比較①：Wilcoxon 順位和検定の結果との対応

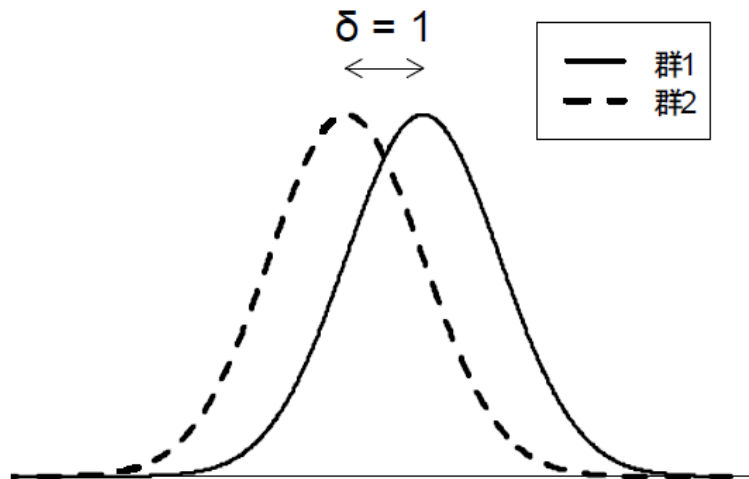


図 6.1 正規分布（分散 $\div 5$ ）

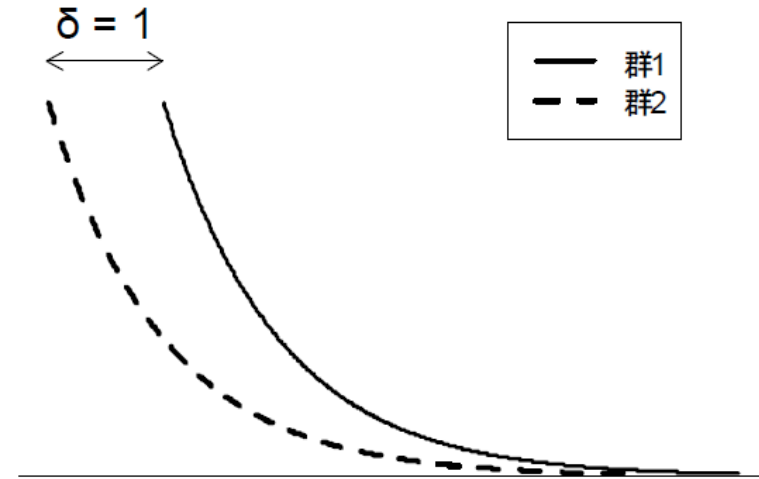


図 6.2 指数分布（平均 $\div 10$ ）

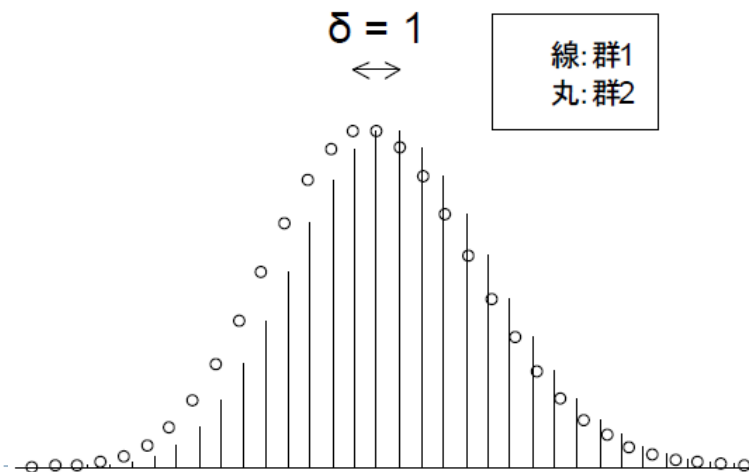


図 6.3 ポアソン分布（平均 $\div 20$ ）

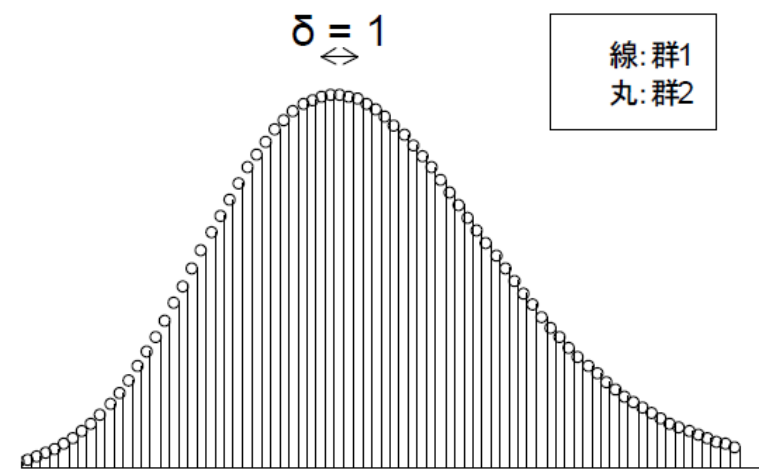


図 6.4 負の二項分布（成功回数 $= 20$ 、 $p \div 0.245$ ）



手法の比較①：Wilcoxon 順位和検定の結果との対応

確率分布 (1群250例)	Wilcoxon 順位和検定	Hodges-Lehmann 型		WMWodds	
		Lower	Upper	Lower	Upper
正規分布 $\hat{\delta} = 0.9$ WMWodds = 1.2	<u>p=0.0480</u>	<u>0.0109</u>	1.8471	<u>1.0011</u>	1.5055
	<u>p=0.0494</u>	<u>0.0028</u>	1.8541	0.9999	1.5037
	p=0.0502	-0.0012	1.8575	0.9991	1.5024
	p=0.0512	-0.0053	1.8610	0.9982	1.5011
指数分布 $\hat{\delta} = 0.9$ WMWodds = 1.2	<u>p=0.0494</u>	<u>0.0010</u>	1.8483	<u>1.0001</u>	1.5034
	<u>p=0.0497</u>	<u>0.0005</u>	1.8487	0.9997	1.5029
	p=0.0501	-0.0009	1.8499	0.9994	1.5025
	p=0.0515	-0.0062	1.8544	0.9981	1.5005
ポアソン分布 $\hat{\delta} = 1.0$ WMWodds = 1.2	<u>p=0.0490</u>	0.0000	2.0000	<u>1.0006</u>	1.5021
	<u>p=0.0495</u>	0.0000	2.0000	<u>1.0001</u>	1.5013
	p=0.0502	0.0000	2.0000	0.9995	1.5005
	p=0.0539	0.0000	2.0000	0.9963	1.4956
負の二項分布 $\hat{\delta} = 3.0$ WMWodds = 1.2	<u>p=0.0481</u>	0.0000	6.0000	<u>1.0014</u>	1.5047
	<u>p=0.0494</u>	0.0000	6.0000	<u>1.0002</u>	1.5029
	p=0.0508	0.0000	6.0000	0.9990	1.5010
	p=0.0515	0.0000	6.0000	0.9984	1.5001

赤字下線部：有意差あり



手法の比較①：Wilcoxon 順位和検定の結果との対応

- ▶ データが連続分布に従っている状況では（タイが生じにくい状況では）、
「2つの分布の形状は同じだが位置がある定数 δ だけずれている」という
仮定が成り立っていれば、Hodges-Lehmann 型による δ の両側 95%信頼区間
は良好な結果となることが分かり、WMWodds の両側 95% 信頼区間では、
p 値が 0.05 をわずかに下回る状況では Hodges-Lehmann 型による推定よりも
わずかに劣ることが示唆された。

（ただし、この問題は実用上はほとんど気にならないと考えられる）

- ▶ データが離散分布に従っている状況では、タイが生じやすい状況である
ため、 $d_{ij} = Y_{1i} - Y_{2j}$ に基づいて δ の推定を行う Hodges-Lehmann 型の方
法では、 d_{ij} として同じ値ばかりが生成されるため 望ましい結果が得られない
ことが伺える。

⇒ さらに考察するため、先ほど用いた正規分布及び指数分布に従うデータの
小数点以下を切り捨て（整数化し）、このデータに対して同様の
シミュレーションを行った。



手法の比較①：Wilcoxon 順位和検定の結果との対応

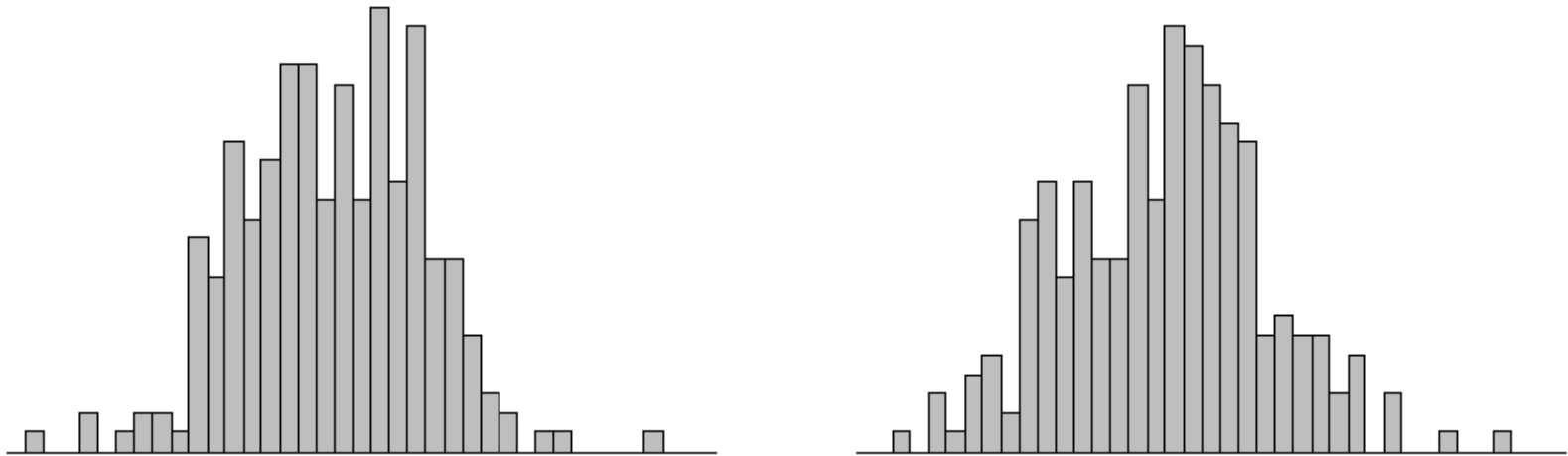


図 6.5 正規分布（図 6.1）を整数化したデータの分布の一例（左：群 1、右：群 2）

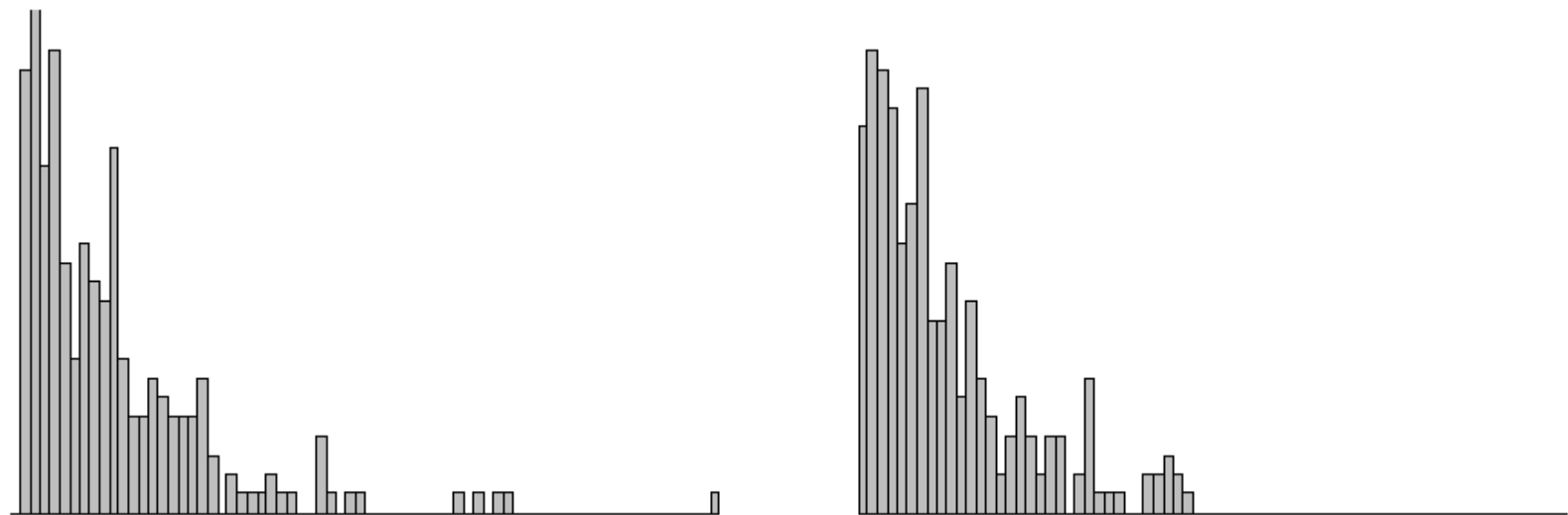


図 6.6 指数分布（図 6.2）を整数化したデータの分布の一例（左：群 1、右：群 2）



手法の比較①：Wilcoxon 順位和検定の結果との対応

確率分布 (1群250例)	Wilcoxon 順位和検定	Hodges-Lehmann 型		WMWodds	
		Lower	Upper	Lower	Upper
正規分布 $\hat{\delta} = 1.0$ $WMWodds = 1.2$	<u>p=0.0441</u>	0.0000	2.0000	<u>1.0049</u>	1.5110
	p=0.0558	0.0000	2.0000	0.9945	1.4941
	p=0.0535	0.0000	2.0000	0.9964	1.4971
	p=0.0501	0.0000	2.0000	0.9992	1.5015
指数分布 $\hat{\delta} = 1.0$ $WMWodds = 1.2$	p=0.0523	0.0000	2.0000	0.9975	1.4982
	p=0.0509	0.0000	2.0000	0.9987	1.5000
	<u>p=0.0492</u>	0.0000	2.0000	<u>1.0002</u>	1.5023
	p=0.0564	0.0000	2.0000	0.9941	1.4930

赤字下線部：有意差あり

- ▶ Hodges-Lehmann 型による δ の両側 95% 信頼区間の下限は全て 0。
- ▶ WMWodds の両側 95% 信頼区間は Wilcoxon 順位和検定の結果と対応が良く取れていることから、データの分布が連続分布でも離散分布でも、Wilcoxon 順位和検定の結果と対応が良く取れていることが分かる。



メニュー

1. Wilcoxon 順位和検定と Hodges-Lehmann 型の推定方法
2. Wilcoxon-Mann-Whitney オッズ (WMWodds)
3. WMWodds と ROC 曲線下面積 (AUROC) との関係
4. 手法の比較①：Wilcoxon 順位和検定の結果との対応
5. 手法の比較②：両側 95% 信頼区間の被覆確率
6. まとめ



手法の比較②：両側 95% 信頼区間の被覆確率

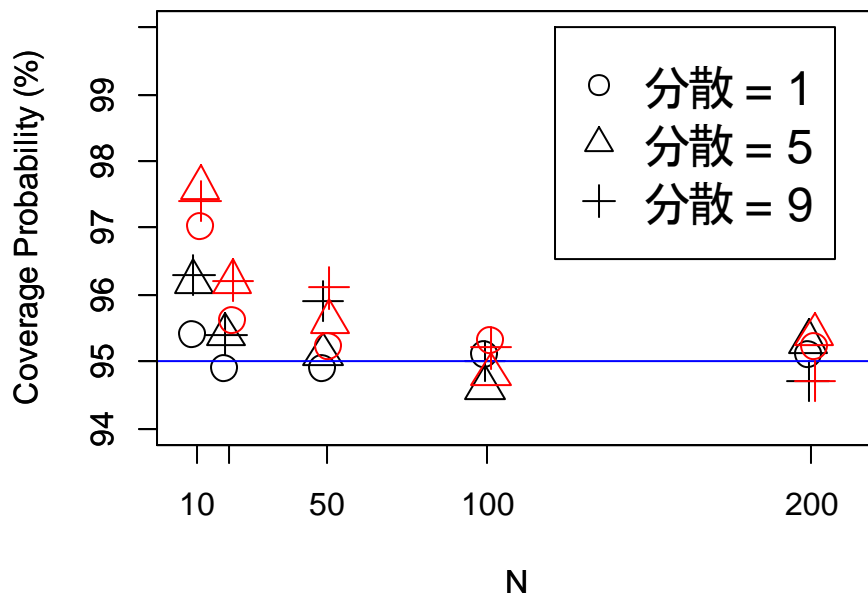
- ▶ 前項と同様の状況ではあるが、本項では各群の応答変数 Y_{1i} 及び Y_{2i} に対して全く同じ確率分布を仮定して ($\delta = 0$ として) シミュレーションデータを生成し、Hodges-Lehmann 型による δ の両側 95% 信頼区間と WMWodds の両側 95% 信頼区間について被覆確率に関する調査を行う
 - ▶ Hodges-Lehmann 型：両側 95% 信頼区間が **0** を含んでいる確率
 - ▶ WMWodds：両側 95% 信頼区間が **1** を含んでいる確率
- ▶ 各群の例数：10、20、50、100 及び 200 例
- ▶ 確率分布：正規分布、指数分布、ポアソン分布及び負の二項分布
- ▶ シミュレーション回数：5000 回

正規分布及び指数分布に関する被覆確率

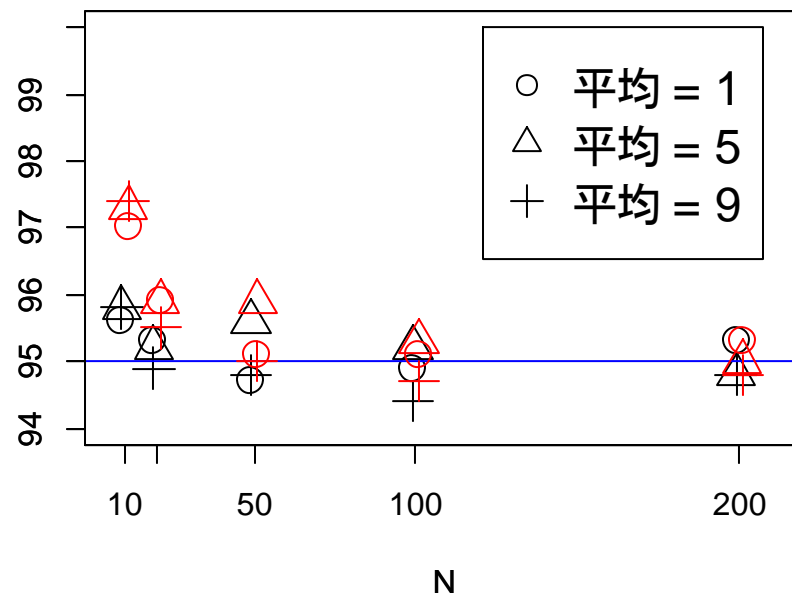
〔黒：Hodges-Lehmann 型、赤：WMWodds〕



正規分布



指数分布

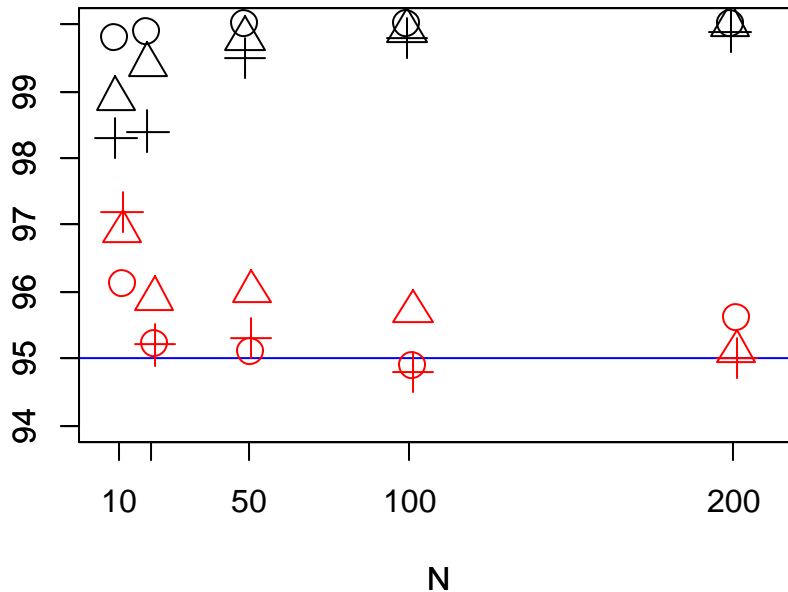


- ▶ 両手法とも 95% を下回る場合が散見されたが、WMWodds の両側 95% 信頼区間の方が頻度は小さかった。
- ▶ また、WMWodds の被覆確率の範囲は 94.7%~97.6% であり、第 1 種の過誤確率は概ね 5% 以内に抑えられており、かつ過度に保守的になっていないことが示唆された。

ポアソン分布及び負の二項分布に関する被覆確率 〔黒：Hodges-Lehmann 型、赤：WMWodds〕

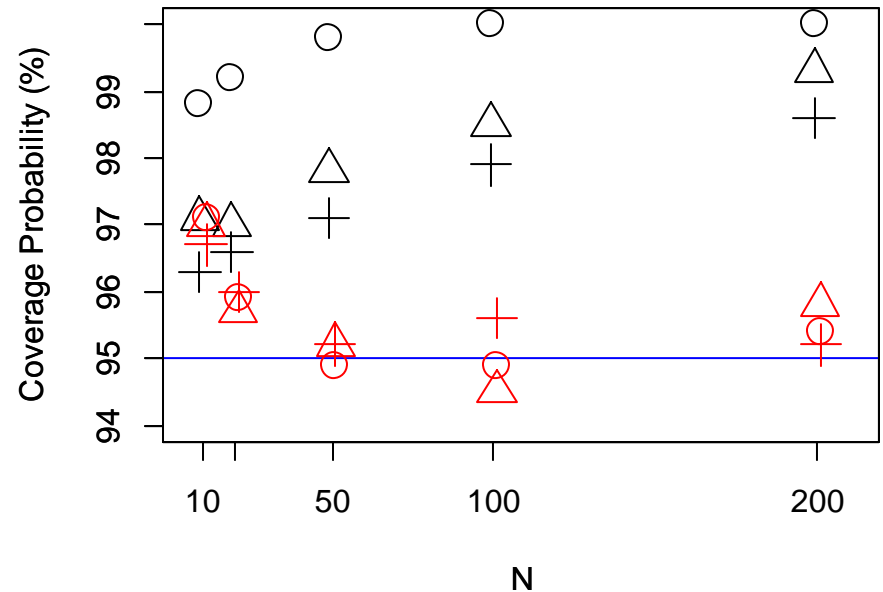


ポアソン分布



○ : 平均 = 1
△ : 平均 = 5
+ : 平均 = 9

負の二項分布 ($p=0.25$)



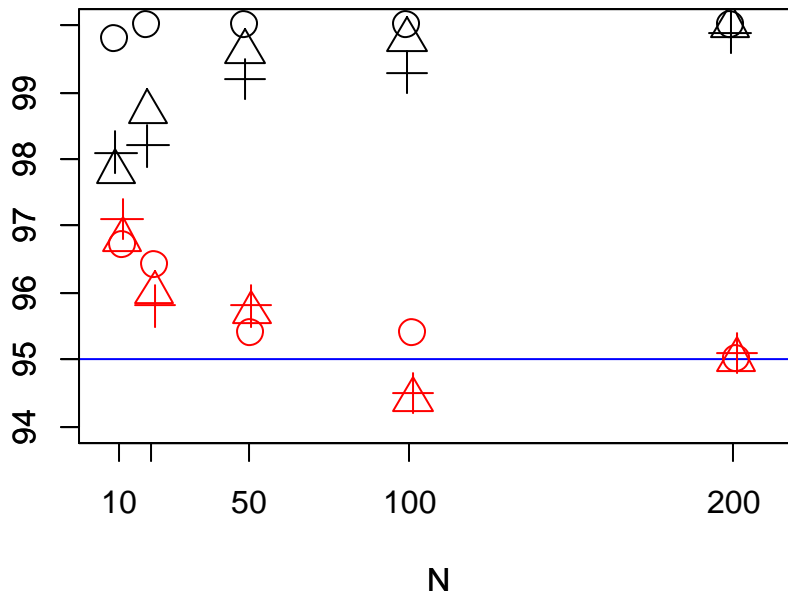
○ : 成功回数 = 1
△ : 成功回数 = 5
+ : 成功回数 = 9

負の二項分布に関する被覆確率

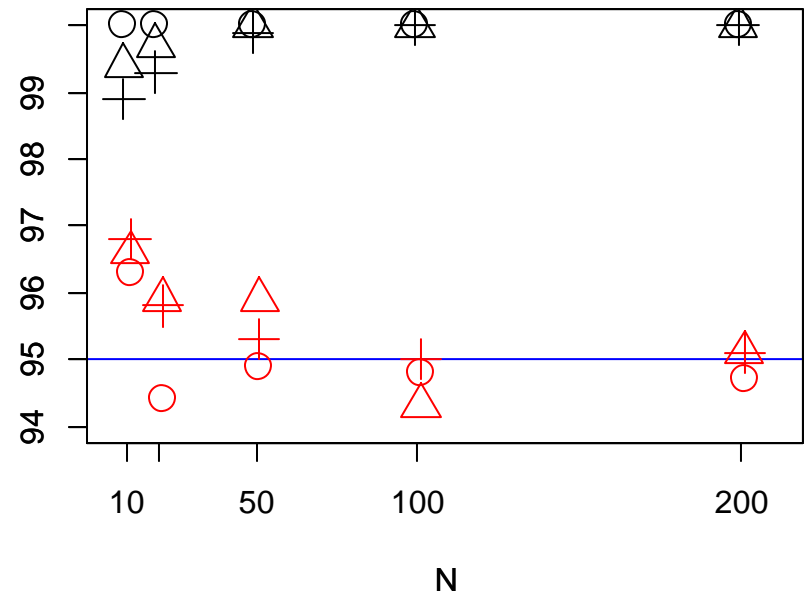
〔黒：Hodges-Lehmann 型、赤：WMWodds〕



負の二項分布($p=0.50$)



負の二項分布($p=0.75$)



- ▶ Hodges-Lehmann 型の被覆確率は過度に大きくなった。これは前項の考察より、離散分布に対する Hodges-Lehmann 型による δ の両側 95% 信頼区間の下限は 0 になりやすいことが原因と思われる。
- ▶ WMWodds の被覆確率は 95% を下回る場合が散見されたが、被覆確率の範囲は 94.3%~97.2% であり、第 1 種の過誤確率は概ね 5% 以内に抑えられており、かつ過度に保守的になっていないことが示唆された。



メニュー

1. Wilcoxon 順位和検定と Hodges-Lehmann 型の推定方法
2. Wilcoxon-Mann-Whitney オッズ (WMWodds)
3. WMWodds と ROC 曲線下面積 (AUROC) との関係
4. 手法の比較①：Wilcoxon 順位和検定の結果との対応
5. 手法の比較②：両側 95% 信頼区間の被覆確率
6. まとめ



まとめ

「Wilcoxon 順位和検定結果との対応」と「両側95%信頼区間の被覆確率」の観点から、連続分布と離散分布の場合において Hodges-Lehmann 型による δ の両側 95% 信頼区間と WMWodds とその両側95%信頼区間の比較を行った。

Hodges-Lehmann 型による両側 95% 信頼区間

データが連続分布に従っている場合は望ましいが、データが離散分布に従っている場合やタイが生じやすい状況では性能が悪くなることが示唆された。

WMWodds の両側95%信頼区間の推定結果

データの分布が連続分布であっても離散分布であっても望ましい結果となることが分かった。また、WMWodds の両側 95% 信頼区間の被覆確率は概ね 95% を上回っており、第 1 種の過誤確率の観点からも WMWodds の両側 95% 信頼区間は望ましい性質を持つことが分かった。



参考文献

- ▶ Acion L, et. al. (2006) 「Probabilistic index: an intuitive non-parametric approach to measuring the size of treatment effects (Statistics in Medicine, Volume 25: 591-602) 」
- ▶ DeLong ER, et. al. (1988) 「Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach (Biometrics, Volume 44(3): 837-45) 」
- ▶ Divine G, et al. (2013) 「A review of analysis and sample size calculation considerations for Wilcoxon tests (Anesth Analg, Volume 117(3): 699-710) 」
- ▶ Hanley JA, McNeil BJ (1982) 「The Meaning and Use of the Area under a Receiver Operating (ROC) Curve Characteristic (Radiology 143(1):29-36) 」
- ▶ Lehmann EL, et. al. (1975) 「Nonparametrics: statistical methods based on ranks (Springer) 」
- ▶ Mason SJ, Graham NE (2002) 「Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical significance and interpretation (Quarterly Journal of the Royal Meteorological Society (128): 2145-2166) 」
- ▶ Newcombe RG (2006a) 「Confidence intervals for an effect size measure based on the Mann-Whitney statistic. Part 1: General issues and tail-area-based methods (Statistics in Medicine, Volume 25: 543-557) 」
- ▶ Newcombe RG (2006b) 「Confidence intervals for an effect size measure based on the Mann-Whitney statistic. Part 2: Asymptotic methods and evaluation (Statistics in Medicine, Volume 25: 559-573) 」



参考文献

- ▶ O'Brien RG, et. al. (2006) 「Exploiting the link between the Wilcoxon-Mann-Whitney test and a simple odds statistic (Proceedings of the Thirty-First Annual SAS Users Group International Conference)」
- ▶ Lehmann EL著、鍋谷 清治 他訳 (1978) 「ノンパラメトリックス (森北出版)」
- ▶ 岩崎 学 (2006) 「統計的データ解析入門 ノンパラメトリック法 (東京図書)」
- ▶ SAS/STAT(R) 9.2 User's Guide
<http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm>

ご清聴いただきましてありがとうございます。



武田薬品工業株式会社

Backup Slides



Mann-Whitney の U 統計量と AUROC との関係

タイがない場合

- ▶ $AUROC = 1 - \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} f_i$ ($f_i : y_{1i}$ より大きい群 2 のデータ数)
- ▶ Mann-Whitney の $U = \sum_{i=1}^{n_1} v_i$ ($v_i : y_{1i}$ より小さい群 2 のデータ数)
- ▶ $n_1 n_2 = \sum_{i=1}^{n_1} f_i + \sum_{i=1}^{n_1} v_i$ となるので、よって $AUROC = \frac{U}{n_1 n_2}$

タイがある場合

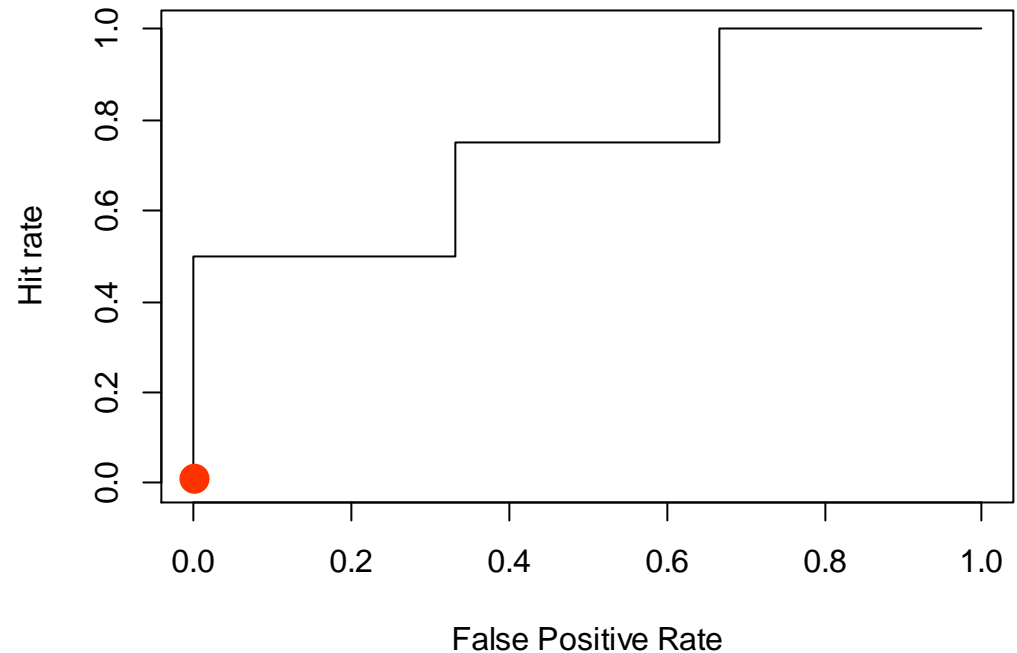
- ▶ $AUROC = 1 - \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} f_i - \frac{1}{2 n_1 n_2} \sum_{i=1}^{n_1} (\tilde{f}_i - f_i)$ ($\tilde{f}_i : y_{1i}$ 以上の群 2 のデータ数)
- ▶ Mann-Whitney の $U = \sum_{i=1}^{n_1} v_i + \frac{1}{2} \sum_{i=1}^{n_1} (\tilde{f}_i - f_i)$
- ▶ $n_1 n_2 = \sum_{i=1}^{n_1} f_i + \sum_{i=1}^{n_1} v_i + \sum_{i=1}^{n_1} (\tilde{f}_i - f_i)$ となるので、よって $AUROC = \frac{U}{n_1 n_2}$



Mann-Whitney の U 統計量と AUROC との関係

表5.1 タイのないデータ

投与群	Y
1	17
1	16
2	15
1	14
2	13
1	12
2	11



▶ $AUROC = 1 - \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} f_i$

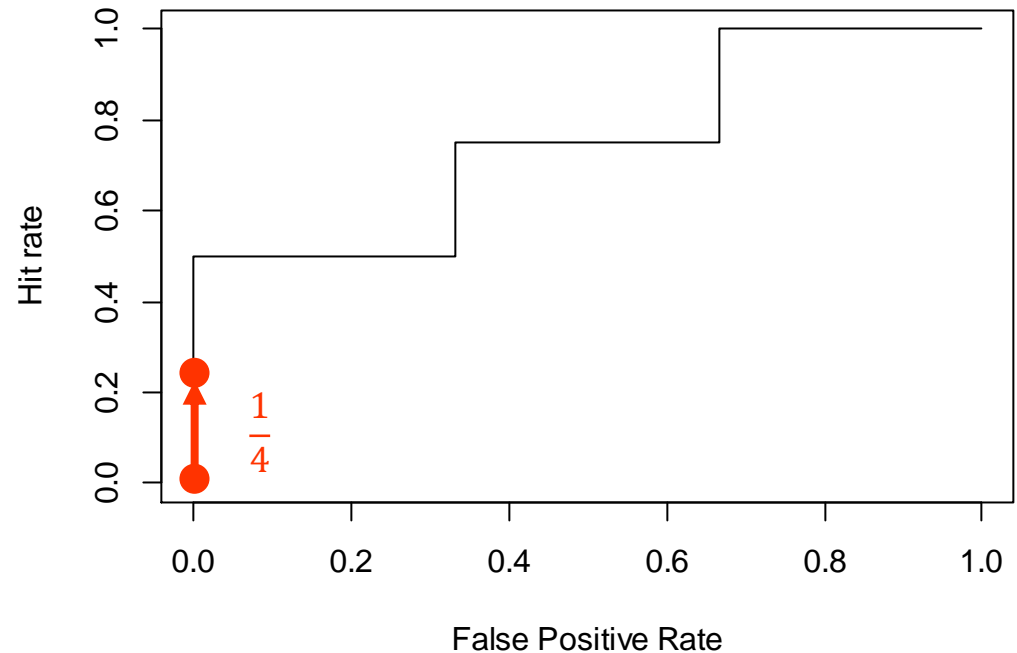
▶ $U_{5.1} = 1 + 2 + 3 + 3 = 9, \pi_{5.1} = \frac{U_{5.1}}{3 \times 4} = 0.75$



Mann-Whitney の U 統計量と AUROC との関係

表5.1 タイのないデータ

投与群	Y
1	17
1	16
2	15
1	14
2	13
1	12
2	11



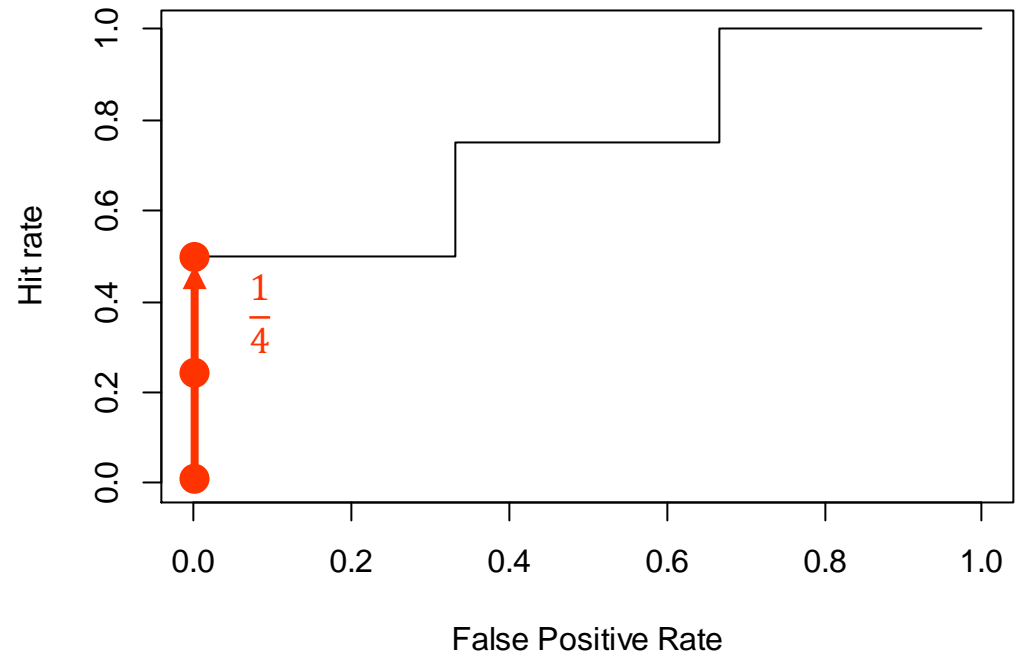
$$\begin{aligned} \text{AUROC} &= 1 - \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} f_i \\ &= 1 - \frac{1}{3 \times 4} (0 \end{aligned}$$



Mann-Whitney の U 統計量と AUROC との関係

表5.1 タイのないデータ

投与群	Y
1	17
1	16
2	15
1	14
2	13
1	12
2	11



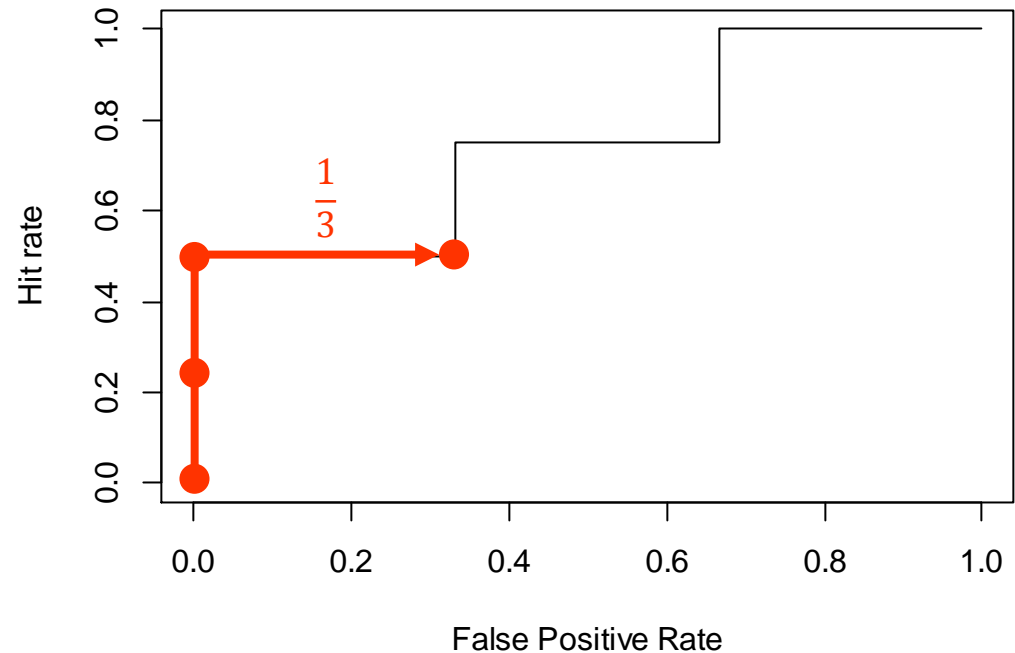
$$\begin{aligned} \text{AUROC} &= 1 - \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} f_i \\ &= 1 - \frac{1}{3 \times 4} (0 + 0) \end{aligned}$$



Mann-Whitney の U 統計量と AUROC との関係

表5.1 タイのないデータ

投与群	Y
1	17
1	16
2	15
1	14
2	13
1	12
2	11



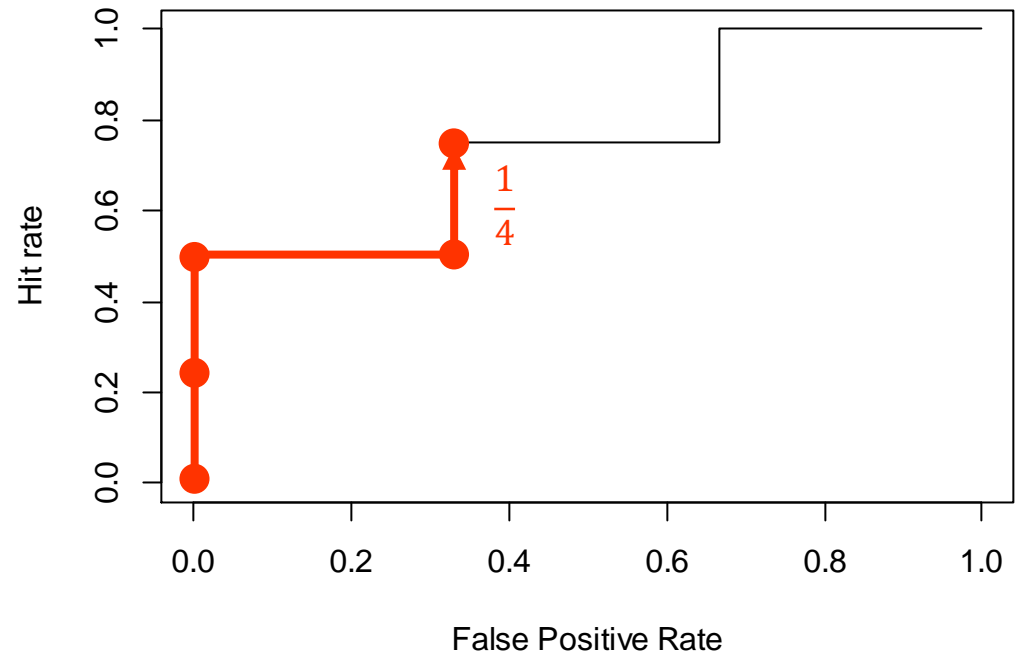
$$\begin{aligned} \text{AUROC} &= 1 - \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} f_i \\ &= 1 - \frac{1}{3 \times 4} (0 + 0 + 0 + 0) \end{aligned}$$



Mann-Whitney の U 統計量と AUROC との関係

表5.1 タイのないデータ

投与群	Y
1	17
1	16
2	15
1	14
2	13
1	12
2	11



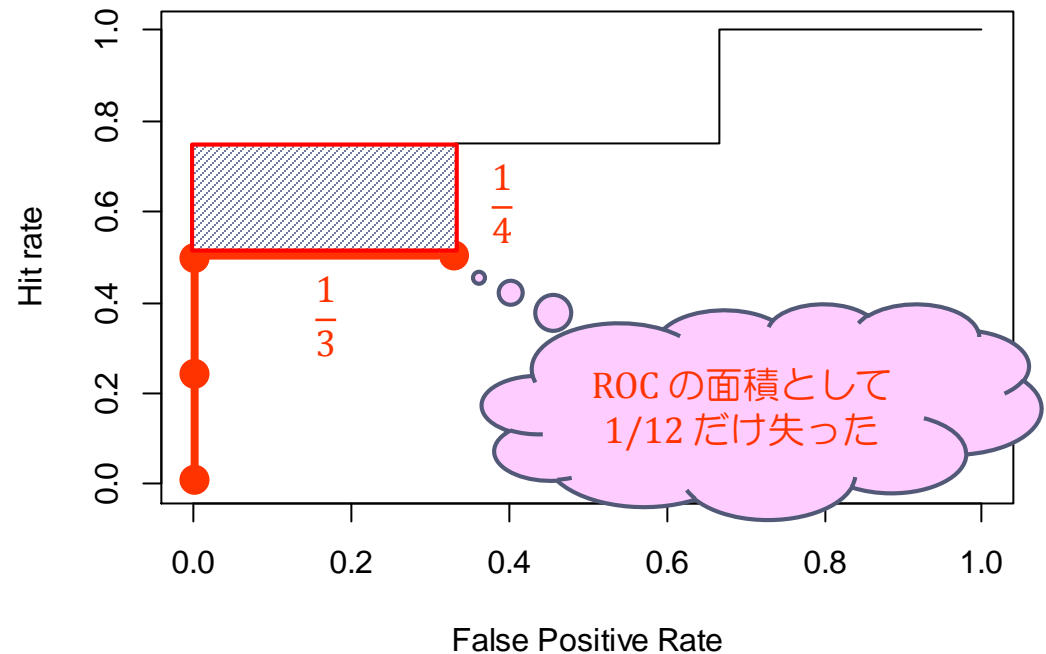
$$\begin{aligned} \text{AUROC} &= 1 - \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} f_i \\ &= 1 - \frac{1}{3 \times 4} (0 + 0 + 1) \end{aligned}$$



Mann-Whitney の U 統計量と AUROC との関係

表5.1 タイのないデータ

投与群	Y
1	17
1	16
2	15
1	14
2	13
1	12
2	11



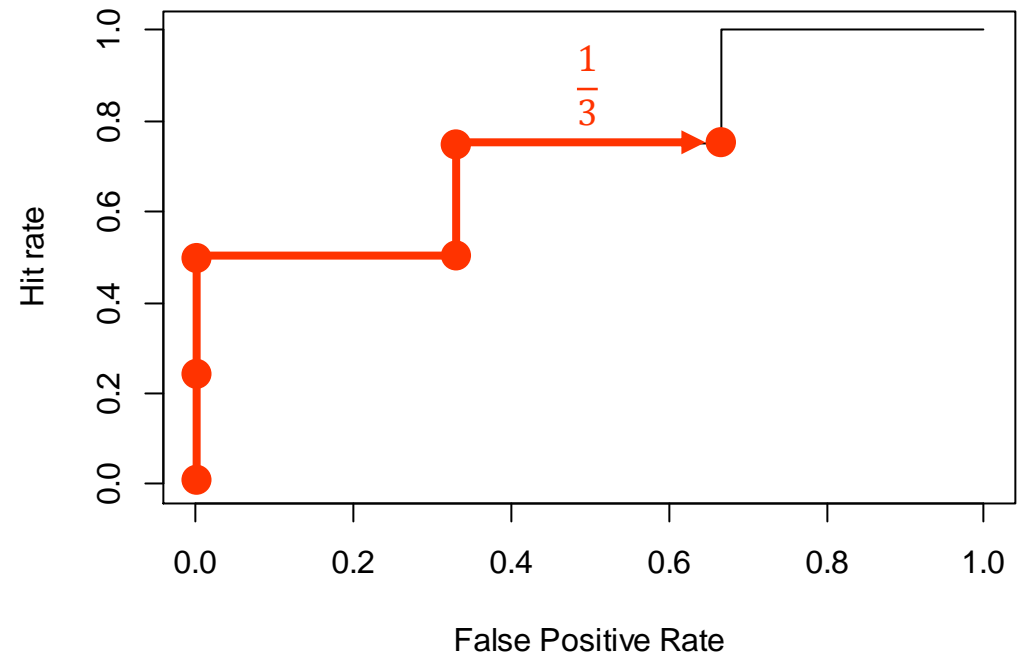
$$\begin{aligned} \text{AUROC} &= 1 - \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} f_i \\ &= 1 - \frac{1}{3 \times 4} (0 + 0 + 1) \end{aligned}$$



Mann-Whitney の U 統計量と AUROC との関係

表5.1 タイのないデータ

投与群	Y
1	17
1	16
2	15
1	14
2	13
1	12
2	11



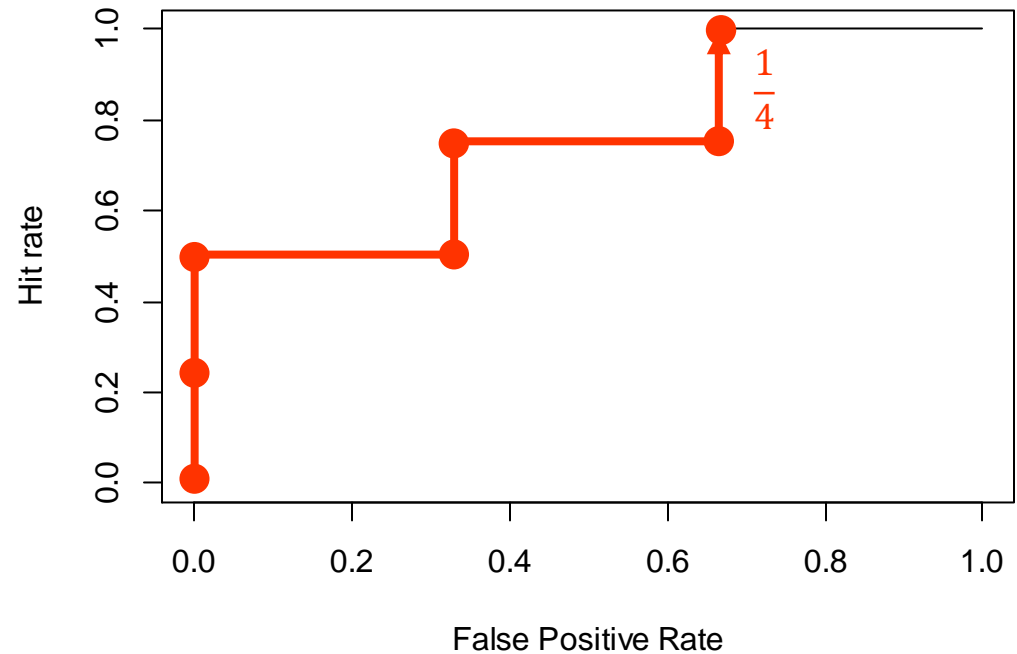
$$\begin{aligned} \text{AUROC} &= 1 - \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} f_i \\ &= 1 - \frac{1}{3 \times 4} (0 + 0 + 1) \end{aligned}$$



Mann-Whitney の U 統計量と AUROC との関係

表5.1 タイのないデータ

投与群	Y
1	17
1	16
2	15
1	14
2	13
1	12
2	11



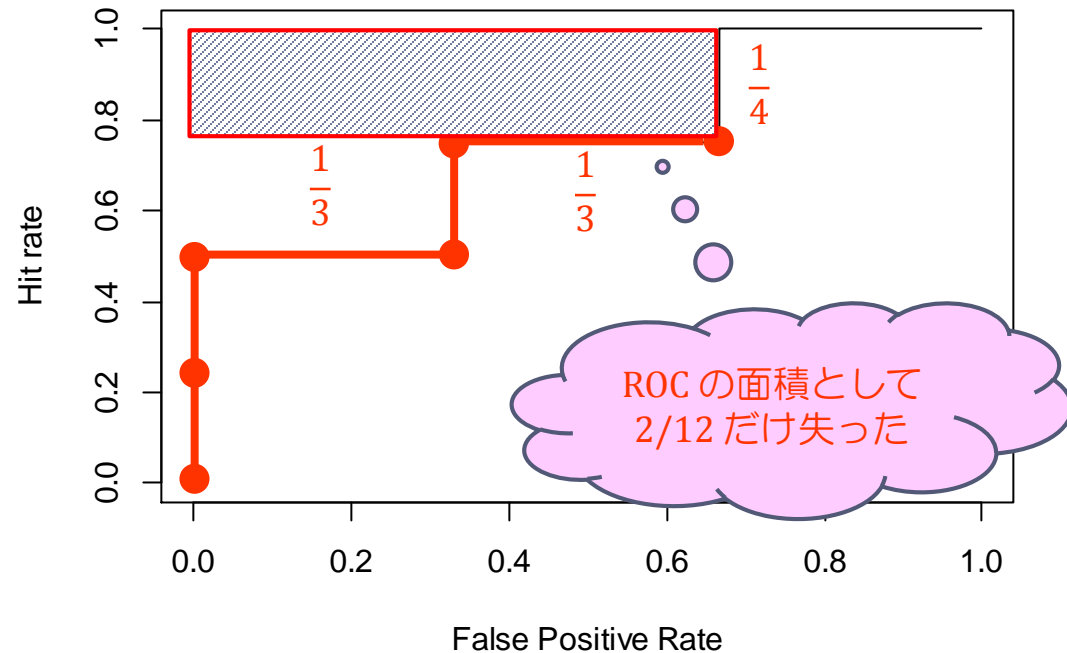
$$\begin{aligned} \text{AUROC} &= 1 - \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} f_i \\ &= 1 - \frac{1}{3 \times 4} (0 + 0 + 1 + 2) \end{aligned}$$



Mann-Whitney の U 統計量と AUROC との関係

表5.1 タイのないデータ

投与群	Y
1	17
1	16
2	15
1	14
2	13
1	12
2	11



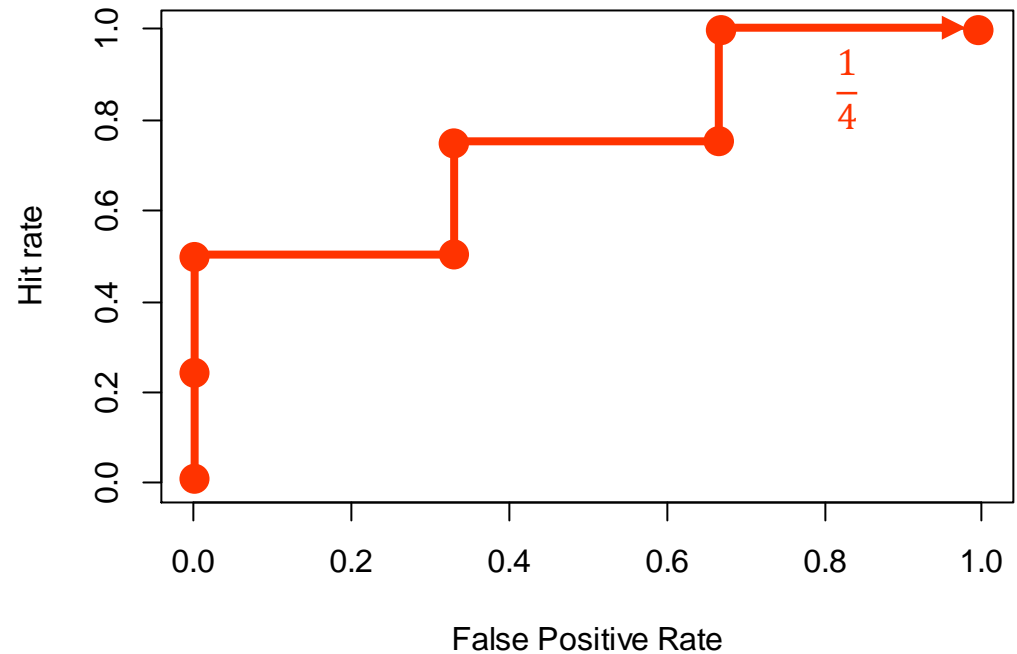
$$\begin{aligned} \text{AUROC} &= 1 - \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} f_i \\ &= 1 - \frac{1}{3 \times 4} (0 + 0 + 1 + 2) \end{aligned}$$



Mann-Whitney の U 統計量と AUROC との関係

表5.1 タイのないデータ

投与群	Y
1	17
1	16
2	15
1	14
2	13
1	12
2	11



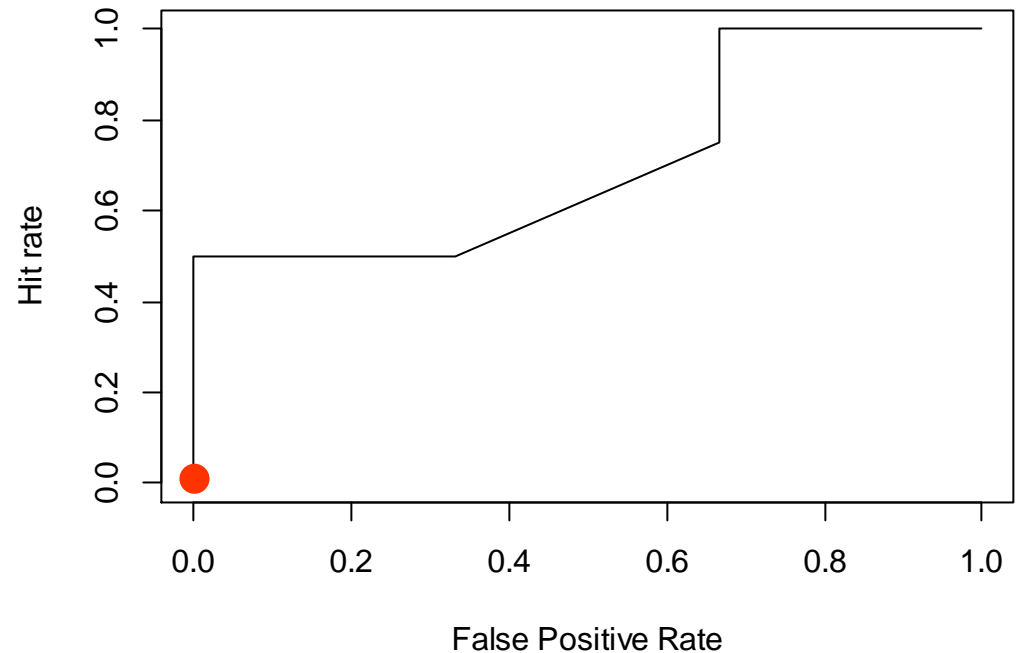
$$\begin{aligned} \text{AUROC} &= 1 - \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} f_i \\ &= 1 - \frac{1}{3 \times 4} (0 + 0 + 1 + 2) = 1 - \frac{3}{12} = 0.75 \end{aligned}$$



Mann-Whitney の U 統計量と AUROC との関係

表5.2 タイのあるデータ

投与群	Y
1	17
1	16
2	15
1	<u>13</u>
2	<u>13</u>
1	12
2	11



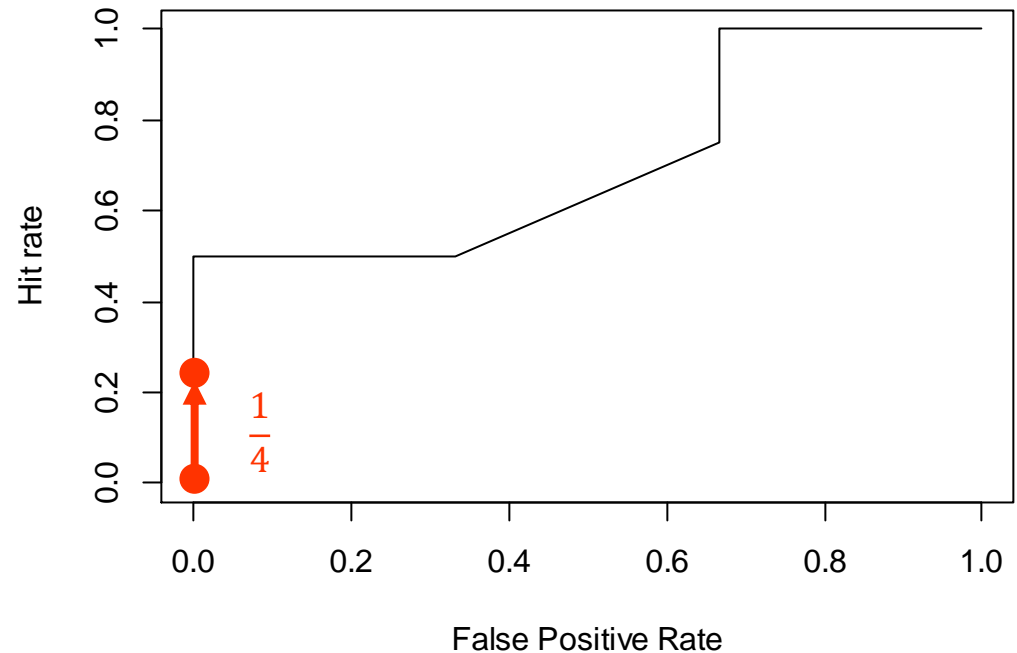
- ▶
$$\text{AUROC} = 1 - \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} f_i - \frac{1}{2 n_1 n_2} \sum_{i=1}^{n_1} (\tilde{f}_i - f_i)$$
- ▶
$$U_{5.2} = 1 + 1.5 + 3 + 3 = 8.5, \pi_{5.2} = \frac{U_{5.2}}{3 \times 4} = 0.708$$



Mann-Whitney の U 統計量と AUROC との関係

表5.2 タイのあるデータ

投与群	Y
1	17
1	16
2	15
1	13
2	13
1	12
2	11



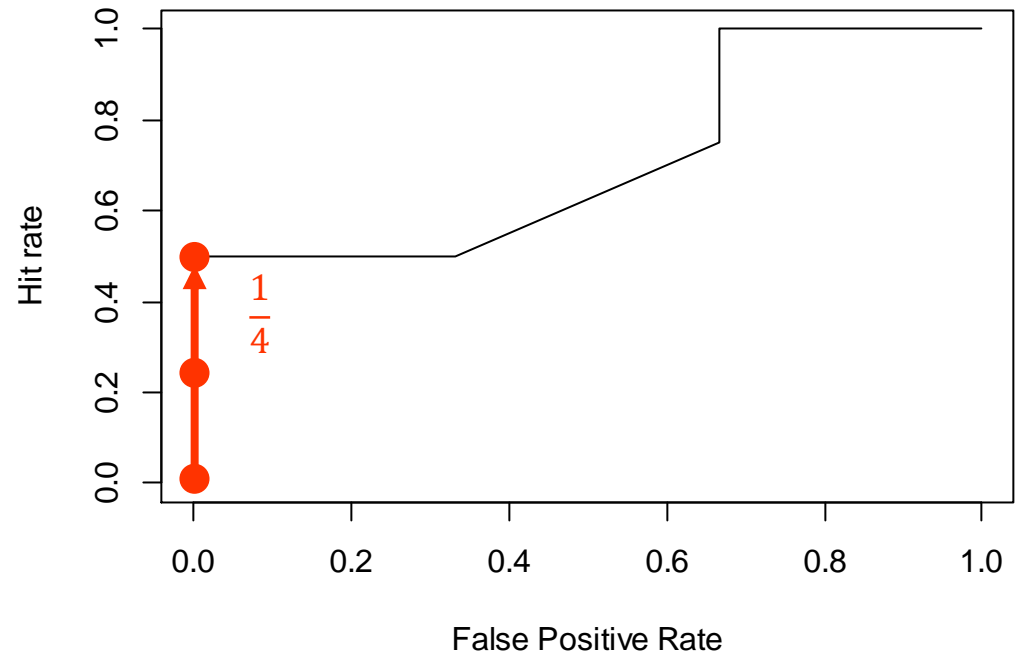
$$\begin{aligned} \text{AUROC} &= 1 - \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} f_i - \frac{1}{2 n_1 n_2} \sum_{i=1}^{n_1} (\tilde{f}_i - f_i) \\ &= 1 - \frac{1}{3 \times 4} (0 \quad \quad \quad) - \frac{1}{2 \times 3 \times 4} (0 \quad \quad \quad) \end{aligned}$$



Mann-Whitney の U 統計量と AUROC との関係

表5.2 タイのあるデータ

投与群	Y
1	17
1	16
2	15
1	13
2	13
1	12
2	11



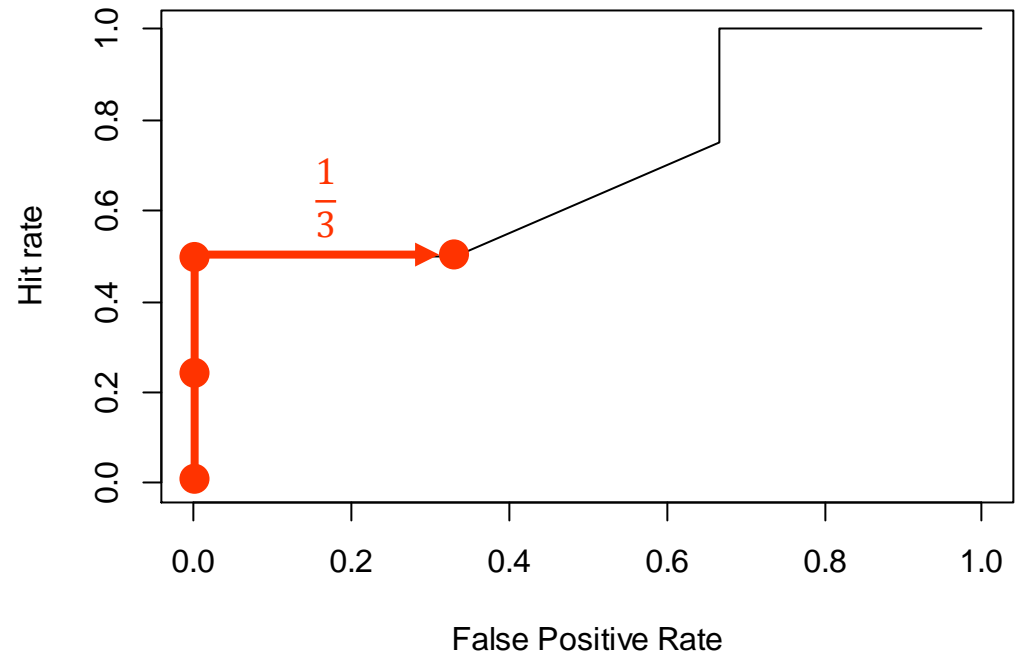
$$\begin{aligned} \text{AUROC} &= 1 - \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} f_i - \frac{1}{2 n_1 n_2} \sum_{i=1}^{n_1} (\tilde{f}_i - f_i) \\ &= 1 - \frac{1}{3 \times 4} (0 + 0) - \frac{1}{2 \times 3 \times 4} (0 + 0) \end{aligned}$$



Mann-Whitney の U 統計量と AUROC との関係

表5.2 タイのあるデータ

投与群	Y
1	17
1	16
2	15
1	13
2	13
1	12
2	11



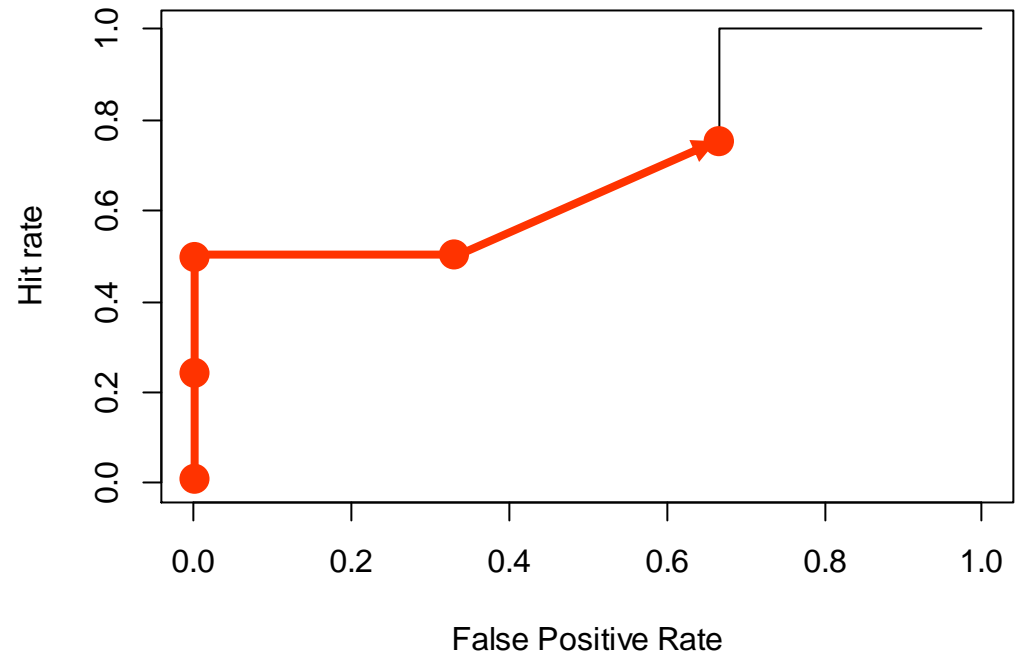
$$\begin{aligned} \text{AUROC} &= 1 - \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} f_i - \frac{1}{2 n_1 n_2} \sum_{i=1}^{n_1} (\tilde{f}_i - f_i) \\ &= 1 - \frac{1}{3 \times 4} (0 + 0) - \frac{1}{2 \times 3 \times 4} (0 + 0) \end{aligned}$$



Mann-Whitney の U 統計量と AUROC との関係

表5.2 タイのあるデータ

投与群	Y
1	17
1	16
2	15
1	13
2	13
1	12
2	11



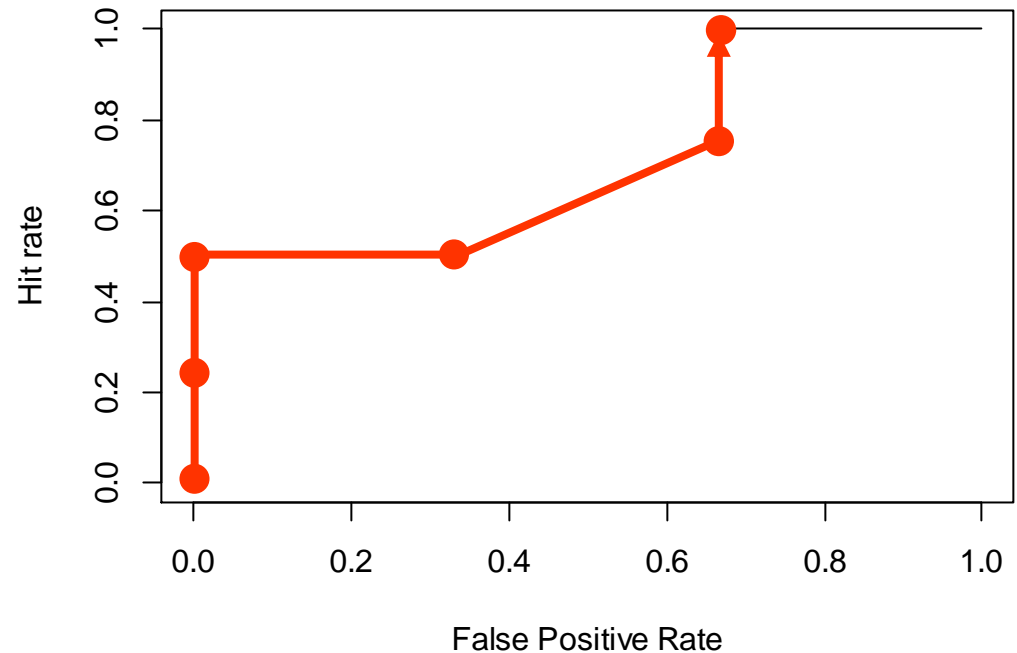
$$\begin{aligned} \text{AUROC} &= 1 - \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} f_i - \frac{1}{2 n_1 n_2} \sum_{i=1}^{n_1} (\tilde{f}_i - f_i) \\ &= 1 - \frac{1}{3 \times 4} (0 + 0 + 1) - \frac{1}{2 \times 3 \times 4} (0 + 0 + 1) \end{aligned}$$



Mann-Whitney の U 統計量と AUROC との関係

表5.2 タイのあるデータ

投与群	Y
1	17
1	16
2	15
1	13
2	13
1	12
2	11



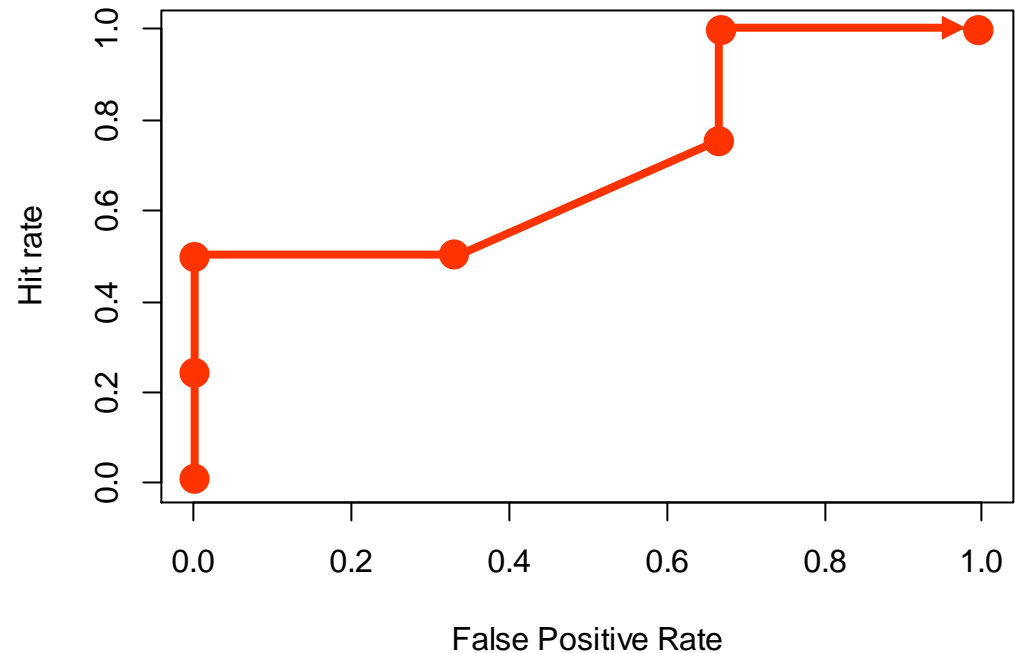
$$\begin{aligned} \text{AUROC} &= 1 - \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} f_i - \frac{1}{2 n_1 n_2} \sum_{i=1}^{n_1} (\tilde{f}_i - f_i) \\ &= 1 - \frac{1}{3 \times 4} (0 + 0 + 1 + 2) - \frac{1}{2 \times 3 \times 4} (0 + 0 + 1 + 0) \end{aligned}$$



Mann-Whitney の U 統計量と AUROC との関係

表5.2 タイのあるデータ

投与群	Y
1	17
1	16
2	15
1	13
2	13
1	12
2	11



$$\begin{aligned} \text{AUROC} &= 1 - \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} f_i - \frac{1}{2 n_1 n_2} \sum_{i=1}^{n_1} (\tilde{f}_i - f_i) \\ &= 1 - \frac{1}{3 \times 4} (0 + 0 + 1 + 2) - \frac{1}{2 \times 3 \times 4} (0 + 0 + 1 + 0) = \frac{17}{24} = 0.708 \end{aligned}$$



generalized odds ratio (genOR)

- ▶ O'Brien (2006) では、WMWodds の標準誤差は Agresti (1980) にて紹介されている generalized odds ratio (genOR) を基に算出している
- ▶ genOR の定義は以下の通り
 - $\text{genOR} = \frac{\text{Prob}(Y_1 > Y_2)}{\text{Prob}(Y_1 < Y_2)} = \frac{1-\gamma}{1+\gamma}$
 - γ は Goodman-Kruskall gamma 統計量であり、freq プロシジャで推定値及び標準誤差の算出が可能
- ▶ タイデータが存在しなければ、 $\text{genOR} = \text{WMWodds}$ となるため、 γ の標準誤差とデルタ法を用いて genOR の標準誤差が算出できる
 - $V[\text{genOR}] = \frac{2}{(1-\hat{\gamma})^2} V[\hat{\gamma}]$



generalized odds ratio (genOR)

- ▶ タイデータが存在する場合、以下の手順で変換を行う
 - ① 値に対する順位付けを行い、頻度集計を行う。
 - ② タイデータが存在する順位について、片方の群の順位を2つの順位に分割する。このとき、分割された順位は元の順位に ± 0.1 をそれぞれ加えたものとなる。
 - ③ 分割された順位の頻度は元の頻度の $1/2$ にする。
- ▶ 上記の変換を行った上で、genORの標準誤差を算出する
- ▶ 次頁以降では、genORを基に WMWodds の両側 95% 信頼区間を算出した場合の（すなわち、O'Brien（2006）で実施していたと予想される計算方法での）シミュレーション結果を紹介する



generalized odds ratio (genOR)

確率分布	1群あたりの例数	分布のパラメータ	Hodges-Lehmann 型	WMWodds	
				本発表の方法	O'Brien
正規分布	10例	分散=1	95.4%	97.0%	95.6%
		分散=5	96.2%	97.6%	96.7%
		分散=9	96.3%	97.4%	96.5%
	20例	分散=1	94.9%	95.6%	95.1%
		分散=5	95.4%	96.2%	95.5%
		分散=9	95.4%	96.2%	95.6%
	50例	分散=1	94.9%	95.2%	95.0%
		分散=5	95.1%	95.6%	95.3%
		分散=9	95.9%	96.1%	95.8%
	100例	分散=1	95.1%	95.3%	95.2%
		分散=5	94.6%	94.8%	94.7%
		分散=9	95.0%	95.2%	95.1%
指数分布	10例	分散=1	95.1%	95.2%	95.1%
	20例	分散=5	95.3%	95.4%	95.3%
		分散=9	94.7%	94.7%	94.7%
	10例	平均=1	95.6%	97.0%	96.0%
	20例	平均=5	95.8%	97.3%	96.1%
		平均=9	95.8%	97.4%	96.2%
	20例	平均=1	95.3%	95.9%	95.4%
	50例	平均=5	95.2%	95.9%	95.3%
		平均=9	94.9%	95.5%	94.9%
	50例	平均=1	94.7%	95.1%	94.8%
	100例	平均=5	95.6%	95.9%	95.7%
		平均=9	94.8%	95.0%	95.0%
	100例	平均=1	94.9%	95.1%	94.9%
	200例	平均=5	95.2%	95.3%	95.2%
		平均=9	94.4%	94.7%	94.5%

シミュレーション回数は 5000 回、確率は小数第 2 位を四捨五入して表示し、95% 未満のものに下線を引いた。



generalized odds ratio (genOR)

確率分布	1群あたりの例数	分布のパラメータ	Hodges-Lehmann 型	WMWodds	
				本発表の方法	O'Brien
ポアソン分布	10例	平均=1	99.8%	96.1%	97.4%
		平均=5	98.9%	96.9%	96.5%
		平均=9	98.3%	97.2%	96.3%
	20例	平均=1	99.9%	95.2%	96.5%
		平均=5	99.4%	95.9%	95.8%
		平均=9	98.4%	95.2%	95.0%
	50例	平均=1	100.0%	95.1%	96.7%
		平均=5	99.8%	96.0%	96.1%
		平均=9	99.5%	95.3%	95.3%
	100例	平均=1	100.0%	94.9%	96.5%
		平均=5	99.9%	95.7%	95.9%
		平均=9	99.8%	94.8%	94.9%
負の二項分布 (成功回数：k)	10例	平均=1	100.0%	95.6%	97.0%
		平均=5	100.0%	95.1%	95.4%
		平均=9	99.9%	95.0%	95.2%
	20例	p=0.25, k=1	98.8%	97.1%	96.7%
		p=0.25, k=5	97.1%	97.0%	95.9%
		p=0.25, k=9	96.3%	96.7%	95.7%
	20例	p=0.25, k=1	99.2%	95.9%	95.9%
		p=0.25, k=5	97.0%	95.7%	95.1%
		p=0.25, k=9	96.6%	96.0%	95.6%
	50例	p=0.25, k=1	99.8%	94.9%	95.2%
		p=0.25, k=5	97.8%	95.2%	95.0%
		p=0.25, k=9	97.1%	95.2%	95.0%
	100例	p=0.25, k=1	100.0%	94.9%	95.3%
		p=0.25, k=5	98.5%	94.5%	94.5%
		p=0.25, k=9	97.9%	95.6%	95.5%
	200例	p=0.25, k=1	100.0%	95.4%	95.8%
		p=0.25, k=5	99.3%	95.8%	95.7%
		p=0.25, k=9	98.6%	95.2%	95.2%

シミュレーション回数は 5000 回、確率は小数第 2 位を四捨五入して表示し、95% 未満のものに下線を引いた。



generalized odds ratio (genOR)

確率分布	1群あたりの例数	分布のパラメータ	Hodges-Lehmann 型	WMWodds	
				本発表の方法	O'Brien
負の二項分布 (成功回数：k)	10例	p=0.50, k=1	99.8%	96.7%	98.1%
		p=0.50, k=5	97.8%	96.8%	96.1%
		p=0.50, k=9	98.1%	97.1%	96.3%
	20例	p=0.50, k=1	100.0%	96.4%	98.0%
		p=0.50, k=5	98.7%	96.0%	95.8%
		p=0.50, k=9	98.2%	95.8%	95.6%
	50例	p=0.50, k=1	100.0%	95.4%	97.2%
		p=0.50, k=5	99.6%	95.7%	95.6%
		p=0.50, k=9	99.2%	95.8%	95.7%
	100例	p=0.50, k=1	100.0%	95.4%	97.4%
		p=0.50, k=5	99.8%	94.4%	94.6%
		p=0.50, k=9	99.3%	94.5%	94.5%
	200例	p=0.50, k=1	100.0%	95.0%	97.4%
		p=0.50, k=5	100.0%	95.0%	95.1%
		p=0.50, k=9	99.9%	95.1%	95.3%
	10例	p=0.75, k=1	100.0%	96.3%	99.7%
		p=0.75, k=5	99.4%	96.6%	96.8%
		p=0.75, k=9	98.9%	96.8%	96.6%
	20例	p=0.75, k=1	100.0%	94.4%	99.5%
		p=0.75, k=5	99.7%	95.9%	96.4%
		p=0.75, k=9	99.3%	95.8%	95.8%
	50例	p=0.75, k=1	100.0%	94.9%	99.5%
		p=0.75, k=5	100.0%	95.9%	96.5%
		p=0.75, k=9	99.9%	95.3%	95.6%
	100例	p=0.75, k=1	100.0%	94.8%	99.5%
		p=0.75, k=5	100.0%	94.3%	95.4%
		p=0.75, k=9	100.0%	95.0%	95.5%
	200例	p=0.75, k=1	100.0%	94.7%	99.6%
		p=0.75, k=5	100.0%	95.1%	96.0%
		p=0.75, k=9	100.0%	95.1%	95.5%

シミュレーション回数は 5000 回、確率は小数第 2 位を四捨五入して表示し、95% 未満のものに下線を引いた。