

# Rで統計解析入門

## (8) 2値データの要約



## 準備：データ「DEP」の読み込み

1. データ「DEP」を以下からダウンロードする  
<http://www.cwk.zaq.ne.jp/fkhud708/files/dep.csv>
2. ダウンロードした場所を把握する　ここでは「c:/temp」とする
3. R を起動し，2. の場所に移動し，データを読み込む
4. データ「DEP」から薬剤 A と B のデータを抽出

```
> setwd("c:/temp") # dep.csv がある場所に移動
> getwd() # 移動できたかどうか確認
> DEP <- read.csv("dep.csv") # dep.csv を読み込む
> AB <- subset(DEP, GROUP != "C") # 薬剤 A と B のデータを抽出
> AB$GROUP <- factor(AB$GROUP) # 薬剤の水準を 2 カテゴリに
> head(AB)
  GROUP QOL EVENT DAY PREDRUG DURATION
1     A  15     1  50        NO         1
2     :   :     :   :         :         :
```



## 準備：架空のデータ「DEP」の変数

---

- ▶ **GROUP**：薬剤の種類（A, B, C）
- ▶ **QOL**：QOL の点数（数値） 点数が大きい方が良い
- ▶ **EVENT**：改善の有無（1：改善あり，2：改善なし）  
QOLの点数が5点以上である場合を「改善あり」とする
- ▶ **DAY**：観察期間（数値，単位は日）
- ▶ **PREDRUG**：前治療薬の有無（YES：他の治療薬を投与したことあり，  
NO：投与したことなし）
- ▶ **DURATION**：罹病期間（数値，単位は年）



## 準備：架空のデータ「DEP」（一部）

GROUP	QOL	EVENT	DAY	PREDRUG	DURATION
A	15	1	50	NO	1
A	13	1	200	NO	3
A	11	1	250	NO	2
A	11	1	300	NO	4
A	10	1	350	NO	2
A	9	1	400	NO	2
A	8	1	450	NO	4
A	8	1	550	NO	2
A	6	1	600	NO	5
A	6	1	100	NO	7
A	4	2	250	NO	4
A	3	2	500	NO	6
A	3	2	750	NO	3
A	3	2	650	NO	7
A	1	2	1000	NO	8
A	6	1	150	YES	6
A	5	1	700	YES	5
A	4	2	800	YES	7
A	2	2	900	YES	12
A	2	2	950	YES	10
B	13	1	380	NO	9
B	12	1	880	NO	5
B	11	1	940	NO	2
B	4	2	20	NO	7
B	4	2	560	NO	2
B	5	1	320	YES	11
B	5	1	940	YES	3
B	4	2	80	YES	6
B	3	2	140	YES	7
B	3	2	160	YES	13



## 本日のメニュー

---

1. イントロ
2. 割合と比
3. リスク差, リスク比, オッズ比
4.  $\chi^2$  検定



## 「改善あり」の比較

- ▶ 「薬剤 A (GROUP=A) の改善ありの数」と  
「薬剤 B (GROUP=B) の改善ありの数」の比較を行う
- ▶ 薬剤ごとの「改善あり」の数を算出する

```
> # 見たい指標 薬剤の順で指定する
> ( TABLE1 <- xtabs(~ EVENT + GROUP, data=AB) )
      GROUP
EVENT  A  B
  1  12  5
  2   8 15
```

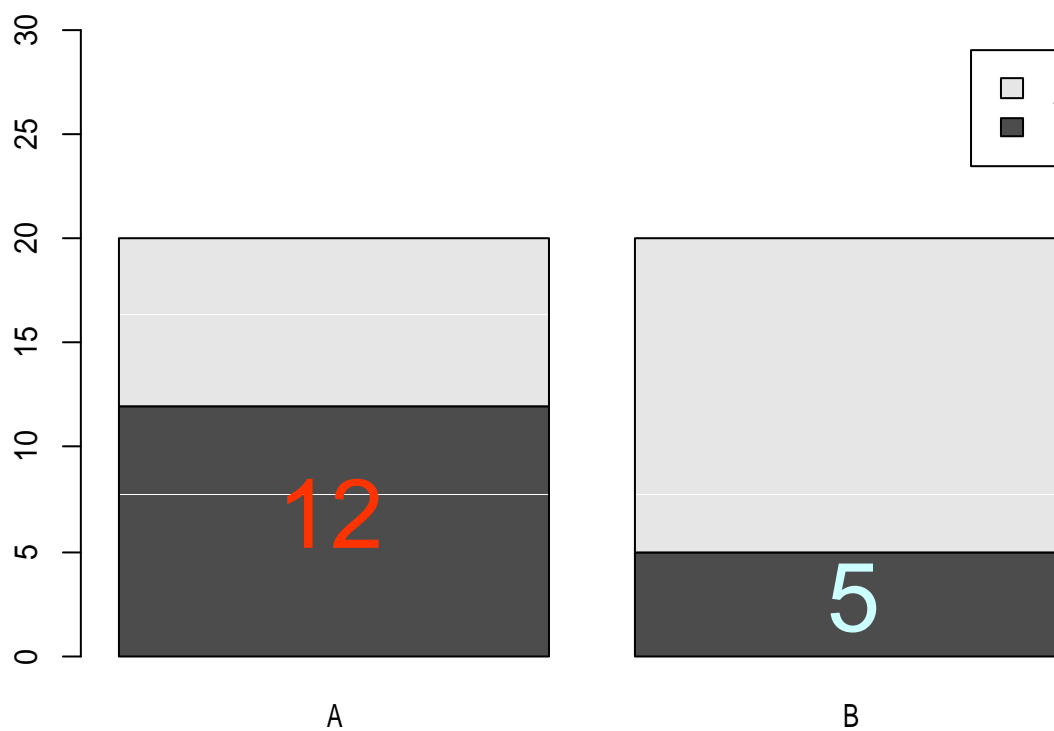
- ▶ 薬剤 A (12 例) の方が薬剤 B (8 例) よりも多い
- ▶ 今回は良いが、もし分母が異なる場合は「数の比較」を行うと良くない



## 「改善あり」の比較

- ▶ 薬剤ごとの「改善あり」の数に関するグラフ〔棒グラフ〕を描く

```
> barplot(TABLE1, legend=rownames(TABLE1), ylim=c(0,30))
```





## 「改善あり」の比較

- ▶ 「薬剤 A (GROUP=A) の改善ありの割合」と  
「薬剤 B (GROUP=B) の改善ありの割合」の比較を行う
- ▶ 薬剤ごとの「改善あり」の割合を算出する

```
> ( TABLE2 <- prop.table(TABLE1, 2) )
```

```
      GROUP  
EVENT   A   B  
  1 0.60 0.25  
  2 0.40 0.75
```

- ▶ 薬剤 A ( 60% ) の方が薬剤 B ( 25% ) よりも多い
- ▶ 60% や 25% って何? **これが今回のトピック**

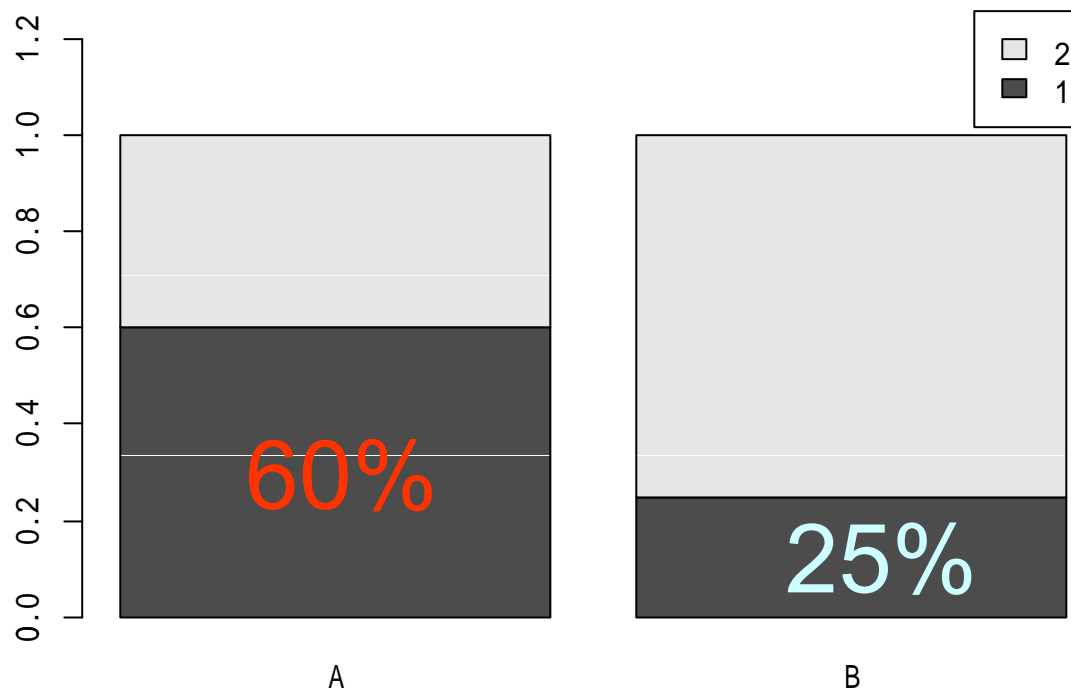




## 「改善あり」の比較

- ▶ 薬剤ごとの「改善あり」の割合に関するグラフ〔棒グラフ〕を描く

```
> barplot(TABLE2, legend=rownames(TABLE2), ylim=c(0,1.3))
```



1:改善あり  
2:改善なし



## 「改善あり」の比較【前治療の有無別】

- ▶ 前治療の有無別に、薬剤ごとの「改善あり」の数を求める

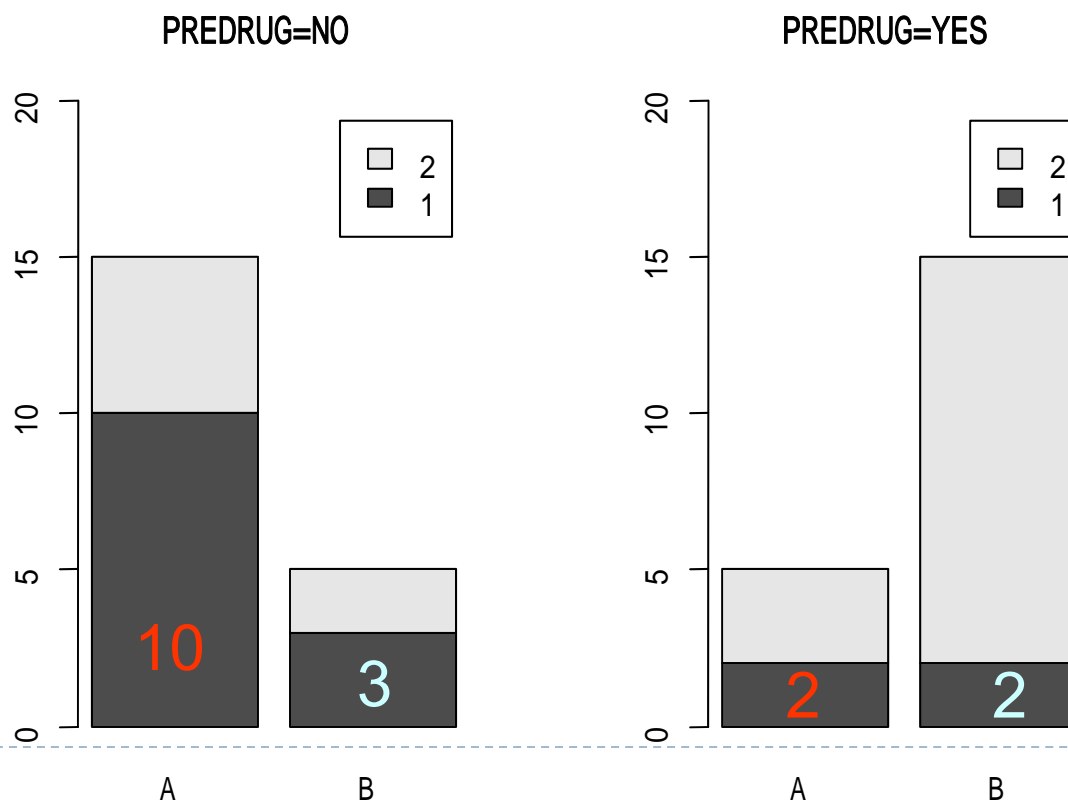
```
> ( TABLE3 <- xtabs(~ EVENT + GROUP + PREDRUG, data=AB)
, , PREDRUG = NO
  GROUP
EVENT  A  B
  1 10  3
  2  5  2
, , PREDRUG = YES
  GROUP
EVENT  A  B
  1  2  2
  2  3 13
```



## 「改善あり」の比較【前治療の有無別】

- ▶ 前治療の有無別・薬剤別の「改善あり」の数に関する棒グラフを描く

```
> par(mfrow=c(1,2))  
> barplot(TABLE3[,1], legend=rownames(TABLE1), main="PREDRUG=NO", ylim=c(0,20))  
> barplot(TABLE3[,2], legend=rownames(TABLE1), main="PREDRUG=YES", ylim=c(0,20))
```



1: 改善あり  
2: 改善なし



## 「改善あり」の比較【前治療の有無別】

- ▶ 前治療の有無別に、薬剤ごとの「改善あり」の割合を求める

```
> ( TABLE4_NO <- prop.table(TABLE3[, ,1], 2) ) # 前治療なし
      GROUP
EVENT      A      B
  1 0.6666667 0.6
  2 0.3333333 0.4

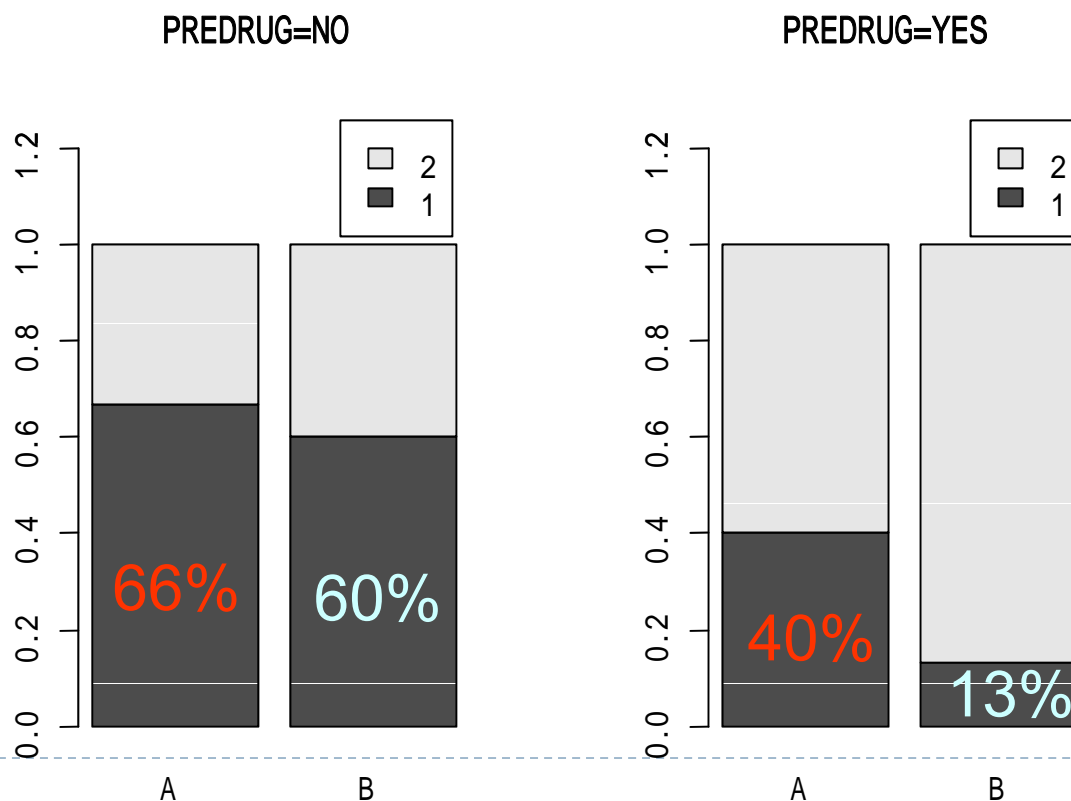
> ( TABLE4_YES <- prop.table(TABLE3[, ,2], 2) ) # 前治療あり
      GROUP
EVENT      A      B
  1 0.4 0.1333333
  2 0.6 0.8666667
```



## 「改善あり」の比較【前治療の有無別】

- ▶ 前治療の有無別・薬剤別の「改善あり」の割合に関する棒グラフを描く

```
> par(mfrow=c(1,2))  
> barplot(TABLE4_NO, legend=rownames(TABLE4_NO), main="PREDRUG=NO", ylim=c(0,1.3))  
> barplot(TABLE4_YES, legend=rownames(TABLE4_YES), main="PREDRUG=YES", ylim=c(0,1.3))
```



1:改善あり  
2:改善なし



## 本日のメニュー

---

1. イントロ

2. 割合と比

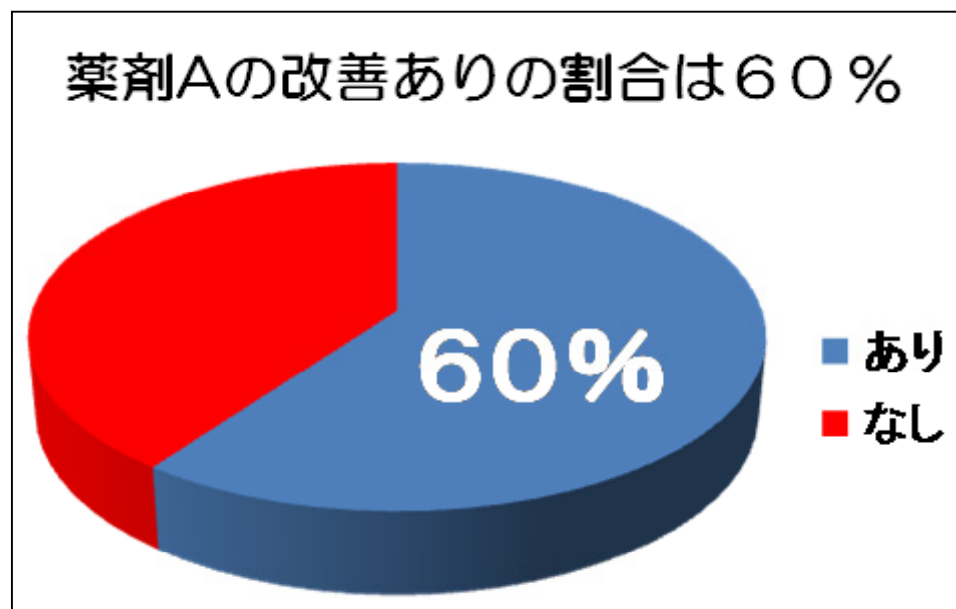
3. リスク差, リスク比, オッズ比

4.  $\chi^2$  検定



## 「割合」とは

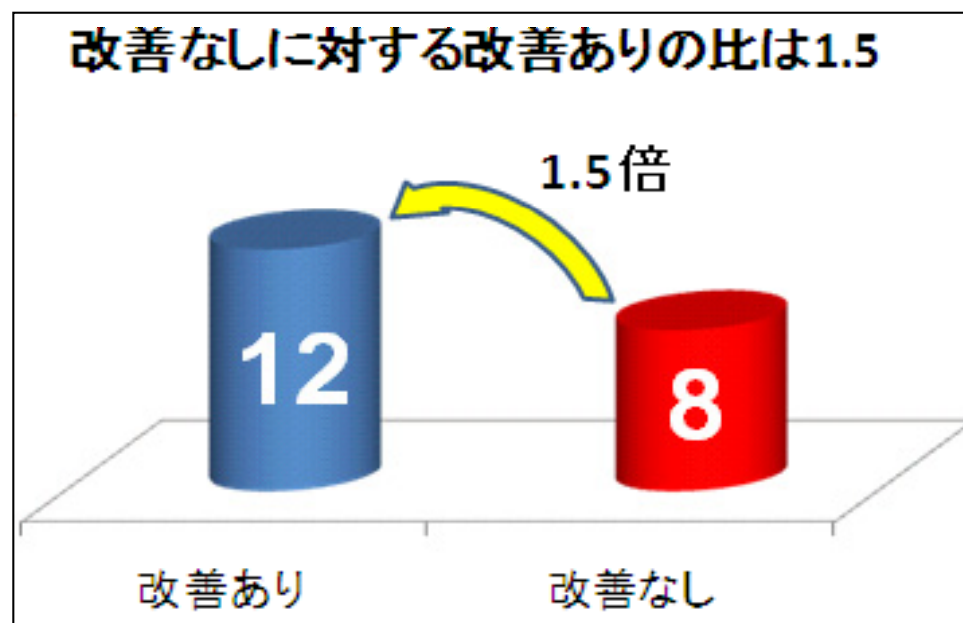
- ▶ 分母を「全体」、分子を「分母の一部」として割り算した値
- ▶ 薬剤 A の改善ありの割合：「薬剤 A の改善ありとなった患者さんの数」を「薬剤 A を投与した患者さん全員の数」で割り算した値  
 $12 \div 20 = 0.6$
- ▶ 割合に 100 をかけた値がパーセント（ $0.6 \times 100 = 60\%$ ）となる





## 「比」とは

- ▶ 性質が異なるもの同士を割り算した値
- ▶ 「男女の比」：「男性の数」と「女性の数」を割り算した値
- ▶ 「薬剤 A の改善なしの数」に対する「薬剤 A の改善ありの数」の比：  
 $12 \div 8 = 1.5$







## 【参考】 「率」とは

- ▶ ある事象が単位時間（例えば1年）の間に起こった数  
時速（1時間あたりに走る距離）のようなものとイメージ出来る
- ▶ 「200人を1年間観察した結果、6人が死亡した」場合：  
「死亡数（6人）」を「のべ観察時間（200人×1年＝200人年）」で  
割り算した値＝0.03（人/年）が率  
「1人年あたり0.03人が死亡する」と解釈する
- ▶ 上記の「0.03（人/年）」に1000をかけて  
「1000人年あたり30人が死亡する」  
「1000人を1年間観察した場合、6人が死亡する」とも解釈出来る

「割合」は「時間」の概念がないものに関する指標であるのに対し、  
「率」は単位時間の間に起こった頻度、と「時間」の概念が入った  
指標となっているのが相違点



## 本日のメニュー

---

1. イントロ
2. 割合と比
3. リスク差, リスク比, オッズ比
4.  $\chi^2$  検定



## 準備：パッケージのインストール等

### 1. パッケージ「epiR」と「lawstat」をインストールする

```
> options(repos="http://cran.md.tsukuba.ac.jp") # CRANの設定
> install.packages("epiR", dep=T) # インストール
> install.packages("epitools", dep=T) # インストール
```

### 2. とりあえず頻度集計を行う

```
> library(epiR); library(epitools)
> ( TABLE5 <- xtabs(~ GROUP + EVENT, data=AB) ) # 薬剤 見たい指標の順
      EVENT
GROUP  1  2
   A 12  8
   B  5 15
> table.margins(TABLE5) # 周辺の合計値を算出
      EVENT
GROUP  1  2 Total
   A   12  8   20
   B   5 15   20
 Total 17 23   40
```



## リスク差, リスク比, オッズ比の算出

### 3. リスク差, リスク比, オッズ比を算出する 詳細な説明は次頁以降

```
> epi .2by2(TABLE5)
```

	Disease +	Disease -	Total	Inc risk *	Odds
Exposed +	12	8	20	60.0	1.500
Exposed -	5	15	20	25.0	0.333
Total	17	23	40	42.5	0.739

Point estimates and 95 % CIs:

<b>Inc risk ratio</b>	<b>2.4 (1.04, 5.55)</b>
<b>Odds ratio</b>	<b>4.5 (1.17, 17.37)</b>
<b>Attrib risk *</b>	<b>35 (6.34, 63.66)</b>
Attrib risk in population *	17.5 (-6.89, 41.89)
Attrib fraction in exposed (%)	58.33 (3.56, 82)
Attrib fraction in population (%)	41.18 (-7.83, 67.91)

\* Cases per 100 population units



## 【参考】リスク比やオッズ比の算出

### 3. リスク比を算出する いろいろなオプションあり (詳細はヘルプ参照)

```
> epitab(TABLE5, rev="both", method="riskratio", pvalue="chi2")
```

```
$tab
```

```
      EVENT
GROUP 2  p0  1  p1 riskratio  lower  upper  p.value
  B 15 0.75  5 0.25      1.0      NA      NA      NA
  A  8 0.40 12 0.60      2.4 1.036928 5.554871 0.02516076
```

```
$measure
```

```
[1] "wald"
```

```
$conf.level
```

```
[1] 0.95
```

```
$pvalue
```

```
[1] "chi2"
```



## リスクとは

▶ リスク：ありの数÷全体の数

各薬剤の「改善ありのリスク」は「改善ありの数÷全体の数」で算出  
「各薬剤の改善ありの割合」を求めていることに相当する

「リスク」というと「悪いこと」が起きる割合，のようなイメージがあるが，ここでの「リスク」は「注目する事象がありの数÷全体の数」なので，「注目する事象」が例えば病状が改善すること（良いこと）であれば，「リスク」は「良いこと」が起きる割合となる点に注意

	改善あり	改善なし	リスク
薬剤 A	12	8	$12 \div 20 = 0.60 (60\%)$
薬剤 B	5	15	$5 \div 20 = 0.25 (25\%)$



## リスク差とは

▶ リスク差：薬剤間の割合の差

「薬剤 A の改善ありの割合」と「薬剤 B の改善ありの割合」の差となり，結果は「薬剤 A は薬剤 B よりも改善ありの割合が 35% 高い」

▶ 「割合が 35% 高い」という表現は誤解を招く？

「薬剤 B の割合 (25%) + 35%」という意味だが，人によっては「薬剤 B の割合 (25%) × 135%」と解釈するかもしれない

▶ リスク差を表す際は「薬剤 A は薬剤 B よりも割合が 35ポイント高い」と「ポイント」という単位を使う分野もある

	リスク	リスク差
薬剤 A	0.60 (60%)	0.60 - 0.25 = 0.35 (35%)
薬剤 B	0.25 (25%)	



## リスク比とは

- ▶ リスク比：薬剤間の割合の比（ある薬剤に対してリスクが何倍か）

「薬剤 A の改善ありの割合」と「薬剤 B の改善ありの割合」の比となり、結果は「薬剤 A は薬剤 B よりも改善ありの割合が 2.4 倍高い」

	リスク	リスク比
薬剤 A	0.60 (60%)	0.60 ÷ 0.25 = 2.4 (倍)
薬剤 B	0.25 (25%)	





## リスク差とリスク比

- ▶ 「薬剤 A は薬剤 B よりもリスク比は 2.4（割合が 2.4 倍）」とう場合  
下表の場合が考えられる
  - ▶ リスク比はいずれも 2.4 だが、リスク差は 0.35%（ほとんど差がない）  
から 35%（かなり差がある）まで様々
- リスク比だけでは発生割合にどの位の差があるか良く分からない場合  
があるので、リスク比に加えて各薬剤のリスクも確認する必要あり

薬剤 A のリスク	薬剤 B のリスク	リスク比	リスク差
0.6%	0.25%	2.4	0.35%
6%	2.5%	2.4	3.5%
30%	12.5%	2.4	17.5%
60%	25%	2.4	35%



## オッズとは

▶ オッズ：ありの数÷なしの数

各薬剤の「改善ありのオッズ」は「改善ありの数÷改善なしの数」  
で算出

	改善あり	改善なし	オッズ
薬剤 A	12	8	$12 \div 8 = 1.50$
薬剤 B	5	15	$5 \div 15 = 0.33$



## オッズ比とは

- ▶ オッズ比：薬剤間のオッズの比（ある薬剤に対してオッズが何倍か）

「薬剤 A の改善ありのオッズ」と「薬剤 B の改善ありのオッズ」の比となり，結果は「薬剤 A は薬剤 B よりもオッズが 4.5 倍高い」

	オッズ	オッズ比
薬剤 A	1.50	$1.50 \div 0.33 = 4.5$ (倍)
薬剤 B	0.33	



## リスク比とオッズ比

---

▶ リスクは「改善ありの割合」

- ▶ 例えば「薬剤 A の改善ありの割合が 60%」となった場合、この結果に対して解釈に困るようなことはない

▶ オッズは「ありの割合÷なしの割合」

- ▶ オッズは計算結果の解釈に困ることが少なくない
- ▶ 例えば「薬剤 A の改善ありのオッズが 1.5」となった場合、薬剤 A は効果があるのかわからないのか、薬剤 A により治療された患者さんは何割が改善したのかどうか、「1.5」という数字からは良く分からない
- ▶ ただ、「改善ありの割合」が小さい場合は「オッズ比 リスク比」となるため、オッズ比の計算結果をリスク比のように解釈することが出来る



## リスク比とオッズ比

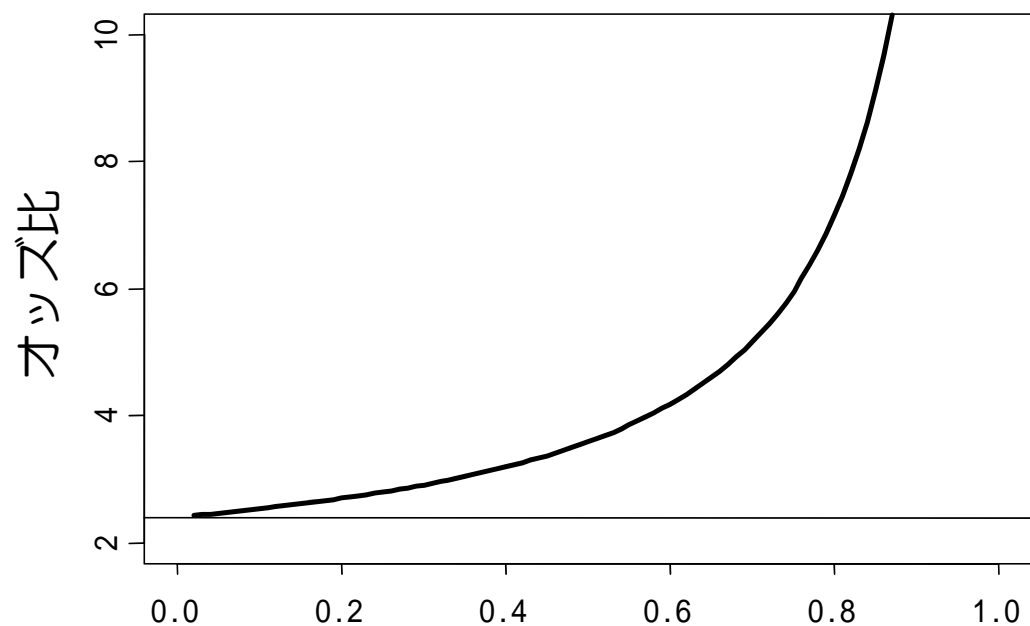
- ▶ 例として、リスク比を 2.4 と固定し、薬剤 A の「改善ありの割合」を変化させた時にオッズ比がどのように変化するかをみる
- ▶ 「改善ありの割合」が大きくなるとリスク比とオッズ比の差が開くが、「改善ありの割合」が小さい場合は「オッズ比 = リスク比」となっている点に注意

薬剤 A のリスク	薬剤 B のリスク	リスク比	オッズ比	
6%	2.5%	2.4	2.49	←リスク比 ≒オッズ比
24%	10%	2.4	2.84	
60%	25%	2.4	4.50	
90%	40%	2.4	36.0	←リスク比 ≠オッズ比



## リスク比とオッズ比

- ▶ 例として、リスク比を 2.4 と固定し、薬剤 A の「改善ありの割合」を変化させた時にオッズ比がどのように変化するかをみる
- ▶ 「改善ありの割合」が大きくなるとリスク比とオッズ比の差が開くが、「改善ありの割合」が小さい場合は「オッズ比 = リスク比」となっている点に注意





## リスク比とオッズ比

---

- ▶ オッズは「ありの割合÷なしの割合」
- ▶ 計算結果の解釈に困ることが少なくないような「オッズ比」を出力する理由は、「リスク比」はデータの集め方によってはバイアスが入る場合があるから
- ▶ 例えば、ケース・コントロールスタディなどの後ろ向き研究（過去に得られたデータをかき集めて解析を行う研究）では、収集したデータの集め方によって「リスク」が変わってしまう
- ▶ 例えば、QOL データの「改善あり」の数を 10 倍にしてみる



## リスク比とオッズ比

	改善あり	改善なし	リスク	リスク比
薬剤 A	120	8	$120 \div 128 = 0.94$	0.94 ÷ 0.77 = 1.22
薬剤 B	50	15	$50 \div 65 = 0.77$	
	改善あり	改善なし	オッズ	オッズ比
薬剤 A	120	8	$120 \div 8 = 15.00$	15.00 ÷ 3.33 = 4.50
薬剤 B	50	15	$50 \div 15 = 3.33$	

- ▶ リスク比：2.4 → 1.22 に変化，オッズ比：4.5 のまま
- ▶ もし，後ろ向き研究の結論を「リスク比は小さい」としたい場合，「改善ありの数」を増やすことでリスク比を小さくすることが出来る  
オッズ比ではこのような操作が出来ません





## リスク比とオッズ比のまとめ

---

- ▶ リスク比はオッズ比に比べて結果の解釈が容易で前向き研究（未来に向かって調査を行いデータを取って解析する研究）などの場合にはよく用いられる
- ▶ リスク比は後ろ向き研究（過去に得られたデータをかき集めて解析を行う研究）の場合にバイアスが入る余地があるため、そのような場合はオッズ比の方が望ましくなる
- ▶ 「改善ありの割合」が小さい場合は「オッズ比 リスク比」となるためオッズ比の計算結果をリスク比のように解釈することが出来る
- ▶ 次回紹介する「ロジスティック回帰」はオッズ比に対する手法  
2 値データに対するモデル解析は「ロジスティック回帰」なので、オッズ比を用いる、という考え方も出来る



## リスク差, リスク比, オッズ比の算出

- ▶ リスク差 (Attrib risk) : 35% (95%信頼区間 : [6.34%, 63.66%])
- ▶ リスク比 (Inc risk ratio) : 2.4 (95%信頼区間 : [1.04, 5.55])
- ▶ オッズ比 (Odds ratio) : 4.5 (95%信頼区間 : [1.17, 17.37])

Disease +	Disease -	Total	Inc risk *	Odds	
Exposed +	12	8	20	60.0	1.500
Exposed -	5	15	20	25.0	0.333
Total	17	23	40	42.5	0.739

Point estimates and 95 % CIs:

---

Inc risk ratio	2.4 (1.04, 5.55)
Odds ratio	4.5 (1.17, 17.37)
Attrib risk *	35 (6.34, 63.66)
Attrib risk in population *	17.5 (-6.89, 41.89)
Attrib fraction in exposed (%)	58.33 (3.56, 82)
Attrib fraction in population (%)	41.18 (-7.83, 67.91)

---

\* Cases per 100 population units



## 本日のメニュー

---

1. イントロ
2. 割合と比
3. リスク差, リスク比, オッズ比
4.  $\chi^2$  検定



## 改善ありの割合に関する $\chi^2$ 検定

- ▶ 「薬剤 A の改善ありの割合」と「薬剤 B の改善ありの割合」が等しいかどうかを検定する
  - ▶  $p = 2.5\%$ , 有意水準  $5\%$  で検定すると結果は有意
  - ▶ 有意なので改善ありの割合は等しくない

```
> ( TABLE5 <- xtabs(~ GROUP + EVENT, data=AB) )
```

```
      EVENT
GROUP  1  2
  A  12  8
  B   5 15
```

```
> chisq.test(TABLE5, correct=F)
```

```
      Pearson's Chi-squared test
```

```
data:  TABLE5
```

```
X-squared = 5.0128, df = 1, p-value = 0.02516  検定結果 ( p 値 = 約 2.5 %)
```



## 改善ありの割合に関する $\chi^2$ 検定

1. 比較の枠組み 薬剤 A と薬剤 B の改善ありの割合を比較する
2. 比較するものの間に差がないという仮説（帰無仮説  $H_0$ ）を立てる  
帰無仮説  $H_0$  : 薬剤 A の割合 = 薬剤 B の割合
3. 帰無仮説とは裏返しの仮説（対立仮説  $H_1$ ）を立てる  
対立仮説  $H_1$  : 薬剤 A の割合  $\neq$  薬剤 B の割合
4. 帰無仮説が成り立つという条件の下で、手元にあるデータ（よりも極端なこと）が起こる確率（= p 値）を計算  $p = 0.02516$  (2.5%)
5. 「確率が 2.5 %の珍しいデータが得られた」と考えずに  
「帰無仮説  $H_0$ が間違っている」と考え、対立仮説  $H_1$  が正しいと結論  
「改善ありの割合は異なる」と解釈する
6. 「割合は異なる」 & 「薬剤 A の割合 = 60% > 薬剤 B の割合 = 25%」  
の合わせ技で「薬剤 A の割合 > 薬剤 B の割合」と結論付ける



## 改善ありの割合に関する $\chi^2$ 検定

---

- ▶  $\chi^2$  検定の帰無仮説と対立仮説
  - ▶ 帰無仮説  $H_0$  : 薬剤 A の割合 = 薬剤 B の割合 (割合に違いが無い)
  - ▶ 対立仮説  $H_1$  : 薬剤 A の割合  $\neq$  薬剤 B の割合
- ▶ 「割合に違いがない」は以下のような場合に相当する
  - ▶ 帰無仮説  $H_0$  : 薬剤間のリスク差が 0 である
  - ▶ 帰無仮説  $H_0$  : 薬剤間のリスク比が 1 である
  - ▶ 帰無仮説  $H_0$  : 薬剤間のオッズ比が 1 である

$\chi^2$  検定は、上記の帰無仮説に対して検定を行っていることになる
- ▶ リスク差, リスク比, オッズ比の関係を次頁に挙げる



## 改善ありの割合に関する $\chi^2$ 検定

	薬剤 A の ありの割合	薬剤 B の ありの割合	リスク差	リスク比	オッズ比
①	0.60	0.59	0.01	1.02	1.04
②	0.60	0.10	0.50	6.00	13.50
③	0.60	0.95	-0.35	0.63	0.08

- ▶ ①：薬剤間の「ありの割合」にほとんど違いがない例  
リスク差はほぼ 0 に等しく，リスク比とオッズ比はほぼ 1 に等しい
- ▶ ②と③：薬剤間の「ありの割合」に違いがある例  
リスク差は 0 から遠ざかり，リスク比とオッズ比は 1 から遠ざかる
- ▶ 薬剤間の「ありの割合」に違いがない  
リスク差は 0 に，リスク比とオッズ比は 1 に近づく
- ▶ 薬剤間の「ありの割合」に違いがある  
リスク差は 0 から，リスク比とオッズ比は 1 から遠ざかる



## 雑談

改善ありの割合に関する  $\chi^2$  検定の場合について、

- ▶ 「改善ありの割合に関するリスク差の 95% 信頼区間」が「0」を 含んでいる場合は、 $\chi^2$  検定 の結果は有意にならない場合がほとんど
- ▶ 「改善ありの割合に関するリスク差の 95% 信頼区間」が「0」を 含んでいない場合は、 $\chi^2$  検定 の結果は有意である場合がほとんど
- ▶ 以下に例を挙げる（最初の例は今回の QOL データの結果）
  - ▶ リスク差の 95% 信頼区間が [6.34, 63.66]  $\chi^2$  検定の p 値 = 0.02516 (有意)
  - ▶ リスク差の 95% 信頼区間が [1.83, 61.33]  $\chi^2$  検定の p 値 = 0.04868 (有意)
  - ▶ リスク差の 95% 信頼区間が [-0.11, 58.75]  $\chi^2$  検定の p 値 = 0.06101 (有意でない)

$\chi^2$  検定（連続修正なし）の場合の話

ただし、信頼区間を計算する際の標準誤差と検定統計量を計算する際の標準誤差が若干異なるので、完全には対応しない





## 【参考】Fisherの正確検定

- ▶ Fisher の正確検定の帰無仮説  $H_0$  : 薬剤 A の割合 = 薬剤 B の割合
- ▶ データが超幾何分布に従っていると仮定した上で、得られたデータがどの位の確率で得られるものかを計算する手法 ( p 値 = 5.3%)

```
> fisher.test(TABLE5)

      Fisher's Exact Test for Count Data
data:  TABLE5
p-value = 0.05355
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.981708 21.946820
sample estimates:
odds ratio
 4.319064
```

- ▶  $\chi^2$  検定 (連続修正あり, 次頁) よりも有意差が出にくい



## 【参考】 $\chi^2$ 検定（連続修正あり）

- ▶  $\chi^2$  検定の帰無仮説  $H_0$  : 薬剤 A の割合 = 薬剤 B の割合
- ▶  $\chi^2$  検定（連続修正あり）は、 $\chi^2$  検定（連続修正なし）の結果を Fisher の正確検定に近づけるために補正したもので「連続修正なし」よりも有意差が出にくくなる特徴がある（ $\chi^2$  では有意差がなくなっている）

```
> chisq.test(TABLE5, correct=T)
```

```
      Pearson's Chi-squared test with Yates' continuity  
      correction
```

```
data:  TABLE5
```

```
X-squared = 3.6829, df = 1, p-value = 0.05497
```

- ▶ クロス表の各セルの中に 5 未満となっているものがある場合は  $\chi^2$  分布への収束が悪くなるため、 $\chi^2$  検定（連続修正あり）を用いた方がよいという話もある



## 参考文献

---

- ▶ 統計学（白旗 慎吾 著，ミネルヴァ書房）
- ▶ 宇宙怪人しまりす医療統計を学ぶ（佐藤 俊哉，岩波書店）
- ▶ Applied Logistic Regression（Hosmer & Lemeshow, Wiley）
- ▶ The R Tips 第2版（オーム社）
- ▶ R流！イメージで理解する統計処理入門（カットシステム）

# Rで統計解析入門

終