

R と RStudio の使い方

芳賀敏郎 (2016) 医薬品開発のための統計解析 第3部 非線形モデル
3 計数値の解析
3.4 多項分布 (名義尺度)

テキストと利用上の注意

●テキスト

芳賀敏郎（2016）医薬品開発のための統計解析

第3部 非線形モデル 改訂版、サイエンティスト社、p.288

（サイトへアップすることに対して、サイエンティスト社の了解を得ています）

●Rによる解析事例を紹介

R スクリプトの出力結果を紹介します（tidyverse 系には次期バージョンで対応します）

R スクリプト（文字コードUTF-8に設定）を、このサイトから[ダウンロード](#)できます

R スクリプトを [Compile Report] することにより、Word または HTML で見ることが出来ます

R と RStudio の設定と基本的な使い方は「[R と RStudio の使い方](#)」を参照してください

R の出力結果の見方は、テキストとそれを解説した [PDF ファイル](#)を参照してください

本PDF ファイルをダウンロードし、Adobe Acrobat Reader DC で注釈のメモを表示してください

●自己責任で利用

上記のことを理解した上で、自己責任により利用してください

第3部 非線形モデル

1. 非線形最小2乗法（基礎）

- 1.1 線形と非線形、1.2 非線形最小2乗法の基本的な考え方、1.3 指数曲線のあてはめ、
1.4 Emaxモデルとロジスティック曲線

2. 非線形最小2乗法（応用）

- 2.1 誤差を考慮した解析、2.2 効力比、2.3 併用効果（相乗・拮抗交換）、
2.4 モデルの探索（複数の曲線の同時あてはめ）、2.5 薬物動態の解析

3. 計数値の解析

- 3.1 2項分布、3.2 割合の推定・検定と区間推定、3.3 割合の差の推定・検定と区間推定、
3.4 多項分布（名義尺度）、3.5 多項分布（順序尺度）、3.6 要因が複数の場合

4. ロジスティック回帰分析

- 4.1 復習、4.2 ロジスティック回帰分析（基本）、4.3 ロジスティック回帰分析（応用）



クロス集計表の作成

xtabs 関数の使い方

分割表の作成

- スクリプトファイル：Green3-3-4d.R

利用した関数：xtabs、factor

方法：Excel ファイルから読み込んだデータフレームと、それをクロス集計して作成した分割表

```
##      group event
## 1      A   yes
## 2      A   no
## 3      B   no
## 4      A   yes
.....
## 8      B   no
## 9      A   yes
## 10     A   yes
## 11     B   yes
## 12     B   yes
## 13     A   yes
## 14     A   no
## 15     B   yes
```

データフレーム



集計した結果

```
##      group event freq
## 1      A   yes    6
## 2      A   no     2
## 3      B   yes    3
## 4      A   no     4
```

度数



```
##           event
## group yes no
## A      6  2
## B      3  4
```

分割表

分割表の作成

- スクリプトファイル：Green3-3-4d.R

利用した関数：xtabs、factor

方法：Excel ファイルから読み込んだデータフレームと、それをクロス集計して作成した分割表

```
##      group event
## 1      A   yes
## 2      A   no
## 3      B   no
## 4      A   yes
.....
## 8      B   no
## 9      A   yes
## 10     A   yes
## 11     B   yes
## 12     B   yes
## 13     A   yes
## 14     A   no
## 15     B   yes
```

データフレーム



集計した結果

```
##      group event freq
## 1      A   yes    6
## 2      A   no     2
## 3      B   yes     3
## 4      A   no     4
```

度数



```
##           event
## group yes  no
##  A     6   2
##  B     3   4
```

分割表

データの形

度数としてまとめてない場合
度数としてまとめてある場合

カテゴリの数と順序関係

2つ（二値データ）→ 2項分布
「yes, no」

3つ以上 → 多項分布

順序がない場合

「typeA, typeB, typeC」

順序がある場合

「none, low, medium, high」

分割表の作成

- スクリプトファイル：Green3-3-4d.R

利用した関数：xtabs、factor

方法：Excel ファイルから読み込んだデータフレームと、それをクロス集計して作成した分割表

```
##      group event
## 1      A   yes
## 2      A   no
## 3      B   no
## 4      A   yes
.....
## 8      B   no
## 9      A   yes
## 10     A   yes
## 11     B   yes
## 12     B   yes
## 13     A   yes
## 14     A   no
## 15     B   yes
```

データフレーム df1

集計した結果：データフレーム df2

```
##      group event Freq
## 1      A   yes    6
## 2      A   no     2
## 3      B   yes    3
## 4      A   no     4
```

度数の列名は、
自動的に「Freq」になる

```
df2 <- data.frame(table(
  group = df1$group,
  event = df1$event))
```

分割表の作成（クロス集計）

- 事例 1 : カテゴリカル変数 1 個、カテゴリー 2 個
スクリプトファイル : Green3-3-4d.R

```
df1
##      group event
## 1      A   yes
## 2      A   no
## 3      B   no
## 4      A   yes
## .....
## 8      B   no
## 9      A   yes
## 10     A   yes
## 11     B   yes
## 12     B   yes
## 13     A   yes
## 14     A   no
## 15     B   yes
```



```
tb11 <- xtabs(~ event, data = df1)
```

```
tb11
```

```
## event
##  no yes
##   6  9
```

```
df1$group <- factor(df1$group, levels = c("A", "B"))
df1$event <- factor(df1$event, levels = c("yes", "no"))
```

```
tb12 <- xtabs(~ group + event, data = df1)
```

```
tb12
```

```
##      event
## group yes no
##   A   6  2
##   B   3  4
```

チルダ「~」 event

変数名が反映

文字コードの順番

分割表の作成（クロス集計）

- 事例 2 : カテゴリカル変数 2 個、カテゴリーがそれぞれ 2 個
スクリプトファイル : Green3-3-4d.R

```
df1
##      group event
## 1      A   yes
## 2      A   no
## 3      B   no
## 4      A   yes
.....
## 8      B   no
## 9      A   yes
## 10     A   yes
## 11     B   yes
## 12     B   yes
## 13     A   yes
## 14     A   no
## 15     B   yes
```



```
tb11 <- xtabs(~ event, data = df1)
tb11
## event
## no yes
## 6 9
```

因子型に設定

位置を設定

```
df1$group <- factor(df1$group, levels = c("A", "B"))
df1$event <- factor(df1$event, levels = c("yes", "no"))
```

```
tb12 <- xtabs(~ group + event, data = df1)
tb12
##      event
## group yes no
## A      6 2
## B      3 4
```

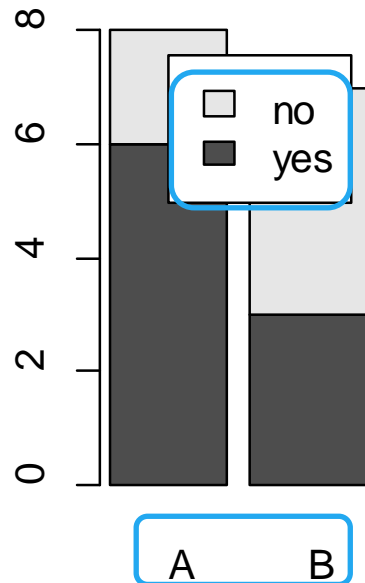
~ group + event
表側 表頭

2 × 2 分割表

levels の設定順

分割表の作成 (クロス集計)

- 事例 2: カテゴリカル変数 2 個、
カテゴリーがそれぞれ 2 個



```
##          event
## group  yes  no
##    A    6  2
##    B    3  4
```

```
levels = c("A", "B")
levels = c("yes", "no")
```

```
barplot(t(tb12), legend = TRUE)
```

```
Epi::twoby2(tb12)
```

t 関数で転置

```
## 2 by 2 table analysis:
## -----
## Outcome      : yes
## Comparing    : A vs. B
```

outcome: 第 1 列

yes の割合 →
yes/no に対するオッズ

Aの割合 / Bの割合

Aのオッズ / Bのオッズ

Aの割合 - Bの割合

```
##      yes no      P(yes) 95% conf. interval
## A     6  2     0.7500     0.3771     0.9370
## B     3  4     0.4286     0.1437     0.7702
##
##                                     95% conf. interval
##                                     Relative Risk: 1.7500     0.6806     4.4994
##                                     Sample Odds Ratio: 4.0000     0.4471    35.7876
##                                     Probability difference: 0.3214    -0.1467     0.6454
```

分割表の作成 (クロス集計)

- 事例 3 : カテゴリカル変数 3 個、カテゴリがそれぞれ 2 個

スクリプトファイル

Green3-3-4d.R

df2

```
##      sex    age event freq
## 1  male  young   yes   16
## 2  male  young   no    34
## 3  male  mature  yes    5
## 4  male  mature  no   43
## 5 female  young   yes   13
## 6 female  young   no   41
## 7 female  mature  yes    6
## 8 female  mature  no   43
```

度数

ftable

```
df2$sex <- factor(df2$sex, levels = c("male", "female"))
df2$age <- factor(df2$age, levels = c("young", "mature"))
df2$event <- factor(df2$event, levels = c("yes", "no"))
```

```
tb21 <- xtabs(freq ~ age + event, data = df2)
ftable(tb21)
```

度数 ~ 変数 + 変数
行 列

```
##           event yes no
## age
## young          29 75
## mature          11 86
```

```
tb22 <- xtabs(freq ~ sex + event, data = df2)
ftable(tb22)
```

適度に空白が入って
見やすくなっている

```
##           event yes no
## sex
## male          21 77
## female        19 84
```

分割表の作成 (クロス集計)

- 事例 3 : カテゴリカル変数 3 個、カテゴリーがそれぞれ 2 個

スクリプトファイル : Green3-3-4d.R

df2

##	sex	age	event	freq
## 1	male	young	yes	16
## 2	male	young	no	34
## 3	male	mature	yes	5
## 4	male	mature	no	43
## 5	female	young	yes	13
## 6	female	young	no	41
## 7	female	mature	yes	6
## 8	female	mature	no	43

度数



```
tb23 <- xtabs(freq ~ I(age:sex) + event, data = df2)
ftable(tb23)
```

I 関数の利用

##	event	yes	no
## I(age:sex)			
## young:male		16	34
## young:female		13	41
## mature:male		5	43
## mature:female		6	43

2次元の分割表

```
tb24 <- xtabs(freq ~ age + sex + event, data = df2)
ftable(tb24)
```

##	(1)	(2)	(3)	event	yes	no
## age		sex				
## young		male			16	34
##		female			13	41
## mature		male			5	43
##		female			6	43

集計した順番

3次元の分割表

分割表の作成（クロス集計）

- 事例 3：カテゴリカル変数 3 個、カテゴリーがそれぞれ 2 個
スクリプトファイル：Green3-3-4d.R

3次元分割表の指定方法

tb24[age, sex, event] のように、集計した順番で指定

(1) tb24["young", "male", "yes"]・・・ブルー矢印の「16」

(2) tb24["young", ,]・・・オレンジ枠の 2×2 分割表（age="young" に限定）

(3) tb24[, "male",]・・・ブルー枠の 2×2 分割表（sex="male" に限定）

```
GTest(tb24[, "male", ],  
      correct = "none")
```

ブルー枠の
2×2分割表の
尤度比検定

```
tb24 <- xtabs(freq ~ age + sex + event, data = df2)  
ftable(tb24)
```

##		event	yes	no
##	age	sex		
##	young	male	16	34
##		female	13	41
##	mature	male	5	43
##		female	6	43

集計した順番

3次元の分割表

分割表の作成（クロス集計）

- 事例 4 : カテゴリカル変数 3 個、カテゴリ 2, 4, 4 個、1 つのカテゴリに順序関係あり
カテゴリに順序を指定（クロス集計による分割表の作成方法は同じ）

df3

##	group	event	degree	freq
## 1	control	none	none	52
## 2	control	type1	low	10
## 3	control	type2	med	5
## 4	control	type3	high	3
## 5	treated	none	none	33
## 6	treated	type1	low	8
## 7	treated	type2	med	10
## 8	treated	type3	high	9

カテゴリ 4 つ

順序に意味なし

順序に意味あり

```
df3$event <- factor(df$event,  
  levels = c("none", "type1", "type2", "type3"))  
  
df3$degree <- factor(df$degree,  
  levels = c("high", "med", "low", "none"),  
  ordered = TRUE)  
  
df3$event  
## Levels: none type1 type2 type3  
  
df3$degree  
## Levels: high < med < low < none
```

levels で位置を指定

levels の位置で順序を指定

順序関係がない

順序関係がある

分割表の作成

- 集計結果を格納するデータ構造：テーブル
スクリプトファイル：Green3-3-4d.R

```
df1
##      group event
## 1      A   yes
## 2      A   no
## 3      B   no
## 4      A   yes
.....
## 8      B   no
## 9      A   yes
## 10     A   yes
## 11     B   yes
## 12     B   yes
## 13     A   yes
## 14     A   no
## 15     B   yes
```

データ構造：データフレーム

データ構造

- ベクトル
- マトリックス（行列）
- データフレーム
- リスト
- 配列
- テーブル・・・xtabsなどでの集計結果、整数

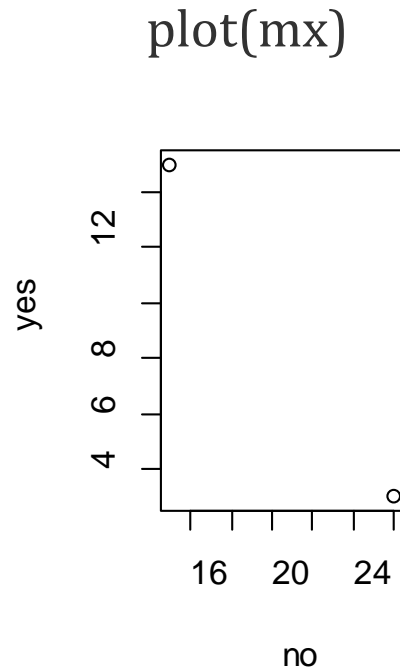
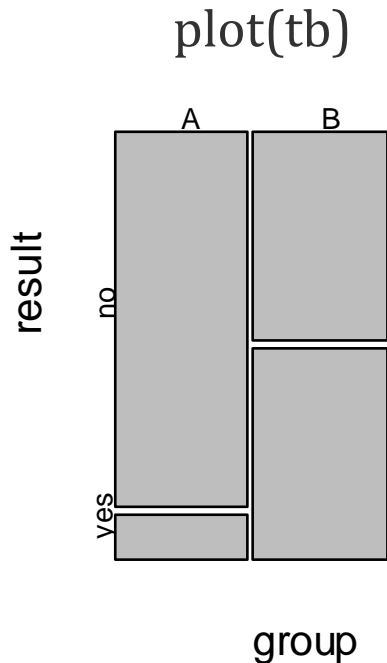
```
tb12 <- xtabs(~ group + event, data = df1)
```

```
tb12
##      event
## group yes no
##   A   6  2
##   B   3  4
```

データ構造：テーブル

分割表の作成

- 集計結果を格納するデータ構造：テーブル
スクリプトファイル：Green3-3-4d.R



```
mx <- matrix(c( 26, 4,
               15, 15),
             nrow = 2, ncol = 2,
             byrow = TRUE)
dimnames(mx) <- list(group = c("A", "B"),
                    result = c("yes", "no"))
```

ベクトル c(26, ...)

```
tb <- as.table(mx)
```

テーブルに変換

```
mx
##      result
## group yes no
##      A  26  3
##      B  15 15
```

マトリックス mx

```
tb
##      result
## group yes no
##      A  26  3
##      B  15 15
```

テーブル tb

分割表の作成（分割表からデータフレームに変換）

- スクリプトファイル：Green3-3-4d.R

利用した関数：reshape2::melt、lapply、function、rep

方法：分割表からデータフレームに変換

```
tb12
##      event
## group yes no
##      B   3  4
##      A   6  2
```

```
df5
##      group event value
## 1      B   yes     3
## 2      A   yes     6
## 3      B   no      4
## 4      A   no      2
```

自動的に命名

```
df5 <- melt(tb12)

df6 <- data.frame(lapply(
  df5, function(i) rep(i, df5[, "value"]))) [-3]
```

value を除去
(3列目)

青木(2009)

```
df6
##      group event
## 1      B   yes
## 2      B   yes
## 3      B   yes
## 4      A   yes
## 5      A   yes
## 6      A   yes
## 7      A   yes
## 8      A   yes
## 9      A   yes
## 10     B   no
## 11     B   no
## 12     B   no
## 13     B   no
## 14     A   no
## 15     A   no
```



2 項分布と多項分布

2 項分布と多項分布

● 2 項分布と多項分布

2 項分布

目的変数が 2 つのカテゴリ（2 値データ）
2×2 分割表、a×2 分割表

2×2 分割表

処理	副作用	
	なし	あり
対照群	52	18
比較群	33	27

a×2 分割表

処理	副作用	
	なし	あり
対照群	42	18
比較群1	33	27
比較群2	24	36

多項分布

目的変数が 3 つ以上のカテゴリ
2×b 分割表、a×b 分割表 (b≥3)

2×b 分割表

処理	副作用		
	なし	発熱	下痢
対照群	52	10	5
比較群	33	8	10

a×b 分割表

処理	副作用		
	なし	発熱	下痢
対照群	22	23	15
比較群1	29	21	10
比較群2	37	13	10

分割表の解析方法

割合の差の検定 (prop.test)

尤度比検定 (GTest、assocstats)

Pearson のカイ 2 乗検定 (chisq.test、assocstats)

Fisher の正確検定 (fisher.test)

2×b 分割表

処理	薬効		
	-	+	++
対照群	52	10	5
比較群	33	8	10

a×b 分割表

処理	薬効		
	-	+	++
対照群	22	23	15
比較群1	29	21	10
比較群2	37	13	10

2 項分布と多項分布

● 2 項分布と多項分布

2 項分布

2 項ロジスティック回帰 → glm

2×2 分割表

処理	副作用	
	なし	あり
対照群	52	18
比較群	33	27

a×2 分割表

処理	副作用	
	なし	あり
対照群	42	18
比較群1	33	27
比較群2	24	36

多項分布

多項ロジスティック回帰

名義ロジスティック回帰

b に順序がない → nnet::multinom

など

順序ロジスティック回帰

b に順序がある → MASS::polr

など

2×b 分割表

処理	副作用		
	なし	発熱	下痢
対照群	52	10	5
比較群	33	8	10

a×b 分割表

処理	副作用		
	なし	発熱	下痢
対照群	22	23	15
比較群1	29	21	10
比較群2	37	13	10

2×b 分割表

処理	薬効		
	-	+	++
対照群	52	10	5
比較群	33	8	10

a×b 分割表

処理	薬効		
	-	+	++
対照群	22	23	15
比較群1	29	21	10
比較群2	37	13	10

2 項分布と多項分布

● 2 項分布と多項分布

2 項分布

2 項ロジスティック回帰 → glm

2×2 分割表

処理	副作用	
	なし	あり
対照群	52	18
比較群	33	27

a×2 分割表

処理	副作用	
	なし	あり
対照群	42	18
比較群1	33	27
比較群2	24	36

多項分布

多項ロジスティック回帰

名義ロジスティック

本節

b に順序がない → nnet::multinom

など

順序ロジスティック

次節

b に順序がある → MASS::polr

など

2×b 分割表

処理	副作用		
	なし	発熱	下痢
対照群	52	10	5
比較群	33	8	10

a×b 分割表

処理	副作用		
	なし	発熱	下痢
対照群	22	23	15
比較群1	29	21	10
比較群2	37	13	10

2×b 分割表

処理	薬効		
	-	+	++
対照群	52	10	5
比較群	33	8	10

a×b 分割表

処理	薬効		
	-	+	++
対照群	22	23	15
比較群1	29	21	10
比較群2	37	13	10



多項分布： $2 \times b$ 分割表

独立性の検定

多項分布：2 × b 分割表

度数

● 表示3.4.1 2 × 4 の分割表の仮想データ

スクリプトファイル：Green3-3-4a.R

利用した関数：factor

方法

Excel ファイルからデータを読み込み、

データフレームとして付値

(2つの異なる実験データを合わせて表示)

df	##	group	event	degree	y	freq	d	dd
	## 1	control	none	none	1	52	0	-1
	## 2	control	type1	low	2	10	0	-1
	## 3	control	type2	med	3	5	0	-1
	## 4	control	type3	high	4	3	0	-1
	## 5	treated	none	none	1	33	1	1
	## 6	treated	type1	low	2	8	1	1
	## 7	treated	type2	med	3	10	1	1
	## 8	treated	type3	high	4	9	1	1

4つのカテゴリーに
順序関係はない

4つのカテゴリーに
順序関係がある

```
df <- read_excel("Green3-3.xlsx", sheet = "34-2xb")  
df <- data.frame(df)
```

```
df$group <- factor(df$group, levels = c("control", "treated"))
```

```
df$event1 <- factor(df$event,  
                    levels = c("none", "type1", "type2", "type3"))
```



多項分布：2 × b 分割表

● 表示3.4.1 2 × 4 の分割表の仮想データ

スクリプトファイル：Green3-3-4a.R

利用した関数：factor

方法

Excel ファイルからデータを読み込み、
データフレームとして付値
(2つの異なる実験データを合わせて表示)

df	##	group	event	degree	y	freq	d	dd
	## 1	control	none	none	1	52	0	-1
	## 2	control	type1	low	2	10	0	-1
	## 3	control	type2	med	3	5	0	-1
	## 4	control	type3	high	4	3	0	-1
	## 5	treated	none	none	1	33	1	1
	## 6	treated	type1	low	2	8	1	1
	## 7	treated	type2	med	3	10	1	1
	## 8	treated	type3	high	4	9	1	1

度数

4つのカテゴリーに
順序関係はない

levels で
カテゴリーの位置を
設定

```
df <- read_excel("Green3-3.xlsx", sheet = "34-2xb")
df <- data.frame(df)

df$group <- factor(df$group, levels = c("control", "treated"))

df$event1 <- factor(df$event,
                    levels = c("none", "type1", "type2", "type3"))
```


多項分布：2 × b 分割表

● 表示3.4.1 2 × 4 の分割表の仮想データ

スクリプトファイル：Green3-3-4a.R

利用した関数：xtabs

方法

データフレームからクロス集計して
分割表を作成

##	group	event	degree	y	freq	d	dd
## 1	control	none	none	1	52	0	-1
## 2	control	type1	low	2	10	0	-1
## 3	control	type2	med	3	5	0	-1
## 4	control	type3	high	4	3	0	-1
## 5	treated	none	none	1	33	1	1
## 6	treated	type1	low	2	8	1	1
## 7	treated	type2	med	3	10	1	1
## 8	treated	type3	high	4	9	1	1

```
tb1 <- xtabs(freq ~ group + event, data = df)
```

```
tb1
```

```
##          event
## group    none type1 type2 type3
## control   52   10    5    3
## treated   33    8   10    9
```

クロス集計

2 × 4 分割表

多項分布：2 × b 分割表

- 表示3.4.2 2 × 4 の分割表のグラフ化

表示3.4.3 JMP による棒グラフ

スクリプトファイル：Green3-3-4a.R

利用した関数：proportions、round、barplot

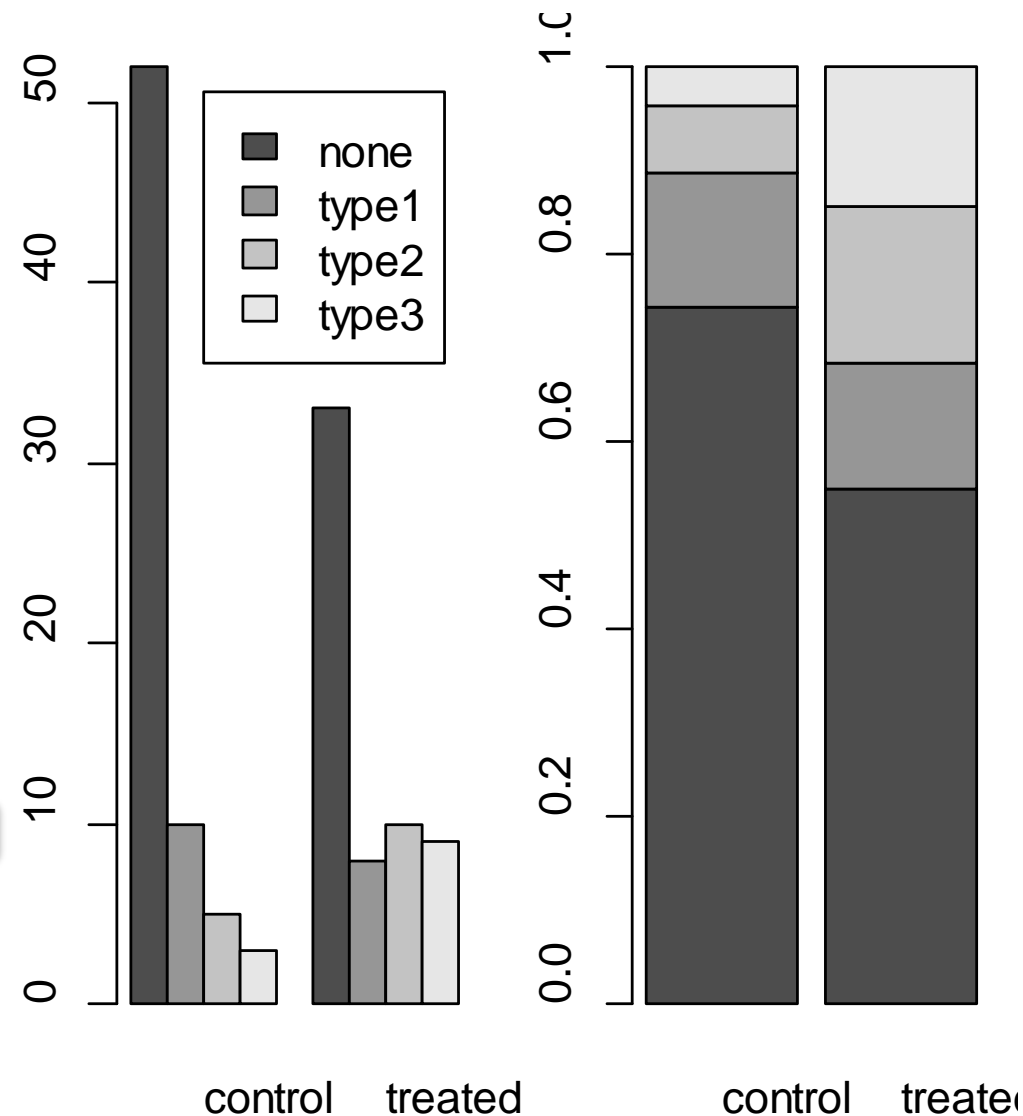
```
tb2 <- proportions(tb1, 1) # 行方向の割合  
round(addmargins(tb2, 2), digits = 2)
```

```
##          event  
## group    none type1 type2 type3  Sum  
## control 0.74  0.14  0.07  0.04  1.00  
## treated 0.55  0.13  0.17  0.15  1.00
```

```
barplot(t(tb1), beside = TRUE, legend = TRUE)  
barplot(t(tb2))
```

行の割合

凡例の表示



- 表示3.4.4 JMP [二変量の関係]の出力、表示3.4.5 分割表の尤度比検定

スクリプトファイル：Green3-3-4a.R

利用した関数：vcd::assocstats

DescTools::GTest

方法：尤度比検定、Pearson のカイ 2 乗検定

```
tb1 ##           event
## group      none type1 type2 type3
## control   52    10    5    3
## treated   33     8   10    9
```

```
assocstats(tb1) # 尤度比検定、カイ 2 乗検定
```

```
##           X^2 df P(> X^2)
## Likelihood Ratio 8.5743  3 0.035520
## Pearson          8.4165  3 0.038144
##
## Phi-Coefficient   : NA
## Contingency Coeff.: 0.247
## Cramer's V       : 0.254
```

```
GTest(tb1, correct = "none") # 尤度比検定 (G 検定)
```

```
##
## Log likelihood ratio (G-test) test of independence without correction
##
## data:  tb1
## G = 8.5743, X-squared df = 3, p-value = 0.03552
```

多項分布：2 × b 分割表

p.191

- 表示3.4.4 JMP [二変量の関係]の出力

表示3.4.6 分割表のカイ2乗検定

スクリプトファイル：Green3-3-4a.R

利用した関数：chisq.test、round

方法

Pearson のカイ2乗検定

期待値、相対的外れを表示

```
chi_out <- chisq.test(tb1, correct = FALSE)
chi_out

##
## Pearson's Chi-squared test
##
## data:  tb1
## X-squared = 8.4165, df = 3, p-value = 0.03814
round(chi_out$expected, digits = 3)

##          event
## group      none type1 type2 type3
## control 45.769 9.692 8.077 6.462
## treated 39.231 8.308 6.923 5.538

round(chi_out$residuals, digits = 3)

##          event
## group      none type1 type2 type3
## control  0.921  0.099 -1.083 -1.362
## treated -0.995 -0.107  1.169  1.471
```

連続修正なし

期待度数

相対的外れ

- 表示3.4.4 JMP [二変量の関係]の出力 (補足)

スクリプトファイル：Green3-3-4a.R

利用した関数：fisher.test

方法

Fisher の正確検定を 2 × 4 分割表に適用
(両側検定)

tb1

##	event				
##	group	none	type1	type2	type3
##	control	52	10	5	3
##	treated	33	8	10	9

```
fisher.test(tb1)
```

```
##  
## Fisher's Exact Test for Count Data  
##  
## data: tb1  
## p-value = 0.03968  
## alternative hypothesis: two.sided
```

両側検定

多項分布： $2 \times b$ 分割表

名義ロジスティック回帰

- 表示3.4.7 JMP [モデルのあてはめ] の出力

スクリプトファイル：Green3-3-4b.R

利用した関数：nnet::multinom、factor

```
df
##      group event degree y freq d dd
## 1 control  none   none  1  52  0 -1
## 2 control type1    low  2  10  0 -1
## 3 control type2    med  3   5  0 -1
## 4 control type3    high  4   3  0 -1
## 5 treated  none   none  1  33  1  1
## 6 treated type1    low  2   8  1  1
## 7 treated type2    med  3  10  1  1
## 8 treated type3    high  4   9  1  1
```

質的変数「group」を
2種類のダミー変数で表す
(前節 [§3.3](#) 参照)
(第2部 [§2.3](#) 参照)

先頭

```
df$group <- factor(df$group, levels = c("control", "treated"))
df$event <- factor(df$event, levels = c("none", "type1", "type2", "type3"))
mn_out <- multinom(event ~ group, weights = freq, data = df, Hess = TRUE)
```

多項分布：2 × b 分割表 名義ロジスティック

- 表示3.4.7 JMP [モデルのあてはめ] の出力

スクリプトファイル：Green3-3-4b.R

利用した関数：nnet::multinom、factor

```
df
##      group event degree y freq d dd
## 1 control  none   none  1   52 0 -1
## 2 control type1    low  2   10 0 -1
## 3 control type2    med  3    5 0 -1
## 4 control type3   high  4    3 0 -1
## 5 treated  none   none  1   33 1  1
## 6 treated type1    low  2    8 1  1
## 7 treated type2    med  3   10 1  1
## 8 treated type3   high  4    9 1  1
```

"contr.treatment" (デフォルト) で処理する場合、基準の control を先頭に置く

none を基準としたオッズ比を計算
none を先頭に置く

```
df$group <- factor(df$group, levels = c("control", "treated"))
df$event <- factor(df$event, levels = c("none", "type1", "type2", "type3"))
mn_out <- multinom(event ~ group, weights = freq, data = df, Hess = TRUE)
```


- 表示3.4.7 JMP [モデルのあてはめ] の出力
スクリプトファイル

Green3-3-4b.R

利用した関数

nnet::multinom、summary

方法

説明変数に質的変数のgroupを使う

contrasts の設定はデフォルト

("contr.treatment")

表示3.4.7 の JMP の結果と一致しない

質的変数

```
mn_out <- multinom(event ~ group, weights = freq,  
                    data = df, Hess = TRUE)
```

```
summary(mn_out)
```

```
## Coefficients:
```

```
##      (Intercept)  grouptreated
```

```
## type1   -1.648666    0.2313055
```

```
## type2   -2.341664    1.1475662
```

```
## type3   -2.852134    1.5525715
```

```
##
```

```
## Std. Errors:
```

```
##      (Intercept)  grouptreated
```

```
## type1    0.3453044    0.5239826
```

```
## type2    0.4681978    0.5911982
```

```
## type3    0.5936406    0.7027366
```

- 表示3.4.7 JMP [モデルのあてはめ] の出力

スクリプトファイル

Green3-3-4b.R

利用した関数

nnet::multinom、summary

方法

説明変数にダミー変数 1 を使う

説明変数に group を使った結果と一致

(contrasts の設定はデフォルト)

group はダミー変数 1 で

処理されていることがわかる

表示3.4.7 の JMP の結果と一致しない

ダミー変数 1

```
mn_out1 <- multinom(event ~ d, weights = freq,
                    data = df, Hess = TRUE)

summary(mn_out1)

## Coefficients:
##           (Intercept)                d
## type1    -1.648666  0.2313055
## type2    -2.341664  1.1475662
## type3    -2.852134  1.5525715
##
## Std. Errors:
##           (Intercept)                d
## type1     0.3453044  0.5239826
## type2     0.4681978  0.5911982
## type3     0.5936406  0.7027366
```

- 表示3.4.7 JMP [モデルのあてはめ] の出力

スクリプトファイル

Green3-3-4b.R

利用した関数

nnet::multinom、summary

方法

説明変数にダミー変数2を使う

表示3.4.7 の JMP の結果と一致

JMP はダミー変数2 で処理している

ダミー変数2

```
mn_out2 <- multinom(event ~ dd, weights = freq,
                    data = df, Hess = TRUE)

summary(mn_out2)

## Coefficients:
##      (Intercept)          dd
## type1   -1.532938  0.1158710
## type2   -1.767889  0.5739790
## type3   -2.075985  0.7766544
##
## Std. Errors:
##      (Intercept)          dd
## type1    0.2619852  0.2619852
## type2    0.2956090  0.2956090
## type3    0.3514172  0.3514172
```

多項分布：2 × b 分割表 名義ロジスティック

p.192

基準(-1)

- 表示3.4.7 JMP [モデルのあてはめ] の出力
スクリプトファイル

Green3-3-4b.R

利用した関数

nnet::multinom、summary、factor

方法

group の制約条件を conr.sum に
変更して multinom 関数を実行
ダミー変数 2 を使った結果と一致
表示3.4.7 の JMP の結果と一致

JMP は質的変数をダミー変数 2 で
処理している

このパラメータ推定値は役立たないと
思われる (テキスト参照)

```
levels(df$group) <- c("treated", "control")
mn_out3 <- multinom(event ~ group,
                    weights = freq,
                    data = df, Hess = TRUE,
                    contrasts = list(group = "contr.sum"))
summary(mn_out3)
```

質的変数

group の制約条件

```
## Coefficients:
##      (Intercept)      group1
## type1  -1.532938  0.1158710
## type2  -1.767889  0.5739790
## type3  -2.075985  0.7766544
##
## Std. Errors:
##      (Intercept)      group1
## type1  0.2619852  0.2619852
## type2  0.2956090  0.2956090
## type3  0.3514172  0.3514172
```

- 表示3.4.7 JMP [モデルのあてはめ] の出力
スクリプトファイル

Green3-3-4b.R

利用した関数

nnet::multinom、summary、
anova、car::Anova

方法

切片のみのモデル (null モデル) と
あてはめたモデルの比較

↓

効果の尤度比検定

car::Anova の利用

null モデル

```
mn_out0 <- multinom(event ~ 1, weights = freq,  
                    data = df, Hess = TRUE)  
anova(mn_out0, mn_out, test = "Chisq")  
## Likelihood ratio tests of Multinomial Models  
## Response: event  
## df Resid. Dev Test Df LR stat. Pr(Chi)  
## 21 265.3756  
## 18 256.8013 1 vs 2 3 8.574321 0.03552007  
  
Anova(mn_out, type = 2)  
## Response: event  
## LR Chisq Df Pr(>Chisq)  
## d 8.5743 3 0.03552
```



多項分布： $a \times b$ 分割表 ($a \geq 3$ 、 $b \geq 3$)

独立性の検定

独立性の検定 (多重比較)

残差分析

多項分布：a × b 分割表

● 表示3.4.9 4 × 4 の分割表のカイ 2 乗検定

スクリプトファイル

Green3-3-4c.R

利用した関数

factor、xtabs、addmargins

```
df1
##      group event freq
## 1      old    A     2
## 2      old    B     5
## 3      old    C     2
## 4      old    D     1
## 5     new1    A     2
## .....
## 14     new3    B     2
## 15     new3    C     7
## 16     new3    D     0
```

```
df1 <- read_excel("Green3-3.xlsx",
                  sheet = "34-4x4")
df1 <- data.frame(df1)
df1$group <- factor(df1$group,
                   levels = c("old", "new1", "new2", "new3"))
df1$event <- factor(df1$event,
                   levels = c("A", "B", "C", "D"))

tb1 <- xtabs(freq ~ group + event, data = df1)
tb2 <- addmargins(tb1, 2)
ftable(tb2)
```

##	event	A	B	C	D	Sum
## group						
## old		2	5	2	1	10
## new1		2	4	2	2	10
## new2		0	2	5	3	10
## new3		1	2	7	0	10

カテゴリーの
位置指定

4 × 4
分割表

多項分布：a × b 分割表

- 表示3.4.9 4 × 4 の分割表のカイ 2 乗検定

スクリプトファイル

Green3-3-4c.R

利用した関数

chisq.test、ftable、round

方法

カイ 2 乗検定

```
ftable(tb1)
```

```
##          event  A  B  C  D
## group
## old          2  5  2  1
## new1         2  4  2  2
## new2         0  2  5  3
## new3         1  2  7  0
```

```
chi_out <- chisq.test(tb1, correct = FALSE)
chi_out
## Pearson's Chi-squared test
## X-squared = 12.11, df = 9, p-value = 0.2072
```

連続修正なし

```
chi_out$expected
##          event
## group      A    B  C  D
## old    1.25 3.25 4 1.5
## new1   1.25 3.25 4 1.5
## new2   1.25 3.25 4 1.5
## new3   1.25 3.25 4 1.5
```

期待度数

```
round(chi_out$residuals, digits = 3)
##          event
## group      A      B      C      D
## old    0.671  0.971 -1.000 -0.408
## new1   0.671  0.416 -1.000  0.408
## new2  -1.118 -0.693  0.500  1.225
## new3  -0.224 -0.693  1.500 -1.225
```

相対的外れ

- 表示3.4.10 分割表の尤度比検定、表示3.4.11 JMP[二変量の関係]の出力

スクリプトファイル

Green3-3-4c.R

利用した関数

chisq.test、ftable

方法

尤度比検定 (G 検定)

```
ftable(tb1)
```

```
##          event  A  B  C  D
## group
## old          2  5  2  1
## new1         2  4  2  2
## new2         0  2  5  3
## new3         1  2  7  0
```

```
GTest(tb1, correct = "none")
```

```
##
##  Log likelihood ratio (G-test) test of
##  independence without correction
##
## data:  tb1
## G = 14.418, X-squared df = 9, p-value = 0.1082
```

```
assocstats(tb1)
```

```
##                X^2 df P(> X^2)
## Likelihood Ratio 14.418  9  0.10821
## Pearson          12.110  9  0.20716
##
## Phi-Coefficient   : NA
## Contingency Coeff.: 0.482
## Cramer's V       : 0.318
```

多項分布：a × b 分割表 多重比較（補足）

● a×b 分割表の多重比較

スクリプトファイル：Green3-3-4e.R

利用した関数

rcompanion::pairwiseNominalIndependence、

lucid::lucid

事例：4群（A1, A2, A3, A4）ごとに100例を調査。
結果を4種類のパターン（B1, B2, B3, B4）に分類。
群間でパターンに違いがあるか

（永田・吉田, 1997）

分割表

tb	##	pattern				
	##	group	B1	B2	B3	B4
	##	A1	14	16	30	40
	##	A2	30	25	18	27
	##	A3	18	32	25	25
	##	A4	8	12	23	57

比較対象
表側 row

```
pw_out1 <- pairwiseNominalIndependence(
  tb,
  compare = "row",
  fisher = FALSE,
  gtest = FALSE,
  chisq = TRUE,
  method = "holm",
  correct = "none",
  cramer = FALSE,
  digits = 3)
```

比較対象

検定方法

有意水準の
調整方法

```
lucid(pw_out1)
##      Comparison      p.Chisq p.adj.Chisq
## 1      A1 : A2  0.004      0.016
## 2      A1 : A3  0.0208     0.0624
## 3      A1 : A4  0.106      0.212
## 4      A2 : A3  0.166      0.212
## 5      A2 : A4  0.00000268  0.0000161
## 6      A3 : A4  0.0000121   0.0000605
```

● a×b 分割表の多重比較

スクリプトファイル：Green3-3-4e.R

利用した関数

rcompanion::pairwiseNominalIndependence、

lucid::lucid

事例：4群（A1, A2, A3, A4）ごとに100例を調査。

結果を4種類のパターン（B1, B2, B3, B4）に分類。

群間でパターンに違いがあるか

（永田・吉田, 1997）

比較対象

tb	##	pattern				
	##	group	B1	B2	B3	B4
	##	A1	14	16	30	40
	##	A2	30	25	18	27
	##	A3	18	32	25	25
	##	A4	8	12	23	57

```
pw_out1 <- pairwiseNominalIndependence(
  tb,
  compare = "row",
  fisher = FALSE,
  gtest = FALSE,
  chisq = TRUE,
  method = "holm",
  correct = "none",
  cramer = FALSE,
  digits = 3)
```

Hol 法の調整 p 値
すべてを使うわけではない
(第2部 [§1.3](#)参照)

```
lucid(pw_out1)
##      Comparison      p.Chisq p.adj.Chisq
## 1      A1 : A2 0.004          0.016
## 2      A1 : A3 0.0208         0.0624
## 3      A1 : A4 0.106          0.212
## 4      A2 : A3 0.166          0.212
## 5      A2 : A4 0.00000268     0.0000161
## 6      A3 : A4 0.0000121      0.0000605
```

多項分布：a × b 分割表 残差分析（補足）

- a×b 分割表の残差分析（カイ 2 乗検定の事後検定）

スクリプトファイル：Green3-3-4e.R

利用した関数：chisq.posthoc.test::chisq.posthoc.test

事例：4 群（A1, A2, A3, A4）ごとに 100 例の結果を 4 種類のパターン（B1, B2, B3, B4）に分類。
群間でパターンに違いがあるか（永田・吉田, 1997）

分割表

有意水準の
調整方法

```
chisq.posthoc.test(tb, method = "none")
```

調整済み標準化残差
p 値

##	Dimension	Value	B1	B2	B3	B4
## 1	A1	Residuals	-1.0636320	-1.481917	1.622214	0.6567984
## 2	A1	p values	0.2874950	0.138362	0.104757	0.5113110
## 3	A2	Residuals	3.7986859	1.058512	-1.622214	-2.4480666
## 4	A2	p values	0.0001450	0.289822	0.104757	0.0143630
## 5	A3	Residuals	0.1519474	3.034402	0.270369	-2.9257381
## 6	A3	p values	0.8792280	0.002410	0.786876	0.0034360
## 7	A4	Residuals	-2.8870013	-2.610997	-0.270369	4.7170064
## 8	A4	p values	0.0038890	0.009028	0.786876	0.0000020

多項分布： $a \times b$ 分割表 ($a \geq 3$ 、 $b \geq 3$)

名義ロジスティック回帰

● 表示3.4.12 JMP [モデルのあてはめ] の出力

スクリプトファイル

Green3-3-4c.R

利用した関数

nnet::multinom

質的変数 group の制約条件は
デフォルトとする

(contr.treatment)

df	説明変数	目的変数	度数
##	group	event	freq
## 1	old	A	2
## 2	old	B	5
## 3	old	C	2
## 4	old	D	1
## 5	new1	A	2
.....			
## 15	new3	C	7
## 16	new3	D	0

```
mn_out <- multinom(event ~ group, weights = freq,  
                    data = df, Hess = TRUE)
```

- 表示3.4.12 JMP [モデルのあてはめ] の出力

スクリプトファイル：Green3-3-4c.R

利用した関数：nnet::multinom、car::Anova

方法

効果の尤度比検定

```
mn_out <- multinom(event ~ group, weights = freq,  
                    data = df, Hess = TRUE)
```

```
Anova(mn_out)
```

```
## # weights: 8 (3 variable)
```

```
## initial value 55.451774
```

```
## final value 51.051671
```

```
## converged
```

```
## Analysis of Deviance Table (Type II tests)
```

```
##
```

```
## Response: event
```

```
##          LR Chisq Df Pr(>Chisq)
```

```
## group    14.418  9    0.1082
```



- 引用文献

永田靖・吉田道弘（1997）統計的多重比較法の基礎
サイエンティスト社

青木繁伸（2009）Rによる統計解析、オーム社

- 作成

片瀬雅彦

- 作成時期

2021年11月20日