

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
3.5 多項分布 (順序尺度)

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 で見ることができます

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）

2 × b 分割表

($b \geq 3$ 、 b が順序尺度)

順序ロジスティック回帰



順序ロジスティック回帰： $2 \times b$ 分割表 ($b \geq 3$)

p.196

● 表示3.5.1 Excelによる基本解析

スクリプトファイル：Green3-3-5a.R

利用した関数：readxl::read_excel

方法：Excel ファイルからデータを読み込

オブジェクト df にデータフレーム

として付値

##	group	event	degree	degree1	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
順序関係がある

二種類の
ダミー変数

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

順序ロジスティック回帰： $2 \times b$ 分割表 ($b \geq 3$)

● 表示3.5.1 Excelによる基本解析

スクリプトファイル：Green3-3-5a.R

利用した関数：factor

方法

factor 関数のlevels 引数とordered 引数で

カテゴリーの順序を指定

Environment タブで確認

先頭が基準

##	group	event	degree	degree1	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

```
df$group <- factor(df$group, levels = c("control", "treated"))
df$degree <- factor(df$degree, levels = c("high", "med", "low", "none"), ordered = TRUE)
```

```
df$group
## [1] control control control control treated treated treated treated
## Levels: control treated
```

順序関係がない

```
df$degree
## [1] none low med high none low med high
## Levels: high < med < low < none
```

順序関係がある

levels で指定した位置を順序とする

順序ロジスティック回帰： $2 \times b$ 分割表 ($b \geq 3$)

● 表示3.5.1 Excelによる基本解析

スクリプトファイル：Green3

利用した関数：factor

方法

factor 関数のlevels 引数とo

カテゴリーの順序を指定

Environment タブで確認

The screenshot shows the R Studio Environment tab for a variable named 'df'. It displays the following structure:

- \$ group : Factor w/ 2 levels "control","treated": 1 1 1...
- \$ event : chr "none" "type1" "type2" "type3" ...
- \$ degree : Ord.factor w/ 4 levels "high"<"med"<"low"<...:...
- \$ degree1 : num 1 2 3 4 1 2 3 4
- \$ freq : num 52 10 5 3 33 8 10 9
- \$ d : num 0 0 0 0 1 1 1 1
- \$ dd : num -1 -1 -1 -1 1 1 1 1

```
df$group <- factor(df$group,  
df$degree <- factor(df$degree,
```

```
df$group  
## [1] control control control control treated treated treated treated  
## Levels: control treated
```

順序関係がない

levels で指定した位置を順序とする

```
df$degree  
## [1] none low med high none low med high  
## Levels: high < med < low < none
```

順序関係がある

順序ロジスティック回帰： $2 \times b$ 分割表 ($b \geq 3$)

● 表示3.5.1 Excelによる基本解析

スクリプトファイル：Green3-3-5a.R

利用した関数：factor、xtabs (前節 [3.5](#) 参照)

cumsum

方法：xtabs 関数でクロス集計、 2×4 分割表

cumsum 関数で累積度数を算出

表示3.5.1 の 1 段目と、2 段目に相当

##	group	event	degree	degree1	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

分割表 (度数) 1 段目

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

tb1

##	degree	high	med	low	none
##	group	high	med	low	none
##	control	3	5	10	52
##	treated	9	10	8	33

列名

行名

2×4
分割表

累積度数 2 段目

```
tb2 <- tb1  
tb2[1, ] <- cumsum(tb1[1, ])  
tb2[2, ] <- cumsum(tb1[2, ])
```

tb2

##	degree	high	med	low	none
##	group	high	med	low	none
##	control	3	8	18	70
##	treated	9	19	27	60

tb1 の属性も移動

順序ロジスティック回帰： $2 \times b$ 分割表 ($b \geq 3$)

● 表示3.5.1 Excelによる基本解析

スクリプトファイル：Green3-3-5a.R

利用した関数：factor、xtabs

方法

表示3.5.1 の3段目の累積割合に相当
4段目のロジットに相当

##	group	event	degree	degree1	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

累積割合 3段目

```
tb3 <- tb2
tb3[1, ] <- tb2[1, ] / sum(tb1[1, ])
tb3[2, ] <- tb2[2, ] / sum(tb1[2, ])
round(tb3, digits = 3)
```

```
##           degree
## group      high  med  low  none
## control 0.043 0.114 0.257 1.000
## treated 0.150 0.317 0.450 1.000
```

対数オッズ (ロジット) 4段目

```
tb4 <- tb3[, 1:3]
tb4 <- log(tb4 / (1 - tb4))
round(tb4, digits = 3)
```

左3列を抽出

ロジットに変換

```
##           degree
## group      high  med  low
## control -3.106 -2.048 -1.061
## treated -1.735 -0.769 -0.201
```

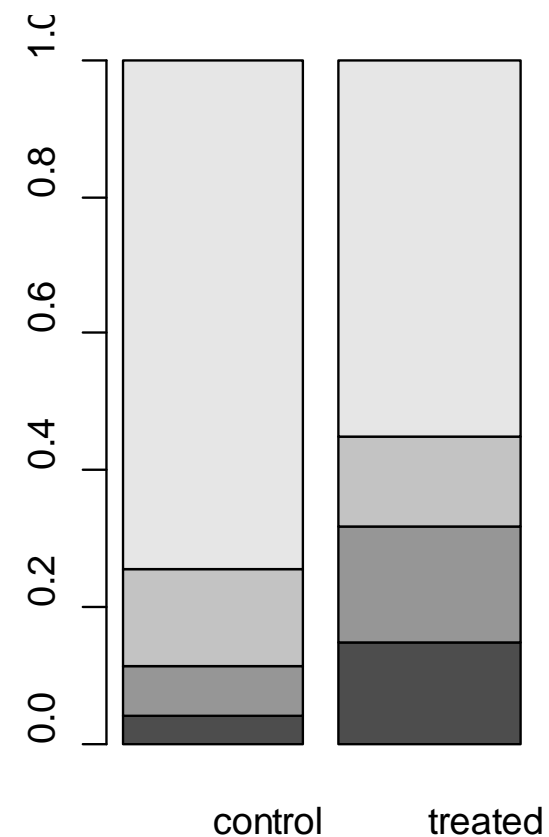
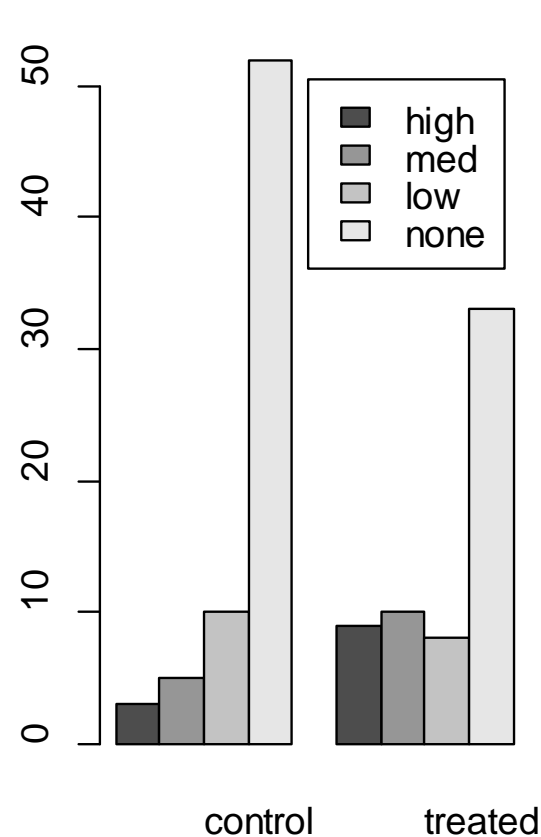
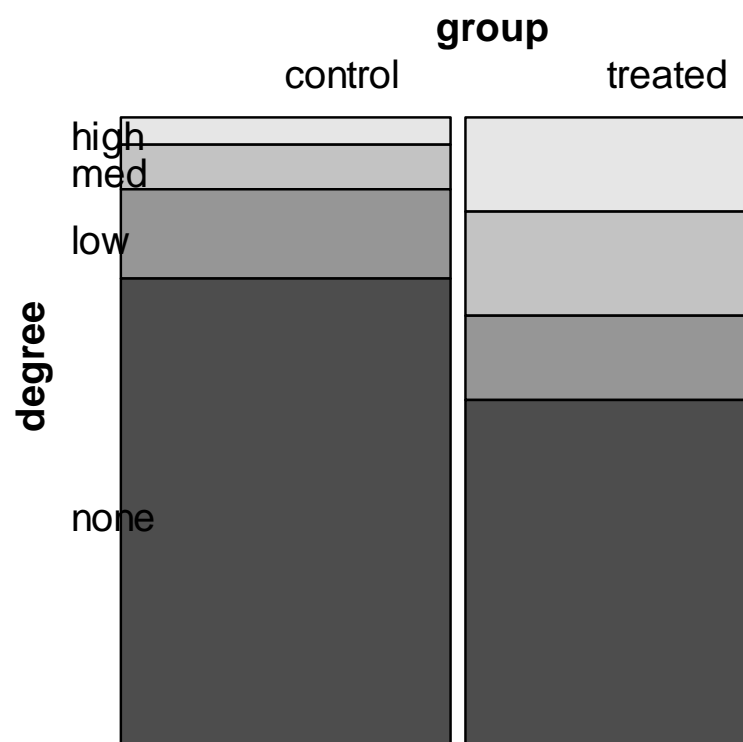
順序ロジスティック回帰： $2 \times b$ 分割表 ($b \geq 3$)

p.197

● 表示3.5.2 基本グラフ

スクリプトファイル：Green3-3-5a.R

利用した関数：vcd::mosaic（次節 [§3.6](#) 参照）、barplot



順序ロジスティック回帰： $2 \times b$ 分割表 ($b \geq 3$)

p.197

● 表示3.5.2 基本グラフ (左、中央)

スクリプトファイル

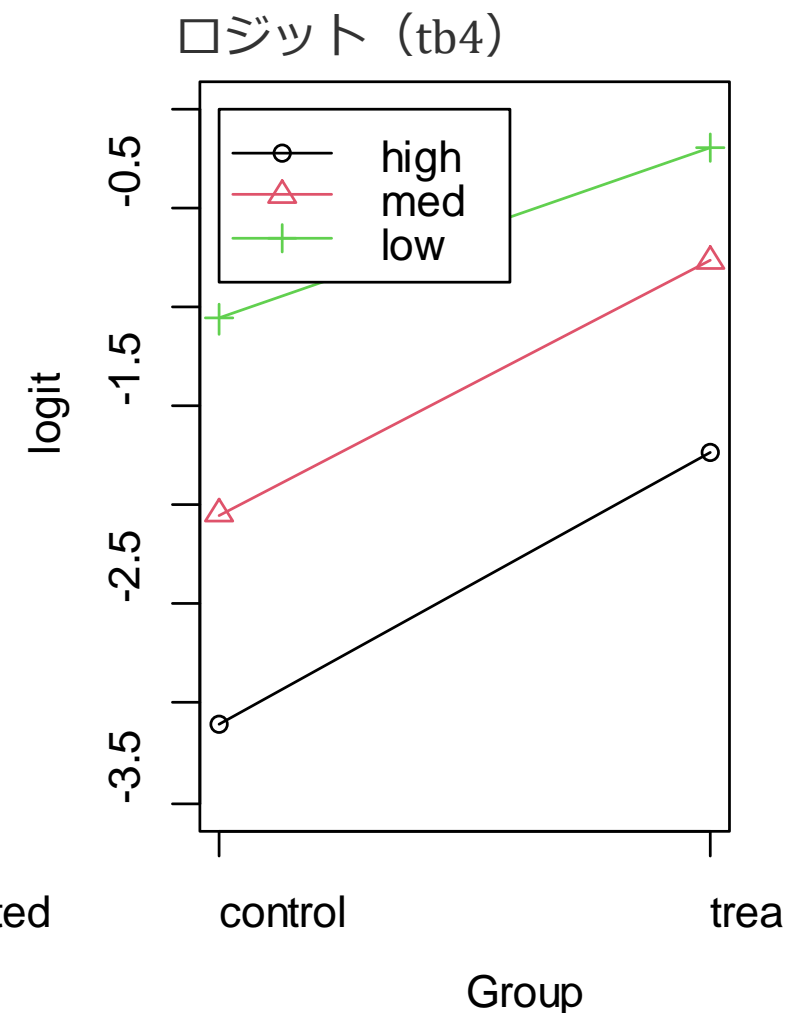
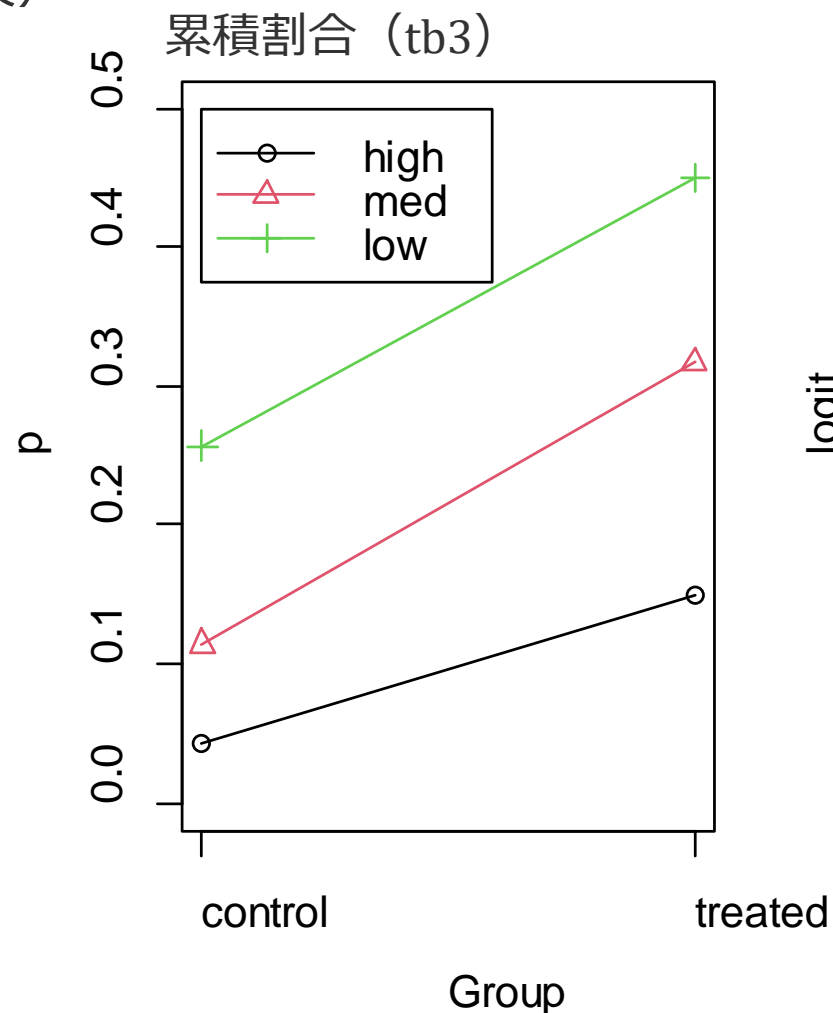
Green3-3-5a.R

利用した関数

plot、axis、segments、
points、legend

方法

累積割合 (tb3) と
ロジット (tb4) の数値
から線分として表示



順序ロジスティック回帰： $2 \times b$ 分割表 ($b \geq 3$)

- 表示3.5.3 JMP[モデルのあてはめ]の出力、表示3.5.4 Excelソルバーによる解析

スクリプトファイル：Green3-3-5a.R

利用した関数：MASS::polr、options、summary、car::Anova

方法：質的因子 group の contrasts を変えずにデフォルトで実行 ("contr.treatment")

テキスト (JMP)
と一致しない

```
df$group <- factor(df$group, levels = c("control", "treated"))
polr_out <- polr(degree ~ group, weights = freq, data = df,
                 method = "logistic", Hess = TRUE)
summary(polr_out)
## Coefficients:
##                Value Std. Error t value
## group.treated -0.9644    0.3687  -2.616
##
## Intercepts:
##                Value Std. Error t value
## high|med -2.8173    0.3785  -7.4433
## med|low  -1.8346    0.3024  -6.0664
## low|none -1.1000    0.2703  -4.0699
```

```
contrasts(df$group)
##                treated
## control          0
## treated          1
```

順序ロジスティック回帰： $2 \times b$ 分割表 ($b \geq 3$)

p.198

- 表示3.5.3 JMP[モデルのあてはめ]の出力、表示3.5.4 Excelソルバーによる解析

スクリプトファイル：Green3-3-5a.R

利用した関数：MASS::polr、options、summary、car::Anova、contrasts

方法：contrasts を contra.sum に設定

```
df1 <- df
```

```
levels(df1$group) <- c("treated", "control")
```

```
contrasts(df1$group) <- "contr.sum"
```

```
polr_out1 <- polr(degree ~ group, weights = freq, data = df1,  
method = "logistic", Hess = TRUE)
```

```
contrasts(df1$group)
```

```
##      [,1]
```

```
## treated    1
```

```
## control   -1
```

この順序で1、-1に
設定され処理される

levels で順序を入れ替え

contrasts を設定

glm のように、polr の引数で
contrasts を指定できるが
エラーが発生するため行わない



順序ロジスティック回帰： $2 \times b$ 分割表 ($b \geq 3$)

- 表示3.5.3 JMP[モデルのあてはめ]の出力、表示3.5.4 Excelソルバーによる解析

スクリプトファイル：Green3-3-5a.R

利用した関数：MASS::polr、options、summary、car::Anova

方法：質的因子 group の制約条件を contra.sum に変更

テキスト (JMP) と

絶対値は一致

正負の符号は逆

```
polr_out1 <- polr(degree ~ group, weights = freq, data = df1,
                  method = "logistic", Hess = TRUE)
summary(polr_out1)

## Coefficients:
##           Value Std. Error t value
## group1 -0.4822    0.1843  -2.616
##
## Intercepts:
##           Value Std. Error t value
## high|med -2.3351    0.3078  -7.5853
## med|low  -1.3525    0.2209  -6.1213
## low|none -0.6179    0.1890  -3.2693
```

JMP と一致

正負の符号が不一致
絶対値は一致



順序ロジスティック回帰： $2 \times b$ 分割表 ($b \geq 3$)

●順序ロジスティック回帰モデル（比例オッズモデル）

切片はカテゴリごとに異なるが、傾きは一定というモデル
係数の解釈は、ソフトウェアにより異なる

JMP のモデル
$$\text{logit}(P(Y \leq j)) = \beta_{j0} + \beta_1 x_1 + \dots + \beta_p x_p$$

R (polr 関数) のモデル

$$\text{logit}(P(Y \leq j)) = \beta_{j0} - \eta_1 x_1 - \dots - \eta_p x_p$$
$$\eta_i = -\beta_i$$

参照 芳賀 (2016) テキスト
藤井 (2010)

UCLA: Statistical Consulting Group.

JMP の結果が°正であれば
polr 関数の結果は負



順序ロジスティック回帰： $2 \times b$ 分割表 ($b \geq 3$)

- 表示3.5.3 JMP[モデルのあてはめ]の出力、表示3.5.4 Excelソルバーによる解析

スクリプトファイル：Green3-3-5a.R

利用した関数：MASS::polr、options、summary、car::Anova

方法：制約条件を変更 (contra.sum)

テキスト (JMP) と
一致

```
polr_out1 <- polr(degree ~ group, weights = freq, data = df1,
                  method = "logistic", Hess = TRUE)
summary(polr_out1)

## Coefficients:
##              Value Std. Error t value
## group1 -0.4822      0.1843  -2.616
##
## Intercepts:
##              Value Std. Error t value
## high|med -2.3351    0.3078  -7.5853
## med|low  -1.3525    0.2209  -6.1213
## low|none -0.6179    0.1890  -3.2693
```

JMP と一致

JMP と一致
ただし符号の表記は逆

順序ロジスティック回帰： $2 \times b$ 分割表 ($b \geq 3$)

- 表示3.5.3、表示3.5.4

スクリプトファイル：Green3-3-5a.R

利用した関数：summary、texreg::screenreg

```
summary(polr_out1)

## Coefficients:
##           Value Std. Error t value
## group1 -0.4822    0.1843  -2.616
##
## Intercepts:
##           Value Std. Error t value
## high|med -2.3351    0.3078  -7.5853
## med|low  -1.3525    0.2209  -6.1213
## low|none -0.6179    0.1890  -3.2693
##
## Residual Deviance: 258.3612
## AIC: 266.3612
```

```
screenreg(polr_out1)
## =====
##                               Model 1
## -----
## group1                        -0.48 **
##                               (0.18)
## high|med                       -2.34 ***
##                               (0.31)
## med|low                        -1.35 ***
##                               (0.22)
## low|none                       -0.62 **
##                               (0.19)
## -----
## AIC                            266.36
## BIC                            277.83
## Log Likelihood                 -129.18
## Deviance                       258.36
## Num. obs.                      130
## =====
```

順序ロジスティック回帰： $2 \times b$ 分割表 ($b \geq 3$)

p.198

- 表示3.5.3 JMP[モデルのあてはめ]の出力、表示3.5.4 Excelソルバーによる解析

スクリプトファイル

Green3-3-5a.R

利用した関数

MASS::polr、car::Anova、anova

方法

効果の尤度比検定

nullモデルとの比較 `polr_out0 <- polr(degree ~ 1, weights = freq, data = df,`

`method = "logistic")`

`anova(polr_out, polr_out0)`

Likelihood ratio tests of ordinal regression models

##

Response: degree

##	Model	Resid. df	Resid. Dev	Test	Df	LR stat.	Pr(Chi)
----	-------	-----------	------------	------	----	----------	---------

##	1	1	127	265.3756			
----	---	---	-----	----------	--	--	--

##	2	group	126	258.3612	1 vs 2	1 7.014367	0.008085822
----	---	-------	-----	----------	--------	------------	-------------

```
Anova(polr_out, type = 2)
```

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

```
##
```

```
## Response: degree
```

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

```
## group 7.0144 1 0.008086 **
```

nullモデル

順序ロジスティック回帰： $2 \times b$ 分割表 ($b \geq 3$)

- 表示3.5.3 JMP[モデルのあてはめ]の出力

表示3.5.4 Excelソルバーによる解析

スクリプトファイル：Green3-3-5a.R

利用した関数：MASS::polr、sumamry、texreg

方法：説明変数に、質的因子の代わりに

dd (ダミー変数2) を使う

(第2部 [§2.3](#) 参照)

##	group	event	degree	degree1	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

```
df2$group <- factor(df2$group, levels = c("treated", "control"))
contrasts(df2$group) <- "contr.sum"
polr_out <- polr(degree ~ group, weights = freq, data = df2, method = "logistic")
polr_out1 <- polr(degree ~ dd, weights = freq, data = df2, method = "logistic")
```

制約条件

末尾が-1

質的因子

ダミー変数2

順序ロジスティック回帰： $2 \times b$ 分割表 ($b \geq 3$)

p.198

- 表示3.5.3 JMP[モデルのあてはめ]の出力、表示3.5.4 Excelソルバーによる解析
スクリプトファイル：Green3-3-5a.R、利用した関数：MASS::polr、nnet::multinom
方法：あてはまりの悪さ (LOF) を、名義ロジスティック回帰との差から算出、重要

```
a <- 2; b <- 4 # axb 分割表
dof <- (a - 1) * (b - 1) - (a - 1)

logL1 <- polr_out$deviance

mn_out <- multinom(degree ~ group, weights = freq, data = df)
logL2 <- mn_out$deviance

chi2 <- logL1 - logL2
p_lof <- min(pchisq(chi2, dof), 1 - pchisq(chi2, dof))

cbind("LOF:dof" = dof, "chisq" = chi2, "p-value" = p_lof)

##      LOF:dof      chisq  p-value
## [1,]         2 1.559952 0.458417
```

LOF の自由度

-2×対数尤度の計
L = 258.376

順序を無視して
名義ロジスティック回帰
(前節 §2.3 参照)

-2×対数尤度の計
L = 256.801

LOF の
カイ2乗値
1.560

LOF の p 値

順序ロジスティック回帰： $2 \times b$ 分割表 ($b \geq 3$)

p.198

● 表示3.5.2 基本グラフ (右)

スクリプトファイル：Green3-3-5a.R

利用した関数：plot、axis、segments、points

方法：パラメータ推定値からロジットの予測値を計算、描画

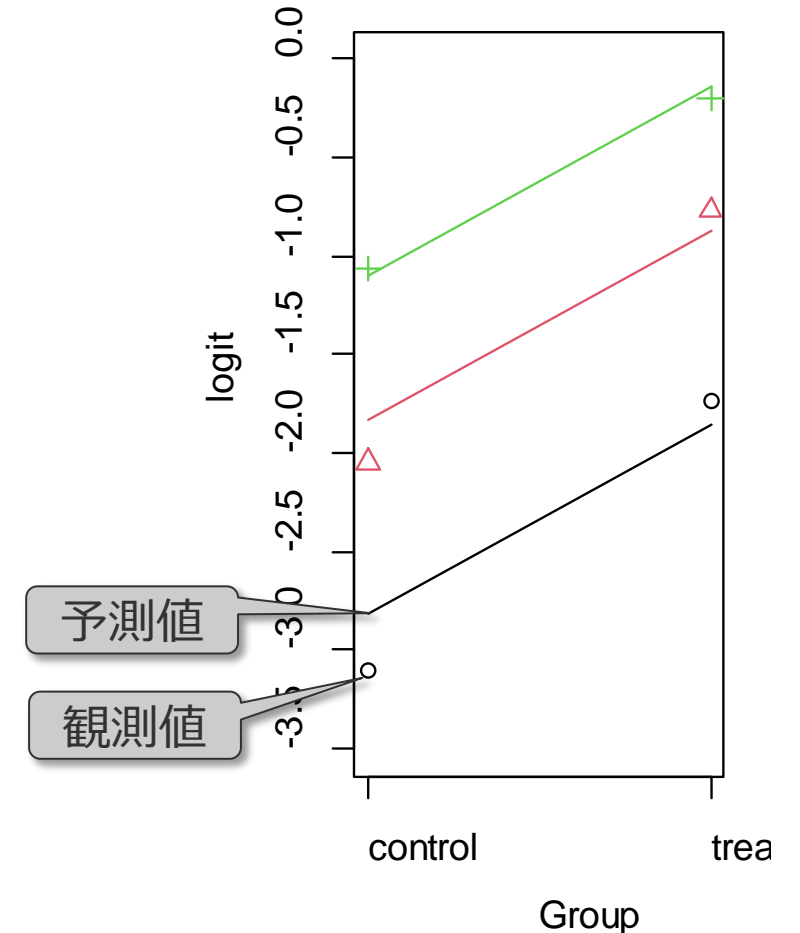
```
a <- polr_out1$zeta
b <- polr_out1$coefficients
x0 <- -1
x1 <- 1
y0 <- a - b * x0
y1 <- a - b * x1

plot(NA, xlim = c(-1, 1), ylim = c(-3.5, 0.0),
     xaxt = "n", ylab = "logit", xlab = "Group")
axis(1, at = c(-1, 1), labels = c("control", "treated"))
segments(x0, y0, x1, y1, col = 1:3)
points(c(-1, -1, -1), tb4[1, ], pch = 1:3, col = 1:3)
points(c(1, 1, 1), tb4[2, ], pch = 1:3, col = 1:3)
```

切片、要素数 3

傾き、要素数 1

-b であることに注意
polr 関数のモデル式





2 × b 分割表 (補足) ($b \geq 3$ 、bが順序尺度)

Wilcoxon の順位和検定による
2 群の単純比較

順序カテゴリー分類データの二群の単純比較（補足）

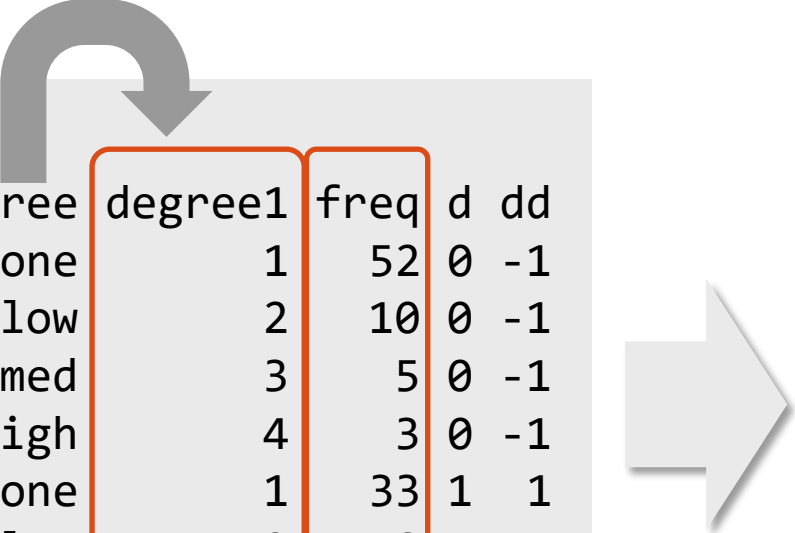
- 表示3.5.1 Excelによる基本解析

スクリプトファイル：Green3-3-5c.R

利用した関数：coin::wilcox_test

方法：Wilcoxon の順位和検定で全体の差を検定（丹後, 1998）

度数のあるデータフレームから度数の列をなくす（[§3.4](#)）



df	group	event	degree	degree1	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

df1	degree1	group
## 1	1	control
## 2	1	control
## 3	1	control
.....		
## 65	3	control
## 66	3	control
## 67	3	control
## 68	4	control
## 69	4	control
## 70	4	control
## 71	1	treated
## 72	1	treated
.....		
## 128	4	treated
## 129	4	treated
## 130	4	treated



順序カテゴリー分類データの二群の単純比較（補足）

● 表示3.5.2 基本グラフ（右）

スクリプトファイル：Green3-3-5c.R

利用した関数：coin::wilcox_test

方法：Wilcoxon の順位和検定で全体の差を検定（丹後, 1998）

```
wilcox_test(degree1 ~ factor(group), data = df1,  
            alternative = "two.sided",  
            ties.method = "mid-ranks",  
            distribution = "asymptotic")  
  
##  
## Asymptotic Wilcoxon-Mann-Whitney Test  
##  
## data: degree1 by factor(group) (control, treated)  
## Z = -2.6252, p-value = 0.008659  
## alternative hypothesis: true mu is not equal to 0
```

```
df1  
##      degree1  group  
## 1          1 control  
## 2          1 control  
## 3          1 control  
## .....  
## 65         3 control  
## 66         3 control  
## 67         3 control  
## 68         4 control  
## 69         4 control  
## 70         4 control  
## 71         1 treated  
## 72         1 treated  
## .....  
## 128        4 treated  
## 129        4 treated  
## 130        4 treated
```


順序カテゴリー分類データの二群の単純比較（補足）

●表20（丹下,1998）

スクリプトファイル：Green3-3-5c.R

利用した関数：coin::wilcox_test、chisq.test

方法：Wilcoxon の順位和検定で全体の差を検定
どこの割合の差が大きいか探索するため、
段階的に 2×2 分割表に組みなおして、
カイ 2 乗検定を繰り返す（丹後, 1998）

全体を
 2×2 分割表に
組みなおして
カイ 2 乗検定

```
wilcox_test(y ~ factor(x), data = df3,  
            alternative = "two.sided",  
            ties.method = "mid-ranks",  
            distribution = "exact")
```

```
##  
## Exact Wilcoxon-Mann-Whitney Test  
##  
## data: y by factor(x) (g1, g2)  
## Z = -2.9009, p-value = 0.003583  
## alternative hypothesis: true mu is  
## not equal to 0
```

```
##      [1,1] [2,1] [1,2] [2,2]      chi2      p.value  
## [1,]      6      4      97      98 0.4002597 0.526955177  
## [2,]     28     13      75      89 6.6782077 0.009759871  
## [3,]     54     33      49      69 8.4541191 0.003642176  
## [4,]     86     77      17      25 2.0159120 0.155657719  
## [5,]    100     96       3       6 1.0767802 0.299418959
```



$a \times b$ 分割表

($a \geq 3$ 、 $b \geq 3$ 、 b が順序尺度)

順序ロジスティック回帰

順序ロジスティック回帰：a×b 分割表 (a≥3、b ≥ 3)

● 表示3.5.5 Excelによる基本解析

スクリプトファイル：Green3-3-5b.R

利用した関数：readxl::read_excel、factor

方法：Excel ファイルからデータを読み込、因子型に指定、
カテゴリーの位置を指定（後述）、順序を指定

```
df <- read_excel("Green3-3.xlsx", sheet = "35-4x4")  
df <- data.frame(df)
```

```
df$group <- factor(df$group,  
  levels = c("old", "new1", "new2", "new3"))
```

```
df$degree <- factor(df$degree,  
  levels = c("high", "med", "low", "none"),  
  ordered = TRUE)
```

4つのカテゴリーに
順序関係がある

順序関係がない

順序関係がある

##	group	degree	freq	d1	...
## 1	old	none	2	0	...
## 2	old	low	5	0	...
## 3	old	med	2	0	...
## 4	old	high	1	0	...
## 5	new1	none	2	1	...
## 6	new1	low	4	1	...
## 7	new1	med	2	1	...
## 8	new1	high	2	1	...
## 9	new2	none	0	0	...
## 10	new2	low	2	0	...
## 11	new2	med	5	0	...
## 12	new2	high	3	0	...
## 13	new3	none	1	0	...
## 14	new3	low	2	0	...
## 15	new3	med	7	0	...
## 16	new3	high	0	0	...

順序ロジスティック回帰：a×b 分割表 (a≥3、b ≥ 3)

● 表示3.5.5 Excelによる基本解析

スクリプトファイル：Green3-3-5b.R

利用した関数：xtabs

方法：クロス集計による分割表の作成 (度数の計算)

4 × 4 分割表

表示3.3.5 1 段目の度数

```
# 分割表 (度数)
tb1 <- xtabs(freq ~ group + degree, data = df)
tb1
```

##	degree				
##	group	high	med	low	none
##	old	1	2	5	2
##	new1	2	2	4	2
##	new2	3	5	2	0
##	new3	0	7	2	1

列名

行名

```
##      group degree freq d1 ...
## 1     old   none    2  0 ...
## 2     old    low    5  0 ...
## 3     old    med    2  0 ...
## 4     old   high    1  0 ...
## 5   new1   none    2  1 ...
## 6   new1    low    4  1 ...
## 7   new1    med    2  1 ...
## 8   new1   high    2  1 ...
## 9   new2   none    0  0 ...
## 10  new2    low    2  0 ...
## 11  new2    med    5  0 ...
## 12  new2   high    3  0 ...
## 13  new3   none    1  0 ...
## 14  new3    low    2  0 ...
## 15  new3    med    7  0 ...
## 16  new3   high    0  0 ...
```

順序ロジスティック回帰：a×b 分割表 (a≥3、b ≥ 3)

● 表示3.5.5 Excelによる基本解析

スクリプトファイル：Green3-3-5b.R

利用した関数：cumsum、for

方法：表示3.5.5 2段目の累積度数

tb1 の属性も移動

```
# 累積度数
tb2 <- tb1
for (i in 1:4) tb2[i, ] <- cumsum(tb1[i, ])
tb2
```

```
##          degree
## group  high med low none
##   old    1  3  8  10
##  new1    2  4  8  10
##  new2    3  8 10  10
##  new3    0  7  9  10
```

ベクトルの
累積度数を計算

```
##      group degree freq d1 ...
## 1     old   none    2  0 ...
## 2     old    low    5  0 ...
## 3     old    med    2  0 ...
## 4     old   high    1  0 ...
## 5   new1   none    2  1 ...
## 6   new1    low    4  1 ...
## 7   new1    med    2  1 ...
## 8   new1   high    2  1 ...
## 9   new2   none    0  0 ...
## 10  new2    low    2  0 ...
## 11  new2    med    5  0 ...
## 12  new2   high    3  0 ...
## 13  new3   none    1  0 ...
## 14  new3    low    2  0 ...
## 15  new3    med    7  0 ...
## 16  new3   high    0  0 ...
```

順序ロジスティック回帰：a×b 分割表 (a≥3、b ≥ 3)

● 表示3.5.5 Excelによる基本解析

スクリプトファイル：Green3-3-5b.R

利用した関数：sum、for

方法：表示3.5.5 3段目の累積割合

ポイント

累積割合

```
tb3 <- tb2
```

```
for (i in 1:4) tb3[i, ] <- tb2[i, ] / sum(tb1[i, ])
```

```
tb3
```

```
##          degree
## group  high med low none
##  old   0.1 0.3 0.8  1.0
## new1  0.2 0.4 0.8  1.0
## new2  0.3 0.8 1.0  1.0
## new3  0.0 0.7 0.9  1.0
```

累積度数

合計

```
##      group degree freq d1 ...
## 1    old   none    2  0 ...
## 2    old   low    5  0 ...
## 3    old   med    2  0 ...
## 4    old   high   1  0 ...
## 5  new1  none    2  1 ...
## 6  new1  low    4  1 ...
## 7  new1  med    2  1 ...
## 8  new1  high   2  1 ...
## 9  new2  none    0  0 ...
## 10 new2  low    2  0 ...
## 11 new2  med    5  0 ...
## 12 new2  high   3  0 ...
## 13 new3  none    1  0 ...
## 14 new3  low    2  0 ...
## 15 new3  med    7  0 ...
## 16 new3  high   0  0 ...
```

順序ロジスティック回帰：a×b 分割表 (a≥3、b ≥ 3)

● 表示3.5.5 Excelによる基本解析

スクリプトファイル：Green3-3-5b.R

利用した関数：log、round

方法：表示3.5.5 4段目の対数オッズ（ロジット）

```
# 対数オッズ
tb4 <- tb3[, 1:3]
tb4 <- log(tb4 / (1 - tb4))
tb4[4, 1] <- -4.0
tb4[3, 3] <- 4.0
round(tb4, digits = 3)
```

左4列

ロジットに変換

仮の値

仮の値

##	degree			
##	group	high	med	low
##	old	-2.197	-0.847	1.386
##	new1	-1.386	-0.405	1.386
##	new2	-0.847	1.386	4.000
##	new3	-4.000	0.847	2.197

##	group	degree	freq	d1	...
##	1	old	none	2	0 ...
##	2	old	low	5	0 ...
##	3	old	med	2	0 ...
##	4	old	high	1	0 ...
##	5	new1	none	2	1 ...
##	6	new1	low	4	1 ...
##	7	new1	med	2	1 ...
##	8	new1	high	2	1 ...
##	9	new2	none	0	0 ...
##	10	new2	low	2	0 ...
##	11	new2	med	5	0 ...
##	12	new2	high	3	0 ...
##	13	new3	none	1	0 ...
##	14	new3	low	2	0 ...
##	15	new3	med	7	0 ...
##	16	new3	high	0	0 ...

● 表示3.5.6 基本グラフ

スクリプトファイル

Green3-3-5b.R

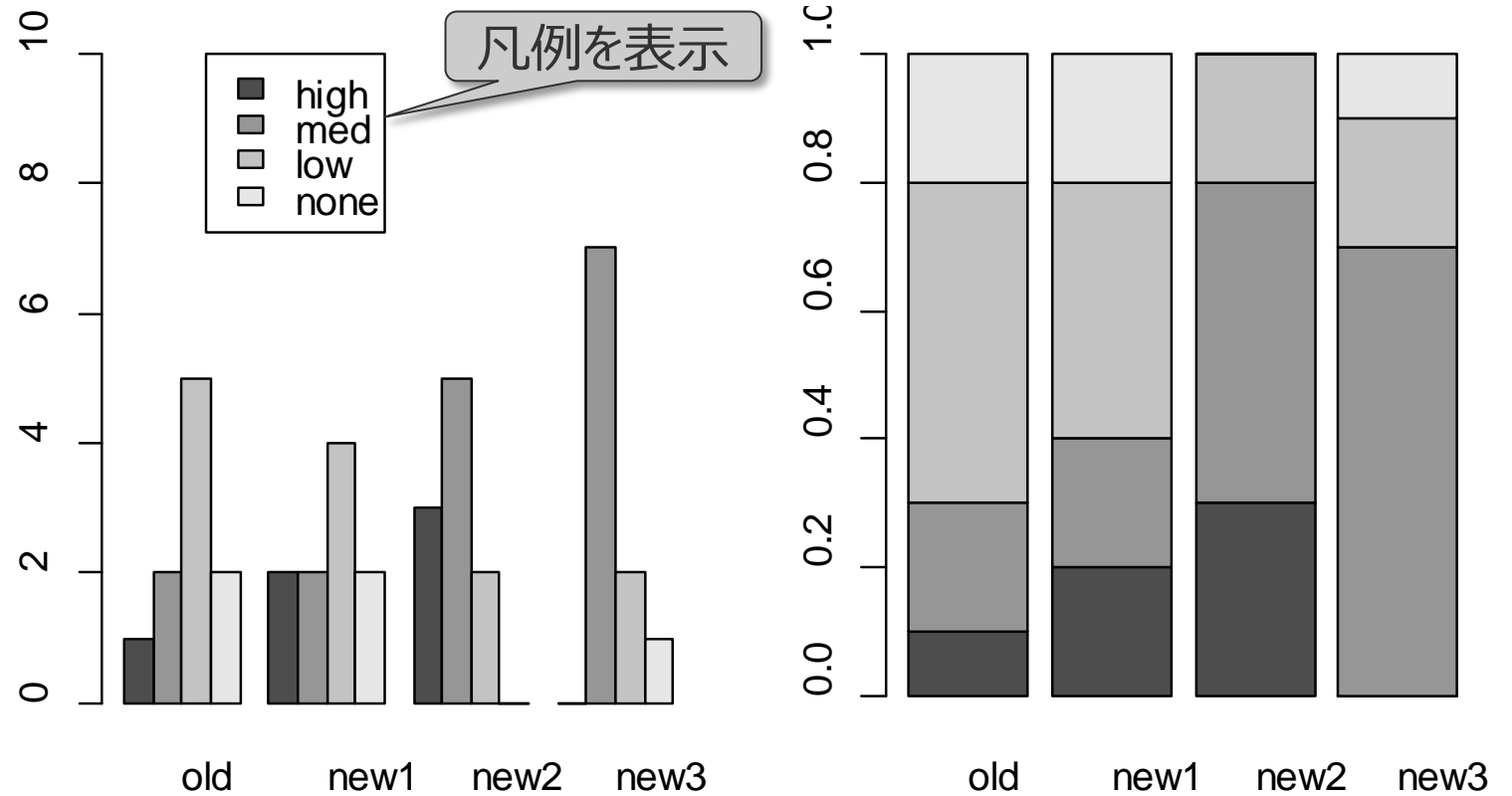
利用した関数

barplot、t、proportions

方法

分割表 (tb1) から棒グラフ
を作成

tb1 を割合に変換した tb11
から積み上げ棒グラフを
作成



凡例を表示

凡例の設定

```
barplot(t(tb1), beside = TRUE, ylim = c(0, 10), legend = TRUE,  
        args.legend = list(x = 10, y = 10, cex = 0.8))  
tb11 <- proportions(tb1, 1)  
barplot(t(tb11))
```

凡例を表示

● 表示3.5.6 基本グラフ

スクリプトファイル

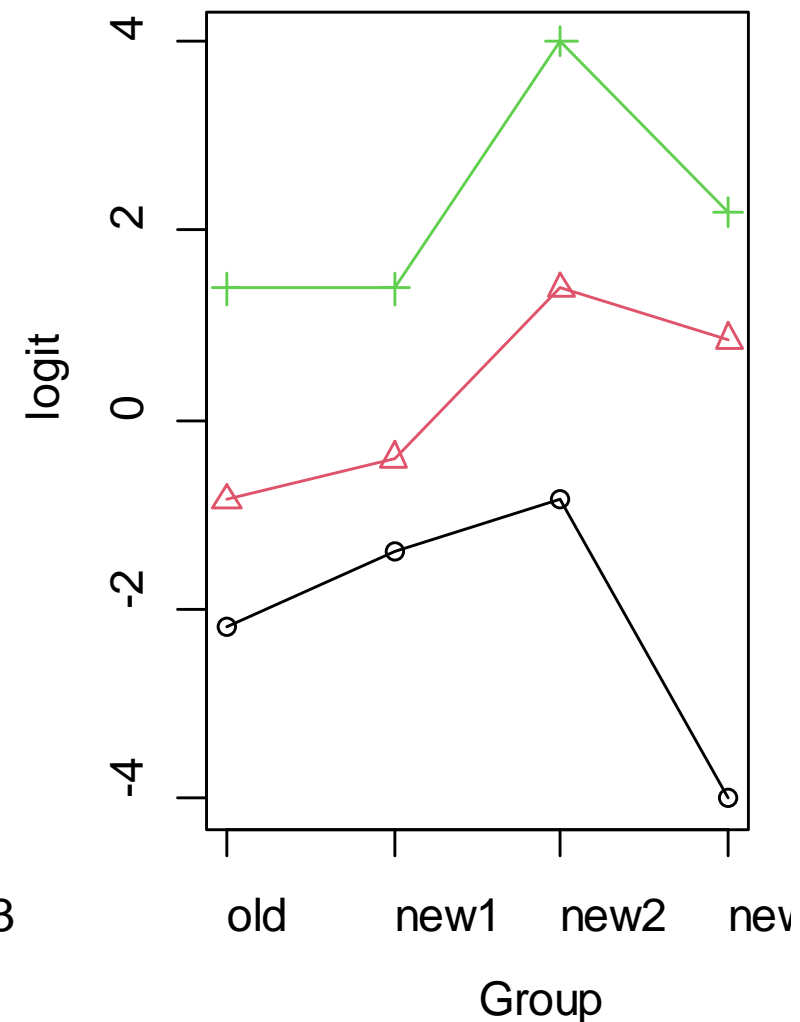
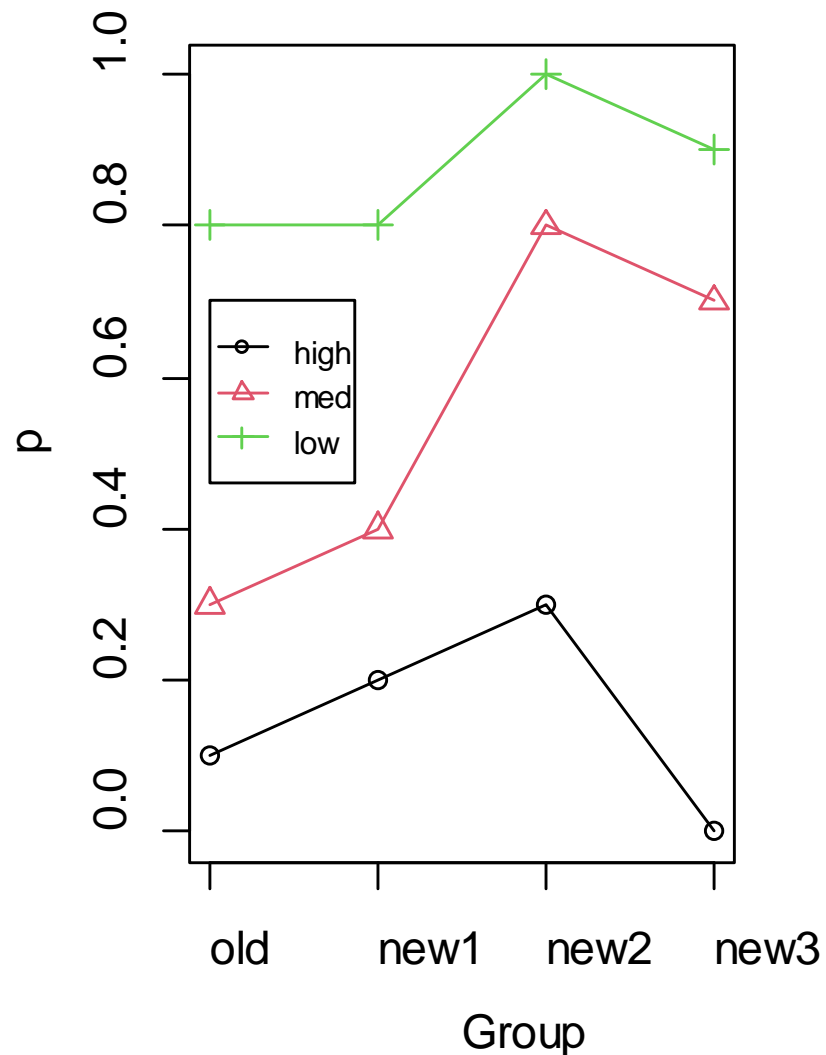
Green3-3-5b.R

利用した関数

plot、segments、lines、
points、axis、legend、
par、for

方法

累積割合の表 (tb3) と
ロジットの表 (tb4) の
データをプロット



順序ロジスティック回帰： $a \times b$ 分割表 ($a \geq 3$ 、 $b \geq 3$)

- 表示3.5.7 JMP[モデルのあてはめ]の出力、表示3.5.8 Excel ソルバーの解析

スクリプトファイル：Green3-3-5b.R

利用した関数：MASS::polr、summary、texreg::screenreg、car::Anova、contrasts

方法：質的因子 group の contrasts を
デフォルトで実行

("contr.treatment"、

ダミー変数 1 で処理される)

テキスト (JMP) と不一致

```
contrasts(df$group)
```

```
##      new1 new2 new3
## old    0    0    0
## new1   1    0    0
## new2   0    1    0
## new3   0    0    1
```

```
polr_out <- polr(degree ~ group, weights = freq,
                 data = df, method = "logistic", Hess = TRUE)
summary(polr_out)
```

```
## Coefficients:
##              Value Std. Error t value
```

"old"がダミー変数 1 の基準 (前節 [§3.4](#)、第2部 [§2.3](#))

```
df$group <- factor(df$group,
                   levels = c("old", "new1", "new2", "new3"))
先頭
```

```
##              Value Std. Error t value
## high|med -2.8443  0.7702  -3.6927
## med|low  -0.6697  0.6204  -1.0795
## low|none  1.2703  0.6537   1.9434
```

順序ロジスティック回帰： $a \times b$ 分割表 ($a \geq 3$ 、 $b \geq 3$)

p.201

- 表示3.5.7 JMP[モデルのあてはめ]の出力、表示3.5.8 Excel ソルバーの解析

スクリプトファイル：Green3-3-5b.R

利用した関数：MASS::polr、summary、texreg::screenreg、car::Anova

方法：質的因子 group の contrasts を
デフォルトで実行

("contr.treatment",
ダミー変数 1 で処理される)

テキスト (JMP) と不一致

```
contrasts(df$group)
```

```
##      new1 new2 new3
## old      0    0    0
## new1     1    0    0
## new2     0    1    0
## new3     0    0    1
```

```
polr_out <- polr(degree ~ group, weights = freq,
                 data = df, method = "logistic", Hess = TRUE)
```

```
summary(polr_out)
```

```
## Coefficients:
```

```
##              Value Std. Error t value
## groupnew1 -0.4128      0.8575 -0.4814
## groupnew2 -2.0453      0.8874 -2.3047
## groupnew3 -1.0270      0.8268 -1.2422
```

```
##
```

```
## Intercepts:
```

```
##              Value Std. Error t value
## high|med -2.8443      0.7702 -3.6927
## med|low  -0.6697      0.6204 -1.0795
## low|none  1.2703      0.6537  1.9434
```

順序ロジスティック回帰： $a \times b$ 分割表 ($a \geq 3$ 、 $b \geq 3$)

p.201

- 表示3.5.7 JMP[モデルのあてはめ]の出力、表示3.5.8 Excel ソルバーの解析

スクリプトファイル：Green3-3-5b.R

利用した関数：factor、contrasts、MASS::polr、summary、texreg::screenreg、car::Anova

方法：質的因子 group の contrasts を

contr.sum に変更して実行

(ダミー変数 2)

テキスト (JMP) と一致

```
contrasts(df1$group)
##      [,1] [,2] [,3]
## new1    1    0    0
## new2    0    1    0
## new3    0    0    1
## old     -1   -1   -1
```

新しいデータフレームを準備

```
df1 <- df
```

基準のカテゴリーを末尾

```
df1$group <- factor(df1$group,
                    levels = c("new1", "new2", "new3", "old"))
```

```
contrasts(df1$group) <- "contr.sum"  contr.sum に変更
```

```
polr_out1 <- polr(degree ~ group, weights = freq,
                  data = df1,
                  method = "logistic", Hess = TRUE)
```

順序ロジスティック回帰： $a \times b$ 分割表 ($a \geq 3$ 、 $b \geq 3$)

- 表示3.5.7 JMP[モデルのあてはめ]の出力、表示3.5.8 Excel ソルバーの解析

スクリプトファイル：Green3-3-5b.R

利用した関数：MASS::polr、summary

方法：質的因子 group の contrasts を
contr.sum に変更して実行

(ダミー変数 2)

テキスト (JMP) と一致

JMP と一致
ただし符号は
逆に表示

JMP と一致

```
polr_out1 <- polr(degree ~ group, weights = freq,
                  data = df1,
                  method = "logistic", Hess = TRUE)
summary(polr_out1)

## Coefficients:
##           Value Std. Error t value
## group1  0.4585    0.5369    0.854
## group2 -1.1742    0.5407   -2.172
## group3 -0.1559    0.4933   -0.316
##
## Intercepts:
##           Value Std. Error t value
## high|med -1.9731    0.4793   -4.1164
## med|low  0.2016    0.3380    0.5964
## low|none 2.1416    0.5010    4.2748
```

- 表示3.5.7 JMP[モデルのあてはめ]の出力、表示3.5.8 Excel ソルバーの解析

スクリプトファイル：Green3-3-5b.R

利用した関数：MASS::polr、summary、texreg::screenreg
car::Anova

方法：質的因子 group の制約条件を contr.sum に変更
(ダミー変数 2 で処理される)

テキスト (JMP) と一致

```
Anova(polr_out, type = 3)
## Analysis of Deviance Table (Type III tests)
##
## Response: degree
##      LR Chisq Df Pr(>Chisq)
## group  6.3015  3  0.09783 .
```

```
screenreg(polr_out1)
## =====
##                               Model 1
## -----
## group1                        0.46
##                               (0.54)
## group2                        -1.17 *
##                               (0.54)
## group3                        -0.16
##                               (0.49)
## high|med                      -1.97 ***
##                               (0.48)
## med|low                        0.20
##                               (0.34)
## low|none                       2.14 ***
##                               (0.50)
## =====
```

順序ロジスティック回帰：a×b 分割表 (a≥3、b ≥ 3)

- 表示3.5.7 JMP[モデルのあてはめ]の出力、表示3.5.8 Excel ソルバーの解析

スクリプトファイル：Green3-3-5b.R

利用した関数：MASS::polr

方法：ダミー変数2を説明変数に使った解析

group の contrasts を"contr.sum"に設定した
結果と一致

```
polr_out3 <- polr(degree ~ dd1 + dd2 + dd3 ,  
  weights = freq, data = df,  
  method = "logistic")  
summary(polr_out3)
```

```
## Coefficients:  
##      Value Std. Error t value  
## dd1  0.4585    0.5369   0.854  
## dd2 -1.1742    0.5407  -2.172  
## dd3 -0.1559    0.4933  -0.316  
.....以下省略
```

ダミー変数2

ダミー変数2

##	group	degree	freq	dd1	dd2	dd3
## 1	old	none	2	-1	-1	-1
## 2	old	low	5	-1	-1	-1
## 3	old	med	2	-1	-1	-1
## 4	old	high	1	-1	-1	-1
## 5	new1	none	2	1	0	0
## 6	new1	low	4	1	0	0
## 7	new1	med	2	1	0	0
## 8	new1	high	2	1	0	0
## 9	new2	none	0	0	1	0
## 10	new2	low	2	0	1	0
## 11	new2	med	5	0	1	0
## 12	new2	high	3	0	1	0
## 13	new3	none	1	0	0	1
## 14	new3	low	2	0	0	1
## 15	new3	med	7	0	0	1
## 16	new3	high	0	0	0	1

順序ロジスティック回帰： $a \times b$ 分割表 ($a \geq 3$ 、 $b \geq 3$)

p.203

- 表示3.5.7 JMP[モデルのあてはめ]の出力、表示3.5.8 Excel ソルバーの解析

スクリプトファイル：Green3-3-5b.R

利用した関数：MASS::polr、nnet::multinom、pchisq

方法：あてはまりの悪さ

(LOF) を算出

順序ロジスティック回帰の結果と、名義ロジスティックとみなして解析した結果と比較

-2 x 対数尤度の差
(カイ2乗値)

$95.802 - 87.685 = 8.117$

```
a <- 4 # axb 分割表
b <- 4
dof <- (a - 1) * (b - 1) - (a - 1)
logL1 <- polr_out$deviance
mn_out <- multinom(degree ~ group,
                    weights = freq, data = df)
logL2 <- mn_out$deviance
# カイ2乗検定
chi2 <- logL1 - logL2
p_lof <- min(pchisq(chi2, dof), 1 - pchisq(chi2, dof))
cbind("LOF:dof" = dof, "chisq" = chi2, "p-value" = p_lof)

##      LOF:dof      chisq      p-value
## [1,]      6 8.116428 0.229697
```

LOF の自由度

表示3.5.8 の $b_0 \sim b_3$ の
変化させるセルの数の差 6

順序ロジスティック解析の
-2x対数尤度 95.802

名義ロジスティック解析の
-2x対数尤度 87.685

順序ロジスティック回帰：a×b 分割表 (a≥3、b ≥ 3)

- 表示3.8.6 JMP[モデルのあてはめ]の出力、表示3.8.7 Excel ソルバーの解析

スクリプトファイル：Green3-3-5b.R

利用した関数：MASS::polr

方法：演習3.5.1 (p.203)

ダミー変数 1 を使った解析

ダミー変数 1

```
polr_out2 <- polr(degree ~ d1 + d2 + d3 ,  
  weights = freq, data = df,  
  method = "logistic")
```

ダミー変数 1

##	group	degree	freq	d1	d2	d3
## 1	old	none	2	0	0	0
## 2	old	low	5	0	0	0
## 3	old	med	2	0	0	0
## 4	old	high	1	0	0	0
## 5	new1	none	2	1	0	0
## 6	new1	low	4	1	0	0
## 7	new1	med	2	1	0	0
## 8	new1	high	2	1	0	0
## 9	new2	none	0	0	1	0
## 10	new2	low	2	0	1	0
## 11	new2	med	5	0	1	0
## 12	new2	high	3	0	1	0
## 13	new3	none	1	0	0	1
## 14	new3	low	2	0	0	1
## 15	new3	med	7	0	0	1
## 16	new3	high	0	0	0	1

順序ロジスティック回帰：a×b 分割表 (a≥3、b ≥ 3)

- 表示3.8.6 JMP[モデルのあてはめ]の出力、表示3.8.7 Excel ソルバーの解析

スクリプトファイル：Green3-3-5b.R

利用した関数：MASS::polr、

方法：ダミー変数 1 による解析

```
summary(polr_out3)
## Coefficients:
##           Value Std. Error t value
## groupnew1 -0.4128    0.8575  -0.4814
## groupnew2 -2.0453    0.8874  -2.3047
## groupnew3 -1.0270    0.8268  -1.2422
##
## Intercepts:
##           Value Std. Error t value
## high|med -2.8443    0.7702  -3.6927
## med|low  -0.6697    0.6204  -1.0795
## low|none  1.2703    0.6537   1.9434
```

群が対応

JMP と
正負が逆

```
screenreg(polr_out3)
## =====
##                               Model 1
## -----
## groupnew1          -0.41
##                   (0.86)
## groupnew2          -2.05 *
##                   (0.89)
## groupnew3          -1.03
##                   (0.83)
## high|med           -2.84 ***
##                   (0.77)
## med|low            -0.67
##                   (0.62)
## low|none           1.27
##                   (0.65)
```



順序ロジスティック回帰： $a \times b$ 分割表 ($a \geq 3$ 、 $b \geq 3$)

p.231

- 表示3.8.6 JMP[モデルのあてはめ]の出力、表示3.8.7 Excel ソルバーの解析
スクリプトファイル：Green3-3-5b.R
利用した関数：MASS::polr、summary、car::Anova
方法：ダミー変数 1 を使った解析

```
Anova(polr_out2)
## Analysis of Deviance Table (Type II tests)
##
## Response: degree
##      LR Chisq Df Pr(>Chisq)
## d1    0.2328  1  0.62946
## d2    5.6581  1  0.01737 *
## d3    1.5707  1  0.21011
```



- 引用文献

藤井良宜 (2010) カテゴリカルデータ解析、共立出版

丹後俊郎 (1998) 統計学のセンス、朝倉書店

UCLA: Statistical Consulting Group.

from <https://stats.idre.ucla.edu/r/faq/ologit-coefficients/> (accessed November 10, 2021)

- 作成

片瀬雅彦

- 作成時期

2021年10月22日、11月29日