

# 数量化Ⅱ類



判別分析+

心理データ解析演習  
2013/6/19 M1 宮坂まみ

# Why 数量化理論?



数量化理論の誕生

# 数量化理論のプロフィール



- ❧ 林知己夫を中心として発展した
- ❧ 多変量解析の理論
- ❧ 林知己夫; 1974-1986年 統計数理研究所七代所長



(統計数理研究所HPより)

# 林先生の疑問と理念



- ☞ サンプル調査を勉強することになった林先生。
  - 「標本を増やせば必ず差が出るので、やらなくても分かっていると感じたのである。」
  - 「世の中には，“数学的に同じ”，“相関係数0”，“数学的に独立なもの”は存在すると考える方がおかしいのである。しかし，これが統計的仮説の土台となっているのである。」
  - 「このあたりで，従来から持っていた私の**“データ”なるものを大事にする**考え方が私の志向する統計学の中で固まりかけてきた。つまり，“データによる現象解析のための統計的方法（データの中で考えデータの中に情報を見い出す）という立場の発祥ということができる。”」（林, 1993）



# ちなみにその頃求められたこと



## ☞ 読み書き能力

アメリカの占領政策の一環として，“日本人の読み書き能力”を調査せよという指令が出た。

→グループ結成。林先生はサンプリング担当。

## ☞ 仮釈放の成功・失敗

法務省の西村氏からの依頼。→数量化Ⅰ類誕生

## ☞ 文化の好み

戦前の学生はフランス文化好み。第二次世界大戦後、アメリカ文化好みが多くなりつつあった。→数量化Ⅱ類

※数量化Ⅲ類は数量化の思想が固まってから出てきた

# 数量化のはじまり



## 仮釈放の予測

- 刑期の1/3を過ぎたとき，累犯の虞のないものを釈放する。
- どう予測するか，判断の成功率を高めなければならない。

それで

とりあえず横浜刑務所を視察

## ∞ 刑務所生活

- よく適応し，素直で明るくよく働く→累犯
- 中くらい→最も累犯しにくい
- すごく反抗的→累犯

## ∞ 裁判に対する態度

- 満足→累犯
- 適度に不満→最も累犯しにくい
- 不満→累犯



+-の方向，直線性が成り立たない!  
なぜ?

☞ なぜこうなるかを知るために、仮釈放に必要な調査項目を練り上げた（殆ど**定質的データ**）

- 諸特性（体型，生活歴，属性，居住歴，性格など）
- 犯した犯罪に対する態度
- 受刑中の行動や身に付けた技能
- 釈放されたとき帰り行く家庭環境
- 社会環境などの生活環境 など

☞ 何を対象とするか？ 罪を犯さないとは何か？

再び罪を犯すものの95%は1年以内に累犯するらしい



1年以内に罪を犯さないもの = 釈放に成功 としよう



# 表にすると，例えばこんな感じ



	被験者	アイテム1 (受刑生活)		アイテム2 (裁判の態度)	
		カテゴリー1 (適応的)	カテゴリー2 (反抗的)	カテゴリー1 (満足)	カテゴリー2 (不満)
外的基準 (仮釈放成功)	1		✓	✓	
	2	✓			✓
	3		✓		✓
	⋮				
	$n_l$	✓		✓	
外的基準2 (仮釈放失敗)	1		✓		✓
	2	✓		✓	
	⋮				
	$n_m$		✓		✓

# ここで問題発生



「カテゴリカルデータ」なんて言葉もなかった時代、定質的データをどう取り扱うか？

釈放の成功、不成功を予測するのだから、これがなるべくよく弁別されるように、**質的データに数量を与えればよい**のではないか。



それで

**数量化されたものを加算して、判別すればよい!**

で,  $\delta_i(j, k_j)$



$\delta_i(j, k_j)$  という記号を用いた。



$i$  という人が  $j$  項目 (アイテム)、 $k_j$  カテゴリーに反応している時に1、そうでないときに0とする  
記号

# 記号の意味



記号	意味
$i$	$i$ という人
$j$	項目 (アイテム)
$k_j$	カテゴリー
$J$	項目 (アイテム) の総数
$K_j$	カテゴリーの総数
$x_{jk_j}$	各質的な調査項目に与える数値
$\alpha_i$	$i$ という人のもつ数値



$J$

	被験者	アイテム1 $j$		アイテム2 $j$	
		カテゴリー1	カテゴリー2	カテゴリー1	カテゴリー2
		$k_j$	$k_j$	$k_j$	$k_j$
外的基準 (仮釈放成功)	$i$ $i$ $i$ $\vdots$ $i$	✓	✓	✓	✓
外的基準2 (仮釈放失敗)	$i$ $i$ $i$ $\vdots$ $i$	✓	✓	✓	✓

$K_j$

1 or 0  
=ダミー変数

$x_{jk_j}$

$x_{jk_j}$

$x_{jk_j}$

$x_{jk_j}$

すると,  $\alpha_i$ は



$$\alpha_i = \sum_j^J \sum_{k_j}^{K_j} \delta_i(j, k_j) x_{jk_j}$$

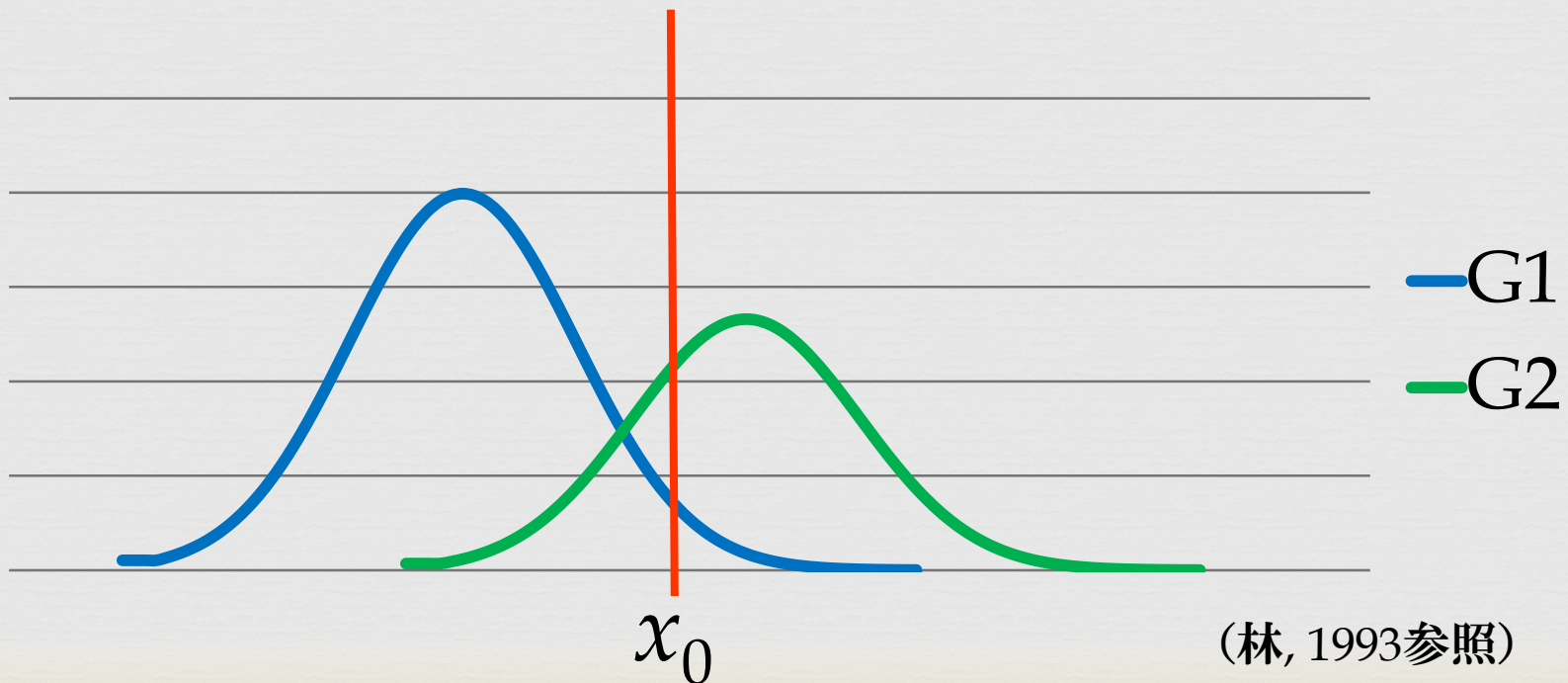
で

$\alpha_i$ が分かると各群の分布が作れる

成功・失敗を判定する点 $x_0$ を考える。

これ以上だと成功，以下だと失敗と判断した時の成功率を最大にするように

$x_{jk_j}$ を求めれば良い。



(林, 1993参照)

# 数値をつける



PCが発達していなかったので、紆余曲折あり...

アイテム別に釈放成功群と失敗群の判別が大きくなるようにした。



- 2群の平均値の差の2乗を最大にするように $x_{jk_j}$ を求める。
- ウェイトをかけ、 $P$ が最大になるように $a$ を算出する。

詳しく知りたい人は

西村克彦・林知己夫: 仮釈放の研究, 東京大学出版会, 1995.



# とにかく、数量化の思想は



- ❧ 「数は対象に内在するものではなく目的の科学的達成のための道具」
- ❧ 「質的なものに数量を与えて分析をほどこし、そうすることによってわからなかった実のあることがわかってくる」

(林, 1993)

# おまけ; 多変量解析比較



対比させてみると...

# 因果関係の手法



原因と結果を識別し，予測や判定を行う。

## 因果関係型の手法分類（大野, 1998参照）

		従属変数	
		名義尺度	間隔尺度
説明変数	名義尺度	数量化理論Ⅱ類	数量化理論Ⅰ類
	間隔尺度	判別分析	重回帰分析

# 類似関係型の手法



- 目的変数がない。何らかの類似関係の構造を明らかにする。

## 類似関係型の手法分類（大野, 1998参照）

類似性	手法
変数間の類似性	主成分分析, 因子分析
個体間の類似性	クラスター分析
変数と個体間の類似性	数量化理論Ⅲ類



# 判別分析



いきなり脱線。そしてかなり重要。

# 数量化Ⅱ類と判別分析



☞ 「これは、当時は気がつかなかったが、いわば線形判別関数の一般化となっている。」(林, 1993)

ということで、判別分析について。

# 判別分析1; 判別分析とは?



- ☞ グループが $G_1, G_2, \dots, G_n$ とある。
- ☞ 1個の新しいデータ $S$ がある。 $S$ がどのグループに属するか分からない。そして知りたい。そんな時に使う。



新しい人  
 $S$



(石村・石村, 2007参照)

# 判別分析2; 判別の基準



∞ 線形判別関数...直線で分ける

∞ マハラノビスの距離...二次曲線で分ける



判別分析 = グループ間に境界線を引くこと。

今回のメインは数量化Ⅱ類なので、  
関連する線形判別関数について



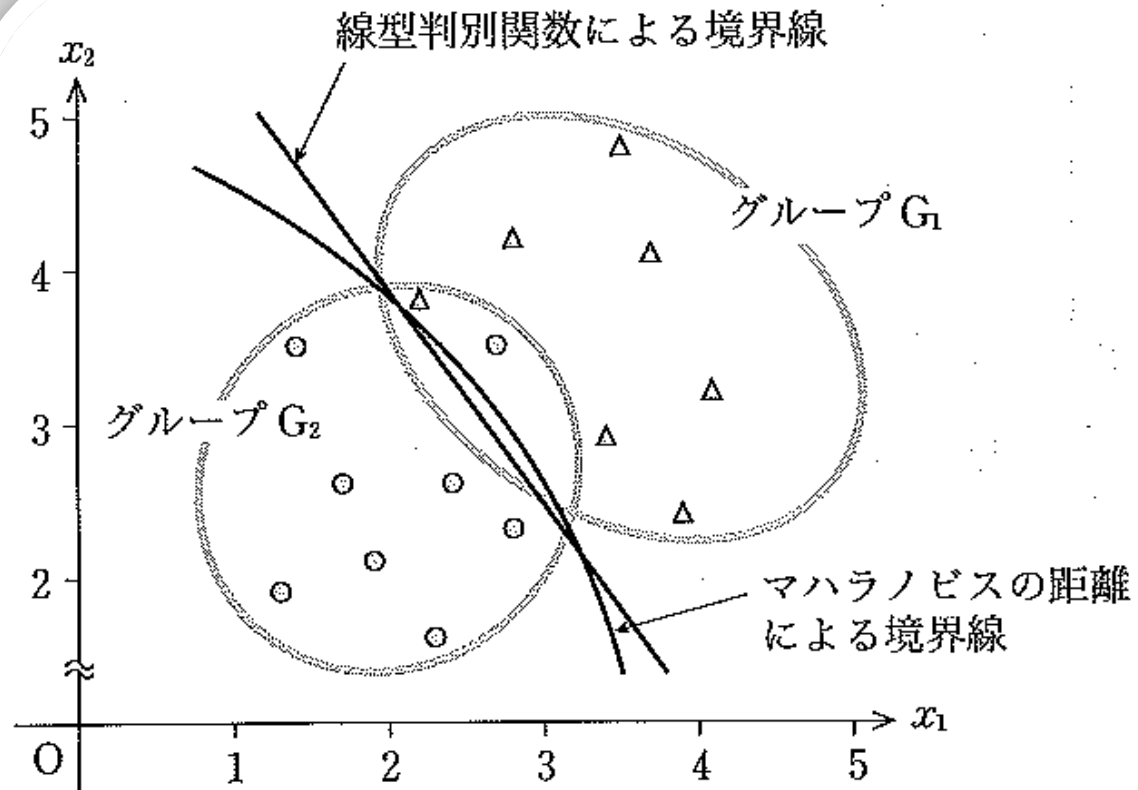


図 5.10.1 2つの境界線の関係

(石村・石村, 2007)

# 判別分析3; 境界線



- ☞ 目標; 最良の1本を引く
- ☞ 最良の1本 = グループを最も良く分離する1次関数 $z$

$$z = a_1x_1 + a_2x_2 + a_0 \quad \dots \text{線形判別関数}$$

- ☞ すべきこと; 係数 $a_1, a_2$ と定数項 $a_0$ を決定する

# 判別分析4; 判別得点



- ☞  $z = a_1x_1 + a_2x_2 + a_0$  の  $x$  にデータを代入した値
- ☞ 各データが境界線 ( $0 = a_1x_1 + a_2x_2 + a_0$ ) からどの程度離れているのかを示す
- ☞ 変数はいくつでもOK

ヒューマンドラマ派		
被験者 No	外向性	情緒 安定性
1	4	10
2	3	7
⋮	⋮	⋮
$n_l$	9	4

サスペンス派		
被験者 No	外向性	情緒 安定性
1	2	1
2	6	5
⋮	⋮	⋮
$n_m$	8	5

被験者2の判別得点  
 $= 6a_1 + 5a_2 + a_0$

(仮想データ)

# 判別分析5; 判別得点の変動1



- 判別得点の変動=判別得点と平均との差の2乗和
- 全変動 $S_T$ =グループ間変動 $S_B$ +グループ内変動 $S_W$



最良の1本を引く  
=全変動の中で グループ間変動 を最大にする

Q 判別得点は出せるようになった。平均は？



# 判別分析6; 判別得点の平均



∞  $G_1$  の平均  $\bar{z}^{(1)}$

$$= x_1 \text{ の } G_1 \text{ 内平均} \times a_1 + x_2 \text{ の } G_1 \text{ 内平均} \times a_2 + a_0$$

∞  $G_2$  の平均  $\bar{z}^{(2)}$

$$= x_1 \text{ の } G_2 \text{ 内平均} \times a_1 + x_2 \text{ の } G_2 \text{ 内平均} \times a_2 + a_0$$

∞ 全体の平均  $\bar{z}$

$$= x_1 \text{ の全体の平均} \times a_1 + x_2 \text{ の全体の平均} \times a_2 + a_0$$

# 判別分析7; 判別得点の変動2



∞ グループ内変動 $S_W = G_1$ 内変動+ $G_2$ 内変動

•  $G_1$ 内変動

$$= (z_1^{(1)} - \bar{z}^{(1)})^2 + (z_2^{(1)} - \bar{z}^{(1)})^2 + \dots + (z_{n_l}^{(1)} - \bar{z}^{(1)})^2$$

•  $G_2$ 内変動

$$= (z_1^{(2)} - \bar{z}^{(2)})^2 + (z_2^{(2)} - \bar{z}^{(2)})^2 + \dots + (z_{n_m}^{(2)} - \bar{z}^{(2)})^2$$

∞ グループ間変動 $S_B$

$$= n_l \times (\bar{z}^{(1)} - \bar{z})^2 + n_m \times (\bar{z}^{(2)} - \bar{z})^2$$

∞ 全変動 $S_T = S_W + S_B$

# 判別分析8; 線形判別関数



✧ ポイント;  $\frac{S_B}{S_T}$ が最大となる $a_1, a_2$ を求める。

$$\begin{cases} \frac{\partial F(a_1, a_2)}{\partial a_1} = 0 \\ \frac{\partial F(a_1, a_2)}{\partial a_2} = 0 \end{cases}$$

この連立方程式を全体の平均を通るように解くと、  
線形判別関数 $z$ が出る

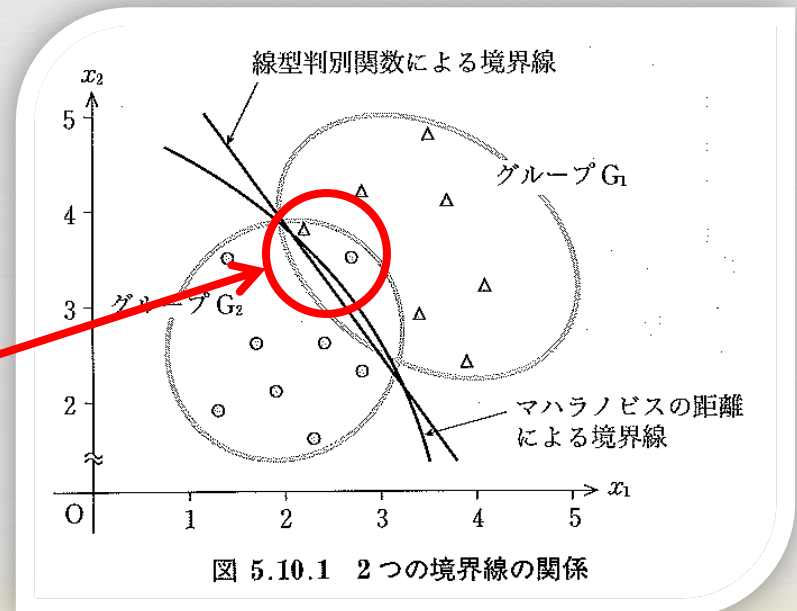
# 判別分析9; 誤判別率



- ❧ どうしても、いくつかのデータは誤って判別されてしまう。
- ❧ どれくらい誤って判別されたか = **誤判別率**

$$\frac{\text{誤って判別されたデータ数}}{\text{全データ数}}$$

このへん



# 判別分析10; Plus One



- ☞ データの精度を上げる方法として、Cross-validation (交差確認, 交差検証, 交差妥当化) もある。
1. データから1つだけ検証用に残す。
  2. 残った $n-1$ 個のデータを使って判別分析をする。
  3. 残した1つのデータを当てはめる。
  4. 1 ~ 3 を $n$ 回繰り返す (全てのデータが1回ずつ検証される)。
  5.  $n$ 回の平均でモデルを評価する。

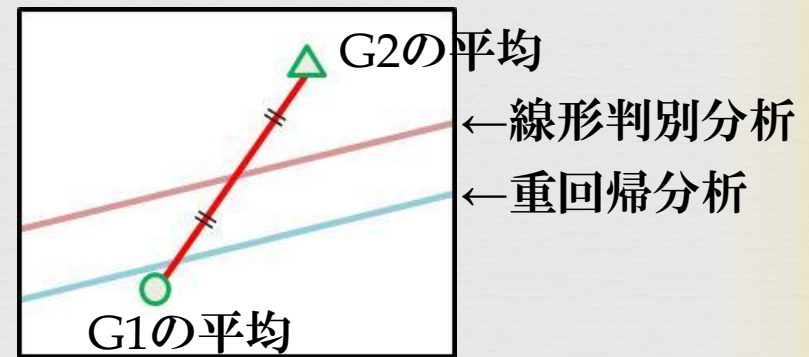


# 判別分析11; Plus One



## 関連する分析

- **重回帰分析**...線形判別分析と傾きは同じ。重回帰式を，座標上の各グループの平均値の中点を通るように縦に移動したものが判別式。



(片所『シリウス先生の心理統計学』，  
青木『おしゃべりな部屋』 参照)

- **ロジスティック回帰分析**...判別スコアを，グループに属する確率に直してロジスティック曲線に当てはめたもの。
- **正準判別分析**...3群以上の判別分析。 などなど

# What's 数量化Ⅱ類?



判別分析を念頭に...

# 質的なデータ



映画の 好み	被験者 No	アイテム1 外向性-内向性		アイテム2 情緒安定性	
		カテゴリ1	カテゴリ2	カテゴリ1	カテゴリ2
		外向的	内向的	安定	不安定
ヒューマン ドラマ	1		✓		✓
	2	✓			✓
	⋮	⋮	⋮	⋮	⋮
	$n_1$	✓			✓
サスペンス	1	✓		✓	
	2		✓	✓	
	⋮	⋮	⋮	⋮	⋮
	$n_m$		✓		✓

(仮想データ)

# ダミー変数に変換



映画の 好み	被験者 No	アイテム1 外向性-内向性		アイテム2 情緒安定性	
		カテゴリ1	カテゴリ2	カテゴリ1	カテゴリ2
		外向的	内向的	安定	不安定
ヒューマン ドラマ	1	0	1	0	1
	2	1	0	0	1
	⋮	⋮	⋮	⋮	⋮
	$n_1$	1	0	0	1
サスペンス	1	1	0	1	0
	2	0	1	1	0
	⋮	⋮	⋮	⋮	⋮
	$n_2$	0	1	1	0

$x_{11}$

$x_{12}$

$x_{21}$

$x_{22}$

# 判別式 $Y$



$$\text{⌘ } Y = a_{11}x_{11} + a_{12}x_{12} + a_{21}x_{21} + a_{22}x_{22} + a_0$$

⌘ 判別式 $Y$ の係数 $a_{ij}$  = **カテゴリ数量**

⌘ 判別式 $Y$ にデータ $x$ を代入した値 = **判別スコア**



# ここでも



- ∞ 全変動 $S_T$ =グループ間変動 $S_B$ +グループ内変動 $S_W$
- ∞ **グループ間変動を最大にするようにカテゴリ数量 $a_{ij}$ を決定する**



- ∞ 相関比

$$\eta^2 = \frac{S_B}{S_T} = \frac{n_l(\bar{Y}^{(1)} - \bar{Y})^2 + n_m(\bar{Y}^{(2)} - \bar{Y})^2}{\sum_{i=1}^{n_l} (Y_i^{(1)} - \bar{Y})^2 + \sum_{i=1}^{n_m} (Y_i^{(2)} - \bar{Y})^2}$$

を最大にするようなカテゴリ数量 $a_{ij}$ を求める。

# ところで



- ☞  $x_{11}+x_{12}=1, x_{21}+x_{22}=1$  なので,  
 $a_{11}=0, a_{21}=0$  として計算する。
- ☞ でも0のままというのも不都合なので、最後に基準化という作業をする。
- ☞ 基準化=各アイテム内のカテゴリ数量の平均が0となるように、カテゴリ数量を変換すること。

# 式ができたなら



## ∞ 判別の程度

$\eta^2$ の値が1に近いほどグループ間変動 $S_B$ が大きい  
→2つの群がよく判別されている。

∞ 外的基準に及ぼす各アイテムの影響の度合い  
範囲が大きいほど影響が大きい。

範囲=各アイテム内のカテゴリ数量の最大値－最小値

## ∞ 判別式を使ってみる

新たなデータを判別式に代入すると、判別スコアが出る。  
それを当てはめると、どちらに入るかが分かる。

Let's analyze カテゴリカルデータ！



ありがとう，統計ソフト。

# その前に Question



Q でも、国際誌へ投稿するには、分析方法はどのように記載すれば良いのか？

と思い、ググってみました。

A 国際誌では数量化理論は認知度が低い。数量化Ⅱ類は、実質、ダミー変数を用いた判別分析である。

→ダミー変数を用いた判別分析をすれば良い。

(青木『おしゃべりな部屋』)



# 得られたデータ



生活	被験者 No	アイテム1 動作		アイテム2 几帳面さ	
		カテゴリ1	カテゴリ2	カテゴリ1	カテゴリ2
		機敏さ	のんびり	几帳面	大雑把
朝型	1		✓	✓	
	2		✓		✓
	3	✓		✓	
	4	✓		✓	
	5	✓		✓	
	6	✓			✓
夜型	1	✓		✓	
	2		✓	✓	
	3		✓	✓	
	4		✓		✓
	5		✓		✓
	6	✓		✓	

# 入力の仕方



## 個票データ

ID	life	movement	methodical
1	朝型	のんびり	几帳面
2	朝型	のんびり	おおざっぱ
3	朝型	機敏	几帳面
4	朝型	機敏	几帳面
5	朝型	機敏	几帳面
6	朝型	機敏	おおざっぱ
7	夜型	機敏	几帳面
8	夜型	のんびり	几帳面
9	夜型	のんびり	几帳面
10	夜型	のんびり	おおざっぱ
11	夜型	のんびり	おおざっぱ
12	夜型	機敏	几帳面

(数量化Ⅱ類はこちらを使う)

## 集計データ

life	movement	methodical	Freq
朝型	機敏	几帳面	3
朝型	機敏	おおざっぱ	1
朝型	のんびり	几帳面	1
朝型	のんびり	おおざっぱ	1
夜型	機敏	几帳面	2
夜型	機敏	おおざっぱ	0
夜型	のんびり	几帳面	2
夜型	のんびり	おおざっぱ	2

(度数)

今回は、Excelで作ってcsv形式で保存しました

# 使う関数

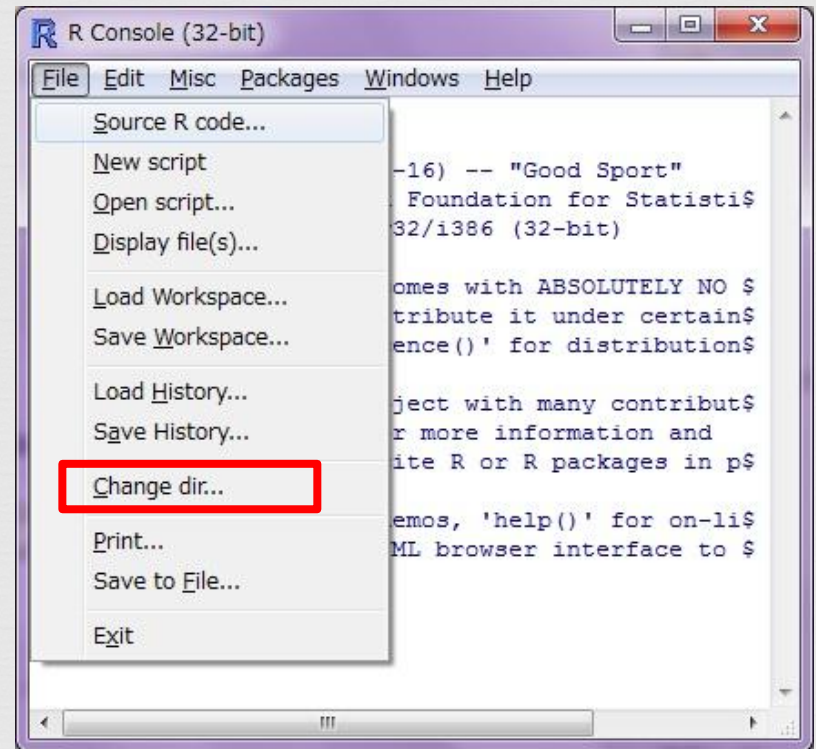


関数名	パッケージ名	関数の意味
class	base	変数の特性を表示する
factor	base	カテゴリカル変数に換える
head	utils	データフレームの最初の6行を表示する
lda	MASS	線形判別分析を行う
read.csv	utils	csvファイルを読み込む
predict	stats	判別分析などの結果を用いて行った予測結果を表示する
tapply	base	グループごとに平均などを求める

# いざ。



1. まず、データの場所を指定する。



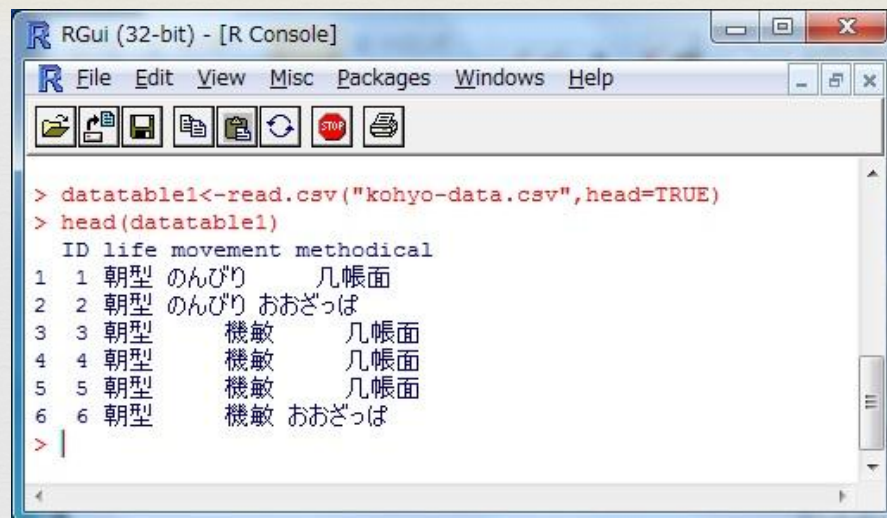


## 2. 個票データを読み込む

```
datatable1<-read.csv("kohyo-  
+data.csv",head=TRUE)
```

## 3. 最初の6行だけを表示

```
head(datatable1)
```



```
RGui (32-bit) - [R Console]  
File Edit View Misc Packages Windows Help  
> datatable1<-read.csv("kohyo-data.csv",head=TRUE)  
> head(datatable1)  
  ID life movement methodical  
1  1 朝型 のんびり   几帳面  
2  2 朝型 のんびり おおざっぱ  
3  3 朝型   機敏   几帳面  
4  4 朝型   機敏   几帳面  
5  5 朝型   機敏   几帳面  
6  6 朝型   機敏 おおざっぱ  
> |
```

## 4. ID以外について、カテゴリー変数であると示す

```
datatable1$life<-
```

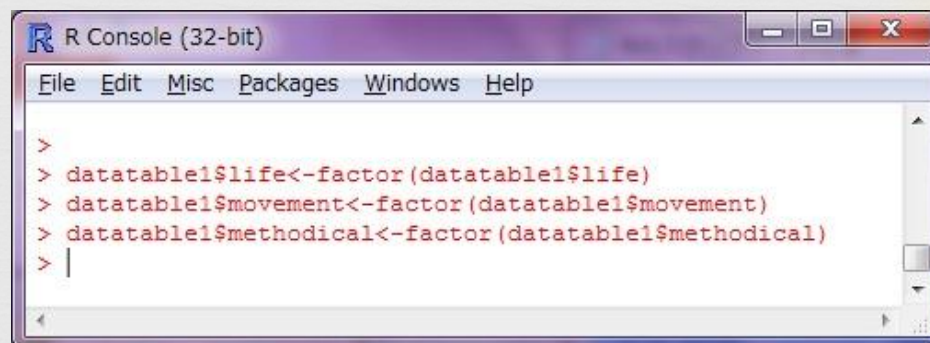
```
+factor(datatable1$life)
```

```
datatable1$movement<-
```

```
+factor(datatable1$movement)
```

```
datatable1$methodical<-
```

```
+factor(datatable1$methodical)
```



```
R Console (32-bit)  
File Edit Misc Packages Windows Help  
>  
> datatable1$life<-factor(datatable1$life)  
> datatable1$movement<-factor(datatable1$movement)  
> datatable1$methodical<-factor(datatable1$methodical)  
> |
```

注; 2行目左側の「+」は、Rで改行した際に自動的に表示されます。敢えて改行する必要はありません。



## 5. 集計

### ① 全体の集計を出す

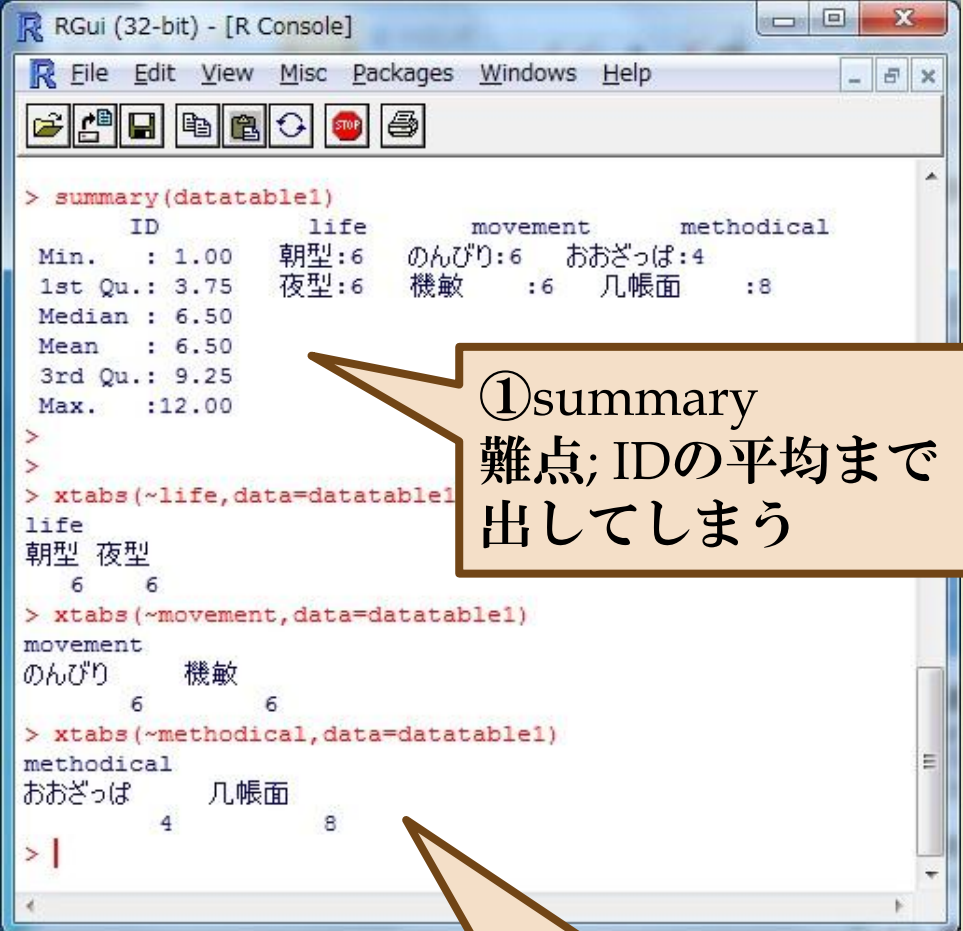
**summary**(datatable1)

### ② 変数毎に表形式で出す

**xtabs**(~life,data=datatable  
+1)

**xtabs**(~movement,data=d  
+atatable1)

**xtabs**(~methodical,data=d  
+atatable1)



```
> summary(datatable1)
      ID      life      movement      methodical
Min.   : 1.00   朝型:6   のんびり:6   おおざっぱ:4
1st Qu.: 3.75   夜型:6   機敏      :6   几帳面      :8
Median : 6.50
Mean   : 6.50
3rd Qu.: 9.25
Max.   :12.00

>
>
> xtabs(~life,data=datatable1)
life
朝型 夜型
   6   6

> xtabs(~movement,data=datatable1)
movement
のんびり  機敏
   6         6

> xtabs(~methodical,data=datatable1)
methodical
おおざっぱ  几帳面
   4         8

> |
```

①summary  
難点; IDの平均まで  
出してしまう

②xtabs  
難点; 変数ごとに指  
定する必要がある

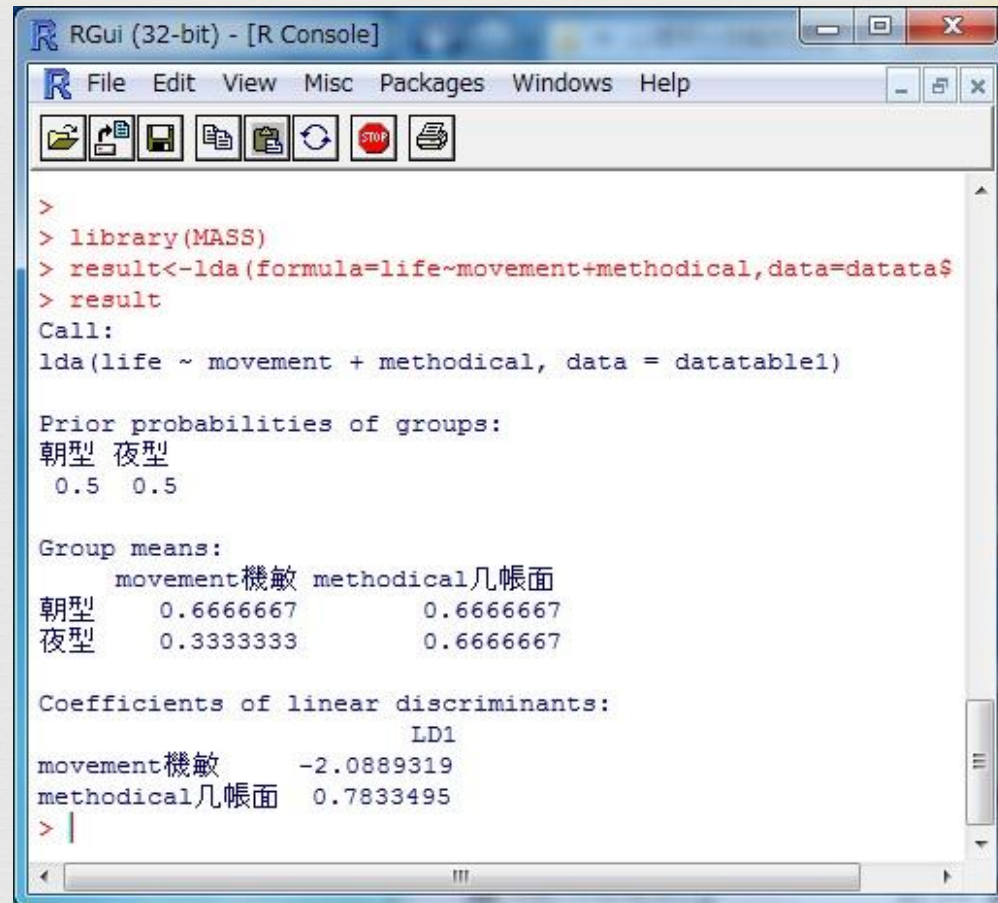
## 6. 線形判別分析をする

① library(MASS)

② result<-

+lda(formula=life~move  
+ment+methodical,data=d  
+atable1)

③ result



```
RGui (32-bit) - [R Console]
File Edit View Misc Packages Windows Help
>
> library(MASS)
> result<-lda(formula=life~movement+methodical,data=datata$
> result
Call:
lda(life ~ movement + methodical, data = datatable1)

Prior probabilities of groups:
朝型 夜型
0.5 0.5

Group means:
      movement機敏 methodical几帳面
朝型    0.6666667    0.6666667
夜型    0.3333333    0.6666667

Coefficients of linear discriminants:
              LD1
movement機敏  -2.0889319
methodical几帳面  0.7833495
> |
```

なんと! ここで行っているのは、判別分析でした。  
ldaは線形判別分析用の関数です。それをカテゴリカルデータであると指示して実行したのでした。Rで行う数量化Ⅱ類の説明では、ldaが用いられていることが多いようです。

```
RGui (32-bit) - [R Console]
File Edit View Misc Packages Windows Help
>
> library(MASS)
> result<-lda(formula=life~movement+methodical,data=datata$
> result
Call:
lda(life ~ movement + methodical, data = datatable1)

Prior probabilities of groups:
 朝型 夜型
 0.5  0.5

Group means:
      movement機敏 methodical几帳面
朝型    0.6666667    0.6666667
夜型    0.3333333    0.6666667

Coefficients of linear discriminants:
              LD1
movement機敏  -2.0889319
methodical几帳面  0.7833495
> |
```

各群の事前確率。各群に属するだろう確率。  
デフォルトの設定は50%ずつ。

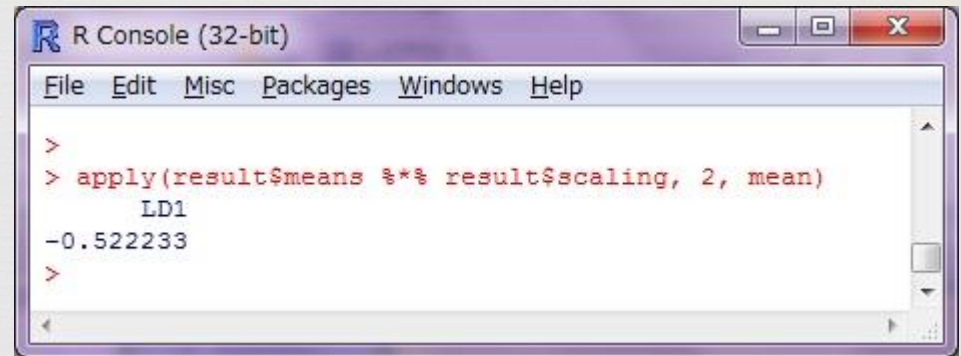
各群ごとの各変数の平均値

判別係数  
$$Z = a_1x_1 + a_2x_2 + a_0$$
$$Z = -2.089x_1 + 0.783x_2 + a_0$$



## 7. 定数項を求める

```
apply(result$means %*%  
+result$scaling, 2, mean)
```



```
R Console (32-bit)  
File Edit Misc Packages Windows Help  
>  
> apply(result$means %*% result$scaling, 2, mean)  
LD1  
-0.522233  
>  
>
```



判別式完成

$$z = -2.089x_1 + 0.783x_2 + (-0.522)$$

## 8. 判別結果を出す

(result2<-predict(result))

各ケースがどの群に判別されたか

各ケースがそれぞれの群に所属する  
事後確率

判別値

判別式

$$z = -2.089x_1 + 0.783x_2 + (-0.522)$$

にデータを入れた値

```
R Console (32-bit)
File Edit Misc Packages Windows Help

> (result2<-predict(result))
$class
[1] 夜型 夜型 朝型 朝型 朝型 朝型 朝型 朝型 朝型 朝型 朝型 朝型
Levels: 朝型 夜型

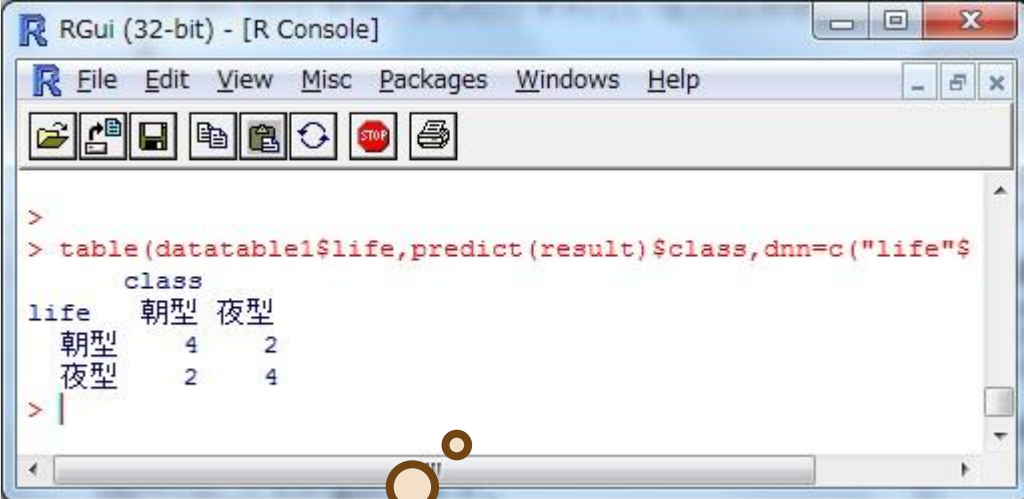
$posterior
      朝型      夜型
1 0.2871859 0.7128141
2 0.4100796 0.5899204
3 0.6330804 0.3669196
4 0.6330804 0.3669196
5 0.6330804 0.3669196
6 0.7485529 0.2514471
7 0.6330804 0.3669196
8 0.2871859 0.7128141
9 0.2871859 0.7128141
10 0.4100796 0.5899204
11 0.4100796 0.5899204
12 0.6330804 0.3669196

$x
      LD1
1  1.3055824
2  0.5222330
3 -0.7833495
4 -0.7833495
5 -0.7833495
6 -1.5666989
7 -0.7833495
8  1.3055824
9  1.3055824
10 0.5222330
11 0.5222330
12 -0.7833495
```



## 9. 判別がうまく行われたか

```
table(datatable1$life,predict(result)$class,dnn=c  
+("life","class"))
```



```
>  
> table(datatable1$life,predict(result)$class,dnn=c("life"  
class  
life 朝型 夜型  
朝型 4 2  
夜型 2 4  
> |
```

本当は朝方だけど夜型に判別された人が2人、本当は夜型だけど朝方に判別された人が2人いるなあ。

$$\rightarrow \text{誤判別率} = \frac{4}{12} = 0.33$$

# 本当の数量化Ⅱ類との比較



## ☞ 使用プログラム

Black-Box...青木繁伸先生（群馬大学）のHPより

<http://aoki2.si.gunma-u.ac.jp/BlackBox/BlackBox.html>

## ☞ 分析プロシージャの指定

数量化Ⅱ類を選択

## ☞ 今回用意したファイルの名前

for-BlackBox.txt

## ☞ オプションに下記を入力

group-variable=2/variables=3-4

何列目に  
何のデータがあるのか

Black-Box --- data anal x

aoki2.si.gunma-u.ac.jp/BI

## 数量化 II 類

Fri Jun 14 03:43:10 2013

データセット名: for-BlackBox.txt  
 ケース数: 12  
 変数の個数: 4  
 有効ケース数: 12

群を表す変数: life  
 群1: life = 1  
 群2: life = 2

説明変数: アイテム カテゴリー数  
 movement 2  
 methodic 2

\*\*\*\* ノーマライズド カテゴリー スコア \*\*\*\*

アイテム-カテゴリー		解1
1 movement	0	1.06904
2	1	-1.06904
3 methodic	0	-0.534522
4	1	0.267261
重心		
群1		-0.356348
群2		0.356348
カットポイント		0.000000
相関比		0.126984

Black-Box --- data anal x

aoki2.si.gunma-u.ac.jp/BI

\*\*\*\* 偏相関係数 \*\*\*\*

アイテム	解1
movement	0.35635
methodic	0.13363

\*\*\*\* サンプルスコア \*\*\*\*

ケース	R	P	解1	
1	1	##	2	1.33631
2	1	##	2	0.53452
3	1		1	-0.80178
4	1		1	-0.80178
5	1		1	-0.80178
6	1		1	-1.60357
7	2	##	1	-0.80178
8	2		2	1.33631
9	2		2	1.33631
10	2		2	0.53452
11	2		2	0.53452
12	2	##	1	-0.80178

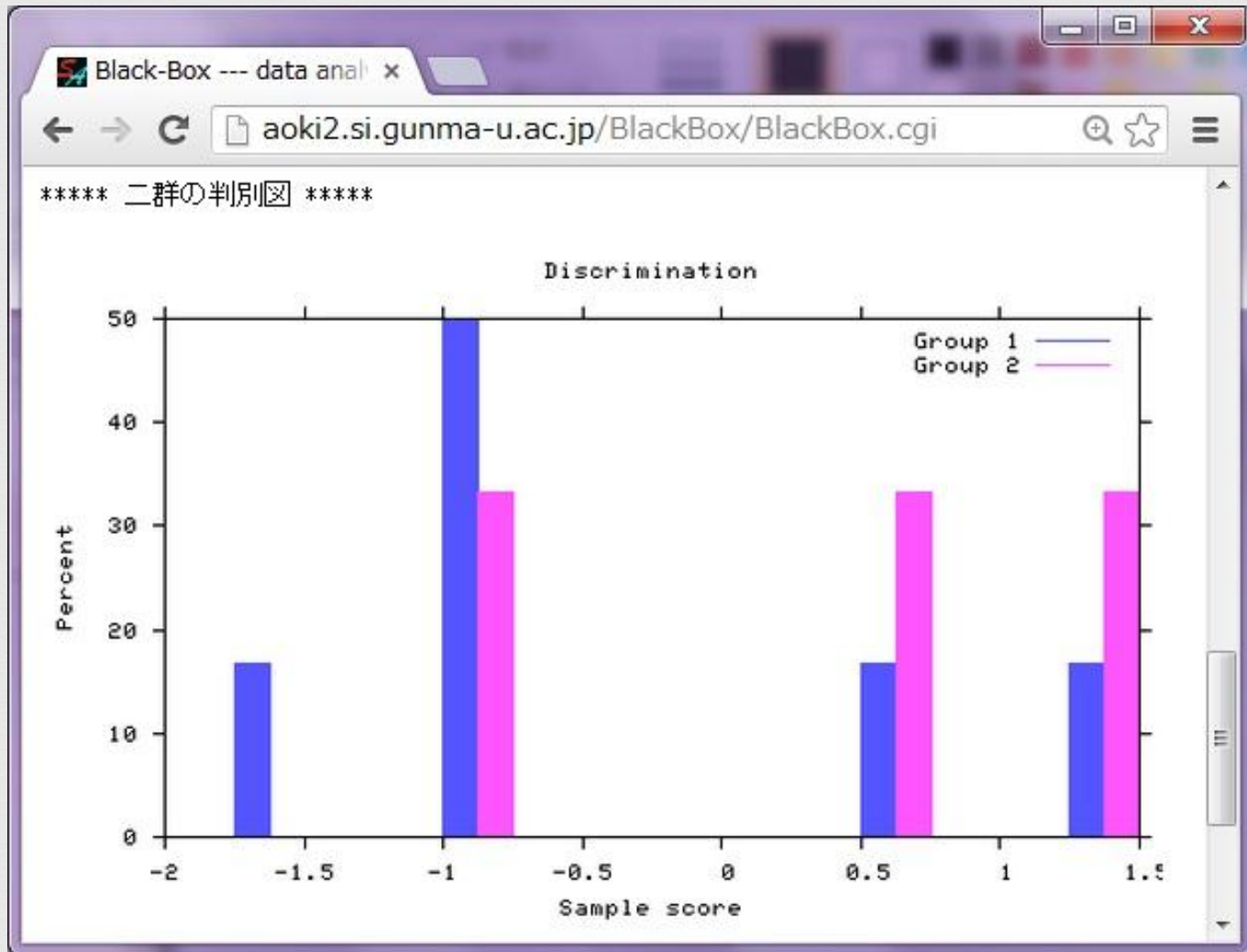
\*\*\*\* 判別結果総括表 \*\*\*\*

実際の群	判別された群		合計
	群1	群2	
群1	4	2	6
群2	2	4	6
%	( 66.7)	( 33.3)	(100.0)
%	( 33.3)	( 66.7)	(100.0)

正判別率: 66.67%

だいたい一緒

一緒





# References



あれこれ



# BOOK



- ☞ 青木繁伸: Rによる統計解析, オーム社, 2009.
- 石村貞夫・石村光資郎: 入門はじめての多変量解析, 東京図書, 2007.
- 大野高裕: 多変量解析入門 自由自在に使いこなすコツ, 同友館, 1998.
- 林知己夫: 数量化—理論と方法—, 朝倉書店, 1993.
- 藤井良宜: Rで学ぶデータサイエンス1 カテゴリーカルデータ解析, 共立出版, 2010.

## ☞ My Bible～数量化理論のデータというものの考え方から派生して～

- 住田幸次郎: 初歩の心理教育統計法, ナカニシヤ出版, 1988.

“統計”というものの考え方自体が理解できるように丁寧に説明されています。検定結果に関してつい誤って書きがちな文章表現例などもついています。

# HP



☞ おしゃべりな部屋（群馬大学・青木先生のHP）

☐ <http://aoki2.si.gunma-u.ac.jp/>

☞ シリウス先生の心理統計（統計アドバイザー・片所先生のHP）

☐ <http://homepage2.nifty.com/nandemoarchive/>

☞ データサイエンス研究室（同志社大学・金先生のHP）

☐ <http://mj.in.doshisha.ac.jp/>

☞ 統計数理研究所

☐ <http://www.ism.ac.jp/>