多変量解析

第10回　数量化2類

萩原•篠田
情報理工学部

数量化2類

質的変数

健常者かどうか吐き気と頭痛の有無から予測可能か？

質的変数

サンプル No.	健常者/患者 y	吐き気 x_1	頭痛 x_2
1	健常者	無	少
2	健常者	少	無
3	健常者	無	無
4	健常者	無	無
5	健常者	無	無
6	患者	少	多
7	患者	多	無
8	患者	少	少
9	患者	少	多
10	患者	多	少

keywords

質的変数、判別分析、ダミー変数、予測式、相関比、外的基準、カテゴリ数量、基準化

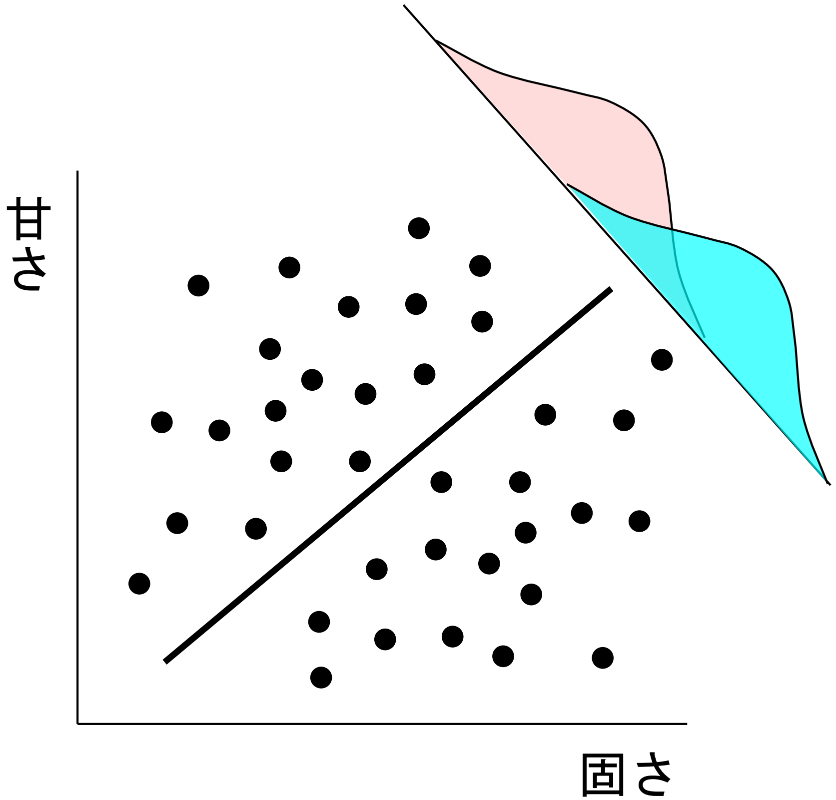
数量化理論とは

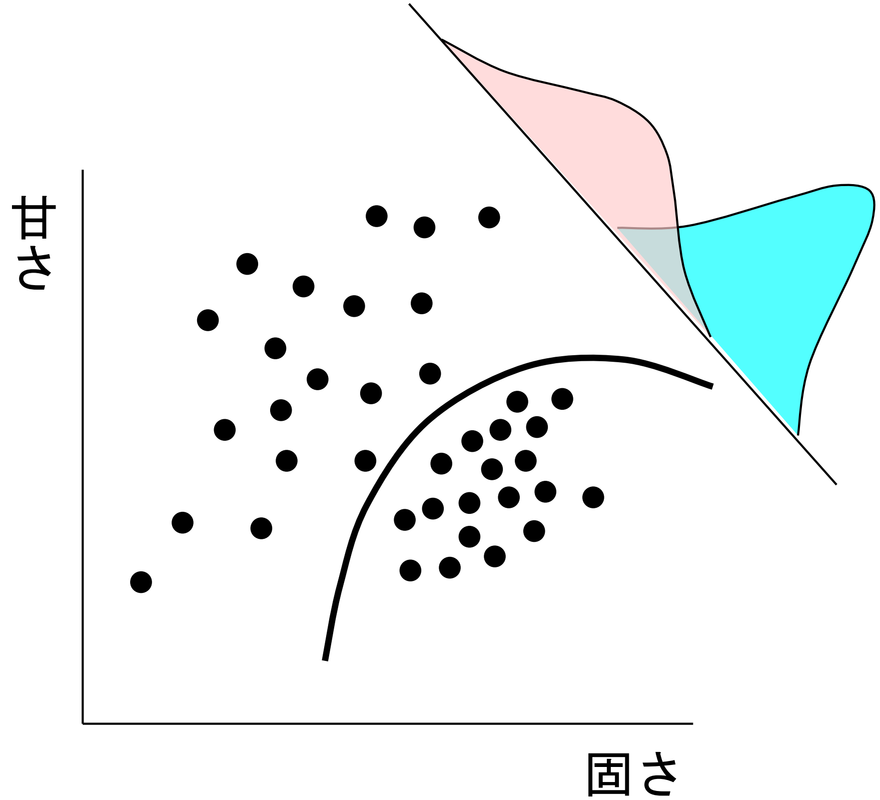
程度、状態　または　はい／いいえといった
質的データに数量を与え多変量データ解析を行う手法

重回帰分析　←→　数量化1類
判別分析　　←→　数量化2類
主成分分析　←→　数量化3類

判別分析の手法

「固さ」と「甘さ」の組み合わせで産地が判別
→ 線引きの方法は？





2群の分散が等しい場合

線型判別関数

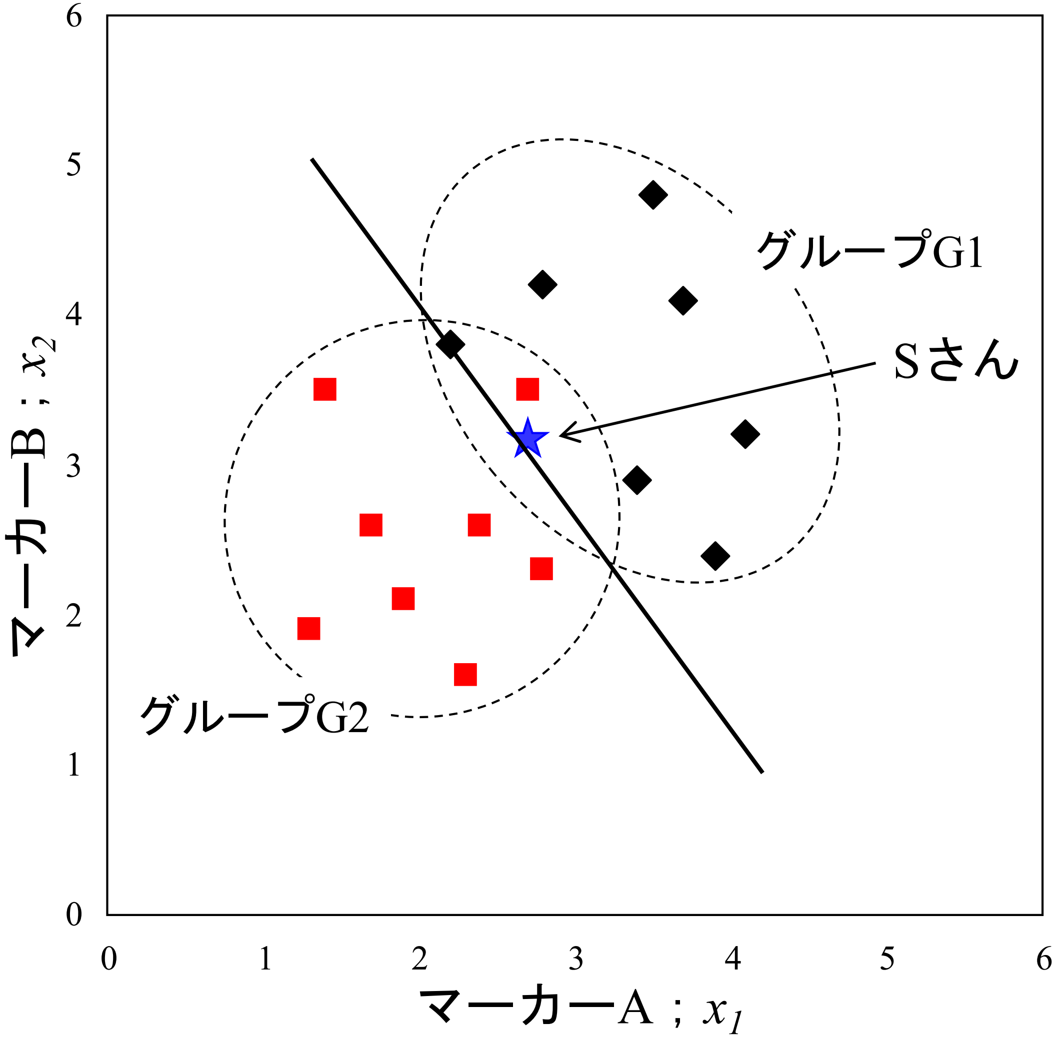
最もよく分離する直線を仮定し
その直線のどちら側に来るかを判別

2群の分散が等しくない場合

マハラノビスの距離

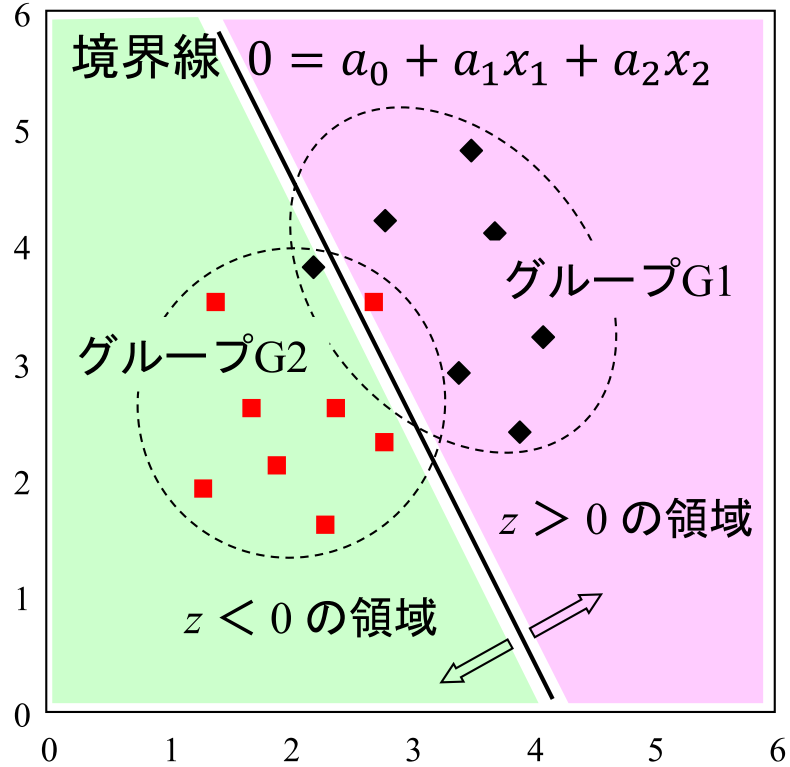
各グループの分布状態を考慮し
各グループの中心からの距離で判別

線形判別：線形判別関数で定義される判別得点の正負で判別



判別得点 z を線形判別関数で定義

$$z = a_0 + a_1 x_1 + a_2 x_2$$



点(p, q)と境界線の距離
ヘッセの標準形 $\dfrac{|a_0 + a_1 p + a_2 q|}{\sqrt{a_1^2 + a_2^2}}$

判別得点
＝直線までの距離×$\sqrt{a_1^2 + a_2^2}$

数量化2類

質的データから質的な形で与えられる外的基準を判別したり予測したりするための手法

A型とB型の2種類の血液型と性格について調べ
➡　性格によるA型・B型の判別をする

アンケート調査票

項目1
あなたは楽天家ですか
　1　はい　　2　いいえ

項目2
あなたは堅実派だと思いますか
　1　はい　　2　いいえ

項目3
あなたの血液型は
　1　A型　　2　B型

アイテム　　カテゴリ　　外的基準

被験者No.	楽天家		堅実派		血液型
	1	2	1	2	
1		レ	レ		A型
2		レ		レ	A型
3		レ	レ		A型
4	レ			レ	A型
1	レ			レ	B型
2	レ		レ		B型
3	レ			レ	B型

数量化1類

質的データから量的に測定される外的基準の予測や関係を調べる

野菜やタンパク質の「好き／嫌い」から

➡　体重を予測する

野菜とタンパク質のうち、どちらのアイテムが

➡　より体重に影響を及ぼしているかを調べる

アンケート調査票

項目1
あなたは野菜が好きですか
1　はい　　2　いいえ

項目2
あなたはタンパク質が好きですか
1　はい　　2　いいえ

項目3
あなたの体重は何Kgですか
________Kg

アイテム　　カテゴリ　　外的基準

被験者No.	野菜		タンパク質		体重
	1	2	1	2	
1	レ			レ	57
2	レ		レ		65
3		レ	レ		51
4	レ		レ		54
5		レ	レ		45
6	レ			レ	67

数量化2類の判別式の求め方

各アイテムのカテゴリの反応から外的基準の判別を行う

ダミー変数 x_{ij} を導入

$$x_{ij} = \begin{cases} 1 & \cdots \text{ アイテム } i \text{ のカテゴリ } j \text{ に反応したとき} \\ 0 & \cdots \text{ その他} \end{cases}$$

アイテム1　楽天家		アイテム2　堅実派	
カテゴリ1 はい	カテゴリ2 いいえ	カテゴリ1 はい	カテゴリ2 いいえ
↑ x_{11}	↑ x_{12}	↑ x_{21}	↑ x_{22}

$$Y = b_{11}x_{11} + b_{12}x_{12} + b_{21}x_{21} + b_{22}x_{22} + b_0$$

b_{ij} : カテゴリ数量

ダミー変数間の多重共線性（$x_{11} + x_{12} = 1,\ x_{21} + x_{22} = 1$）により

$$Y = (b_{12} - b_{11})x_{12} + (b_{22} - b_{21})x_{22} + b_0 + b_{11} + b_{21}$$

$$\boldsymbol{Y = c_1 x_{12} + c_2 x_{22} + c_0}$$

を求めることにする　ただし
$$\begin{cases} c_1 = b_{12} - b_{11} \\ c_2 = b_{22} - b_{21} \\ c_0 = b_0 + b_{11} + b_{21} \end{cases}$$

外的基準を最も良く判別したい➡群間変動を最大にするようカテゴリ数量 c_i を決定

判別分析

$$\boldsymbol{Y = c_1 x_{12} + c_2 x_{22} + c_0}$$

被験者No.	外的基準	アイテム1 x_{11}	アイテム1 x_{12}	アイテム2 x_{21}	アイテム2 x_{22}	判別得点 Y	平均
1	グループ1	0	1	1	0	$Y_1^{(1)} = c_1 + c_0$	$\bar{Y}^{(1)} = \bar{x}_{12}^{(1)} c_1 + \bar{x}_{22}^{(1)} c_2 + c_0 = \frac{3}{4} c_1 + \frac{2}{4} c_2 + c_0$
2		0	1	0	1	$Y_2^{(1)} = c_1 + c_2 + c_0$	
3		0	1	1	0	$Y_3^{(1)} = c_1 + c_0$	
4		1	0	0	1	$Y_4^{(1)} = c_2 + c_0$	
1	グループ2	1	0	0	1	$Y_1^{(2)} = c_2 + c_0$	$\bar{Y}^{(2)} = \bar{x}_{12}^{(2)} c_1 + \bar{x}_{22}^{(2)} c_2 = \frac{0}{3} c_1 + \frac{2}{3} c_2 + c_0$
2		1	0	1	0	$Y_2^{(2)} = c_0$	
3		1	0	0	1	$Y_3^{(2)} = c_2 + c_0$	
						全平均	$\bar{Y} = \bar{x}_{12} c_1 + \bar{x}_{22} c_2 + c_0 = \frac{3}{7} c_1 + \frac{4}{7} c_2 + c_0$

判別スコアの平均 $= \bar{Y}$

グループ1における判別スコアの平均 $= \bar{Y}^{(1)}$

グループ2における判別スコアの平均 $= \bar{Y}^{(2)}$

$$\underbrace{\text{全変動}}_{SS_T} = \underbrace{\text{グループ間変動}}_{SS_B} + \underbrace{\text{グループ内変動}}_{SS_W}$$

全変動SS_T，群間変動SS_B，相関比SS_B/SS_Tを計算

$$SS_T = \sum_{i=1}^{4}(Y_i^{(1)} - \bar{Y})^2 + \sum_{i=1}^{3}(Y_i^{(2)} - \bar{Y})^2 = \frac{2(6c_1^2 - 5c_1c_2 + 6c_2^2)}{7}$$

$$SS_B = 4(\bar{Y}^{(1)} - \bar{Y})^2 + 3(\bar{Y}^{(2)} - \bar{Y})^2 = \frac{(9c_1 - 2c_2)^2}{84}$$

$$\frac{SS_B}{SS_T} = \frac{(9c_1 - 2c_2)^2}{24(6c_1^2 - 5c_1c_2 + 6c_2^2)}$$

全変動SS_Tと群間変動SS_Bのc_1やc_2による偏微分を計算

$$\frac{\partial SS_T}{\partial c_1} = \frac{2}{7}(12c_1 - 5c_2), \qquad \frac{\partial SS_B}{\partial c_1} = \frac{9}{42}(9c_1 - 2c_2),$$

$$\frac{\partial SS_T}{\partial c_2} = \frac{2}{7}(12c_2 - 5c_1), \qquad \frac{\partial SS_B}{\partial c_2} = \frac{-2}{42}(9c_1 - 2c_2)$$

訂正

$F(c_1, c_2) = \frac{SS_B}{SS_T}$ の最大値を与える(c_1, c_2)

$$\begin{cases} \dfrac{\partial F}{\partial c_1} = \dfrac{1}{SS_T}\left(\dfrac{\partial SS_B}{\partial c_1} - \dfrac{SS_B}{SS_T}\dfrac{\partial SS_T}{\partial c_1}\right) = 0 \\ \dfrac{\partial F}{\partial c_2} = \dfrac{1}{SS_T}\left(\dfrac{\partial SS_B}{\partial c_2} - \dfrac{SS_B}{SS_T}\dfrac{\partial SS_T}{\partial c_2}\right) = 0 \end{cases} \Rightarrow \quad \frac{\partial SS_T}{\partial c_1}\cdot\frac{\partial SS_B}{\partial c_2} - \frac{\partial SS_B}{\partial c_1}\cdot\frac{\partial SS_T}{\partial c_2} = 0$$

これを解いて $\frac{c_1}{c_2} = \frac{14}{3} = 4.667$

そのときの相関比 $\frac{SS_B}{SS_T} = \frac{10}{17} = 0.5882$

$$u(x) = y(x) \cdot z(x)$$

$$u'(x) = \frac{du}{dx} = \frac{dy}{dx} \cdot z + y \cdot \frac{dz}{dx} = y'(x) \cdot z(x) + y(x) \cdot z'(x)$$

$$v(x) = \frac{1}{y(x)} = \{y(x)\}^{-1}$$

$$v'(x) = \frac{dv}{dx} = \frac{dv}{dy} \cdot \frac{dy}{dx} = (-1) \cdot \{y(x)\}^{-2} \cdot y'(x) = -\frac{y'(x)}{\{y(x)\}^2} = -\{v(x)\}^2 \cdot y'(x)$$

$$w(x) = \frac{z(x)}{y(x)} = z(x) \cdot \{y(x)\}^{-1}$$

$$w'(x) = \frac{dw}{dx} = \frac{dz}{dx} \cdot \frac{1}{y} + z \cdot \frac{d}{dx}\left(\frac{1}{y}\right) = \frac{z'(x)}{y(x)} + z(x) \cdot \frac{-y'(x)}{\{y(x)\}^2} = \frac{z'(x) - w(x) \cdot y'(x)}{y(x)}$$

$$Y = c_1 x_{12} + c_2 x_{22} + c_0 = c_2\left(\frac{c_1}{c_2}x_{12} + x_{22} + \frac{c_0}{c_2}\right) = c_2\left(\frac{14}{3}x_{12} + x_{22} + \frac{c_0}{c_2}\right)$$

全平均$\bar{x}_{12}$と$\bar{x}_{22}$において$\bar{Y} = 0$ ➡ $\bar{Y} = \bar{x}_{12}c_1 + \bar{x}_{22}c_2 + c_0 = c_2\left(\frac{3}{7}\frac{c_1}{c_2} + \frac{4}{7} + \frac{c_0}{c_2}\right) = 0$

これを解いて $\frac{c_0}{c_2} = -\frac{18}{7}$

$$Y = c_1 x_{12} + c_2 x_{22} + c_0 = c_2\left(\frac{c_1}{c_2}x_{12} + x_{22} + \frac{c_0}{c_2}\right) = c_2\left(\frac{14}{3}x_{12} + x_{22} - \frac{18}{7}\right)$$

Yの正負で判別するためc_2は任意の値でよい　ここでは $c_2 = 1$とする

以上より判別式は

$$Y = c_1 x_{12} + c_2 x_{22} + c_0 = \frac{14}{3}x_{12} + x_{22} - \frac{18}{7} = 4.667x_{12} + x_{22} - 2.571$$

$Y>0$ ・・・・グループ1(A型)

$Y<0$ ・・・・グループ2(B型)

基準化（アイテム内のカテゴリ数量の平均が0となるようにする）

被験者 No.	外的基準	アイテム1		アイテム2	
		x_{11}	x_{12}	x_{21}	x_{22}
1	グループ1	0	1	1	0
2		0	1	0	1
3		0	1	1	0
4		1	0	0	1
1	グループ2	1	0	0	1
2		1	0	1	0
3		1	0	0	1
		$\frac{4}{7}b_{11}+\frac{3}{7}b_{12}$		$\frac{3}{7}b_{21}+\frac{4}{7}b_{22}$	

アイテム内のカテゴリ数量平均＝0

$$\begin{cases} \bar{x}_{11}b_{11}+\bar{x}_{12}b_{12}=\frac{4}{7}b_{11}+\frac{3}{7}b_{12}=0 \\ \bar{x}_{21}b_{21}+\bar{x}_{22}b_{22}=\frac{3}{7}b_{21}+\frac{4}{7}b_{22}=0 \end{cases}$$

b_{ij} , b_0とc_{ij} , c_0の関係式

$$\begin{cases} b_{12}-b_{11}=c_1=\frac{14}{3}=4.667 \\ b_{22}-b_{21}=c_2=1 \\ b_0+b_{11}+b_{21}=c_0=-\frac{18}{7}=-2.571 \end{cases}$$

上記の5式を解いてb_{ij} , b_0を得る

$$Y=-2x_{11}+\frac{8}{3}x_{12}-\frac{4}{7}x_{21}+\frac{3}{7}x_{22}=-2x_{11}+2.67x_{12}-0.571x_{21}+0.429x_{22}$$

$$(Y=\frac{14}{3}x_{12}+x_{22}-\frac{18}{7},\ x_{11}+x_{12}=1,\ x_{21}+x_{22}=1)$$

新たな被験者P さんのアンケート結果

被験者	楽天家		堅実派		グループ（血液型）
	1	2	1	2	
P		レ	レ		？

判別式

$$Y = \frac{14}{3}x_{12} + x_{22} - \frac{18}{7}, \qquad x_{11} + x_{12} = 1,\ x_{21} + x_{22} = 1$$

基準化された判別式

$$Y = -2x_{11} + \frac{8}{3}x_{12} - \frac{4}{7}x_{21} + \frac{3}{7}x_{22} = -2x_{11} + 2.67x_{12} - 0.571x_{21} + 0.429x_{22}$$

x_{12}=1 , $x_{22} = 0$ を代入して被験者P さんの判別スコアを計算

$$Y = \frac{14}{3}x_{12} + x_{22} - \frac{18}{7} = \frac{14}{3} - \frac{18}{7} = \frac{44}{21} > 0$$

$$Y = -2x_{11} + \frac{8}{3}x_{12} - \frac{4}{7}x_{21} + \frac{3}{7}x_{22} = \frac{8}{3} - \frac{4}{7} = \frac{44}{21} > 0$$

グループ1(A型)と予想（判定）

各被験者の判別得点を求めてみる　$Y=\frac{14}{3}x_{12}+x_{22}-\frac{18}{7}=4.667x_{12}+x_{22}-2.571$

被験者 No.	外的基準	楽天家		堅実派		判別得点 Y	平均
		はい	いいえ	はい	いいえ		
1	グループ1（A型）	0	1	1	0	2.095	1.429
2		0	1	0	1	3.095	
3		0	1	1	0	2.095	
4		1	0	0	1	-1.571	
1	グループ2（B型）	1	0	0	1	-1.571	-1.905
2		1	0	1	0	-2.571	
3		1	0	0	1	-1.571	
平均		4/7	3/7	3/7	4/7	0.0	0.0

基準化された判別式

$$Y=-2x_{11}+\frac{8}{3}x_{12}-\frac{4}{7}x_{21}+\frac{3}{7}x_{22}=-2x_{11}+2.67x_{12}-0.571x_{21}+0.429x_{22}$$

基準化されたカテゴリ数量と範囲

アイテム	カテゴリ	カテゴリ数量	範囲
楽天家	1	-2	4.67
	2	2.67	
堅実派	1	-0.571	1.00
	2	0.429	
相関比		0.5882	

範囲：アイテム内のカテゴリ数量最大値ー最小値

範囲の大小により
“各アイテムの外的基準に及ぼす影響の度合”
を知ることができる

楽天家の範囲の方が堅実派の範囲より大きいので
血液型に及ぼす影響は楽天家の性格の方が大きい

別解

ダミー変数による数量化 → 判別分析（線形判別式による）

G_i の j 番目データの判別得点　$z_j^{(i)} = a_0 + x_{1j}^{(i)} a_1 + x_{2j}^{(i)} a_2$

G_i の判別得点の平均　$\bar{z}^{(i)} = a_0 + \bar{x}_1^{(i)} a_1 + \bar{x}_2^{(i)} a_2$

判別得点の全平均　$\bar{z} = a_0 + \bar{x}_1 a_1 + \bar{x}_2 a_2$

全変動 SS_T = 群間変動 SS_B + 群内変動 SS_W

$$SS_T = \sum_{i=1}^{2}\sum_{j=1}^{n_i}\left(z_j^{(i)} - \bar{z}\right)^2 = \sum_{i=1}^{2}\sum_{j=1}^{n_i}\left\{\left(x_{1j}^{(i)} - \bar{x}_1\right)a_1 + \left(x_{2j}^{(i)} - \bar{x}_2\right)a_2\right\}^2$$

$$SS_B = \sum_{i=1}^{2}\sum_{j=1}^{n_i}\left(\bar{z}^{(i)} - \bar{z}\right)^2 = \sum_{i=1}^{2}\sum_{j=1}^{n_i}\left\{\left(\bar{x}_1^{(i)} - \bar{x}_1\right)a_1 + \left(\bar{x}_2^{(i)} - \bar{x}_2\right)a_2\right\}^2$$

$$SS_W = \sum_{i=1}^{2}\sum_{j=1}^{n_i}\left(z_j^{(i)} - \bar{z}^{(i)}\right)^2 = \sum_{i=1}^{2}\sum_{j=1}^{n_i}\left\{\left(x_{1j}^{(i)} - \bar{x}_1^{(i)}\right)a_1 + \left(x_{2j}^{(i)} - \bar{x}_2^{(i)}\right)a_2\right\}^2$$

2群の分離を良くするには群間変動SS_B/全変動SS_T（=相関比）を最大にする

$F(a_1, a_2) = \dfrac{SS_B}{SS_T}$ の最大化 ⇔ $G(a_1, a_2) = \dfrac{SS_W}{SS_B}$ の最小化

$$F(a_1, a_2) = \frac{SS_B}{SS_T} = \frac{SS_B}{SS_B + SS_W} = \frac{1}{1 + \dfrac{SS_W}{SS_B}} = \frac{1}{1 + G(a_1, a_2)}$$

別解　ベクトル表現　$\boldsymbol{x}_j{}^{(i)} = \begin{pmatrix} x_{1j}{}^{(i)} \\ x_{2j}{}^{(i)} \end{pmatrix}$　$\overline{\boldsymbol{x}}^{(i)} = \begin{pmatrix} \bar{x}_1^{(i)} \\ \bar{x}_2^{(i)} \end{pmatrix}$　$\overline{\boldsymbol{x}} = \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \end{pmatrix}$　$\boldsymbol{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$

$$SS_B = \sum_{i=1}^{2}\sum_{j=1}^{n_i}\left(\bar{z}^{(i)} - \bar{z}\right)^2 = \sum_{i=1}^{2}\sum_{j=1}^{n_i}\left\{\left(\bar{x}_1^{(i)} - \bar{x}_1\right)a_1 + \left(\bar{x}_2^{(i)} - \bar{x}_2\right)a_2\right\}^2 = \sum_{i=1}^{2}\sum_{j=1}^{n_i}\boldsymbol{a}'\left(\overline{\boldsymbol{x}}^{(i)} - \overline{\boldsymbol{x}}\right)\left(\overline{\boldsymbol{x}}^{(i)} - \overline{\boldsymbol{x}}\right)'\boldsymbol{a}$$

$$SS_W = \sum_{i=1}^{2}\sum_{j=1}^{n_i}\left(z_j^{(i)} - \bar{z}^{(i)}\right)^2 = \sum_{i=1}^{2}\sum_{j=1}^{n_i}\left\{\left(x_{1j}^{(i)} - \bar{x}_1^{(i)}\right)a_1 + \left(x_{2j}^{(i)} - \bar{x}_2^{(i)}\right)a_2\right\}^2 = \sum_{i=1}^{2}\sum_{j=1}^{n_i}\boldsymbol{a}'\left(\boldsymbol{x}_j{}^{(i)} - \overline{\boldsymbol{x}}^{(i)}\right)\left(\boldsymbol{x}_j{}^{(i)} - \overline{\boldsymbol{x}}^{(i)}\right)'\boldsymbol{a}$$

偏微分=0

$$\frac{\partial G}{\partial \boldsymbol{a}} = \frac{1}{SS_B}\left(\frac{\partial SS_W}{\partial \boldsymbol{a}} - \frac{SS_W}{SS_B}\frac{\partial SS_B}{\partial \boldsymbol{a}}\right) = \begin{pmatrix} \dfrac{\partial G}{\partial a_1} \\ \dfrac{\partial G}{\partial a_2} \end{pmatrix} = \frac{1}{SS_B}\left\{\begin{pmatrix} \dfrac{\partial SS_W}{\partial a_1} \\ \dfrac{\partial SS_W}{\partial a_2} \end{pmatrix} - \frac{SS_W}{SS_B}\begin{pmatrix} \dfrac{\partial SS_B}{\partial a_1} \\ \dfrac{\partial SS_B}{\partial a_2} \end{pmatrix}\right\} = 0$$

$$\therefore \ \frac{\partial SS_W}{\partial \boldsymbol{a}} = G\frac{\partial SS_B}{\partial \boldsymbol{a}} \ \Rightarrow \ \boxed{\sum_{i=1}^{2}\sum_{j=1}^{n_i}\left(\boldsymbol{x}_j{}^{(i)} - \overline{\boldsymbol{x}}^{(i)}\right)\left(\boldsymbol{x}_j{}^{(i)} - \overline{\boldsymbol{x}}^{(i)}\right)'\boldsymbol{a} = G\sum_{i=1}^{2}\sum_{j=1}^{n_i}\left(\overline{\boldsymbol{x}}^{(i)} - \overline{\boldsymbol{x}}\right)\left(\overline{\boldsymbol{x}}^{(i)} - \overline{\boldsymbol{x}}\right)'\boldsymbol{a}}$$

$$\frac{\partial SS_W}{\partial \boldsymbol{a}} = 2\sum_{i=1}^{2}\sum_{j=1}^{n_i}\left(\boldsymbol{x}_j{}^{(i)} - \overline{\boldsymbol{x}}^{(i)}\right)\left(\boldsymbol{x}_j{}^{(i)} - \overline{\boldsymbol{x}}^{(i)}\right)'\boldsymbol{a}$$

$$\frac{\partial SS_W}{\partial a_1} = 2\sum_{i=1}^{2}\sum_{j=1}^{n_i}\left(x_{1j}^{(i)} - \bar{x}_1^{(i)}\right)\left\{\left(x_{1j}^{(i)} - \bar{x}_1^{(i)}\right)a_1 + \left(x_{2j}^{(i)} - \bar{x}_2^{(i)}\right)a_2\right\}$$

$$\frac{\partial SS_W}{\partial a_2} = 2\sum_{i=1}^{2}\sum_{j=1}^{n_i}\left(x_{2j}^{(i)} - \bar{x}_2^{(i)}\right)\left\{\left(x_{1j}^{(i)} - \bar{x}_1^{(i)}\right)a_1 + \left(x_{2j}^{(i)} - \bar{x}_2^{(i)}\right)a_2\right\}$$

$$\frac{\partial SS_B}{\partial \boldsymbol{a}} = 2\sum_{i=1}^{2}\sum_{j=1}^{n_i}\left(\overline{\boldsymbol{x}}^{(i)} - \overline{\boldsymbol{x}}\right)\left(\overline{\boldsymbol{x}}^{(i)} - \overline{\boldsymbol{x}}\right)'\boldsymbol{a}$$

$$\frac{\partial SS_B}{\partial a_1} = 2\sum_{i=1}^{2}\sum_{j=1}^{n_i}\left(\bar{x}_1^{(i)} - \bar{x}_1\right)\left\{\left(\bar{x}_1^{(i)} - \bar{x}_1\right)a_1 + \left(\bar{x}_2^{(i)} - \bar{x}_2\right)a_2\right\}$$

$$\frac{\partial SS_B}{\partial a_2} = 2\sum_{i=1}^{2}\sum_{j=1}^{n_i}\left(\bar{x}_2^{(i)} - \bar{x}_2\right)\left\{\left(\bar{x}_1^{(i)} - \bar{x}_1\right)a_1 + \left(\bar{x}_2^{(i)} - \bar{x}_2\right)a_2\right\}$$

別解

前ページより $\displaystyle \sum_{i=1}^{2}\sum_{j=1}^{n_i}\left(\boldsymbol{x_j}^{(i)}-\overline{\boldsymbol{x}}^{(i)}\right)\left(\boldsymbol{x_j}^{(i)}-\overline{\boldsymbol{x}}^{(i)}\right)'\boldsymbol{a}=G\sum_{i=1}^{2}\sum_{j=1}^{n_i}\left(\overline{\boldsymbol{x}}^{(i)}-\overline{\boldsymbol{x}}\right)\left(\overline{\boldsymbol{x}}^{(i)}-\overline{\boldsymbol{x}}\right)'\boldsymbol{a}$

左辺

$$\sum_{i=1}^{2}\sum_{j=1}^{n_i}\left(\boldsymbol{x_j}^{(i)}-\overline{\boldsymbol{x}}^{(i)}\right)\left(\boldsymbol{x_j}^{(i)}-\overline{\boldsymbol{x}}^{(i)}\right)'\boldsymbol{a}=\sum_{i=1}^{2}\sum_{j=1}^{n_i}\begin{pmatrix}\left(x_{1j}^{(i)}-\bar{x}_1^{(i)}\right)^2 & \left(x_{1j}^{(i)}-\bar{x}_1^{(i)}\right)\left(x_{2j}^{(i)}-\bar{x}_2^{(i)}\right)\\ \left(x_{2j}^{(i)}-\bar{x}_1^{(i)}\right)\left(x_{1j}^{(i)}-\bar{x}_1^{(i)}\right) & \left(x_{2j}^{(i)}-\bar{x}_2^{(i)}\right)^2\end{pmatrix}\begin{pmatrix}a_1\\a_2\end{pmatrix}$$

$$=\sum_{i=1}^{2}(n_i-1)\begin{pmatrix}Var\left(x_1^{(i)}\right) & Cov\left(x_1^{(i)},x_2^{(i)}\right)\\ Cov\left(x_2^{(i)},x_1^{(i)}\right) & Var\left(x_2^{(i)}\right)\end{pmatrix}\begin{pmatrix}a_1\\a_2\end{pmatrix}=\boldsymbol{S_W a}$$

右辺

$$G\sum_{i=1}^{2}\sum_{j=1}^{n_i}\left(\overline{\boldsymbol{x}}^{(i)}-\overline{\boldsymbol{x}}\right)\left(\overline{\boldsymbol{x}}^{(i)}-\overline{\boldsymbol{x}}\right)'\boldsymbol{a}=G\left\{n_1\left(\overline{\boldsymbol{x}}^{(1)}-\overline{\boldsymbol{x}}\right)\left(\overline{\boldsymbol{x}}^{(1)}-\overline{\boldsymbol{x}}\right)'+n_2\left(\overline{\boldsymbol{x}}^{(2)}-\overline{\boldsymbol{x}}\right)\left(\overline{\boldsymbol{x}}^{(2)}-\overline{\boldsymbol{x}}\right)'\right\}\boldsymbol{a}$$

$$=G\left(\frac{n_1n_2}{n_1+n_2}\right)^2\left(\frac{1}{n_1}-\frac{1}{n_2}\right)\left(\overline{\boldsymbol{x}}^{(1)}-\overline{\boldsymbol{x}}^{(2)}\right)\left(\overline{\boldsymbol{x}}^{(1)}-\overline{\boldsymbol{x}}^{(2)}\right)'\boldsymbol{a}=k\left(\overline{\boldsymbol{x}}^{(1)}-\overline{\boldsymbol{x}}^{(2)}\right)$$

ただし

$$\overline{\boldsymbol{x}}=\frac{n_1\overline{\boldsymbol{x}}^{(1)}+n_2\overline{\boldsymbol{x}}^{(2)}}{n_1+n_2}$$

$$k=G\left(\frac{n_1n_2}{n_1+n_2}\right)^2\left(\frac{1}{n_1}-\frac{1}{n_2}\right)\left(\overline{\boldsymbol{x}}^{(1)}-\overline{\boldsymbol{x}}^{(2)}\right)'\boldsymbol{a}=G\left(\frac{n_1n_2}{n_1+n_2}\right)^2\left(\frac{1}{n_1}-\frac{1}{n_2}\right)\left\{\left(\bar{x}_1^{(1)}-\bar{x}_1^{(2)}\right)a_1+\left(\bar{x}_2^{(1)}-\bar{x}_2^{(2)}\right)a_2\right\}$$

したがって

$$\boldsymbol{S_W a}=k\left(\overline{\boldsymbol{x}}^{(1)}-\overline{\boldsymbol{x}}^{(2)}\right)$$

$k=1$として

$$\boldsymbol{a}=\boldsymbol{S_W}^{-1}\left(\overline{\boldsymbol{x}}^{(1)}-\overline{\boldsymbol{x}}^{(2)}\right)$$

以上より

$$\begin{pmatrix}a_1\\a_2\end{pmatrix}=\boldsymbol{S_W}^{-1}\begin{pmatrix}\bar{x}_1^{(1)}-\bar{x}_1^{(2)}\\ \bar{x}_2^{(1)}-\bar{x}_2^{(2)}\end{pmatrix}$$

別解

G1（A型）

患者＼説明変数	楽天家 No $x_1^{(1)}$	堅実派 No $x_2^{(1)}$
1	1	0
2	1	1
3	1	0
4	0	1
$\bar{x}^{(1)}$	0. 750	0.500
Var	0.250	0.333
Cov	-0.167	

G2（B型）

患者＼説明変数	楽天家 No $x_2^{(2)}$	堅実派 No $x_2^{(2)}$
1	0	1
2	0	0
3	0	1
$\bar{x}^{(2)}$	0	0.667
Var	0	0.333
Cov	0	

全体総平均

	楽天家 No $\bar{x}_1$	堅実派 No $\bar{x}_2$
$\bar{x}$	0.429	0.571

$$\boldsymbol{S_W} = \sum_{i=1}^{2}(n_i - 1)\begin{pmatrix} Var\left(x_1^{(i)}\right) & Cov\left(x_1^{(i)}, x_2^{(i)}\right) \\ Cov\left(x_2^{(i)}, x_1^{(i)}\right) & Var\left(x_2^{(i)}\right) \end{pmatrix} = 3\begin{pmatrix} 0.250 & -0.167 \\ -0.167 & 0.333 \end{pmatrix} + 2\begin{pmatrix} 0.000 & 0.000 \\ 0.000 & 0.333 \end{pmatrix}$$

$$= \begin{pmatrix} 0.75 & -0.5 \\ -0.5 & 1.667 \end{pmatrix}$$

$$\begin{pmatrix} \bar{x}_1^{(1)} - \bar{x}_1^{(2)} \\ \bar{x}_2^{(1)} - \bar{x}_2^{(2)} \end{pmatrix} = \begin{pmatrix} 0.750 - 0.000 \\ 0.500 - 0.667 \end{pmatrix} = \begin{pmatrix} 0.750 \\ -0.167 \end{pmatrix}$$

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \boldsymbol{S_W}^{-1}\begin{pmatrix} \bar{x}_1^{(1)} - \bar{x}_1^{(2)} \\ \bar{x}_2^{(1)} - \bar{x}_2^{(2)} \end{pmatrix} = \begin{pmatrix} 1.667 & 0.500 \\ 0.500 & 0.7501 \end{pmatrix}\begin{pmatrix} 0.750 \\ -0.167 \end{pmatrix} = \begin{pmatrix} 1.167 \\ 0.250 \end{pmatrix}$$

総平均では判別得点$\boldsymbol{z} = \boldsymbol{0}$となる　　$a_0 = -a_1\bar{x}_1 - a_2\bar{x}_2 = -0.623$

以上より　$z = a_1x_1 + a_2x_2 + a_0 = 1.167x_1 + 0.25x_2 - 0.623$

$z>0$ ････グループ1（A型）

$z<0$ ････グループ2（B型）

先ほどの判別式　$Y = 4.667x_{12} + x_{22} - 2.571$

$Y = 4z$　正負を判定するため実質的に同等

数量化2類

① 数量化2類とは何か
② アンケート調査で下記のデータを得た。
各アイテムのカテゴリの反応から外的基準の判別を行う式を1次式で表現せよ。
判別式：Y 、ダミー変数： x_{ij}、カテゴリ数量： a_{ij}、定数項：a_0を用いよ。
③ カテゴリ数量 a_{ij}と定数項a_0の決定方法を述べよ。
④ カテゴリー数量の基準化を行い予測式を求めよ。
⑤ アイテムの外的基準に与える影響の大きさを検討せよ。

アンケート調査票

項目1
あなたは楽天家ですか
1 はい　2 いいえ

項目2
あなたは堅実派だと思いますか
1 はい　2 いいえ

項目3
あなたの血液型は
1 A型　2 B型

アイテム　カテゴリ　外的基準

被験者No.	楽天家		堅実派		血液型
	1	2	1	2	
1		レ	レ		A型
2		レ		レ	
3		レ	レ		
4	レ			レ	
1	レ			レ	B型
2	レ		レ		
3	レ			レ	