入門 統計学　第13章

ロジスティック回帰分析とクラスター分析
― 多変量解析② ―

『入門　統計学　第2版　ー検定から多変量解析・実験計画法・ベイズ統計学までー』（オーム社）

※注：本書を購入された方へのサービスですので，教科書指定（参考図書は不可）していない授業での使用はお控えください。

13.1　ロジスティック回帰分析

- 結果（従属変数）が2値（ダミー変数）の回帰分析

→結果の値よりも，買う・買わない，壊れる・壊れない，病気・病気でない，の「どちらなのか」を知りたいことが多い

$y(1/0) = \alpha + \beta x_1 + \beta x_2 + \cdots$

普通の回帰分析は量的な結果しか扱えない

- 回帰モデルが特定できれば，未知の個体がどちらの群に分類されるかを確率で予測可

いろいろな離散選択モデル

（従属変数が質的尺度）

- 多項ロジスティック/プロビット：

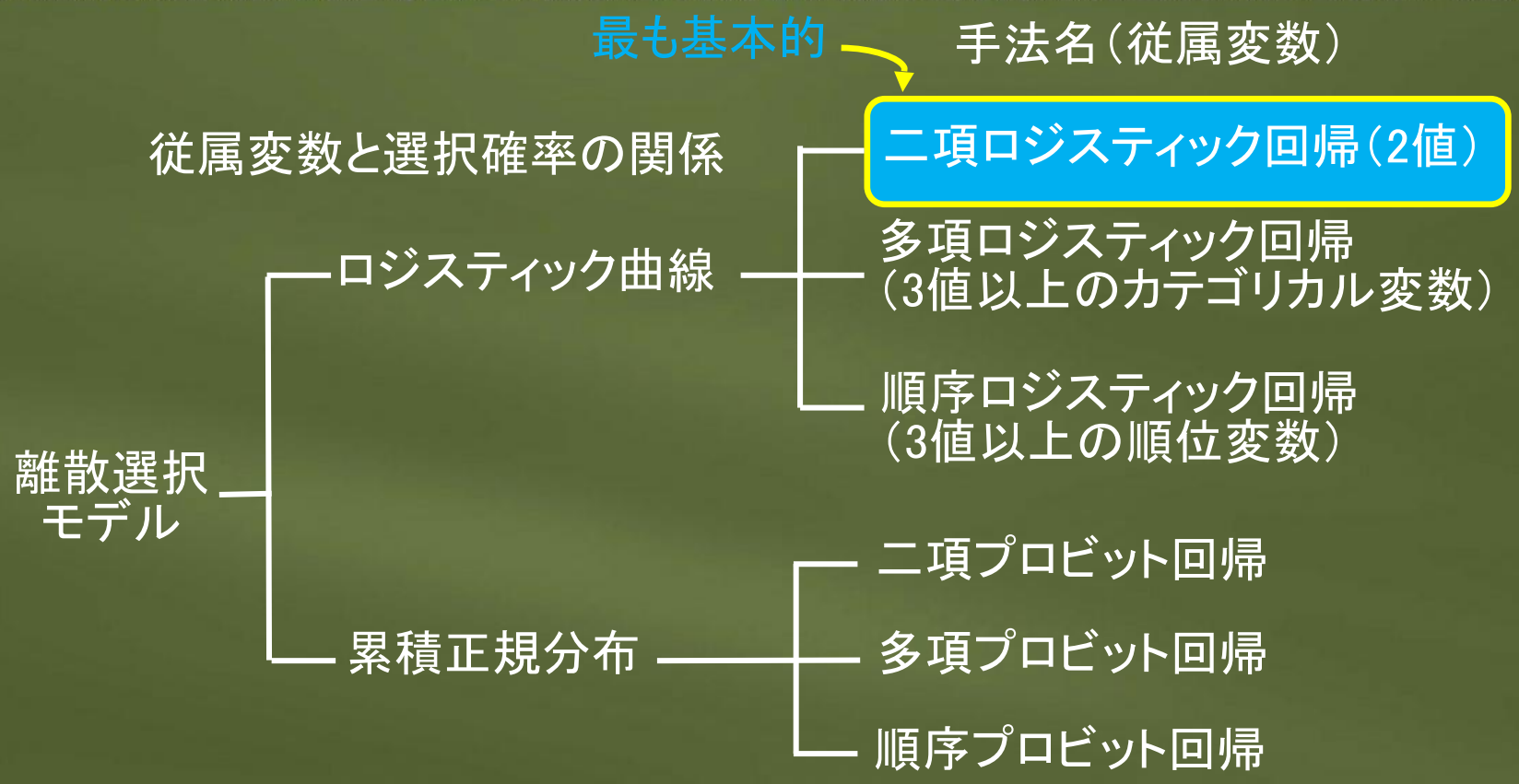
（消費者が購入するのは）ブランドA，ブランドB，ブランドC…といった従属変数がカテゴリカルなデータ

- 順序ロジスティック/プロビット：

（回答者の評価は）満足，やや満足，やや不満，不満など，従属変数が順位データ

注：誤差項の分布（従属変数と選択確率の関係）に想定する曲線によって，それぞれロジスティック回帰とプロビット回帰の2種類がある。

離散選択モデルの一覧


最も基本的
手法名（従属変数）
従属変数と選択確率の関係
離散選択モデル
ロジスティック曲線
累積正規分布
二項ロジスティック回帰（2値）
多項ロジスティック回帰
（3値以上のカテゴリカル変数）
順序ロジスティック回帰
（3値以上の順位変数）
二項プロビット回帰
多項プロビット回帰
順序プロビット回帰

2値データに線形回帰は適さない

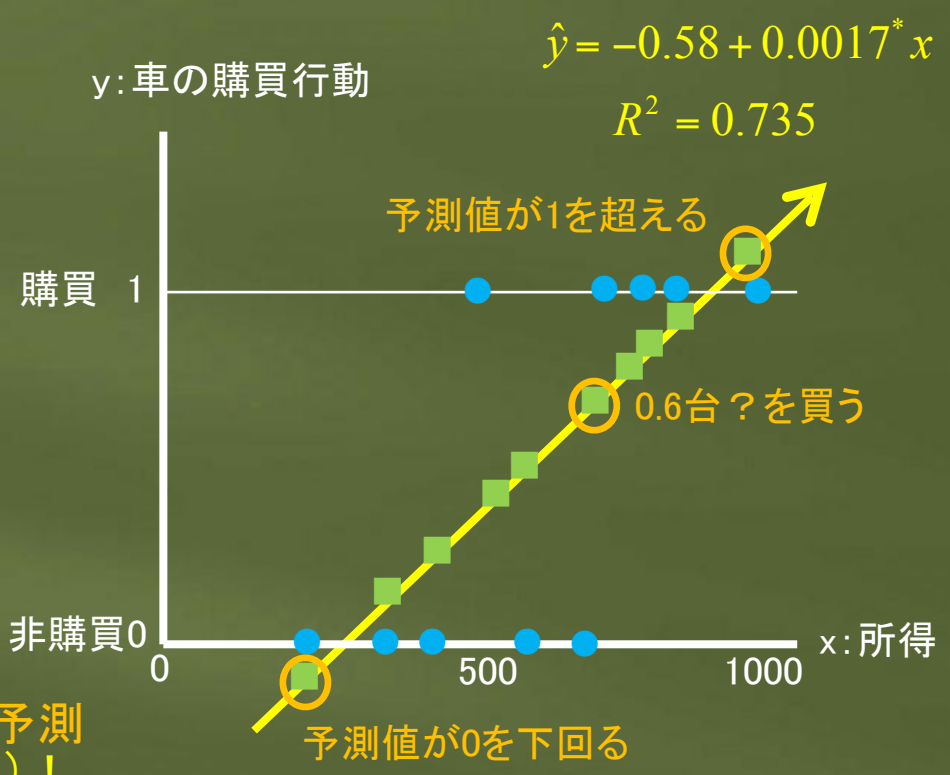
- 第12章で学んだ普通の重回帰モデル（線形回帰）は
  2値データに適さない
  →予測値がおかしくなるから（次に図示）

- 線形回帰モデルとは…
  ①図で言えば，従属変数と説明変数とが直線関係
  ②式で言えば，足し算の形になっている
  ③条件で言えば，データ（誤差項）が正規分布している

2値を線形回帰で予測すると…

所得と車の購買行動

	x: 所得 (万円/年)	y: 購買行動 (購入=1)	□:予測値
Aさん	300	0	-0.07
Bさん	400	0	0.09
Cさん	450	0	0.18
Dさん	550	0	0.35
Eさん	700	0	0.60
Fさん	500	1	0.26
Gさん	800	1	0.77
Hさん	850	1	0.85
Iさん	900	1	0.94
Jさん	950	1	1.02

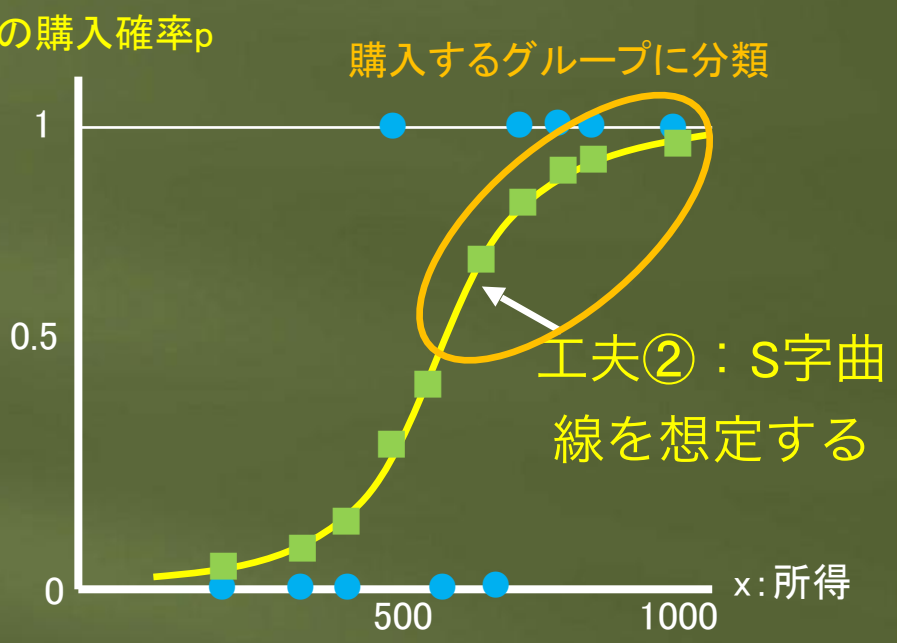


予測事例：所得2000万円の外挿予測
0.58+0.0017×2000=2.8（台）！

2つの工夫で0～1に収める

工夫①：y=1を選択する（事例では車を買う）確率を考える

	x: 所得	y: 購買行動	π:購入確率p
Aさん	300	0	0.03
Bさん	400	0	0.08
Cさん	450	0	0.12
Dさん	550	0	0.28
Eさん	700	0	0.65
Fさん	500	1	0.19
Gさん	800	1	0.84
Hさん	850	1	0.90
Iさん	900	1	0.94
Jさん	950	1	0.96



こちらの方が単純

注：S字曲線にはロジスティック曲線と累積正規分布関数の2種類がある

ロジスティック関数

i番目のデータy_iが，x_iのときに1を選択する確率

ロジスティック関数
（ロジスティック回帰モデル）

$$P(y_i = 1) = p_i = \frac{e^{\alpha + \beta x_i}}{1 + e^{\alpha + \beta x_i}}$$

パラメータは最尤法で推定（後述）

オッズに変形してパラメータを解釈しやすくする

1を選択する確率

オッズ $\dfrac{p_i}{1 - p_i}$

1を選択しない確率

$$= \frac{\dfrac{e^{\alpha + \beta x_i}}{1 + e^{\alpha + \beta x_i}}}{\dfrac{1 + e^{\alpha + \beta x_i}}{1 + e^{\alpha + \beta x_i}} - \dfrac{e^{\alpha + \beta x_i}}{1 + e^{\alpha + \beta x_i}}} = \frac{\dfrac{e^{\alpha + \beta x_i}}{1 + e^{\alpha + \beta x_i}}}{\dfrac{1}{1 + e^{\alpha + \beta x_i}}} = e^{\alpha + \beta x_i}$$

単純な形になる
（β の解釈法は…）

オッズ比（$\hat{\beta}$の解釈）

$\hat{\beta}$の解釈：x+1 （xが1単位増加）の，オッズに対する効果

オッズ

オッズ比
（xのオッズと比べたx+1のオッズ）

解釈には不要

$$\frac{\hat{p}_i}{1-\hat{p}_i} = e^{\hat{\alpha}+\hat{\beta}(x_i+1)} = e^{\hat{\alpha}} \times e^{\beta(x_i+1)} \xrightarrow{\hat{\alpha}=0} e^{0} \times e^{\hat{\beta}(x_i+1)} = 1 \times e^{\beta(x_i+1)} = e^{\beta x_i} \times e^{\hat{\beta}}$$

オッズに対する乗法効果（オッズを●倍に変化させる）

事例：
pが車を買う確率で，所得のオッズ比が“2”と推定された場合
→所得xが1単位増加すると，車を買わない確率に対する買う確率の比が2倍になる。

ロジット変換による一般化線形モデル

オッズ $\dfrac{p_i}{1-p_i} = e^{\alpha+\beta x_i}$

↓ ロジット変換（対数をとる）

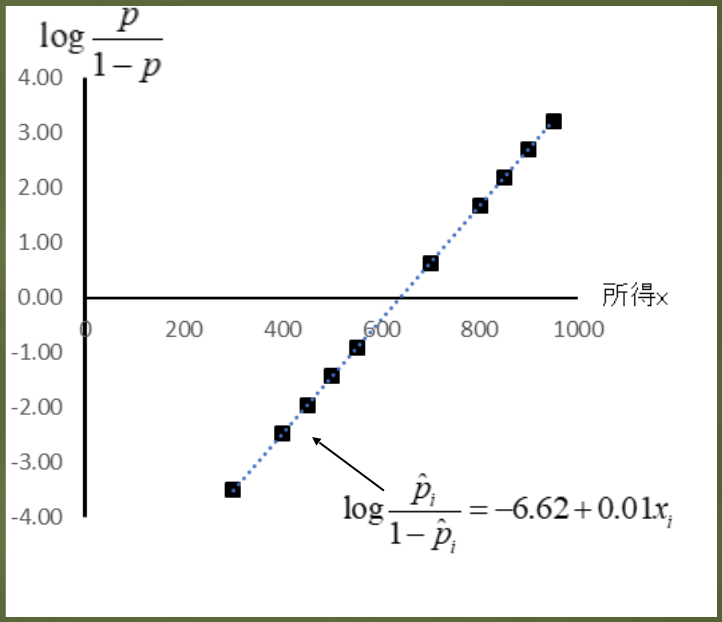
ロジットモデル $\log\dfrac{p_i}{1-p_i} = \alpha+\beta x_i$

対数オッズ（確率pの関数と見なせばロジット関数）

線形モデルになる（一般化線形モデル）

カテゴリカルな従属変数の各水準の確率を連続尺度に橋渡しするリンク関数の1つ

車の購入確率の事例をロジット変換



13.2　パラメータの推定

- ロジスティック関数のパラメータ推定にOLSは適さない

→OLSは結果（誤差）が連続値で正規分布の場合用

→ロジスティック関数は離散値でベルヌーイ分布

（試行回数n=1の二項分布）

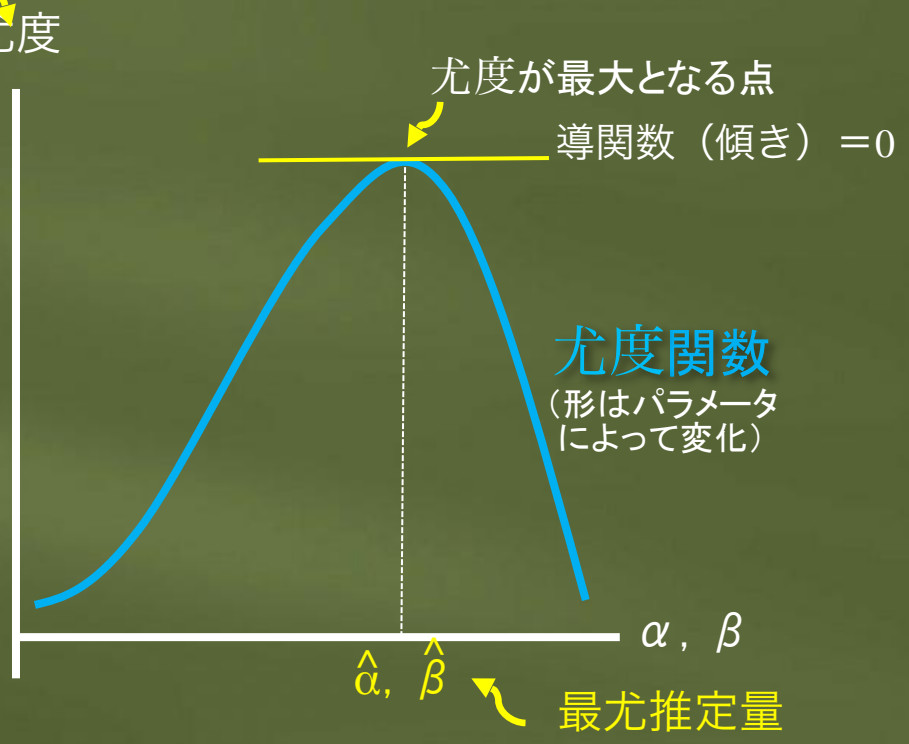
最尤法（さいゆうほう）

maximum likelihood estimation (MLE)

- 手元にある観測データから，それが従う確率分布のパラメータを点推定
- 一番，尤（もっと）もらしくデータを説明できるパラメータを見つける

尤度（もっともらしさ）

データ全体の得られやすさ
（個別データの生起確率を
全て掛け合わせたもの）



尤度関数

（2値変数が従うベルヌーイ分布の場合）

2値変数：1になるか0になるかのどちらか

↓

試行回数が1の二項分布であるベルヌーイ分布に従う

i番目のデータyiは，1か0のどちらかをとる（piか，1−piになる）

ベルヌーイ分布
の確率質量関数　$P(y_i) = p_i^{y_i}(1-p_i)^{1-y_i}$

↓ 確率を掛け合わせる（Πは総乗記号）

尤度関数　$L(p) = \prod_{i=1}^{n} P(y_i) = p_1^{y_1}(1-p_1)^{1-y_1} \cdot p_2^{y_2}(1-p_2)^{1-y_2} \cdot \cdots\cdots \cdot p_n^{y_n}(1-p_n)^{1-y_n}$

尤度計算事例（ロジスティック関数）①

観測データが3つで，$y_1=1$，$y_2=0$，$y_3=1$の場合の尤度：

$$L(p) = p_1^1(1-p_1)^0 \cdot p_2^0(1-p_2)^1 \cdot p_3^1(1-p_3)^0 = p_1 \cdot (1-p_2) \cdot p_3$$

p_iはロジスティック関数で予測される確率 $p_i = \dfrac{e^{\alpha+\beta x_i}}{1+e^{\alpha+\beta x_i}}$

$$L(p) = \frac{e^{\alpha+\beta x_1}}{1+e^{\alpha+\beta x_1}} \cdot \left(1 - \frac{e^{\alpha+\beta x_2}}{1+e^{\alpha+\beta x_2}}\right) \cdot \frac{e^{\alpha+\beta x_3}}{1+e^{\alpha+\beta x_3}} = \frac{e^{\alpha+\beta x_1}}{1+e^{\alpha+\beta x_1}} \cdot \frac{1}{1+e^{\alpha+\beta x_2}} \cdot \frac{e^{\alpha+\beta x_3}}{1+e^{\alpha+\beta x_3}}$$

乗算のままだと扱い難いので自然対数をとって足し算の形（対数尤度）にする

尤度計算事例（ロジスティック関数）②

対数尤度 $\log\left(L(p)\right) = \log\left(\frac{e^{\alpha+\beta x_1}}{1+e^{\alpha+\beta x_1}}\right) + \log\left(\frac{1}{1+e^{\alpha+\beta x_2}}\right) + \log\left(\frac{e^{\alpha+\beta x_3}}{1+e^{\alpha+\beta x_3}}\right)$

$= \left\{\log(e^{\alpha+\beta x_1}) - \log(1+e^{\alpha+\beta x_1})\right\} - \log(1+e^{\alpha+\beta x_2}) + \left\{\log(e^{\alpha+\beta x_3}) - \log(1+e^{\alpha+\beta x_3})\right\}$

$= (\alpha+\beta x_1) - \log(1+e^{\alpha+\beta x_1}) - \log(1+e^{\alpha+\beta x_2}) + (\alpha+\beta x_3) - \log(1+e^{\alpha+\beta x_3})$

$= -\log(1+e^{\alpha+\beta x_1}) - \log(1+e^{\alpha+\beta x_2}) - \log(1+e^{\alpha+\beta x_3}) + 2\alpha + \beta(x_1+x_3)$

↓ α と β で偏微分

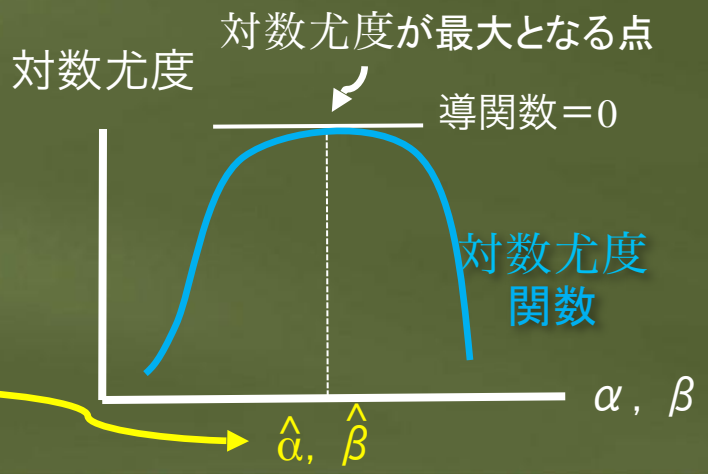
$$\begin{cases} \dfrac{\partial \log L}{\partial \alpha} = -\dfrac{e^{\alpha+\beta x_1}}{1+e^{\alpha+\beta x_1}} - \dfrac{e^{\alpha+\beta x_2}}{1+e^{\alpha+\beta x_2}} - \dfrac{e^{\alpha+\beta x_3}}{1+e^{\alpha+\beta x_3}} + 2 \\ \dfrac{\partial \log L}{\partial \beta} = -\dfrac{x_1 e^{\alpha+\beta x_1}}{1+e^{\alpha+\beta x_1}} - \dfrac{x_2 e^{\alpha+\beta x_2}}{1+e^{\alpha+\beta x_2}} - \dfrac{x_2 e^{\alpha+\beta x_3}}{1+e^{\alpha+\beta x_3}} + (x_1+x_3) \end{cases}$$

尤度計算事例（ロジスティック関数）③

それぞれを=0と置けば尤度方程式を得る

$$\begin{cases} -\dfrac{e^{\hat{\alpha}+\hat{\beta}x_1}}{1+e^{\hat{\alpha}+\hat{\beta}x_1}} - \dfrac{e^{\alpha+\beta x_2}}{1+e^{\alpha+\beta x_2}} - \dfrac{e^{\hat{\alpha}+\hat{\beta}x_3}}{1+e^{\alpha+\beta x_3}} + 2 = 0 \\ -\dfrac{x_1 e^{\hat{\alpha}+\hat{\beta}x_1}}{1+e^{\hat{\alpha}+\hat{\beta}x_1}} - \dfrac{x_2 e^{\alpha+\beta x_2}}{1+e^{\alpha+\beta x_2}} - \dfrac{x_3 e^{\hat{\alpha}+\hat{\beta}x_3}}{1+e^{\alpha+\beta x_3}} + (x_1+x_3) = 0 \end{cases}$$

この連立方程式を解けば最尤推定量 $\hat{\alpha}$ と $\hat{\beta}$ が求まる



現実には複雑過ぎて代数的に解けないため，適当な初期値から少しずつ値を変えて最尤推定量を探し出す（ニュートン法）。

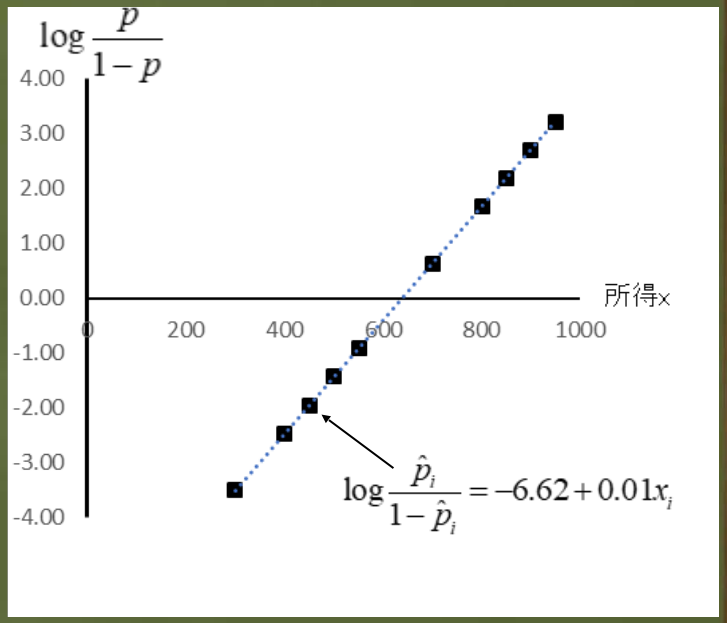
ロジット変換後ならOLSが使える？

ロジット変換後のロジットモデル

$$\log \frac{p_i}{1-p_i} = \alpha + \beta x_i$$

質問：ロジット変換した後なら線形なので，OLSが使えるのでは？

回答：pが1や0のとき，オッズの対数が計算できないので適さない。ただし，2値データを集計した割合データならば0はないため可能（分母を同じにすることに注意）。
→病院毎の発症割合や，工場毎の不良品割合のデータなど。

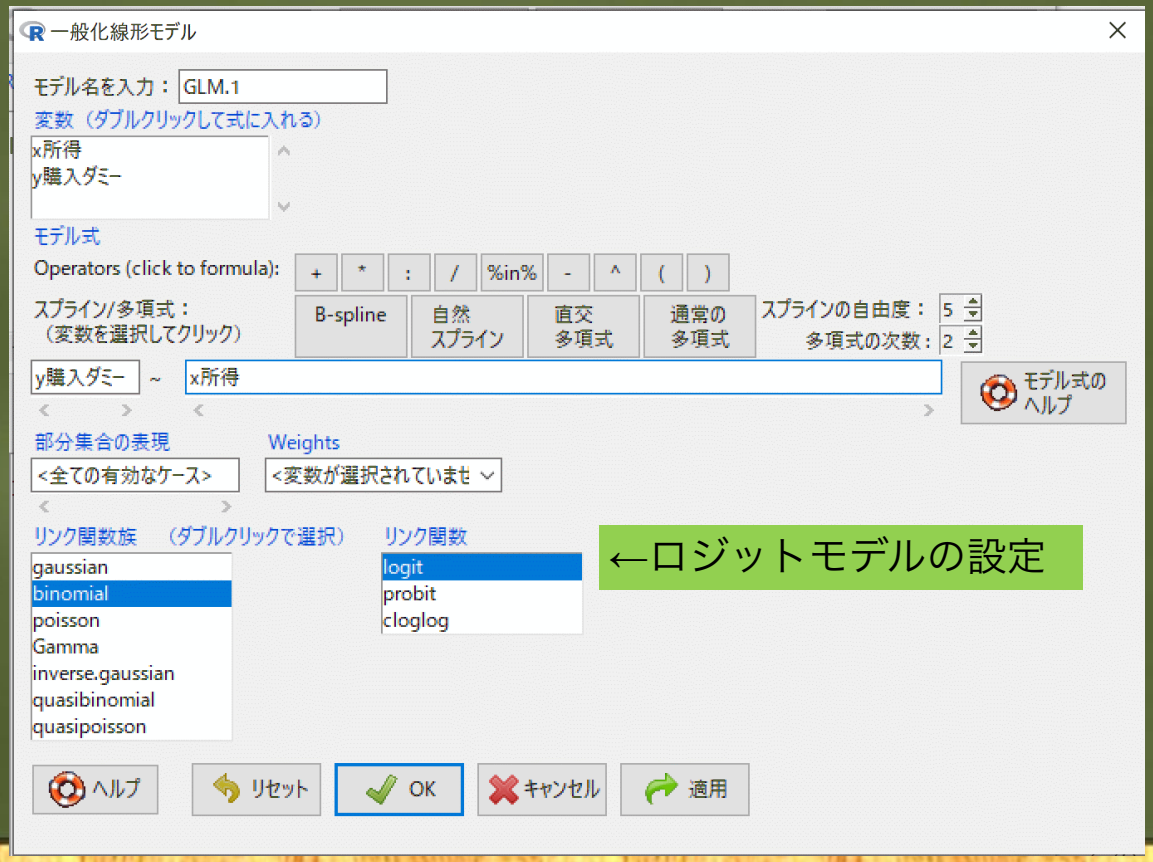


ソフト（Rコマンダー）による分析

メニュー[統計量]→[モデルへの適合]→[一般化線形モデル]を選択

従属変数と説明変数を設定→

二項分布→



ソフト（Rコマンダー）の出力

```
Call:
glm(formula = y購入ダミー ~ x所得, family = binomial(logit), data = Dataset)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.4555  -0.4861   0.0184   0.4358   1.8175

Coefficients:
             Estimate   Std. Error   z value   Pr(>|z|)
(Intercept) -6.619136    3.685590    -1.796     0.0725 .
x所得        0.010361    0.005618     1.844     0.0652 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 13.8629  on 9  degrees of freedom
Residual deviance:  7.3422  on 8  degrees of freedom
AIC: 11.342

Number of Fisher Scoring iterations: 5


> exp(coef(GLM.1))  # Exponentiated coefficients ('odds ratios')
(Intercept)        x所得
0.001334583  1.010414674
```

推定されたロジスティック回帰モデル

$$\hat{p}_i = \frac{e^{-6.62+0.01x_i}}{1-e^{-6.62+0.01x_i}}$$

回帰係数の検定（この後解説）

予測能力（この後解説）

パラメータの解釈（オッズ比）

$$\frac{\hat{p}_i}{1-\hat{p}_i} = e^{0.01(x_i+1)} = e^{0.01x_i} \times e^{0.01} = e^{0.01x_i} \times 1.01$$

所得が1単位（万円）増加する度に，車を購入しない確率に対する購入する確率が1.01倍に増加する

13.3　推定モデルの評価①

赤池情報量規準(Akaike's Information Criterion；AIC)

- 決定係数は使えない：予測しているのは確率なので，観測値（1と0）の適合度を考えても無意味
- 最大対数尤度（logL）も適さない：モデルが複雑になる（説明変数が増える）だけで大きくなってしまう欠点

赤池情報量規準（AIC）：複雑さを考慮しつつ推定モデルの予測能力の“悪さ”を評価

パラメータの数k

$$\mathrm{AIC} = \boxed{-2\log L} + 2k$$

逸脱度（あてはまりの悪さ）

モデルの評価②

判別的中率

- AICは複数モデルを比較するのに適している

  →一番小さいAICのモデルを採用する

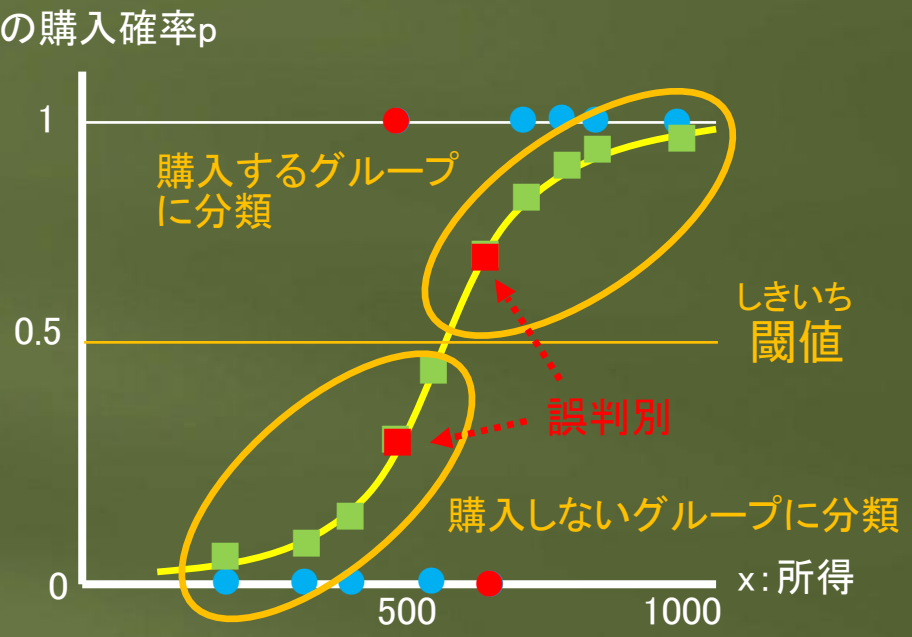
- 単独のモデル評価には判別的中率か疑似決定係数を用いる

0.5を閾値として分類

$$\text{判別的中率} = \frac{\text{正しく分類されたデータ数}}{\text{観測データ総数}} \times 100(\%)$$

判別的中率の計算事例

	x: 所得	y: 購買行動	𠮟:購入確率p
Aさん	300	0	0.03
Bさん	400	0	0.08
Cさん	450	0	0.12
Dさん	550	0	0.28
Eさん	700	0	0.65
Fさん	500	1	0.19
Gさん	800	1	0.84
Hさん	850	1	0.90
Iさん	900	1	0.94
Jさん	950	1	0.96



$$\text{判別的中率} = \frac{\text{正しく分類されたデータ数}}{\text{観測データ総数}} = \frac{8}{10} \times 100 = 80(\%)$$

モデルの評価③

疑似決定係数（Pseudo R^2)

- 対数尤度logLを使った擬似的な決定係数

推定されたモデルの対数尤度

McFaddenの疑似決定係数（マクファーデン）

$$R^2 = 1 - \frac{\log L_1}{\log L_0}$$

説明変数xを含めたことによる対数尤度の改善度合いを示す

切片（定数項）のみのモデルの対数尤度

Rコマンダーの出力結果からの計算：

通常の決定係数よりも低めに出る

$$McFadden's\ Pseudo\ R^2 = 1 - \frac{Residual\ deviance(推}{Null\ deviance(切片モ} = \frac{7.3422}{13.8629}$$

0.47

回帰係数の検定

$$\hat{p}_i = \frac{e^{\hat{\alpha}+\hat{\beta}x_i}}{1+e^{\hat{\alpha}+\hat{\beta}x_i}}$$

最尤推定量は漸近的に（大標本ならば）正規分布に従う性質を利用してz検定を実施（帰無仮説は$H_0: \beta = 0$）

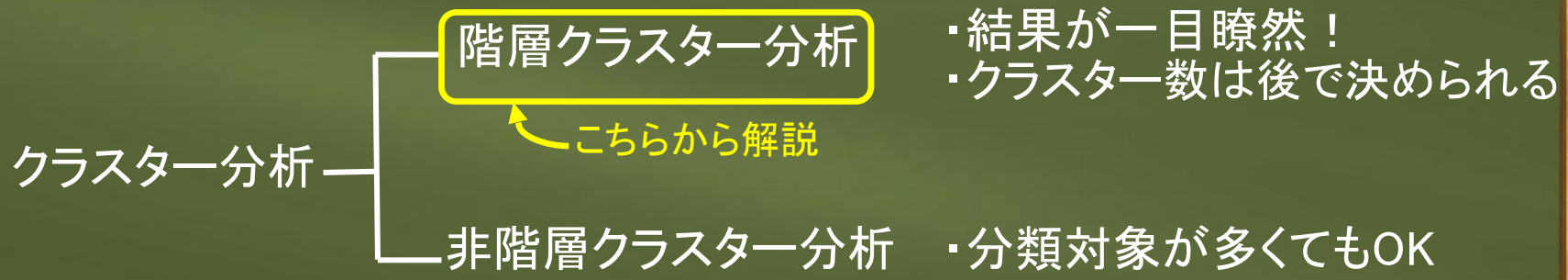
Rコマンダーで出力される

事例の回帰係数の検定統計量z＝$\frac{\text{回帰係数値}}{\text{標準誤差}}$＝1.844 （p値=0.0652）

注：最尤法は，OLSに比べて大きな標本を必要とする
→従属変数の1か0のうち少ない方の数が説明変数の数の10倍が最低ライン（事例ならばn=50）

13.4　クラスター分析

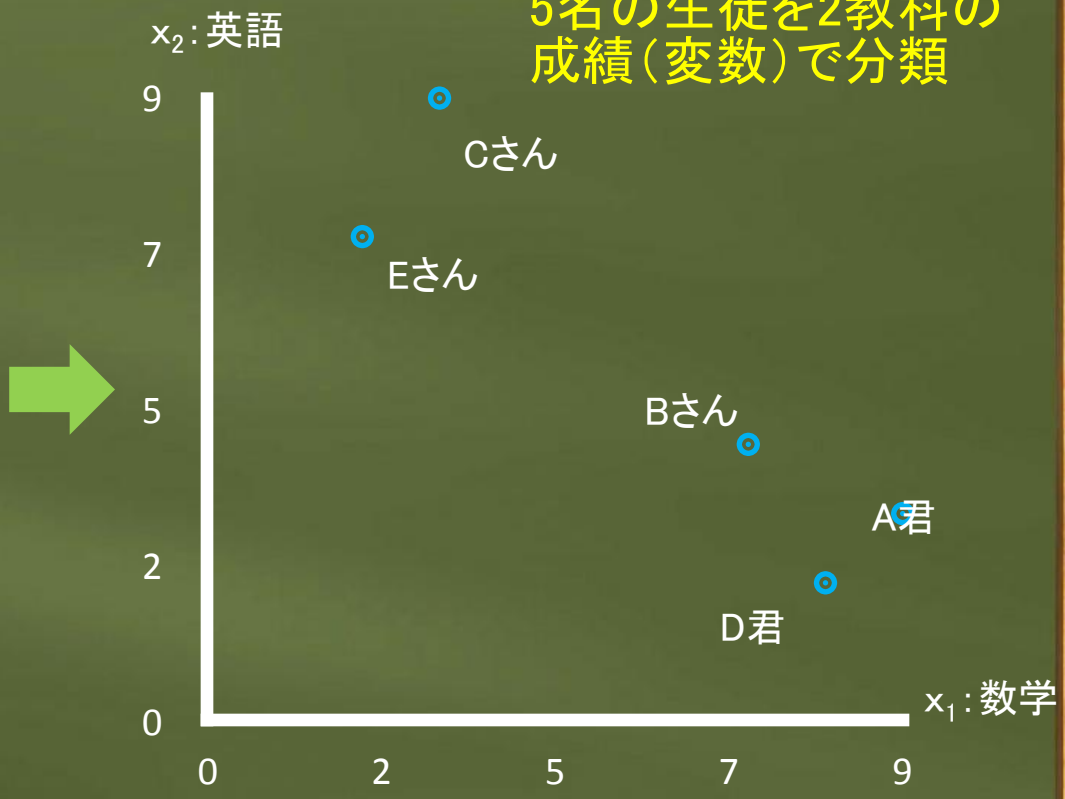
- ロジスティック回帰分析では，分類の手本（教師）が従属変数（買う/買わない）として与えられていた
- 教師なしの場合はどうする？　→クラスター分析



クラスター分析の事例

数学と英語の成績（10点満点）

生徒	x_1：数学	x_2：英語
A君	9	3
Bさん	7	4
Cさん	3	9
D君	8	2
Eさん	2	7

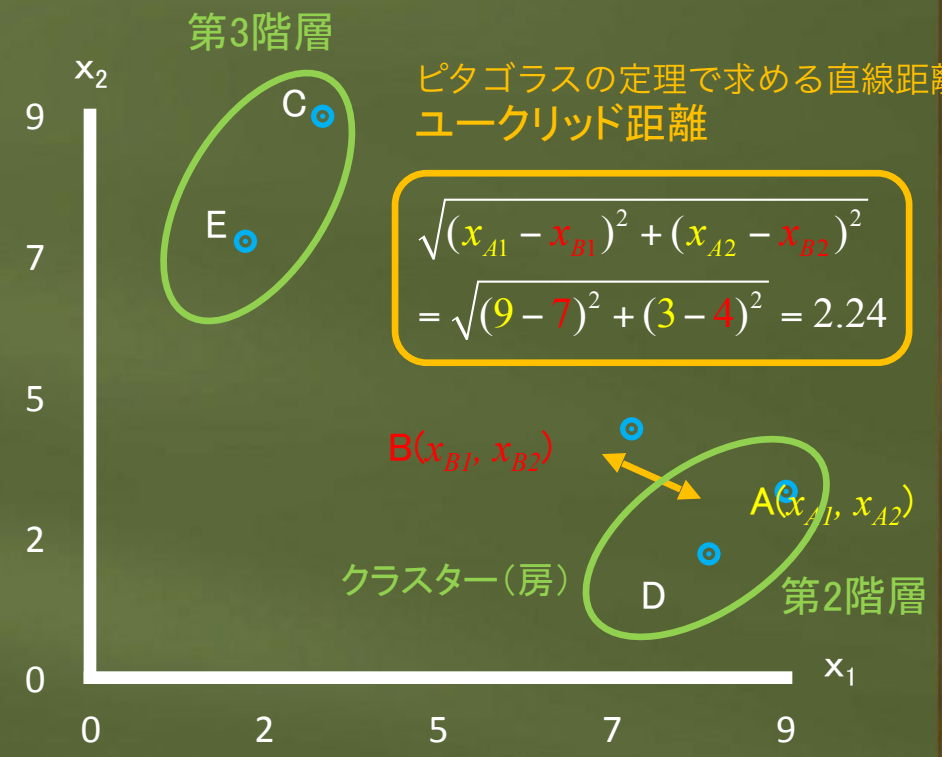


階層クラスター分析

手順①　対象間の距離

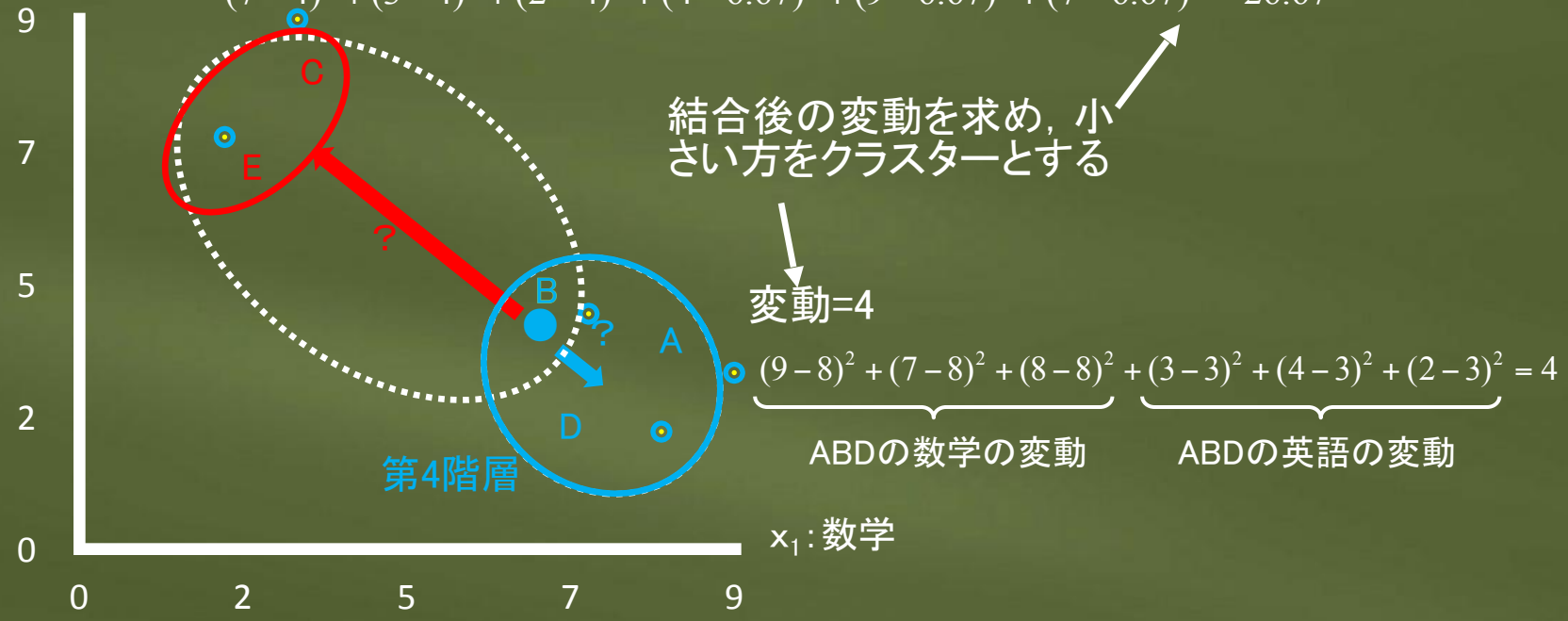
対象（個体）間の距離行列

	A	B	C	D	E
A					
B	2.24				
C	8.49	6.40			
D	1.41	2.24	8.60		
E	8.06	5.83	2.24	7.81	



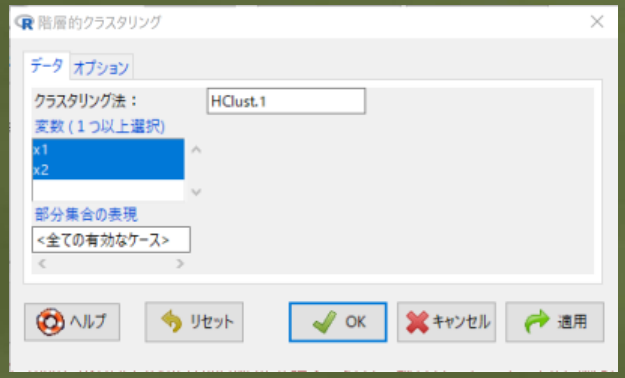
階層クラスター分析

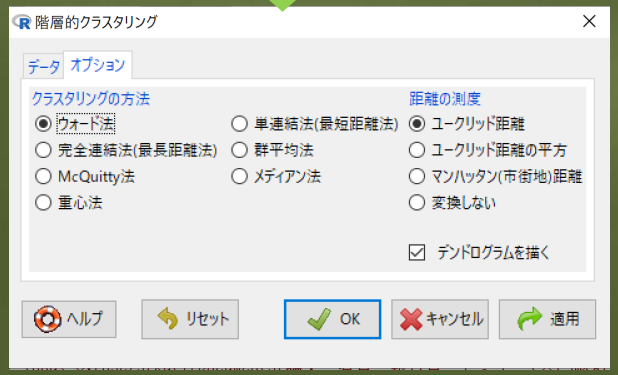
手順②　クラスター間の距離（ウォード法）

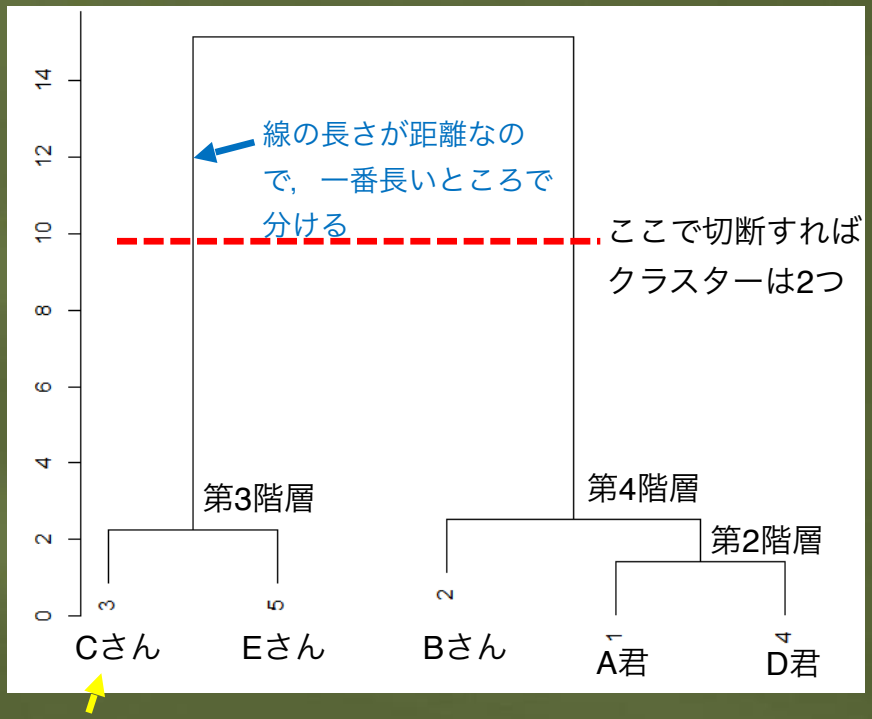


Rコマンダーで描く
デンドログラム（樹形図）

［統計量］→［次元解析］→［クラスタ分析］→［階層的クラスタ分析］

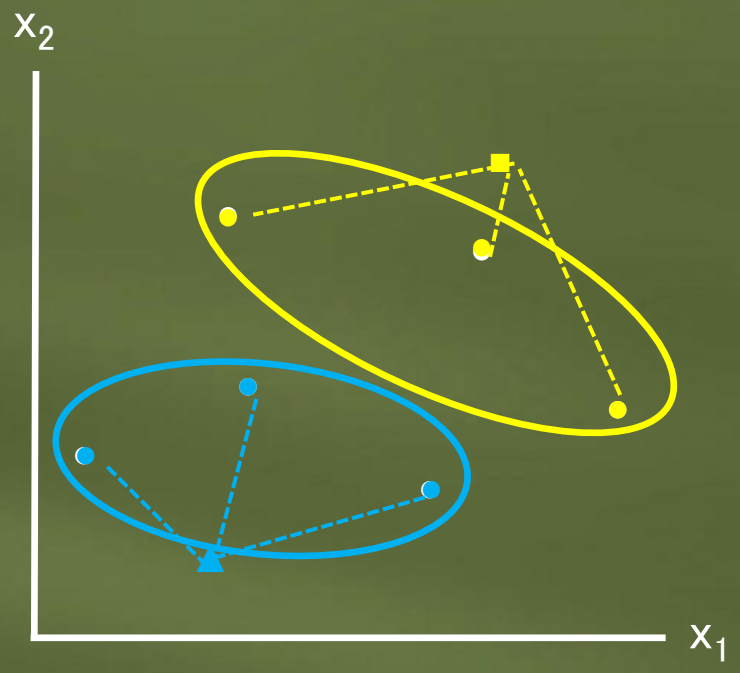






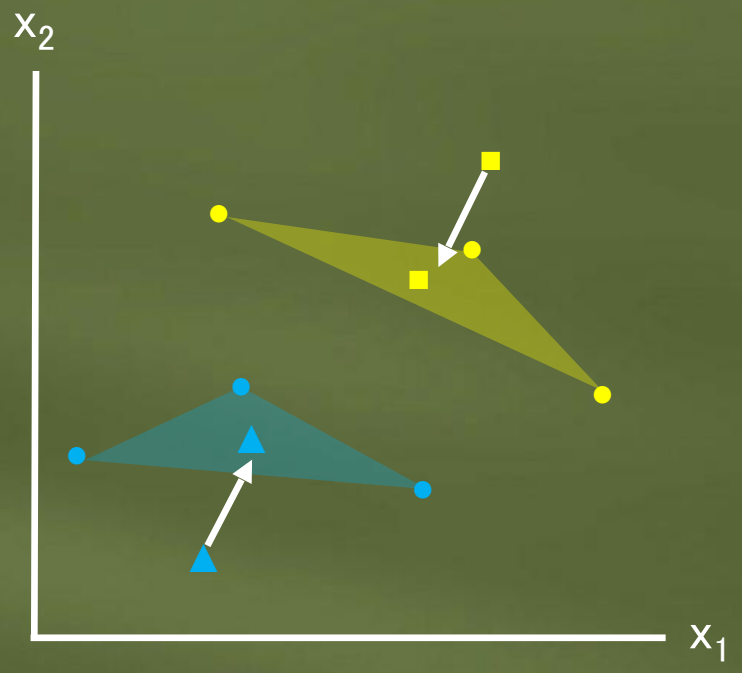
注：個体ではなく変数を分類することも可能（Rコマンダーでは不可）

13.5　非階層クラスター分析（k-means法）①



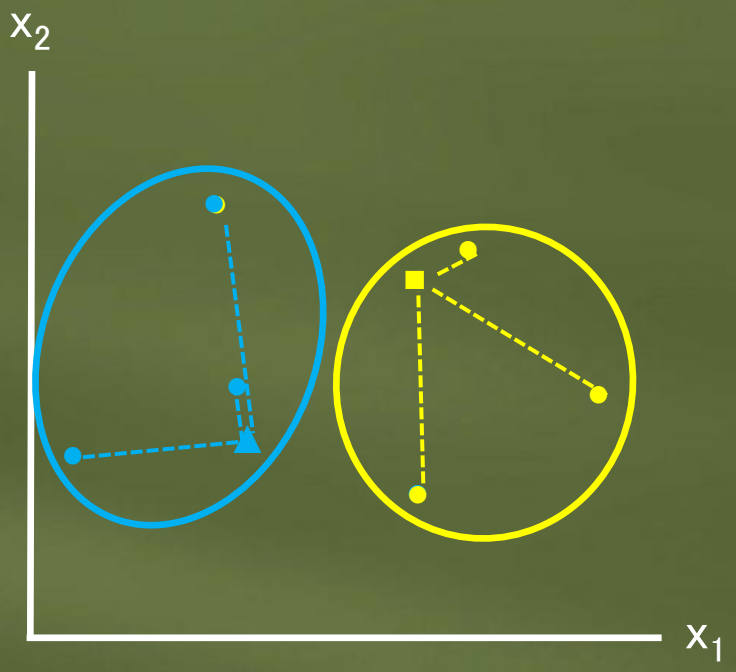
手順①：事前に決めたクラスターの数kだけ基点（■▲）
を無作為に配置して，近い個体でクラスターを作成

非階層クラスター分析（k-means法）②



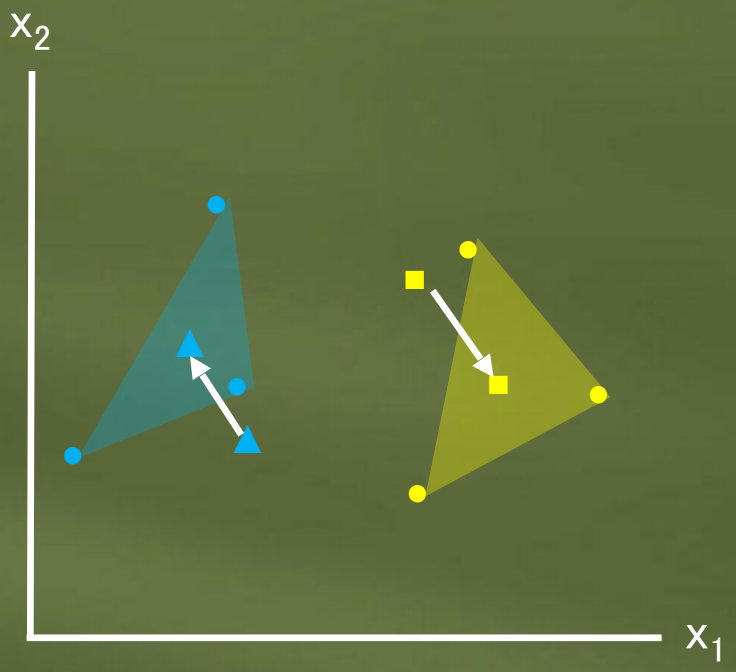
手順②：クラスターの重心を計算して基準を移動
（meansは直線の重心から来ている）

非階層クラスター分析（k-means法）③



手順③：新しい基準から近い個体で新クラスターを作成

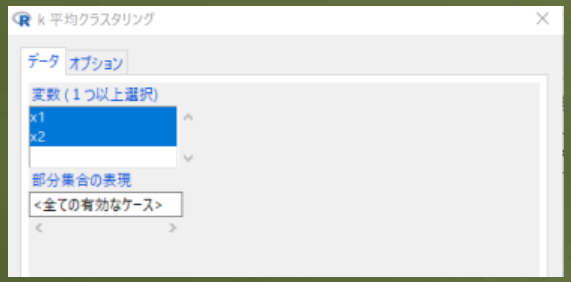
非階層クラスター分析（k-means法）④

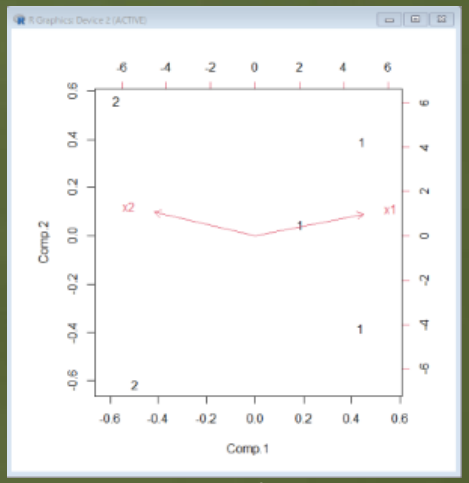


手順④：再び新クラスターの重心を計算して基準を移動する。
分類に変更がなくなるまで繰り返す。

Rコマンダーによるk平均クラスタ分析

［統計量］→［次元解析］→［クラスタ分析］→［k平均クラスタ分析］





バイプロット

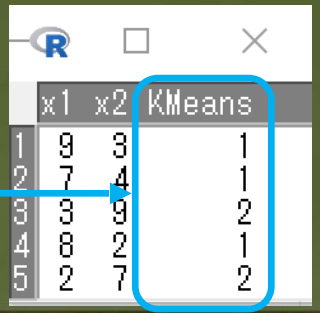
最初にクラスターの数を決めておかねばならない

最初に基準を置くための種になる値（適当で良い）

手順②～④を最大何回まで繰り返すかを設定

変数と分類された個体の関係を図示

データセットに分類されたクラスターの番号を保存



	x1	x2	KMeans
1	9	3	1
2	7	4	1
3	3	9	2
4	8	2	1
5	2	7	2

クラスター分析の注意点

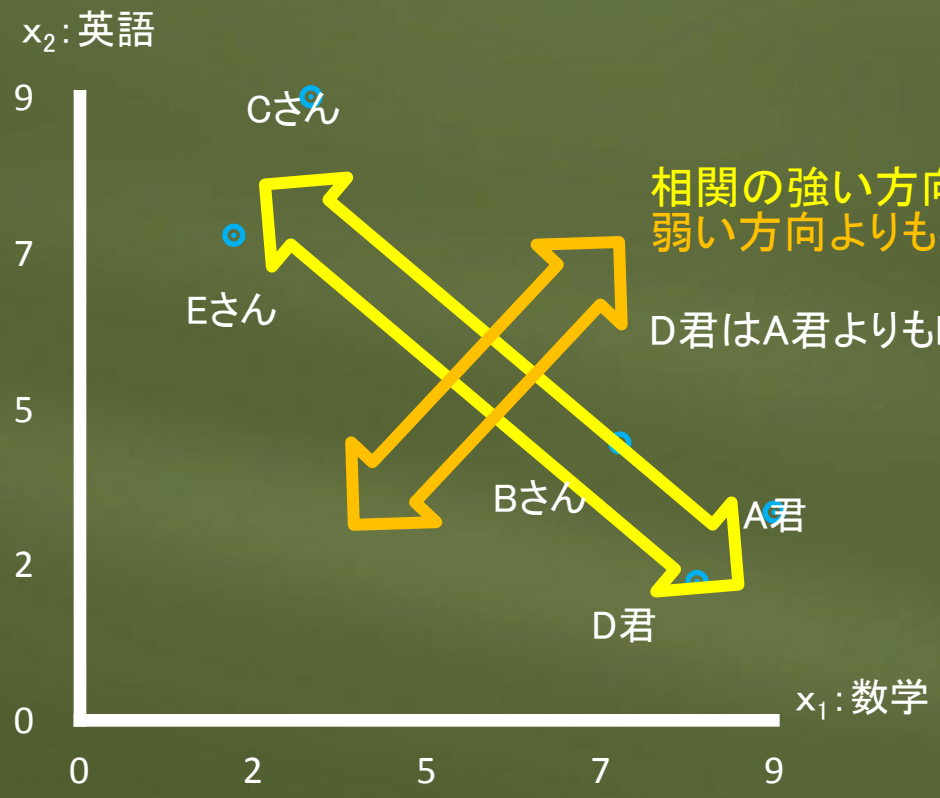
①変数間でバラツキが異なっていると，バラツキが大きな方の変数の影響が強く出る

→変数毎に標準化しておく

②変数間に強い相関がある場合には，ユークリッド距離では正しい分類ができない（次に図で解説）

→マハラノビス距離などを用いる

変数間の相関の影響



相関の強い方向のユークリッドの距離は
弱い方向よりも相対的に長く計算される

D君はA君よりもBさんと近かったかも…

以上で第13章は終了です。