データ解析

http://coconut.sys.eng.shizuoka.ac.jp/data/06/

静岡大学工学部
安藤和敏

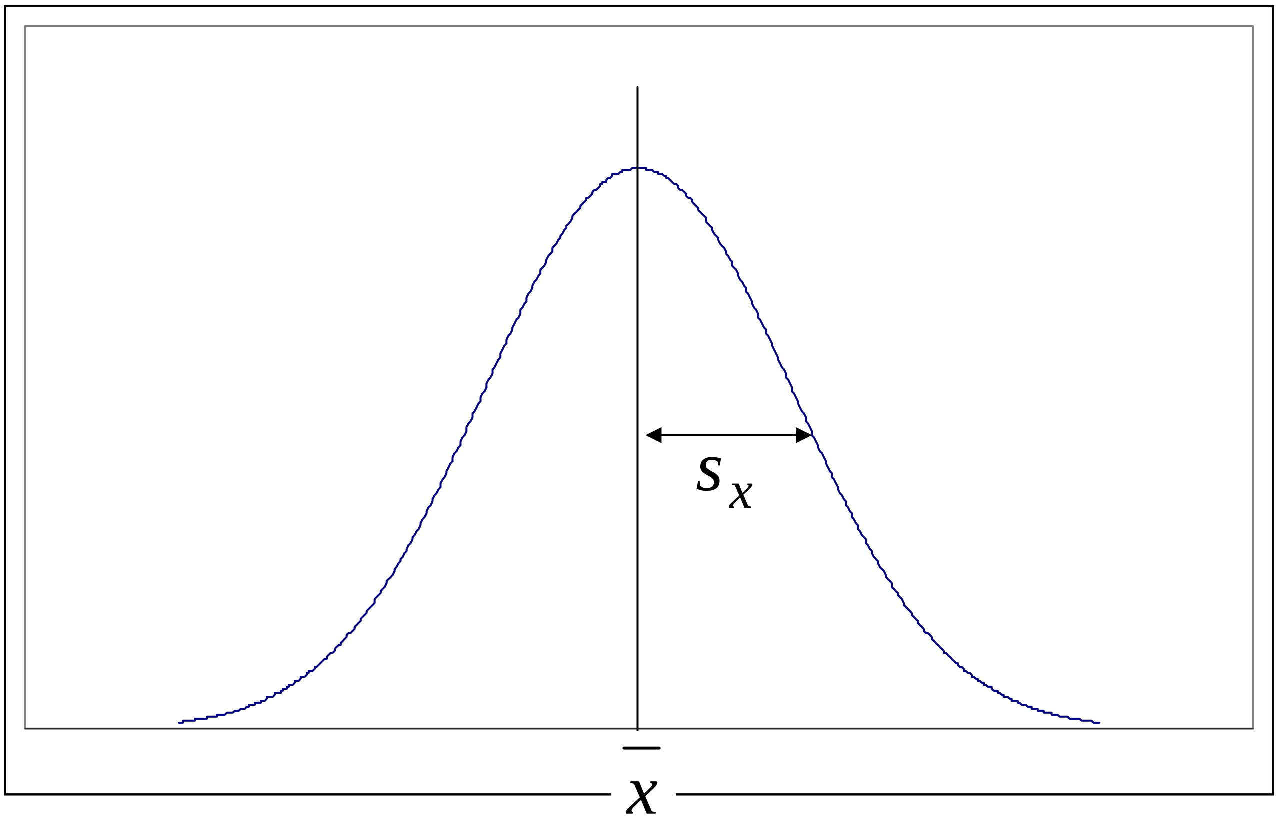
2006.11.16

第5章 Excelで学ぶ判別分析

5-1 相関図で判別分析
5-2 線形判別関数を利用した判別分析
5-3 マハラノビスの距離を利用した判別分析

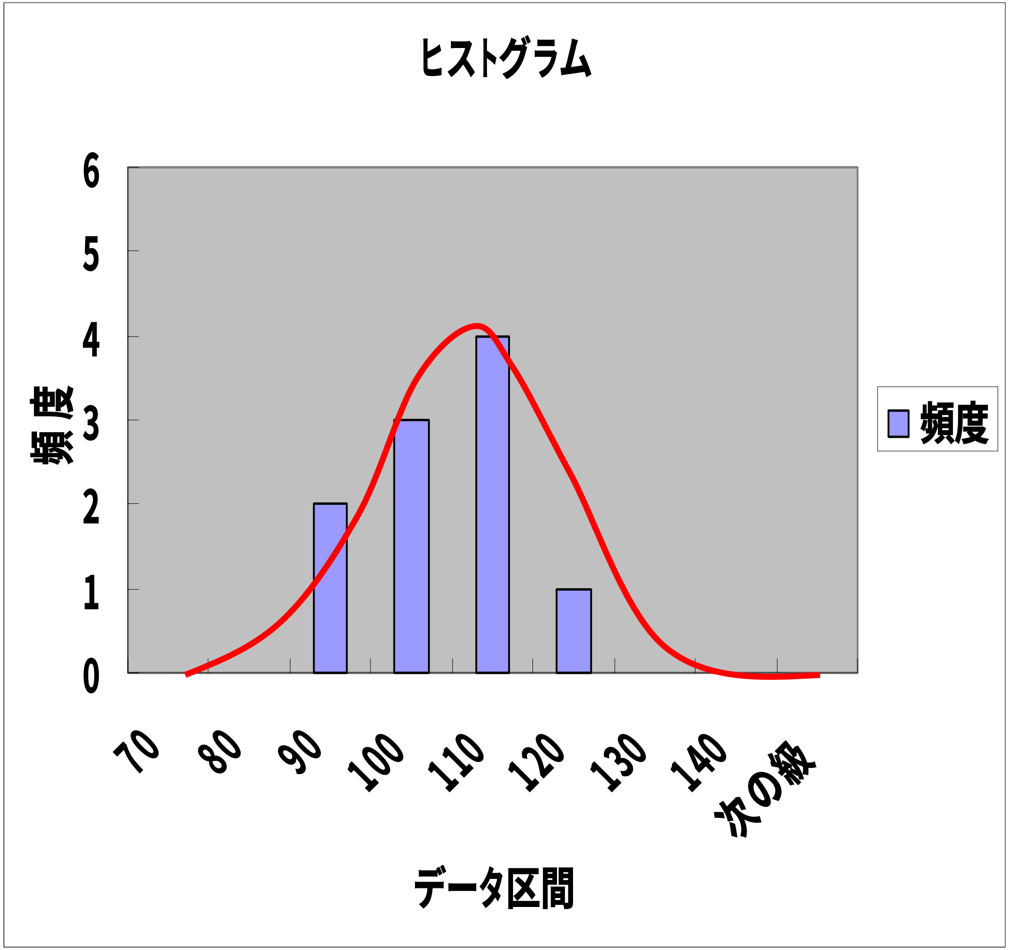
標準偏差　s_x

$$s_x = \sqrt{s_x^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$



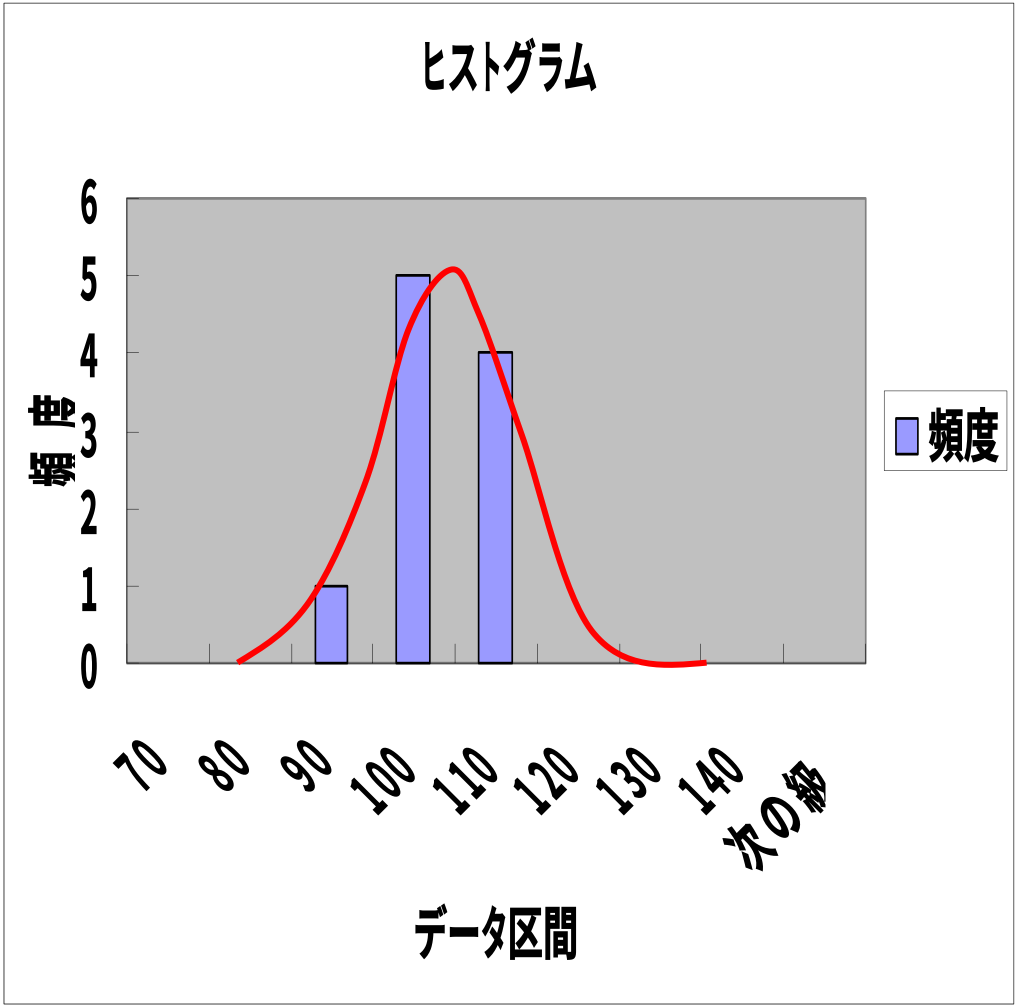
1変数のマハラノビスの距離

No	x
1	109
2	82
3	103
4	87
5	102
6	100
7	93
8	101
9	115
10	99
平均	99.1
分散	85.49
標準偏差	9.24608



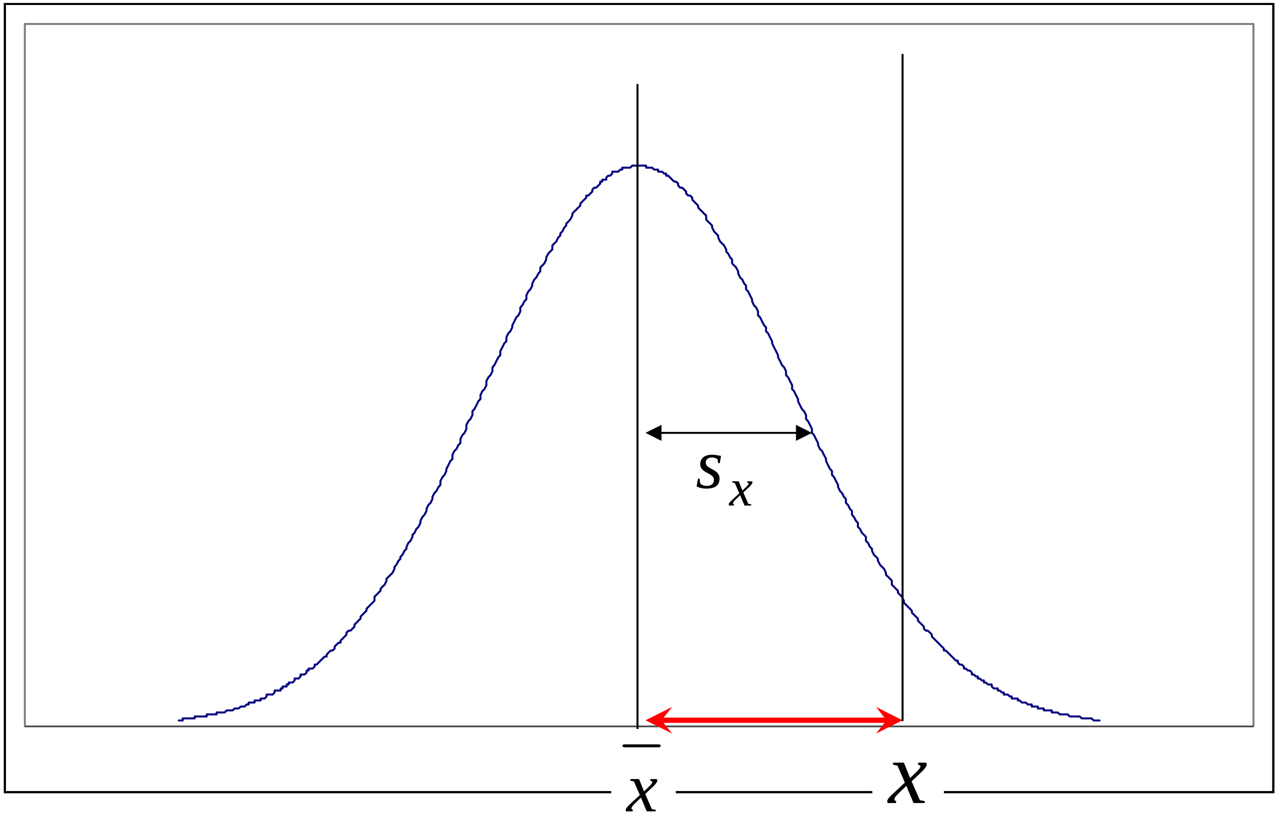
1変数のマハラノビスの距離

No	x
1	95
2	89
3	98
4	91
5	103
6	102
7	97
8	99
9	103
10	107
平均	98.4
分散	28.64
標準偏差	5.351635



1変数のマハラノビスの距離

$$D = \frac{|x - \bar{x}|}{s_x}$$



多変数のマハラノビスの距離

1変数のマハラノビスの距離は，

$$D^2 = \frac{\left|x - \overline{x}\right|^2}{s_x^2}$$

$$= (x - \overline{x})\left(s_x^2\right)^{-1}(x - \overline{x})$$

と書くことができる．

多変数のマハラノビスの距離

2変数(x,y)のマハラノビスの距離Dは，

$$D^2 = (x - \overline{x},\ y - \overline{y})\begin{pmatrix} s_x^2 & s_{xy} \\ s_{xy} & s_y^2 \end{pmatrix}^{-1}\begin{pmatrix} x - \overline{x} \\ y - \overline{y} \end{pmatrix}$$

と定義される．

多変数のマハラノビスの距離

3変数(x,y,z)のマハラノビスの距離Dは，

$$D^2 = (x-\bar{x},\ y-\bar{y},\ z-\bar{z})\begin{pmatrix} s_x^2 & s_{xy} & s_{xz} \\ s_{xy} & s_y^2 & s_{yz} \\ s_{xz} & s_{yz} & s_z^2 \end{pmatrix}^{-1}\begin{pmatrix} x-\bar{x} \\ y-\bar{y} \\ z-\bar{z} \end{pmatrix}$$

と定義される．

マハラノビスの距離による判別

グループPの平均ベクトル

$$(\overline{x}_P\ ,\ \overline{y}_P)$$

とグループPの分散共分散行列S_Pを計算する.

グループQの平均ベクトル

$$(\overline{x}_Q\ ,\ \overline{y}_Q)$$

とグループQの分散共分散行列S_Qを計算する.

マハラノビスの距離による判別

各個体(x,y)に対してマハラノビスの距離

$$D_P^2 = (x - \overline{x}_P, \mathrm{y} - \overline{y}_P) S_P^{-1} \begin{pmatrix} x - \overline{x}_P \\ y - \overline{y}_P \end{pmatrix}$$

$$D_Q^2 = (x - \overline{x}_Q, \mathrm{y} - \overline{y}_Q) S_Q^{-1} \begin{pmatrix} x - \overline{x}_Q \\ y - \overline{y}_Q \end{pmatrix}$$

を計算し,

$D_P^2 < D_Q^2$ **のとき(x,y)をPに分類し,**

$D_P^2 > D_Q^2$ **のとき(x,y)をQに分類する.**

Excelで学ぼう

ファイル：第5章/5_3