


相関係数

$-1 \leq \text{相関係数 } r \leq 1$

X'	Y'
$x_1 - \bar{x}$	$y_1 - \bar{y}$
$x_2 - \bar{x}$	$y_2 - \bar{y}$
$x_3 - \bar{x}$	$y_3 - \bar{y}$
$x_4 - \bar{x}$	$y_4 - \bar{y}$
$x_5 - \bar{x}$	$y_5 - \bar{y}$
$x_6 - \bar{x}$	$y_6 - \bar{y}$
$x_7 - \bar{x}$	$y_7 - \bar{y}$
$x_8 - \bar{x}$	$y_8 - \bar{y}$
\vdots	\vdots
$x_{N-3} - \bar{x}$	$y_{N-3} - \bar{y}$
$x_{N-3} - \bar{x}$	$y_{N-3} - \bar{y}$
$x_{N-2} - \bar{x}$	$y_{N-2} - \bar{y}$
$x_{N-1} - \bar{x}$	$y_{N-1} - \bar{y}$
$x_N - \bar{x}$	$y_N - \bar{y}$

$\theta = 0^\circ \quad \cos\theta = 1 \quad r = 1$



$\theta = 90^\circ \quad \cos\theta = 0 \quad r = 0$



相関がない

直交している

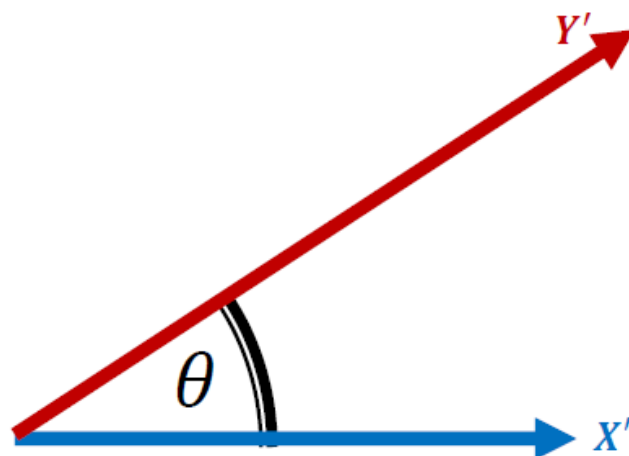
相関係数 r ベクトル X' 、 Y' の内積

$$= \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \sqrt{\sum (y_i - \bar{y})^2}}$$

ベクトル X' 、 Y' のノルム
(スカラー、長さ・大きさ)

$$= \frac{X' \cdot Y'}{\|X'\| \|Y'\|}$$

$$= \cos \theta$$



相関係数 $r = \frac{\vec{\text{馬D}} \cdot \vec{\text{馬E}}}{|\vec{\text{馬D}}| |\vec{\text{馬E}}|} = \cos \theta = \frac{1 \times 2 + 3 \times 4 + 2 \times 4}{\sqrt{1^2 + 3^2 + 2^2} \cdot \sqrt{2^2 + 4^2 + 4^2}} = 0.98$

内積

$\vec{\text{馬D}} \cdot \vec{\text{馬E}} = |\vec{\text{馬D}}| |\vec{\text{馬E}}| \cos \theta = 2 \times 0 + 1 \times 2 + 0 \times 4 = 2$

相関係数

	~1600	~2200	2200~		馬A	馬B	馬C	馬D	馬E	馬F	ベクトルの大きさ
馬A	1	3	2	馬A							3.74
馬B	2	2	2	馬B	0.93						3.46
馬C	2	4	4	馬C	0.98	0.96					6.00
馬D	2	1	0	馬D	0.60	0.39	0.60				2.24
馬E	0	2	4	馬E	0.84	0.77	0.89	0.20			4.47
馬F	3	4	2	馬F	0.94	0.96	0.93	0.83	0.66		5.39
	~1600	~2200	2200~		馬AC	馬B	馬D	馬E	馬F		
馬AC	1.5	3.5	3	馬AC							4.85
馬B	2	2	2	馬B	0.95						3.46
馬D	2	1	0	馬D	0.60	0.77					2.24
馬E	0	2	4	馬E	0.88	0.77	0.20				4.47
馬F	3	4	2	馬F	0.94	0.96	0.83	0.66			5.39
	~1600	~2200	2200~		馬AC	馬BF	馬D	馬E			
馬AC	2	2	2	馬AC							3.46
馬BF	2.5	3	2	馬BF	0.99						4.39
馬D	2	1	0	馬D	0.77	0.82					2.24
馬E	0	2	4	馬E	0.77	0.71	0.20				4.47
	~1600	~2200	2200~		馬ABCF	馬D	馬E				
馬ABCF	2.25	2.5	2	馬ABCF							3.91
馬D	2	1	0	馬D	0.80						2.24
馬E	0	2	4	馬E	0.74	0.20					4.47
	~1600	~2200	2200~		馬ABCDF	馬E					
馬ABCDF	2.25	2.5	2	馬ABCDF							2.93
馬E	2	1	0	馬E	1.07						2.24

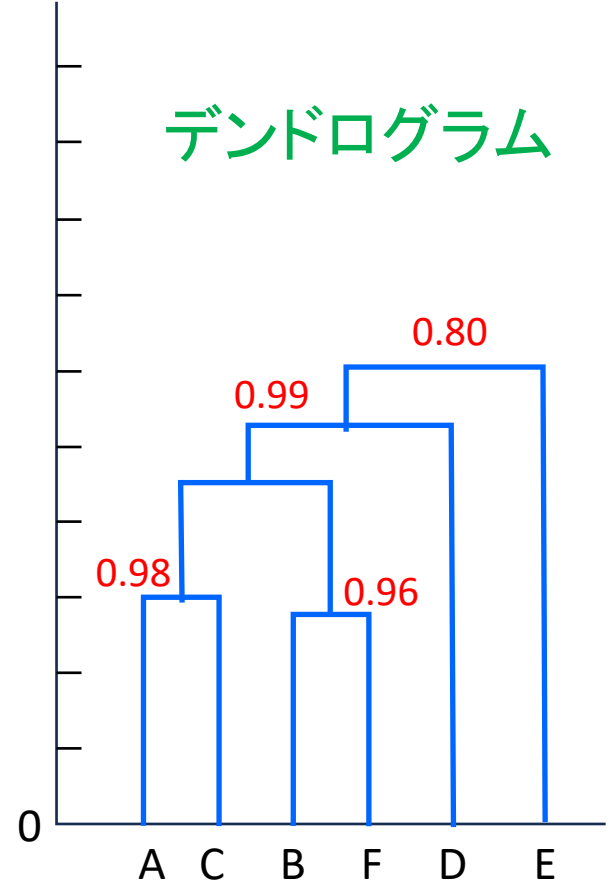
平均値

平均値

平均値

平均値

デンドログラム



$$\vec{馬C} \cdot \vec{馬F} = |\vec{馬C}| |\vec{馬F}| \cos \theta = 2 \times 3 + 4 \times 4 + 4 \times 2 = 30$$

$$\vec{馬CF} \cdot \vec{馬A} = |\vec{馬CF}| |\vec{馬A}| \cos \theta = 2.5 \times 1 + 4 \times 3 + 3 \times 2 = 20.5$$

$$\vec{馬ACF} \cdot \vec{E} = |\vec{馬ACF}| |\vec{馬E}| \cos \theta = 1.75 \times 0 + 3.5 \times 2 + 2.5 \times 4 = 17$$

類似性

平均値

平均値

平均値

平均値

	~1600	~2200	2200~		馬A	馬B	馬C	馬D	馬E	馬F
馬A	1	3	2	馬A						
馬B	2	2	2	馬B	12					
馬C	2	4	4	馬C	22	20				
馬D	2	1	0	馬D	5	6	8			
馬E	0	2	4	馬E	14	12	24	2		
馬F	3	4	2	馬F	19	18	30	10	16	

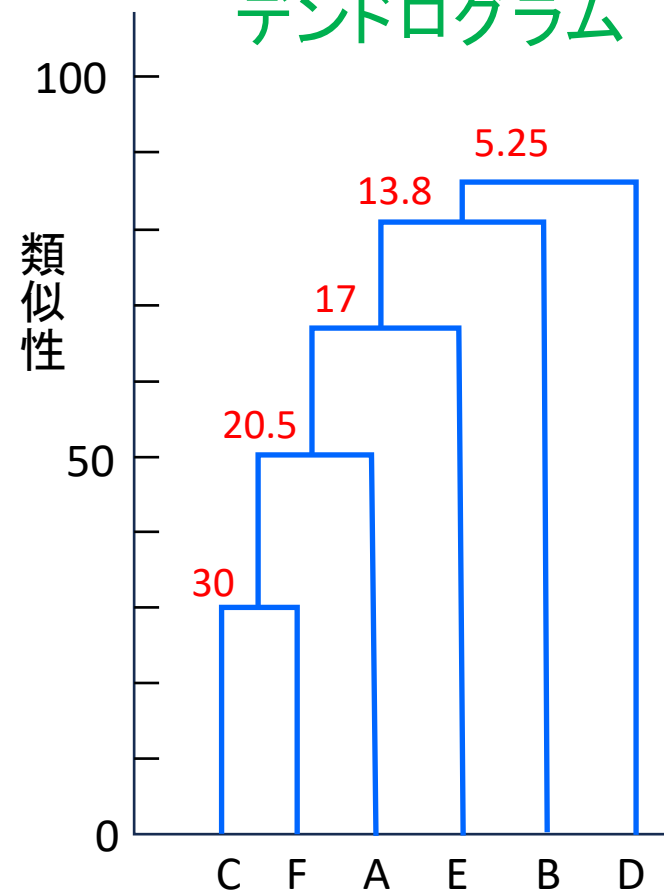
	~1600	~2200	2200~		馬CF	馬A	馬B	馬D	馬E
馬CF	2.5	4	3	馬CF					
馬A	1	3	2	馬A	20.5				
馬B	2	2	2	馬B	19	12			
馬D	2	1	0	馬D	9	5	6		
馬E	0	2	4	馬E	20	14	12	2	

	~1600	~2200	2200~		馬ACF	馬B	馬D	馬E
馬ACF	1.75	3.5	2.5	馬ACF				
馬B	2	2	2	馬B	15.5			
馬D	2	1	0	馬D	7	6		
馬E	0	2	4	馬E	17	12	2	

	~1600	~2200	2200~		馬ACEF	馬B	馬D
馬ACEF	0.9	2.8	3.3	馬ACEF			
馬B	2	2	2	馬B	13.8		
馬D	2	1	0	馬D	4.5	6	

	~1600	~2200	2200~		馬ABCEF	馬D
馬ABCEF	1.4	2.4	2.6	馬ABCEF		
馬D	2	1	0	馬D	5.25	

デンドログラム



A(1 3 2)とB(2 2 2)のユークリッド距離 = $\sqrt{(1-2)^2 + (3-2)^2 + (2-2)^2} = 1.4$

AB(1.5 2.5 2)とF(3 4 2)のユークリッド距離 = $\sqrt{(1.5-3)^2 + (2.5-4)^2 + (2-2)^2} = 2.1$

ABF(2.3 3.3 2)とC(2 4 4)のユークリッド距離 = $\sqrt{(2.3-2)^2 + (3.3-4)^2 + (2-4)^2} = 2.2$

平均値

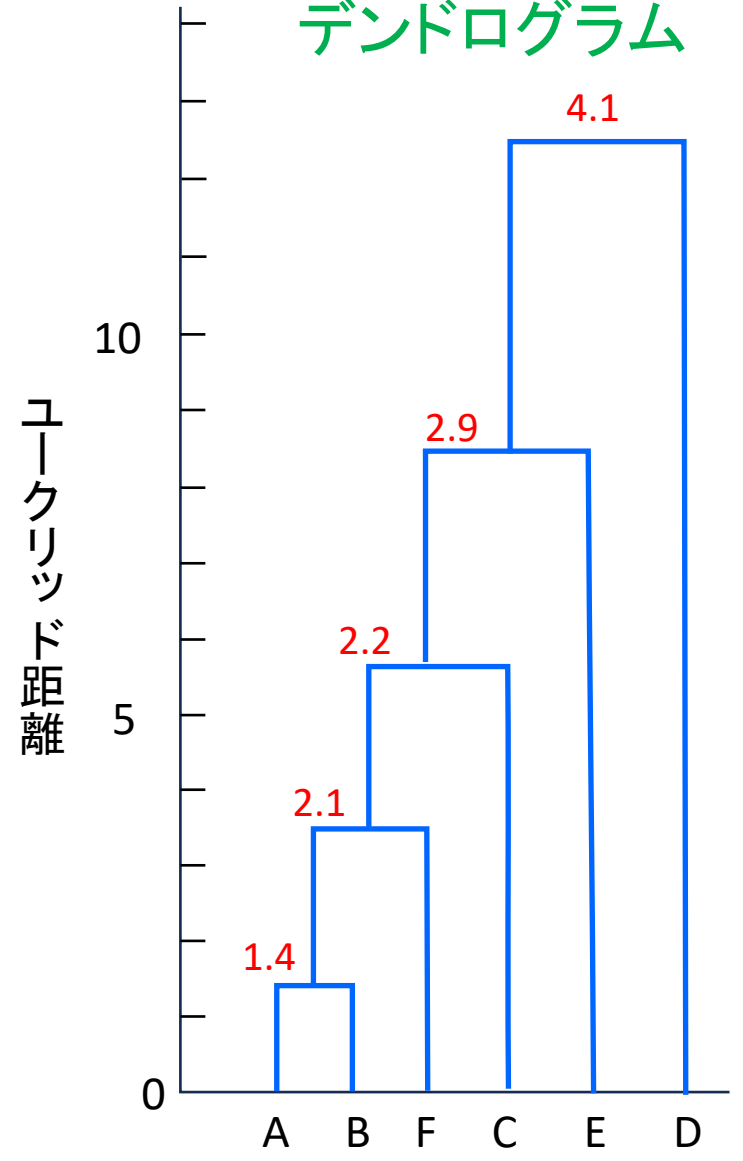
平均値

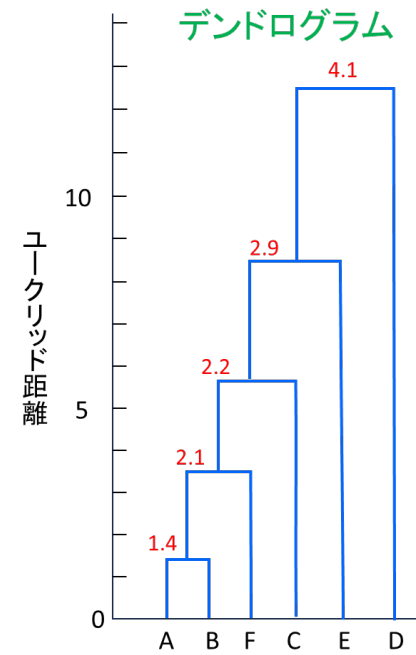
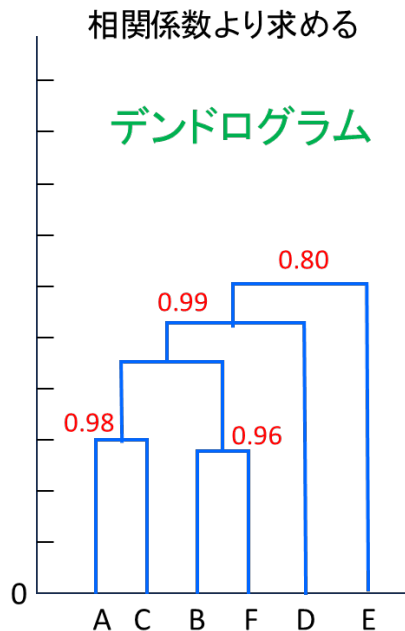
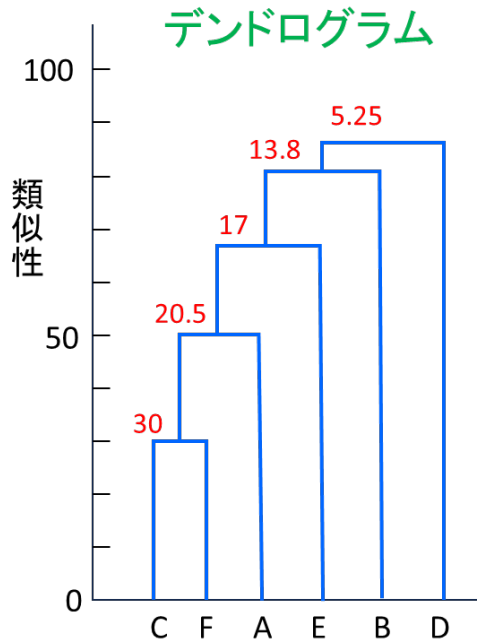
平均値

平均値

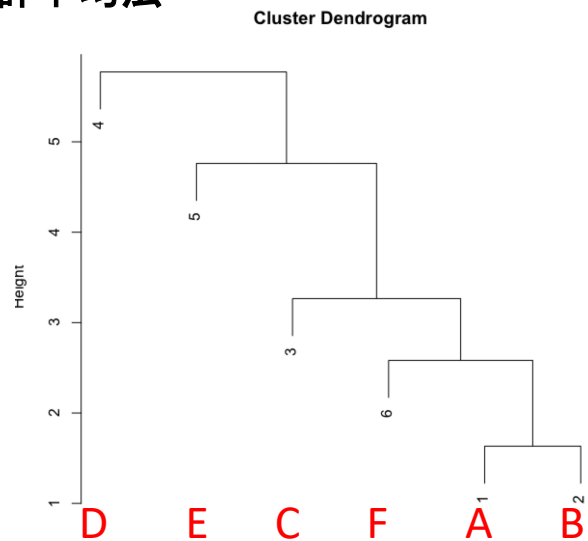
	~1600	~2200	2200~		馬A	馬B	馬C	馬D	馬E	馬F
馬A	1	3	2	馬A						
馬B	2	2	2	馬B	1.4					
馬C	2	4	4	馬C	2.4	2.8				
馬D	2	1	0	馬D	3.0	2.2	5.0			
馬E	0	2	4	馬E	2.4	2.8	2.8	4.6		
馬F	3	4	2	馬F	2.2	2.2	2.2	3.7	4.1	
馬AB	1.5	2.5	2	馬AB						
馬C	2	4	4	馬C	2.5					
馬D	2	1	0	馬D	2.5	5.0				
馬E	0	2	4	馬E	2.5	2.8	4.6			
馬F	3	4	2	馬F	2.1	2.2	3.7	4.1		
馬ABF	2.3	3.3	2	馬ABF						
馬C	2	4	4	馬C	2.2					
馬D	2	1	0	馬D	3.0	5.0				
馬E	0	2	4	馬E	3.3	2.8	4.6			
馬ABCF	2.1	3.6	3.0	馬ABCF						
馬D	2	1	0	馬D	4.0					
馬E	0	2	4	馬E	2.9	4.6				
馬ABCEF	1.1	2.8	3.5	馬ABCEF						
馬D	2	1	0	馬D	4.1					

デンドログラム



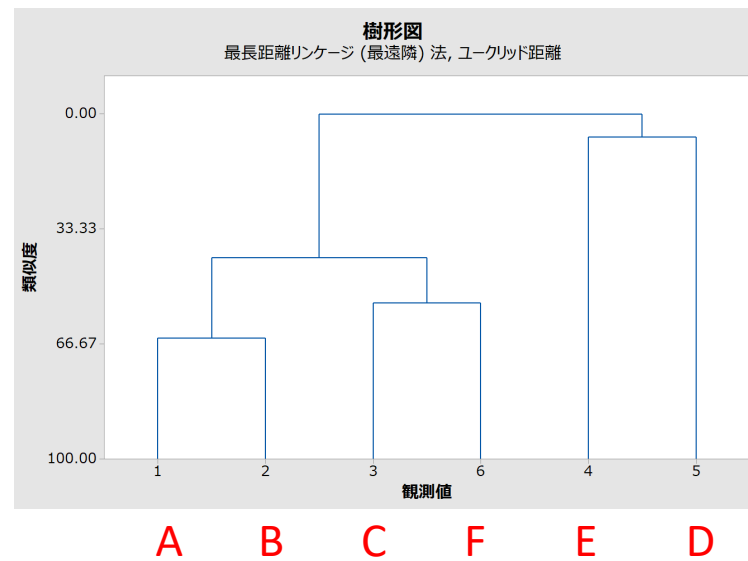


最長距離法(完全連結法)
群平均法



統計ソフトRを用いて実行

Minitab実行結果



統計ソフトRで実行

cluster.csv

	1600	2200	2201
A	1	3	2
B	2	2	2
C	2	4	4
D	2	1	0
E	0	2	4
F	3	4	2

```
> Data <- read.csv("cluster.csv")
```

```
> Data
```

```
X X1600 X2200 X2201
```

```
1 A 1 3 2
2 B 2 2 2
3 C 2 4 4
4 D 2 1 0
5 E 0 2 4
6 F 3 4 2
```

階層クラスター分析計算

```
> d <- dist(data)
```

```
> hc <- hclust(d, method="single")
```

```
> plot(hc)
```

```
> hc <- hclust(d^2, method="complete")
```

```
> plot(hc)
```

```
> hc <- hclust(d, method="average")
```

```
> plot(hc)
```

```
> hc <- hclust(d, method="centroid")
```

```
> plot(hc)
```

```
> hc <- hclust(d^2, method="median")
```

```
> plot(hc)
```

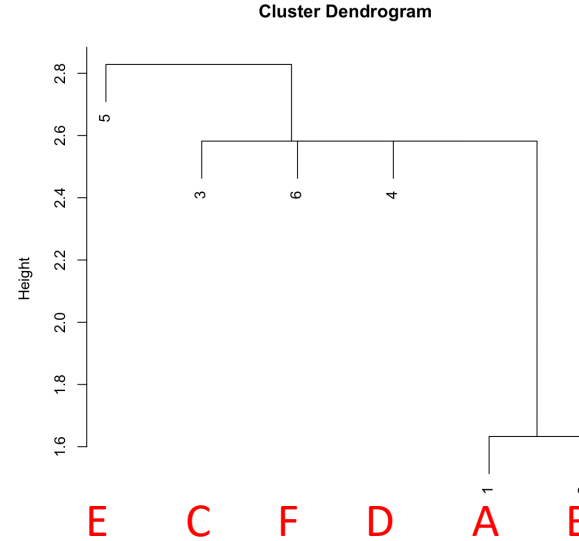
```
> hc <- hclust(d^2, method="ward.D2")
```

```
> plot(hc)
```

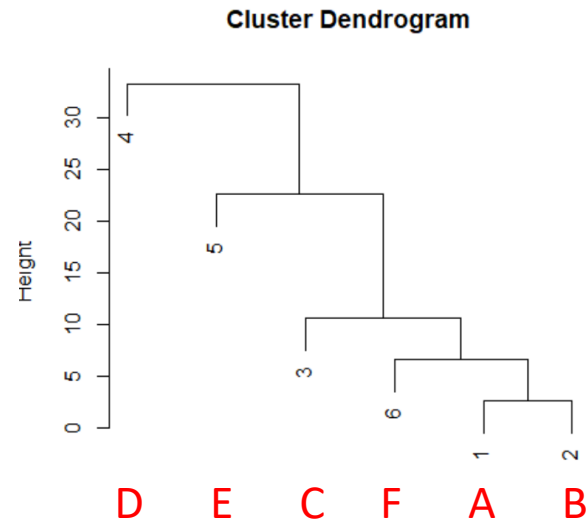
方法	引数methodで設定する文字列
最短距離法 (単連結法)	single
最長距離法 (完全連結法)	complete
群平均法	average
重心法	centroid
メディアン法	median
McQuitty 法	mcquitty
ウォード法 (最小分散法)	ward.D2

デンドログラムを描く

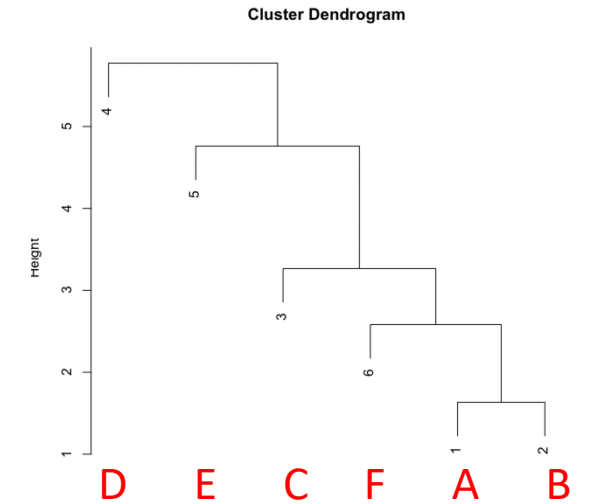
最短距離法 (単連結法)



重心法 メディアン法



最長距離法 (完全連結法) 群平均法



ウォード法 (最小分散法)

