

1 4. 多変量解析 5
(クラスター分析)

クラスター分析

2

階層的クラスター分析

非階層的クラスター分析

3

クラスター分析

サンプルクラスター分析

変数クラスター分析

4

例) 階層的クラスター分析 サンプルクラスター分析

英語	国語	社会	数学	理科	英語	国語	社会	数学	理科
60	65	76	54	45	85	73	77	85	91
70	68	67	66	56	90	82	84	93	94
43	56	54	23	34	77	63	75	74	84
82	78	78	45	69	69	58	62	76	67
89	89	93	78	82	56	63	45	72	82
67	78	92	56	45	59	73	67	54	62
87	92	70	69	72	79	87	65	83	85
72	58	74	78	68	68	66	74	80	78
67	82	77	67	62	82	72	81	89	93
58	67	84	78	83	72	78	83	95	89

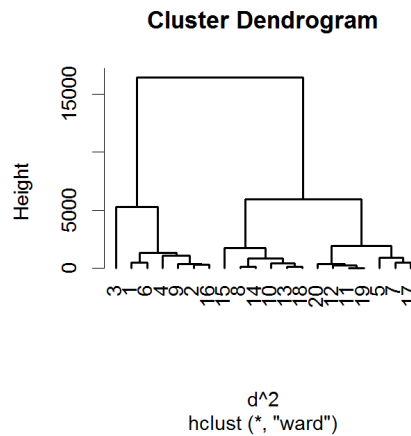
5

```
eigo <-
c(60, 70, 43, 82, 89, 67, 87, 72, 67, 58, 85, 90, 77, 69, 56, 59, 79, 68, 82, 72)
kokugo <-
c(65, 68, 56, 78, 89, 78, 92, 58, 82, 67, 73, 82, 63, 58, 63, 73, 87, 66, 72, 78)
syakai <-
c(76, 67, 54, 78, 93, 92, 70, 74, 77, 84, 77, 84, 75, 62, 45, 67, 65, 74, 81, 83)
sugaku <-
c(54, 66, 23, 45, 78, 56, 69, 78, 67, 78, 87, 93, 74, 76, 72, 54, 83, 80, 89, 95)
rika <-
c(45, 56, 34, 69, 82, 45, 72, 68, 62, 83, 91, 94, 84, 67, 82, 62, 85, 78, 93, 89)
kyoka <- cbind(eigo, kokugo, syakai, sugaku, rika)
colnames(kyoka) <- c("英語", "国語", "社会", "数学", "理科")
rownames(kyoka) <- 1:20
```

6

```
d <- dist(kyoka) ユークリッド距離
ans <- hclust(d^2, method="ward")
par(cex=1.8, lwd=3)
plot(ans, hang=-1)
```

↑ ワード法による
クラスタ分析



7

距離、類似度、非類似度

ユークリッド距離 $\sqrt{\sum (x_k - y_k)^2}$

マハラノビスの汎距離

相関係数

マンハッタン距離

ミンコフスキー距離

⋮

8

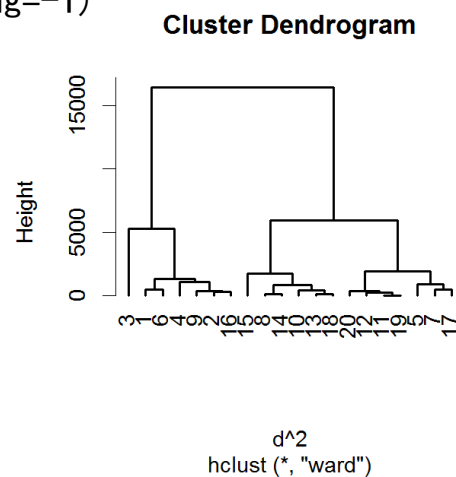
クラスター作成方法

ウォード法	最もバランスのとれたクラスターを作る
最短距離法	鎖状のクラスターを作る
最長距離法	空間の拡散が生じやすい
メディアン法	最短距離法と最長距離法の折衷 クラスター間の距離の逆転が生じることがある
重心法	クラスター間の距離の逆転が生じることがある

ワード法

9

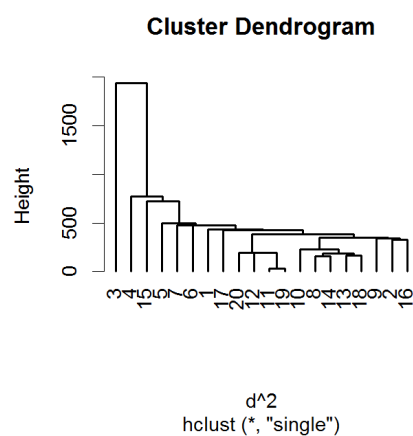
```
d <- dist(kyoka)
ans <- hclust(d^2, method="ward")
par(cex=1.8, lwd=3)
plot(ans, hang=-1)
```



最短距離法（最近隣法、単連結法）

10

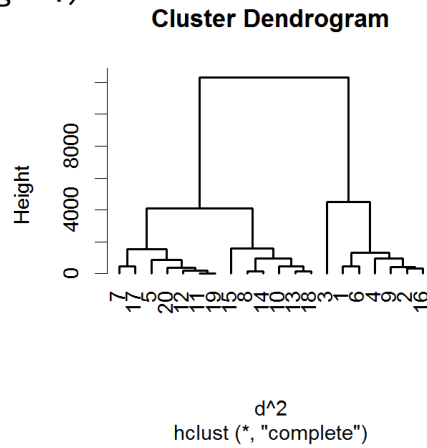
```
d <- dist(kyoka)
ans <- hclust(d^2, method="single")
par(cex=1.8, lwd=3)
plot(ans)
```



最長距離法（最遠隣法、完全連結法）

11

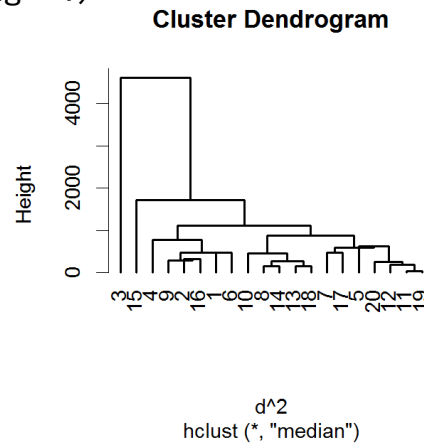
```
d <- dist(kyoka)
ans <- hclust(d^2, method="complete")
par(cex=1.8, lwd=3)
plot(ans, hang=-1)
```



メディアン法

12

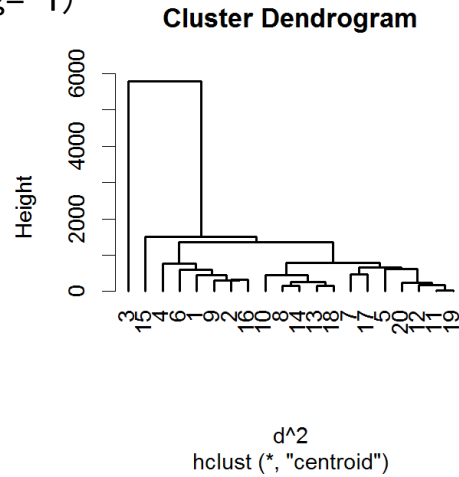
```
d <- dist(kyoka)
ans <- hclust(d^2, method="median")
par(cex=1.8, lwd=3)
plot(ans, hang=-1)
```



重心法

13

```
d <- dist(kyoka)
ans <- hclust(d^2, method="centroid")
par(cex=1.8, lwd=3)
plot(ans, hang=-1)
```

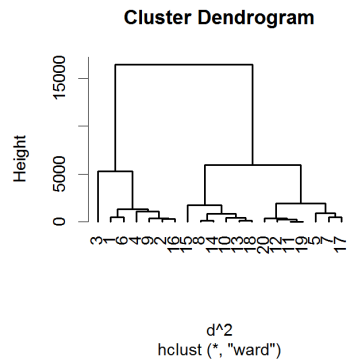


樹形図の枝の長さ (高さ)

14

```
d <- dist(kyoka)
ans <- hclust(d^2, method="ward")
ans$height
```

[1]	34.0000	158.0000
	163.0000	248.6667
	326.0000	359.3333
[7]	395.3333	415.6667
	478.0000	479.0000
	854.5333	937.0000
[13]	1097.1667	1319.1667
	1761.1333	1912.2857
	5276.3333	5944.1502
[19]	16403.0308	



変数のクラスター分析

15

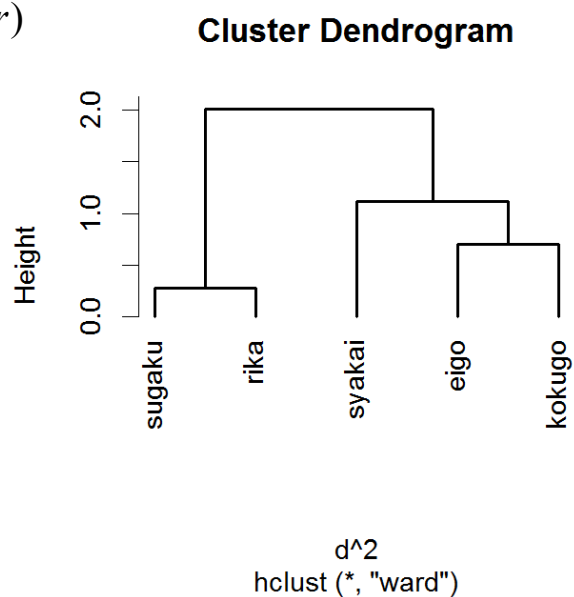
$$d = \sqrt{2(1-r)}$$

```
r <- cor(kyoka)
round(r, 3)
d <- sqrt(2*(1-r))
d <- as.dist(d)
round(d, 3)
par(cex=1.8, lwd=3)
plot(hclust(d^2, method="ward"), hang=-1)
```

変数のクラスター分析

16

$$d = \sqrt{2(1-r)}$$



例題)

17

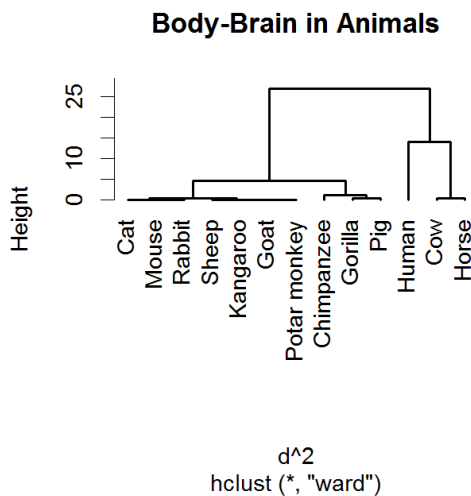
```
library(MASS)
> data(Animals)
> Animals
```

	body	brain		body	brain
Mountain beaver	1.350	8.1	African elephant	6654.000	5712.0
Cow	465.000	423.0	Triceratops	9400.000	70.0
Grey wolf	36.330	119.5	Rhesus monkey	6.800	179.0
Goat	27.660	115.0	Kangaroo	35.000	56.0
Guinea pig	1.040	5.5	Golden hamster	0.120	1.0
Dipliodocus	11700.000	50.0	Mouse	0.023	0.4
Asian elephant	2547.000	4603.0	Rabbit	2.500	12.1
Donkey	187.100	419.0	Sheep	55.500	175.0
Horse	521.000	655.0	Jaguar	100.000	157.0
Potar monkey	10.000	115.0	Chimpanzee	52.160	440.0
Cat	3.300	25.6	Rat	0.280	1.9
Giraffe	529.000	680.0	Brachiosaurus	87000.000	154.5
Gorilla	207.000	406.0	Mole	0.122	3.0

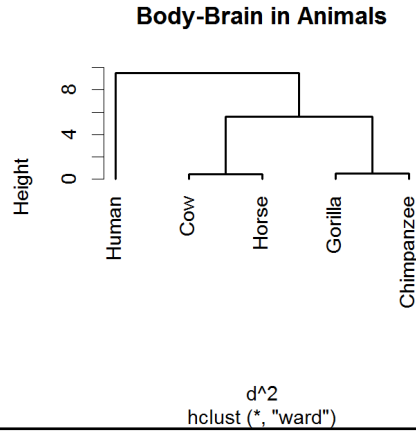
```
> dat <-
Animals[c(2, 4, 9, 10, 11, 13, 14, 18, 20, 21, 22, 24, 28), ]
> d <- dist(scale(dat)) 標準化
```

```
> ans <- hclust(d^2, method="ward")
> par(cex=1.8, lwd=3)
> plot(ans, hang=-1, main="Body-Brain in Animals")
```

18



```
> dat <- Animals[c(2, 9, 13, 14, 24), ]
> d <- dist(scale(dat))
> ans <- hclust(d^2, method="ward")
> plot(ans, hang=-1, main="Body-Brain in Animals")
```



37	香川県	高松市 地図 >>	北緯 34°20'25" 東経 134°02'36"	34.34028,134.04333	34.34028	134.04333	20
38	愛媛県	松山市 地図 >>	北緯 33°50'30" 東経 132°45'58"	33.84167,132.76611	33.84167	132.76611	
39	高知県	高知市 地図 >>	北緯 33°33'35" 東経 133°31'52"	33.55972,133.53111	33.55972	133.53111	
40	福岡県	福岡市 地図 >>	北緯 33°36'23" 東経 130°25'05"	33.60639,130.41806	33.60639	130.41806	
41	佐賀県	佐賀市 地図 >>	北緯 33°14'58" 東経 130°17'56"	33.24944,130.29889	33.24944	130.29889	
42	長崎県	長崎市 地図 >>	北緯 32°44'41" 東経 129°52'25"	32.74472,129.87361	32.74472	129.87361	
43	熊本県	熊本市 地図 >>	北緯 32°47'23" 東経 130°44'30"	32.78972,130.74167	32.78972	130.74167	
44	大分県	大分市 地図 >>	北緯 33°14'17" 東経 131°36'45"	33.23806,131.6125	33.23806	131.6125	
45	宮崎県	宮崎市 地図 >>	北緯 31°54'40" 東経 131°25'26"	31.91111,131.42389	31.91111	131.42389	
46	鹿児島県	鹿児島市 地図 >>	北緯 31°33'37" 東経 130°33'29"	31.56028,130.55806	31.56028	130.55806	
47	沖縄県	那覇市 地図 >>	北緯 26°12'45" 東経 127°40'52"	26.2125,127.68111	26.2125	127.68111	

21

```

Ido <- c(34.1, 34.3, 33.8, 33.6, 33.6, 33.2, 32.7, 32.7,
33.2, 31.9, 31.5, 26.2)
Keido <- c(134.6, 134.0, 132.8, 133.5, 130.4, 130.3,
129.9, 130.7, 131.6, 131.4, 130.6, 127.6)
Ichi <- cbind(Ido, Keido)
rownames(Ichi) <- c("徳島","高松","松山","高知","福岡","
佐賀","長崎","熊本","大分","宮崎","鹿児島","那覇")
d <- dist(Ichi)
ans <- hclust(d^2, method="ward")
plot(ans, hang=-1)

```

22

