資料探勘

711378912 蔡宜諠

27/09/2025

```
# 創造一個x
x <- sample(1:39, 10, replace = FALSE)
x
```

```
##  [1] 13 23 15 36 19 16 11 29 31 24
```

```
x2 = 1 + (x >= 10) + (x >= 20) + (x >= 30)
x2
```

```
##  [1] 2 3 2 4 2 2 2 3 4 3
```

```
# 將上述變數的數字編碼改為文字編碼：
labels = c("A", "B", "C")
x3 = labels[x2]
x3
```

```
##  [1] "B" "C" "B" NA  "B" "B" "B" "C" NA  "C"
```

```
# 將年齡資料轉換為年齡群組 1~20、21~40、41~60、61以上，並編碼為 1、2、3、4：
age = c(36, 53, 13, 3, 6, 36, 48, 6, 22, 1, 7, 18, 9, 41, 68)

age2 = 1 + (age >= 21) + (age >= 41) + (age >= 61)
age2
```

```
##  [1] 2 3 1 1 1 2 3 1 2 1 1 1 1 3 4
```

```
x = c('貓', '狗', '狼')
animals = c('家庭寵物', '野生動物')
x.index = 1 * (x %in% c('貓', '狗')) + 2 * (x == '狼')
x = animals[x.index]
x
```

```
## [1] "家庭寵物" "家庭寵物" "野生動物"
```

```
income = c(130065, 82961, 133076, 123028, 108945, 173466, 17477)
newcodes = c("低收入", "中等收入", "高收入")
index = 1 * (income < 20000) +
        2 * (income >= 20000 & income < 60000) +
        3 * (income >= 60000)
income = newcodes[index]
income
```

```
## [1] "高收入" "高收入" "高收入" "高收入" "高收入" "高收入" "低收入"
```

```
income = 1 + 1 * (income >= 20000) + 1 * (income >= 60000)
income
```

```
## [1] 3 3 3 3 3 3 3
```

```
x = c(4, 12, 50, 18, 50, 22, 23, 46, 8, 46, 36, 18, 10, 14, 35, 48, 23, 17, 29, 30)
x2 = ifelse(x <= 30, 1, 2)
x2
```

```
##  [1] 1 1 2 1 2 1 1 2 1 2 2 1 1 1 2 2 1 1 1 1
```

```
x3 = ifelse(x <= 30, "A", "B")
x3
```

```
##  [1] "A" "A" "B" "A" "B" "A" "A" "B" "A" "B" "B" "A" "A" "A" "B" "B" "A" "A" "A"
## [20] "A"
```

```
y = c("B", "A", "C", "C", "B", "A", "D", "B", "D", "C")
y2 = ifelse(y %in% c("A", "C"), "Group1", "Group2")
y2
```

```
##  [1] "Group2" "Group1" "Group1" "Group1" "Group2" "Group1" "Group2" "Group2"
##  [9] "Group2" "Group1"
```

```
x = c(4, 12, 50, 18, 50, 22, 23, 46, 8, 46, 36, 18, 10, 14, 35, 48, 23, 17, 29, 30)
x2 = ifelse(x <= 10, 1, ifelse(x <= 20, 2, 3))
x2
```

```
##  [1] 1 2 3 2 3 3 3 3 1 3 3 2 1 2 3 3 3 2 3 3
```

```
y = c("D", "A", "B", "D", "A", "E", "B", "E", "D", "A", "D", "D", "D", "B", "A", "C", "E",
"C", "D", "E")

y2 = ifelse(y %in% c("A", "E"), 1,
            ifelse(y == "C", 2, 3))
y2
```

```
##  [1] 3 1 3 3 1 1 3 1 3 1 3 3 3 3 1 2 1 2 3 1
```

```
x = c(4, 12, 50, 18, 50, 22, 23, 46, 8, 46, 36, 18, 10, 14, 35, 48, 23, 17, 29, 30)

x2 = cut(x,
         breaks = c(0, 10, 20, max(x)),
         labels = c(1, 2, 3))

x2
```

```
##  [1] 1 2 3 2 3 3 3 3 1 3 3 2 1 2 3 3 3 2 3 3
## Levels: 1 2 3
```

```
as.vector(x2)
```

```
##  [1] "1" "2" "3" "2" "3" "3" "3" "3" "1" "3" "3" "2" "1" "2" "3" "3" "3" "2" "3"
## [20] "3"
```

```
x3 = cut(x, breaks = c(0, 10, 20, max(x)))
x3
```

```
##  [1] (0,10]  (10,20] (20,50] (10,20] (20,50] (20,50] (20,50] (20,50] (0,10]
## [10] (20,50] (20,50] (10,20] (0,10]  (10,20] (20,50] (20,50] (20,50] (10,20]
## [19] (20,50] (20,50]
## Levels: (0,10] (10,20] (20,50]
```

```
x5 = cut(x,
         breaks = c(0, 10, 20, max(x)),
         labels = c("A", "B", "C"))
x5
```

```
##  [1] A B C B C C C C A C C B A B C C C B C C
## Levels: A B C
```

```
score = round(rnorm(10, 60, 10))
score
```

```
##  [1] 66 47 56 57 49 71 79 60 71 66
```

```
score.cut = cut(score, breaks = 5)
score.cut
```

```
##  [1] (59.8,66.2] (47,53.4]   (53.4,59.8] (53.4,59.8] (47,53.4]   (66.2,72.6]
##  [7] (72.6,79]   (59.8,66.2] (66.2,72.6] (59.8,66.2]
## Levels: (47,53.4] (53.4,59.8] (59.8,66.2] (66.2,72.6] (72.6,79]
```

```
# 載入套件
library(car)
```

```
## Loading required package: carData
```

```
# 範例資料
x = c(1, 3, 1, 2, 3, 1, 1, 3)

# 將 1、2 編碼為 "A"，3 編碼為 "B"
recode(x, "c(1,2) = 'A'; else = 'B'")
```

```
## [1] "A" "B" "A" "A" "B" "A" "A" "B"
```

```
recode(x, "1:2 = 'A'; 3 = 'B'")
```

```
## [1] "A" "B" "A" "A" "B" "A" "A" "B"
```

```
# 成績資料
score = c(61, 85, 79, 62, 29, 100, 71, 25, 40, 54)

# 將不及格 (<60) 改為 0，及格 (>=60) 改為 1
recode(score, "lo:59 = 0; 60:100 = 1")
```

```
##  [1] 1 1 1 1 0 1 1 0 0 0
```

```
# 使用 else 的寫法
recode(score, "lo:59 = 0; else = 1")
```

```
##  [1] 1 1 1 1 0 1 1 0 0 0
```

```
recode(score, "lo:59 = '不及格'; 60:100 = '及格'")
```

```
##  [1] "及格"   "及格"   "及格"   "及格"   "不及格" "及格"   "及格"   "不及格"
##  [9] "不及格" "不及格"
```

```
recode(score, "lo:40 = 1; 41:60 = 2; 61:80 = 3; 81:hi = 4; else = NA")
```

```
##  [1] 3 4 3 3 1 4 3 1 1 2
```

```
recode(score, "0:40 = 'A'; 40:60 = 'B'; 60:80 = 'C'; 80:100 = 'D'; else = NA")
```

```
##  [1] "C" "D" "C" "C" "A" "D" "C" "A" "A" "B"
```

第二部分~

```
matrix(c(1, 2, 3, 4, 5, 6), nrow = 2, ncol = 3)
```

```
##      [,1] [,2] [,3]
## [1,]    1    3    5
## [2,]    2    4    6
```

```
c1 = c(1, 2, 3, 4)
c2 = c(5, 6, 7, 8)
X = data.frame(c1, c2)
X
```

```
##   c1 c2
## 1  1  5
## 2  2  6
## 3  3  7
## 4  4  8
```

```
as.matrix(X)
```

```
##      c1 c2
## [1,]  1  5
## [2,]  2  6
## [3,]  3  7
## [4,]  4  8
```

```
# w 是文字向量
w = c("100", "0200", "0030.52")

# 轉換為整數
as.integer(w)
```

```
## [1] 100 200  30
```

```
# 轉換為數值（可保留小數）
as.numeric(w)
```

```
## [1] 100.00 200.00  30.52
```

```
c1 = c(1, 2, 3)
c2 = c(11, 12, 13)
c3 = c(31, 32, 33)

c123 = c(c1, c2, c3)
c123
```

```
## [1]  1  2  3 11 12 13 31 32 33
```

```
union(c1, c2)
```

```
## [1]  1  2  3 11 12 13
```

```
c1 = c(1, 2, 3)
c2 = c(11, 12, 13)
c3 = c(31, 32, 33)

c123.rows = rbind(c1, c2, c3)
c123.rows
```

```
##    [,1] [,2] [,3]
## c1    1    2    3
## c2   11   12   13
## c3   31   32   33
```

```
c123.columns = cbind(c1, c2, c3)
c123.columns
```

```
##      c1 c2 c3
## [1,]  1 11 31
## [2,]  2 12 32
## [3,]  3 13 33
```

```
c(c123.rows, c123.columns)
```

```
##  [1]  1 11 31  2 12 32  3 13 33  1  2  3 11 12 13 31 32 33
```

```
rbind(c123.rows, c123.columns)
```

```
##    c1 c2 c3
## c1  1  2  3
## c2 11 12 13
## c3 31 32 33
##     1 11 31
##     2 12 32
##     3 13 33
```

```
c4 = cbind(c123.rows, c123.columns)
c4
```

```
##             c1 c2 c3
## c1  1  2  3  1 11 31
## c2 11 12 13  2 12 32
## c3 31 32 33  3 13 33
```