

2020/11/13(五), 109 學年第一學期 資料科學應用 R 作業(3)

學號:A106260020 姓名:楊錄綺

```
> #ex1.25(a)
> score0 <- read_excel("data/R-score.xlsx", sheet = "工作表 1", na = "NA", skip = 1)
New names:
* `0.15` -> `0.15...6`
* `0.15` -> `0.15...7`
> head(score0)
# A tibble: 6 x 10
  No 系級 學號 姓名 `0.1` `0.15...6` `0.15...7` `0.2`
  <dbl> <chr> <dbl> <chr> <dbl> <dbl> <dbl> <dbl>
1 1 統計系 1... 3.26e7 周小如... 55 95 100 100
2 2 統計系 1... 3.26e7 周抒如... 30 65 70 100
3 3 會計系 1... 3.26e7 林育安... 10 5 25 10
4 4 會計系 1... 3.26e7 林育辰... 10 20 45 40
5 5 會計系 1... 3.26e7 黃季晴... 5 15 20 25
6 6 統計系 1... 3.26e7 詹宜瑄... 10 35 60 0
# ... with 2 more variables: `0.4` <dbl>, `10 分` <dbl>
```

```
> #ex1.25(b)
> lapply(score0[5:9], mean)
```

```
$`0.1`
```

```
[1] 25
```

```
$`0.15...6`
```

```
[1] 36.15385
```

```
$`0.15...7`
```

```
[1] 51.15385
```

```
$`0.2`
```

```
[1] 51.15385
```

```
$`0.4`
```

```
[1] 77.23077
```

```
> lapply(score0[5:9], sd)
```

```
$`0.1`
```

```
[1] 18.37117
```

```
$`0.15...6`
```

```
[1] 33.05008
```

```
$`0.15...7`
```

```
[1] 26.7047
```

```
$`0.2`
```

```
[1] 38.57643
```

```
$`0.4`
```

```
[1] 23.89963
```

```
> #exl1.25(c)
```

```
> score1 <- 0.1*score0[,5]+0.15*score0[,6]+0.15*score0[,7]+0.20*score0[,8]
```

```
> +0.40*score0[,9]
```

```
0.4
```

```
1 34.4
```

```
2 37.6
```

```
3 30.8
```

```
4 34.8
```

```
5 34.4
```

```
6 30.8
```

```
7 34.8
```

```
8 31.6
```

```
9 34.8
```

```
10 30.4
```

```
11 31.2
```

```
12 36.0
```

```
13 0.0
```

```
> total <- data.frame(score0[,3], score1)
```

```
> names(total) <- c("學號", "總平均")
```

```
> #exl1.29(a)
```

```
> r.score <- read_excel("data/R-score.xlsx", na = "NA", skip = 1)
```

```
New names:
```

```
* `0.15` -> `0.15...6`
```

```
* `0.15` -> `0.15...7`
```

```
> head(r.score)
```

```
# A tibble: 6 x 10
```

```
      No 系級 學號 姓名 `0.1` `0.15...6` `0.15...7` `0.2`  
  <dbl> <chr> <dbl> <chr> <dbl>      <dbl> <dbl> <dbl>  
1     1 統計系 1... 3.26e7 周小如...    55      95    100    100  
2     2 統計系 1... 3.26e7 周抒如...    30      65     70    100  
3     3 會計系 1... 3.26e7 林育安...    10       5     25     10  
4     4 會計系 1... 3.26e7 林育辰...    10      20     45     40  
5     5 會計系 1... 3.26e7 黃季晴...     5      15     20     25  
6     6 統計系 1... 3.26e7 詹宜瑄...    10      35     60      0
```

```
# ... with 2 more variables: `0.4` <dbl>, `10 分` <dbl>
```

```
> tail(r.score)
```

```
# A tibble: 6 x 10
```

```
      No 系級 學號 姓名 `0.1` `0.15...6` `0.15...7` `0.2`  
  <dbl> <chr> <dbl> <chr> <dbl>      <dbl> <dbl> <dbl>  
1     8 統計系 1... 3.26e7 蔡旻憫...    55      45     75    100  
2     9 統計系 1... 3.26e7 黎奕璇...    10      15     55     55  
3    10 會計系 1... 3.25e7 蕭偲賢...    15       5     30     45  
4    11 會計系 1... 3.25e7 謝涵融...    35      10      5      0  
5    12 會計系 1... 3.26e7 羅順霓...    50     100     65    100  
6    13 統計系 1... 3.26e7 顧瀚薇...    15      10     75     30
```

```
# ... with 2 more variables: `0.4` <dbl>, `10 分` <dbl>
```

```
> str(r.score)
```

```
tibble [13 × 10] (S3: tbl_df/tbl/data.frame)
```

```
$ No      : num [1:13] 1 2 3 4 5 6 7 8 9 10 ...  
$ 系級    : chr [1:13] "統計系 1" "統計系 1" "會計系 1" "會計系 1" ...  
$ 學號    : num [1:13] 32578012 32578014 32578016 32578018 32578020 ...  
$ 姓名    : chr [1:13] "周小如" "周抒如" "林育安" "林育辰" ...  
$ 0.1     : num [1:13] 55 30 10 10 5 10 25 55 10 15 ...  
$ 0.15...6: num [1:13] 95 65 5 20 15 35 50 45 15 5 ...  
$ 0.15...7: num [1:13] 100 70 25 45 20 60 40 75 55 30 ...  
$ 0.2     : num [1:13] 100 100 10 40 25 0 60 100 55 45 ...  
$ 0.4     : num [1:13] 86 94 77 87 86 77 87 79 87 76 ...  
$ 10 分   : num [1:13] 10 10 10 10 0 0 10 10 4 7 ...
```

```
> #ex1.29(b)
```

```
> tdwether <- read.table("data/20140714-weather.txt", sep = "", header = T, fileEncoding = 'BIG5')
```

```
> head(tdwether)
```

```
locationName  lat      lon stationId TEMP ELEV  
1      基隆 25.1348 121.7321   466940 29.1   27  
2      淡水 25.1656 121.4400   466900 28.5   19
```

```

3      板橋 24.9993 121.4338    466880 29.0    10
4      竹子湖 25.1650 121.5363    466930 25.2   607
5      新竹 24.8300 121.0061    467571 29.8    34
6      臺中 24.1475 120.6759    467490 29.4    84

```

```
> tail(tdwether)
```

```

      locationName      lat      lon stationId TEMP ELEV
24      新店 24.9608 121.5165    A0A9M0 26.8    24
25      臺北 25.0396 121.5067    466920 30.4     5
26      臺南 22.9952 120.1970    467410 30.0    41
27      金門 24.4074 118.2893    467110 28.4    48
28      馬祖 26.1694 119.9232    467990 28.0    98
29      新屋 25.0067 121.0475    467050 29.3    21

```

```
> str(tdwether)
```

```
'data.frame': 29 obs. of 6 variables:
```

```

 $ locationName: chr  "基隆" "淡水" "板橋" "竹子湖" ...
 $ lat         : num   25.1 25.2 25 25.2 24.8 ...
 $ lon         : num   122 121 121 122 121 ...
 $ stationId   : chr   "466940" "466900" "466880" "466930" ...
 $ TEMP        : num   29.1 28.5 29 25.2 29.8 29.4 29.2 27.8 22.8 14.4 ...
 $ ELEV        : int    27 19 10 607 34 84 7 11 1015 2413 ...

```

```
> #ex1.29(c)
```

```
> delayw <- read.csv("data/weather_delays14.csv", header = T)
```

```
> head(delayw)
```

```

      year month day dep_time arr_time carrier tailnum flight
1 2014      1   1    1733     2024      AA   N3HPAA     199
2 2014      1   1    1718     1840      B6   N324JB    1734
3 2014      1   1     624     946      DL   N3751B     479
4 2014      1   1     910    1203      DL   N910DL    1174
5 2014      1   1    1850    2052      MQ   N1EAMQ    2839
6 2014      1   2    2049     45      AA   N319AA     21

```

```
      origin dest carrier_delay weather_delay nas_delay
```

```

1   JFK   ORD              0              7          51
2   JFK   BTV              0             18           6
3   JFK   ATL              0              9          45
4   LGA   PBI              0             52           0
5   LGA   STL              0             35          12
6   JFK   LAX              0             87          41

```

```
      aircraft_delay
```

```
1              11
```

```

2          0
3          0
4          0
5          0
6         22

```

```
> tail(delayw)
```

```

      year month day dep_time arr_time carrier tailnum flight
4654 2014    10  26    1039    1413      VX  N634VA      23
4655 2014    10  26    1135    1451      VX  N836VA     409
4656 2014    10  27    1042    1416      VX  N642VA     187
4657 2014    10  29    1507    1808      DL  N321NB    1923
4658 2014    10  31    1500    1751      DL  N338NB    1685
4659 2014    10  31    1323    1502      AA  N3KNAA     329

```

```

      origin dest carrier_delay weather_delay nas_delay
4654   JFK   SFO              0              9          9
4655   JFK   LAX              5             11          0
4656   EWR   SFO             12              9          0
4657   LGA   MIA              0             81          0
4658   LGA   MCO              0             28          0
4659   LGA   ORD              0            113          4

```

```

      aircraft_delay
4654                0
4655                0
4656                0
4657                0
4658                0
4659                0

```

```
> str(delayw)
```

```
'data.frame': 4659 obs. of 14 variables:
```

```

 $ year      : int  2014 2014 2014 2014 2014 2014 2014 2014 2014 2014 ...
 $ month     : int   1 1 1 1 1 1 1 1 1 1 ...
 $ day       : int   1 1 1 1 1 2 2 2 2 2 ...
 $ dep_time  : int  1733 1718 624 910 1850 2049 738 5 1618 1657 ...
 $ arr_time  : int  2024 1840 946 1203 2052 45 1124 339 1958 2050 ...
 $ carrier   : chr   "AA" "B6" "DL" "DL" ...
 $ tailnum   : chr   "N3HPAA" "N324JB" "N3751B" "N910DL" ...
 $ flight    : int   199 1734 479 1174 2839 21 33 185 133 145 ...
 $ origin    : chr   "JFK" "JFK" "JFK" "LGA" ...
 $ dest      : chr   "ORD" "BTV" "ATL" "PBI" ...
 $ carrier_delay : int   0 0 0 0 0 0 0 0 0 0 ...

```

```
$ weather_delay : int  7 18 9 52 35 87 8 53 32 6 ...
$ nas_delay      : int  51 6 45 0 12 41 26 14 5 18 ...
$ aircraft_delay: int  11 0 0 0 0 22 0 97 1 101 ...
```

```
> #exl2.10
```

```
> score <- sample(1:100, 50, replace = TRUE)
```

```
> score
```

```
[1] 47 79 47 79 59 62 33 62 90 69 14 75 67 31 8
[16] 46 86 60 62 66 1 86 74 38 77 88 87 56 15 10
[31] 82 94 81 4 40 34 100 67 63 67 59 45 57 95 73
[46] 12 31 40 8 24
```

```
> if(any(score > 95)){
```

```
+   cat("老師請同學吃飯")
```

```
+ }else{
```

```
+   cat("老師很生氣")}
```

```
老師請同學吃飯>
```

```
> #exl2.21(a)
```

```
> score02 <- read.csv("data/score02.csv", header = T, fileEncoding = 'BIG5')
```

```
> score02[1:7,]
```

```
學號 期中考 期末考
```

| | | | |
|---|-----------|----|----|
| 1 | 410072106 | 80 | 60 |
| 2 | 410073023 | 50 | 73 |
| 3 | 410079062 | 45 | 35 |
| 4 | 410079090 | 77 | 54 |
| 5 | 410079118 | 62 | 54 |
| 6 | 410079120 | 67 | 45 |
| 7 | 410079121 | 72 | 78 |

```
> #exl2.21(b)
```

```
> names(score02) <- c("id", "mid", "final")
```

```
> names(score02)
```

```
[1] "id" "mid" "final"
```

```
> #exl2.21(c)
```

```
> progress <- subset(score02, final > mid)
```

```
> progress$id
```

```
[1] 410073023 410079121 410172016 410172027 410173072 410173136
[7] 410174210 410273014 410273016 410273042 410273048 410273062
[13] 410273067 410273073 410273076 410273108 410273116 410275016
[19] 410275029 410275051 410279018 410279049 410279054 410279063
```

[25] 410279075 49981011

```
> #exl2.21(d)
> A <- nrow(subset(score02, mid >= 60 & final >=60))
> B <- nrow(subset(score02, mid >= 60 & final <=60))
> C <- nrow(subset(score02, mid <= 60 & final >=60))
> D <- nrow(subset(score02, mid <= 60 & final <=60))
> library(Hmisc)
> label(A) <- "期中期末都及格"
> label(B) <- "期中及格，期末不及格"
> label(C) <- "期中不及格，期末及格"
> label(D) <- "期中不及格，期末不及格"
```

```
> A
```

期中期末都及格

```
[1] 38
```

```
> B
```

期中及格，期末不及格

```
[1] 33
```

```
> C
```

期中不及格，期末及格

```
[1] 10
```

```
> D
```

期中不及格，期末不及格

```
[1] 19
```

```
> #exl2.21(e)
```

```
> meanscore <- 0.5*score02[,2]+0.5*score02[,3]
```

```
> sort(meanscore, decreasing = T)
```

```
[1] 100.0 100.0 96.5 92.5 92.0 91.0 88.5 88.0 88.0 87.0
[11] 86.0 86.0 85.0 85.0 84.0 83.0 82.5 81.5 81.0 79.0
[21] 79.0 78.5 78.0 77.5 77.5 77.5 77.5 77.0 77.0 76.5
[31] 75.5 75.0 75.0 75.0 74.0 73.5 73.0 70.0 69.5 69.0
[41] 68.5 68.5 68.5 68.5 68.5 68.0 65.5 65.5 65.0 64.0
[51] 64.0 63.5 63.5 63.0 62.5 62.5 62.5 61.5 61.5 61.0
[61] 60.5 59.0 58.5 58.0 57.5 57.5 57.0 56.5 56.0 56.0
[71] 56.0 55.0 54.0 53.5 52.0 51.0 51.0 51.0 50.0 49.5
[81] 49.0 48.5 47.5 45.5 45.0 44.5 43.0 42.5 40.0 37.0
[91] 35.0 26.5 26.5 23.0
```

```
>
```