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

學號: A106260052　　姓名: 劉映妤

(請依照規定)貼上執行程式碼及執行結果。
詳見: R 程式作業繳交方式
http://www.hmwu.idv.tw/web/teaching/doc/R-how-homework.pdf

```
> #2020/11/13
>
> # ex1.25(a)
>   #讀取資料檔，印出前 5 位同學成績紀錄
> library(readxl)
> xlsx_file <- "data/R-score.xlsx"
> excel_sheets(xlsx_file)
[1] "工作表 1"
> mydata <- read_excel(xlsx_file, sheet = "工作表 1", na = "NA")
New names:
* `` -> ...2
* `` -> ...3
* `` -> ...4
> head(mydata, 6)
# A tibble: 6 x 10
  `115-2-R 程式設計`~ ...2  ...3  ...4  `小考(1)`
  <chr>                <chr> <chr> <chr>       <dbl>
1 No                   系級  學號  姓名          0.1
2 1                    統計系 1~ 3257~ 周小如~      55
3 2                    統計系 1~ 3257~ 周抒如~      30
4 3                    會計系 1~ 3257~ 林育安~      10
5 4                    會計系 1~ 3257~ 林育辰~      10
6 5                    會計系 1~ 3257~ 黃季晴~       5
# ... with 5 more variables: `小考(2)` <dbl>,
#   `小考(3)` <dbl>, 作業 <dbl>, 期末考 <dbl>,
#   點名 <chr>
>
> # ex1.25(b)
>   #計算各項考試 (不含點名) 平均分數及標準差
```

```
>   #小考一
> quiz1 <- mean(rowMeans(mydata[2:14 ,5]))
> quiz1
[1] 25
> quiz1.2 <- sd(rowMeans(mydata[2:14 ,5]))
> quiz1.2
[1] 18.37117
>   #小考二
> quiz2 <- mean(rowMeans(mydata[2:14 ,6]))
> quiz2
[1] 36.15385
> quiz2.2 <- sd(rowMeans(mydata[2:14 ,6]))
> quiz2.2
[1] 33.05008
>   #小考三
> quiz3 <- mean(rowMeans(mydata[2:14 ,7]))
> quiz3
[1] 51.15385
> quiz3.2 <- sd(rowMeans(mydata[2:14 ,7]))
> quiz3.2
[1] 26.7047
>   #作業
> quiz4 <- mean(rowMeans(mydata[2:14 ,8]))
> quiz4
[1] 51.15385
> quiz4.2 <- sd(rowMeans(mydata[2:14 ,8]))
> quiz4.2
[1] 38.57643
>   #期末考
> quiz5 <- mean(rowMeans(mydata[2:14 ,9]))
> quiz5
[1] 77.23077
> quiz5.2 <- sd(rowMeans(mydata[2:14 ,9]))
> quiz5.2
[1] 23.89963
>
> # ex1.25(c)
> Semester_grades <-
```

```
(mydata[2:14 ,5])*0.1+(mydata[2:14 ,6])*0.15+(mydata[2:14 ,7])*0.15+(mydata[2:14
,8])*0.2+(mydata[2:14 ,9])*0.4
> names(Semester_grades) <- "學期成績"
>
> Semester_grades.1 <- data.frame(mydata[2:14,3],Semester_grades)
> names(Semester_grades.1) <- "學號"
> colnames(Semester_grades.1)[2] <- "學期成績"
>
> #ex1.29(a)
> library(readxl)
> xlsx_file <- "data/R-score.xlsx"
> excel_sheets(xlsx_file)
[1] "工作表 1"
> mydata.1 <- read_excel(xlsx_file,    range = "A2:I15")
New names:
* `0.15` -> `0.15...6`
* `0.15` -> `0.15...7`
> front.1 <-head(mydata.1, 5)
> front.1
# A tibble: 5 x 9
        No  系級       學號  姓名    `0.1` `0.15...6`
   <dbl> <chr>    <dbl> <chr> <dbl>          <dbl>
1        1  統計系 1~ 3.26e7  周小如~       55             95
2        2  統計系 1~ 3.26e7  周抒如~       30             65
3        3  會計系 1~ 3.26e7  林育安~       10              5
4        4  會計系 1~ 3.26e7  林育辰~       10             20
5        5  會計系 1~ 3.26e7  黃季晴~        5             15
# ... with 3 more variables: `0.15...7` <dbl>,
#      `0.2` <dbl>, `0.4` <dbl>
> str(front.1)
tibble [5 x 9] (S3: tbl_df/tbl/data.frame)
  $ No           : num [1:5] 1 2 3 4 5
  $  系級        : chr [1:5] "統計系 1" "統計系 1" "會計系 1" "會計系 1" ...
  $  學號        : num [1:5] 32578012 32578014 32578016 32578018 32578020
  $  姓名        : chr [1:5] "周小如" "周抒如" "林育安" "林育辰" ...
  $ 0.1          : num [1:5] 55 30 10 10 5
  $ 0.15...6: num [1:5] 95 65 5 20 15
  $ 0.15...7: num [1:5] 100 70 25 45 20
```

```
 $ 0.2         : num [1:5] 100 100 10 40 25
 $ 0.4         : num [1:5] 86 94 77 87 86
> end.1 <- tail(mydata.1, 5)
> end.1
# A tibble: 5 x 9
       No 系級       學號 姓名   `0.1` `0.15...6`
   <dbl> <chr>    <dbl> <chr> <dbl>          <dbl>
1      9 統計系 1~ 3.26e7 黎奕璇~       10             15
2     10 會計系 1~ 3.25e7 蕭偲賢~       15              5
3     11 會計系 1~ 3.25e7 謝涵融~       35             10
4     12 會計系 1~ 3.26e7 羅順霓~       50            100
5     13 統計系 1~ 3.26e7 顧瀚薇~       15             10
# ... with 3 more variables: `0.15...7` <dbl>,
#     `0.2` <dbl>, `0.4` <dbl>
> str(end.1)
tibble [5 x 9] (S3: tbl_df/tbl/data.frame)
 $ No          : num [1:5] 9 10 11 12 13
 $ 系級        : chr [1:5] "統計系 1" "會計系 1" "會計系 1" "會計系 1" ...
 $ 學號        : num [1:5] 32578030 32474226 32475032 32578002 32578004
 $ 姓名        : chr [1:5] "黎奕璇" "蕭偲賢" "謝涵融" "羅順霓" ...
 $ 0.1         : num [1:5] 10 15 35 50 15
 $ 0.15...6: num [1:5] 15 5 10 100 10
 $ 0.15...7: num [1:5] 55 30 5 65 75
 $ 0.2         : num [1:5] 55 45 0 100 30
 $ 0.4         : num [1:5] 87 76 78 90 0
>
> #ex1.29(b)
> weather.1 <- read.delim("data/20140714-weather.txt")
> front.2 <- head(weather.1,5)
> front.2
   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
> str(front.2)
'data.frame':   5 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
 $ ELEV         : int  27 19 10 607 34
> end.2 <- tail(weather.1,5)
> end.2
    locationName      lat      lon stationId TEMP ELEV
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(end.2)
'data.frame':   5 obs. of  6 variables:
 $ locationName: chr  "臺北" "臺南" "金門" "馬祖" ...
 $ lat          : num  25 23 24.4 26.2 25
 $ lon          : num  122 120 118 120 121
 $ stationId    : chr  "466920" "467410" "467110" "467990" ...
 $ TEMP         : num  30.4 30 28.4 28 29.3
 $ ELEV         : int  5 41 48 98 21
>
> #ex1.29(c)
> library(haven)
> weather.2 <- read.csv("data/weather_delays14.csv")
> front.3 <- head(weather.2,5)
> front.3
  year month day dep_time arr_time carrier tailnum
1 2014     1   1     1733     2024      AA  N3HPAA
2 2014     1   1     1718     1840      B6  N324JB
3 2014     1   1      624      946      DL  N3751B
4 2014     1   1      910     1203      DL  N910DL
5 2014     1   1     1850     2052      MQ  N1EAMQ
  flight origin dest carrier_delay weather_delay
1    199    JFK  ORD             0             7
2   1734    JFK  BTV             0            18
3    479    JFK  ATL             0             9
```

```
4     1174       LGA    PBI                  0                  52
5     2839       LGA    STL                  0                  35
   nas_delay aircraft_delay
1             51                   11
2              6                    0
3             45                    0
4              0                    0
5             12                    0
> str(front.3)
'data.frame':   5 obs. of   14 variables:
 $ year                : int   2014 2014 2014 2014 2014
 $ month               : int    1 1 1 1 1
 $ day                 : int   1 1 1 1 1
 $ dep_time            : int   1733 1718 624 910 1850
 $ arr_time            : int   2024 1840 946 1203 2052
 $ carrier             : chr   "AA" "B6" "DL" "DL" ...
 $ tailnum             : chr   "N3HPAA" "N324JB" "N3751B" "N910DL" ...
 $ flight              : int   199 1734 479 1174 2839
 $ origin              : chr   "JFK" "JFK" "JFK" "LGA" ...
 $ dest                : chr   "ORD" "BTV" "ATL" "PBI" ...
 $ carrier_delay : int   0 0 0 0 0
 $ weather_delay : int   7 18 9 52 35
 $ nas_delay           : int   51 6 45 0 12
 $ aircraft_delay: int   11 0 0 0 0
> end.3 <- tail(weather.2,5)
> end.3
        year month day dep_time arr_time carrier
4655 2014       10   26        1135        1451         VX
4656 2014       10   27        1042        1416         VX
4657 2014       10   29        1507        1808         DL
4658 2014       10   31        1500        1751         DL
4659 2014       10   31        1323        1502         AA
        tailnum flight origin dest carrier_delay
4655    N836VA       409       JFK    LAX                   5
4656    N642VA       187       EWR    SFO                  12
4657    N321NB      1923       LGA    MIA                   0
4658    N338NB      1685       LGA    MCO                   0
4659    N3KNAA       329       LGA    ORD                   0
```

```
      weather_delay nas_delay aircraft_delay
4655            11         0              0
4656             9         0              0
4657            81         0              0
4658            28         0              0
4659           113         4              0
> str(end.3)
'data.frame':   5 obs. of  14 variables:
 $ year          : int  2014 2014 2014 2014 2014
 $ month         : int  10 10 10 10 10
 $ day           : int  26 27 29 31 31
 $ dep_time      : int  1135 1042 1507 1500 1323
 $ arr_time      : int  1451 1416 1808 1751 1502
 $ carrier       : chr  "VX" "VX" "DL" "DL" ...
 $ tailnum       : chr  "N836VA" "N642VA" "N321NB" "N338NB" ...
 $ flight        : int  409 187 1923 1685 329
 $ origin        : chr  "JFK" "EWR" "LGA" "LGA" ...
 $ dest          : chr  "LAX" "SFO" "MIA" "MCO" ...
 $ carrier_delay : int  5 12 0 0 0
 $ weather_delay : int  11 9 81 28 113
 $ nas_delay     : int  0 0 0 0 4
 $ aircraft_delay: int  0 0 0 0 0
>
> #ex2.10
> score <- sample(1:100, 50, replace = TRUE)
> score
 [1]  82  56  49  97  23  33  10  82  72  23  63  62
[13]   8  99   7   1  39  74  66  73  91  65  84  65
[25]  33  50  12  34  18  27  90  45   5  41  27  73
[37]  68  41  72  25  72  60  60 100  53  12  25  41
[49]  35   3
> y <- numeric(length(score))
> y[score>=95] <- 1
> y[score<95] <- 0
> y
 [1] 0 0 0 1 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0
[25] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0
[49] 0 0
```

```
> if (sum(y)>1) cat("老師請同學吃飯") else cat("老師很生氣")
老師請同學吃飯>
> #ex2.21(a)
> library(haven)
> score02 <- read.csv("data/score02.csv")
> front.score02 <- head(score02,7)
> front.score02
        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
>
> #ex2.21(b)
> colnames(score02)[1] <- "id"
> colnames(score02)[2] <- "mid"
> colnames(score02)[3] <- "final"
> score02
          id mid final
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
8  410172016  62    75
9  410172027  82    95
10 410172103  92    66
11 410173029  42    11
12 410173072  55    73
13 410173101  82    64
14 410173134  92    78
15 410173135 100    55
16 410173136  80    88
```

17	410174210	50	63
18	410183004	95	90
19	410183012	67	35
20	410184012	75	16
21	410184015	52	45
22	410273002	100	25
23	410273004	99	56
24	410273005	60	55
25	410273007	100	76
26	410273010	72	40
27	410273011	55	45
28	410273014	45	57
29	410273016	62	100
30	410273018	100	25
31	410273019	70	67
32	410273020	95	55
33	410273024	75	55
34	410273031	85	68
35	410273032	75	64
36	410273034	70	47
37	410273040	67	56
38	410273041	57	28
39	410273042	70	85
40	410273048	52	62
41	410273049	72	40
42	410273050	57	42
43	410273051	47	6
44	410273057	80	70
45	410273060	50	40
46	410273062	60	76
47	410273065	85	70
48	410273067	70	86
49	410273069	82	65
50	410273070	100	72
51	410273073	75	88
52	410273075	87	40
53	410273076	47	75
54	410273081	90	31

55	410273094	100	8
56	410273095	90	64
57	410273096	87	70
58	410273102	100	100
59	410273105	85	52
60	410273106	80	71
61	410273108	90	94
62	410273109	90	80
63	410273110	87	87
64	410273116	82	100
65	410275001	61	9
66	410275005	92	73
67	410275015	52	43
68	410275016	55	60
69	410275017	57	47
70	410275020	95	81
71	410275029	79	93
72	410275032	85	33
73	410275033	60	29
74	410275034	85	81
75	410275036	72	26
76	410275040	70	57
77	410275051	35	90
78	410275055	85	53
79	410275058	100	100
80	410279001	100	48
81	410279006	32	14
82	410279018	47	55
83	410279021	42	32
84	410279039	90	41
85	410279049	47	60
86	410279054	32	54
87	410279063	72	82
88	410279075	38	90
89	410279080	90	36
90	49973086	82	76
91	49979003	85	25
92	49979046	82	55

```
93    49981006    82       55
94    49981011    95       98
>
> #ex2.21(c)
> improve1 <- ifelse((score02$final-score02$mid)>0,"1","0")
> score03 <- data.frame(score02, improve1)
> improve3 <- ifelse(score03$improve1==1, score03$id,'')
> improve3
  [1] ""           "410073023" ""           ""
  [5] ""           ""          "410079121" "410172016"
  [9] "410172027" ""           ""          "410173072"
[13] ""           ""          ""          "410173136"
[17] "410174210" ""           ""          ""
[21] ""           ""          ""          ""
[25] ""           ""          ""          "410273014"
[29] "410273016" ""           ""          ""
[33] ""           ""          ""          ""
[37] ""           ""          "410273042" "410273048"
[41] ""           ""          ""          ""
[45] ""           "410273062" ""          "410273067"
[49] ""           ""          "410273073" ""
[53] "410273076" ""           ""          ""
[57] ""           ""          ""          ""
[61] "410273108" ""           ""          "410273116"
[65] ""           ""          ""          "410275016"
[69] ""           ""          "410275029" ""
[73] ""           ""          ""          ""
[77] "410275051" ""           ""          ""
[81] ""           "410279018" ""          ""
[85] "410279049" "410279054" "410279063" "410279075"
[89] ""           ""          ""          ""
[93] ""           "49981011"
>
> #ex2.21(d)
> grade1 <- ifelse(score02$mid>=60,ifelse(score02$final>=60,"都及格","期中及格，
期末不及格"),ifelse(score02$final>=60,"期中不及格，期末及格","都不及格"))
> grade1
  [1] "都及格"                    "期中不及格，期末及格"
```

```
 [3] "都不及格"                 "期中及格，期末不及格"
 [5] "期中及格，期末不及格" "期中及格，期末不及格"
 [7] "都及格"                   "都及格"
 [9] "都及格"                   "都及格"
[11] "都不及格"                 "期中不及格，期末及格"
[13] "都及格"                   "都及格"
[15] "期中及格，期末不及格" "都及格"
[17] "期中不及格，期末及格" "都及格"
[19] "期中及格，期末不及格" "期中及格，期末不及格"
[21] "都不及格"                 "期中及格，期末不及格"
[23] "期中及格，期末不及格" "期中及格，期末不及格"
[25] "都及格"                   "期中及格，期末不及格"
[27] "都不及格"                 "都不及格"
[29] "都及格"                   "期中及格，期末不及格"
[31] "都及格"                   "期中及格，期末不及格"
[33] "期中及格，期末不及格" "都及格"
[35] "都及格"                   "期中及格，期末不及格"
[37] "期中及格，期末不及格" "都不及格"
[39] "都及格"                   "期中不及格，期末及格"
[41] "期中及格，期末不及格" "都不及格"
[43] "都不及格"                 "都及格"
[45] "都不及格"                 "都及格"
[47] "都及格"                   "都及格"
[49] "都及格"                   "都及格"
[51] "都及格"                   "期中及格，期末不及格"
[53] "期中不及格，期末及格" "期中及格，期末不及格"
[55] "期中及格，期末不及格" "都及格"
[57] "都及格"                   "都及格"
[59] "期中及格，期末不及格" "都及格"
[61] "都及格"                   "都及格"
[63] "都及格"                   "都及格"
[65] "期中及格，期末不及格" "都及格"
[67] "都不及格"                 "期中不及格，期末及格"
[69] "都不及格"                 "都及格"
[71] "都及格"                   "期中及格，期末不及格"
[73] "期中及格，期末不及格" "都及格"
[75] "期中及格，期末不及格" "期中及格，期末不及格"
[77] "期中不及格，期末及格" "期中及格，期末不及格"
```

```
[79] "都及格"                    "期中及格，期末不及格"
[81] "都不及格"                  "都不及格"
[83] "都不及格"                  "期中及格，期末不及格"
[85] "期中不及格，期末及格" "都不及格"
[87] "都及格"                    "期中不及格，期末及格"
[89] "期中及格，期末不及格" "都及格"
[91] "期中及格，期末不及格" "期中及格，期末不及格"
[93] "期中及格，期末不及格" "都及格"
> table(grade1)
grade1
                 都不及格                      都及格
                       15                          38
期中不及格，期末及格  期中及格，期末不及格
                        9                          32
>
> #ex2.21(e)
> library(dplyr)
> final <- (score02$mid+score02$final)/2
> final.1 <- data.frame(score02$id,final)
> arrange(final.1,desc(final))
    score02.id final
1    410273102 100.0
2    410275058 100.0
3     49981011  96.5
4    410183004  92.5
5    410273108  92.0
6    410273116  91.0
7    410172027  88.5
8    410273007  88.0
9    410275020  88.0
10   410273110  87.0
11   410273070  86.0
12   410275029  86.0
13   410173134  85.0
14   410273109  85.0
15   410173136  84.0
16   410275034  83.0
17   410275005  82.5
```

18	410273073	81.5
19	410273016	81.0
20	410172103	79.0
21	49973086	79.0
22	410273096	78.5
23	410273067	78.0
24	410173135	77.5
25	410273004	77.5
26	410273042	77.5
27	410273065	77.5
28	410273095	77.0
29	410279063	77.0
30	410273031	76.5
31	410273106	75.5
32	410079121	75.0
33	410273020	75.0
34	410273057	75.0
35	410279001	74.0
36	410273069	73.5
37	410173101	73.0
38	410072106	70.0
39	410273032	69.5
40	410275055	69.0
41	410172016	68.5
42	410273019	68.5
43	410273105	68.5
44	49979046	68.5
45	49981006	68.5
46	410273062	68.0
47	410079090	65.5
48	410279039	65.5
49	410273024	65.0
50	410173072	64.0
51	410279075	64.0
52	410273075	63.5
53	410275040	63.5
54	410279080	63.0
55	410273002	62.5

56	410273018	62.5
57	410275051	62.5
58	410073023	61.5
59	410273040	61.5
60	410273076	61.0
61	410273081	60.5
62	410275032	59.0
63	410273034	58.5
64	410079118	58.0
65	410273005	57.5
66	410275016	57.5
67	410273048	57.0
68	410174210	56.5
69	410079120	56.0
70	410273010	56.0
71	410273049	56.0
72	49979003	55.0
73	410273094	54.0
74	410279049	53.5
75	410275017	52.0
76	410183012	51.0
77	410273014	51.0
78	410279018	51.0
79	410273011	50.0
80	410273050	49.5
81	410275036	49.0
82	410184015	48.5
83	410275015	47.5
84	410184012	45.5
85	410273060	45.0
86	410275033	44.5
87	410279054	43.0
88	410273041	42.5
89	410079062	40.0
90	410279021	37.0
91	410275001	35.0
92	410173029	26.5
93	410273051	26.5

94 410279006 23.0
>