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

學號:A106260093　　　姓名:王緯華

(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 作業
>
> #ex1.25(a) 讀取資料檔，印出前 5 位同學成績紀錄
> #讀 excel 檔
> library(readxl)
> student_test <- 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`
> colnames(student_test) <- c("NO", "系級", "學號", "姓名", "Quiz1", "Quiz2", "Quiz3", "HomeWork", "finaltest", "RollCall")
> head(student_test, 5)
# A tibble: 5 x 10
      NO 系級        學號 姓名   Quiz1 Quiz2 Quiz3 HomeWork finaltest RollCall
   <dbl> <chr>      <dbl> <chr> <dbl> <dbl> <dbl>    <dbl>     <dbl>    <dbl>
1      1 統計系 1 32578012 周小如    55    95   100      100        86       10
2      2 統計系 1 32578014 周抒如    30    65    70      100        94       10
3      3 會計系 1 32578016 林育安    10     5    25       10        77       10
4      4 會計系 1 32578018 林育辰    10    20    45       40        87       10
5      5 會計系 1 32578020 黃季晴     5    15    20       25        86        0
>
> #ex1.25(b)  計算各項考試 (不含點名) 平均分數及標準差
```

```
> str(student_test)
tibble [13 x 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" ...
 $ Quiz1      : num [1:13] 55 30 10 10 5 10 25 55 10 15 ...
 $ Quiz2      : num [1:13] 95 65 5 20 15 35 50 45 15 5 ...
 $ Quiz3      : num [1:13] 100 70 25 45 20 60 40 75 55 30 ...
 $ HomeWork : num [1:13] 100 100 10 40 25 0 60 100 55 45 ...
 $ finaltest: num [1:13] 86 94 77 87 86 77 87 79 87 76 ...
 $ RollCall : num [1:13] 10 10 10 10 0 0 10 10 4 7 ...
> #Quiz1
> mean(student_test$Quiz1)
[1] 25
> sd(student_test$Quiz1)
[1] 18.37117
> #Quiz2
> mean(student_test$Quiz2)
[1] 36.15385
> sd(student_test$Quiz2)
[1] 33.05008
> #Quiz3
> mean(student_test$Quiz3)
[1] 51.15385
> sd(student_test$Quiz3)
[1] 26.7047
> #HomeWork
> mean(student_test$HomeWork)
[1] 51.15385
> sd(student_test$HomeWork)
[1] 38.57643
> #finaltest
> mean(student_test$finaltest)
[1] 77.23077
> sd(student_test$finaltest)
[1] 23.89963
>
```

```
> #ex1.25(c)　依照各項考試配分 (小考 1(10%), 小考 2(15%), 小考 3(15%), 作
業 (20%), 期末考 (40%)) 計算每位同學之學期成績，並以 data.frame 的類別型
式印出學號及學期成績。(其它項目不用列出)
> student_test_matrix <- data.frame(student_test$Quiz1, student_test$Quiz2,
student_test$Quiz3, student_test$HomeWork, student_test$finaltest)
> str(student_test_matrix)
'data.frame':   13 obs. of   5 variables:
 $ student_test.Quiz1    : num   55 30 10 10 5 10 25 55 10 15 ...
 $ student_test.Quiz2    : num   95 65 5 20 15 35 50 45 15 5 ...
 $ student_test.Quiz3    : num   100 70 25 45 20 60 40 75 55 30 ...
 $ student_test.HomeWork : num   100 100 10 40 25 0 60 100 55 45 ...
 $ student_test.finaltest: num   86 94 77 87 86 77 87 79 87 76 ...
> student_test$Quiz1 <- (student_test$Quiz1)*0.1
> student_test$Quiz2 <- (student_test$Quiz2)*0.15
> student_test$Quiz3 <- (student_test$Quiz3)*0.15
> student_test$HomeWork <- (student_test$HomeWork)*0.2
> student_test$finaltest <-(student_test$finaltest)*0.4
> mean <- rowMeans(student_test_matrix)
> str(mean)
 num [1:13] 87.2 71.8 25.4 40.4 30.2 36.4 52.4 70.8 44.4 34.2 ...
>
> student_test_matrix_all <-   data.frame(student_test$學號, mean)
> student_test_matrix_all
    student_test.學號  mean
1              32578012 87.2
2              32578014 71.8
3              32578016 25.4
4              32578018 40.4
5              32578020 30.2
6              32578022 36.4
7              32578026 52.4
8              32578028 70.8
9              32578030 44.4
10             32474226 34.2
11             32475032 25.6
12             32578002 81.0
13             32578004 26.0
>
```

```
> 
> #ex1.29(a)
> student_test <- 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`
> colnames(student_test) <- c("NO", "系級", "學號", "姓名", "Quiz1",  "Quiz2", "Quiz3", "HomeWork", "finaltest", "RollCall")
> head(student_test, 5)
# A tibble: 5 x 10
      NO 系級          學號 姓名    Quiz1 Quiz2 Quiz3 HomeWork finaltest RollCall
   <dbl> <chr>        <dbl> <chr>   <dbl> <dbl> <dbl>    <dbl>     <dbl>    <dbl>
1      1 統計系 1 32578012 周小如     55    95   100      100        86       10
2      2 統計系 1 32578014 周抒如     30    65    70      100        94       10
3      3 會計系 1 32578016 林育安     10     5    25       10        77       10
4      4 會計系 1 32578018 林育辰     10    20    45       40        87       10
5      5 會計系 1 32578020 黃季晴      5    15    20       25        86        0
> lapply(student_test,class)
$NO
[1] "numeric"

$系級
[1] "character"

$學號
[1] "numeric"

$姓名
[1] "character"
```

$Quiz1
[1] "numeric"

$Quiz2
[1] "numeric"

$Quiz3
[1] "numeric"

$HomeWork
[1] "numeric"

$finaltest
[1] "numeric"

$RollCall
[1] "numeric"

>
> #ex1.29(b)
> weather <- read.table("data/20140714-weather.txt",   header = TRUE, sep="\t")
> head(weather, 5)
   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
> lapply(weather,class)
$locationName
[1] "character"

$lat
[1] "numeric"

$lon
[1] "numeric"

```
$stationId
[1] "character"

$TEMP
[1] "numeric"

$ELEV
[1] "integer"

>
> #ex1.29(c)
> weather_delays14 <- read.csv("data/weather_delays14.csv", na = "NA", header =
TRUE, sep=",")
> head(weather_delays14, 5)
  year month day dep_time arr_time carrier tailnum flight origin dest carrier_delay
1 2014    1   1    1733     2024      AA   N3HPAA    199    JFK
ORD              0
2 2014    1   1    1718     1840      B6   N324JB   1734    JFK  BTV
0
3 2014    1   1     624      946      DL   N3751B    479    JFK  ATL
0
4 2014    1   1     910     1203      DL   N910DL   1174    LGA  PBI
0
5 2014    1   1    1850     2052      MQ   N1EAMQ   2839    LGA
STL              0
  weather_delay nas_delay aircraft_delay
1             7        51             11
2            18         6              0
3             9        45              0
4            52         0              0
5            35        12              0
> lapply(weather_delays14,class)
$year
[1] "integer"

$month
[1] "integer"
```

```
$day
[1] "integer"

$dep_time
[1] "integer"

$arr_time
[1] "integer"

$carrier
[1] "character"

$tailnum
[1] "character"

$flight
[1] "integer"

$origin
[1] "character"

$dest
[1] "character"

$carrier_delay
[1] "integer"

$weather_delay
[1] "integer"

$nas_delay
[1] "integer"

$aircraft_delay
[1] "integer"


> 
> 
```

```
> #ex2.10 用 ifelse 來做所有的 T/F 判斷
> score <- sample(1:100, 50, replace = TRUE)
> ifelse(score>95, "老師請同學吃飯", "老師很生氣")
 [1] "老師很生氣"     "老師很生氣"     "老師很生氣"     "老師很生氣"
 [5] "老師很生氣"     "老師很生氣"     "老師很生氣"     "老師很生氣"
 [9] "老師很生氣"     "老師很生氣"     "老師很生氣"     "老師很生氣"
[13] "老師很生氣"     "老師很生氣"     "老師很生氣"     "老師很生氣"
[17] "老師請同學吃飯" "老師很生氣"     "老師很生氣"     "老師很生氣"
[21] "老師很生氣"     "老師很生氣"     "老師很生氣"     "老師很生氣"
[25] "老師很生氣"     "老師請同學吃飯" "老師很生氣"     "老師很生氣"
[29] "老師很生氣"     "老師很生氣"     "老師很生氣"     "老師很生氣"
[33] "老師很生氣"     "老師很生氣"     "老師很生氣"     "老師很生氣"
[37] "老師很生氣"     "老師很生氣"     "老師很生氣"     "老師很生氣"
[41] "老師很生氣"     "老師很生氣"     "老師很生氣"     "老師很生氣"
[45] "老師很生氣"     "老師很生氣"     "老師很生氣"     "老師很生氣"
[49] "老師很生氣"     "老師很生氣"
>
>
> #ex2.21(a)
> score02_data <- read.csv("data/score02.csv", na = "NA", header = TRUE, sep=",")
> head(score02_data, 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
>
> #ex2.21(b)
> colnames(score02_data) <- c("id", "mid", "final")
> str(score02_data)
'data.frame':   94 obs. of    3 variables:
 $ id    : int    410072106 410073023 410079062 410079090 410079118 410079120
410079121 410172016 410172027 410172103 ...
 $ mid   : int    80 50 45 77 62 67 72 62 82 92 ...
 $ final: int    60 73 35 54 54 45 78 75 95 66 ...
```

```
> 
> #ex2.21(c)
> A <- score02_data$mid
> B <- score02_data$final
> ifelse(B>A, score02_data$id, 0)
 [1]         0 410073023         0         0         0         0
410079121 410172016
 [9] 410172027         0         0 410173072         0         0
0 410173136
[17] 410174210         0         0         0         0         0
0         0
[25]         0         0         0 410273014 410273016         0
0         0
[33]         0         0         0         0         0         0
410273042 410273048
[41]         0         0         0         0         0 410273062
0 410273067
[49]         0         0 410273073         0 410273076         0
0         0
[57]         0         0         0         0 410273108         0
0 410273116
[65]         0         0         0 410275016         0         0
410275029         0
[73]         0         0         0         0 410275051         0
0         0
[81]         0 410279018         0         0 410279049 410279054
410279063 410279075
[89]         0         0         0         0         0  49981011
> 
> #ex2.21(d) 先分"及格"再分"期中及格期末不及格"、"期末及格期中不及格",若
為"不及格"則都不及格
> ifelse(A>=60 & B>=60, "ALL_pass",
+          ifelse(A>=60, "mid_pass but final_fail",
+                 ifelse(B>=60, "final_pass but mid_fail",
+                        "All_fail")
+                 )
+          )
 [1] "ALL_pass"                "final_pass but mid_fail" "All_fail"
```

```
 [4] "mid_pass but final_fail" "mid_pass but final_fail" "mid_pass but final_fail"
 [7] "ALL_pass"                "ALL_pass"                "ALL_pass"
[10] "ALL_pass"                "All_fail"                "final_pass but
mid_fail"
[13] "ALL_pass"                "ALL_pass"                "mid_pass but
final_fail"
[16] "ALL_pass"                "final_pass but mid_fail" "ALL_pass"
[19] "mid_pass but final_fail" "mid_pass but final_fail" "All_fail"
[22] "mid_pass but final_fail" "mid_pass but final_fail" "mid_pass but final_fail"
[25] "ALL_pass"                "mid_pass but final_fail" "All_fail"
[28] "All_fail"                "ALL_pass"                "mid_pass but
final_fail"
[31] "ALL_pass"                "mid_pass but final_fail" "mid_pass but final_fail"
[34] "ALL_pass"                "ALL_pass"                "mid_pass but
final_fail"
[37] "mid_pass but final_fail" "All_fail"                "ALL_pass"
[40] "final_pass but mid_fail" "mid_pass but final_fail" "All_fail"
[43] "All_fail"                "ALL_pass"                "All_fail"
[46] "ALL_pass"                "ALL_pass"                "ALL_pass"
[49] "ALL_pass"                "ALL_pass"                "ALL_pass"
[52] "mid_pass but final_fail" "final_pass but mid_fail" "mid_pass but final_fail"
[55] "mid_pass but final_fail" "ALL_pass"                "ALL_pass"
[58] "ALL_pass"                "mid_pass but final_fail" "ALL_pass"
[61] "ALL_pass"                "ALL_pass"                "ALL_pass"
[64] "ALL_pass"                "mid_pass but final_fail" "ALL_pass"
[67] "All_fail"                "final_pass but mid_fail" "All_fail"
[70] "ALL_pass"                "ALL_pass"                "mid_pass but
final_fail"
[73] "mid_pass but final_fail" "ALL_pass"                "mid_pass but final_fail"
[76] "mid_pass but final_fail" "final_pass but mid_fail" "mid_pass but final_fail"
[79] "ALL_pass"                "mid_pass but final_fail" "All_fail"
[82] "All_fail"                "All_fail"                "mid_pass but
final_fail"
[85] "final_pass but mid_fail" "All_fail"                "ALL_pass"
[88] "final_pass but mid_fail" "mid_pass but final_fail" "ALL_pass"
[91] "mid_pass but final_fail" "mid_pass but final_fail" "mid_pass but final_fail"
[94] "ALL_pass"
>
```

```
> #ifelse(A>=60 & B>=60, A, B)
> #A = ALL_pass
> #B = ifelse(A>B, "mid_pass but final_fail",C)
> #C = ifelse(A<B, "final_pass but mid_fail", D)
> #D = All_fail
>
> #ex2.21(e)
> score02_data_AB <- data.frame(A, B)
> score02_data_mean <- rowMeans(score02_data_AB)
> score02_data_mean
 [1]   70.0  61.5  40.0  65.5  58.0  56.0  75.0  68.5  88.5  79.0  26.5
64.0   73.0
[14]   85.0  77.5  84.0  56.5  92.5  51.0  45.5  48.5  62.5  77.5  57.5
88.0   56.0
[27]   50.0  51.0  81.0  62.5  68.5  75.0  65.0  76.5  69.5  58.5  61.5
42.5   77.5
[40]   57.0  56.0  49.5  26.5  75.0  45.0  68.0  77.5  78.0  73.5  86.0
81.5   63.5
[53]   61.0  60.5  54.0  77.0  78.5 100.0  68.5  75.5  92.0  85.0  87.0
91.0   35.0
[66]   82.5  47.5  57.5  52.0  88.0  86.0  59.0  44.5  83.0  49.0  63.5
62.5   69.0
[79] 100.0   74.0  23.0  51.0  37.0  65.5  53.5  43.0  77.0  64.0  63.0
79.0   55.0
[92]   68.5  68.5  96.5
>
> score02_data_new <- data.frame(score02_data$id, score02_data_mean)
> score02_data_new
    score02_data.id score02_data_mean
1         410072106              70.0
2         410073023              61.5
3         410079062              40.0
4         410079090              65.5
5         410079118              58.0
6         410079120              56.0
7         410079121              75.0
8         410172016              68.5
9         410172027              88.5
```

10	410172103	79.0
11	410173029	26.5
12	410173072	64.0
13	410173101	73.0
14	410173134	85.0
15	410173135	77.5
16	410173136	84.0
17	410174210	56.5
18	410183004	92.5
19	410183012	51.0
20	410184012	45.5
21	410184015	48.5
22	410273002	62.5
23	410273004	77.5
24	410273005	57.5
25	410273007	88.0
26	410273010	56.0
27	410273011	50.0
28	410273014	51.0
29	410273016	81.0
30	410273018	62.5
31	410273019	68.5
32	410273020	75.0
33	410273024	65.0
34	410273031	76.5
35	410273032	69.5
36	410273034	58.5
37	410273040	61.5
38	410273041	42.5
39	410273042	77.5
40	410273048	57.0
41	410273049	56.0
42	410273050	49.5
43	410273051	26.5
44	410273057	75.0
45	410273060	45.0
46	410273062	68.0
47	410273065	77.5

48	410273067	78.0
49	410273069	73.5
50	410273070	86.0
51	410273073	81.5
52	410273075	63.5
53	410273076	61.0
54	410273081	60.5
55	410273094	54.0
56	410273095	77.0
57	410273096	78.5
58	410273102	100.0
59	410273105	68.5
60	410273106	75.5
61	410273108	92.0
62	410273109	85.0
63	410273110	87.0
64	410273116	91.0
65	410275001	35.0
66	410275005	82.5
67	410275015	47.5
68	410275016	57.5
69	410275017	52.0
70	410275020	88.0
71	410275029	86.0
72	410275032	59.0
73	410275033	44.5
74	410275034	83.0
75	410275036	49.0
76	410275040	63.5
77	410275051	62.5
78	410275055	69.0
79	410275058	100.0
80	410279001	74.0
81	410279006	23.0
82	410279018	51.0
83	410279021	37.0
84	410279039	65.5
85	410279049	53.5

86	410279054	43.0
87	410279063	77.0
88	410279075	64.0
89	410279080	63.0
90	49973086	79.0
91	49979003	55.0
92	49979046	68.5
93	49981006	68.5
94	49981011	96.5

```
>
> #install.packages("dplyr")  安裝套件 dplyr
> #library(dplyr)  導入套件 dplyr
> #install.packages("dplyr")
> library(dplyr)
>
> #arrange(資料位置, 排序名稱)  需要遞減就用 desc()將遞減資料包起來
> score02_data_newarrange <- arrange(score02_data_new,
desc(score02_data_mean))
> score02_data_newarrange
     score02_data.id score02_data_mean
```

	score02_data.id	score02_data_mean
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

\>