109-1 資料科學應用 HW3

班級:經濟四乙

學號:A106260018

姓名:張鈞茹

```
> #第 1.25 題
> #某班「R 程式設計」一科學期各項成績總表紀錄於「R-score.xlsx」。
> #(a) 讀取資料檔，印出前 5 位同學成績紀錄。
> library(xlsx)
>
> data1 <- read.xlsx("R-score.xlsx",1, startRow=2,encoding="UTF-8")
> head(data1,5)
  No        系級      學號    姓名 X0.1 X0.15 X0.15.1 X0.2 X0.4 X10 分
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
> str(data1)
'data.frame':       13 obs. of  10 variables:
 $ No        : num  1 2 3 4 5 6 7 8 9 10 ...
 $ 系級      : chr  "統計系 1 " "統計系 1 " " 會計系 1 " " 會計系 1 " ...
 $ 學號      : num  32578012 32578014 32578016 32578018 32578020 ...
 $ 姓名      : chr  "周小如" "周抒如" "林育安" "林育辰" ...
 $ X0.1      : num  55 30 10 10 5 10 25 55 10 15 ...
 $ X0.15     : num  95 65 5 20 15 35 50 45 15 5 ...
 $ X0.15.1: num  100 70 25 45 20 60 40 75 55 30 ...
 $ X0.2      : num  100 100 10 40 25 0 60 100 55 45 ...
 $ X0.4      : num  86 94 77 87 86 77 87 79 87 76 ...
 $ X10 分    : num  10 10 10 10 0 0 10 10 4 7 ...
>
> #(b) 計算各項考試 (不含點名) 平均分數及標準差。
> x0.1 <- mean(c(55,30,10,10,5,10,25,55,10,15,35,50,15))
> x0.15 <- mean(c(95,65,5,20,15,35,50,45,15,5,10,100,10 ))
> x0.15.1 <- mean(c(100,70,25,45,20,60,40,75,55,30,5,65,75))
> x0.2 <- mean(c(100,100,10,40,25,0,60,100,55,45,0,100,30))
> x0.4 <- mean(c(86,94,77,87,86,77,87,79,87,76,78,90,0))
> cat("平均分別為",x0.1,x0.15,x0.15.1,x0.2,x0.4)
```

平均分別為 25 36.15385 51.15385 51.15385 77.23077> x0.1sd <-
sd(c(55,30,10,10,5,10,25,55,10,15,35,50,15))
> x0.15sd <- sd(c(95,65,5,20,15,35,50,45,15,5,10,100,10))
> x0.15.1sd <- sd(c(100,70,25,45,20,60,40,75,55,30,5,65,75))
> x0.2sd <- sd(c(100,100,10,40,25,0,60,100,55,45,0,100,30))
> x0.4sd <- sd(c(86,94,77,87,86,77,87,79,87,76,78,90,0))
> cat("標準差分別為",x0.1sd,x0.15sd,x0.15.1sd,x0.2sd,x0.4sd)
標準差分別為 18.37117 33.05008 26.7047 38.57643 23.89963>
> #品君的
> #第一次
> test.1 <- mean(rowMeans(data1[5]))
> test.1
[1] 25
> test.1.2 <- sd(rowMeans(data1[5]))
> test.1.2
[1] 18.37117
> #第二次
> test.2 <- mean(rowMeans(data1[6]))
> test.2
[1] 36.15385
> test.2.2 <- sd(rowMeans(data1[6]))
> test.2.2
[1] 33.05008
> #第三次
> test.3 <- mean(rowMeans(data1[7]))
> test.3
[1] 51.15385
> test.3.2 <- sd(rowMeans(data1[7]))
> test.3.2
[1] 26.7047
> #作業
> hw.1 <- mean(rowMeans(data1[8]))
> hw.1
[1] 51.15385
> hw.1.2 <- sd(rowMeans(data1[8]))
> hw.1.2
[1] 38.57643
> #期中

```
> testm.1 <- mean(rowMeans(data1[9]))
> testm.1
[1] 77.23077
> testm.1.2 <- sd(rowMeans(data1[9]))
> testm.1.2
[1] 23.89963
>
> #(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)
> 學期成績 <- (data1$X0.1)*0.1+(data1$X0.15)*0.15+(data1$X0.15.1)*0.15+(data1$X0.2)*0.2+(data1$X0.4)*0.4
> data1.1 <- data.frame(data1$學號,學期成績)
> data1.1
   data1.學號 學期成績
1    32578012    89.15
2    32578014    80.85
3    32578016    38.30
4    32578018    53.55
5    32578020    45.15
6    32578022    46.05
7    32578026    62.80
8    32578028    75.10
9    32578030    57.30
10   32474226    46.15
11   32475032    36.95
12   32578002    85.75
13   32578004    20.25
>
> ##############################################
> #第 1.29 題 data
> #讀取下列檔案，列印出資料前 5 筆，及後 5 筆紀錄; 同時檢查 (印出) 資料每一變數(欄位)是否有符合 R 的類別物件，若沒有，請更改。
> #(a) R-score.xlsx
> data1 <- read.xlsx("R-score.xlsx",1, startRow=2,encoding="UTF-8")
> str(data1)
'data.frame':    13 obs. of  10 variables:
```

```
 $ No         : num    1 2 3 4 5 6 7 8 9 10 ...
 $ 系級       : chr    "統計系 1 " "統計系 1 " "  會計系 1 " "  會計系 1 " ...
 $ 學號       : num    32578012 32578014 32578016 32578018 32578020 ...
 $ 姓名       : chr    "周小如" "周抒如" "林育安" "林育辰" ...
 $ X0.1       : num    55 30 10 10 5 10 25 55 10 15 ...
 $ X0.15      : num    95 65 5 20 15 35 50 45 15 5 ...
 $ X0.15.1: num    100 70 25 45 20 60 40 75 55 30 ...
 $ X0.2       : num    100 100 10 40 25 0 60 100 55 45 ...
 $ X0.4       : num    86 94 77 87 86 77 87 79 87 76 ...
 $ X10 分     : num    10 10 10 10 0 0 10 10 4 7 ...
> head(data1,5)
   No         系級        學號      姓名 X0.1 X0.15 X0.15.1 X0.2 X0.4 X10 分
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
> tail(data1,5)
    No         系級        學號      姓名 X0.1 X0.15 X0.15.1 X0.2 X0.4 X10 分
9    9   統計系 1   32578030  黎奕璇     10       15          55     55     87        4
10 10   會計系 1   32474226  蕭偲賢     15        5          30     45     76        7
11 11   會計系 1   32475032  謝涵融     35       10           5      0     78       10
12 12   會計系 1   32578002  羅順霓     50      100          65    100     90       10
13 13   統計系 1   32578004  顧瀚薇     15       10          75     30      0       10
>
> #(b) 20140714-weather.txt
> data2 <- read.delim("data/20140714-weather.txt")
> str(data2)
'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 ...
> head(data2,5)
   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
> tail(data2,5)
     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
>
> #(c) weather_delays14.csv
> data3 <- read.csv("weather_delays14.csv")
> str(data3)
'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 ...
> head(data3,5)
   year month day dep_time arr_time carrier tailnum flight origin dest carrier_delay weather_delay
nas_delay
1 2014     1   1     1733     2024      AA  N3HPAA    199    JFK  ORD
0                 7            51
2 2014     1   1     1718     1840      B6  N324JB   1734    JFK  BTV                0
18             6
3 2014     1   1      624      946      DL  N3751B    479    JFK  ATL                0
```

```
9             45
4 2014        1     1         910        1203         DL    N910DL     1174       LGA    PBI                  0
52              0
5 2014        1     1        1850        2052         MQ    N1EAMQ     2839       LGA    STL
0                35            12
   aircraft_delay
1                  11
2                   0
3                   0
4                   0
5                   0
> tail(data3,5)
        year month day dep_time arr_time carrier tailnum flight origin dest carrier_delay
weather_delay
4655 2014       10    26        1135        1451         VX    N836VA       409       JFK    LAX
5                  11
4656 2014       10    27        1042        1416         VX    N642VA       187       EWR    SFO
12                  9
4657 2014       10    29        1507        1808         DL    N321NB      1923       LGA    MIA
0                  81
4658 2014       10    31        1500        1751         DL    N338NB      1685       LGA    MCO
0                  28
4659 2014       10    31        1323        1502         AA    N3KNAA       329       LGA    ORD
0                 113
        nas_delay aircraft_delay
4655              0                 0
4656              0                 0
4657              0                 0
4658              0                 0
4659              4                 0
>
> ##############################################
> #第 2.10 題
> #有一 50  筆成績資料如下
> score <- sample(1:100, 50, replace = TRUE)
> #判別此資料中是否有高於  95  分的同學，若有，印出「老師請同學吃飯」，若沒有印出「老師很生氣」。
> score
```

```
 [1]  73  86  55  17   3  18  69  74   7  95  74  20  12  98  68  35   6  23  68
59  30  50  31  99  93  25
[27]  48  25  28  76  92 100   8  78  18  20  85  21  47  77  11  51  71  44  24
63  62  72  56  84
> y <- numeric(length(score))
> y[score>=95] <- 1
> y[score<95] <- 0
> y
 [1] 0 0 0 0 0 0 0 0 0 1 0 0 0 1 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 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>
> ##############################################
> #第 2.21 題
> #檔案  score02.csv  記錄某班的統計學期中和期末成績。
> #(a)  讀入資料  (使其具有欄位名稱:  「學號、期中考、期末考」),  印出前  7  筆紀錄。
> score02 <- read.csv("score02.csv")
> head(score02,7)
        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
>
> #(b)  將欄位名稱依序更改為: id, mid, final。
> colnames(score02) <- c("id", "mid", "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
>
> #(c)  印出期末成績比期中成績進步的同學  id。
> improve1 <- ifelse((score02$final-score02$mid)>0,"1","O")
> score03 <- data.frame(score02,improve1)
> improve3 <- ifelse(score03$improve1==1,score03$id,"")
> improve3
  [1] ""                "410073023" ""                ""                ""                ""
"410079121" "410172016"
  [9] "410172027" ""                ""                "410173072" ""                ""                ""
"410173136"
[17] "410174210" ""                ""                ""                ""                ""                ""
""
[25] ""                ""                ""                "410273014" "410273016" ""                ""
""
[33] ""                ""                ""                ""                ""                ""
"410273042" "410273048"
[41] ""                ""                ""                ""                ""                "410273062" ""
"410273067"
[49] ""                ""                "410273073" ""                "410273076" ""                ""
""
[57] ""                ""                ""                ""                "410273108" ""                ""
"410273116"
[65] ""                ""                ""                "410275016" ""                ""
"410275029" ""
[73] ""                ""                ""                ""                "410275051" ""                ""
""
[81] ""                "410279018" ""                ""                "410279049" "410279054" "410279063"
"410279075"
```

[89] "" "" "" "" "" "49981011"

>

> #(d) 將期中及期末成績，各分成及格和不及格兩組，則會有四種狀況

> #(例如其中一種: 期中及格，但期末不及格)。印出四種狀況之人數。

> 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" "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" "都及格" "都及格"

[9] "都及格" "都及格" "都不及格" "期中不及格,期末及格"

[13] "都及格" "都及格" "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" "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" "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" "都不及格" "都不及格"

[29] "都及格" "期中及格,期末不及格" "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" "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" "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" "都不及格" "都及格"

[45] "都不及格" "都及格" "都及格" "都及格"

[49] "都及格" "都及格" "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" "期中及格,期末不及格" "都及格"

[57] "都及格" "都及格" "期中及格,期末不及格" "都及格"

[61] "都及格" "都及格" "都及格" "都及格"

[65] "期中及格,期末不及格" "都及格" "都不及格" "期中不及格,期末及格"

[69] "都不及格" "都及格" "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" "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" "都及格" "期中及格,期末不

及格"
[81] "都不及格" "都不及格" "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" "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" "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"
>
> #(e) 學期成績的計算方式為期中考和期末考的平均成績，請將資料依學期成績由高分至低份排序印出。
> 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

>