

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

學號:A107260086 姓名:張允銓

(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
> library(readxl)
> readxl_example()
[1] "clippy.xls" "clippy.xlsx" "datasets.xls"
[4] "datasets.xlsx" "deaths.xls" "deaths.xlsx"
[7] "geometry.xls" "geometry.xlsx" "type-me.xls"
[10] "type-me.xlsx"
> # ex1.25(a)
> xlsx_file <- "R-score.xlsx"
> excel_sheets(xlsx_file)
[1] "工作表 1"
> mydata <- read_excel(xlsx_file, sheet = "工作表 1", na = "NA", skip = 1)
New names:
* `0.15` -> `0.15...6`
* `0.15` -> `0.15...7`
> head(mydata, 5)
# A tibble: 5 x 10
  No 系級 學號 姓名 `0.1` `0.15...6` `0.15...7`
  <dbl> <chr> <dbl> <chr> <dbl> <dbl> <dbl>
1 1 統計系 1~3.26e7 周小如~ 55 95 100
2 2 統計系 1~3.26e7 周抒如~ 30 65 70
3 3 會計系 1~3.26e7 林育安~ 10 5 25
4 4 會計系 1~3.26e7 林育辰~ 10 20 45
5 5 會計系 1~3.26e7 黃季晴~ 5 15 20
# ... with 3 more variables: `0.2` <dbl>, `0.4` <dbl>,
# `10 分` <dbl>
> str(mydata)
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" ...
 $ 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 ...
>
> # ex1.25(b)
> list1 <- (read_excel(xlsx_file, range = "E2:E15"))
> list11 <- as.data.frame(list1)
> list2 <- (read_excel(xlsx_file, range = "F2:F15"))
> list22 <- as.data.frame(list2)
> list3 <- (read_excel(xlsx_file, range = "G2:G15"))
> list33 <- as.data.frame(list3)
> list4 <- (read_excel(xlsx_file, range = "H2:H15"))
> list44 <- as.data.frame(list4)
> list5 <- (read_excel(xlsx_file, range = "I2:I15"))
> list55 <- as.data.frame(list5)
>
> q <- sum(list11) / 13
> q
[1] 25
> (sum((list11-q)^2)/(13-1))^(1/2)
[1] 18.37117
>
> w <- sum(list22) / 13
> w
[1] 36.15385
> (sum((list22-w)^2)/(13-1))^(1/2)
[1] 33.05008
>
> e <- sum(list33) / 13
> e
[1] 51.15385
> (sum((list33-e)^2)/(13-1))^(1/2)
[1] 26.7047
>
> r <- sum(list44) / 13
> r

```

```

[1] 51.15385
> (sum((list44-r)^2)/(13-1))^(1/2)
[1] 38.57643
>
> t <- sum(list55) / 13
> t
[1] 77.23077
> (sum((list55-t)^2)/(13-1))^(1/2)
[1] 23.89963
>
> # ex1.25(c)
> A <- (list11[1:13, ]*0.1 + list22[1:13, ]*0.15 + list33[1:13, ]*0.15 + list44[1:13, ]*0.2
+ list55[1:13, ]*0.4)
> data.frame(read_excel(xlsx_file, range = "C2:C15"), "學期成績" = A)
  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
>
> # ex1.29(a)
> xlsx_file <- "R-score.xlsx"
> excel_sheets(xlsx_file)
[1] "工作表 1"
> mydata <- read_excel(xlsx_file, sheet = "工作表 1", na = "NA", skip = 1)
New names:
* `0.15` -> `0.15...6`
* `0.15` -> `0.15...7`
> g <- as.data.frame(head(mydata, 5)) # 返回前 n 行

```

```

> G <- as.data.frame(tail(mydata, 5)) # 返回後 n 行
> str(g)
'data.frame': 5 obs. of 10 variables:
 $ No : num 1 2 3 4 5
 $ 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" ...
 $ 0.1 : num 55 30 10 10 5
 $ 0.15...6: num 95 65 5 20 15
 $ 0.15...7: num 100 70 25 45 20
 $ 0.2 : num 100 100 10 40 25
 $ 0.4 : num 86 94 77 87 86
 $ 10 分 : num 10 10 10 10 0
> str(G)
'data.frame': 5 obs. of 10 variables:
 $ No : num 9 10 11 12 13
 $ 系級 : chr "統計系 1" "會計系 1" "會計系 1" "會計系 1" ...
 $ 學號 : num 32578030 32474226 32475032 32578002 32578004
 $ 姓名 : chr "黎奕璇" "蕭偲賢" "謝涵融" "羅順霓" ...
 $ 0.1 : num 10 15 35 50 15
 $ 0.15...6: num 15 5 10 100 10
 $ 0.15...7: num 55 30 5 65 75
 $ 0.2 : num 55 45 0 100 30
 $ 0.4 : num 87 76 78 90 0
 $ 10 分 : num 4 7 10 10 10
> g
  No 系級 學號 姓名 0.1 0.15...6 0.15...7 0.2
1 1 統計系 1 32578012 周小如 55 95 100 100
2 2 統計系 1 32578014 周抒如 30 65 70 100
3 3 會計系 1 32578016 林育安 10 5 25 10
4 4 會計系 1 32578018 林育辰 10 20 45 40
5 5 會計系 1 32578020 黃季晴 5 15 20 25
  0.4 10 分
1 86 10
2 94 10
3 77 10
4 87 10
5 86 0

```

```

> G
No 系級 學號 姓名 0.1 0.15...6 0.15...7 0.2
1 9 統計系 1 32578030 黎奕璇 10 15 55 55
2 10 會計系 1 32474226 蕭偲賢 15 5 30 45
3 11 會計系 1 32475032 謝涵融 35 10 5 0
4 12 會計系 1 32578002 羅順寬 50 100 65 100
5 13 統計系 1 32578004 顧瀚薇 15 10 75 30
0.4 10 分
1 87 4
2 76 7
3 78 10
4 90 10
5 0 10
>
> # ex1.29(b)
> my.data1 <- read.table("20140714-weather.txt", header = T, sep="\t")
> a <- factor(c(my.data1[,2]))
> a1 <- as.numeric(as.character(a))
>
> b <- factor(c(my.data1[,3]))
> b1 <- as.numeric(as.character(b))
>
> d <- factor(c(my.data1[,5]))
> d1 <- as.numeric(as.character(d))
>
> e <- factor(c(my.data1[,6]))
> e1 <- as.numeric(as.character(e))
>
> str(my.data1)
'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 ...
> my.data1[c(1:5, 25:29), ]
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
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
>
> # ex1.29(c)
> my.data2 <- read.csv("weather_delays14.csv")
> str(my.data2)
'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 ...
> my.data2[c(1:5, 67:71), ]
  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
67 2014 1 2 1920 2256 B6 N629JB
68 2014 1 2 2027 104 B6 N630JB

```

```

69 2014 1 2 2058 242 B6 N641JB
70 2014 1 2 1915 2250 B6 N644JB
71 2014 1 2 2334 337 B6 N649JB
  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
67 1801 JFK FLL 0 41
68 263 JFK SEA 69 31
69 803 JFK SJU 0 79
70 669 JFK SJC 0 26
71 1901 JFK FLL 0 41
  nas_delay aircraft_delay
1 51 11
2 6 0
3 45 0
4 0 0
5 12 0
67 18 163
68 77 0
69 48 7
70 0 19
71 62 63
>
> # ex2.10
> 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"
[4] "老師很生氣" "老師很生氣" "老師很生氣"
[7] "老師很生氣" "老師很生氣" "老師很生氣"
[10] "老師很生氣" "老師很生氣" "老師請同學吃飯"
[13] "老師很生氣" "老師很生氣" "老師很生氣"
[16] "老師很生氣" "老師很生氣" "老師請同學吃飯"
[19] "老師很生氣" "老師很生氣" "老師很生氣"
[22] "老師很生氣" "老師很生氣" "老師請同學吃飯"
[25] "老師很生氣" "老師很生氣" "老師很生氣"

```

```
[28] "老師很生氣" "老師很生氣" "老師很生氣"
[31] "老師很生氣" "老師很生氣" "老師請同學吃飯"
[34] "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"
[40] "老師很生氣" "老師很生氣" "老師很生氣"
[43] "老師很生氣" "老師很生氣" "老師很生氣"
[46] "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)
```

```
> my.data3 <- read.csv("score02.csv")
```

```
> head(my.data3, 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(my.data3) <- c("id", "mid", "final")
```

```
> my.data3
```

```
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)
> ifelse(my.data3[,3] > my.data3[,2], my.data3[,1], NA)
[1] NA 410073023 NA NA NA
[6] NA 410079121 410172016 410172027 NA
[11] NA 410173072 NA NA NA
[16] 410173136 410174210 NA NA NA
[21] NA NA NA NA NA
[26] NA NA 410273014 410273016 NA
[31] NA NA NA NA NA
[36] NA NA NA 410273042 410273048
[41] NA NA NA NA NA
[46] 410273062 NA 410273067 NA NA
[51] 410273073 NA 410273076 NA NA
[56] NA NA NA NA NA
[61] 410273108 NA NA 410273116 NA
[66] NA NA 410275016 NA NA
[71] 410275029 NA NA NA NA
[76] NA 410275051 NA NA NA
[81] NA 410279018 NA NA 410279049
[86] 410279054 410279063 410279075 NA NA
[91] NA NA NA 49981011
>
> # ex2.21(d)
> group.id <- ifelse(my.data3[,2] < 60 & my.data3[,3] < 60, "期中不及格，且期末不及格",
+ ifelse(my.data3[,2] < 60 & my.data3[,3] >= 60, "期中不及格，但期末及格 ",
+ ifelse(my.data3[,2] >= 60 & my.data3[,3] < 60, "期中及格，但期末不及格 ",
+ ifelse(my.data3[,2] >= 60 & my.data3[,3] >= 60, "期中及格，且期末及格", NA))))
> group.id

```

- [1] "期中及格，且期末及格"
- [2] "期中不及格，但期末及格 "
- [3] "期中不及格，且期末不及格"
- [4] "期中及格，但期末不及格 "
- [5] "期中及格，但期末不及格 "
- [6] "期中及格，但期末不及格 "
- [7] "期中及格，且期末及格"
- [8] "期中及格，且期末及格"
- [9] "期中及格，且期末及格"
- [10] "期中及格，且期末及格"
- [11] "期中不及格，且期末不及格"
- [12] "期中不及格，但期末及格 "
- [13] "期中及格，且期末及格"
- [14] "期中及格，且期末及格"
- [15] "期中及格，但期末不及格 "
- [16] "期中及格，且期末及格"
- [17] "期中不及格，但期末及格 "
- [18] "期中及格，且期末及格"
- [19] "期中及格，但期末不及格 "
- [20] "期中及格，但期末不及格 "
- [21] "期中不及格，且期末不及格"
- [22] "期中及格，但期末不及格 "
- [23] "期中及格，但期末不及格 "
- [24] "期中及格，但期末不及格 "
- [25] "期中及格，且期末及格"
- [26] "期中及格，但期末不及格 "
- [27] "期中不及格，且期末不及格"
- [28] "期中不及格，且期末不及格"
- [29] "期中及格，且期末及格"
- [30] "期中及格，但期末不及格 "
- [31] "期中及格，且期末及格"
- [32] "期中及格，但期末不及格 "
- [33] "期中及格，但期末不及格 "
- [34] "期中及格，且期末及格"
- [35] "期中及格，且期末及格"
- [36] "期中及格，但期末不及格 "
- [37] "期中及格，但期末不及格 "
- [38] "期中不及格，且期末不及格"

- [39] "期中及格，且期末及格"
- [40] "期中不及格，但期末及格 "
- [41] "期中及格，但期末不及格 "
- [42] "期中不及格，且期末不及格"
- [43] "期中不及格，且期末不及格"
- [44] "期中及格，且期末及格"
- [45] "期中不及格，且期末不及格"
- [46] "期中及格，且期末及格"
- [47] "期中及格，且期末及格"
- [48] "期中及格，且期末及格"
- [49] "期中及格，且期末及格"
- [50] "期中及格，且期末及格"
- [51] "期中及格，且期末及格"
- [52] "期中及格，但期末不及格 "
- [53] "期中不及格，但期末及格 "
- [54] "期中及格，但期末不及格 "
- [55] "期中及格，但期末不及格 "
- [56] "期中及格，且期末及格"
- [57] "期中及格，且期末及格"
- [58] "期中及格，且期末及格"
- [59] "期中及格，但期末不及格 "
- [60] "期中及格，且期末及格"
- [61] "期中及格，且期末及格"
- [62] "期中及格，且期末及格"
- [63] "期中及格，且期末及格"
- [64] "期中及格，且期末及格"
- [65] "期中及格，但期末不及格 "
- [66] "期中及格，且期末及格"
- [67] "期中不及格，且期末不及格"
- [68] "期中不及格，但期末及格 "
- [69] "期中不及格，且期末不及格"
- [70] "期中及格，且期末及格"
- [71] "期中及格，且期末及格"
- [72] "期中及格，但期末不及格 "
- [73] "期中及格，但期末不及格 "
- [74] "期中及格，且期末及格"
- [75] "期中及格，但期末不及格 "
- [76] "期中及格，但期末不及格 "

```

[77] "期中不及格，但期末及格 "
[78] "期中及格，但期末不及格 "
[79] "期中及格，且期末及格"
[80] "期中及格，但期末不及格 "
[81] "期中不及格，且期末不及格"
[82] "期中不及格，且期末不及格"
[83] "期中不及格，且期末不及格"
[84] "期中及格，但期末不及格 "
[85] "期中不及格，但期末及格 "
[86] "期中不及格，且期末不及格"
[87] "期中及格，且期末及格"
[88] "期中不及格，但期末及格 "
[89] "期中及格，但期末不及格 "
[90] "期中及格，且期末及格"
[91] "期中及格，但期末不及格 "
[92] "期中及格，但期末不及格 "
[93] "期中及格，但期末不及格 "
[94] "期中及格，且期末及格"
>
> # ex2.21(e)
> SCORE <- (my.data3[,2] + my.data3[,3]) / 2
> rev(sort(SCORE))
[1] 100.0 100.0 96.5 92.5 92.0 91.0 88.5 88.0
[9] 88.0 87.0 86.0 86.0 85.0 85.0 84.0 83.0
[17] 82.5 81.5 81.0 79.0 79.0 78.5 78.0 77.5
[25] 77.5 77.5 77.5 77.0 77.0 76.5 75.5 75.0
[33] 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
[49] 65.0 64.0 64.0 63.5 63.5 63.0 62.5 62.5
[57] 62.5 61.5 61.5 61.0 60.5 59.0 58.5 58.0
[65] 57.5 57.5 57.0 56.5 56.0 56.0 56.0 55.0
[73] 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
[89] 40.0 37.0 35.0 26.5 26.5 23.0
>

```