

2020/11/06(五), 109學年第一學期 資料科學應用 R作業(2)

學號:A106260082姓名:姜品君

# (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>

```
> # ex1.13(a)
> lm.obj <- lm(airquality$Wind ~ airquality$Temp)
> lm.anova <- anova(lm.obj)
> lm.summary <- summary(lm.obj)
> class(lm.anova)
[1] "anova"      "data.frame"
> str(lm.anova)
Classes 'anova' and 'data.frame':   2 obs. of  5 variables:
 $ Df    : int  1 151
 $ Sum Sq: num  396 1491
 $ Mean Sq: num  395.71 9.87
 $ F value: num  40.1 NA
 $ Pr(>F) : num  2.64e-09 NA
 - attr(*, "heading")= chr [1:2] "Analysis of Variance Table\n" "Response: airquality$Wind"

> # ex1.13(b)
> attributes(lm.summary)
$names
[1] "call"      "terms"      "residuals"
[4] "coefficients" "aliased"    "sigma"
[7] "df"        "r.squared"  "adj.r.squared"
[10] "fstatistic" "cov.unscaled"

$class
[1] "summary.lm"

> summary(lm.obj)$r.squared
[1] 0.2097529

#1.20
> data1 <- read.delim("data/statlog_vehicle_846x18.txt")
> #資料框維度
> nrow(data1)
[1] 846
> ncol(data1)
[1] 20
> dim(data1)
[1] 846 20
> #前後各 5 筆紀錄
```

```
> head(data1,5)
  no class compactness circularity distance radiusratio
1 1 0 96 55 103 201
2 2 0 101 56 100 215
3 3 0 93 35 66 154
4 4 0 101 48 107 222
5 5 0 87 38 85 177
  pr.axis max.length scatterratio elongatedness
1 65 9 204 32
2 69 10 208 32
3 59 6 142 46
4 68 10 208 32
5 61 8 164 40
  pr.axis.1 max.length.1 scaledvmi scaledvma
1 23 166 227 624
2 24 169 227 651
3 18 128 162 304
4 24 154 232 641
5 20 129 186 402
  scaledradius skewness skewness.1 kurtosis kurtosis.1
1 246 74 6 2 186
2 223 74 6 5 186
3 120 64 5 13 197
4 204 70 5 38 190
5 130 63 1 25 198
  hollows
1 194
2 193
3 202
4 202
5 205
```

```
> tail(data1,5)
  no class compactness circularity distance
842 842 3 87 45 66
843 843 3 95 43 76
844 844 3 90 44 72
845 845 3 89 46 84
846 846 3 85 36 66
  radiusratio pr.axis max.length scatterratio
842 139 58 8 140
843 142 57 10 151
844 157 64 8 137
845 163 66 11 159
846 123 55 5 120
  elongatedness pr.axis.1 max.length.1 scaledvmi
842 47 18 148 168
843 44 19 149 173
844 48 18 144 159
```

```

845      43      20      159      173
846      56      17      128      140
      scaledvma scaledradius skewness skewness.1 kurtosis
842      294      175      73      3      12
843      339      159      71      2      23
844      283      171      65      9      4
845      368      176      72      1      20
846      212      131      73      1      18

```

```

      kurtosis.1 hollows

```

```

842      188      196
843      187      200
844      196      203
845      186      197
846      186      190

```

```

> #儲存此資料框物件所佔用的記憶體
> print(object.size(data1), units = "Mb")
0.1 Mb

```

```

# ex1.28

```

```

> data2 <- read.delim("data/stock-data.txt")
> head(data2,5)

```

```

民國100年5家半導體公司股票月成交資訊.元.股. X X.1

```

```

1      半導體公司 年度 月份
2      台積電 100 1
3      台積電 100 2
4      台積電 100 3
5      台積電 100 4

```

```

      X.2 X.3 X.4 X.5 X.6
1 最高價 最低價 加權平均價 成交筆數 成交金額
2 78.3 69.6 74.3 263,999 100,578,274,926
3 77 69.9 72.54 235,159 74,985,055,548
4 72.2 65.7 69.74 276,434 88,459,924,495
5 73.9 68 71.37 211,611 70,177,023,098

```

```

      X.7 X.8
1 成交股數 週轉率百分比
2 1,353,616,348 5.22
3 1,033,654,452 3.98
4 1,268,289,393 4.89
5 983,177,475 3.79

```

```

> tail(data2,5)

```

```

民國100年5家半導體公司股票月成交資訊.元.股. X X.1

```

```

57      旺宏 100 8
58      旺宏 100 9
59      旺宏 100 10
60      旺宏 100 11
61      旺宏 100 12

```

```

      X.2 X.3 X.4 X.5 X.6 X.7
57 14.5 10.25 11.84 152,177 8,137,500,167 687,167,610

```

```
58 12.65 10.4 11.55 108,879 5,542,998,380 479,779,350
59 12 10.25 11.31 68,571 3,041,525,834 268,710,697
60 13.65 10.85 12.54 167,018 9,538,526,797 760,264,306
61 12.85 11.15 12.17 115,192 5,070,210,532 416,455,073
```

```
  X.8
```

```
57 20.31
```

```
58 14.18
```

```
59 7.94
```

```
60 22.47
```

```
61 12.31
```

```
>
```

```
> lapply(data2, class)#幾個變數與各類別
```

```
$民國100年5家半導體公司股票月成交資訊.元.股.
```

```
[1] "character"
```

```
$X
```

```
[1] "character"
```

```
$X.1
```

```
[1] "character"
```

```
$X.2
```

```
[1] "character"
```

```
$X.3
```

```
[1] "character"
```

```
$X.4
```

```
[1] "character"
```

```
$X.5
```

```
[1] "character"
```

```
$X.6
```

```
[1] "character"
```

```
$X.7
```

```
[1] "character"
```

```
$X.8
```

```
[1] "character"
```

```
>
```

```
>
```

```
> #將成交筆數從character變為numeric
```

```
> data2.1 <- data2$X.5
```

```
> data2.2 <- data2.1[c(2:61)]
```

```
> data2.3 <- as.numeric(gsub(",", "", data2.2))
```

```

> class(data2.3)
[1] "numeric"
>
> #將成交金額從character變為numeric
> data3.1 <- data2$X.6
> data3.2 <- data3.1[c(2:61)]
> data3.3 <- as.numeric(gsub(",", "", data3.2))
> class(data3.3)
[1] "numeric"
>
> #將成交股數從character變為numeric
> data4.1 <- data2$X.7
> data4.2 <- data4.1[c(2:61)]
> data4.3 <- as.numeric(gsub(",", "", data4.2))
> class(data4.3)
[1] "numeric"

> # ex1.33(a)
> #Dates: 0924, 1112, 1231, 1105, 0604, 0219, 0416, 0611, 0813, 1029
> #Time: 01:00, 04:00, 16:00, 23:00, 08:00, 09:00, 07:00, 17:00, 03:00, 14:00
> #Items: shirt, shirt, pants, jacket, jacket, shirt, jacket, jacket, shoes, shirt
> #Volume: 7951, 159,1958, 6848, 3762, 3678, 8696, 9045, 6208, 1425
> dates <- c("0924", "1112", "1231", "1105", "0604", "0219", "0416", "0611", "0813", "1029")
> time <- c("01:00", "04:00", "16:00", "23:00", "08:00", "09:00", "07:00", "17:00", "03:00",
"14:00")
> items0 <- c("shirt", "shirt", "pants", "jacket", "jacket", "shirt", "jacket", "jacket", "shoes",
"shirt")
> volume <- c(7951, 159,1958, 6848, 3762, 3678, 8696, 9045, 6208, 1425)
> #日期時間格式設定
> library(lubridate)
> dates1 <- as.POSIXct(dates, format="%m%d")
> year(dates1) <- 2018
> class(dates1)
[1] "POSIXct" "POSIXt"
> time1 <- as.POSIXct(time, format="%H:%M")
> class(time1)
[1] "POSIXct" "POSIXt"
> time1
[1] "2020-11-12 01:00:00 CST" "2020-11-12 04:00:00 CST"
[3] "2020-11-12 16:00:00 CST" "2020-11-12 23:00:00 CST"
[5] "2020-11-12 08:00:00 CST" "2020-11-12 09:00:00 CST"
[7] "2020-11-12 07:00:00 CST" "2020-11-12 17:00:00 CST"
[9] "2020-11-12 03:00:00 CST" "2020-11-12 14:00:00 CST"
> time2 <- format(time1, format="%H:%M")
> DateTime0 <- paste(dates1, time2)
> DateTime0
[1] "2018-09-24 01:00" "2018-11-12 04:00"
[3] "2018-12-31 16:00" "2018-11-05 23:00"

```

```

[5] "2018-06-04 08:00" "2018-02-19 09:00"
[7] "2018-04-16 07:00" "2018-06-11 17:00"
[9] "2018-08-13 03:00" "2018-10-29 14:00"
> DateTime <- ymd_hm(datetime)
> DateTime
[1] "2018-09-24 01:00:00 UTC" "2018-11-12 04:00:00 UTC"
[3] "2018-12-31 16:00:00 UTC" "2018-11-05 23:00:00 UTC"
[5] "2018-06-04 08:00:00 UTC" "2018-02-19 09:00:00 UTC"
[7] "2018-04-16 07:00:00 UTC" "2018-06-11 17:00:00 UTC"
[9] "2018-08-13 03:00:00 UTC" "2018-10-29 14:00:00 UTC"
>
> #item設定
> items <- as.factor(items0)
> #類別檢查
> class(DateTime)
[1] "POSIXct" "POSIXt"
> class(items)
[1] "factor"
> class(volume)
[1] "numeric"
>
> #儲存資料框
> mySale <- data.frame(DateTime, items, volume)
> mySale
      DateTime items volume
1 2018-09-24 01:00:00 shirt  7951
2 2018-11-12 04:00:00 shirt   159
3 2018-12-31 16:00:00 pants  1958
4 2018-11-05 23:00:00 jacket  6848
5 2018-06-04 08:00:00 jacket  3762
6 2018-02-19 09:00:00 shirt  3678
7 2018-04-16 07:00:00 jacket  8696
8 2018-06-11 17:00:00 jacket  9045
9 2018-08-13 03:00:00 shoes  6208
10 2018-10-29 14:00:00 shirt  1425

> #ex1.33(b)
> #(b) 本資料中，七月 (含) 之後的銷售品項為何？其總銷售量為多少？
> library(dplyr)
> dates2 <- format(dates1, format="%m")
> dates3 <- as.numeric(dates2)
> mySale1 <- data.frame(dates3, items, volume)
> mySale1
  dates3 items volume
1     9 shirt  7951
2    11 shirt   159
3    12 pants  1958
4    11 jacket  6848

```

```

5    6 jacket  3762
6    2 shirt  3678
7    4 jacket  8696
8    6 jacket  9045
9    8 shoes  6208
10   10 shirt  1425
> july <- filter(mySale1, dates3 > 07 )
> july
  dates3 items volume
1     9 shirt  7951
2    11 shirt   159
3    12 pants  1958
4    11 jacket 6848
5     8 shoes  6208
6    10 shirt  1425
> levels(july$items)
[1] "jacket" "pants" "shirt" "shoes"
> summarise(july, sum(volume))
  sum(volume)
1    24549

```