2020/10/6(五), 109 學年第一學期 資料科學應用 R 作業(2)

學號:A106260020　　姓名:楊鎵綺

```
> #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)
> str(lm.summary)
List of 11
 $ call          : language lm(formula = airquality$Wind ~ airquality$Temp)
 $ terms         :Classes 'terms', 'formula'   language airquality$Wind ~ airquality$Temp
  .. ..- attr(*, "variables")= language list(airquality$Wind, airquality$Temp)
  .. ..- attr(*, "factors")= int [1:2, 1] 0 1
  .. .. ..- attr(*, "dimnames")=List of 2
  .. .. .. ..$ : chr [1:2] "airquality$Wind" "airquality$Temp"
  .. .. .. ..$ : chr "airquality$Temp"
  .. ..- attr(*, "term.labels")= chr "airquality$Temp"
  .. ..- attr(*, "order")= int 1
  .. ..- attr(*, "intercept")= int 1
  .. ..- attr(*, "response")= int 1
  .. ..- attr(*, ".Environment")=<environment: R_GlobalEnv>
  .. ..- attr(*, "predvars")= language list(airquality$Wind, airquality$Temp)
  .. ..- attr(*, "dataClasses")= Named chr [1:2] "numeric" "numeric"
  .. .. ..- attr(*, "names")= chr [1:2] "airquality$Wind" "airquality$Temp"
 $ residuals     : Named num [1:153] -4.413 -2.96 1.981 -1.165 0.612 ...
  ..- attr(*, "names")= chr [1:153] "1" "2" "3" "4" ...
 $ coefficients : num [1:2, 1:4] 23.2337 -0.1705 2.1124 0.0269 10.9987 ...
```

```
  ..- attr(*, "dimnames")=List of 2
  .. ..$ : chr [1:2] "(Intercept)" "airquality$Temp"
  .. ..$ : chr [1:4] "Estimate" "Std. Error" "t value" "Pr(>|t|)"
 $ aliased           : Named logi [1:2] FALSE FALSE
  ..- attr(*, "names")= chr [1:2] "(Intercept)" "airquality$Temp"
 $ sigma             : num 3.14
 $ df                : int [1:3] 2 151 2
 $ r.squared         : num 0.21
 $ adj.r.squared: num 0.205
 $ fstatistic        : Named num [1:3] 40.1 1 151
  ..- attr(*, "names")= chr [1:3] "value" "numdf" "dendf"
 $ cov.unscaled : num [1:2, 1:2] 4.52e-01 -5.72e-03 -5.72e-03 7.34e-05
  ..- attr(*, "dimnames")=List of 2
  .. ..$ : chr [1:2] "(Intercept)" "airquality$Temp"
  .. ..$ : chr [1:2] "(Intercept)" "airquality$Temp"
 - attr(*, "class")= chr "summary.lm"

> lm.summary$r.squared
[1] 0.2097529

> #ex1.20
> vehicle <- read.delim("data/statlog_vehicle_846x18.txt")
> dim(vehicle)
[1] 846  20
> head(vehicle)
  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
6  6     0          95          48      104         214
  pr.axis max.length scatterratio elongatedness pr.axis.1
1      65          9          204           32         23
2      69         10          208           32         24
3      59          6          142           46         18
4      68         10          208           32         24
5      61          8          164           40         20
6      67          9          205           32         23
  max.length.1 scaledvmi scaledvma scaledradius skewness
```

```
1             166         227         624             246         74
2             169         227         651             223         74
3             128         162         304             120         64
4             154         232         641             204         70
5             129         186         402             130         63
6             151         227         628             202         74
   skewness.1 kurtosis kurtosis.1 hollows
1             6           2           186         194
2             6           5           186         193
3             5           13          197         202
4             5           38          190         202
5             1           25          198         205
6             5           9           186         193
> tail(vehicle)
       no class compactness circularity distance radiusratio
841 841      3          95          49          82          139
842 842      3          87          45          66          139
843 843      3          95          43          76          142
844 844      3          90          44          72          157
845 845      3          89          46          84          163
846 846      3          85          36          66          123
      pr.axis max.length scatterratio elongatedness pr.axis.1
841       56          11          159             43          20
842       58           8          140             47          18
843       57          10          151             44          19
844       64           8          137             48          18
845       66          11          159             43          20
846       55           5          120             56          17
      max.length.1 scaledvmi scaledvma scaledradius skewness
841           162         173         365             185         75
842           148         168         294             175         73
843           149         173         339             159         71
844           144         159         283             171         65
845           159         173         368             176         72
846           128         140         212             131         73
      skewness.1 kurtosis kurtosis.1 hollows
841           7           10          182         191
842           3           12          188         196
843           2           23          187         200
844           9            4          196         203
```

```
845              1           20            186        197
846              1           18            186        190
> print(object.size(vehicle), units = "Mb")
0.1 Mb

> #ex1.28
> stock <- read.table("data/stock-data.txt", header=TRUE, skip=1,
+ fileEncoding="big5")
>
> head(stock)
   半導體公司 年度 月份 最高價 最低價 加權平均價 成交筆數
1        台積電  100     1   78.3   69.6       74.30  263,999
2        台積電  100     2   77.0   69.9       72.54  235,159
3        台積電  100     3   72.2   65.7       69.74  276,434
4        台積電  100     4   73.9   68.0       71.37  211,611
5        台積電  100     5   76.9   73.0       74.96  213,185
6        台積電  100     6   78.2   70.4       74.70  260,507
              成交金額       成交股數 週轉率百分比
1 100,578,274,926 1,353,616,348          5.22
2  74,985,055,548 1,033,654,452          3.98
3  88,459,924,495 1,268,289,393          4.89
4  70,177,023,098   983,177,475          3.79
5  74,005,599,560   987,256,484          3.80
6  96,761,306,205 1,295,262,736          4.99
> tail(stock)
    半導體公司 年度 月份 最高價 最低價 加權平均價 成交筆數
55         旺宏  100     7   18.50   14.40       17.09  125,851
56         旺宏  100     8   14.50   10.25       11.84  152,177
57         旺宏  100     9   12.65   10.40       11.55  108,879
58         旺宏  100    10   12.00   10.25       11.31   68,571
59         旺宏  100    11   13.65   10.85       12.54  167,018
60         旺宏  100    12   12.85   11.15       12.17  115,192
            成交金額     成交股數 週轉率百分比
55 8,571,233,298 501,422,845         14.82
56 8,137,500,167 687,167,610         20.31
57 5,542,998,380 479,779,350         14.18
58 3,041,525,834 268,710,697          7.94
59 9,538,526,797 760,264,306         22.47
60 5,070,210,532 416,455,073         12.31
> stock$成交筆數  <- as.numeric(stock$成交筆數)
```

```
Warning message:
強制變更過程中產生了 NA
> stock$成交金額 <- as.numeric(stock$成交金額)
Warning message:
強制變更過程中產生了 NA
> stock$成交股數 <- as.numeric(stock$成交股數)
Warning message:
強制變更過程中產生了 NA
> lapply(stock, class)
$半導體公司
[1] "character"

$年度
[1] "integer"

$月份
[1] "integer"

$最高價
[1] "numeric"

$最低價
[1] "numeric"

$加權平均價
[1] "numeric"

$成交筆數
[1] "numeric"

$成交金額
[1] "numeric"

$成交股數
[1] "numeric"

$週轉率百分比
[1] "numeric"

> #ex1.33(a)
```

```
> mysale <- scan(what=list(Dates="POSIXct", Time="UTC", Items="factor",
+ Volume="numeric"))
1: 0924 01:00 shirt 7951
2: 1112 04:00 shirt 159
3: 1231 16:00 pants 1968
4: 1105 23:00 jacket 6848
5: 0604 08:00 jacket 3762
6: 0219 09:00 shirt 3678
7: 0416 07:00 jacket 8696
8: 0611 17:00 jacket 9045
9: 0813 03:00 shoes 6208
10: 1029 14:00 shirt 1425
11:
Read 10 records
> mySale <- data.frame(mysale)
> mySale$Items <- as.factor(mySale$Items)
> mySale$Volume <- as.numeric(mySale$Volume)
> lapply(mySale, class)
$Dates
[1] "character"

$Time
[1] "character"

$Items
[1] "factor"

$Volume
[1] "numeric"

> mySale
      Dates    Time    Items Volume
1     0924 01:00    shirt      7951
2     1112 04:00    shirt       159
3     1231 16:00    pants      1968
4     1105 23:00 jacket      6848
5     0604 08:00 jacket      3762
6     0219 09:00    shirt      3678
7     0416 07:00 jacket      8696
8     0611 17:00 jacket      9045
```

```
9     0813 03:00    shoes      6208
10    1029 14:00    shirt      1425
>
```