```
> #2020/11/06(五), 109 學年第一學期 資料科學應用 R 作業(2)
> #學號:A107260102      姓名:熊家濬
> # 2020/11/03
> ## 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"         "coefficients"
 [5] "aliased"             "sigma"               "df"                  "r.squared"
 [9] "adj.r.squared" "fstatistic"       "cov.unscaled"

$class
[1] "summary.lm"

> attr(lm.summary, "names")
 [1] "call"                "terms"               "residuals"         "coefficients"
 [5] "aliased"             "sigma"               "df"                  "r.squared"
 [9] "adj.r.squared" "fstatistic"       "cov.unscaled"
> names(lm.summary)
 [1] "call"                "terms"               "residuals"         "coefficients"
 [5] "aliased"             "sigma"               "df"                  "r.squared"
 [9] "adj.r.squared" "fstatistic"       "cov.unscaled"
> R <- lm.summary["r.squared"]
> class(R)
```

```
[1] "list"
> R2 <- as.numeric(R)
> class(R2)
[1] "numeric"
> R2^2
[1] 0.04399628
>
> # ex1.20
> my.data <- read.table("statlog_vehicle_846x18.txt", row.names=1,header = T ,
sep="\t")
> str(my.data)
'data.frame':    846 obs. of    19 variables:
 $ class              : int    0 0 0 0 0 0 0 0 0 0 ...
 $ compactness    : int    96 101 93 101 87 95 98 107 103 77 ...
 $ circularity    : int    55 56 35 48 38 48 55 53 50 38 ...
 $ distance         : int    103 100 66 107 85 104 101 103 98 63 ...
 $ radiusratio    : int    201 215 154 222 177 214 228 221 212 135 ...
 $ pr.axis           : int    65 69 59 68 61 67 70 66 63 59 ...
 $ max.length     : int    9 10 6 10 8 9 9 11 9 5 ...
 $ scatterratio : int    204 208 142 208 164 205 210 209 193 130 ...
 $ elongatedness: int    32 32 46 32 40 32 31 32 34 52 ...
 $ pr.axis.1       : int    23 24 18 24 20 23 24 24 22 18 ...
 $ max.length.1 : int    166 169 128 154 129 151 168 163 161 130 ...
 $ scaledvmi       : int    227 227 162 232 186 227 236 222 214 145 ...
 $ scaledvma       : int    624 651 304 641 402 628 661 653 567 247 ...
 $ scaledradius : int    246 223 120 204 130 202 245 212 185 139 ...
 $ skewness         : int    74 74 64 70 63 74 72 66 64 79 ...
 $ skewness.1     : int    6 6 5 5 1 5 1 0 5 13 ...
 $ kurtosis         : int    2 5 13 38 25 9 6 1 5 21 ...
 $ kurtosis.1     : int    186 186 197 190 198 186 188 191 198 183 ...
 $ hollows          : int    194 193 202 202 205 193 197 201 204 187 ...
> dim(my.data)
[1] 846    19
> my.data[c(1:6, 843:847), ]
      class compactness circularity distance radiusratio pr.axis max.length
1         0          96          55      103         201      65
9
2         0         101          56      100         215      69
```

10
3 0 93 35 66 154 59
6
4 0 101 48 107 222 68
10
5 0 87 38 85 177 61
8
6 0 95 48 104 214 67
9
843 3 95 43 76 142 57
10
844 3 90 44 72 157 64
8
845 3 89 46 84 163 66
11
846 3 85 36 66 123 55
5
NA NA NA NA NA NA NA
NA

	scatterratio	elongatedness	pr.axis.1	max.length.1	scaledvmi	scaledvma
1	204	32	23	166	227	624
2	208	32	24	169	227	651
3	142	46	18	128	162	304
4	208	32	24	154	232	641
5	164	40	20	129	186	402
6	205	32	23	151	227	628
843	151	44	19	149	173	339
844	137	48	18	144	159	283
845	159	43	20	159	173	368

```
846          120          56          17          128          140
212
NA           NA           NA          NA          NA           NA
NA
     scaledradius skewness skewness.1 kurtosis kurtosis.1 hollows
1            246       74          6          2        186      194
2            223       74          6          5        186      193
3            120       64          5         13        197      202
4            204       70          5         38        190      202
5            130       63          1         25        198      205
6            202       74          5          9        186      193
843          159       71          2         23        187      200
844          171       65          9          4        196      203
845          176       72          1         20        186      197
846          131       73          1         18        186      190
NA           NA        NA          NA        NA        NA       NA
> n <- nrow(my.data)
> p <- ncol(my.data)
> myData <- matrix(rnorm(n*p), ncol = p, nrow=n)
> print(object.size(myData), units = "Mb")
0.1 Mb
>
> ## ex1.28
> my.data2 <- read.table("stock-data.txt", header = TRUE, skip = 2, sep="\t")
> my.data2[c(1:5, 56:60), ]
    半導體公司 年度 月份 最高價 最低價 加權平均價 成交筆數          成交
金額
1       台積電  100    1  78.30  69.60     74.30  263,999
100,578,274,926
2       台積電  100    2  77.00  69.90     72.54  235,159
74,985,055,548
3       台積電  100    3  72.20  65.70     69.74  276,434
88,459,924,495
4       台積電  100    4  73.90  68.00     71.37  211,611
70,177,023,098
5       台積電  100    5  76.90  73.00     74.96  213,185
74,005,599,560
56        旺宏  100    8  14.50  10.25     11.84  152,177
```

```
8,137,500,167
57          旺宏   100      9   12.65   10.40         11.55   108,879
5,542,998,380
58          旺宏   100     10   12.00   10.25         11.31    68,571
3,041,525,834
59          旺宏   100     11   13.65   10.85         12.54   167,018
9,538,526,797
60          旺宏   100     12   12.85   11.15         12.17   115,192
5,070,210,532
          成交股數  週轉率百分比
1   1,353,616,348              5.22
2   1,033,654,452              3.98
3   1,268,289,393              4.89
4     983,177,475              3.79
5     987,256,484              3.80
56    687,167,610             20.31
57    479,779,350             14.18
58    268,710,697              7.94
59    760,264,306             22.47
60    416,455,073             12.31
>
> attributes(my.data2)
$names
 [1] "半導體公司"     "年度"             "月份"             "最高價"           "最低價
"
 [6] "加權平均價"     "成交筆數"         "成交金額"         "成交股數"         "週轉率
百分比"

$class
[1] "data.frame"

$row.names
 [1]  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
25 26
[27] 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52
[53] 53 54 55 56 57 58 59 60

>
```

```
> str(my.data2)
'data.frame':    60 obs. of    10 variables:
 $ 半導體公司    : chr    "台積電" "台積電" "台積電" "台積電" ...
 $ 年度          : int    100 100 100 100 100 100 100 100 100 100 ...
 $ 月份          : int    1 2 3 4 5 6 7 8 9 10 ...
 $ 最高價        : num    78.3 77 72.2 73.9 76.9 78.2 73.9 72.8 72.1 74 ...
 $ 最低價        : num    69.6 69.9 65.7 68 73 70.4 68.5 62.2 65.9 68.1 ...
 $ 加權平均價    : num    74.3 72.5 69.7 71.4 75 ...
 $ 成交筆數      : chr    "263,999" "235,159" "276,434" "211,611" ...
 $ 成交金額      : chr    "100,578,274,926" "74,985,055,548" "88,459,924,495"
"70,177,023,098" ...
 $ 成交股數      : chr    "1,353,616,348" "1,033,654,452" "1,268,289,393"
"983,177,475" ...
 $ 週轉率百分比: num    5.22 3.98 4.89 3.79 3.8 4.99 3.96 4.9 4.14 3.27 ...
> n <- factor(c(my.data2[,7]))
> n_clean = gsub('[,]', '', n)
> n1 <- as.numeric(as.character(n_clean ))
> class(n1)
[1] "numeric"
> m <- factor(c(my.data2[,8]))
> m_clean = gsub('[,]', '', m)
> m1 <- as.numeric(as.character(m_clean ))
> class(m1)
[1] "numeric"
> s <- factor(c(my.data2[,9]))
> s_clean = gsub('[,]', '', s)
> s1 <- as.numeric(as.character(s_clean ))
> class(s1)
[1] "numeric"
>    ## ex1.33(a)
> 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")
> Items1 <-c ( "shirt", "shirt", "pants", "jacket", "jacket", "shirt", "jacket", "jacket",
"shoes", "shirt")
> Volume1 <-c ("7951", "159","1958", "6848", "3762", "3678", "8696", "9045",
"6208", "1425")
```

```
> DateTime1 <- paste("2018", Dates, Time)
> DateTime <- strptime(DateTime1, format="%Y %m%d %H:%M",    tz = "UTC")
> Items <- as.factor(Items1)
> Volume <- as.numeric(Volume1)
> 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)
> id <- 1:length(Dates)
> Q <- id [Dates >=    "0701"]
> mysale[Q, ]
                    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
9    2018-08-13 03:00:00    shoes      6208
10 2018-10-29 14:00:00    shirt      1425
>
```