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

> ##學號:A106260073 姓名:張雅雯

> ##1.13(a)

> lm.obj <- lm(airquality$Wind ~ airquality$Temp)

> lm.anova <- anova(lm.obj)

> lm.summary <- summary(lm.obj)

> lapply(lm.anova,class )

$Df

[1] "integer"

$`Sum Sq`

[1] "numeric"

$`Mean Sq`

[1] "numeric"

$`F value`

[1] "numeric"

$`Pr(>F)`

[1] "numeric"

> ##1.13(b)

> lm.summary$r.squared

[1] 0.2097529

> ##1.20

> getwd()

[1] "C:/Users/張啟 生/Desktop"

> dir("data")

[1] "20140714-weather.txt"

[2] "A106260073-張雅雯-SHU-R-exam1-run.txt"

[3] "Calculus-score-A.csv"

[4] "Calculus-score-B.xls"

[5] "elections-2000.csv"

[6] "iris-data0.txt"

[7] "iris-data1.txt"

[8] "iris-data2.txt"

[9] "iris-data3.txt"

[10] "R-score.xlsx"

[11] "sale.txt"

[12] "score02.csv"

[13] "statlog\_vehicle\_846x18.txt"

[14] "stock-data.txt"

[15] "test2.txt"

[16] "weather\_delays14.csv"

[17] "題目.png"

> statlog\_vehicle <- read.delim("data/statlog\_vehicle\_846x18.txt")

> head(statlog\_vehicle, n=5)

no class compactness circularity

1 1 0 96 55

2 2 0 101 56

3 3 0 93 35

4 4 0 101 48

5 5 0 87 38

distance radiusratio pr.axis max.length

1 103 201 65 9

2 100 215 69 10

3 66 154 59 6

4 107 222 68 10

5 85 177 61 8

scatterratio elongatedness pr.axis.1

1 204 32 23

2 208 32 24

3 142 46 18

4 208 32 24

5 164 40 20

max.length.1 scaledvmi scaledvma

1 166 227 624

2 169 227 651

3 128 162 304

4 154 232 641

5 129 186 402

scaledradius skewness skewness.1

1 246 74 6

2 223 74 6

3 120 64 5

4 204 70 5

5 130 63 1

kurtosis kurtosis.1 hollows

1 2 186 194

2 5 186 193

3 13 197 202

4 38 190 202

5 25 198 205

> tail(statlog\_vehicle, n=5)

no class compactness circularity

842 842 3 87 45

843 843 3 95 43

844 844 3 90 44

845 845 3 89 46

846 846 3 85 36

distance radiusratio pr.axis

842 66 139 58

843 76 142 57

844 72 157 64

845 84 163 66

846 66 123 55

max.length scatterratio elongatedness

842 8 140 47

843 10 151 44

844 8 137 48

845 11 159 43

846 5 120 56

pr.axis.1 max.length.1 scaledvmi

842 18 148 168

843 19 149 173

844 18 144 159

845 20 159 173

846 17 128 140

scaledvma scaledradius skewness

842 294 175 73

843 339 159 71

844 283 171 65

845 368 176 72

846 212 131 73

skewness.1 kurtosis kurtosis.1 hollows

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

> dim(statlog\_vehicle)

[1] 846 20

> print(object.size(statlog\_vehicle),units='auto')

69.2 Kb

> ##1.28

> getwd()

[1] "C:/Users/張啟 生/Desktop"

> dir("data")

[1] "20140714-weather.txt"

[2] "A106260073-張雅雯-SHU-R-exam1-run.txt"

[3] "Calculus-score-A.csv"

[4] "Calculus-score-B.xls"

[5] "elections-2000.csv"

[6] "iris-data0.txt"

[7] "iris-data1.txt"

[8] "iris-data2.txt"

[9] "iris-data3.txt"

[10] "R-score.xlsx"

[11] "sale.txt"

[12] "score02.csv"

[13] "statlog\_vehicle\_846x18.txt"

[14] "stock-data.txt"

[15] "test2.txt"

[16] "weather\_delays14.csv"

[17] "題目.png"

> stock <- read.delim("data/stock-data.txt")

> head(stock, n=6,skep=1)

民國100年5家半導體公司股票月成交資訊.元.股.

1 半導體公司

2 台積電

3 台積電

4 台積電

5 台積電

6 台積電

X X.1 X.2 X.3 X.4

1 年度 月份 最高價 最低價 加權平均價

2 100 1 78.3 69.6 74.3

3 100 2 77 69.9 72.54

4 100 3 72.2 65.7 69.74

5 100 4 73.9 68 71.37

6 100 5 76.9 73 74.96

X.5 X.6 X.7

1 成交筆數 成交金額 成交股數

2 263,999 100,578,274,926 1,353,616,348

3 235,159 74,985,055,548 1,033,654,452

4 276,434 88,459,924,495 1,268,289,393

5 211,611 70,177,023,098 983,177,475

6 213,185 74,005,599,560 987,256,484

X.8

1 週轉率百分比

2 5.22

3 3.98

4 4.89

5 3.79

6 3.8

> tail(stock)

民國100年5家半導體公司股票月成交資訊.元.股.

56 旺宏

57 旺宏

58 旺宏

59 旺宏

60 旺宏

61 旺宏

X X.1 X.2 X.3 X.4 X.5

56 100 7 18.5 14.4 17.09 125,851

57 100 8 14.5 10.25 11.84 152,177

58 100 9 12.65 10.4 11.55 108,879

59 100 10 12 10.25 11.31 68,571

60 100 11 13.65 10.85 12.54 167,018

61 100 12 12.85 11.15 12.17 115,192

X.6 X.7 X.8

56 8,571,233,298 501,422,845 14.82

57 8,137,500,167 687,167,610 20.31

58 5,542,998,380 479,779,350 14.18

59 3,041,525,834 268,710,697 7.94

60 9,538,526,797 760,264,306 22.47

61 5,070,210,532 416,455,073 12.31

> lapply(stock,class)

$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"

>

> stock$成交筆數<- as.numeric(gsub(",","", stock$成交筆數))

Error in `$<-.data.frame`(`\*tmp\*`, 成交筆數, value = numeric(0)) :

replacement has 0 rows, data has 61

>

> ##1.33(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")

> Items <- c('shirt', 'shirt', 'pants', 'jacket', 'jacket', 'shirt', 'jacket', 'jacket', 'shoes', 'shirt')

> Volume <- c(7951, 159,1958, 6848, 3762, 3678, 8696, 9045, 6208, 1425)

> DateTime <-strptime(paste("2018",Dates, Time),format= "%Y %m%d %H:%M" , tz="UTC")

> DateTime

[1] "2018-09-24 01:00:00 UTC"

[2] "2018-11-12 04:00:00 UTC"

[3] "2018-12-31 16:00:00 UTC"

[4] "2018-11-05 23:00:00 UTC"

[5] "2018-06-04 08:00:00 UTC"

[6] "2018-02-19 09:00:00 UTC"

[7] "2018-04-16 07:00:00 UTC"

[8] "2018-06-11 17:00:00 UTC"

[9] "2018-08-13 03:00:00 UTC"

[10] "2018-10-29 14:00:00 UTC"

> 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

> str(mySale)

'data.frame': 10 obs. of 3 variables:

$ DateTime: POSIXct, format: ...

$ Items : chr "shirt" "shirt" "pants" "jacket" ...

$ Volume : num 7951 159 1958 6848 3762 ...

> ##1.33(b)

> Item\_July <- mySale$Items[DateTime >= "0416 07:00"]

Error in as.POSIXlt.character(x, tz, ...) :

character string is not in a standard unambiguous format

> Item\_July

錯誤: 找不到物件 'Item\_July'

> Volume\_July <- sum(mySale$Volume[DateTime >= "0416 07:00"])

Error in as.POSIXlt.character(x, tz, ...) :

character string is not in a standard unambiguous format

> Volume\_July

錯誤: 找不到物件 'Volume\_July'

>