

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

學號: A106260052 姓名: 劉映妤

(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>

```
> # 2020/11/06
>
> # 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"
 [3] "residuals" "coefficients"
 [5] "aliased"    "sigma"
 [7] "df"         "r.squared"
 [9] "adj.r.squared" "fstatistic"
[11] "cov.unscaled"

$class
```

```
[1] "summary.lm"
```

```
> summary(lm.obj)$r.squared
```

```
[1] 0.2097529
```

```
>
```

```
> # ex1.20
```

```
> mydata1 <- read.delim("data/statlog_vehicle_846x18.txt")
```

```
> #資料框維度
```

```
> nrow(mydata1)
```

```
[1] 846
```

```
> ncol(mydata1)
```

```
[1] 20
```

```
> dim(mydata1)
```

```
[1] 846 20
```

```
> #前後各 5 筆紀錄
```

```
> head(mydata1,5)
```

```
no class compactness circularity distance
```

1	1	0	96	55	103
2	2	0	101	56	100
3	3	0	93	35	66
4	4	0	101	48	107
5	5	0	87	38	85

```
radiusratio pr.axis max.length scatterratio
```

1	201	65	9	204
2	215	69	10	208
3	154	59	6	142
4	222	68	10	208
5	177	61	8	164

```
elongatedness pr.axis.1 max.length.1
```

1	32	23	166
2	32	24	169
3	46	18	128
4	32	24	154
5	40	20	129

```
scaledvmi scaledvma scaledradius skewness
```

1	227	624	246	74
2	227	651	223	74
3	162	304	120	64

4	232	641	204	70
5	186	402	130	63

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

> tail(mydata1,5)

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

842	47	18	148
843	44	19	149
844	48	18	144
845	43	20	159
846	56	17	128

scaledvmi scaledvma scaledradius skewness

842	168	294	175	73
843	173	339	159	71
844	159	283	171	65
845	173	368	176	72
846	140	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

> #儲存此資料框物件所佔用的記憶體

> print(object.size(mydata1), units = "Mb")

0.1 Mb

>

> # ex1.28

> mydata2 <- read.delim("data/stock-data.txt")

> #前後各 5 筆紀錄

> head(mydata2,5)

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

	X	X.1	X.2	X.3	X.4	X.5
1	年度	月份	最高價	最低價	加權平均價	成交筆數
2	100	1	78.3	69.6	74.3	263,999
3	100	2	77	69.9	72.54	235,159
4	100	3	72.2	65.7	69.74	276,434
5	100	4	73.9	68	71.37	211,611

		X.6	X.7	X.8
1		成交金額	成交股數	週轉率百分比
2	100,578,274,926	1,353,616,348		5.22
3	74,985,055,548	1,033,654,452		3.98
4	88,459,924,495	1,268,289,393		4.89
5	70,177,023,098	983,177,475		3.79

> tail(mydata2,5)

民國 100 年 5 家半導體公司股票月成交資訊.元.股.

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
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

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(mydata2, 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"
```

```
>
```

```
> #將成交筆數從 character 轉為 numeric
```

```
> mydata2.1 <- mydata2$X.5
```

```
> mydata2.2 <- mydata2.1[c[2:61]]
```

```
> mydata2.3 <- as.numeric(gsub(",", "", mydata2.2))
```

```
> class(mydata2.3)
```

```
[1] "numeric"
```

```
>
```

```
> #將成交金額從 character 轉為 numeric
```

```
> mydata2.4 <- mydata2$X.6
```

```
> mydata2.5 <- mydata2.4[c[2:61]]
```

```
> mydata2.6 <- as.numeric(gsub(",", "", mydata2.5))
```

```
> class(mydata2.6)
```

```
[1] "numeric"
```

```
>
```

```
> #將成交股數從 character 轉為 numeric
```

```
> mydata2.7 <- mydata2$X.7
```

```
> mydata2.8 <- mydata2.7[c[2:61]]
```

```
> mydata2.9 <- as.numeric(gsub(",", "", mydata2.8))
```

```
> class(mydata2.9)
```

```
[1] "numeric"
```

```
>
```

```
> # ex1.33(a)
```

```
> Dates <- c("2018/09/24", "2018/11/12", "2018/12/31", "2018/11/05",  
"2018/06/04", "2018/02/19", "2018/04/16", "2018/06/11", "2018/08/13",  
"2018/10/29")
```

```
> 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 <- paste(Dates, Time)
```

```
> library(lubridate)
```

```
> class(ymd_hm(DateTime))
```

```
[1] "POSIXct" "POSIXt"
```

```
>
> class(factor(Items))
[1] "factor"
>
> class(Volume)
[1] "numeric"
>
> mySale <- data.frame(DateTime, Items, Volume)
> mySale
```

	DateTime	Items	Volume
1	2018/09/24 01:00	shirt	7951
2	2018/11/12 04:00	shirt	159
3	2018/12/31 16:00	pants	1958
4	2018/11/05 23:00	jacket	6848
5	2018/06/04 08:00	jacket	3762
6	2018/02/19 09:00	shirt	3678
7	2018/04/16 07:00	jacket	8696
8	2018/06/11 17:00	jacket	9045
9	2018/08/13 03:00	shoes	6208
10	2018/10/29 14:00	shirt	1425

```
>
```