

#2020/10/23(五) 109 學年第一學期 資料科學應用 R 作業(1)

> #學號: A107260088

姓名: 施珮慈

> #1.13(a)

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

> class(lm.obj)

[1] "lm"

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

> class(lm.anova)

[1] "anova" "data.frame"

> #1.13(b)

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

> attributes(lm.summary)

\$names

[1] "call" "terms" "residuals" "coefficients" "aliased"

[6] "sigma" "df" "r.squared" "adj.r.squared"

"fstatistic"

[11] "cov.unscaled"

\$class

[1] "summary.lm"

> lm.summary\$r.squared

[1] 0.2097529

> #1.20

> a<-read.delim("data/statlog\_vehicle\_846x18.txt",header=TRUE,sep="\t")

> dim(a)

[1] 846 20

> head(a,5)

	no	class	compactness	circularity	distance	radiusratio	pr.axis	max.length
1	1	0	96	55	103	201	65	
9								
2	2	0	101	56	100	215	69	
10								
3	3	0	93	35	66	154	59	
6								
4	4	0	101	48	107	222	68	
10								
5	5	0	87	38	85	177	61	

8

	scatterratio	elongatedness	pr.axis.1	max.length.1	scaledvmi	scaledvma
1	204	32	23	166	227	
624	246					
2	208	32	24	169	227	
651	223					
3	142	46	18	128	162	
304	120					
4	208	32	24	154	232	
641	204					
5	164	40	20	129	186	
402	130					

	skewness	skewness.1	kurtosis	kurtosis.1	hollows
1	74	6	2	186	194
2	74	6	5	186	193
3	64	5	13	197	202
4	70	5	38	190	202
5	63	1	25	198	205

> tail(a,5)

	no class	compactness	circularity	distance	radiusratio	pr.axis	max.length
842	842	3	87	45	66	139	58
8							
843	843	3	95	43	76	142	57
10							
844	844	3	90	44	72	157	64
8							
845	845	3	89	46	84	163	66
11							
846	846	3	85	36	66	123	55
5							

	scatterratio	elongatedness	pr.axis.1	max.length.1	scaledvmi	scaledvma
842	140	47	18	148	168	
294						
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

```

scaledradius skewness skewness.1 kurtosis kurtosis.1 hollows

```

842          175          73          3          12          188          196
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

```

```
> print(object.size(a), units = "Kb")
```

```
69.2 Kb
```

```
> #1.28
```

```
> b<-read.delim("data/stock-data.txt",header=TRUE,skip=1)
```

```
> head(b,5)
```

	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	78.3	69.6	74.30	263,999	100,578,274,926
2	台積電	100	2	77.0	69.9	72.54	235,159	74,985,055,548
3	台積電	100	3	72.2	65.7	69.74	276,434	88,459,924,495
4	台積電	100	4	73.9	68.0	71.37	211,611	70,177,023,098
5	台積電	100	5	76.9	73.0	74.96	213,185	74,005,599,560

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

```
> tail(b,5)
```

	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	成交股數
56	旺宏	100	8	14.50	10.25	11.84	152,177	8,137,500,167	687,167,610
57	旺宏	100	9	12.65	10.40	11.55	108,879	5,542,998,380	479,779,350
58	旺宏	100	10	12.00	10.25	11.31	68,571	3,041,525,834	268,710,697
59	旺宏	100	11	13.65	10.85	12.54	167,018	9,538,526,797	

760,264,306

60 旺宏 100 12 12.85 11.15 12.17 115,192 5,070,210,532

416,455,073

週轉率百分比

56 20.31

57 14.18

58 7.94

59 22.47

60 12.31

> #1.33

```
> Dates <- c("180924", "181112", "181231", "181105", "180604", "180219",  
"180416", "180611", "180813", "181029")
```

```
> Time <- c("01:00", "04:00", "16:00", "23:00", "08:00", "09:00", "07:00", "17:00",  
"03:00", "14:00")
```

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

```
> DateTime <- as.POSIXlt(strptime(d, format = "%y%m%d %H:%M", tz = "UTC" ))
```

```
> class(DateTime)
```

```
[1] "POSIXlt" "POSIXt"
```

```
> Items <- as.factor( c("shirt", "shirt", "pants", "jacket", "jacket", "shirt", "jacket",  
"jacket", "shoes", "shirt"))
```

```
> class(Items)
```

```
[1] "factor"
```

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

```
> class(Volume)
```

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

```
> mySale <- data.frame(DateTime, Items, Volume)
```

```
> print(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)
> Items[Dates >= "0700"]
[1] shirt  shirt  pants  jacket jacket shirt  jacket jacket shoes  shirt
Levels: jacket pants shirt shoes
> sum(Volume[Dates >= "0700"], na.rm=T)
[1] 49730
```