

109-1 資料科學應用 HW2

班級:經濟四乙

學號:A106260018

姓名:張鈞茹

#第 1.13 題

> #某學生分析空氣品質資料 airquality 之風速 (Wind) 與溫度 (Temp) 的關係，

> #他採用迴歸分析及共變異數分析，步驟如下：

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

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

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

```
>
```

> #(a) 物件 lm.anova 是屬於何種類別，其儲存結構如何？

```
> 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"
```

```
>
```

> #(b) 物件 lm.summary 有哪一些屬性可供存取？試取出 R2 值。

```
> attributes(lm.anova)
```

```
$names
```

```
[1] "Df"      "Sum Sq"  "Mean Sq" "F value" "Pr(>F)"
```

```
$row.names
```

```
[1] "airquality$Temp" "Residuals"
```

```
$class
```

```
[1] "anova"      "data.frame"
```

```
$heading
```

```
[1] "Analysis of Variance Table\n" "Response: airquality$Wind"
```

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

```
[1] 0.2097529
```

```
>
```

```
>
```

```
> #####
```

```
> #第 1.20 題
```

```
> # 「statlog_vehicle_846x18.txt」是以 tab 為分隔的資料，具有 18 個變數，請讀入 R 之後，
```

```
> #列出資料框維度、前後各 5 筆紀錄及儲存此資料框物件所佔用的記憶體。
```

```
> data1 <- read.table("data/statlog_vehicle_846x18.txt",sep = "\t")
```

```
> data1
```

	V1	V2	V3	V4	V5	V6	V7	V8	
V9			V10						
1	no class	compactness	circularity	distance	radiusratio	pr.axis	max.length	scatterratio	elongatedness
2	1	0	96	55	103	201	65	9	
204			32						
3	2	0	101	56	100	215	69	10	
208			32						
4	3	0	93	35	66	154	59	6	
142			46						
5	4	0	101	48	107	222	68	10	
208			32						
6	5	0	87	38	85	177	61	8	
164			40						
7	6	0	95	48	104	214	67	9	
205			32						
8	7	0	98	55	101	228	70	9	
210			31						
9	8	0	107	53	103	221	66	11	
209			32						
10	9	0	103	50	98	212	63	9	
193			34						
11	10	0	77	38	63	135	59	5	
130			52						
12	11	0	89	41	75	143	56	7	
146			46						
13	12	0	98	55	101	219	69	11	
225			30						
14	13	0	96	55	98	161	54	10	

215		31						
15 14	0		97	59	108	227	70	11
224		30						
16 15	0		92	39	91	191	62	8
176		37						
17 16	0		73	37	53	111	54	6
126		55						
18 17	0		101	53	103	203	63	9
195		34						
19 18	0		79	40	80	133	55	7
147		47						
20 19	0		80	37	57	116	55	6
125		54						
21 20	0		94	38	84	158	55	9
169		39						
22 21	0		97	50	108	211	65	10
214		31						
23 22	0		95	46	105	219	68	9
201		33						
24 23	0		99	46	105	209	64	11
197		34						
25 24	0		85	39	77	151	59	8
150		45						
26 25	0		77	38	75	144	59	6
147		46						
27 26	0		88	35	50	121	58	5
114		59						
28 27	0		100	45	100	209	65	8
201		32						
29 28	0		102	54	100	163	53	10
213		31						
30 29	0		106	49	107	194	57	11
214		31						
31 30	0		95	45	80	186	62	7
164		40						
32 31	0		103	54	107	218	64	12
222		30						
33 32	0		93	35	72	172	62	7

149		44						
34 33	0		85	36	78	149	55	7
147		45						
35 34	0		91	45	75	154	57	6
150		44						
36 35	0		82	38	53	125	59	5
133		51						
37 36	0		107	52	101	218	64	11
202		33						
38 37	0		98	54	104	186	59	10
213		32						
39 38	0		103	54	91	179	57	11
220		31						
40 39	0		108	51	103	197	60	11
211		31						
41 40	0		84	39	90	180	60	7
177		37						
42 41	0		78	36	60	116	56	6
123		55						
43 42	0		98	45	76	166	60	7
157		42						
44 43	0		101	51	105	212	68	10
209		32						
45 44	0		90	36	78	179	64	8
157		42						
46 45	0		97	48	94	198	63	9
181		36						
47 46	0		111	54	103	171	50	11
221		30						
48 47	0		103	55	100	194	62	11
212		31						
49 48	0		92	46	79	176	64	8
162		41						
50 49	0		101	56	100	168	55	11
214		31						
	V11	V12	V13	V14	V15	V16	V17	V18
V19	V20							

1 pr.axis max.length scaledvmi scaledvma scaledradius skewness skewness kurtosis kurtosis hollows

2	23	166	227	624	246	74	6	2
186	194							
3	24	169	227	651	223	74	6	5
186	193							
4	18	128	162	304	120	64	5	13
197	202							
5	24	154	232	641	204	70	5	38
190	202							
6	20	129	186	402	130	63	1	25
198	205							
7	23	151	227	628	202	74	5	9
186	193							
8	24	168	236	661	245	72	1	6
188	197							
9	24	163	222	653	212	66	0	1
191	201							
10	22	161	214	567	185	64	5	5
198	204							
11	18	130	145	247	139	79	13	21
183	187							
12	19	137	170	317	156	76	18	5
184	188							
13	25	178	231	748	216	74	6	14
187	195							
14	24	175	226	683	221	76	3	6
185	193							
15	25	186	225	732	218	70	10	25
186	198							
16	21	137	196	466	151	67	3	23
192	200							
17	18	128	135	227	147	82	1	15
176	184							
18	22	162	210	571	210	68	5	5
191	198							
19	19	135	172	311	144	76	8	30
181	193							
20	18	125	142	229	132	81	8	5
178	184							

21	20	130	196	430	155	69	9	15
190	195							
22	24	156	232	683	218	72	7	29
188	197							
23	23	148	223	602	201	69	5	38
191	202							
24	23	152	212	575	159	65	0	33
194	205							
25	19	134	176	331	133	73	0	16
184	193							
26	19	132	167	315	136	80	16	20
181	187							
27	17	122	132	192	138	74	21	4
182	187							
28	23	147	231	611	189	72	5	5
189	195							
29	24	173	219	669	201	76	12	27
187	195							
30	24	161	224	670	172	67	0	39
192	206							
31	20	145	188	406	178	65	11	18
199	204							
32	25	174	221	728	199	67	0	18
189	200							
33	19	124	169	334	125	62	5	30
203	210							
34	19	128	168	321	134	64	10	24
197	203							
35	19	146	170	335	180	66	16	2
193	198							
36	18	128	152	259	146	87	0	0
177	183							
37	23	164	219	610	192	65	17	2
197	206							
38	24	172	223	665	217	73	1	26
186	195							
39	25	170	220	707	198	72	1	32
186	198							


```

3  2    0      101      56      100      215      69      10
208      32      24
4  3    0      93      35      66      154      59      6
142      46      18
5  4    0      101      48      107      222      68      10
208      32      24
      V12      V13      V14      V15      V16      V17      V18      V19
V20

```

```

1 max.length scaledvmi scaledvma scaledradius skewness skewness kurtosis kurtosis hollows
2      166      227      624      246      74      6      2      186
194
3      169      227      651      223      74      6      5      186
193
4      128      162      304      120      64      5      13      197
202
5      154      232      641      204      70      5      38      190
202

```

```

> tail(data1,5)
      V1 V2 V3 V4 V5  V6 V7 V8  V9 V10 V11 V12 V13 V14 V15 V16 V17 V18 V19 V20
843 842  3 87 45 66 139 58  8 140  47  18 148 168 294 175  73  3  12 188 196
844 843  3 95 43 76 142 57 10 151  44  19 149 173 339 159  71  2  23 187 200
845 844  3 90 44 72 157 64  8 137  48  18 144 159 283 171  65  9  4 196 203
846 845  3 89 46 84 163 66 11 159  43  20 159 173 368 176  72  1  20 186 197
847 846  3 85 36 66 123 55  5 120  56  17 128 140 212 131  73  1  18 186 190

```

```

>
> print(object.size(data1),units = "Mb")
0.3 Mb

```

```

>
>
> #####

```

```

> #第 1.28 題

```

```

> #8 讀取「stock-data.txt」資料檔，印出資料前 5 筆紀錄、後 5 筆紀錄。
> #檢查 (印出)資料每一變數 (欄位) 是否有符合 R 的類別物件，若沒有，請更改。
> data2 <- read.table("data/stock-data.txt",sep = "\t")
> data2

```

```

      V1  V2  V3  V4  V5      V6
V7
      V8

```

```

1  民國 100 年 5 家半導體公司股票月成交資訊(元,股)

```


2		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						
3		台積電	100	1	78.3	69.6	74.3
263,999	100,578,274,926						
4		台積電	100	2	77	69.9	72.54
235,159	74,985,055,548						
5		台積電	100	3	72.2	65.7	69.74
276,434	88,459,924,495						
6		台積電	100	4	73.9	68	71.37
211,611	70,177,023,098						
7		台積電	100	5	76.9	73	74.96
213,185	74,005,599,560						
8		台積電	100	6	78.2	70.4	74.7
260,507	96,761,306,205						
9		台積電	100	7	73.9	68.5	71.59
238,386	73,569,965,426						
10		台積電	100	8	72.8	62.2	66.61
305,409	84,617,942,159						
11		台積電	100	9	72.1	65.9	69.11
266,720	74,225,030,814						
12		台積電	100	10	74	68.1	70.7
181,361	59,947,670,693						
13		台積電	100	11	76	71.3	74.03
197,579	65,432,526,407						
14		台積電	100	12	76.8	72	75
179,107	53,687,756,290						
15		威盛	100	1	33.4	29.3	30.97
55,107	4,580,913,795						
16		威盛	100	2	32.65	28.35	30.54
26,901	2,060,809,696						
17		威盛	100	3	35.45	28.5	32.01
55,802	4,355,434,679						
18		威盛	100	4	32.8	27.55	30.35
27,568	1,815,454,798						
19		威盛	100	5	32.6	25.95	29.4
37,516	2,758,375,085						
20		威盛	100	6	37.25	31.2	34.68
89,247	7,828,188,732						

21		威盛	100	7	38.15	32.45	35.47
67,463	5,968,464,729						
22		威盛	100	8	35.4	26.6	30.13
45,393	3,364,616,892						
23		威盛	100	9	29	23.1	26.17
24,781	1,477,865,479						
24		威盛	100	10	25.15	20.4	23.39
25,791	1,528,259,415						
25		威盛	100	11	25.7	18.7	22.74
29,099	1,687,413,881						
26		威盛	100	12	20.2	14.8	16.96
21,092	856,362,397						
27		聯發科	100	1	424	378	403.55
106,530	57,621,649,341						
28		聯發科	100	2	380	325.5	348.98
97,339	46,409,931,806						
29		聯發科	100	3	355	312.5	339.96
117,960	52,887,228,668						
30		聯發科	100	4	354	301	328.65
87,638	39,442,097,346						
31		聯發科	100	5	362.5	305.5	335.42
128,717	60,665,847,316						
32		聯發科	100	6	331	295	311.57
110,521	50,190,673,665						
33		聯發科	100	7	316.5	244	274.39
161,471	67,807,228,929						
34		聯發科	100	8	298	221	262.09
249,066	99,279,007,797						
35		聯發科	100	9	348	268	309.66
240,792	110,850,615,666						
36		聯發科	100	10	345	310.5	329.66
185,407	86,245,899,331						
37		聯發科	100	11	326	268	302.52
160,330	66,694,256,195						
38		聯發科	100	12	292	243	268.01
135,509	50,261,172,442						
39		聯電	100	1	18.2	15.5	17.19
258,572	31,112,735,815						

40		聯電	100	2	18.3	15.3	16.38
150,872	14,737,456,282						
41		聯電	100	3	16.1	13.9	14.92
209,011	19,678,194,951						
42		聯電	100	4	15.65	14.55	15.21
125,663	11,339,720,871						
43		聯電	100	5	15.3	14.25	14.76
116,087	10,613,932,085						
44		聯電	100	6	15.15	13.85	14.51
125,348	11,651,143,825						
45		聯電	100	7	14.5	12.95	13.89
122,812	11,900,583,208						
46		聯電	100	8	13.15	10.15	11.13
169,781	13,165,667,283						
47		聯電	100	9	12.05	10.65	11.25
127,617	9,214,851,731						
48		聯電	100	10	13.7	11.05	12.39
113,378	7,702,645,861						
49		聯電	100	11	13.3	11.6	12.68
107,400	7,641,319,053						
50		聯電	100	12	13.6	11.7	12.51
99,760	6,317,139,669						
51		旺宏	100	1	23.75	20.2	22.19
241,726	24,488,010,731						
52		旺宏	100	2	22.95	20.3	21.49
113,440	10,237,820,122						
53		旺宏	100	3	22.4	17.65	19.48
208,006	16,814,336,067						
54		旺宏	100	4	19.65	18.05	18.88
107,292	7,081,789,345						
55		旺宏	100	5	18.9	17.4	18.25
103,567	7,221,174,001						
56		旺宏	100	6	18.15	16.9	17.6
72,617	4,294,383,140						
57		旺宏	100	7	18.5	14.4	17.09
125,851	8,571,233,298						
58		旺宏	100	8	14.5	10.25	11.84
152,177	8,137,500,167						

59		旺宏	100	9	12.65	10.4	11.55
108,879	5,542,998,380						
60		旺宏	100	10	12	10.25	11.31
68,571	3,041,525,834						
61		旺宏	100	11	13.65	10.85	12.54
167,018	9,538,526,797						
62		旺宏	100	12	12.85	11.15	12.17
115,192	5,070,210,532						

V9

V10

1		
2	成交股數	週轉率百分比
3	1,353,616,348	5.22
4	1,033,654,452	3.98
5	1,268,289,393	4.89
6	983,177,475	3.79
7	987,256,484	3.8
8	1,295,262,736	4.99
9	1,027,567,656	3.96
10	1,270,302,342	4.9
11	1,073,997,108	4.14
12	847,821,278	3.27
13	883,753,804	3.41
14	715,808,271	2.76
15	147,912,893	21.54
16	67,459,942	9.82
17	136,059,651	19.81
18	59,799,382	8.7
19	93,810,158	13.66
20	225,687,324	32.86
21	168,228,930	24.5
22	111,649,410	16.26
23	56,460,496	8.22
24	65,336,840	9.51
25	74,175,097	10.8
26	50,464,211	7.34
27	142,786,216	12.98
28	132,985,689	12.08
29	155,567,203	14.14

30	120,011,172	10.91
31	180,862,384	16.44
32	161,084,547	14.64
33	247,119,699	22.46
34	378,794,148	34.43
35	357,971,048	30.97
36	261,616,653	22.64
37	220,461,694	19.21
38	187,529,947	16.34
39	1,809,650,075	13.93
40	899,524,191	6.92
41	1,318,563,860	10.15
42	745,385,215	5.73
43	718,857,838	5.53
44	802,571,097	6.17
45	856,247,283	6.55
46	1,182,650,262	9.04
47	818,390,302	6.25
48	621,343,297	4.74
49	602,169,179	4.6
50	504,611,921	3.85
51	1,103,457,390	32.81
52	476,337,345	14.13
53	863,074,087	25.58
54	374,989,300	11.1
55	395,658,986	11.7
56	243,965,636	7.22
57	501,422,845	14.82
58	687,167,610	20.31
59	479,779,350	14.18
60	268,710,697	7.94
61	760,264,306	22.47
62	416,455,073	12.31

```
> class(data2)
```

```
[1] "data.frame"
```

```
>
```

```
> head(data2,5)
```

V1

V2

V3

V4

V5

V6

```

V7              V8
1  民國 100 年 5 家半導體公司股票月成交資訊(元,股)
2              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
3              台積電 100    1   78.3   69.6           74.3
263,999 100,578,274,926
4              台積電 100    2    77    69.9           72.54
235,159  74,985,055,548
5              台積電 100    3   72.2   65.7           69.74
276,434  88,459,924,495

```

```

          V9          V10
1
2      成交股數 週轉率百分比
3 1,353,616,348          5.22
4 1,033,654,452          3.98
5 1,268,289,393          4.89

```

```
> tail(data2,5)
```

```

      V1  V2 V3    V4    V5    V6    V7          V8          V9  V10
58 旺宏 100  8  14.5 10.25 11.84 152,177 8,137,500,167 687,167,610 20.31
59 旺宏 100  9  12.65  10.4 11.55 108,879 5,542,998,380 479,779,350 14.18
60 旺宏 100 10   12 10.25 11.31  68,571 3,041,525,834 268,710,697  7.94
61 旺宏 100 11 13.65 10.85 12.54 167,018 9,538,526,797 760,264,306 22.47
62 旺宏 100 12 12.85 11.15 12.17 115,192 5,070,210,532 416,455,073 12.31

```

```
>
```

```
> lapply(data2,class)
```

```
$V1
```

```
[1] "character"
```

```
$V2
```

```
[1] "character"
```

```
$V3
```

```
[1] "character"
```

```
$V4
```

```
[1] "character"
```

```
$V5
```

[1] "character"

\$V6

[1] "character"

\$V7

[1] "character"

\$V8

[1] "character"

\$V9

[1] "character"

\$V10

[1] "character"

>

>

> #####

> #第 1.33 題

> #某銷售人員在 2018 年的網路銷售紀錄從公司資料庫隨機抽樣 10 筆如下:

> #Dates: 0924, 1112, 1231, 1105, 0604, 0219, 0416, 0611, 0813, 1029

> #Time: 01:00, 04:00, 16:00, 23:00, 08:00, 09:00, 07:00, 17:00, 03:00, 14:00

> #Items: shirt, shirt, pants, jacket, jacket, shirt, jacket, jacket, shoes, shirt

> #Volume: 7951, 159,1958, 6848, 3762, 3678, 8696, 9045, 6208, 1425

>

> #(a) 請將上述資料儲存成一資料框 (data.frame) 類別之物件，命名 mySale，

> #使得第一個欄位為銷售日期時間 (DateTime)，類別為 POSIXct，時區為世界協調時間 (UTC);

> #第二個欄位為銷售品項 (Items)，類別為 factor;

> #第三個欄位為銷售量 (Volume)，類別為 numeric。印出 mySale。

> 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] "character"
>
> 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
>

```