

2020/11/20(五), 109 學年第一學期 資料科學應用 R 小考(1)

> #

> # 學號: A107260032 姓名: 歐佩姍

> #

> # 本檔案為各題之程式碼檔，無執行結果

>

>

> # ex1(a)

> library(readxl)

> scoreB <- read_excel("data/Calculus-score-B.xls", na = "NA", skip=2)

New names:

* `0.070000000000000007` -> `0.070000000000000007...5`

* `0.070000000000000007` -> `0.070000000000000007...6`

* `0.080000000000000002` -> `0.080000000000000002...7`

* `0.080000000000000002` -> `0.080000000000000002...8`

> names(scoreB)[1:12] <- c("No", "ID", "Name", "Gender", "Quiz1",
+ "Quiz2", "Quiz3", "Quiz4", "TA", "Midterm",
+ "Final", "ATT")

> scoreA <- read_csv("data/Calculus-score-A.csv", header = TRUE, na = "NA", skip=2)

> names(scoreA)[1:12] <- c("No", "ID", "Name", "Gender", "Quiz1",
+ "Quiz2", "Quiz3", "Quiz4", "TA", "Midterm",
+ "Final", "ATT")

> head(scoreA, 5)

	No	ID	Name	Gender	Quiz1	Quiz2	Quiz3	Quiz4
1	1	401405008	希瑄彥	男	10	0	5	20
2	2	401550880	張泓丞	男	25	40	70	87
3	3	404550061	張安婕	女	18	15	48	33
4	4	404550042	柯政學	男	10	10	NA	NA
5	5	404550023	謝文躍	女	35	45	52	97

	TA	Midterm	Final	ATT
1	0.0	55	50	2
2	80.0	46	68	9
3	86.7	54	79	9
4	13.3	2	0	7
5	86.7	55	67	9

> tail(scoreA, 5)

	No	ID	Name	Gender	Quiz1	Quiz2	Quiz3	Quiz4
36	36	404550369	陳王霖	女	55	73	92	73

37	37	404550420	何瑄穎	男	28	10	35	3
38	38	404550431	沈泓霏	女	15	25	53	67
39	39	404550442	許安霏	女	53	60	80	72
40	40	404550453	李政宜	男	80	100	85	100

TA Midterm Final ATT

36	100.0	72	81	9
37	66.7	30	0	7
38	93.3	29	42	9
39	100.0	61	62	9
40	100.0	95	100	3

> head(scoreB, 5)

A tibble: 5 x 12

	No	ID	Name	Gender	Quiz1	Quiz2	Quiz3	Quiz4
	<dbl>	<dbl>	<chr>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>
1	1	4.05e8	史文羽~	男	60	81	100	97
2	2	4.05e8	鄭樺好~	男	80	100	100	92
3	3	4.05e8	張敬安~	男	10	40	62	93
4	4	4.05e8	何筑亦~	女	15	25	40	13
5	5	4.05e8	張儀~	女	30	45	70	61

... with 4 more variables: TA <dbl>, Midterm <dbl>,

Final <dbl>, ATT <dbl>

> tail(scoreB, 5)

A tibble: 5 x 12

	No	ID	Name	Gender	Quiz1	Quiz2	Quiz3	Quiz4
	<dbl>	<dbl>	<chr>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>
1	51	4.05e8	鄭鈺尤~	女	80	85	100	85
2	52	4.05e8	楊宜路~	男	48	35	48	98
3	53	4.05e8	張渝好~	男	0	38	60	40
4	54	4.05e8	廖暄安~	男	50	70	20	85
5	55	5.00e8	楊毅亦~	女	5	35	45	55

... with 4 more variables: TA <dbl>, Midterm <dbl>,

Final <dbl>, ATT <dbl>

>

> # ex1(b)

> options("max.print"=10000)

> scoreA\$class <- "A"

> scoreB\$class <- "B"

> score <- rbind(scoreA, scoreB)

```

>
> # ex1(c)
> score[is.na(score)] <- 0
> score$"學期成績" <- (0.07*score$"Quiz1" + 0.07*score$"Quiz2"
+
+ 0.08*score$"Quiz3" + 0.08*score$"Quiz4"
+
+ 0.15*score$"TA" + 0.25*score$"Midterm" +
0.3*score$"Final") +score$"ATT"
> score$"學期成績" <- ifelse(score$"學期成績" >= 100, 100, score$"學期成績")
>
> # ex1(d)
> subset(score, score$"學期成績" >= 55 & score$"學期成績" < 60)

```

No	ID	Name	Gender	Quiz1	Quiz2	Quiz3	Quiz4
38	38 404550431	沈泓霏	女	15	25	53	67
45	5 404685100	張儀	女	30	45	70	61
68	28 404720722	楊佳聿	女	30	35	20	50
70	30 404720527	馨飛羽	男	15	0	45	65
95	55 499555916	楊毅亦	女	5	35	45	55

TA	Midterm	Final	ATT	class	學期成績	
38	93.3	29	42	9	A	55.245
45	93.3	29	48	4	B	55.375
68	60.0	45	63	9	B	58.300
70	93.0	44	44	7	B	55.000
95	87.0	58	60	3	B	59.350

```

>
> # ex1(e)
> mean(score$"學期成績"[1:40])
[1] 58.84575
> mean(score$"學期成績"[41:95])
[1] 61.123
> mean(score$"學期成績"[score$"Gender"=="女"])
[1] 58.95292
> mean(score$"學期成績"[score$"Gender"=="男"])
[1] 60.90322
>
> # ex1(f)
> unfailA <- length(score$"學期成績"[score$"學期成績" <60 & score$"class"=="A"])
> A <- unfailA/length(score$"學期成績"[score$"class"=="A"])
> sprintf("%.2f%%", A * 100)

```

```

[1] "35.00%"
> unfailB <- length(score$"學期成績"[score$"學期成績" <60 & score$"class"== "B"])
> B <- unfailB/length(score$"學期成績"[score$"class"== "B"])
> sprintf("%.2f%%", B * 100)
[1] "45.45%"
>
> # ex1(g)
> G <- subset(score, Gender == "女")
> g <- G[order(-G$"學期成績"),c(13,2,3,14)]
> g$rank <- c(1:36)
> head(g, 5)
  class      ID   Name 學期成績 rank
89    B 404720541 詹傑仙  100.000    1
91    B 404685407 鄭鈺尤  100.000    2
11    A 404550189 丁易偉   96.435    3
36    A 404550369 陳王霖   88.460    4
79    B 404720436 曼李儷   85.140    5
> B <- subset(score, Gender == "男")
> b <- B[order(-B$"學期成績"),c(13,2,3,14)]
> b$rank <- c(1:59)
> head(b, 5)
  class      ID   Name 學期成績 rank
40    A 404550453 李政宜   99.15    1
42    B 404685071 鄭樺妤   97.06    2
90    B 404685109 許 何   95.72    3
58    B 404720161 劉莞韋   94.07    4
41    B 404550465 史文羽   94.03    5
>
> # ex2(a)
> set.seed(123456)
> Letters.code <- sample(LETTERS[1:5], 20, replace=T)
> z <- c("A", "E")
> x <- "C"
> y <- c("B", "D")
> Numbers.code <- ifelse(Letters.code %in% z, 1,
+   ifelse(Letters.code %in% x, 2, 3)
+ )
>

```

```
> # ex2(b)
> data.frame(Letters.code, Numbers.code)
  Letters.code Numbers.code
1           D             3
2           B             3
3           B             3
4           A             1
5           E             1
6           D             3
7           B             3
8           C             2
9           C             2
10          E             1
11          C             2
12          D             3
13          B             3
14          E             1
15          A             1
16          B             3
17          E             1
18          C             2
19          D             3
20          C             2
>
```