

109-1 資料科學應用 HW7

班級:經濟四乙

學號:A106260018

姓名:張鈞茹

> #2.30

> #有某一試卷之測驗結果，紀錄於"answer.txt"。試卷中 10 題選擇題之正確答案依序為

> #B, D, B, D, D, A, C, D, C, B

> #(a) 請讀取此資料，並列印前 5 筆紀錄。

> answer <- read.table("data/answer.txt", na = "NA", skip = 1)

> head(answer)

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11
1 s1	C	D	D	A	D	A	B	C	C	B	
2 s2	B	D	B	D	D	A	C	D	B	B	
3 s3	B	A	A	B	D	A	C	B	C	B	
4 s4	B	D	B	A	B	C	C	D	C	B	
5 s5	B	D	D	D	A	C	C	D	A	B	
6 s6	D	B	B	B	D	A	B	C	C	B	

>

> #2.51

> #(a) 資料壓縮: 將字串"AAABBBCCCC" 表示成"3A3B4C"。

> x <- "AAABBBCCCC"

> y <- c("A", "B", "C")

> num <- 0

> for (i in 1:length(y)) {

+ num[i] <- length(gregexpr(y[i], x)[[1]])

+ }

> cat(num[1],"A",num[2],"B",num[3],"C")

3 A 3 B 4 C>

> #2.51

> #(b) 資料解壓縮: 將字串"3A3B4C" 表示"AAABBBCCCC"。

> x <- "3A3B4C"

> times <- substr(rep("3A3B4C",3), c(1,3,5),c(1,3,5))

> cat(rep(c("A", "B", "C"),times))

A A A B B B C C C C>

> #2.52

> #寫一「簡單資料壓縮」之 R 函式 (命名為 compress): 輸入為 ABC 三個字

母組成之字串,

> #例如: 字串"ABAABBAABCCCAC", 輸出為每個字母出現之次數, (例

如:"6A4B4C")。

```
> compress <- function(x){  
+   for(i in 1:3){  
+     text <- gregexpr(LETTERS[i], x)  
+     cat(paste0(length(text[[1]]), LETTERS[i]))  
+   }  
+   cat("\n")  
+ }  
> compress("ABAABBAABCCCAC")  
6A4B4C  
>  
> #5.2
```

> #一袋中有 6 顆白球 4 顆紅球, 隨機從中抽取 3 球 (取出不放回),

> #若 $P(A)$ 代表抽中 2 顆白球及 1 顆紅球的機率, 試求 $P(A)$ 。

```
> set.seed(123456)
```