2020/12/18(五), 109 學年第一學期 資料科學應用 R 作業(6)

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

```
> > #2020/1/7
>
> #2.30(a)
> answer <- read.delim("data/answer.txt")
> first5.records <- head(answer,5)
>
> #2.30(b)
> correct.ans <- c('B', 'D', 'B', 'D', 'D', 'A', 'C', 'D', 'C', 'B')
> stu.a <- c("A", "D", "B", "D", "B", "A", "B", "D", "C", "B")
> correct.item <- which(correct.ans == stu.a)
> correct.item
[1]  2  3  4  6  8  9 10
> n.correct <- 10*length(correct.item)
> n.correct
[1] 70
>
> #2.30(c)
> v1.1 <-ifelse(answer$V1 == correct.ans[1],1,0)
> v2.1 <-ifelse(answer$V2 == correct.ans[2],1,0)
> v3.1 <-ifelse(answer$V3 == correct.ans[3],1,0)
> v4.1 <-ifelse(answer$V4 == correct.ans[4],1,0)
> v5.1 <-ifelse(answer$V5 == correct.ans[5],1,0)
> v6.1 <-ifelse(answer$V6 == correct.ans[6],1,0)
> v7.1 <-ifelse(answer$V7 == correct.ans[7],1,0)
> v8.1 <-ifelse(answer$V8 == correct.ans[8],1,0)
> v9.1 <-ifelse(answer$V9 == correct.ans[9],1,0)
> v10.1 <-ifelse(answer$V10 == correct.ans[10],1,0)
> ans.new <-
data.frame(answer$Student,v1.1,v2.1,v3.1,v4.1,v5.1,v6.1,v7.1,v8.1,v9.1,v10.1)
```

```
> stu.ans <-
10*(ans.new$v1.1+ans.new$v2.1+ans.new$v3.1+ans.new$v4.1+ans.new$v5.1+ans.n
ew$v6.1+ans.new$v7.1+ans.new$v8.1+ans.new$v9.1+ans.new$v10.1)
> ans.news <- data.frame(ans.new,stu.ans)
> table(ans.news$stu.ans)

  0  10  20  30  40  50  60  70  80  90 100
  3  10   9  11  19  23  28  40  30  12   7
>
> #ex2.30(d)
> #若設定總得分前 25% 為高分組，總得分後 25% 為低分組，則哪些學生是
高分組，哪些學生是低分組，而人數各為多少人。
> top <- ans.news$answer.Student[which(ans.news$stu.ans > 100*0.75)]
> top
 [1] "s2"   "s12"  "s16"  "s19"  "s20"  "s21"  "s24"
 [8] "s25"  "s27"  "s31"  "s41"  "s43"  "s44"  "s47"
[15] "s50"  "s52"  "s54"  "s55"  "s66"  "s69"  "s73"
[22] "s79"  "s80"  "s81"  "s86"  "s95"  "s96"  "s108"
[29] "s110" "s112" "s123" "s125" "s128" "s129" "s131"
[36] "s135" "s136" "s139" "s143" "s146" "s152" "s157"
[43] "s159" "s165" "s171" "s187" "s189" "s190" "s192"
> low <- ans.news$answer.Student[which(ans.news$stu.ans < 100*0.25)]
> low
 [1] "s17"  "s32"  "s65"  "s71"  "s74"  "s82"  "s87"
 [8] "s90"  "s97"  "s105" "s107" "s120" "s132" "s142"
[15] "s160" "s161" "s163" "s168" "s169" "s174" "s177"
[22] "s178"
> n.top <- length(top)
> n.top
[1] 49
> n.low <- length(low)
> n.low
[1] 22
>
> #ex2.30(e)
> #試計算高分組及低分組在每?題答對的人數百分比, 記為 PH 及 PL。
> #高分組
> top1 <- round(sum(ans.news$v1.1[which(ans.news$stu.ans > 100*0.75)])/n.top,2)
```

```
> top2 <- round(sum(ans.news$v2.1[which(ans.news$stu.ans > 100*0.75)])/n.top,2)
> top3 <- round(sum(ans.news$v3.1[which(ans.news$stu.ans > 100*0.75)])/n.top,2)
> top4 <- round(sum(ans.news$v4.1[which(ans.news$stu.ans > 100*0.75)])/n.top,2)
> top5 <- round(sum(ans.news$v5.1[which(ans.news$stu.ans > 100*0.75)])/n.top,2)
> top6 <- round(sum(ans.news$v6.1[which(ans.news$stu.ans > 100*0.75)])/n.top,2)
> top7 <- round(sum(ans.news$v7.1[which(ans.news$stu.ans > 100*0.75)])/n.top,2)
> top8 <- round(sum(ans.news$v8.1[which(ans.news$stu.ans > 100*0.75)])/n.top,2)
> top9 <- round(sum(ans.news$v9.1[which(ans.news$stu.ans > 100*0.75)])/n.top,2)
> top10 <- round(sum(ans.news$v10.1[which(ans.news$stu.ans >
100*0.75)])/n.top,2)
> PH <- c(top1,top2,top3,top4,top5,top6,top7,top8,top9,top10)
> PH
 [1] 0.88 0.88 0.78 0.80 0.88 0.82 0.96 0.76 0.86 0.94
>
> #低分組
> low1 <- round(sum(ans.news$v1.1[which(ans.news$stu.ans < 100*0.25)])/n.low,2)
> low2 <- round(sum(ans.news$v2.1[which(ans.news$stu.ans < 100*0.25)])/n.low,2)
> low3 <- round(sum(ans.news$v3.1[which(ans.news$stu.ans < 100*0.25)])/n.low,2)
> low4 <- round(sum(ans.news$v4.1[which(ans.news$stu.ans < 100*0.25)])/n.low,2)
> low5 <- round(sum(ans.news$v5.1[which(ans.news$stu.ans < 100*0.25)])/n.low,2)
> low6 <- round(sum(ans.news$v6.1[which(ans.news$stu.ans < 100*0.25)])/n.low,2)
> low7 <- round(sum(ans.news$v7.1[which(ans.news$stu.ans < 100*0.25)])/n.low,2)
> low8 <- round(sum(ans.news$v8.1[which(ans.news$stu.ans < 100*0.25)])/n.low,2)
> low9 <- round(sum(ans.news$v9.1[which(ans.news$stu.ans < 100*0.25)])/n.low,2)
> low10 <- round(sum(ans.news$v10.1[which(ans.news$stu.ans <
100*0.25)])/n.low,2)
> PL <- c(low1,low2,low3,low4,low5,low6,low7,low8,low9,low10)
> PL
 [1] 0.14 0.14 0.27 0.18 0.09 0.14 0.14 0.09 0.00 0.09
>
> #ex2.30(f)
> #請計算每?題之難度 (公式 P = (PH + PL)/2) 及鑑別度 (公式 D = PH ? PL)。
> P <- round((PH+PL)/2,2)
> P
 [1] 0.51 0.51 0.52 0.49 0.48 0.48 0.55 0.42 0.43 0.52
> D <-(PH-PL)
> D
 [1] 0.74 0.74 0.51 0.62 0.79 0.68 0.82 0.67 0.86 0.85
```

```
>
> #ex2.51(a)
> #(a) 資料壓縮: 將字串"AAABBBCCCC" 表示成"3A3B4C"。 (提示:
gregexpr,cat。)
> 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>
> #ex2.51(b)
> #(b) 資料解壓縮: 將字串"3A3B4C" 表示"AAABBBCCCC"。 (提示: substr,
cat,rep)
> x.1 <- c("3A3B4C")
> t4 <- substr(rep(x.1,3),c(1,3,5),c(1,3,5))
> a1 <- rep(LETTERS[1:3],times=t4)
> a1
 [1] "A" "A" "A" "B" "B" "B" "C" "C" "C" "C"
>
> #ex2.52
> #輸入為 ABC 三個字?組成之字串, 例如: 字串"ABAABBAABCCCAC"，輸出為每
個字?出現之次數"6A4B4C"
> compress <- function(x){
+     c1 <-LETTERS[1:3]
+     for(i in c1){
+         c2 <- length(gregexpr(i,x)[[1]])
+         cat(c(c2,i,''))
+   }
+     cat("\n")
+ }
> compress("ABAABBAABCCCAC")
6 A 4 B 4 C
>
> #ex5.2(a)
> #白球:w <- 0
> #紅球:R <- 1
```

```
> ball <- rep(c(0,1),c(6,4))
> ball
 [1] 0 0 0 0 0 0 1 1 1 1
> play <- sample(ball,3,replace = FALSE)
> set.seed(123456)
> cat('紅球:',sum(play),'白球:',(length(play)-sum(play)))
紅球: 2  白球: 1>
> #ex5.2(b)
> set.seed(123456)
> play1 <- replicate(10, sample(ball, 3, replace = FALSE))
> R <- colSums(play1)
> W <- rep(3,times=10)-R
> cbind(R,W)
      R W
 [1,] 1 2
 [2,] 1 2
 [3,] 1 2
 [4,] 2 1
 [5,] 1 2
 [6,] 1 2
 [7,] 2 1
 [8,] 1 2
 [9,] 1 2
[10,] 2 1
>
> #ex5.2(c)
> #重覆上述實驗 100 次，計算抽中 2 顆白球及 1 顆紅球的機率
> set.seed(123456)
> play2 <- replicate(100, sample(ball, 3, replace = FALSE))
> R1 <- colSums(play2)
> sum(R1==1)/100
[1] 0.51
>
```