

```
--- title: "109-1 資料科學應用 - R語言篇" subtitle: "期末考" author: "學號:A107260063 姓名:關嘉良"
date: "r format(Sys.time(), '%d %B %Y')"
```

注意事項 {unnumbered} ### 下載題目卷 {-} * 於課程網站()下載題目卷。### 考試期間 {-} * 請按照平時上課之座位入座。* 可參考課本、上課講義(包含電子檔)及其它資料，但不能與別人討論。* 可使用計算機、自己的筆記型電腦、平板電腦及手機。* 全程可上網查詢，但不能用通訊軟體(例如: FB/LINE/IG)討論，也不可抄襲網路上之程式碼。* 不按照規定作答者，酌量扣分。* 有問題者，請舉手發問，勿與同學交談。* 不可使用它人之隨身碟。* 「作弊」或「疑似作弊」，本學期總成績不予評分。* 程式請隨時存檔，避免突然意外發生，程式檔不見。### 答題檔案原則 {-} * 若程式碼直接複製(或照抄)講義上的以不給分為原則。* 程式碼請直接寫在本'Rmd'檔。經由'Knit'編譯出'.html'檔。* 作答檔案，請隨時存檔並備份，勿直接存於公用電腦(例如: 桌面)。### 上傳答題檔案 {-} * 上傳方式同平時作業方式。* 請上傳三個檔案: 「學號-姓名-SHU-R-FinalExam.Rmd」、「學號-姓名-SHU-R-FinalExam.html」及「學號-姓名-SHU-R-FinalExam.pdf」。其中'pdf'檔為使用瀏覽器(IE、Edge、Firefox、Chrome)開啟'.html'印出成PDF檔。* 若上傳檔案格式錯誤，內容亂碼，空檔等等問題。請自行負責。

```
# 抽球算機率 一袋中有5顆紅球及3顆白球，小明由袋中隨機抽球，每次取一球，共取4次，令$A$為
抽出2次白球的事件，計算此事件 分別在放回(replacement)、不放回(without replacement)兩種情況下
之機率$P(A_r)$, $P(A_w)$。 \[ \mbox{放回}: P(A_r)=C^4_2\left(\frac{5}{8}\right)^2\left(\frac{3}{8}\right)^2 \] \[ \mbox{不放回}: P(A_w)=\frac{C^5_2C^3_2}{C^8_4} \] ## 直接算機率 請利用上式
($C$為組合數)，使用'R'指令直接計算上述之機率(分別命名為'Prob.Ar','Prob.Aw')並印出。 ``{r} #
your source code here Prob.Ar <- funtion(x){P(A_r)} Prob.Aw <- funtion(x){P(A_w)} `` ## 模擬抽球 小明
今天想要以寫'R'程式的方式來模擬此隨機實驗，計算抽球的機率，若設定{tt set.seed(123456)}，列
出「一袋中有5顆紅球及3顆白球，小明由袋中隨機抽球，分別在放回(replacement)、不放回(without
replacement)兩種情況下，每次取一球，共取4次」實驗一次的結果，並計數印出白球出現之個數。
(不需寫成'R'函式)(提示: 'sample','table') ``{r} # your source code here set.seed(123456) ball <- c("白
球","紅球") bag <- rep(ball, c(3, 5)) s <- sample(bag, 4) table(s) `` ## 重覆實驗10次 同上小題，寫一'R'
函式(命名為'Draw_Ball')，沒有輸入，輸出為白球 分別在放回、不放回兩種情況下的個數。重覆上
述實驗10次，印出分別在放回、不放回兩種情況下白球出現的個數。(提示: 'as.data.frame',
'replicate') ``{r} # your source code here Draw_Ball <- function(){ ball <- c("白球","紅球") bag <-
rep(ball, c(3, 5)) ans <- sample(bag, 4) table(factor(ans, levels=ball)) } DrawResult <-
as.data.frame(t(replicate(10, Draw_Ball()))) DrawResult `` ## 重覆實驗1000次 同上小題，重覆上述實驗
1000次，計算 在放回、不放回兩種情況下，抽到2顆白球的機率。(提示: 'as.data.frame','replicate',
'sum','==') ``{r} # your source code here n <- 1000 set.seed(123456) DrawResult <-
as.data.frame(t(replicate(n, Draw_Ball()))) sum((DrawResult$白球==2)/n) `` # 字串處理 某商業公司舉行
抽獎活動，中獎名單紀錄於'award-list.xlsx'檔中，包含 會員姓名、會員卡號及得獎金額。## 讀取資
料 請讀取此檔案，並印出全部中獎名單。 ``{r} # your source code here library(readxl) award.list <-
read.table("award-list.xlsx", header=T, sep="\t") `` ## 屏蔽字元 因考量個資法，公告名單不能將全名公
開，請你幫此名單，每一中獎者的姓名及會員卡號，部份字元打上'*'，例如第一筆紀錄為「沈俞予
7113235607」，請改為「沈*予7113***607」，印出修改後可公告之名單。(提示: 'substr') ``{r} #
your source code here substr(award.list, "") `` ## 排序資料 承上小題，請將修改後之名單，依照「得獎
金額」由多至少的順序，全部印出。 ``{r} # your source code here `` # 屬質變異指數(IQV) 計算名目
變數(nominal variable)的變異分散程度，其中Index of Qualitative Variation (IQV)是一個指標(其數值是
介於0與1中間)。公式如下: 
$$[IQV = \frac{k(n^2 - \sum f^2)}{n^2(k-1)},]$$
 其中$k$是類別數或組數，$n$是
樣本數，$\sum \{f^2\}$是將各類別次數之平方加起來之總和。假設有一名目變數資料('nv')如下，試寫
一'R'函式，計算IQV。(提示: 'table') `` set.seed(12345) no <- sample(20:100, 1) nv <-
LETTERS[sample(1:26, 5)][sample(1:5, no, replace=T)] `` ``{r} # your source code here ``
```