

20210319

統計學演習課內容

```
#####109 學年度第二學期 R 語言#####
#####2021/03/19#####
##運算與指派運算##
1+1

## [1] 2

x <- 2+1
y <-3+4
z <- 4+9

##Removes all items in Environment!##
#rm(list=ls()) #ls 當前所有物件
#rm(x)
#rm(y, z)
#rm(x, y, z)
#ls() #顯示當前物件
#objects() #顯示當前物件

#####
##常用函式##
x <- c(4, 9, 16, 25) #建構一個向量
mean(x) #x 的平均數

## [1] 13.5

median(x) #x 的中位數

## [1] 12.5

range(x)

## [1] 4 25

var(x) #x 的變異數

## [1] 83

sd(x) #x 的標準差

## [1] 9.110434

max(x) #x 的最大值

## [1] 25
```

```
min(x) #x 的最小值
## [1] 4

sqrt(x) #x 的開根號
## [1] 2 3 4 5

abs(x) #絕對值
## [1] 4 9 16 25

y <- 0.12345678
round(y, 2) #對y 做四捨五入到第2 位
## [1] 0.12

log(x) #取對數，底為 e
## [1] 1.386294 2.197225 2.772589 3.218876

log(x, 10) #底為 10
## [1] 0.6020600 0.9542425 1.2041200 1.3979400

exp(x) #指數函數，e^x
## [1] 5.459815e+01 8.103084e+03 8.886111e+06 7.200490e+10

choose(5, 2) #二項式係數
## [1] 10

7%/3 #商
## [1] 2

7%3 #餘數
## [1] 1

#####
##資料模式(mode)##
a<-3; class(a) #numeric(實數)
## [1] "numeric"

b<-3+2i; class(b) #complex(複數)
## [1] "complex"

c<-"hello!"; class(c) #character(文字)
## [1] "character"
```

```

d<-TRUE; class(d) #logical(邏輯值)
## [1] "logical"
e<-NA; class(e) #missing value(遺失值)
## [1] "logical"
#是不是一個...
x<-c("A", "B", NA, NaN)
is.numeric(x)
## [1] FALSE
is.complex(x)
## [1] FALSE
is.character(x)
## [1] TRUE
is.na(x)
## [1] FALSE FALSE TRUE FALSE
#模式轉換
x <- pi; x
## [1] 3.141593
mode(x)
## [1] "numeric"
as.numeric(x)
## [1] 3.141593
as.integer(x) #整數
## [1] 3
as.complex(x)
## [1] 3.141593+0i
as.character(x)
## [1] "3.14159265358979"
#####
##建構向量, 矩陣, data.frame##
#向量

```

```

a1 <- c(2, 5, 10) #建構一向量物件
a2 <- seq(from=0, to=1, by=0.2) #由0 到1 且間隔為0.2 的向量
a3 <- seq(from=1, by=10, length=5) #5 個由1 開始間隔為10..
aaa <- c(1, 2, 3, 4, 5, 6, 7, 8, 90, 1, 2, 345, 123)
length(aaa) #查看個數

## [1] 13

a4 <- rep(c(2, 3, 4), 1:3); a4 #這三個數分別重複1, 2, 3 次

## [1] 2 3 3 4 4 4

a5 <- "AB"; rep(a5, times=5); #times 可省略

## [1] "AB" "AB" "AB" "AB" "AB"

a5 <- "AB"; rep(a5, 5)

## [1] "AB" "AB" "AB" "AB" "AB"

a6 <- c("A", "B"); rep(a6, times=5)

## [1] "A" "B" "A" "B" "A" "B" "A" "B" "A" "B"

a7 <- c("A", "B"); rep(a7, each=5) #依序

## [1] "A" "A" "A" "A" "A" "B" "B" "B" "B" "B"

a8 <- c("C", "D"); rep(a8, 6)

## [1] "C" "D" "C" "D" "C" "D" "C" "D" "C" "D" "C" "D"

#矩陣
z <- matrix(c(1, 2, 3, 4, 5, 6), nrow=2, byrow=T); z

##      [,1] [,2] [,3]
## [1,]    1    2    3
## [2,]    4    5    6

#byrow=T 根據矩陣列的方向來排序，默認為欄位方向，如下。
z1 <- matrix(c(1, 2, 3, 4, 5, 6), nrow=2); z1

##      [,1] [,2] [,3]
## [1,]    1    3    5
## [2,]    2    4    6

z2 <- matrix(c(1, 2, 3, 4, 5, 6), ncol=2); z2

##      [,1] [,2]
## [1,]    1    4
## [2,]    2    5
## [3,]    3    6

```

```

z3 <- matrix(z1, nrow=3, ncol=2); z3

##      [,1] [,2]
## [1,]    1    4
## [2,]    2    5
## [3,]    3    6

#z2*%z3 #不能互乘
z2*%z1 #方正

##      [,1] [,2] [,3]
## [1,]    9   19   29
## [2,]   12   26   40
## [3,]   15   33   51

#Data.frame
x<-c("男","男","女","女","男","女","男","女","女","男")
y<-c(100, 90, 95, 85, 44, 86, 23, 67, 99, 33)
z<-c(55, 66, 78, 95, 33, 22, 15, 16, 33, 88)
q1 <- data.frame(x, y, z); q1

##      x  y  z
## 1  男 100 55
## 2  男  90 66
## 3  女  95 78
## 4  女  85 95
## 5  男  44 33
## 6  女  86 22
## 7  男  23 15
## 8  女  67 16
## 9  女  99 33
## 10 男  33 88

#cbind()
a12 <- cbind(x, y, z)
class(a12)

## [1] "matrix"

class(q1)

## [1] "data.frame"

#####
##邏輯值##
#關係比較
x <- seq(from=-1, to=1, by=1); x

## [1] -1  0  1

x<0 #(x 是否<0)

```

```

## [1] TRUE FALSE FALSE
x>0
## [1] FALSE FALSE TRUE
x==0 #(x 是否=0)
## [1] FALSE TRUE FALSE
x<=0
## [1] TRUE TRUE FALSE
x>=0
## [1] FALSE TRUE TRUE
x!=0 #(x 是否不等於0)
## [1] TRUE FALSE TRUE
#邏輯向量操作
x <- seq(from=0, to=5, by=1); x
## [1] 0 1 2 3 4 5
y <- seq(from=3, to=8, by=1); y
## [1] 3 4 5 6 7 8
(x>4)
## [1] FALSE FALSE FALSE FALSE FALSE TRUE
(y>4)
## [1] FALSE FALSE TRUE TRUE TRUE TRUE
(x>4)&(y>4) #&比較嚴格，兩個true 才是true。
## [1] FALSE FALSE FALSE FALSE FALSE TRUE
(x>4)|(y>4)
## [1] FALSE FALSE TRUE TRUE TRUE TRUE
!(x>4)
## [1] TRUE TRUE TRUE TRUE TRUE FALSE
#&適用於向量式的邏輯判斷，&&適用於單一值的邏輯判斷。|同理。
a13 <- c(10, 20, 30);a13
## [1] 10 20 30

```

```

a14 <- c(2, 22, 0); a14

## [1] 2 22 0

a13 > 4 && a14 > 10

## [1] FALSE

a13 > 4 || a14 > 10

## [1] TRUE

#(運算順序 !>&>|)
x <- 3; y<- -3
(x>2)&(y<2)|!(x+y!=3)

## [1] TRUE

#####
##if, switch, repeat 迴圈, for 迴圈, while 迴圈##
#if
x<-1
if(x==1){x}

## [1] 1

#switch{數值, 程式1, 程式2, ..., 程式N}
switch(1, 4, 1+1, "A")

## [1] 4

switch(2, 5, 2+2, "B")

## [1] 4

switch(3, 6, 3+3, "C")

## [1] "C"

#repeat
x<-0
#repeat{x<-x+1; print(x)} #無窮
repeat{
  if(x==10){break}
  x<-x+1; print(x)} #break 停止

## [1] 1
## [1] 2
## [1] 3
## [1] 4
## [1] 5
## [1] 6
## [1] 7

```

```
## [1] 8
## [1] 9
## [1] 10

#if 搭配for 迴圈
x<-c("男","男","女","女","男","女","男","女","女","男")
y<-c(100, 90, 95, 85, 44, 86, 23, 67, 99, 33)
z<-c(55, 66, 78, 95, 33, 22, 15, 16, 33, 88)
q1 <- data.frame(x, y, z); q1
```

```
##      x   y   z
## 1  男 100 55
## 2  男  90 66
## 3  女  95 78
## 4  女  85 95
## 5  男  44 33
## 6  女  86 22
## 7  男  23 15
## 8  女  67 16
## 9  女  99 33
## 10 男  33 88
```

```
q2 <- data.frame(性別=x, 數學=y, 國文=z); q2
```

```
##      性別 數學 國文
## 1     男  100   55
## 2     男   90   66
## 3     女   95   78
## 4     女   85   95
## 5     男   44   33
## 6     女   86   22
## 7     男   23   15
## 8     女   67   16
## 9     女   99   33
## 10    男   33   88
```

```
#男生全考0 分
for(i in 1:10)
{if(q2[i, 1]=="男"){
  q2[i, 3] <- 0}}
```

```
#女生考99
for(i in 1:10){
  if(q2[i, 1]=="男"){
    q2[i, 3] <- 0}
  else{q2[i, 3] <- 99}}
```



*#while 迴圈*

x <- 0

while(x<=10){print(x); x <- x+1}

## [1] 0

## [1] 1

## [1] 2

## [1] 3

## [1] 4

## [1] 5

## [1] 6

## [1] 7

## [1] 8

## [1] 9

## [1] 10

while(x<=10){x <- x+1; print(x)}

#####

*#其他*

rnorm(100) *#100 個呈常態分配的資料*

## [1] 1.313615102 0.129407076 0.898820018 0.821276337 0.0068890

91

## [6] 0.673266344 -0.525320763 1.732805732 0.490077109 -0.6765646

78

## [11] 0.533772622 0.575223581 -0.868745302 0.378583355 -1.3098669

33

## [16] -0.694994081 -1.439850883 0.197742120 -1.108716161 1.5581230

95

## [21] -0.256553905 0.128186475 -0.176678644 -0.794204252 0.0025061

76

## [26] 0.654553221 -1.988144258 -0.534172841 0.086347216 -1.2065872

82

## [31] -0.100061747 1.432049150 0.163338865 -0.324969186 2.0772775

25

## [36] -0.161452762 -0.788470480 1.803765394 -1.350604476 1.7498436

23

## [41] -0.610079981 -0.182890621 0.835156603 -0.588208895 -0.1639935

79

## [46] -0.047745518 0.602714621 0.184663256 1.656028509 0.4774063

99

## [51] -0.288440677 0.152515024 1.884296067 0.240922090 -1.5453790

07

## [56] 0.295385443 1.111865966 1.078128303 0.894361115 -0.4011420

86

## [61] -0.491725069 -0.833145490 -0.472275228 -0.073567543 0.2098311

35

## [66] -0.435273340 -1.057379237 1.092261737 0.482896184 -0.4273773

26

```
## [71] -0.595270335 -0.581199063 -0.039529198 -0.579851082 -1.3202818
23
## [76] -0.376255035 1.083076792 -1.313704477 -1.015280743 -0.1869518
34
## [81] -0.471802756 0.889710424 -0.167792521 -1.277508088 -1.3190725
60
## [86] -0.604153084 -0.777167899 0.960324121 -0.702630129 1.0415886
59
## [91] 2.181607820 0.865135470 0.389047384 1.234933888 0.6277367
45
## [96] -1.049231134 -1.014513873 0.428116363 -2.320621304 0.8647439
99
```

`runif(5)` #五個介於0與1的均勻分佈隨機變數

```
## [1] 0.8749218 0.9669750 0.6021622 0.4267016 0.7124219
```

`runif(5,2,10)` #產生五個介於2到10的均勻分佈隨機變數

```
## [1] 3.638573 6.191284 4.914379 7.422602 5.018611
```

`rnorm(100, 40, 10)` #100個平均40標準差為10

```
## [1] 55.35284 29.68762 32.65542 40.86514 44.48568 16.64900 40.71525
38.10948
## [9] 53.33645 57.64661 31.40907 53.43201 30.81161 39.00360 39.75428
34.44413
## [17] 35.29673 34.79317 36.27608 25.62810 54.30682 32.00220 60.97954
46.18403
## [25] 53.60030 60.35257 46.36974 39.14334 48.57126 28.97497 48.87637
40.08792
## [33] 30.38138 29.76351 45.83149 30.05909 25.39034 67.40555 57.05451
39.99845
## [41] 51.77881 44.79913 51.69400 55.47083 34.54789 56.94579 54.57341
44.97717
## [49] 27.87855 50.77207 40.42169 35.32258 27.99361 46.70726 14.00658
38.92224
## [57] 44.81954 47.46104 45.10859 41.79636 60.60210 40.07610 21.22495
32.16434
## [65] 44.90524 42.51285 35.49736 30.92087 35.94773 38.11155 39.08693
44.40325
## [73] 40.90557 46.27671 48.29559 28.58714 48.17640 28.13587 42.31087
39.94717
## [81] 38.69118 33.78708 36.15177 20.04013 47.17744 41.41384 29.95069
40.44731
## [89] 43.61828 21.32692 32.58896 52.97469 45.21546 33.82417 33.95454
53.08562
## [97] 29.43458 36.33598 41.29980 20.67727
```

`sample(6,3)` #sample(n,m)產生三個介於1到6的隨機整數

```
## [1] 5 4 2

sample(6,8,replace=TRUE) #8 個介於 1 到 6 的數字

## [1] 4 4 5 3 4 6 4 2

#sample(6,8)#錯誤，默認為不放回
#####
#套件
#library(AER)
#####
rm(list=ls()) #Removes all items in Environment!
##基本敘述統計##
# 次數計算
## set environment & import data
setwd("C:/Users/User/Desktop/R/R-project/R Statistics/data")
tw <- read.table("tw_marriage.txt") #會有狀況
tw <- read.table("tw_marriage.txt", header = T) #first row -> colname

# 彙整分組表格數據
xtabs(widowed ~ area, data=tw) #widowed(喪偶)依照區域分類計算人數

## area
##      中      北      外島      東      南
## 269693 465706 11557 38566 348013

xtabs(married ~ area + gender, data=tw) #已婚者依照區域與性別分組

##      gender
## area      F      M
## 中    1124686 1186626
## 北    2342430 2358321
## 外島    41091  48636
## 東     110083 120633
## 南    1366661 1429582

t(xtabs(married ~ area + gender, data=tw)) #轉置

##      area
## gender      中      北      外島      東      南
##      F 1124686 2342430 41091 110083 1366661
##      M 1186626 2358321 48636 120633 1429582

# 也可以如下
tw.xtabs <- xtabs(married ~ gender + area, data=tw) #更簡潔

margin.table(tw.xtabs,margin=1) #橫列邊際總和
```

```

## gender
##      F      M
## 4984951 5143798

margin.table(tw.xtabs,margin=2) #直行邊際總和

## area
##      中      北      外島      東      南
## 2311312 4700751  89727  230716 2796243

# 也可以
rowSums(xtabs(married ~ gender + area, data = tw)) #橫列總和

##      F      M
## 4984951 5143798

colMeans(xtabs(married ~ gender + area, data = tw))

##      中      北      外島      東      南
## 1155656.0 2350375.5  44863.5  115358.0 1398121.5

colSums(xtabs(married ~ gender + area, data = tw))

##      中      北      外島      東      南
## 2311312 4700751  89727  230716 2796243

#####
# tapply 可以計算各格的平均數等資料
# 各區未婚人士依照性別與區域分組的平均(因為一區域有包含很多縣市)
tapply(tw$unmarried, list(tw$gender,tw$area), FUN = mean)

##      中      北      外島      東      南
## F 129248.8 180047.9 7378.333 33620.5 116046.9
## M 159233.8 209691.1 9317.667 50041.5 144451.9

# 以下則是最大值
tapply(tw$unmarried, list(tw$gender,tw$area),FUN = max)

##      中      北      外島      東      南
## F 199040 543174 11091 41160 212854
## M 248357 637368 14441 59123 246285

# aggregate 可以達成類似 tapply 的效果，但輸出結果為每一組一列
# 可以用來做後續處理
aggregate(tw$unmarried, by=list(tw$gender, tw$area), FUN=mean)

##      Group.1 Group.2      x
## 1      F      中 129248.800
## 2      M      中 159233.800
## 3      F      北 180047.875
## 4      M      北 209691.125

```

```
## 5      F    外島    7378.333
## 6      M    外島    9317.667
## 7      F     東    33620.500
## 8      M     東    50041.500
## 9      F     南   116046.857
## 10     M     南   144451.857
```

*# prop.table 比例*

*# round 去小數後進位 prop.table 比例*

```
round(prop.table(xtabs(married ~ gender + area,data=tw)), 2) #小數點後2位
```

```
##      area
## gender 中  北 外島  東  南
##      F 0.11 0.23 0.00 0.01 0.13
##      M 0.12 0.23 0.00 0.01 0.14
```

```
round(prop.table(xtabs(married ~ gender + area, data=tw),
                    margin=1),2) #列總和100%
```

```
##      area
## gender 中  北 外島  東  南
##      F 0.23 0.47 0.01 0.02 0.27
##      M 0.23 0.46 0.01 0.02 0.28
```

```
round(prop.table(xtabs(married ~ gender + area, data=tw),
                    margin=2),2) #行總和100%
```

```
##      area
## gender 中  北 外島  東  南
##      F 0.49 0.50 0.46 0.48 0.49
##      M 0.51 0.50 0.54 0.52 0.51
```

*xtabs(married ~ city + gender , data=tw) #依縣市、性別分組*

```
##      gender
## city      F      M
## 宜蘭縣   98737 107117
## 花蓮縣   65982  71455
## 金門縣   18887  23376
## 南投縣  113063 126201
## 屏東縣  187474 202809
## 苗栗縣  119849 134653
## 桃園縣  422715 427139
## 高雄市  324075 323007
## 高雄縣  267598 283640
## 基隆市   79712  83714
## 連江縣   1839   2902
```

```
## 雲林縣 155479 184870
## 新竹市 90412 88596
## 新竹縣 111944 119320
## 嘉義市 58515 57474
## 嘉義縣 120412 139453
## 彰化縣 288130 311580
## 臺中市 230629 219307
## 臺中縣 337385 344668
## 臺北市 591588 578989
## 臺北縣 827473 818793
## 臺東縣 44101 49178
## 臺南市 163828 162218
## 臺南縣 244759 260981
## 澎湖縣 20365 22358
```

```
aggregate(tw$married, by=list(tw$city, tw$gender), FUN=mean) #無意義
```

```
##      Group.1 Group.2      x
## 1  宜蘭縣      F  98737
## 2  花蓮縣      F  65982
## 3  金門縣      F  18887
## 4  南投縣      F 113063
## 5  屏東縣      F 187474
## 6  苗栗縣      F 119849
## 7  桃園縣      F 422715
## 8  高雄市      F 324075
## 9  高雄縣      F 267598
## 10 基隆市      F  79712
## 11 連江縣      F   1839
## 12 雲林縣      F 155479
## 13 新竹市      F  90412
## 14 新竹縣      F 111944
## 15 嘉義市      F  58515
## 16 嘉義縣      F 120412
## 17 彰化縣      F 288130
## 18 臺中市      F 230629
## 19 臺中縣      F 337385
## 20 臺北市      F 591588
## 21 臺北縣      F 827473
## 22 臺東縣      F  44101
## 23 臺南市      F 163828
## 24 臺南縣      F 244759
## 25 澎湖縣      F  20365
## 26 宜蘭縣      M 107117
## 27 花蓮縣      M  71455
```

```
## 28 金門縣      M 23376
## 29 南投縣      M 126201
## 30 屏東縣      M 202809
## 31 苗栗縣      M 134653
## 32 桃園縣      M 427139
## 33 高雄市      M 323007
## 34 高雄縣      M 283640
## 35 基隆市      M 83714
## 36 連江縣      M 2902
## 37 雲林縣      M 184870
## 38 新竹市      M 88596
## 39 新竹縣      M 119320
## 40 嘉義市      M 57474
## 41 嘉義縣      M 139453
## 42 彰化縣      M 311580
## 43 臺中市      M 219307
## 44 臺中縣      M 344668
## 45 臺北市      M 578989
## 46 臺北縣      M 818793
## 47 臺東縣      M 49178
## 48 臺南市      M 162218
## 49 臺南縣      M 260981
## 50 澎湖縣      M 22358
```

```
#####
```

```
cards <- c("金卡", "白金卡", "普通卡", "普通卡", "普通卡",
           "金卡", "普通卡", "普通卡", "普通卡", "普通卡",
           "普通卡", "金卡", "普通卡", "白金卡", "金卡",
           "普通卡", "普通卡", "普通卡", "白金卡", "普通卡")
```

```
area <- c("南", "北", "南", "南", "北",
          "南", "南", "南", "南", "南",
          "北", "南", "南", "南", "南",
          "北", "北", "北", "北", "南")
```

```
gender <- c("女", "女", "女", "女", "女",
            "男", "女", "男", "男", "男",
            "男", "女", "女", "男", "女",
            "女", "男", "男", "男", "男")
```

```
x1 <- data.frame(cards, area, gender); x1
```

```
##      cards area gender
## 1   金卡   南    女
## 2  白金卡   北    女
## 3  普通卡   南    女
## 4  普通卡   南    女
```

```

## 5  普通卡  北  女
## 6  金卡  南  男
## 7  普通卡  南  女
## 8  普通卡  南  男
## 9  普通卡  南  男
## 10 普通卡  南  男
## 11 普通卡  北  男
## 12 金卡  南  女
## 13 普通卡  南  女
## 14 白金卡  南  男
## 15 金卡  南  女
## 16 普通卡  北  女
## 17 普通卡  北  男
## 18 普通卡  北  男
## 19 白金卡  北  男
## 20 普通卡  南  男

#想知道各卡片種類人數
#xtabs(cards~area, data=x1) #錯誤，因為都是文字串
tabulate(factor(x1$cards, levels = c("普通卡",
                                     "金卡", "白金卡"))))

## [1] 13  4  3

table(cards)

## cards
## 白金卡  金卡 普通卡
##      3      4     13

table(cards, dnn = "會員名稱")

## 會員名稱
## 白金卡  金卡 普通卡
##      3      4     13

x2 <- table(cards, gender); x2

##      gender
## cards  女  男
## 白金卡  1  2
## 金卡    3  1
## 普通卡  6  7

x3 <- table(cards, gender, dnn = c("會員分類", "性別")); x3

##      性別
## 會員分類 女  男

```



```
## 白金卡 1 2
## 金卡 3 1
## 普通卡 6 7
```

####練習題#### 求到小點第二位

##1. 計算 row, col 的邊際總和？

##2. 普通卡的會員佔全體會員多少比例？

##3. 白金卡的會員中男女各佔多少比例？

##4. 有多少比例的女生持有金卡？