Introduction Program R

R เป็นภาษาโปรแกรมทางคอมพิวเตอร์ใช้สำหรับคำนวณทางสถิติ คิดริเริ่มโดย Ross Ihaka and Robert Gentleman at the University of Auckland, New Zealand และปัจจุบันได้รับการดูแลและจัดการโดย R Development Core Team. R is freely available under the GNU General Public License, and pre-compiled binary versions are provided for various operating systems like Linux, Windows and Mac.

เอกสารคู่มือนี้อ้างอิงเนื้อหามาจาก

https://www.tutorialspoint.com/r/index.htm

http://www.r-tutor.com/r-introduction

การติดตั้ง

Download program r

www.google.com -----> program r download

https://www.r-project.org/ -----> CRAN mirror. (https://cran.r-project.org/mirrors.html)

Thailand http://mirrors.psu.ac.th/pub/cran/ -----> Download R for Windows ---> base ติดตั้ง double-click ---> NEXT

Basic Data Types

- Numeric - - -> Decimal values are called numerics in R (ตัวเลขที่มีทศนิยม)

> **a1 = 24.5** หรือ **a1 <- 24.5**

```
> a1 = 24.5
> a1
[1] 24.5
> class(a1)
[1] "numeric"
>
> a2 = 12
> a2
[1] 12
> class(a2)
[1] "numeric"
>
```

```
>
> is.integer(a2)
[1] FALSE
> is.numeric(a2)
[1] TRUE
>
```

- Integer - - -> จำนวนเต็มไม่คิดทศนิยม create an integer variable in R จะเรียกใช้ฟังก์ชันจำนวนเต็ม

```
> a3 = as.integer(5)
> a3
[1] 5
> class(a3)
[1] "integer"
> is.integer(a3)
[1] TRUE
>
> a4 = 6
> a4
[1] 6
> class(a4)
[1] "numeric"
> is.integer(a4)
[1] FALSE
>
>
```

```
> a4 = 6L
> class(a4)
[1] "integer"
> is.integer(a4)
[1] TRUE
>
> a3 = as.integer(3.25)
> a3
[1] 3
> a4 = as.integer(3.65)
> a4
[1] 3
> a4 = as.integer("3.75")
> a4
[1] 3
> is.integer(a4)
[1] TRUE
>
```

- Complex ---> A complex value in R is defined via the pure imaginary value i

```
> b1 = 3 + 5i
> b1
[1] 3+5i
> class(b1)
[1] "complex"
> is.complex(b1)
[1] TRUE
>

> sqrt(-1)
Warning in sqrt(-1) : NaNs produced
[1] NaN
> sqrt(-1+0i)
[1] 0+1i
> sqrt(as.complex(-1))
[1] 0+1i
>
```

- Logical --> A logical value is often created via comparison between variables.

```
> a = 3.25; b = as.integer(3.75)
> a
[1] 3.25
> b
[1] 3
> a > b
[1] TRUE
> a < b
[1] FALSE
> c = b < a
> c
[1] TRUE
> class(c)
[1] "logical"
> class(a)
[1] "numeric"
> class(b)
[1] "integer"
> 
> x = T; y = F
> x
[1] TRUE
> y
[1] FALSE
> x & y
[1] FALSE
> x | y
[1] TRUE
> !x
[1] FALSE
> 
```

- Character --> A character object is used to represent string values in R.

Convert objects into character values with the as.character()

function

```
> x1 = as.character(4.25)
> x1
[1] "4.25"
> class(x1)
[1] "character"
> is.character(x1)
[1] TRUE
> 
> name = "Chotitham"
> sname = "Thanrak"
> name
[1] "Chotitham"
> sname
[1] "Thanrak"
>
> paste(name,sname)
[1] "Chotitham Thanrak"
> 
```

Vector

A vector is a sequence of data elements of the same basic type. Members in a vector are officially called components

```
> c(2)
[1] 2
> c(2,1)
[1] 2 1
> c(2,1,9,8)
[1] 2 1 9 8
```

```
> a = c(2)
> b = c(2,1)
> c = c(2,1,9,8)
> a
[1] 2
> b
[1] 2 1
> c
[1] 2 1 9 8
>
```

- vector of logical values

```
> vl = c(T,F,T,T,F,F,F)
> vl
[1]  TRUE FALSE  TRUE  TRUE FALSE FALSE FALSE
>
```

- vector contain character strings

```
> vc = c("ID","Student","Nmae","SName")
> vc
[1] "ID"      "Student" "Nmae"    "SName"
>
```

- Number of members in a vector is given by the length function

```
> length(vl)
[1] 7
> length(vc)
[1] 4
> length(a)
[1] 1
> length(b)
[1] 2
> length(c)
[1] 4
>
```

- Combining Vectors --> Vectors can be combined via the function c

```
> a = c(2.25,3,5.56)
> b = c("id","student","name")
> c = c(T,F,T)
> d = c("F","T","F")
> c(a,a)
[1] 2.25 3.00 5.56 2.25 3.00 5.56
> c(a,b)
[1] "2.25"    "3"       "5.56"    "id"      "student" "name"
> c(a,c)
[1] 2.25 3.00 5.56 1.00 0.00 1.00
> c(a,d)
[1] "2.25" "3"    "5.56" "F"    "T"    "F"
> c(b,c)
[1] "id"      "student" "name"    "TRUE"    "FALSE"   "TRUE"
> c(a,c,d)
[1] "2.25"  "3"     "5.56"  "TRUE"  "FALSE" "TRUE"  "F"     "T"     "F"
>
```

- Vector Arithmetics --> Arithmetic operations of vectors are performed

member-by-member

```
> x = c(2,2.36,5.5,3,4)
> y = c(1,0,5,6.3,8)
> z = c("a","b","c","d","e")
> w = c(T,F,T,T,F)
>
> x + y
[1]  3.00  2.36 10.50  9.30 12.00
> x - y
[1]  1.00  2.36  0.50 -3.30 -4.00
> 2*y
[1]  2.0  0.0 10.0 12.6 16.0
> 3.5*x
[1]  7.00  8.26 19.25 10.50 14.00
> y/x
[1] 0.5000000 0.0000000 0.9090909 2.1000000 2.0000000
> x + z
Error in x + z : non-numeric argument to binary operator
> z + w
Error in z + w : non-numeric argument to binary operator
>
```

- Vector Index --> Retrieve values in a vector by declaring an index inside a single square bracket "[ ]" operator

```
> x[1]
[1] 2
> y[3]
[1] 5
> z[4]
[1] "d"
> x[-2]
[1] 2.0 5.5 3.0 4.0
> x[-1]
[1] 2.36 5.50 3.00 4.00
> x
[1] 2.00 2.36 5.50 3.00 4.00
> x[6]
[1] NA
>
```

Matrix

A matrix is a collection of data elements arranged in a two-dimensional rectangular layout

```
> U = matrix(c(3,5,4,7,5,6,-2,0,-4),nrow=3,ncol=3,byrow=T)
> U
     [,1] [,2] [,3]
[1,]    3    5    4
[2,]    7    5    6
[3,]   -2    0   -4
> 
```

- Element at the m row, n column of A can be accessed by the expression A[m, n].

```
> U[1,2]
[1] 5
> U[1,1]
[1] 3
> U[3,2]
[1] 0
> U[3,3]
[1] -4
> 
```

```
> U[2,]
[1] 7 5 6
> U[3,]
[1] -2  0 -4
> U[,3]
[1]  4  6 -4
> U[,1]
[1]  3  7 -2
> U[c(2,3)]
[1]  7 -2
> U[,c(2,3)]
     [,1] [,2]
[1,]    5    4
[2,]    5    6
[3,]    0   -4
> 
```

- Construct the transpose of a matrix by interchanging its columns and rows

with the function t

```
> U
     [,1] [,2] [,3]
[1,]    3    5    4
[2,]    7    5    6
[3,]   -2    0   -4
> t(U)
     [,1] [,2] [,3]
[1,]    3    7   -2
[2,]    5    5    0
[3,]    4    6   -4
>
```

- Columns of two matrices having the same number of rows can be combined into a larger matrix หรือ สร้างเมทริกซ์แต่งเติม

```
> U
     [,1] [,2] [,3]
[1,]    3    5    4
[2,]    7    5    6
[3,]   -2    0   -4
> B
     [,1]
[1,] -1.0
[2,]  0.5
[3,]  1.8
> cbind(U,B)
     [,1] [,2] [,3] [,4]
[1,]    3    5    4 -1.0
[2,]    7    5    6  0.5
[3,]   -2    0   -4  1.8
>
```

List

list is a generic vector containing other objects.

```
> x = c(2,2.36,5.5,3,4)
> y = c(1,0,5,6.3,8)
> z = c("a","b","c","d","e")
> w = c(T,F,T,T,F)
> A = list(x,y,z,w)
> A
[[1]]
[1] 2.00 2.36 5.50 3.00 4.00

[[2]]
[1] 1.0 0.0 5.0 6.3 8.0

[[3]]
[1] "a" "b" "c" "d" "e"

[[4]]
[1]  TRUE FALSE  TRUE  TRUE FALSE

>
```

- List Slicing --> retrieve a list slice with the single square bracket "[ ]" operator

```
> A[1]
[[1]]
[1] 2.00 2.36 5.50 3.00 4.00

> A[4]
[[1]]
[1]  TRUE FALSE  TRUE  TRUE FALSE

> A[5]
[[1]]
NULL
```

Data Frame

Data frame is used for storing data tables. It is a list of vectors of equal length.

```
> x = c(2,2.36,5.5,3,4)
> y = c(1,0,5,6.3,8)
> z = c("a","b","c","d","e")
> w = c(T,F,T,T,F)
> df = data.frame(x,y,z,w)
> df
     x   y z     w
1 2.00 1.0 a  TRUE
2 2.36 0.0 b FALSE
3 5.50 5.0 c  TRUE
4 3.00 6.3 d  TRUE
5 4.00 8.0 e FALSE
```

ตัวอย่างข้อมูล data frame ของ Program R

```
> mtcars
                   mpg cyl  disp  hp drat    wt  qsec vs am gear carb
Mazda RX4         21.0   6 160.0 110 3.90 2.620 16.46  0  1    4    4
Mazda RX4 Wag     21.0   6 160.0 110 3.90 2.875 17.02  0  1    4    4
Datsun 710        22.8   4 108.0  93 3.85 2.320 18.61  1  1    4    1
Hornet 4 Drive    21.4   6 258.0 110 3.08 3.215 19.44  1  0    3    1
Hornet Sportabout 18.7   8 360.0 175 3.15 3.440 17.02  0  0    3    2
Valiant           18.1   6 225.0 105 2.76 3.460 20.22  1  0    3    1
```

```
> mtcars["Merc 450SLC","cyl"]
[1] 8
```

- Data Import

Excel File (install gdata) --> library(gdata)

data = read.xls("filedata.xls")

data = read.xls(file.choose(),header=T)

Minitab File --> library(foreign); data = read.mtp("mydata.mtp")

data = read.mtp(file.choose())

SPSS File --> library(foreign)

data = read.spss("file", to.data.frame=TRUE)

Table File --> data = read.table("data.txt")

CSV File --> data = read.csv("mydata.csv")

ข้อมูลเชิงคุณภาพ (Qualitative Data)

ใช้ตัวอย่างข้อมูลในProgram R (R built-in data frame named painters)

ตัวอย่างคำสั่ง

```
> library(MASS)      # load the MASS package
> painters
```

	Composition	Drawing	Colour	Expression	School
Da Udine	10	8	16	3	A
Da Vinci	15	16	4	14	A
Del Piombo	8	13	16	7	A
Del Sarto	12	16	9	8	A
Fr. Penni	0	15	8	0	A
Guilio Romano	15	16	4	14	A
Michelangelo	8	17	4	8	A

```
> painters$School # เรียกดูข้อมูลในคอลัมภ์ School
```

แจกแจงความถี่ (Frequency Distribution of Qualitative Data) The frequency distribution of a data variable is a summary of the data occurrence in a collection of non-overlapping categories.

ตัวอย่างคำสั่ง

```
> s = painters$School
> s.fr = table(s)
> s.fr
s
 A  B  C  D  E  F  G  H
10  6  6 10  7  4  7  4
```

ใช้ cbind function to print the result in column format

```
>cbind(s.fr)
 s.fr
A  10
B   6
C   6
D  10
E   7
F   4
G   7
H   4
```

การแจกแจงความถี่สัมพัทธ์ (relative frequency distribution)

```
> s = painters$School ; s.fr = table(s)
> s.relfr = s.fr/nrow(painters)
> s.relfr
s
        A         B         C         D         E         F
0.18518519 0.11111111 0.11111111 0.18518519 0.12962963 0.07407407
        G         H
0.12962963 0.07407407
> old = options(digits=2)
> cbind(s.relfr)
 s.relfr
A  0.185
B  0.111
C  0.111
D  0.185
E  0.130
F  0.074
G  0.130
H  0.074
> options(old)
```

แสดงข้อมูลเป็นกราฟต่างๆ

Bar graph A bar graph of a qualitative data sample consists of vertical parallel bars that shows the frequency distribution graphically.

```
> library(gdata)
> data = read.xls(file.choose(),header=T)
> data
department input retire graduate normal location
1  health_sciA    47      6        0        41       B8
2  health_sciB    47     12        0        35       B9
3  chemA          35     15        0        20       B10
4  bioA           40     17        0        23       B10
5  homecoA        47     14        0        33       B52
6  envA           37      9        0        28       B8
7  envB           37     17        0        20       B9
8  phyA           27     13        0        14       B10
9  comsciA        29     12        0        17       B8
10  comsciB       30     16        0        14       B9
11   itA          33     11        0        22       B10
12   itB          33     11        0        22       B11

> locate=data$location
> locate.fr=table(locate)
> cbind(locate.fr)
 locate.fr
B10         4
B11         1
B52         1
B8          3
B9          3
> barplot(locate.fr)
> colors = c("red", "yellow", "green", "violet", "orange")
> barplot(locate.fr,col=colors)
```

Pie Chart

A pie chart of a qualitative data sample consists of pizza wedges that shows the frequency distribution graphicaally.

```
>pie(locate.fr)
> pie(locate.fr,col=colors)
```

Category Statistics

```
> data
> build= locate == "B10" # สร้าง logical index
> dep = data[build, ] # สร้าง data set ลูก
> dep
```

	department	input	retire	graduate	normal	location
3	chemA	35	15	0	20	B10
4	bioA	40	17	0	23	B10
8	phyA	27	13	0	14	B10
11	itA	33	11	0	22	B10

```
> mean(dep$input)
[1] 33.75
```

ข้อมูลเชิงปริมาณ (Quantitative Data) as continuous data, consists of numeric data that support arithmetic operations

```
> celsius = c(25,26,27,28,29,30)
> fahrenheit = (9/5)*celsius+32
> fahrenheit
[1] 77.0 78.8 80.6 82.4 84.2 86.0
>data.frame(Celsius=celsius, Fahrenheit=fahrenheit)
>result=data.frame(Celsius=celsius, Fahrenheit=fahrenheit)
>cbind(result)
```

แจกแจงความถี่ (Frequency Distribution of Quantitative Data)

สร้างตารางแจกแจงความถี่นักศึกษารับเข้า

```
>data
> duration = data$input
> range(duration)
[1] 27 47
> breaks = seq(25,50,by=5)
> breaks
[1] 25 30 35 40 45 50
> duration.cut = cut(duration,breaks,right=F)
> duration.fr = table(duration.cut)
> duration.fr
duration.cut
[25,30) [30,35) [35,40) [40,45) [45,50)
    2       3       3       1       3
>cbind(duration.fr)
```

แสดงข้อมูลด้วย Histogram

```
> duration
[1] 47 47 35 40 47 37 37 27 29 30 33 33
> hist(duration,right=F)
> hist(duration,right=F,col=colors,main="Student Input",xlab="Number of students")
```

การแจกแจงความถี่สัมพัทธ์ Relative Frequency Distribution of Quantitative Data

```
> duration = data$input
> breaks = seq(25,50,by=5)
> duration.cut = cut(duration,breaks,right=F)
> duration.fr = table(duration.cut)
> duration.relfr = duration.fr/nrow(data)
> duration.relfr
duration.cut
[25,30)     [30,35)     [35,40)     [40,45)     [45,50)
0.16666667 0.25000000 0.25000000 0.08333333 0.25000000
> old = options(digits=1)
> duration.relfr
duration.cut
[25,30) [30,35) [35,40) [40,45) [45,50)
  0.17    0.25    0.25    0.08    0.25
> cbind(duration.fr,duration.relfr)
duration.fr duration.relfr
[25,30)           2           0.17
[30,35)           3           0.25
[35,40)           3           0.25
[40,45)           1           0.08
[45,50)           3           0.25
> options(old)
```

การแจกแจงความถี่สะสม Cumulative Frequency Distribution

```
> duration.cumfr = cumsum(duration.fr)
> duration.cumfr
[25,30) [30,35) [35,40) [40,45) [45,50)
    2       5       8       9      12
> cbind(duration.cumfr)
duration.cumfr
[25,30)             2
[30,35)             5
[35,40)             8
[40,45)             9
[45,50)            12
```

กราฟความถี่สะสม Cumulative Frequency Graph

```
> cumfr = c(0, cumsum(duration.fr))
> plot(breaks, cumfr, main="Input student",
+ xlab="number of student",ylab="Cumulative input")
> lines(breaks,cumfr)
```

การแจกแจงความถี่สัมพัทธ์สะสม Cumulative Relative Frequency Distribution

```
> duration.cumrelfr = duration.cumfr/nrow(data)
> duration.cumrelfr
[25,30)   [30,35)   [35,40)   [40,45)   [45,50)
0.1666667 0.4166667 0.6666667 0.7500000 1.0000000
> old = options(digits=2)
> cbind(duration.cumfr,duration.cumrelfr)
duration.cumfr duration.cumrelfr
[25,30)             2            0.17
[30,35)             5            0.42
[35,40)             8            0.67
[40,45)             9            0.75
[45,50)            12             1.00
```

options(old)

กราฟความถี่สัมพัทธ์สะสม Cumulative Relative Frequency Graph

> plot(breaks, cumrelfr, main="Input student", xlab="number of student",ylab="Cumulative input")

> lines(breaks,cumrelfr)

มาตรเชิงตัวเลข Numerical Measures

ตัวอย่างข้อมูล

weight : 50,56,49,62,43,61,54,58,59,60

height : 1.75,1.65,1.50,1.52,1.63,1.80,1.90,1.78,1.67,1.52

bmi = weight/height^2

mean()

median()

quantile()

quantile(, c(.32, .57, .98))

max( ) – min( )

IQR( ) #The interquartile range

boxplot(, horizontal=TRUE)

var( )

sd( )

การแจกแจงความน่าจะเป็น Probability Distributions

- การแจกแจงแบบทวินาม Binomial Distribution

dbinom(x, size, prob)

pbinom(x, size, prob)

qbinom(p, size, prob)

dbinom(2,size=4,prob=.30) ความน่าจะเป็นแบบเฉพาะค่าเดี่ยวที่จะเกิดผลสำเร็จเท่ากับ 2 ครั้ง $\mathbf{x = 2}$ ในการทดลอง n = 4 ครั้ง และ ผลสำเร็จในแต่ละครั้งเท่ากับ 0.30 (p = 0.30) หรือ **P(X = 2)**

> dbinom(2,size=4,prob=.30)
[1] 0.2646

pbinom(2,size=4,prob=.30) สะสม ที่จะเกิดผลสำเร็จน้อยกว่าหรือเท่ากับ 2 ครั้ง $\mathbf{x \leq 2}$ ในการทดลอง n = 4 ครั้ง และ ผลสำเร็จในแต่ละครั้งเท่ากับ 0.30 (p = 0.30) หรือ $\mathbf{P(X \leq 2)}$= P(x=0) + P(x=1) + P(x=2)

> pbinom(2,size=4,prob=.30)
[1] 0.9163

> qbinom(0.9163,size=4,prob=.30)
[1] 2

- การแจกแจงแบบปกติ Normal Distribution

pnorm(x, mean, sd)

qnorm(p, mean, sd)

pnorm(10,mean=7,sd=5) หรือ > pnorm(z) ความน่าจะเป็นแบบของตัวแปรเชิงสุ่มแบบปกติ หรือ z = (x - mean)/sd

> pnorm(10,mean=7,sd=5)
[1] 0.7257469
> qnorm(0.72575,mean=7,sd=5)
[1] 10.00005

```
> pnorm(0.6)
[1] 0.7257469
> qnorm(0.72575)
[1] 0.6000094
```

- การแจกแจงแบบไคสแควร์ Chi-squared Distribution

pchisq(chi , df)
qchisq(p , df)

```
>pchisq(9.04,df=7)
[1] 0.750203
> qchisq(0.750203,df=7)
[1] 9.039999
```

- การแจกแจงแบบ t Student t Distribution
pt(t , df)
qt(p , df)

```
> pt(1.37,df=10)
[1] 0.8996706
> qt(0.8996706,df=10)
[1] 1.37
```

การทดสอบสมมติฐาน Hypothesis Testing

ตัวอย่างข้อมูลแบบ เวกเตอร์

ชุด1 10t

Pen=c(786.34,810.89,793.86,777.83,782.45,779.28,788.87,791.96,795.23,780.05)

HAProxy=c(741.08,744.96,743.70,747.97,742.93,741.93,748.29,744.23,744.12,743.97)

ชุด2 50t

Pen=c(1612.10,1610.02,1608.60,1609.69,1607.28,1606.75,1609.36,1608.70,1610.60,1611.72)

HAProxy=c(1611.56,1610.27,1606.48,1603.70,1603.01,1605.46,1600.69,1600.14,1603.95,1609.36)

t.test(Pen,HAProxy,paired=F, alternative='two.sided',conf.level=.95,var.equal=F)

ตัวอย่างข้อมูลแบบ เมทริกซ์