

ปฏิบัติการที่ 2

การใช้คำสั่งต่าง ๆ ในการจัดการข้อมูล

- วัตถุประสงค์:**
1. เพื่อให้นักศึกษาฝึกปฏิบัติใช้คำสั่ง และเทคนิคต่าง ๆ ในการจัดการข้อมูลด้วยโปรแกรม R
 2. เพื่อให้นักศึกษาฝึกปฏิบัติการแปลงข้อมูล และปรับโครงสร้างข้อมูล

Data file: “CBR64-3Prv_Practice.xlsx”, “ncdSuicide.xlsx”

สรุปเนื้อหา:

ในการสำรวจข้อมูลว่ามีข้อมูลซ้ำซ้อน มีค่าผิดปกติ ไม่สมบูรณ์ หรือค่าที่ผิดปกติหรือไม่ นอกจากจะสำรวจทีละตัวแปร บางครั้งอาจจะต้องสำรวจโดยใช้ 2 ตัวแปร หรือมากกว่า 2 ตัวแปร ในกรณีตัวแปรประเภทต่อเนื่องต้องมีการตรวจสอบการแจกแจงของข้อมูลเนื่องจากหนึ่งในข้อตกลง (assumption) ของสถิติพาราเมตริก คือ ข้อมูลมีการแจกแจงแบบปกติ เทคนิคที่ใช้ตรวจสอบ คือ ใช้กราฟฮิสโตแกรม หากตรวจสอบว่าข้อมูลที่ต้องการวิเคราะห์นั้นไม่มีการแจกแจงแบบปกติ ต้องมีการปรับข้อมูลด้วยเทคนิคทางคณิตศาสตร์ ถ้าให้ x แทนข้อมูล ถ้า x มีการแจกแจงแบบเบ้ขวา (Moderate skew) ให้ใช้เทคนิค \sqrt{x} ถ้า x มีการแจกแจงแบบเบ้ซ้าย (Moderate skew) ให้ใช้เทคนิค $\sqrt{\max(x+1) - x}$ ถ้า x มีการแจกแจงแบบเบ้ขวาค่อนข้างมาก (Greater skew) ให้ใช้เทคนิค $\log_{10}(x)$ ถ้า x มีการแจกแจงแบบเบ้ซ้ายค่อนข้างมาก (Greater skew) ให้ใช้เทคนิค $\log_{10}(\max(x+1) - x)$ ถ้า x มีการแจกแจงแบบเบ้ขวาอย่างเห็นได้ชัด (Severe skew) ให้ใช้เทคนิค $1/x$ และถ้า x มีการแจกแจงแบบเบ้ซ้ายอย่างเห็นได้ชัด (Severe skew) ให้ใช้เทคนิค $1/(\max(x+1) - x)$

อย่างไรก็ตามในการจัดการข้อมูลอย่างรวมถึงการปรับโครงสร้างของข้อมูลให้เหมาะสมกับวิธีการวิเคราะห์ข้อมูลอีกด้วยไม่ว่าจะเป็นการปรับโครงสร้างจากแนวคอลัมน์ให้เป็นแถว (Wide to long) หรือจากแนวแถวให้เป็นแนวคอลัมน์ (Long to wide) ทั้งนี้ขึ้นอยู่กับสถานการณ์ของการวิเคราะห์ข้อมูล แม้กระทั่งตัวแปร หรือ วิธีการทางสถิติที่ใช้ ก็อาจจะมีผลที่นักวิจัยต้องมีการปรับโครงสร้างของข้อมูลก่อนทำการวิเคราะห์ข้อมูล

1. ให้หาคำตอบของข้อต่อไปนี โดยใช้คำสั่ง if()

> x <- c(3, 12, 5, 89, 90, 147, 123, 4, 35, 44, 30, 25, 93, 78, 65, 63, 43, 98, 87)

a) นับจำนวนค่าของ x ว่ามีจำนวนเท่าไร ถ้าจำนวนที่นับได้มีค่าน้อยกว่า 25 ให้แสดงข้อความ "Number of x is less than 25"

คำตอบ (แสดงคำสั่งและผลลัพธ์)

```
> if(d<25){  
+ print("Number of x is less than 25")  
+ }  
[1] "Number of x is less than 25"
```

b) ถ้าจำนวนที่นับได้มีค่าน้อยกว่า 10 ให้แสดงข้อความ "Number of x is less than 10"

คำตอบ (แสดงคำสั่งและผลลัพธ์)

```
> if(d<10){  
+ print("Number of x is less than 10")  
+ }
```

c) แสดงข้อความ "Number of d is less than 100" **มิเช่นนั้น** ให้แสดงข้อความ "Number of d is more than 100"

คำตอบ (แสดงคำสั่งและผลลัพธ์)

```
> if(d<100){  
+ print("Number of d is less than 100")  
+ }else{  
+ print("Number of d is more than 100")  
+ }  
[1] "Number of d is less than 100"
```

d) ถ้าจำนวนที่นับได้มีค่าน้อยกว่า 22 ให้แสดงข้อความ "Number of x is less than 22"

คำตอบ (แสดงคำสั่งและผลลัพธ์)

```
> if(d<22){  
+ print("Number of x is less than 22")  
+ }  
[1] "Number of x is less than 22"
```

2. ให้หาคำตอบข้อต่อไปนี โดยใช้คำสั่ง if...else

> d <- rep(c(seq(1,76,2), seq(1,96,3)),2)

คำถาม	คำตอบ (เขียนคำอธิบาย)
a) ให้นักศึกษาอธิบายคำสั่ง rep()	ทำให้ค่าเป็นซ้ำๆโดยให้ทำกี่ครั้งก็ได้
b) ให้นักศึกษาอธิบายคำสั่ง seq()	สร้างตัวเลขโดยเรียงลำดับ
c) ให้นักเรียนจำนวนค่าของ d ว่ามีจำนวนเท่าไร	140

d) ถ้าจำนวนที่นับได้มีค่าน้อยกว่า 50 ให้แสดงข้อความ "Number of d is less than 50"

มิเช่นนั้นให้แสดงข้อความ "Number of d is more than 50"

คำตอบ (แสดงคำสั่งและผลลัพธ์)

```
> if(d<50){
+ print("Number of d is less than 50")
+ }else{
+ print("Number of d is more than 50")
+ }
[1] "Number of d is more than 50"
```

e) ถ้าจำนวนที่นับได้มีค่าน้อยกว่า 50 ให้แสดงข้อความ "Number of d is less than 50"

มิเช่นนั้น ถ้าจำนวนที่นับได้มีค่าน้อยกว่า100 ให้

คำตอบ (แสดงคำสั่งและผลลัพธ์)

```
> if(d<50){
+ print("Number of d is less than 50")
+ }else if(d<100){
+ print("Number of d is less than 100")
+ }
```

f) ถ้าจำนวนข้อมูล (จำนวนค่า) ของ d น้อยกว่า 50 ให้แสดงข้อความ "Number of d is less than 50"

มีเช่นนั้น ถ้าจำนวนข้อมูลของ d น้อยกว่า 100 ให้แสดงข้อความ "Number of d is less than 100"

มีเช่นนั้น ถ้าจำนวนข้อมูลของ d น้อยกว่า 150 ให้แสดงข้อความ "Number of d is less than 150"

มีเช่นนั้น ให้แสดงข้อความ "Number of d is more than 150"

คำตอบ (แสดงคำสั่งและผลลัพธ์)

```
> if(d<50){
+ print("Number of d is less than 50")
+ }else if(d<100){
+ print("Number of d is less than 100")
+ }else if(d<150){
+ print("Number of d is less than 150")
+ }else{
+ print("Number of d is more than 150")
+ }
[1] "Number of d is less than 150"
```

3. ให้หาคำตอบข้อต่อไปนี โดยใช้คำสั่ง ifelse()

```
> k <- c(145, 63.2, 58, 132.65, 87, 154, 132.50)
```

a) ถ้า k มีค่าน้อยกว่า 100 ให้ค่าเป็น 1 และถ้า k มีค่ามากกว่า 100 ให้ค่าเป็น 0

คำตอบ (แสดงคำสั่งและผลลัพธ์)

```
> ifelse(k<100,1,0)
[1] 0 1 1 0 1 0 0
```

b) ถ้า k มีค่าน้อยกว่า 100 ให้ค่าเป็น 1 และถ้า k มีค่ามากกว่า 100 ให้ค่าเป็น 0 โดยให้เก็บค่าใหม่นี้ในตัวแปรใหม่ คือ m

คำตอบ (แสดงคำสั่งและผลลัพธ์)

```
> ifelse(k<100,1,0)->m
> m
[1] 0 1 1 0 1 0 0
```

c) หาค่าเฉลี่ยของตัวแปร k เก็บไว้ในตัวแปร avg.k

คำตอบ (แสดงคำสั่งและผลลัพธ์)

```
> avg.k<- mean(k)
> avg.k
[1] 110.3357
```

d) ถ้า k มีค่าน้อยกว่า 100 ให้ลบด้วยค่าเฉลี่ยของ k ถ้า k มีค่ามากกว่า 100 ให้บวกด้วยค่าเฉลี่ยของ k

คำตอบ (แสดงคำสั่งและผลลัพธ์)

```
> ifelse(k<100,k-avg.k,k+avg.k)
[1] 255.33571 -47.13571 -52.33571 242.98571 -23.33571 264.33571
[7] 242.83571
```

4. ให้หาค่าสะสมตั้งแต่ 1 ถึง 100 โดยใช้คำสั่ง loop for

คำตอบ (แสดงคำสั่งและผลลัพธ์)

```
> sum <- 0
> for(i in c(1:100)){
+ sum <- sum+i
+ }
> sum
[1] 5050
```

5. ให้เขียนคำสั่ง หากตัวเลขใดต่อไปนี้เป็นเลขคี่ให้แสดงผลลัพธ์บนหน้าจอ โดยใช้คำสั่ง loop for

92, 33, 14, 38, 71, 32, 2, 35, 86, 82, 91, 23, 22, 39, 38, 57, 42, 9, 90, 96, 53, 91, 17, 81, 6, 49, 43

คำตอบ (แสดงคำสั่งและผลลัพธ์)

```
> sum<- 0
> for(i in c(1:h)){
+ if(num[i]%2==1){
+ sum <- sum+1
+ }
+ }
> sum
[1] 14
```

6. จากตัวเลขต่อไปนี้ ให้เขียนคำสั่งเพื่อนับจำนวนตัวเลขที่เป็นเลขคู่

92, 33, 14, 38, 71, 32, 2, 35, 86, 82, 91, 23, 22, 39, 38, 57, 42, 9, 90, 96, 53, 91, 17, 81, 6, 49, 43

คำตอบ (แสดงคำสั่งและผลลัพธ์)

```
> num<- c(92, 33, 14, 38, 71, 32, 2, 35, 86, 82, 91, 23, 22, 39, 38, 57, 42, 9, 90, 96, 53, 91, 17, 81, 6, 49, 43)
> h<- length(num)
> h
[1] 27
>
> sum<- 0
> for(i in c(1:h)){
+ if(num[i]%2==0){
+ sum <- sum+1
+ }
+ }
> sum
[1] 13
> |
```

7. ให้นักศึกษาเขียนฟังก์ชันเพื่อตรวจสอบว่าเลขที่ส่งไปเป็น “เลขคู่” หรือ “เลขคี่”

คำตอบ (แสดงคำสั่งและผลลัพธ์)

```
> myFunction <- function(num) {
+ if (num%2==0) {
+ print("Even Number")
+ }else{
+ print("Odd Number")
+ }
+ }
> myFunction(11)
[1] "Odd Number"
> myFunction(10)
[1] "Even Number"
. |
```

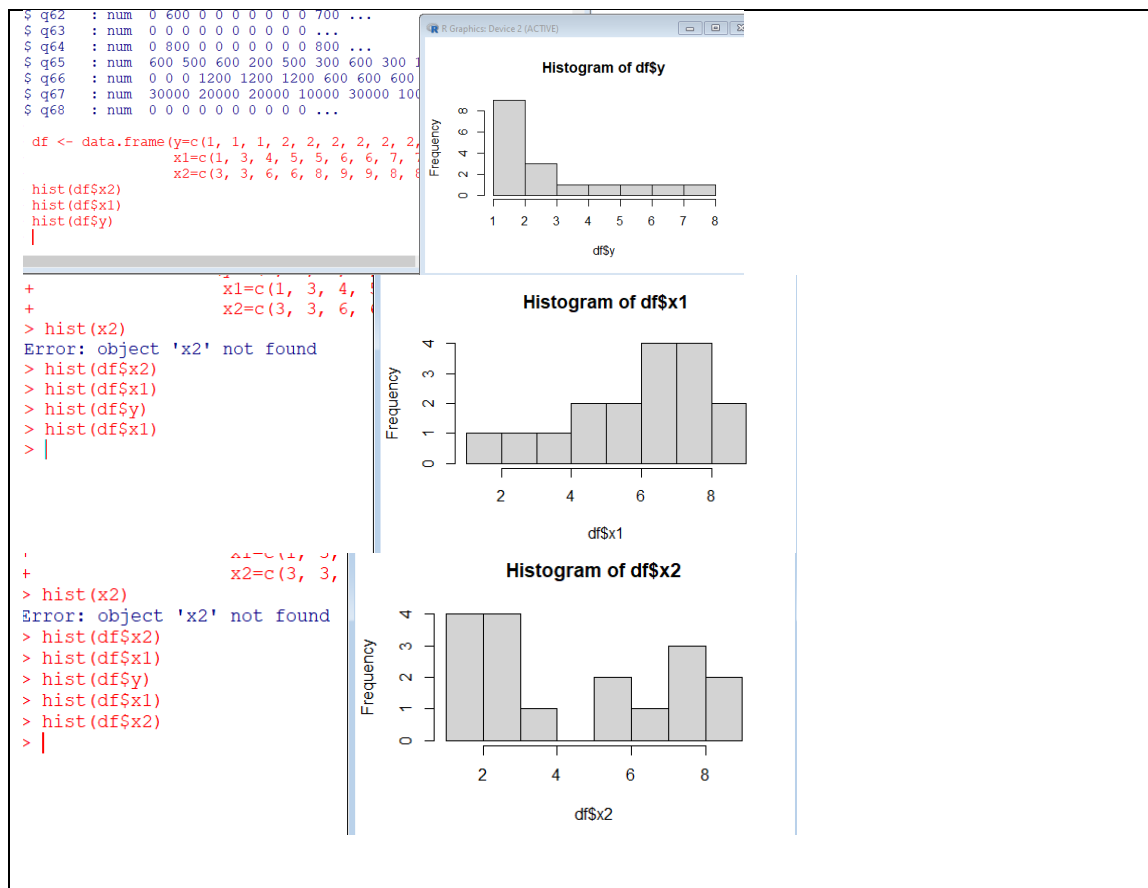
8. การแปลงข้อมูล (Transformation) จากข้อมูลไม่ได้มีการแจกแจงแบบปกติปรับข้อมูลให้มีการแจกแจงแบบปกติ

```
> df <- data.frame(y=c(1, 1, 1, 2, 2, 2, 2, 2, 2, 3, 3, 3, 4, 5, 6, 7, 8),
```

```
  x1=c(1, 3, 4, 5, 5, 6, 6, 7, 7, 7, 7, 8, 8, 8, 8, 9, 9),
```

```
  x2=c(3, 3, 6, 6, 8, 9, 9, 8, 8, 7, 4, 3, 3, 2, 2, 1, 1))
```

คำตอบ (แสดงคำสั่งและผลลัพธ์)



9. จาก data file “CBR64-3Prv_Practice.xlsx” ให้นักศึกษาฝึกปฏิบัติตรวจสอบการแจกแจงของข้อมูล และแปลงข้อมูลให้อยู่ในรูปของการแจกแจงแบบปกติ ข้อมูลประกอบด้วยตัวแปรดังต่อไปนี้

ตัวแปร
1.2 ความเกี่ยวข้องของโครงการ
1.3 คำนาคำหน้า
1.4 เพศ
1.5 อายุ
วันเดือนปีเกิด
1.6 สถานภาพ
จังหวัด
1.8 ระดับการศึกษาสูงสุด
1.9 อาชีพหลัก
3.1 ค่าจ้างแรงงาน หรือ เงินเดือน ของ "ครัวเรือน" เฉลี่ยเดือนละกี่บาท
5.3 รายเดือน หากท่านได้รายได้จากการเกษตรเป็นรายเดือน ครัวเรือนของท่านมีรายได้"เดือนละกี่บาท"
5.5 รายปี หากท่านได้รับรายได้จากการเกษตรเป็นรายปี ครัวเรือนของท่านมีรายได้"ปีละกี่บาท"
5.7 รายรอบ หากท่านมีรายได้จากการเกษตรเป็นรายรอบ ท่านมีรายได้ปีละกี่รอบ"
5.8 รายรอบ หากท่านได้รายได้จากการเกษตร ครัวเรือนของท่านมีรายได้ รอบละกี่บาท"
6.1 ท่านได้รับเงินบำนาญจากราชการ เดือนละกี่บาท
6.2 ครัวเรือนของท่านได้รับเบี้ยยังชีพผู้สูงอายุ เดือนละกี่บาท
6.3 ครัวเรือนของท่านได้รับเงินสำรองเลี้ยงชีพสำหรับผู้สูงอายุ เดือนละกี่บาท
6.4 ครัวเรือนของท่านได้รับเบี้ยความพิการ เดือนละกี่บาท
6.5 ครัวเรือนของท่านได้รับความช่วยเหลือจากบัตรสวัสดิการแห่งรัฐ เดือนละกี่บาท
6.6 ครัวเรือนของท่านได้รับความช่วยเหลือเงินสนับสนุนเด็กเล็ก เดือนละกี่บาท
6.7 ครัวเรือนของท่านได้รับความช่วยเหลือจากรัฐบาล ในสถานการณ์โควิด-19 ปีละกี่บาท
6.8 ครัวเรือนของท่านมีรายได้จากเงินช่วยเหลือจากคนอื่น่า หน่วยงานอื่น่า ที่ไม่ใช่รัฐบาลและสมาชิกในครอบครัว เดือนละกี่บาท

a) ให้นักศึกษาทำการเปลี่ยนชื่อตัวแปรดังนี้

ตัวแปร	เปลี่ยนเป็น
1.2 ความเกี่ยวข้องของโครงการ	stt
1.3 คำน่าหน้า	tit
1.4 เพศ	sex
1.5 อายุ	age
วันเดือนปีเกิด	brtdte
1.6 สถานภาพ	mar
จังหวัด	prv
1.8 ระดับการศึกษาสูงสุด	edu
1.9 อาชีพหลัก	occ
3.1 ค่าจ้างแรงงาน หรือ เงินเดือน ของ "ครัวเรือน"	sal
5.3 รายเดือน หากท่านได้รายได้จากการเกษตรเป็นรายเดือน	q53
5.5 รายปี หากท่านได้รับรายได้จากการเกษตรเป็นรายปี	q54
5.7 รายรอบ หากท่านมีรายได้จากการเกษตรเป็นรายรอบ	q57
5.8 รายรอบ หากท่านได้รายได้จากการเกษตร ครัวเรือน	q58
6.1 ท่านได้รับเงินบำนาญจากราชการ เดือนละกี่บาท	q61
6.2 ครัวเรือนของท่านได้รับเบี้ยยังชีพผู้สูงอายุ เดือนละกี่บาท	q62
6.3 ครัวเรือนของท่านได้รับเงินสำรองเลี้ยงชีพสำหรับตนเอง	q63
6.4 ครัวเรือนของท่านได้รับเบี้ยความพิการ เดือนละกี่บาท	q64
6.5 ครัวเรือนของท่านได้รับความช่วยเหลือจากมูลนิธิ	q65
6.6 ครัวเรือนของท่านได้รับความช่วยเหลือเงินสนับสนุน	q66
6.7 ครัวเรือนของท่านได้รับความช่วยเหลือจากรัฐบาล	q67
6.8 ครัวเรือนของท่านมีรายได้จากเงินช่วยเหลือจากครอบครัว	q68

คำตอบ (แสดงคำสั่งและผลลัพธ์)

```
> dt<- read_excel("CBR64-3Prv_Practice (1).xlsx")
> dt<- data.frame(dt)
> str(dt)
'data.frame': 203 obs. of 22 variables:
 $ X1.2. ความเกี่ยวข้องของโครงการ      $
 $ X1.3. คำน่าหน้า                      $
 $ X1.4. เพศ                          $
 $ X1.5. . อายุ                        $
 $ วันเดือนปีเกิด                      $
 $ X1.6. สถานภาพ                      $
 $ จังหวัด                            $
 $ X1.8. ระดับการศึกษาสูงสุด          $
 $ X1.9. อาชีพหลัก                    $
 $ X3.1. ค่าจ้างแรงงานหรือเงินเดือนของครัวเรือนเฉลี่ยเดือนละ$
 $ X5.3. รายเดือนหากท่านได้รายได้จากการเกษตรเป็นรายเดือนครัวเรือน$
 $ X5.5. รายปีหากท่านได้รายได้จากการเกษตรเป็นรายรอบครัวเรือน$
 $ X5.7. รายรอบหากท่านได้รายได้จากการเกษตรเป็นรายรอบครัวเรือน$
 $ X5.8. รายรอบหากท่านได้รายได้จากการเกษตรครัวเรือนของครัวเรือน$
 $ X6.1. ท่านได้รับเงินบำนาญจากราชการเดือนละบาท$
 $ X6.2. ครัวเรือนของท่านได้รับเบี้ยยังชีพผู้สูงอายุเดือนละบาท$
 $ X6.3. ครัวเรือนของท่านได้รับเงินสำรองเลี้ยงชีพสำหรับตนเอง$
 $ X6.4. ครัวเรือนของท่านได้รับความช่วยเหลือจากมูลนิธิ$
 $ X6.5. ครัวเรือนของท่านได้รับความช่วยเหลือเงินสนับสนุน$
 $ X6.6. ครัวเรือนของท่านได้รับความช่วยเหลือจากรัฐบาล$
 $ X6.7. ครัวเรือนของท่านได้รับความช่วยเหลือจากครอบครัว$
 $ X6.8. ครัวเรือนของท่านมีรายได้จากเงินช่วยเหลือจากครอบครัว$
> colnames(dt) <- c("stt",
+ "tit",
+ "sex",
+ "age",
+ "brtdte",
+ "mar",
```

```

+ "q63",
+ "q64",
+ "q65",
+ "q66",
+ "q67",
+ "q68"
+ )
> str(dt)
'data.frame': 203 obs. of 22 variables:
 $ stt : chr "1) กลุ่มเป้าหมาย" "1) กลุ่มเป้าหมาย" "1) กลุ่มเป้าหมาย" "$
 $ tit : chr "2) นางสาว" "2) นางสาว" "2) นางสาว" "2) นางสาว" "$
 $ sex : chr "2) หญิง" "2) หญิง" "2) หญิง" "2) หญิง" ...
 $ age : num 56 66 59 35 36 28 34 29 47 57 ...
 $ brtde: chr "26042508" "1012498" "17102505" "2112528" ...
 $ mar : chr "2) สมรส" "4) หม้าย" "4) หม้าย" "4) หม้าย" ...
 $ prv : chr "นครราชสีมา" "นครราชสีมา" "นครราชสีมา" "นครราชสีมา" ...
 $ edu : chr "1) ปริญญาตรี" "ไม่เรียนหนังสือ" "ไม่เรียนหนังสือ" "2) มัธยมศึกษา"
 $ occ : chr "กรรณิการ์ ปลูกพืชผักสวนครัว" "ทำสวน" "ทำสวน" "ทำสวน" ...
 $ sal : num 3500 3000 NA 3000 NA NA NA NA NA 2700 ...
 $ q53 : num 3000 NA NA NA 4000 NA NA 9000 4000 2000 ...
 $ q54 : num NA 5000 NA NA NA NA NA NA NA NA ...
 $ q57 : num NA NA 24 NA NA 10 22 NA NA NA ...
 $ q58 : num NA NA 300 NA NA 1900 2000 NA NA NA ...
 $ q61 : num 0 0 0 0 0 0 0 0 0 0 ...
 $ q62 : num 0 600 0 0 0 0 0 0 0 700 ...
 $ q63 : num 0 0 0 0 0 0 0 0 0 0 ...
 $ q64 : num 0 800 0 0 0 0 0 0 0 800 ...
 $ q65 : num 600 500 600 200 500 300 600 300 1000 600 ...
 $ q66 : num 0 0 0 1200 1200 1200 600 600 600 0 ...
 $ q67 : num 30000 20000 20000 10000 30000 10000 10000 20000 30000
 $ q68 : num 0 0 0 0 0 0 0 0 0 0 ...

```

- b) ให้ตรวจสอบการแจกแจงของข้อมูลตัวแปร sal ด้วยกราฟฮิสโตแกรม หากข้อมูลไม่มีการแจกแจงแบบปกติให้ทำการแปลงข้อมูลเพื่อให้ข้อมูลมีการแจกแจงแบบปกติ

คำตอบ (แสดงคำสั่งและผลลัพธ์)

```

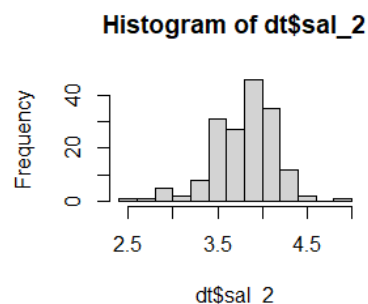
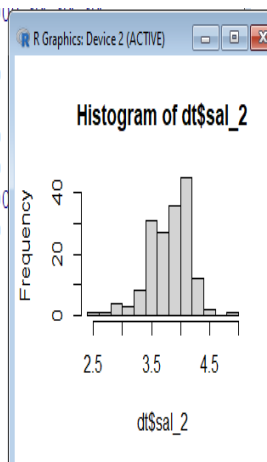
$ q58 : num NA NA 300 NA NA 1900 2000 ...
$ q61 : num 0 0 0 0 0 0 0 0 0 0 ...
$ q62 : num 0 600 0 0 0 0 0 0 0 700 ...
$ q63 : num 0 0 0 0 0 0 0 0 0 0 ...
$ q64 : num 0 800 0 0 0 0 0 0 0 800 ...
$ q65 : num 600 500 600 200 500 300 600 300 1000 600 ...
$ q66 : num 0 0 0 1200 1200 1200 600 600 600 0 ...
$ q67 : num 30000 20000 20000 10000 30000 10000 10000 20000 30000
$ q68 : num 0 0 0 0 0 0 0 0 0 0 ...

```

```

> hist(dt$sal)
> dt$sal_1<- sqrt(dt$sal)
> hist(dt$sal_1)
> dt$sal_2<- log10(dt$sal+0.001)
> hist(dt$sal_2)
> |

```



เขียนคำอธิบาย

ข้อมูลตัวแปร sal มีการแจกแจงแบบปกติโดยการแปลงข้อมูลโดย Log10

c) ให้จัดการข้อมูลรายได้ดังนี้

- กำหนดหาผลคูณของรายได้ที่เป็นรอบตัวแปร q57xq58 ตั้งชื่อตัวแปรใหม่เป็น q5ro_total
- กำหนดหาผลรวมของรายได้ที่เป็นรายเดือนทั้งหมด ได้แก่ ตัวแปร sal, q53 q61, q62, q63, q64, q65, q66, q68 ตั้งชื่อตัวแปรใหม่เป็น qinc_mtotal
- กำหนดรายได้จากรายเดือนเป็นรายปี qinc_mtotal x 12 ตั้งชื่อตัวแปรใหม่เป็น qinc_ytotal (qinc_ytotal= qinc_mtotal x 12)
- กำหนดหาผลรวมรายได้รายปี q5ro_total+ qinc_ytotal+ q54+ q67 ตั้งชื่อตัวแปรใหม่เป็น inctotal

คำตอบ (แสดงคำสั่งและบางส่วนของผลลัพธ์)

```
$ sal      : num  3500 3000 NA 3000 NA NA NA NA 2700 ...
$ q53      : num  3000 NA NA NA 4000 NA NA 9000 4000 2000 ...
$ q54      : num  NA 5000 NA NA NA NA NA NA NA ...
$ q57      : num  NA NA 24 NA NA 10 22 NA NA NA ...
$ q58      : num  NA NA 300 NA NA 1900 2000 NA NA NA ...
$ q61      : num  0 0 0 0 0 0 0 0 0 0 ...
$ q62      : num  0 600 0 0 0 0 0 0 0 700 ...
$ q63      : num  0 0 0 0 0 0 0 0 0 0 ...
$ q64      : num  0 800 0 0 0 0 0 0 0 800 ...
$ q65      : num  600 500 600 200 500 300 600 300 1000 600 ...
$ q66      : num  0 0 0 1200 1200 1200 600 600 600 0 ...
$ q67      : num  30000 20000 20000 10000 30000 10000 10000 20000 ...
$ q68      : num  0 0 0 0 0 0 0 0 0 0 ...
$ sal_1    : num  59.2 54.8 NA 54.8 NA ...
$ sal_2    : num  3.54 3.48 NA 3.48 NA ...
$ sal_3    : num  0.000286 0.000333 NA 0.000333 NA ...
$ q5ro_total : num  NA NA 7200 NA NA 19000 44000 NA NA NA ...
$ qinc_mtotal : num  7100 4900 600 4400 5700 1500 1200 9900 5600 ...
$ qinc_ytotal : num  85200 58800 7200 52800 68400 ...
$ inctotal  : num  115200 83800 34400 62800 98400 ...
> |
```

```
dt$sal_3<- 1/(dt$sal)
hist(dt$sal_3)

log10
log10(0)

dt$q5ro_total <- dt$q57*dt$q58
str(dt)

dt$qinc_mtotal <- rowSums(dt[,c("sal","q53","q61","q62","q63","q64","q65","q66","q68")],na.rm=T)
str(dt)

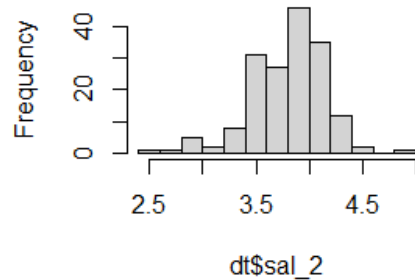
dt$qinc_ytotal<- dt$qinc_mtotal*12
str(dt)

dt$inctotal<- rowSums(dt[,c("q5ro_total","qinc_ytotal","q54","q67")],na.rm=T)
str(dt)
```

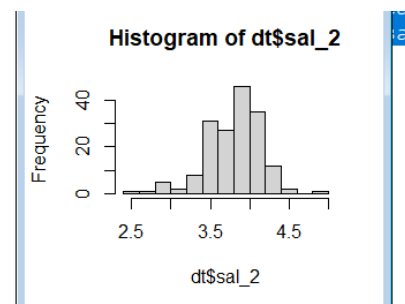

- d) ให้ตรวจสอบการแจกแจงของข้อมูลตัวแปร `inctotal` ด้วยกราฟฮิสโตแกรม หากข้อมูลไม่มีการแจกแจงแบบปกติให้ทำการแปลงข้อมูลเพื่อให้ข้อมูลมีการแจกแจงแบบปกติ

คำตอบ (แสดงคำสั่งและผลลัพธ์)

Histogram of dt\$sal_2



```
> dt$inctotal <- sqrt(dt$inctotal)
> hist(dt$sal)
> dt$inctotal <- log10(dt$inctotal)
> hist(dt$sal)
> dt$inctotal_1 <- log10(dt$inctotal)
> hist(dt$sal_1)
> dt$inctotal_2 <- 1/(dt$inctotal)
Error: object '1' not found
> hist(dt$sal_2)
> dt$inctotal_1 <- log10(dt$inctotal+0.001)
> hist(dt$sal_1)
> dt$inctotal_2 <- 1/(dt$inctotal)
Error: object '1' not found
```



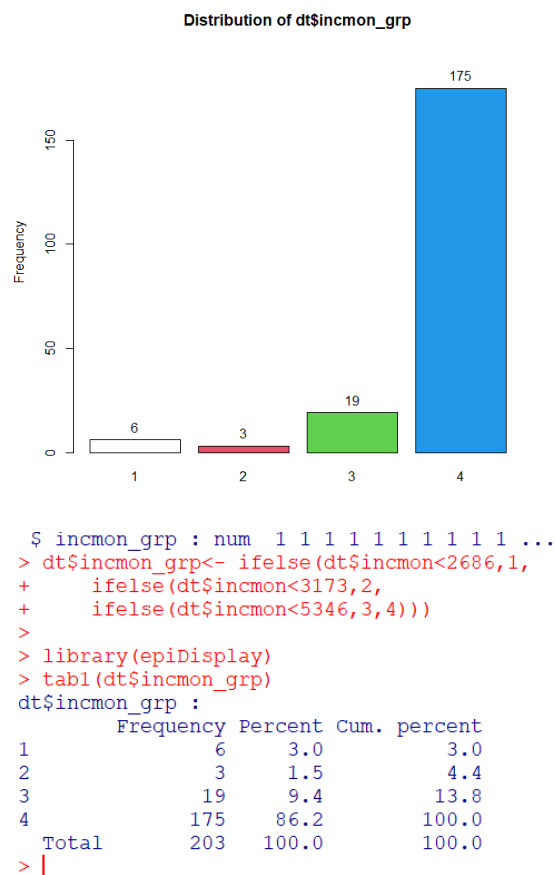
เขียนคำอธิบาย

ข้อมูลตัวแปร `sal` มีการแจกแจงแบบปกติโดยการแปลงข้อมูลโดย `1/`

- e) จัดข้อมูลรายได้ต่อปี `inctotal` ให้เป็นรายได้ต่อเดือนดังนี้
- `incmon = inctotal/12` (`incmon` คือ ข้อมูลรายได้ต่อเดือน)
 - จัดกลุ่มรายได้ต่อเดือน ดังนี้ โดยตั้งชื่อตัวแปรใหม่เป็น `incmon_grp`
 1. $< 2,686$ --> ต่ำกว่าเส้นยากจน

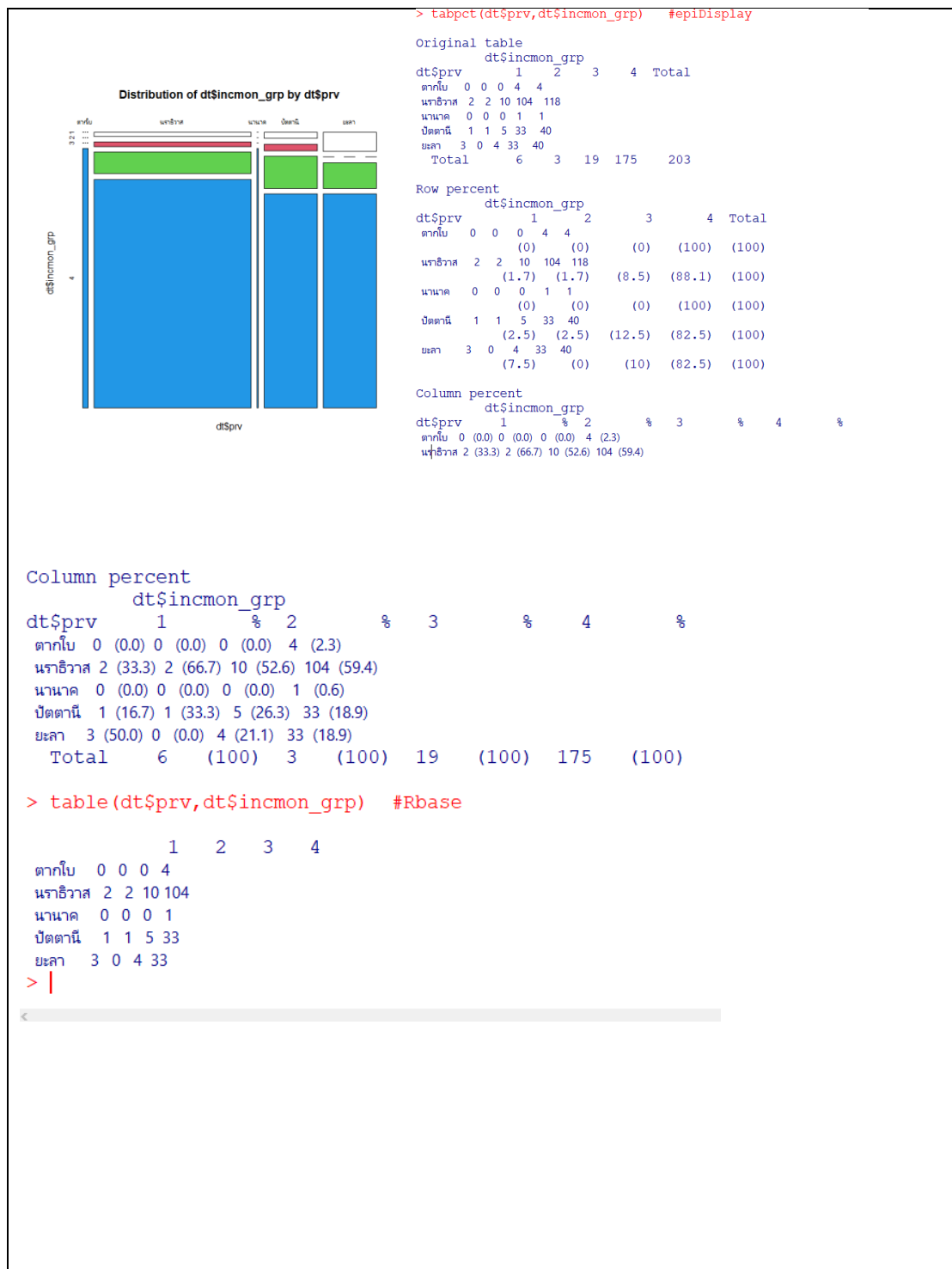
2. $< 3,173$ --> ต่ำกว่า 40% ล่าง
3. $< 5,346$ --> ต่ำกว่าเส้นมัธยฐาน
4. $\geq 5,346$ --> สูงกว่ามัธยฐาน

คำตอบ (แสดงคำสั่งและบางส่วนของผลลัพธ์)



- f) ให้สำรวจข้อมูลกลุ่มตัวอย่างแต่ละจังหวัดมีกลุ่มรายได้ต่อเดือนเป็นอย่างไรบ้าง

คำตอบ (แสดงคำสั่งและผลลัพธ์)



10. จาก data file “ncdSuicide.xlsx” ให้นักศึกษาฝึกปฏิบัติการจัดการข้อมูล

Country	Year	Probability (Crude suicid		
		Both sexes	Male	Female	Both sexes	Male	Female
Afghanistan	2016	29.8	31.8	27.7	4.7	7.6	1.5
Afghanistan	2015	29.8	31.9	27.8	4.8	7.8	1.5
Afghanistan	2010	31.7	34.1	29.4	5.1	8.6	1.4
Afghanistan	2005	34.1	36.5	31.6	6.3	10.8	1.5
Afghanistan	2000	34.4	36.6	32.1	5.7	10	1
Albania	2016	17	20.7	13.1	6.3	7.9	4.7
Albania	2015	17.4	20.9	13.7	6	7.6	4.4
Albania	2010	18.6	22.3	14.6	7.8	9.5	6.1
Albania	2005	18.8	22.1	15	6.7	7.7	5.7
Albania	2000	19.4	23.1	14.9	5.5	7.4	3.5
Algeria	2016	14.2	15.2	13.2	3.2	4.7	1.7
Algeria	2015	14.4	15.4	13.3	3.2	4.8	1.7
Algeria	2010	15.4	16.4	14.4	3.3	4.9	1.8
Algeria	2005	17.6	18.9	16.2	3.8	5.3	2.2
Algeria	2000	21.1	23.1	19.1	4.1	5.6	2.5
Angola	2016	16.5	17.2	16	4.7	7	2.5
Angola	2015	16.9	17.6	16.3	5	7.5	2.6
Angola	2010	18.1	18.6	17.8	5.7	8.7	2.8
Angola	2005	21.3	21.5	21.2	7.2	11	3.7
Angola	2000	24	24.2	23.8	7.9	11.9	4

a) จากชุดข้อมูลต้นมีจำนวนกี่ประเทศ

คำตอบ (แสดงคำสั่งและผลลัพธ์)

```
> n_c <- length(unique(dt$Country))
> n_c
[1] 184
> |
```

```
setwd("E:\\Prince of Songkla University Pattani Campus In Thai")
library(readxl)
n_c <- length(unique(dt$Country))
n_c
```

b) แต่ละประเทศบันทึกปีเหมือนกันหรือแตกต่างกันอย่างไร

คำตอบ (แสดงคำสั่งและผลลัพธ์)

```
> n_c <- unique(dt$Country)
> for (country in n_c) {
+   data_country <- dt[dt$Country == country, "Year"]
+   if (length(unique(data_country)) == 1) {
+     cat(country, "has consistent years:", unique(data_country), "\n")
+   } else {
+     cat(country, "has different years:", paste(unique(data_country), collapse = ", ")
+   }
+ }
Country has different years:
Afghanistan has different years:
Albania has different years:
Algeria has different years:
Angola has different years:
Antigua and Barbuda has different years:
Argentina has different years:
Armenia has different years:
Australia has different years:
Austria has different years:
Azerbaijan has different years:
Bahamas has different years:
Bahrain has different years:
Bangladesh has different years:
Barbados has different years:
Belarus has different years:
Belgium has different years:
Belize has different years:
Benin has different years:
Bhutan has different years:
Bolivia (Plurinational State of) has different years:
```

```
setwd("E:\\Prince of Songkla University Pattani Campus In Thai")
library(readxl)
dt <- read_excel("ncdSuicide.xlsx")
dt <- data.frame(dt)
str(dt)
str(dt)
colnames(dt) <- c("Country", "Year", "Both Sexes", "Male", "Female")
str(dt)
n_c <- length(unique(dt$Country))
n_c
n_c <- unique(dt$Country)
n_c
n_c <- unique(dt$Country)
for (country in n_c) {
  data_country <- dt[dt$Country == country, "Year"]
  if (length(unique(data_country)) == 1) {
    cat(country, "has consistent years:", unique(data_country), "\n")
  } else {
    cat(country, "has different years:", paste(unique(data_country), collapse = ", ")
  }
}
```

c) จากคอลัมน์ Crude suicide rates (per 100 000 population) ประกอบด้วยข้อมูล 3 ตัวแปร ดังนี้

Both sexes, Male, and Female ให้นักศึกษาฝึกจัดการข้อมูลจาก wide to long

ตัวอย่างข้อมูลหลังจากจัดการข้อมูลแล้ว

Country	Year	Gender	SuicideRate
Afghanistan	2016	Both sexes	4.7
Afghanistan	2016	Male	7.6
Afghanistan	2016	Female	1.5
Afghanistan	2015	Both sexes	4.8
Afghanistan	2015	Male	7.8
Afghanistan	2015	Female	1.5
...

คำตอบ (แสดงคำสั่งและบางส่วนของผลลัพธ์)

```
> long_data <- arrange(
+   mutate(
+     pivot_longer(
+       data,
+       cols = starts_with("Crude_suicide_rates"),
+       names_to = "Gender",
+       values_to = "SuicideRate"
+     ),
+     Gender = sub("Crude_suicide_rates_", "", Gender)
+   ),
+   Country, Year, Gender
+ )
>
> print("Data in Long Format:")
[1] "Data in Long Format:"
> print(long_data)
# A tibble: 6 x 4
  Country    Year Gender    SuicideRate
  <chr>      <dbl> <chr>      <dbl>
1 Afghanistan 2015 Both sexes    4.8
2 Afghanistan 2015 Female      1.5
3 Afghanistan 2015 Male       7.8
4 Afghanistan 2016 Both sexes    4.7
5 Afghanistan 2016 Female      1.5
```

```
names_to = "Gender",
values_to = "SuicideRate"
) %>%
mutate(Gender = sub("Crude_suicide_rates_", "", Gender),
       Country, Year, Gender)

print("Data in Long Format:")
print(long_data)

long_data <- arrange(
  mutate(
    pivot_longer(
      data,
      cols = starts_with("Crude_suicide_rates"),
      names_to = "Gender",
      values_to = "SuicideRate"
    ),
    Gender = sub("Crude_suicide_rates_", "", Gender)
  ),
  Country, Year, Gender
)

print("Data in Long Format:")
```

- d) ให้นักศึกษาดัดข้อมูลที่ Gender มีค่าเป็น Both sexes จากนั้นให้สร้างกราฟฮิสโตแกรมของ Crude suicide rates (per 100 000 population) หากข้อมูลไม่มีการแจกแจงแบบปกติให้ทำการแปลงข้อมูลให้อยู่ในรูปการแจกแจงแบบปกติ

คำตอบ (แสดงคำสั่งและผลลัพธ์)

```
> ggplot(long_data, aes(x = SuicideRate)) +
+   geom_histogram(binwidth = 1, fill = "skyblue", color = "black") +
+   labs(title = "Histogram of Suicide Rates", x = "Suicide Rate", y = "Frequency")
Error in ggplot(long_data, aes(x = SuicideRate)) :
could not find function "ggplot"
>
> shapiro_test_result <- shapiro.test(long_data$SuicideRate)
> print(paste("Shapiro-Wilk test p-value:", shapiro_test_result$p.value))
[1] "Shapiro-Wilk test p-value: 0.208586995912467"
>
> if (shapiro_test_result$p.value < 0.05) {
+   cat("Data is not normally distributed. Applying log transformation.\n")
+   long_data$TransformedRate <- log(long_data$SuicideRate + 1)
+   ggplot(long_data, aes(x = TransformedRate)) +
+     geom_histogram(binwidth = 0.2, fill = "skyblue", color = "black") +
+     labs(title = "Histogram of Transformed Suicide Rates", x = "Log Transformed Suicide Rate", y = "Frequency")
+   shapiro_test_transformed <- shapiro.test(long_data$TransformedRate)
+   print(paste("Shapiro-Wilk test p-value for transformed data:", shapiro_test_transformed$p.value))
+   } else {
+   cat("Data is normally distributed. No transformation needed.\n")
+   }
Data is normally distributed. No transformation needed.
>
> |
```

```
Country, Year, Gender
)

print("Data in Long Format:")
print(long_data)

shapiro_test_result <- shapiro.test(long_data$SuicideRate)
print(paste("Shapiro-Wilk test p-value:", shapiro_test_result$p.value))

if (shapiro_test_result$p.value < 0.05) {
  cat("Data is not normally distributed. Applying log transformation.\n")
  long_data$TransformedRate <- log(long_data$SuicideRate + 1)
  ggplot(long_data, aes(x = TransformedRate)) +
    geom_histogram(binwidth = 0.2, fill = "skyblue", color = "black") +
    labs(title = "Histogram of Transformed Suicide Rates", x = "Log Transformed Suicide Rate", y = "Frequency")
  shapiro_test_transformed <- shapiro.test(long_data$TransformedRate)
  print(paste("Shapiro-Wilk test p-value for transformed data:", shapiro_test_transformed$p.value))
  } else {
  cat("Data is normally distributed. No transformation needed.\n")
  }
}
```

