**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 มีค่าผิดพลาด 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 log10(x) ถ้า x มีการแจกแจงแบบเบ้ซ้ายค่อนข้างมาก (Greater skew) ให้ใช้เทคนิค log10(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. ให้หาคำตอบของข้อต่อไปนี้ โดยใช้คำสั่ง if()

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

1. นับจำนวนค่าของ x ว่ามีจำนวนเท่าไหร่ ถ้าจำนวนที่นับได้มีค่าน้อยกว่า 25 ให้แสดงข้อความ “Number of x is less than 25”

|  |
| --- |
| คำตอบ (แสดงคำสั่งและผลลัพธ์) |

1. ถ้าจำนวนที่นับได้มีค่าน้อยกว่า 10 ให้แสดงข้อความ “Number of x is less than 10”

|  |
| --- |
| คำตอบ (แสดงคำสั่งและผลลัพธ์) |

1. แสดงข้อความ “Number of d is less than 100” **มิเช่นนั้น** ให้แสดงข้อความ “Number of d is more than 100”

|  |
| --- |
| คำตอบ (แสดงคำสั่งและผลลัพธ์) |

1. ถ้าจำนวนที่นับได้มีค่าน้อยกว่า 22 ให้แสดงข้อความ “Number of x is less than 22”

|  |
| --- |
| คำตอบ (แสดงคำสั่งและผลลัพธ์) |

1. ให้หาคำตอบข้อต่อไปนี้ โดยใช้คำสั่ง if...else

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

| คำถาม | คำตอบ (เขียนคำอธิบาย) |
| --- | --- |
| 1. ให้นักศึกษาอธิบายคำสั่ง rep() | ทำให้ค่าเป็นซ่ำๆโดยให้ทำกี่ครั้งก็ได้ |
| 1. ให้นักศึกษาอธิบายคำสั่ง seq() | สร้างตัวเล่ขโดยเรียกลำดับ |
| 1. ให้นับจำนวนค่าของ d ว่ามีจำนวนเท่าไหร่ | 140 |

1. **ถ้า**จำนวนที่นับได้มีค่าน้อยกว่า 50 ให้แสดงข้อความ “Number of d is less than 50”   
   **มิเช่น**นั้นให้แสดงข้อความ “Number of d is more than 50”

|  |
| --- |
| คำตอบ (แสดงคำสั่งและผลลัพธ์) |

1. **ถ้า**จำนวนที่นับได้มีค่าน้อยกว่า 50 ให้แสดงข้อความ “Number of d is less than 50”   
   **มิเช่นนั้น** ถ้าจำนวนที่นับได้มีค่าน้อยกว่า100 ให้

|  |
| --- |
| คำตอบ (แสดงคำสั่งและผลลัพธ์) |

1. **ถ้า**จำนวนข้อมูล (จำนวนค่า) ของ d น้อยกว่า 50 ให้แสดงข้อความ “Number of d is less than 50”   
   **มิเช่นนั้น** ถ้าจำนวนข้อมูลของ d น้อยกว่า 100 ให้แสดงข้อความ “Number of d is less than 100”  
   **มิเช่นนั้น** ถ้าจำนวนข้อมูลของ d น้อยกว่า 150 ให้แสดงข้อความ “Number of d is less than 150”  
   **มิเช่นนั้น** ให้แสดงข้อความ “Number of d is more than 150”

|  |
| --- |
| คำตอบ (แสดงคำสั่งและผลลัพธ์) |

1. ให้หาคำตอบข้อต่อไปนี้ โดยใช้คำสั่ง ifelse()

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

1. ถ้า k มีค่าน้อยกว่า 100 ให้ค่าเป็น 1 และถ้า k มีค่ามากกว่า 100 ให้ค่าเป็น 0

|  |
| --- |
| คำตอบ (แสดงคำสั่งและผลลัพธ์) |

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

|  |
| --- |
| คำตอบ (แสดงคำสั่งและผลลัพธ์) |

1. หาค่าเฉลี่ยของตัวแปร k เก็บไว้ในตัวแปร avg.k

|  |
| --- |
| คำตอบ (แสดงคำสั่งและผลลัพธ์) |

1. ถ้า k มีค่าน้อยกว่า 100 ให้ลบด้วยค่าเฉลี่ยของ k ถ้า k มีค่ามากกว่า 100 ให้บวกด้วยค่าเฉลี่ยของ k

|  |
| --- |
| คำตอบ (แสดงคำสั่งและผลลัพธ์) |

1. ให้หาค่าสะสมตั้งแต่ 1 ถึง 100 โดยใช้คำสั่ง loop for

|  |
| --- |
| คำตอบ (แสดงคำสั่งและผลลัพธ์) |

1. 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) |

1. 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) |

1. ให้นักศึกษาเขียนฟังก์ชั่นเพื่อตรวจสอบว่าเลขที่ส่งไปเป็น “เลขคู่” หรือ “เลขคี่”

|  |
| --- |
| คำตอบ (แสดงคำสั่งและผลลัพธ์) |

1. 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) |

1. จาก data file “CBR64-3Prv\_Practice.xlsx” ให้นักศึกษาฝึกปฏิบัติตรวจสอบการแจกแจงของข้อมูล และแปลงข้อมูลให้อยู่ในรูปของการแจกแจงแบบปกติ ข้อมูลประกอบด้วยตัวแปรดังต่อไปนี้



1. ให้นักศึกษาทำการเปลี่ยนชื่อตัวแปรดังนี้



|  |
| --- |
| คำตอบ (แสดงคำสั่งและผลลัพธ์) |

1. ให้ตรวจสอบการแจกแจงของข้อมูลตัวแปร sal ด้วยกราฟฮิสโตแกรม หากข้อมูลไม่มีการแจกแจงแบบปกติให้ทำการแปลงข้อมูลเพื่อให้ข้อมูลมีการแจกแจงแบบปกติ

|  |
| --- |
| คำตอบ (แสดงคำสั่งและผลลัพธ์)    เขียนคำอธิบาย  ข้อมูลตัวแปร sal มีการแจกแจงแบบปกติโดยการแปลงข้อมูลโดย Log10 |

1. ให้จัดการข้อมูลรายได้ดังนี้  
   - คำนวณหาผลคูณของรายได้ที่เป็นรอบตัวแปร q57xq58 ตั้งชื่อตัวแปรใหม่เป็น q5ro\_total  
   - คำนวณหาผลรวมของรายได้ที่เป็นรายเดือนทั้งหมด ได้แก่ ตัวแปร sal, q53 q61, q62, q63, q64, q65, q66, q68 ตั้งชื่อตัวแปรใหม่เป็นqinc\_mtotal  
   - คำนวณรายได้จากรายเดือนเป็นรายปี qinc\_mtotal x 12 ตั้งชื่อตัวแปรใหม่เป็น qinc\_ytotal   
   (qinc\_ytotal= qinc\_mtotal x 12)  
   - คำนวณหาผลรวมรายได้รายปี q5ro\_total+ qinc\_ytotal+ q54+ q67 ตั้งชื่อตัวแปรใหม่เป็น inctotal

|  |
| --- |
| คำตอบ (แสดงคำสั่งและบางส่วนของผลลัพธ์) |

1. ให้ตรวจสอบการแจกแจงของข้อมูลตัวแปร inctotal ด้วยกราฟฮิสโตแกรม หากข้อมูลไม่มีการแจกแจงแบบปกติให้ทำการแปลงข้อมูลเพื่อให้ข้อมูลมีการแจกแจงแบบปกติ

|  |
| --- |
| คำตอบ (แสดงคำสั่งและผลลัพธ์)      เขียนคำอธิบาย  ข้อมูลตัวแปร sal มีการแจกแจงแบบปกติโดยการแปลงข้อมูลโดย 1/ |

1. 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. >= 5,346 --> สูงกว่ามัธยฐาน

|  |
| --- |
| คำตอบ (แสดงคำสั่งและบางส่วนของผลลัพธ์) |

1. ให้สำรวจข้อมูลกลุ่มตัวอย่างแต่ละจังหวัดมีกลุ่มรายได้ต่อเดือนเป็นอย่างไรบ้าง

|  |
| --- |
| คำตอบ (แสดงคำสั่งและผลลัพธ์) |

1. จาก data file “ncdSuicide.xlsx” ให้นักศึกษาฝึกปฏิบัติการจัดการข้อมูล



1. จากชุดข้อมูลต้นมีจำนวนกี่ประเทศ

|  |
| --- |
| คำตอบ (แสดงคำสั่งและผลลัพธ์) |

1. แต่ละประเทศบันทึกปีเหมือนกันหรือแตกต่างกันอย่างไร

|  |
| --- |
| คำตอบ (แสดงคำสั่งและผลลัพธ์) |

1. 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) |

1. ให้นักศึกษาตัดข้อมูลที่ Gender มีค่าเป็น Both sexes จากนั้นให้สร้างกราฟฮิสโตแกรมของ Crude suicide rates (per 100 000 population) หากข้อมูลไม่มีการแจกแจงแบบปกติให้ทำการแปลงข้อมูลให้อยู่ในรูปการแจกแจงแบบปกติ

|  |
| --- |
| คำตอบ (แสดงคำสั่งและผลลัพธ์) |