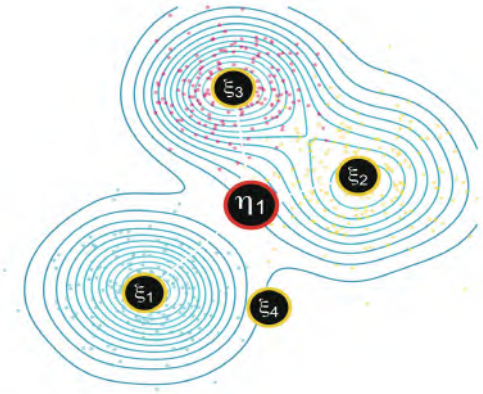
สถิติขั้นสูง

สำหรับการวิจัยทางพฤติกรรมศาสตร์และสังคมศาสตร์

อิศรัฏฐ์ รินไธสง

คณะวิทยาการจัดการ มหาวิทยาลัยสงขลานครินทร์



สถิติขั้นสูงสำหรับการวิจัยทางพฤติกรรมศาสตร์และสังคมศาสตร์

Advanced Statistics for the Behavioral and Social Sciences Research

อิศรัฏฐ์ รินไธสง

ได้รับทุนการเขียนหนังสือและตำรา
จากคณะวิทยาการจัดการ มหาวิทยาลัยสงขลานครินทร์
ปีงบประมาณ พ.ศ. 2567

ชื่อหนังสือ สถิติขั้นสูงสำหรับการวิจัยทางพฤติกรรมศาสตร์และสังคมศาสตร์

ผู้เขียน ผู้ช่วยศาสตราจารย์ ดร.อิศรัฏฐ์ รินไธสง
จัดทำโดย ผู้ช่วยศาสตราจารย์ ดร.อิศรัฏฐ์ รินไธสง

พิมพ์ครั้งที่ 1(e-Book) พ.ศ. 2569
จำนวนหน้า 478 หน้า
ขนาด B5
ออกแบบปก อิศรัฏฐ์ รินไธสง
ออกแบบภายใน บริษัท นีโอพ้อยท์ (1995) จำกัด

พิมพ์ที่ บริษัท นีโอพ้อยท์ (1995) จำกัด
1/59 ถ.ราษฎ์อุทิศ ต.หาดใหญ่ จ.สงขลา

ราคา(e-Book) 390 บาท

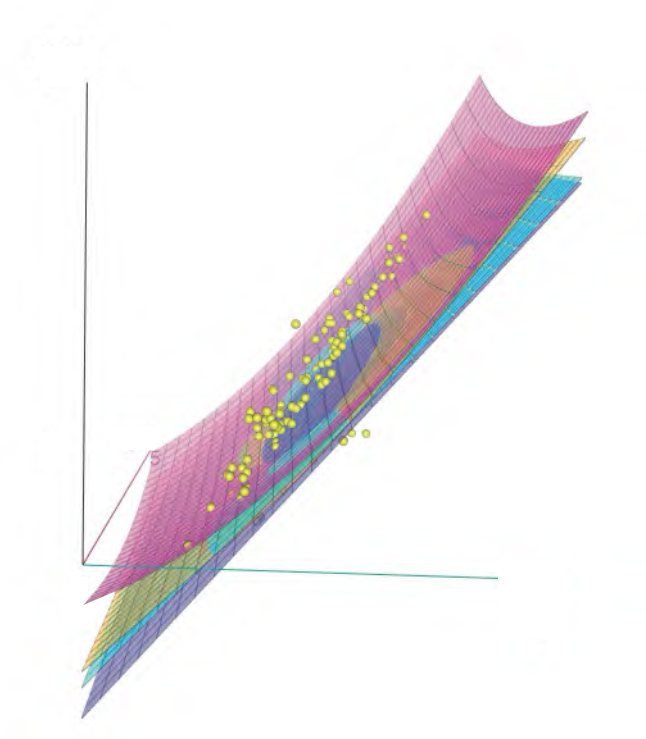
ข้อมูลทางบรรณานุกรมหนังสือ

อิศรัฏฐ์ รินไธสง.
สถิติขั้นสูงสำหรับการวิจัยทางพฤติกรรมศาสตร์และสังคมศาสตร์.--สงขลา : [ม.ป.พ.],2569.
478 หน้า.

1.พฤติกรรมศาสตร์--ระเบียบวิธีทางสถิติ. 2.สังคมศาสตร์--ระเบียบวิธีทางสถิติ. I.ชื่อเรื่อง.

300.727
ISBN 978-616-631-444-1

สงวนลิขสิทธิ์ตามพระราชบัญญัติลิขสิทธิ์ พ.ศ. 2537



สถิติเป็นเครื่องมือสำคัญหนึ่งสำหรับการวิจัยในการแสวงความรู้
ถ้าเรามีเครื่องมือจำกัด การแสวงหาความรู้ก็จะถูกจำกัดไปด้วย

กิตติกรรมประกาศ

หนังสือสถิติขั้นสูงสำหรับการวิจัยทางพฤติกรรมศาสตร์และสังคมศาสตร์ มีวัตถุประสงค์เพื่อยกระดับการใช้สถิติสำหรับการวิจัยเพื่อพัฒนาสมรรถนะการทำวิจัยของของนักศึกษาในระดับบัณฑิตศึกษา รวมทั้งนักวิจัยทางพฤติกรรมศาสตร์ ซึนักวิจัยมักใช้สถิติขั้นสูงเป็นเครื่องมือสำคัญในการเป็นส่วนหนึ่งของการออกแบบงานวิจัยได้หลากหลายและตอบคำถามวิจัยได้ตรงกับความต้องการของตนเองมากขึ้น แต่จากประสบการณ์พบว่านักวิจัยจำนวนมากยังมีปัญหาในการนำสถิติไปใช้ในการตอบคำถามวิจัยของตนเองไม่ถูกต้อง รวมทั้งการแปลความหมายจากผลการวิเคราะห์ นอกจากนี้ปัญหาสำคัญหนึ่งคือการใช้โปรแกรมการวิเคราะห์ที่มีประสิทธิภาพมักสุ่มเสี่ยงปัญหาลิขสิทธิ์ของโปรแกรม ด้วยเหตุผลดังกล่าวจึงได้เขียนหนังสือนี้ขึ้นมาโดยใช้โปรแกรม R ซึ่งเป็นโปรแกรมไม่มีลิขสิทธิ์ แต่มีประสิทธิภาพสูงในการวิเคราะห์ข้อมูลเมื่อใช้สถิติขั้นสูง ผู้เขียนจึงหวังว่าหนังสือนี้จะเป็นส่วนหนึ่งที่ช่วยยกระดับนักวิจัยภาพการทำวิจัยทางรัฐประศาสนศาสตร์ และการประยุกต์ใช้ในศาสตร์อื่นได้มีประสิทธิภาพมากขึ้น

ผู้เขียนขอขอบคุณผู้ทรงคุณวุฒิที่ใช้ให้คำแนะนำที่ดีและเป็นกัลยาณมิตรผ่านกระบวนการ double-blind peer review และขอบคุณเป็นพิเศษสำหรับคณะวิทยาการจัดการที่สนับสนุน

คุณค่าใด ๆ ที่เกิดจากหนังสือเล่มนี้ ผู้เขียนขอมอบแด่อาจารย์ทุกท่านที่มีส่วนประสิทธิ์ประสาทความรู้ทางสถิติสำหรับการวิจัยต่อผู้เขียน โดยเฉพาะคณาจารย์สถาบันวิจัยพฤติกรรมศาสตร์ประยุกต์ มหาวิทยาลัยศรีนครินทรวิโรฒ อันเป็นสถานที่บ่มเพาะผู้เขียนทางด้านสถิติสำหรับการวิจัย ส่วนความผิดพลาดใด ๆ ที่เกิดขึ้นในหนังสือเล่มนี้ผู้เขียนน้อมรับแต่เพียงผู้เดียว

อิศรัฏฐ์ รินไธสง

สารบัญเนื้อหา

กิตติกรรมประกาศ ii
สารบัญเนื้อหา iii
สารบัญตาราง xi
สารบัญภาพ xii
ความนำ xviii
ความนิยม xx

บทที่ 1 บทนำ 1

1. ความนำ 1
2. แนะนำ R และ RStudio 2
   - *2.1 ทำไมต้อง R และ RStuio 2*
   - *2.2 การติดตั้ง R และ RSTudio 3*
   - *2.3 User Interface ของ RStuido 5*
   - *2.4 การกำหนด Theme 7*
   - *2.5 การทำงานของ RStuio 7*
   - *2.6 ชุดโปรแกรม (Packages) 9*
3. การเตรียมและตรวจสอบข้อมูล 20
4. ค่าผิดพลาด 22
   - *4.1 ประเภทของค่าผิดพลาด 22*
   - *4.2 วิธีการจัดการกับค่าผิดพลาด 23*
5. ค่าสุดโต่ง 34
   - *5.1 ลักษณะของค่าสุดโต่ง 34*
   - *5.2 การจัดการกับค่าสุดโต่ง 41*
6. การวิเคราะห์ข้อตกลงเบื้องต้น 42
   - *6.1 การแจกแจงปกติ (Normality) 42*
   - *6.2 การทดสอบความเป็นเอกพันธุ์ของเมทริกซ์ความแปรปรวน–ความแปรปรวนร่วม 50*
   - *6.3 Homoscedasticity 52*
   - *6.4 Sphericity 54*
   - *6.5 ความสัมพันธ์เชิงเส้นตรง 55*
   - *6.6 การร่วมกันเชิงเส้นพหุ 55*
7. สรุป 56
8. เอกสารอ้างอิง 56

บทที่ 2 การวิเคราะห์ถดถอยพหุคูณ 61

1. ความนำ 61
2. วัตถุประสงค์ของการวิเคราะห์ถดถอยพหุคูณ 61

3. การถดถอยอย่างง่าย **62**
3.1 สมการถดถอยอย่างง่าย *62*
3.2 การประมาณค่ากำลังสองน้อยที่สุด (Least Square Estimation) *64*
3.3 การทดสอบนัยสำคัญทางสถิติและการประมาณค่าแบบช่วงของค่าสัมประสิทธิ์ถดถอย *67*
3.4 ค่าความคลาดเคลื่อนในการทำนาย *69*
3.5 ค่าความแปรปรวนโดยรวม *69*
3.6 ค่าความแปรปรวนที่ถูกทำนายได้และค่าสัมประสิทธิ์การตัดสินใจ *70*
3.7 ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่า *71*
4. การถดถอยพหุคูณ **72**
4.1 สมการและวิธีการประมาณในโมเดลถดถอยพหุคูณ *72*
4.2 การทดสอบค่าสัมประสิทธิ์ถดถอย *75*
4.3 สหสัมพันธ์พหุคูณ *77*
4.4 ความคลาดเคลื่อนมาตรฐาน *78*
4.5 ข้อตกลงเบื้องต้นของการวิเคราะห์การถดถอยพหุคูณ *79*
4.6 วิธีการวิเคราะห์ถดถอยพหุคูณ *90*
4.7 การเลือกตัวแปรทำนายที่เหลือเข้าสู่สมการ *90*
4.8 ขนาดตัวอย่างในการวิเคราะห์ถดถอยพหุคูณ *91*
5. การประยุกต์การวิเคราะห์ถดถอยพหุคูณสำหรับการวิจัย **94**
5.1 คำถามวิจัย *94*
5.2 ตัวแปรและข้อมูล *94*
5.3 ออกแบบการวิเคราะห์ *94*
5.4 ผลการวิเคราะห์ *94*
5.5 การวินิจฉัยโมเดลถดถอย *97*
5.6 การตรวจสอบความผิดปกติของค่าสังเกต *103*
5.7 การวิเคราะห์โมเดลด้วยชุดข้อมูลใหม่ *108*
5.8 การคัดเลือกโมเดล *110*
5.9 การเลือกโมเดลโดยพิจารณาจาก AIC *111*
6. สรุป **113**
7. เอกสารอ้างอิง **113**

บทที่ 3 การวิเคราะห์คั่นกลาง กำกับ และกระบวนการแบบมีเงื่อนไข **117**

1. บทนำ **117**
2. วัตถุประสงค์ของการวิเคราะห์ **118**
3. รูปแบบของการวิเคราะห์คั่นกลาง กำกับ และกระบวนการแบบมีเงื่อนไข **118**
4. ซอฟต์แวร์สำหรับการวิเคราะห์คั่นกลาง กำกับ และการวิเคราะห์กระบวนการแบบมีเงื่อนไข **119**
5. การวิเคราะห์คั่นกลางอย่างง่าย **121**

5.1 แนวคิดพื้นฐานการวิเคราะห์คั่นกลาง 121
5.2 อิทธิพลรวม ทางตรง และทางอ้อม 122
5.3 แนวคิดการวิเคราะห์คั่นกลางสมัยใหม่ 123
5.4 รูปแบบการเป็นตัวแปรคั่นกลาง 124
5.5 การประยุกต์การวิเคราะห์คั่นกลางอย่างง่ายสำหรับการวิจัย 125
6. โมเดลตัวแปรคั่นกลางหลายตัวแบบขนาน **129**
6.1 แนวคิดพื้นฐาน 129
6.2 ตัวอย่างการวิเคราะห์ตัวแปรคั่นกลางหลายตัวแบบขนาน 131
7. โมเดลตัวแปรคั่นกลางหลายตัวแบบอนุกรม **134**
7.1 แนวคิดพื้นฐาน 134
7.2 การประยุกต์การวิเคราะห์ตัวแปรคั่นกลางหลายตัวแบบอนุกรมสำหรับการวิจัย 135
8. การวิเคราะห์กำกับ **138**
8.1 แนวคิดพื้นฐาน 138
8.2 อิทธิพลของตัวแปรในฐานะฟังก์ชันตัวแปรกำกับ 138
8.3 การวิเคราะห์ Spotlight และ Floodlight 139
8.4 การประยุกต์การวิเคราะห์กำกับสำหรับการวิจัย 140
9. การวิเคราะห์กระบวนการแบบมีเงื่อนไข **145**
9.1 แนวคิดพื้นฐาน 145
9.2 การประยุกต์ใช้สำหรับการวิจัย 148
10. บทสรุป **153**
11. เอกสารอ้างอิง **157**

บทที่ 4 การวิเคราะห์อนุกรมเวลาแบบขัดขวาง 157

1. บทนำ **157**
2. วัตถุประสงค์ของ ITSA **157**
3. แบบการวิจัย ITS **158**
4. การวิเคราะห์ถดถอยแบบแบ่งช่วง **159**
5. การทดสอบสหสัมพันธ์อัตโนมัติ **165**
6. ขนาดตัวอย่าง **168**
7. ข้อตกลงเบื้องต้น **168**
8. การระบุโมเดลการถดถอยแบบแบ่งช่วงสุดท้าย **169**
9. การประยุกต์ใช้การถดถอยแบบแบ่งช่วง **170**
9.1 คำถามวิจัย 170
9.2 ลักษณะข้อมูลและตัวแปร 170
9.3 การวิเคราะห์ข้อมูล 172
9.4 ผลการวิเคราะห์ข้อมูล 173
9.5 ผลการวินิจฉัยโมเดล 176
10. สรุป **178**
11. เอกสารอ้างอิง **178**

บทที่ 5 การวิเคราะห์ถดถอยโลจิสติก 183

1. ความนำ 183
2. วัตถุประสงค์ของการวิเคราะห์ถดถอยโลจิสติก 184
3. การวิเคราะห์ถดถอยโลจิสติกแบบตัวแปรทวิ 184
4. ข้อตกลงเบื้องต้นของการวิเคราะห์ถดถอยโลจิสติก 188
5. โมเดลการวิเคราะห์ถดถอยโลจิสติก 188
   - *5.1 โมเดลการถดถอยโลจิสติกตัวแปรทวิ* *188*
   - *5.2 โมเดลการถดถอยโลจิสติกมัลติโนเมียล* *190*
6. การแปลความหมายของสัมประสิทธิ์ถดถอยโลจิสติก 192
   - *6.1 แนวคิดเบื้องต้น* *192*
   - *6.2 กรณีตัวแปรทำนายมี 2 ค่า* *193*
   - *6.3 กรณีที่ตัวแปรทำนายมีหลายค่า* *196*
   - *6.4 กรณีตัวแปรทำนายมีค่าต่อเนื่อง* *198*
7. ขั้นตอนการวิเคราะห์ถดถอยโลจิสติก 200
8. ขนาดตัวอย่างในการวิเคราะห์ถดถอยโลจิสติก 200
9. การประเมินความกลมกลืนของโมเดลโลจิสติก 201
   - *9.1 การวัดสัมบูรณ์ (Absolute Measures)* *201*
   - *9.2 การวัดสัมพัทธ์ (Relative Measures)* *202*
10. การประยุกต์วิเคราะห์ถดถอยโลจิสติกสำหรับการวิจัย 203
    - *10.1 คำถามวิจัย* *203*
    - *10.2 ตัวแปรและข้อมูล* *203*
    - *10.3 ออกแบบการวิเคราะห์* *203*
    - *10.4 ผลการวิเคราะห์* *203*
    - *10.5 การวินิจฉัยโมเดลโลจิสติก* *205*
11. สรุป 211
12. เอกสารอ้างอิง 212
13. ภาคผนวกท้ายบท 214

บทที่ 6 การวิเคราะห์ฟังก์ชันการจำแนก 215

1. ความนำ 215
2. วัตถุประสงค์ของการวิเคราะห์ฟังก์ชันการจำแนก 215
3. ลักษณะของตัวแปรในการวิเคราะห์ฟังก์ชันการจำแนก 216
4. ข้อตกลงเบื้องต้นของการวิเคราะห์ฟังก์ชันการจำแนก 216
5. ขนาดตัวอย่างสำหรับการวิเคราะห์ฟังก์ชันการจำแนก 217
6. ฟังก์ชันการจำแนก 217
   - *6.1 ฟังก์ชันการจำแนกสำหรับ 2 กลุ่ม* *218*
   - *6.2 ฟังก์ชันการจำแนกสำหรับหลายกลุ่ม* *221*
   - *6.3 การวัดความสัมพันธ์ของฟังก์ชันการจำแนก* *223*

7. การแปลความหมายฟังก์ชันการจำแนก 223
7.1 ค่าสัมประสิทธิ์มาตรฐาน *223*
7.2 ความสัมพันธ์ระหว่างตัวแปรกับฟังก์ชันการจำแนก *224*
8. การทดสอบนัยสำคัญ 224
8.1 กรณีสองกลุ่ม *225*
8.2 กรณีมีหลายกลุ่ม *226*
9. การจัดกลุ่ม 227
9.1 การจัดกลุ่มเพื่อเข้ากลุ่มสองกลุ่ม 227
9.2 การจัดเข้ากลุ่มเมื่อมีหลายกลุ่ม 230
9.3 การประเมินการจัดเข้ากลุ่ม 231
10. การประยุกต์การวิเคราะห์ฟังก์ชันการจำแนกกรณีสองกลุ่ม 231
10.1 คำถามวิจัย *231*
10.2 ข้อมูลและตัวแปร *231*
10.3 ออกแบบการวิเคราะห์ข้อมูล *232*
10.4. ผลการวิเคราะห์ข้อมูล *232*
11. การวิเคราะห์ฟังก์ชันการจำแนกกรณีหลายกลุ่ม 243
11.1 คำถามวิจัย *243*
11.2 ข้อมูลและตัวแปร *243*
11.3 ออกแบบการวิเคราะห์ข้อมูล *243*
11.4. ผลการวิเคราะห์ข้อมูล *243*
12. สรุป 255
13. เอกสารอ้างอิง 255

บทที่ 7 การวิเคราะห์ความแปรปรวนพหุตัวแปร 257

1. ความนำ 257
2. วัตถุประสงค์ของ MANOVA 258
3. การใช้ MANOVA 258
3.1 เหตุผลการใช้ MANOVA *258*
3.2 ความสัมพันธ์ระหว่างตัวแปรตาม *259*
3.3 ขนาดตัวอย่าง *260*
3.4 ตัวแปรที่ใช้ในการวิเคราะห์ MANOVA *260*
4. หลักการวิเคราะห์ MANOVA 260
5. ข้อตกลงเบื้องต้นของ MANOVA 263
6. ขั้นตอนการวิเคราะห์ MANOVA 264
7. วิธีการวิเคราะห์ความแปรปรวนพหุตัวแปรแบบทางเดียว 265
7.1 ประชากร 2 กลุ่ม *265*
7.2 ประชากร 3 กลุ่มขึ้นไป *268*
8. การวิเคราะห์ two way MANOVA 274

9. **การทดสอบหลังการวิเคราะห์ MANOVA มีนัยสำคัญทางสถิติ** **276**
9.1 ตัวแปรเดี่ยว *277*
9.2 พหุตัวแปร *277*
10. **การเลือกสถิติทดสอบสำหรับ MANOVA** **277**
10.1 กรณีความแกร่ง (Robustness) *278*
10.2 กรณีอำนาจการทดสอบ (Power of the Test) *278*
11. **การประยุกต์ MANOVA สำหรับการวิจัย: กรณี 2 กลุ่ม** **278**
11.1 คำถามวิจัย *278*
11.2 ตัวแปรและข้อมูล *278*
11.3 ออกแบบการวิเคราะห์ข้อมูล *279*
11.4 ผลการวิเคราะห์ *279*
11.5 การวินิจฉัยโมเดล *282*
12. **การประยุกต์ MANOVA สำหรับการวิจัย: กรณี 3 กลุ่ม** **287**
12.1 คำถามวิจัย *287*
12.2 ตัวแปรและข้อมูล *288*
12.3 ออกแบบการวิเคราะห์ข้อมูล *288*
12.4 ผลการวิเคราะห์ *288*
12.5 การวินิจฉัย MANOVA *290*
13. **การประยุกต์ MANOVA สำหรับการวิจัย: two way MANOVA** **291**
13.1 คำถามวิจัย *291*
13.2 ตัวแปรและข้อมูล *291*
13.3 ออกแบบการวิเคราะห์ข้อมูล *291*
13.4 ผลการวิเคราะห์ *291*
14. **สรุป** **294**
15. **เอกสารอ้างอิง** **294**

บทที่ 8 การวิเคราะห์องค์ประกอบเชิงสำรวจ 297

1. **ความนำ** **297**
2. **วัตถุประสงค์ของการวิเคราะห์องค์ประกอบ** **298**
3. **แนวคิดพื้นฐานการวิเคราะห์องค์ประกอบ** **298**
3.1 ความหมาย และวัตถุประสงค์ของการวิเคราะห์องค์ประกอบ *298*
3.2 ข้อตกลงเบื้องต้นการวิเคราะห์องค์ประกอบ *300*
3.3 ขั้นตอนการวิเคราะห์องค์ประกอบเชิงสำรวจ *301*
4. **การสกัดองค์ประกอบ** **302**
4.1 วิธีการสกัดองค์ประกอบ: PCA vs FA *302*
4.2 การเลือกวิธีสกัดองค์ประกอบ *307*
5. **การกำหนดจำนวนองค์ประกอบ** **308**
5.1 เทคนิคการกำหนดจำนวนองค์ประกอบ *308*

5.2 การเลือกเทคนิคการกำหนดจำนวนองค์ประกอบ 314
6. การหมุนแกนองค์ประกอบ 315
6.1 เทคนิคการหมุนองค์ประกอบ 315
6.2 การหมุนแกนแบบออร์โธกอนอลกับออบลิค 320
7. ค่าน้ำหนักองค์ประกอบ 321
8. คะแนนองค์ประกอบ 322
8.1 วิธีการไม่มีแก้ปรับ 323
8.2 วิธีการมีแก้ปรับ 324
9. ขนาดตัวอย่างสำหรับการวิเคราะห์องค์ประกอบ 328
10. ประยุกต์ PCA สำหรับการวิจัย 329
10.1 คำถามวิจัย 329
10.2 ตัวแปรและข้อมูล 329
10.3 ออกแบบการวิเคราะห์ 330
10.4 ค่าสถิติพื้นฐานกับตัวแปร 330
10.5 การตรวจสอบข้อตกลงเบื้องต้นเกี่ยวกับความสัมพันธ์ระหว่างตัวแปร 331
10.6 การสกัดองค์ประกอบ 333
10.7 การกำหนดจำนวนองค์ประกอบ 337
10.8 การหมุนแกน 342
10.9 การแปลความหมายองค์ประกอบ 345
10.10 การสร้างคะแนนองค์ประกอบ 346
11. ประยุกต์ FA สำหรับการวิจัย 348
11.1 คำถามวิจัย 348
11.2 ตัวแปรและข้อมูล 348
11.3 ออกแบบการวิเคราะห์ 349
11.4 การสกัดองค์ประกอบ 349
11.5 จำนวนองค์ประกอบ 351
11.6 การหมุนแกนองค์ประกอบและการแปลความหมาย 352
11.7 การสร้างองค์ประกอบ 357
12. สรุป 358
13. เอกสารอ้างอิง 358

บทที่ 9 การวิเคราะห์โมเดลสมการโครงสร้าง 365

1. ความนำ 365
2. วัตถุประสงค์ของการวิเคราะห์โมเดลสมการโครงสร้าง 366
3. โมเดลสมการโครงสร้าง (SEM Model) 367
4. ข้อตกลงเบื้องต้นสำหรับโมเดลสมการโครงสร้าง 370
5. ขั้นตอนการวิเคราะห์โมเดลสมการโครงสร้าง 370
5.1 การกำหนดลักษณะเฉพาะของโมเดล (Model Specification) 371

5.2 การระบุความเป็นไปได้ค่าเดียวของโมเดล (Model Identification) 373
5.3 การประมาณค่าพารามิเตอร์ของโมเดล (Parameter Estimation of the Model) 376
5.4 การประเมินความกลมกลืนของโมเดล 377
5.5 การปรับโมเดล 381
6. ปัญหาที่พบในการประมาณค่าพารามิเตอร์ **383**
7. การประยุกต์ใช้โมเดลสมการโครงสร้างสำหรับการวิจัย **385**
7.1 คำถามวิจัย 385
7.2 ข้อมูลและตัวแปร 385
7.3 ออกแบบการวิเคราะห์ข้อมูล 386
7.4 ผลการวิเคราะห์ 386
8. สรุป **401**
9. เอกสารอ้างอิง **401**

บทที่ 10 การวิเคราะห์โมเดลถดถอยพหุระดับ **405**

1. ความนำ **405**
2. วัตถุประสงค์ในการวิเคราะห์โมเดลพหุระดับ **406**
3. โมเดลถดถอยพหุระดับ **406**
3.1 โมเดลพหุระดับแบบ 2 ระดับ **407**
3.2 โมเดลพหุระดับแบบ 3 ระดับ **409**
4. การประมาณค่าพารามิเตอร์ **410**
5. การพัฒนาโมเดลพหุระดับ **411**
6. การกำหนดค่ากลาง **420**
7. ขนาดของตัวอย่างในการวิเคราะห์พหุระดับ **420**
8. การประยุกต์ใช้โมเดลถดถอยพหุระดับสำหรับการวิจัย **421**
8.1 คำถามวิจัย 421
8.2 ตัวแปรและข้อมูล 421
8.3 การวิเคราะห์ข้อมูล 422
8.4 ผลการวิเคราะห์ข้อมูล 422
8.5 การวินิจฉัยโมเดล 437
9. สรุป **448**
10. เอกสารอ้างอิง **449**

ภาคผนวกดัชนีคำสำคัญ **453**

สารบัญตาราง

ตาราง 1-1 ฟังก์ชันการบันทึกผลลัพธ์กราฟิก 15
ตาราง 1-2 เกณฑ์ค่า Z-score เพื่อพิจารณาในการกำหนดข้อมูลให้เป็นค่าสุดโต่ง 34
ตาราง 2-1 แสดงคะแนนการรับรู้ความยุติธรรม (POJ) และพฤติกรรมการป็นสมาชิกที่ดี (OCB) 65
ตาราง 2-2 การวิเคราะห์ความแปรปรวนของการวิเคราะห์ถดถอย 75
ตาราง 2-3 เงื่อนไขการตัดสินใจยอมรับหรือปฏิเสธ H0 ในการทดสอบค่า d 81
ตาราง 2-4 ค่าสถิติและเกณฑ์บ่งชี้การมีปัญหาการร่วมกันเชิงเส้นพหุ 83
ตาราง 4-1 ลักษณะข้อมูลอนุกรมเวลาแบบขัดขวางสำหรับการถดถอยแบบแบ่งช่วง 161
ตาราง 4-2 ข้อมูลอนุกรมเวลาของจำนวนเฉลี่ยของใบสั่งยาที่จ่ายต่อคนไข้ต่อเดือน 170
ตาราง 5-1 แสดงข้อมูลตัวแปร AGE กับ CHD 185
ตาราง 5-2 แสดงความถี่และสัดส่วนของการมีภาวะซึมเศร้าจำแนกตามกลุ่มอายุ 186
ตาราง 5-3 แสดงสมการสำหรับโมเดลโลจิสติกเมื่อตัวแปรทำนายเป็นตัวแปรทวิ 193
ตาราง 5-4 แสดงการจำแนกกลุ่มตามตัวแปร age กับ dps 194
ตาราง 5-5 การกำหนดรหัสให้กับตัวแปรใหม่เมื่อมีตัวแปรทำนายมีหลายค่า 197
ตาราง 6-1 ตารางข้อมูลการจำแนกองค์กรด้วย VIS และ EMO 220
ตาราง 6-2 สัมประสิทธิ์การจำแนกคะแนนไม่มาตรฐานและมาตรฐานของฟังก์ชันการจำแนก 221
ตาราง 6-3 ค่า Eigenvalue และ Canonical Correlation 223
ตาราง 6-4 ค่า Wilks' lambda และการทดสอบนัยสำคัญด้วย Chi-Square 226
ตาราง 6-5 Functions at Group Centroids 229
ตาราง 6-6 การใช้ฟังก์ชันการจำแนกทำนายตัวอย่างเข้ากลุ่ม 229
ตาราง 7-1 แสดงสูตรการคำนวณเวกเตอร์ค่าเฉลี่ยจากสองประชากร 266
ตาราง 7-2 แสดงการคำนวณเมทริกซ์ SSCP และค่าองศาอิสระ 270
ตาราง 7-3 แสดงการแจกแจงของ F ตามจำนวนตัวแปรและจำนวนกลุ่ม 271
ตาราง 7-4 แสดงการวิเคราะห์ two way MANOVA Fixed Effect with Interaction Model 275
ตาราง 8-1 เป้าหมายการสกัดองค์ประกอบในแต่ละวิธี 307
ตาราง 8-2 ตัวอย่างค่าไอเกน 310
ตาราง 8-3 เปรียบเทียบโครงสร้างแบบแผนของน้ำหนักองค์ประกอบ 317
ตาราง 8-4 สรุปกระบวนการ ข้อดี และข้อควรพิจารณาของวิธีคิดคะแนนองค์ประกอบ 327
ตาราง 8-5 อัตราส่วนของตัวอย่างต่อตัวแปรในการวิเคราะห์องค์ประกอบ 328
ตาราง 8-6 แสดงการตั้งชื่อองค์ประกอบ 345
ตาราง 9-1 ค่าดัชนีวัดความกลมกลืนโมเดลสมการโครงสร้าง 480
ตาราง 10-1 โครงสร้างตัวแปร 412
ตาราง 10-2 ค่าสหสัมพันธ์ระหว่างตัวแปร 412
ตาราง 10-3 แสดงผลการวิเคราะห์ one way ANOVA with Random Effects Model 413
ตาราง 10-4 ตัวอย่างผลการวิเคราะห์ Level-1 Random Intercept Model 416
ตาราง 10-5 แสดงผลการวิเคราะห์ด้วย Group Level Random Intercept Model 417
ตาราง 10-6 ผลการวิเคราะห์ด้วย Random-Intercept –and –Slope Model 418
ตาราง 10-7 แสดงผลการวิเคราะห์ด้วย Intercept-and-Slopes-as-Outcomes Model 419

สารบัญภาพ

ภาพ 1-1	หน้าต่างแสดงการติดติดตั้ง R สำเร็จและพร้อมใช้งาน	4
ภาพ 1-2	การดาวน์โหลด RStudio และ Icons ของ R และ RStudio ที่พร้อมใช้งาน	4
ภาพ 1-3	ส่วนประกอบของ User interface ของ RStudio	5
ภาพ 1-4	การปรับ Theme	8
ภาพ 1-5	คำสั่งใน Source และผลลัพธ์ที่ Console	9
ภาพ 1-6	การใช้คำสั่ง search () และผลลัพธ์ที่ Console	10
ภาพ 1-7	การติดตั้งชุดโปรแกรม gclus ด้วยคำสั่ง install.package()	10
ภาพ 1-8	การติดตั้งชุดโปรแกรม gclus ด้วย Install tap	11
ภาพ 1-9	การ Update ชุดโปรแกรม	12
ภาพ 1-10	การนำชุดข้อมูลเข้า	13
ภาพ 1-11	การนำชุดข้อมูลเข้าโดย Browse และรายละเอียดของชุดข้อมูล	13
ภาพ 1-12	การชุดข้อมูลเข้าและการแสดงข้อมูลในแผง Sources	13
ภาพ 1-13	เทียบคำสั่งแผง Source กับ ผลการ run ที่แสดงในแผง Console	14
ภาพ 1-14	การนำเข้า R Script	14
ภาพ 1-15	การนำเข้าโดยฟังก์ชัน source() และผลลัพธ์โดยใช้ฟังก์ชัน sink()	17
ภาพ 1-16	การบันทึกข้อมูลภาพจากแผง Output	17
ภาพ 1-17	รายการคำสั่ง และผลลัพธ์จากหนังสือ และการนำเขียนคำสั่งในแผง Sources	18
ภาพ 1-18	แสดงข้อมูลจากคำสั่ง View ()	19
ภาพ 1-19	การใช้เอกสารช่วยเหลือเกี่ยวกับการฟังก์ชัน mean	19
ภาพ 1-20	รายละเอียดโครงสร้างของวิธีใช้จากเอกสารช่วยเหลือเกี่ยวกับการฟังก์ชัน mean	20
ภาพ 1-21	ตรวจสอบแบบแผนค่าผิดพลาด	27
ภาพ 1-22	ตรวจสอบจำนวนค่าผิดพลาดด้วยข้อมูลภาพ	28
ภาพ 1-23	ตรวจสอบการกระจายของข้อมูลค่าผิดพลาดกับข้อมูลเดิม	31
ภาพ 1-24	ตรวจสอบการแจกแจงชุดข้อมูลค่าผิดพลาดกับข้อมูลเดิม	32
ภาพ 1-25	แผนภาพจาการวิเคราะห์ Box plot	35
ภาพ 1-26	แสดงค่าสุดโต่ง DMi จากภาพเรขาคณิตแบบ 3 มิติ	36
ภาพ 1-27	กราฟตรวจสอบค่าสุดโต่งพหุตัวแปรจากวิธีมาฮาลาโนบิส	37
ภาพ 1-28	การตรวจสอบค่าสุดโต่งด้วย Cook's distance	39
ภาพ 1-29	การตรวจสอบค่าสุดโต่งด้วยวิธีหาค่าส่วนเหลือมาตรฐาน	40
ภาพ 1-30	ลักษณะการแจกแจงแบบเบ้ซ้ายและขวา	43
ภาพ 1-31	ลักษณะการแจกแจงเมื่อมีความโด่งต่างกัน	43
ภาพ 1-32	Q-Q plot	46
ภาพ 1-33	ภาพการแจกแจงแบบสองตัวแปรแบบ perspective (A) และ contour (B)	50
ภาพ 1-34	ความเป็นเอกพันธ์ของเมทริกซ์ความแปรปรวน	51
ภาพ 1-35	ความสัมพันธ์ค่าจากสมการถดถอยกับค่าส่วนเหลือ	52
ภาพ 2-1	แสดงเส้นถดถอยอย่างง่าย	63
ภาพ 2-2	แสดงความสัมพันธ์ระหว่างค่าตัวแปรตามที่แท้จริงกับค่าที่ถูกทำนาย	66

ภาพ 2-3	การประมาณแบบช่วงของ Y	69
ภาพ 2-4	ระนาบที่ได้จากกำลังสองน้อยที่สุดของโมเดลถดถอยพหุคูณ	73
ภาพ 2-5	ขอบเขตการยอมรับและปฏิเสธ H0 ในการทดสอบสหสัมพันธ์ออโตในการทดสอบ d test	82
ภาพ 2-6	การกำหนดขนาดตัวอย่างจากการวิเคราะห์อำนาจการทดสอบจาก G*power	92
ภาพ 2-7	กราฟความสัมพันธ์ระหว่างอำนาจการทดสอบกับขนาดตัวอย่างจาก G*power	93
ภาพ 2-8	อำนาจการทดสอบกับขนาดตัวอย่างจาก G*power	93
ภาพ 2-9	กรอบความคิดการวิจัยสำหรับการวิเคราะห์ถดถอยพหุคูณ	94
ภาพ 2-10	เมทริกซ์แผนภาพการกระจายของตัวแปรตามกรอบแนวคิด เส้นตรงแสดงความพอดี และค่าสหสัมพันธ์	96
ภาพ 2-11	แผนภาพการวินิจฉัยค่าสุดโต่ง	99
ภาพ 2-12	แผนภาพการวินิจฉัยโมเดลถดถอย	100
ภาพ 2-13	การตรวจความสัมพันธ์เชิงเส้นตรงระหว่างสองตัวแปรด้วย component plus plots	102
ภาพ 2-14	แผนภาพแสดงค่า hat และดัชนีระบุค่าสังเกตผิดปกติ	105
ภาพ 2-15	แผนภาพวิเคราะห์อิทธิพลจากค่าสังเกตโดยวิธี Cook's distance	106
ภาพ 2-16	แผนภาพ Added-variable	107
ภาพ 2-17	แผนภาพอิทธิพลจากค่าสังเกตจาก Cook's D และ Leverage	108
ภาพ 2-18	โมเดลจากวิเคราะห์ถดถอยพหุคูณพฤติกรรมการเป็นสมาชิกที่ดีในองค์การ	110
ภาพ 3-1	โมเดลตามกรอบแนวคิดและทางสถิติการวิเคราะห์คั่นกลาง (A)	118
ภาพ 3-2	โมเดลจากการใช้คำสั่ง pmacroModel และ statisticalDiagram	121
ภาพ 3-3	โมเดลการวิเคราะห์คั่นกลางอย่างง่ายและที่มีตัวแปรร่วม	121
ภาพ 3-4	โมเดลทางสถิติตัวแปรคั่นกลางอย่างง่าย	122
ภาพ 3-5	โมเดลการวิเคราะห์คั่นกลางอย่างง่ายที่มีตัวแปรร่วม	126
ภาพ 3-6	ผลลัพธ์การวิเคราะห์โมเดลตัวแปรคั่นกลางอย่างง่ายของอารมณ์ซึมเศร้า	129
ภาพ 3-7	โมเดลทางสถิติตัวแปรคั่นกลางหลายตัวแบบขนาน	130
ภาพ 3-8	โมเดลกรอบแนวคิดตัวแปรคั่นกลางหลายตัวแบบขนาน	131
ภาพ 3-9	โมเดลผลการวิเคราะห์คั่นกลางหลายตัวแบบขนาน	134
ภาพ 3-10	โมเดลทางสถิติตัวแปรคั่นกลางหลายตัวแบบอนุกรม	135
ภาพ 3-11	โมเดลกรอบแนวคิดสำหรับตัวแปรคั่นกลางหลายตัวแบบอนุกรม	136
ภาพ 3-12	โมเดลผลการวิเคราะห์คั่นกลางหลายตัวแบบอนุกรม	137
ภาพ 3-13	แผนภาพกรอบแนวคิดโมเดลกำกับอย่างง่าย	138
ภาพ 3-14	โมเดลกรอบแนวคิดตัวแปรกำกับอย่างง่าย	141
ภาพ 3-15	แผนภาพแสดงความชันอย่างง่ายของ estresss ต่อ affect เมื่อ ese ณ ตำแหน่ง -1SD, Mean และ +1SD	143
ภาพ 3-16	J-N plot ที่มีการปรับค่ากลาง (A) และใช้คะแนนดิบ (B)	144
ภาพ 3-17	โมเดลผลการวิเคราะห์ตัวแปรกำกับอย่างง่าย	145
ภาพ 3-18	ตัวอย่างโมเดลกรอบแนวคิด (A) และสถิติ (B) การวิเคราะห์กระบวนการแบบมีเงื่อนไข	146

ภาพ 3-19	การวิเคราะห์กระบวนการแบบมีเงื่อนไขตามโมเดล 8 ของ Hayes (2018)	147
ภาพ 3-20	โมเดลการกรอบแนวคิดและโมเดลทางสถิติการวิเคราะห์กระบวนการแบบมีเงื่อนไข กรณีตัวอย่าง	148
ภาพ 3-21	แผนภาพแสดงการทดสอบค่าความชันอย่างง่าย (A) และ J-N plot (B)	152
ภาพ 3-22	โมเดลผลการวิเคราะห์กระบวนการแบบมีเงื่อนไขของอิทธิพลของความเครียดทางเศรษฐกิจต่อความตั้งใจถอนตัวจากการเป็นผู้ประกอบการ	152
ภาพ 4-1	แนวโน้มการใช้การถดถอยแบบแบ่งส่วน ช่วงปี 2015 - 2023	160
ภาพ 4-2	ลักษณะ ITS ผลกระทบนโยบายต่อจำนวนสถานการณ์ความรุนแรงในสามจังหวัดชายแดนใต้	162
ภาพ 4-3	ลักษณะของการเปลี่ยนแปลงระดับและการเปลี่ยนแปลงแนวโน้มความชันการออกแบบ 2 เฟส	164
ภาพ 4-4	ลักษณะแผนภาพข้อมูลอนุกรมเวลาของจำนวนเฉลี่ยของใบสั่งยาที่จ่ายต่อคนไข้ต่อเดือน	172
ภาพ 4-5	ข้อมูลภาพแสดงค่าสหสัมพันธ์ระหว่างตัวแปร	174
ภาพ 4-6	ผลการวิเคราะห์ถดถอยแบบแบ่งช่วงจากข้อมูล ppm	176
ภาพ 4-7	พล็อตกราฟ Residual vs Fitted และ Q-Q plot	177
ภาพ 5-1	แผนภาพการกระจายของ dps ตาม age	186
ภาพ 5-2	แผนภาพการกระจายแสดงสัดส่วนการเป็น dps ในแต่ละกลุ่มอายุ	187
ภาพ 5-3	ข้อมูล dpsdataset และ ตัวแปร age_group ที่เตรียมขึ้นมาใหม่	195
ภาพ 5-4	การวิเคราะห์ค่า exponential จากอินเทอร์เน็ต	199
ภาพ 5-5	การตรวจสอบอิทธิพลของค่าสังเกต	206
ภาพ 5-6	การตรวจสอบอิทธิพลของค่าสังเกตด้วยค่า studentized residuals และ hat values	207
ภาพ 5-7	การตรวจสอบอิทธิพลของเลเวอเรจน์	208
ภาพ 5-8	การตรวจสอบส่วนเหลือของแต่ละตัวแปร	210
ภาพ 5-9	การวิเคราะห์ส่วนเหลือ	211
ภาพ 6-1	การจำแนกกลุ่มโดยฟังก์ชันการจำแนก	219
ภาพ 6-2	แผนภาพการกระจายตัวแปร VIS และ EMO	220
ภาพ 6-3	ภาพฉายตามแกนน X และ Y ของตัวแปร VIS และ EMO	221
ภาพ 6-4	การจัดตัวอย่างเข้ากลุ่มตามวิธีของ Fisher	228
ภาพ 6-5	แสดงการจัดเข้ากลุ่มโดยพิจารณาจากคะแนน Z เทียบกับจุดตัด	229
ภาพ 6-6	ผลการตรวจสอบค่าผิดพลาด	233
ภาพ 6-7	ความสัมพันธ์ระหว่างตัวแปร	233
ภาพ 6-8	การแจกแจงคะแนนจากฟังก์ชันการจำแนกของแต่ละกลุ่ม	237
ภาพ 6-9	ขนาดและทิศทางความสัมพันธ์ระหว่างตัวแปรกับคะแนนคาโนนิคอล	239
ภาพ 6-10	ภาพผลการจัดเข้ากลุ่ม	241
ภาพ 6-11	ผลการตรวจสอบการแจกแจงและค่าสหสัมพันธ์ระหว่างตัวแปรอิสระ	244
ภาพ 6-12	แผนภาพการจำแนกกลุ่มจาก LD1 และ LD2	252

ภาพ 7-1	แสดงค่าเฉลี่ยของตัวแปรแบบ univariate (A) กับ ค่ากลาง (group centroids) ที่ได้ (B) จากการรวมเชิงเส้น (linear combination)	262
ภาพ 7-2	แสดงการเปรียบเทียบค่าเฉลี่ยแบบตัวแปรเดี่ยว (A) และ พหุตัวแปร (B)	262
ภาพ 7-3	ลักษณะของเมทริกซ์ความแปรปรวน-ความแปรปรวนร่วม	263
ภาพ 7-4	แสดงขั้นตอนการวิเคราะห์ MANOVA	265
ภาพ 7-5	แผนภาพแสดงค่าสุดโต่งพหุตัวแปร	280
ภาพ 7-6	แผนภาพแสดงค่าสุดโต่งพหุตัวแปร	283
ภาพ 7-7	แผนภาพแสดงค่าสุดโต่งของตัวอย่างรายตัวแปรด้วย Cook's distance	285
ภาพ 7-8	แผนภาพแสดงความแปรปรวน-ความแปรปรวนร่วม	286
ภาพ 7-9	แผนภาพการตรวจสอบการแจกแจงปกติพหุตัวแปรด้วย Q-Q plot	287
ภาพ 8-1	แสดง common variance กับ uniqe variance	303
ภาพ 8-2	แสดงโมเดล FA (A) และ PCA (B)	303
ภาพ 8-3	สัดส่วนความแปรปรวนในการวิเคราะห์องค์ประกอบ	304
ภาพ 8-4	ตัวอย่าง scree plot	311
ภาพ 8-5	กราฟแสดงค่าไอเกนจากตัวแปรเดิมและที่ได้จากเทคนิค B-S	312
ภาพ 8-6	แสดงตำแหน่งขององค์ประกอบที่อาจไม่ชัดเจนด้วย scree test	315
ภาพ 8-7	แสดงน้ำหนักองค์ประกอบก่อน (A) และหลัง (B) การหมุนแกนองค์ประกอบ	316
ภาพ 8-8	ตำแหน่งและค่าร้อยละก่อน (A) และหลัง (B) การหมุนแกนองค์ประกอบ	319
ภาพ 8-9	แผนภาพแสดงความสัมพันธ์ระหว่างตัวแปรเมื่อมีการจัดกลุ่มความสัมพันธ์	333
ภาพ 8-10	ค่าน้ำหนักองค์ประกอบในรูปของแผนภาพก่อนหมุนแกน	337
ภาพ 8-11	แผนภาพ scree plot จากค่าไอเกน	339
ภาพ 8-12	กราฟแสดงค่าไอเกนที่ได้จากตัวแปรเดิมกับที่ได้จากเทคนิค PA	340
ภาพ 8-13	กราฟแสดงค่าไอเกนที่ได้จากตัวแปรเดิมกับที่ได้จากเทคนิค VSS	342
ภาพ 8-14	แผนภาพโครงสร้างองค์ประกอบก่อนและหลังการหมุนแกนแบบแวริแมกซ์	345
ภาพ 8-15	กราฟแสดงค่าไอเกนที่ได้จากข้อมูลจริงและเมื่อสกัดองค์ประกอบด้วย PA	352
ภาพ 8-16	ข้อมูลภาพน้ำหนักองค์ประกอบจากวิธี FA สกัดด้วย ML	354
ภาพ 8-17	แผนภาพโครงสร้างองค์ประกอบก่อนและหลังการหมุนแกนแบบโพรแมกซ์	355
ภาพ 9-1	รูปแบบโมเดลสมการโครงสร้าง	367
ภาพ 9-2	ขั้นตอนการวิเคราะห์โมเดลสมการโครงสร้าง	371
ภาพ 9-3	โมเดลสมการโครงสร้างเพื่อกำหนดลักษณะเฉพาะ	372
ภาพ 9-4	ลักษณะเฉพาะโมเดลระบุความเป็นไปได้ค่าเดียวเกินพอดี	375
ภาพ 9-5	ลักษณะข้อมูลที่ใช้วิเคราะห์โมเดลสมการโครงสร้าง	386
ภาพ 9-6	ข้อมูลภาพด้วยข้อมูลภาพค่าสหสัมพันธ์ระหว่างตัวแปรในโมเดล	387
ภาพ 9-7	โมเดลการวัดของตัวแปรตามกรอบแนวคิดการวิจัย	393
ภาพ 9-8	โมเดลสมการโครงสร้างจากคะแนนไม่มาตรฐาน (A) และคะแนนมาตรฐาน (B) ก่อนปรับโมเดล	398
ภาพ 9-9	โมเดลสมการโครงสร้างจากคะแนนมาตรฐานหลังปรับโมเดล	400
ภาพ 10-1	แผนภาพโมเดลการวิเคราะห์พหุระดับแบบ 2 ระดับ	408

ภาพ 10-2 แสดงแผนภาพโมเดลการวิเคราะห์พหุระดับแบบ 3 ระดับ 410
ภาพ 10-3 แสดงขั้นตอนการวิเคราะห์โมเดลพหุระดับแบบ 2 ระดับ 411
ภาพ 10-4 แผนภาพแสดงค่าผิดพลาดและค่าสังเกต 422
ภาพ 10-5 แผนภาพค่าสหสัมพันธ์และและการแจกแจง 423
ภาพ 10-6 แผนภาพค่าความชันของ 25 กลุ่มแรก 433
ภาพ 10-7 แผนภาพปฏิสัมพันธ์ระหว่าง G.HRS กับ LEAD 436
ภาพ 10-8 แผนภาพ dotplot ของ Cook's distance ในระดับที่ 1 439
ภาพ 10-9 แผนภาพ modified dotplot ของ Cook's disctance และ MDFFITS ในระดับที่ 2 441
ภาพ 10-10 แผนภาพค่าอิทธิพลจากการเปลี่ยนแปลงความแปรปรวนสัมพัทธ์กับกลุ่ม 445
ภาพ 10-11 Over all leverage ในระดับที่ 2 447
ภาพ 10-12 การตรวจสอบความเป็นเส้นตรง (A) และความคงที่ของความความแปรปรวนส่วนเหลือ (B) 448

ความนำ

ในช่วงไม่กี่ทศวรรษที่ผ่านมาได้มีการนำเสนอเทคนิคการวิเคราะห์ขั้นสูงที่จุดเน้นร่วมกันคือการวิเคราะห์พหุตัวแปร (multivariate analysis) และมีวิวัฒนาการอย่างต่อเนื่องของวิธีการทางสถิติแบบ “ดั้งเดิม” เช่น การถดถอยพหุคูณ การวิเคราะห์ความแปรปรวน/การวิเคราะห์องค์ประกอบเชิงสำรวจ เทคนิคทางสถิติแหล่านี้ได้รับการขยายให้มีขีดความสามารถในการประยุกต์ใช้มากขึ้น เช่น โมเดลพหุระดับและเทคนิคการอนุมานเชิงสาเหตุ วิธีการเหล่านี้ได้รับการเสริมด้วยโมเดลทางสถิติใหม่ที่แสดงด้วยการสร้างโมเดลสมการเชิงโครงสร้างฐานความแปรปรวนร่วม (covarianc based SEM) หรือฐานความแปรปรวน (PLS-SEM) ซึ่งเป็นวิธีการผสานวิธีการก่อนหน้านี้หลาย ๆ วิธี เช่น การถดถอยเชิงพหุคูณ การวิเคราะห์เส้นทาง และการวิเคราะห์องค์ประกอบเชิงสำรวจเข้ากับเทคนิคการวิเคราะห์ใหม่ ๆ ทำให้ได้เทคนิคการวิเคราะห์ที่ประยุกต์ใช้ได้อย่างหลากหลายมากขึ้น ผลคือทำให้ได้เทคนิคใหม่ ๆ ที่ลดข้อจำกัดของเทคนิคแบบดั้งเดิมที่เคยมีมา

การพัฒนาวิธีการวิเคราะห์ขั้นสูงเหล่านี้ได้รับการสนุบสนุนอย่างมากจากพลังการประมวลผลที่เพิ่มขึ้นอย่างมาก ซึ่งมีอยู่ในรูปแบบและแพลตฟอร์มต่าง ๆ มากมายในปัจจุบัน ความพร้อมใช้งานของเทคนิคเหล่านี้ยังขยายตัวออกไปไม่เพียงแค่ผ่านการพัฒนาอย่างต่อเนื่องของแพ็คเกจซอฟต์แวร์แบบดั้งเดิม เช่น SAS และ JMP, IBM SPSS และ STATA ซึ่งเป็นซอฟต์แวร์ที่ได้รับความนิยม MPlus สำหรับ SEM รวมถึง SmartPLS และ ADANCO สำหรับ PLS-SEM นอกจากนี้ยังมีซอฟต์แวร์โอเพนซอร์สฟรีที่แพร่หลายในปัจจุบัน ได้แก่ R project ซึ่งมีมาตั้งแต่ปี 1992 และได้รับพัฒนาให้ใช้งานง่ายขึ้นในแพลตฟอร์ม RStudio ดังนั้น ปัจจุบัน นักวิจัยมีทางเลือกซอฟต์แวร์ที่หลากหลายที่สุดเท่าที่มีมา

สำหรับหนังสือเล่มนี้เขียนขึ้นโดยมีความเชื่อมโยงกับรายวิชาการวิเคราะห์ข้อมูลขั้นสูงสำหรับการวิจัยทางการจัดการ (Advance Analysis in Management Research) ในการศึกษาระดับปริญญาเอก สาขาการจัดการ คณะวิทยาการจัดการ มหาวิทยาลัยสงขลานครินทร์ โดยยังคงแนวทางการวิเคราะห์ดั้งเดิมและผสมเทคนิคใหม่ในบางเทคนิค โดยมุ่งเน้นการออกแบบ การประมาณค่าและการตีความ โดยพยายามใช้สัญลักษณะทางสถิติและคำศัพท์เฉพาะที่จำเป็น และใช้แนวคิดพื้นฐานที่ส่งผลต่อการประยุกต์ใช้เทคนิคเหล่านี้ ผู้เขียนได้ใช้โปรแกรม RStuio มาใช้ในหนังสือนี้ทั้งเล่มด้วยเหตุผลหลายประการ ได้แก่ เป็นโอเพนซอร์สทำให้เป็นโปรแกรมฟรี มีแนวทางและตัวอย่างการใช้คำสั่งที่หาได้ง่ายบนอินเทอร์เนต มีชุดโปรแกรม (package) จำนวนมากและพัฒนาอย่างต่อเนื่อง และที่สำคัญคือสามารถสร้างข้อมูลภาพ (data vistualization) ซึ่งจะช่วยให้นักวิจัยแปลความหมายทางสถิติได้ดีขึ้น หนังสือเล่มนี้พยายามเสนอในแนวทางที่เรียบง่าย ทำให้มีความเหมาะสมสำหรับการวิเคราะห์พหุตัวแปรที่เน้นการประยุกต์ใช้สำหรับผู้ที่ไม่ใช่นักสถิติได้เป็นอย่างดี จากประสบการณ์ในการทดลองการใช้หนังสือนี้กับนักศึกษาปริญญาเอกไม่น้อยกว่าห้าปี จากนักศึกษาที่ไม่พื้นฐานการใช้สถิติขั้นสูงและการใช้ R หรือ RStuido มาก่อน พบว่านักศึกษาสามารถเรียนรู้โดยใช้เวลาไม่นานภายใต้คำแนะนำเพียงเล็กน้อยเท่านั้น นอกจากนี้นักวิจัยสามารถดาวน์โหลดคำสั่งและข้อมูล (chapter resource) สำหรับการวิเคราะห์ในแต่ละบทได้จาก QR code ที่ผู้เขียนได้เตรียมไว้ในแต่ละบท

ผู้เขียนมีความมุ่งมั่นจะสร้างความเข้าใจที่มั่นคงเกี่ยวกับหลักการทางสถิติและการจัดการที่เป็นพื้นฐานของการวิเคราะห์ขั้นสูงพหุตัวแปรในแบบที่ไม่สร้างความหนักใจกับนักวิจัยมากเกินไปและหวังว่าหนังสือเล่มนี้จะเป็นประโยชน์สำหรับนักวิจัยในทุกระดับ

อิศรัฏฐ์ รินไธสง

คำนิยม

การวิจัยทางพฤติกรรมศาสตร์เป็นศาสตร์เชิงบูรณาการที่ผสมผสานองค์ความรู้จากหลายสาขา อาทิ จิตวิทยา สังคมวิทยา การบริหาร และเศรษฐศาสตร์ เพื่ออธิบายพฤติกรรมของมนุษย์อย่างรอบด้านและลึกซึ้ง จำเป็นต้องอาศัยเครื่องมือทางสถิติขั้นสูงที่มีความแม่นยำและยืดหยุ่นในการวิเคราะห์ข้อมูล หนังสือเล่มนี้จึงมีบทบาทสำคัญในการเป็นสะพานเชื่อมระหว่างแนวคิดเชิงทฤษฎีจากหลากหลายศาสตร์กับการประยุกต์ใช้วิธีการทางสถิติอย่างเป็นระบบ ช่วยให้นักวิจัยสามารถสังเคราะห์องค์ความรู้และสร้างข้อค้นพบใหม่ที่มีคุณค่าเชิงวิชาการและเชิงปฏิบัติได้อย่างแท้จริง

จุดเด่นของหนังสือเล่มนี้คือการเชื่อมโยงระหว่าง “ทฤษฎี” และ “การประยุกต์ใช้จริง” ได้อย่างชัดเจน ผู้อ่านจะไม่เพียงเข้าใจหลักการทางสถิติเท่านั้น แต่ยังสามารถนำความรู้นี้ไปใช้ในการวิเคราะห์ข้อมูลและตอบคำถามวิจัยได้อย่างถูกต้องและมีประสิทธิภาพ โดยเฉพาะการนำเสนอขั้นตอนการวิเคราะห์ การตรวจสอบสมมติฐาน และการแปลผล ซึ่งเป็นหัวใจสำคัญของงานวิจัยที่มีคุณภาพ

นอกจากนี้ การเลือกใช้โปรแกรม R เป็นเครื่องมือหลักในการวิเคราะห์ข้อมูลถือเป็นอีกหนึ่งจุดแข็งของหนังสือเล่มนี้ เนื่องจากเป็นเครื่องมือที่มีประสิทธิภาพสูงและเปิดโอกาสให้ผู้เรียนสามารถเข้าถึงได้โดยไม่มีข้อจำกัดด้านลิขสิทธิ์ ส่งเสริมให้เกิดการเรียนรู้และพัฒนาทักษะอย่างยั่งยืน ข้าพเจ้ามีความเชื่อมั่นว่าหนังสือเล่มนี้จะเป็นแหล่งความรู้ที่มีคุณค่าและเป็นประโยชน์อย่างยิ่งต่อ นักวิจัยทางพฤติกรรมศาสตร์ และผู้อ่านทุกท่าน ทั้งในการเพิ่มความรู้และทักษะทางสถิติ และการนำไปประยุกต์ใช้ในงานวิจัยอย่างแท้จริง

ศาสตราจารย์ เกียรติคุณ ดร. ดุษฎี อินทรประเสริฐ
นายกสมาคมพฤติกรรมศาสตร์
2 เมษายน 2569

บทที่ 1

บทนำ
Introduction

1. ความนำ

ในบทนำนี้ผู้เขียนมีวัตถุประสงค์หลักอยู่สองประการด้วยกัน ประการแรก ด้วยหนังสือเล่มนี้ประมวลผลด้วย RStudio ทั้งเล่ม มีความจำเป็นยิ่งที่นักวิจัยต้องเข้าใจพื้นฐานการใช้ RStudio ที่เพียงพอต่อการใช้งาน ผู้เขียนจึงเริ่มด้วยทำไมจึงใช้ RStudio ตามด้วยวิธีการติดตั้ง ส่วนประกอบ การกำหนด theme เพื่ออาจทำให้สภาพแวดล้อมการทำงานของ RStudio น่าทำงานมากขึ้น สิ่งสำคัญที่นักวิจัยต้องควรทำเข้าใจคือการทำงานของ RStudio และชุดโปรแกรม (package) ซึ่งหนังสือให้ความเข้าใจพื้นฐานเพียงพอต่อการใช้งาน โดยไม่จำเป็นต้องรู้การเขียนโค้ดมาก่อน นักวิจัยเพียงทำตามที่หนังสือให้แนวทางไว้ การนำโค้ด และชุดโปรแกรมที่ได้ไว้ในหนังสือจะทำให้นักวิจัยเข้าใจมากขึ้น และเมื่อใช้ไปสักระยะ ผู้เขียนมั่นใจว่านักวิจัยจะชื่นชอบการใช้ RStudio

ประการที่สอง ด้วยเหตุว่าการทดสอบทางสถิติไม่อาจหลีกเลี่ยงการจัดการข้อมูล การเตรียมข้อมูล การทดสอบข้อตกลงเบื้องต้น การตรียมข้อมูลมีความจำเป็นอย่างยิ่งทั้งในแง่ของการตรวจสอบและการชำระข้อมูล (screening and cleaning) โดยเฉพาะการจัดการค่าผิดพลาด (missing value) ค่าสุดโต่ง (outlier) เพื่อให้ได้ข้อมูลที่ดีก่อนเข้าการวิเคราะห์ใด ๆ ขณะเดียวกันในการตรวจสอบข้อตกลงเบื้องต้น มักจะมีข้อตกลงเบื้องต้นร่วมกันหลายประการ เช่น การแจกแจงปกติ (normality) ความเป็นเอกพันธ์ของเมทริกซ์ความแปรปรวน - ความแปรปรวนร่วม (homogeneity of variance) ความคงที่ของความแปรปรวนของความคลาดเคลื่อน (homosedasiticity) เป็นต้น โดยในบทนี้ทำการทดสอบ การแปรความหมาย และแนวทางจัดการไว้เพื่อเป็นแนวทางสำหรับนักวิจัย ส่วนข้อตกลงบางประการ เช่น Sphericity ซึ่งการตรวจสอบเมทริกซ์สหสัมพันธ์ของประชากรว่าเป็นเมทริกซ์เอกลักษณ์หรือไม่ในการวิเคราะห์องค์ประกอบเชิงสำรวจ ความสัมพันธ์เชิงเส้นตรง (linearity) การร่วมกันเชิงเส้นพหุ (collinerity) ซึ่งเป็นเงื่อนไขการวิเคราะห์การถดถอยพหุคูณ ผู้เขียนจะให้แนวคิด (concept) ไว้แต่การทดสอบจะนำเสนอในแต่ละบทที่เกี่ยวข้องเป็นการเฉพาะ

ผู้เขียนเห็นว่านักวิจัยสามารถเลือกอ่านในบทที่ตนเองสนใจได้ เนื่องจากการอธิบายมีความเป็นอิสระของแต่ละบท อาจมีความเกี่ยวข้องกันบ้าง เช่น การวิเคราะห์ฟังก์ชันการจำแนก (discriminnat function analysis) กับการวิเคราะห์ความแปรปรวนพหุตัวแปร (MANOVA) แต่โดยรวมแล้วยังสามารถเลือกก่อนได้ตามความสนใจ ยกเว้นนักวิจัยที่ไม่มีพื้นฐาน R หรือ RStudio มาก่อน แนะนำว่ามีความจำเป็นต้องอ่านบทที่ 1 นี้ และหากให้เกิดความเชี่ยวชาญควรศึกษาจากแหล่งอื่นเพิ่มเติมที่สามารถสืบค้นโดยง่ายในออนไลน์

2. แนะนำ R และ RStudio

2.1 ทำไมต้องต้อง R และ RStudio

R เป็นภาษาโปรแกรมโอเพนซอร์สที่ได้รับความนิยมภาษาหนึ่ง หากนักวิจัยเคยมีความรู้พื้นฐานด้านการเขียนโปรแกรมใน Python, Swift หรือ C มาก่อนแล้ว นักวิจัยสามารถเรียนรู้ไวยากรณ์ของ R ได้อย่างรวดเร็วหาก นักวิจัยเพิ่งเริ่มใช้ R ก็มีแหล่งข้อมูลออนไลน์มากมายที่น่าสนใจและฟรีให้นักวิจัยเรียนรู้ อันที่จริงแล้ว R ไม่เพียงแต่มีชื่อเสียงว่าเป็นภาษาโปรแกรมที่เรียนรู้ได้ง่ายกว่า Python เท่านั้น นักพัฒนา R มีช่อง YouTube ยอดนิยมที่ให้คำแนะนำการเขียนโค้ดแบบสดโดยใช้ R และ RStudio ขณะเดียวกันมีหนังสือโอเพนซอร์ส R แล RStudio จำนวนมาก นอกจากนี้ยังมีแหล่งข้อมูลนคุณวิจัยภาพสูง และฟรีอีกมากมายเกี่ยวกับการสร้างกราฟิกที่สวยงาม

RStudio เป็นแอปพลิเคชันที่เขียนด้วยคำสั่งแบบ R ผู้เขียนใช้ RStudio ตลอดทั้งเล่มเพราะทำให้การใช้ R ง่ายขึ้นมาก นอกจากนี้ อินเทอร์เฟซ RStudio ก็มีลักษณะเหมือนกันสำหรับ Windows, Mac OS และ Linux เมื่อติดตั้ง RStudio หรือแม้แต่ RStudio online นักวิจัยสามารถเปิดได้เช่นเดียวกับโปรแกรมอื่น ๆ บนคอมพิวเตอร์ โดยปกติแล้วจะเปิดโดยคลิกไอคอนบนเดสก์ท็อป RStudio ได้กลายเป็นสภาพแวดล้อมการพัฒนาแบบบูรณาการ (integrated development environment [IDE]) ที่ได้รับความนิยมสูงสุดสำหรับผู้ใช้ R นับตั้งแต่เปิดตัวในปี 2011 โดย Posit ซึ่งเป็นบริษัทวิทยาศาสตร์ข้อมูลโอเพนซอร์ส

การที่ RStudio ได้รับการนำมาใช้ในวงกว้างโดยชุมชนการวิเคราะห์ข้อมูลนั้นเกิดจากการที่ RStudio นำเสนอแนวทางแบบบูรณาการ และเรียบง่ายแก่ผู้ใช้ในการดำเนินการวิเคราะห์ข้อมูลการแสดงภาพ และการสร้างโมเดลทางสถิติ ไม่ว่าจะเป็นนักวิจัย นักวิเคราะห์ข้อมูล หรือสถิติศาสตร์ อินเทอร์เฟซที่เป็นมิตรกับผู้ใช้ในการเรียนรู้และใช้งานได้ง่าย แม้ว่า RStudio จะหมายถึง IDE ที่ใช้เมื่อเขียนโค้ดใน R แต่จะเข้าใจได้ดีกว่าว่าเป็นชุดเครื่องมือที่ช่วยให้นักวิเคราะห์จัดการแสดงภาพสร้างแบบจำลองข้อมูล และปรับใช้แบบจำลองการเรียนรู้ของเครื่องจักร (machine learning) RStudio คือโปรแกรมแก้ไขโค้ดที่มาพร้อมกับเครื่องมือเน้นไวยากรณ์ เติมโค้ด และดีบัก (debug) จึงเป็นเครื่องมือที่นักวิจัยสามารถเขียนโค้ด R ได้โดยตรง ด้วยคุณสมบัติเหล่านี้จะทำให้กระบวนการเขียนโค้ดราบรื่นและมีประสิทธิภาพมากขึ้น ซึ่งจะกลายเป็นสิ่งสำคัญมากขึ้นเมื่อฐานของโค้ดมีความซับซ้อนมากขึ้น นอกจากนี้ยังมาพร้อมกับคอนโซล (console) แบบโต้ตอบที่ให้นักวิจัยเรียกใช้สคริปต์ (script) ทั้งหมดของโค้ด R เพื่อดูผลลัพธ์แบบเรียลไทม์

RStudio มีเบราว์เซอร์พื้นที่ทำงานที่ติดตามตัวแปร ฟังก์ชัน รายการ และเฟรมข้อมูลที่ใช้ในสภาพแวดล้อมปัจจุบันของนักวิจัย RStudio มีหน้าต่างการพล็อต (plot) ในตัวที่แสดงพล็อตใด ๆ ที่นักวิจัยสร้างขึ้นขณะทำการวิเคราะห์ข้อมูลเชิงสำรวจ นักวิจัยสามารถแก้ไขและบันทึกพล็อตเหล่านี้ได้โดยตรง นอกจากนี้ RStudio สามารถบูรณาการกับเครื่องมืออื่น ๆ ได้ดีตัวอย่างเช่น นักวิจัยสามารถนำการควบคุมเวอร์ชันมาใช้กับ Git ซึ่งช่วยให้นักวิจัยติดตามและจัดการการเปลี่ยนแปลงของโค้ดในช่วงเวลาหนึ่ง และกับคนเขียนโค้ด R หลายคนที่ทำงานในโปรเจ็กต์เดียวกันได้ นอกจากนี้ยังรองรับ Shiny อีกด้วย ดังนั้นนักวิจัยจึงสร้างแอปพลิเคชันเว็บ หรือแดชบอร์ดแบบโต้ตอบใน R ได้โดยไม่ต้องรู้เรื่องการพัฒนาหรือการปรับใช้เว็บใด ๆ ทั้งสิ้น RStudio นอกจากทำงานบนเดสก์ท็อปแล้วยังสามารถทำงานบนแพลตฟอร์มออนไลน์ได้

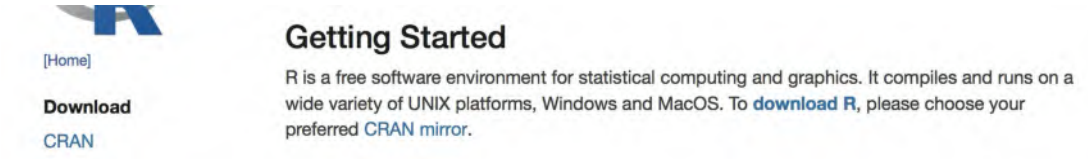
ที่กล่าวมาเป็นความสามารถของ RStudio แต่สำหรับหนังสือเล่มนี้นักวิจัยไม่จำเป็นต้องมีทักษะขั้นสูงใด ๆ เพียงแต่เข้าใจการใช้พื้นฐาน ส่วนการเขียนคำสั่ง ฟังก์ชันต่าง ๆ นักวิจัยสามารถใช้ตามหนังสือนี้ ได้เลย และยังสามารถมีทางเลือกอื่น ๆ สำหรับการวิเคราะห์สถิติเดียวกันและด้วยฟังก์ชันจาก package ที่ต่างกันด้วยการเรียนรู้จากเว็บไซต์ต่าง ๆ จำนวนมาก ในหนังสือนี้จะใช้ RStudio สำหรับการวิเคราะห์ข้อมูล การสร้างข้อมูลภาพ และการแปลความหมายทางสถิติเป็นหลัก

2.2 การติดตั้ง R และ RStudio

RStudio IDE คือ Integrated Development Environment โปรแกรมที่ทำงานร่วมกับโปรแกรม R ซึ่งเอื้อ ให้เหมาะสมกับการใช้งานและการพัฒนาโปรแกรมให้สะดวกมากขึ้น เช่น การทำงานร่วมกับ Markdown, Latex เป็นต้น ข้อดีของ RStudio ยังสามารถทำงานร่วมกับ Version Control เช่น Git และ Github อีกด้วย ซึ่งเหมาะกับงานทางด้าน Data Scientists อย่างยิ่ง โดยการใช้งานโปรแกรม RStudio นั้น ผู้ใช้จะต้องทำการติดตั้งอยู่สองขั้นตอนด้วยกัน คือ การติดตั้งโปรแกรม R และการติดตั้งโปรแกรม RStudio

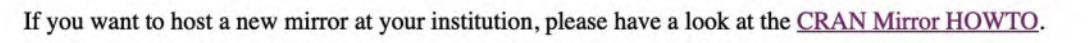
2.2.1 ติดตั้ง R ไปที่ http://www.r-project.org

2.2.2 คลิก ที่ CRAN หัวข้อ Download หรือ คลิกที่ Download R หรือ CRAN Mirror ในข้อความ Getting Started



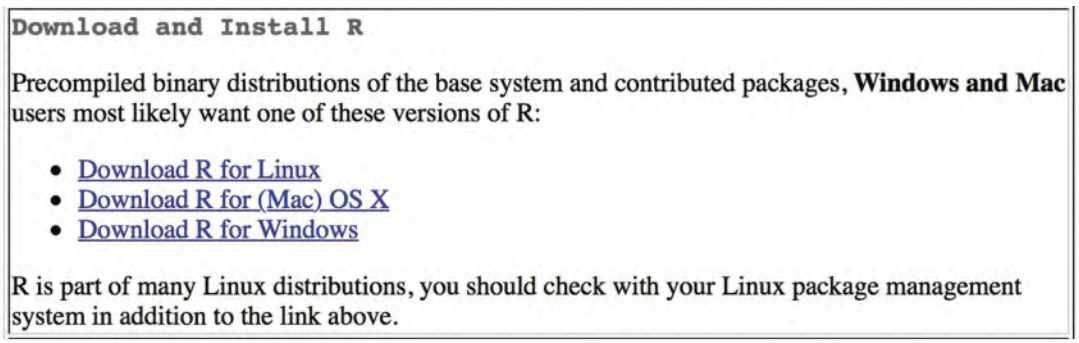
2.2.3 คลิก Link ใต้ข้อความ Thailand

กรณีไม่สามารถติดตั้งผ่านประเทศไทยได้ ให้ไปที่ด้านบนจะพบข้อความดังข้างล่าง ให้คลิกที่ CRAN Mirror HowTo

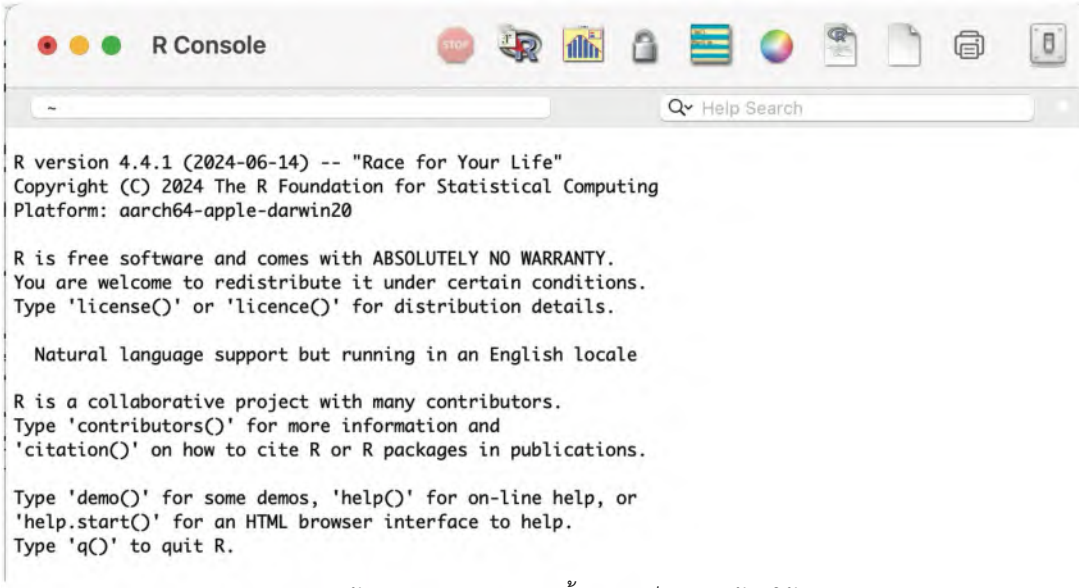


จะไปที่เพจ คลิกที่ https://cran.r-project.org

2.2.4 เลือก version และ OS ที่เหมาะกับเครื่อง หลังจากนั้นเข้าสู่กระบวนการติดตั้งปกติ



โปรแกรม R ที่พร้อมใช้งาน



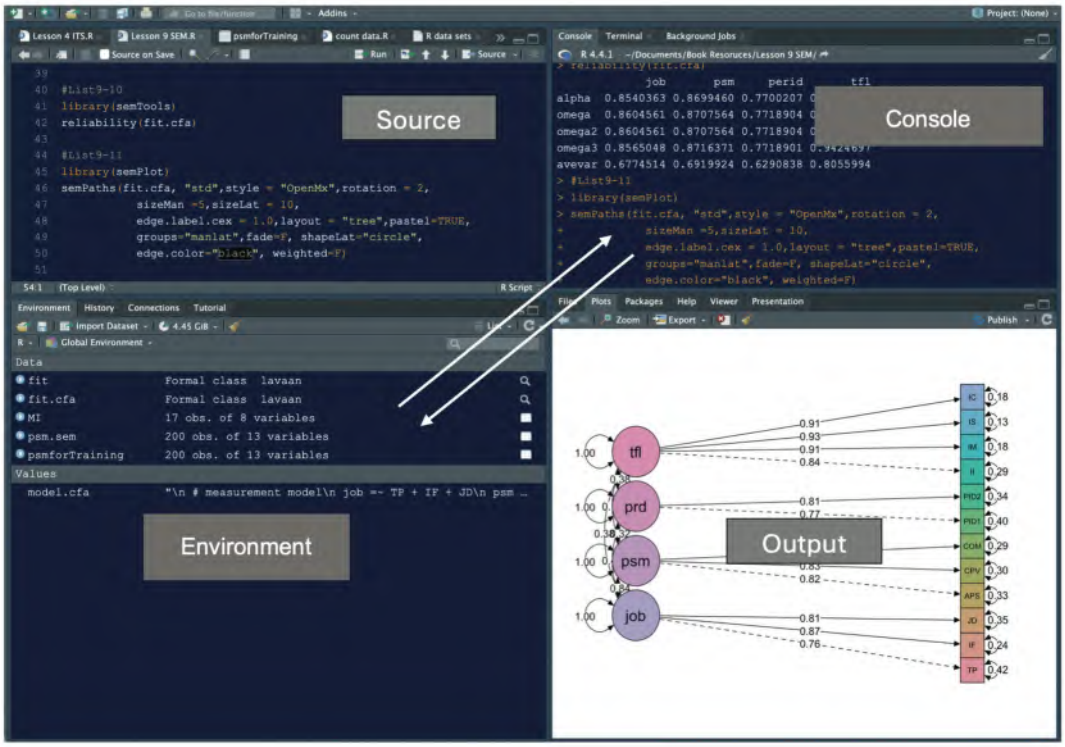
ภาพ 1-1 หน้าต่างแสดงการติดติดตั้ง R สำเร็จและพร้อมใช้งาน

2.2.5 การติดตั้ง Rstudio

RStudio เป็นแอปพลิเคชันเช่นเดียวกับโปรแกรมทั่วไปบนคอมพิวเตอร์ แต่ RStudio จะช่วยนักวิจัยเขียนคำสั่งด้วย R ผู้เขียนใช้ RStudio ตลอดทั้งเล่มเพราะทำให้การใช้ R ง่ายขึ้นมาก นอกจากนี้ อินเทอร์เฟซ RStudio ก็มีลักษณะเหมือนกันสำหรับ Windows, Mac OS และ Linux ซึ่งจะช่วยให้ฉันจับคู่หนังสือกับประสบการณ์ส่วนตัวของนักวิจัยได้ดาวน์โหลด RStudio ได้ฟรีที่เว็บไซต์ http://www.rstudio.com/ide/download/ หรือ https://posit.co/downloads/ แล้วคลิกที่ปุ่ม DOWNLOAD RSTUDIO แล้วทำตามคำแนะนำ เมื่อนักวิจัยติดตั้ง RStudio แล้ว เราสามารถเปิดใช้งานได้เช่นเดียวกับโปรแกรมอื่น ๆ โดยปกติแล้วจะเปิดโดยคลิกไอคอนบนเดสก์ท็อป

ภาพ 1-2 การดาวน์โหลด RStudio และ icons ของ R และ RStudio ที่พร้อมใช้งาน

2.3 User interface ของ RStuido



ภาพ 1-3 ส่วนประกอบของ User interface ของ RStudio

User interface ของ RStudio มีแผงหลัก 4 แผง (pane) ได้แก่

2.3.1 Source pane

Source pane เป็นพื้นที่สำหรับเขียนคำสั่ง หรือแหล่งที่มาข้อมูลและคำสั่ง (script) แผงซอร์ส ช่วยให้ผู้ใช้สามารถดูและแก้ไขไฟล์ที่เกี่ยวข้องกับโค้ดต่าง ๆ เช่น .R, .rmd, .qmd, .py, .css หรือไฟล์ข้อความทั่วไป เช่น .txt หรือ .md โดยค่าเริ่มต้น แผงซอร์สจะอยู่ด้านบนซ้าย และสามารถเปิดใช้งานได้โดยการเปิดไฟล์ที่แก้ไขได้ใน RStudio ไฟล์เพิ่มเติมที่เปิดขึ้นแต่ละไฟล์จะถูกเพิ่มเป็นแท็บใหม่ภายในแผงซอร์ส สามารถเพิ่ม “คอลัมน์” ซอร์สเพิ่มเติมทางด้านซ้ายของแผงหลักทั้งสี่แผงได้หากต้องการเพิ่มคอลัมน์ซอร์สเพิ่มเติม ให้ไปที่เมนู Tools->Global Options -> Pane Layout -> Add Source Column ซึ่งให้ความสามารถในการทำงานกับไฟล์ซอร์สสองไฟล์ (หรือมากกว่า) พร้อมกันภายในแผงซอร์ส

ไฟล์ซอร์สแต่ละไฟล์สามารถเปิดในหน้าต่างของตัวเองได้ แทนที่จะเปิดเฉพาะในแผงซอร์สเท่านั้น หน้าต่างซอร์สช่วยให้นักวิจัยแก้ไขไฟล์นอกหน้าต่างหลักของ RStudio ได้ ซึ่งมีประโยชน์ในการแบ่งงานของนักวิจัยระหว่างมอนิเตอร์หลายตัว หรือเพิ่มพื้นที่ให้กับโปรแกรมแก้ไข ในหนังสือนี้จะเขียนคำสั่งจากแผงซอร์สทั้งหมด

จากภาพ 1-3 Source pane แสดงตัวอย่างการเรียกใช้ semTool package และ semPlot package จากคำสั่ง library และใช้ฟังก์ชัน reliability จะแสดงผลใน Console pane สร้างแผนภาพจากฟังก์ชัน semPaths แสดงผลใน Output pane โดยสั่ง run คำสั่งจากปุ่ม run จากเมนูบาร์ด้านบน

2.3.2 Console pane

Console pane ทำหน้าที่เช่นเดียวกับ console ใน R ซึ่งสามารถเขียนคำสั่ง และสั่ง run ได้ โดยค่าเริ่มต้นแล้ว แผงคอนโซลจะเป็นแผงด้านล่างซ้าย แผงคอนโซลจะให้พื้นที่สำหรับเรียกใช้โค้ดแบบโต้ตอบผลจากการ run คำสั่งจาก Sources pane โดยปกติแล้วแผงคอนโซลจะเชื่อมโยงกับ R ด้วยการใช้แพ็คเกจ reticulate package แผงคอนโซลยังสามารถจัดเตรียมคอนโซล Python ได้ด้วย นอกจากนี้ แผงคอนโซลยังรวมถึงแท็บเทอร์มินัลในตัวสำหรับเรียกใช้คำสั่งระบบเพิ่มหรือลบเทอร์มินัลในตัวเพิ่มเติม รวมถึงการควบคุมทั่วไปของเทอร์มินัลที่เลือกในปัจจุบัน แท็บงานพื้นหลังให้ความสามารถในการส่งสคริปต์ R ที่รันยาวนานไปยังงานเบื้องหลังในพื้นที่และระยะไกล ฟังก์ชันนี้จะช่วยปรับปรุงประสิทธิภาพของนักวิทยาศาสตร์ข้อมูล และนักวิเคราะห์ที่ใช้ R ได้อย่างมาก เนื่องจากพวกเขาสามารถทำงานใน RStudio ต่อไปได้ในขณะที่งานกำลังทำงานอยู่เบื้องหลัง

2.3.3 Environtments pane

Environtments pane โดยค่าเริ่มต้น แผงสภาพแวดล้อมจะอยู่ที่ด้านบนขวา และประกอบด้วยแท็บสภาพแวดล้อม ประวัติ การเชื่อมต่อ การสร้าง และระบบควบคุมเวอร์ชัน (VCS)

Environtments tap

แท็บสภาพแวดล้อมจะแสดงวัตถุ R ที่บันทึกไว้ในปัจจุบัน กล่าวคือเมื่อเราเขียนโค้ดในแผงซอร์ส และสั่ง run โค้ดจะถึง เขียนลงในแผงสภาพแวดล้อมตลอดระยะเวลาของเซสชัน (session) ปัจจุบัน แถบเมนูสภาพแวดล้อมช่วยให้โหลดหรือบันทึกเวิร์กสเปซ R (workspace) นำเข้าชุดข้อมูลแบบโต้ตอบจากไฟล์ข้อความ Excel/SPSS/SAS/Stata หรืออื่น นอกจากนี้ยังแสดงหน่วยความจำที่ใช้โดยเซสชัน R ที่ใช้งานอยู่ และไอคอนไม้กวาดเพื่อลบวัตถุสภาพแวดล้อมปัจจุบันทั้งหมด

History tap

แท็บประวัติจะแสดงคำสั่งที่ดำเนินการในเซสชันปัจจุบันพร้อมกับการค้นหา มีปุ่มสำหรับโหลด/บันทึกประวัติคำสั่งไปยังไฟล์ รวมถึงการส่งคำสั่งที่เลือกไปยังคอนโซล หรือแทรกคำสั่งนั้นลงในเอกสารการทำงานปัจจุบัน มีปุ่มลบสำหรับลบประวัติที่เลือกหรือปุ่มไม้กวาดสำหรับลบประวัติทั้งหมดของเซสชันปัจจุบัน

Connection tap

แท็บการเชื่อมต่อจะแสดงการเชื่อมต่อกับฐานข้อมูลในพื้นที่หรือระยะไกล สามารถเพิ่มการเชื่อมต่อที่เป็นไปได้เพิ่มเติมได้โดยการติดตั้งไดรเวอร์ฐานข้อมูลหรือแพ็กเกจ R เฉพาะ นักวิจัยสามารถคลิกการเชื่อมต่อที่พร้อมใช้งานแบบโต้ตอบได้ และจะมีโค้ด R ทั่วไปให้เพื่อลงทะเบียนการเชื่อมต่อที่ใช้งานอยู่กับแหล่งข้อมูลนั้น เมื่อสร้างการเชื่อมต่อแล้ว นักวิจัยสามารถสำรวจตารางที่พร้อมใช้งานในการเชื่อมต่อนั้นได้ นอกจากนี้ยังมีปุ่มเชื่อมต่อสำหรับสร้างการเชื่อมต่อในคอนโซล R สคริปต์ หรือสมุดบันทึก รวมถึงการคัดลอกโค้ดที่เกี่ยวข้องไปยังคลิปบอร์ดได้หากมี

2.3.4 Output pane

Output pane โดยค่าเริ่มต้น แผงผลลัพธ์จะเป็นแผงด้านล่างขวา และจะแสดงผลลัพธ์ต่าง ๆ เช่น พล็อต เนื้อหา HTML หรือไฟล์บนดิสก์ แผงนี้ประกอบด้วยแท็บ ไฟล์ พล็อต แพ็กเกจ R วิธีใช้ บทช่วยสอน โปรแกรมดู และการนำเสนอ

Files tap

แท็บนี้ให้การสำรวจแบบโต้ตอบของโปรเจ็กต์ R ปัจจุบันพร้อมกับไดเร็กทอรีทั้งหมด มีตัวเลือกแถบเมนูสำหรับการเพิ่มโฟลเดอร์ใหม่ ไฟล์ว่างใหม่ การลบ/เปลี่ยนชื่อไฟล์ ตลอดจนแผงรูปเฟืองสำหรับฟังก์ชันเพิ่มเติม

Plots tap

แท็บนี้จะแสดงรูปภาพคงที่ที่สร้างโดยโค้ดจนกว่าจะเริ่มเซสชันใหม่ มีลูกศรย้อนกลับและไปข้างหน้าสำหรับการนำทางระหว่างพล็อตเก่าและใหม่กว่า คุณลักษณะซูม ปุ่มส่งออกสำหรับบันทึกผลลัพธ์ที่แสดง ปุ่มลบสำหรับลบภาพที่แสดงอยู่ในปัจจุบัน และไอคอนไม้กวาดที่จะล้างพล็อตชั่วคราวทั้งหมดจากแท็บนี้ ซึ่งจากภาพจะเป็นการแสดงแผนภาพเมทริกซ์สหสัมพันธ์จากคำสั่งในแผงซอร์ส

Packages tap

แท็บแพ็คเกจช่วยให้ดูแพ็คเกจ R ที่ติดตั้งอยู่ในปัจจุบันได้ และมีแถบค้นหาสำหรับค้นหาไลบรารีแพ็คเกจปัจจุบัน มีปุ่มติดตั้งและอัปเดตสำหรับการติดตั้งแพ็คเกจใหม่หรืออัปเดตแพ็คเกจที่เลือกไว้แล้ว

Help tap

แท็บความช่วยเหลือใช้เพื่อแสดงเอกสารแพ็คเกจและคำอธิบายสั้น ๆ มีลูกศรสำหรับนำทางไปข้างหน้าและข้างหลังเมื่อดูหน้าความช่วยเหลือเพิ่มเติม ไอคอนหน้าแรกจะกลับไปที่หน้าความช่วยเหลือทั่วไปพร้อมลิงก์ไปยังทรัพยากร คู่มือ การอ้างอิง และการสนับสนุน Posit ในการรับความช่วยเหลือเกี่ยวกับฟังก์ชัน R เฉพาะ ให้ใช้รูปแบบ ?functionName หรือ help(functionName) ตัวอย่างเช่น ? paste เมื่อสั่ง run จะเปิดหน้าความช่วยเหลือสำหรับฟังก์ชัน paste()

Tutorial tap

แท็บบทช่วยสอนใช้สำหรับโหลดบทช่วยสอนการเรียนรู้แบบโต้ตอบ บทช่วยสอนเหล่านี้มอบสภาพแวดล้อมแบบโต้ตอบเพื่อก้าวผ่านบทเรียนที่ผู้เขียนแพ็คเกจหรือผู้สอนเขียนขึ้น

Viewer tap

แท็บตัวแสดงใช้เพื่อแสดงเนื้อหาเว็บ เช่น แอป Shiny หน้าเว็บที่สร้างโดย Quarto หรือ วิดเจ็ต HTML แบบโต้ตอบ เช่น กราฟแบบ Plotly

Presentation tap

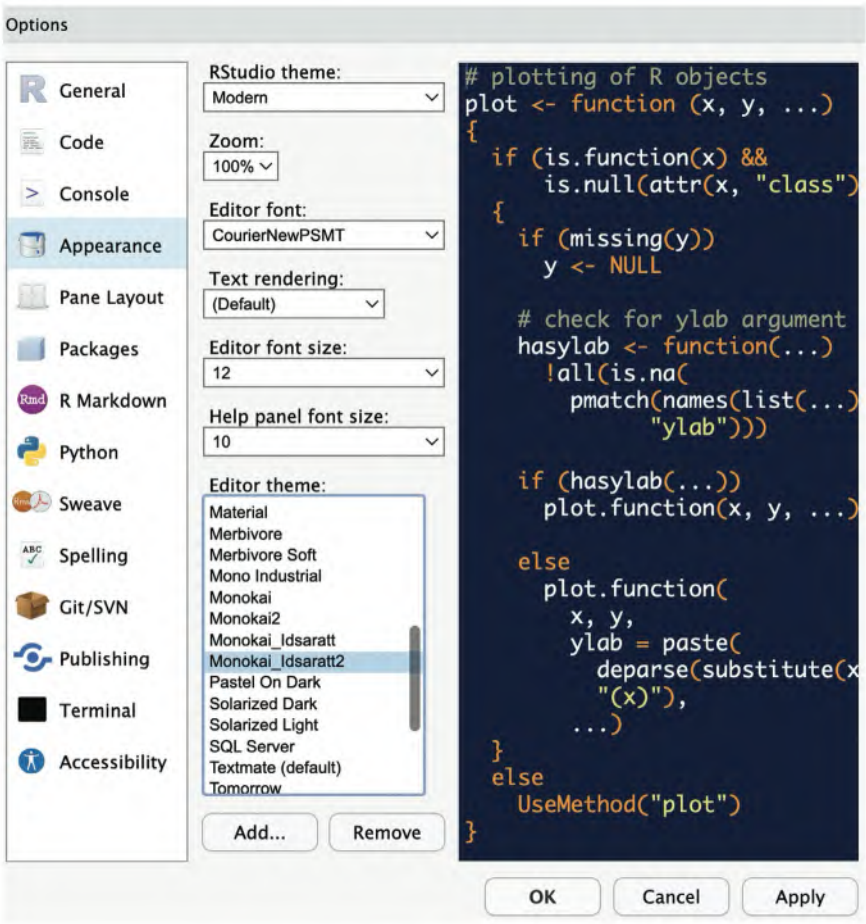
แท็บการนำเสนอใช้เพื่อแสดงสไลด์ HTML ที่สร้างโดยใช้รูปแบบ revealjs ของ Quarto

2.4 การกำหนด theme

ในส่วน Windows ของ RStudio เราสามารถกำหนด themes ได้โดยไปที่ Tools > Global Options... เลือก Appearance แล้วกำหนดลักษณะต่าง ๆ ตามที่ต้องการ ในส่วน Editor theme จะเกี่ยวข้องกับสีพื้น ตัวหนังสือที่ถูกออกแบบไว้แล้ว ในตัวอย่างผู้เขียนเลือก Monokai_Idsaratt2 ซึ่งเป็น theme ผู้เขียนกำหนดขึ้นเอง นอกเหนือไปจาก theme มาตรฐานที่มีอยู่

2.5 การทำงานของ RStudio

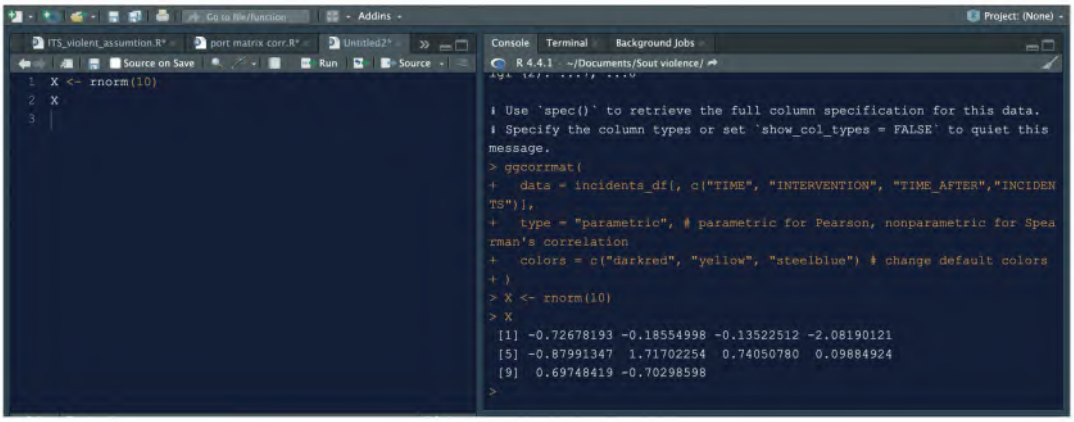
โปรแกรม RStudio จะทำงานก็ต่อเมื่อเรา ป้อนคำสั่งเข้าไป และคำสั่งที่ป้อนเข้าไป โดยโปรแกรมจะมีความไวต่อคำสั่งค่อนข้างมาก กล่าวคือ การเขียนคำสั่งจะต้องอยู่ในรูปของภาษาที่กำหนดขึ้นมาเฉพาะ R นักวิจัยสามารถป้อนคำสั่งในส่วน Source pane หรือเป็นชุดคำสั่งจากแฟ้มที่เตรียมไว้ (script.R) ส่วนข้อมูลสำหรับใช้กับโปรแกรม RStudio มีหลากหลายรูปแบบ เช่น ในรูปของเวคเตอร์ (Vector) เมทริกซ์ (metrics) ดาต้าเฟรม (data frames) ซึ่งเป็นชุดข้อมูล เป็นต้น



ภาพ 1-4 การปรับ theme

การทำงานของโปรแกรม RStudio จะมีการกำหนดให้สิ่งที่เราต้องการคำนวณหรือประมวลผลอยู่ในรูปของวัตถุ (objects) โดยเราสามารถกำหนดค่าต่าง ๆ ให้กับวัตถุนี้ได้ เช่น ข้อมูล ฟังก์ชัน กราฟ ผลการวิเคราะห์ และอื่น ๆ ทุก ๆ วัตถุจะต้องบอกว่าเราจะทำอะไรกับมัน ทุกวัตถุจะถูกบันทึกไว้ในหน่วยความจำ ตลอดช่วงการวิเคราะห์ข้อมูลของเรา เราสามารถเรียกฟังก์ชันพื้นฐานขึ้นมาใช้ได้โดยปกติ แต่ถ้าฟังก์ชันที่เป็นการเฉพาะ เช่น การวิเคราะห์สถิติขั้นสูง หรือการสร้างภาพจากโมเดลจำเป็นต้องใช้ชุดโปรแกรมโดยเฉพาะที่เรียกว่า packages ซึ่งเราจะต้องเรียก packages นั้นมาใช้ก่อนทำการวิเคราะห์ข้อมูล แต่ถ้าในโปรแกรมไม่มี packages นั้น นักวิจัยต้องติดตั้ง (install) เสียก่อนซึ่งทำได้ง่าย เพียงใช้คำสั่ง install.packages () ใส่ชื่อ packages ที่ต้องการติดตั้งใน () โปรแกรมก็จะติดตั้งให้อัตโนมัติ แต่เครื่องคอมพิวเตอร์ของเราต้องเชื่อมต่อกับอินเทอร์เน็ต

ในการกำหนดฟังก์ชันต่าง ๆ นั้น ในโปรแกรม R มักนิยมใช้สัญลักษณ์ <- แทนการใช้เครื่องหมาย = ตัวอย่างคำสั่ง X <- rnom(10)



ภาพ 1-5 คำสั่งใน Source และผลลัพธ์ที่ Console

จากคำสั่ง หมายความว่าเราสร้างเวกเตอร์ข้อมูลให้กับวัตถุ ชื่อ X โดยมีเลขอย่างสุ่ม (random) จำนวน 10 ค่า ที่มีการกระจายแบบแจกแจงปกติ (normal distribution) เราอาจใช้เครื่องหมาย = แทนได้บางกรณี แต่อาจไม่สามารถใช้ได้ทุกฟังก์ชัน จึงไม่แนะนำ (ภาพ 1-5 ผู้เขียนย้าย Console pane จากล่างซ้ายมากอยู่บนขวา เนื่องจากเห็นว่าสะดวกมากกว่า โดยสามารถย้ายได้จาก เมนู View->Panse->Console on Rigth)

สำหรับหนังสือนี้จะมี source code หรือ R script ซึ่งนักวิจัยสามารถดาวน์โหลดและนำเข้าหรือคัดลอกแล้วนำมาวางในส่วน Source pane ได้เลย แต่อาจมีการปรับแต่งเล็กน้อย เช่น การกำหนดพื้นที่ทำงาน ตำแหน่งของข้อมูล การนำเข้าเนื่องจากสภาพแวดล้อมต่างกัน แต่สิ่งเหล่านี้เป็นทักษะพื้นฐานทั่วไปของนักวิจัยที่มีอยู่แล้ว

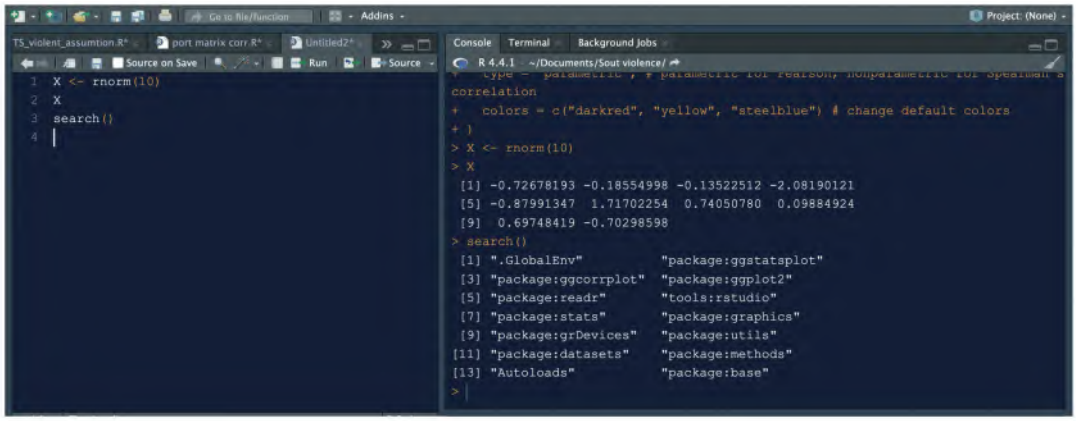
2.6 ชุดโปรแกรม (Packages)

2.6.1 ชุดโปรแกรมคืออะไร

กล่าวได้ว่าชุดโปรแกรมคือการรวบรวมฟังก์ชันของ R ข้อมูล และรวบรวมรหัสให้อยู่ในรูปแบบที่เข้าใจง่าย ไดเรกทอรีที่เราจะเข้าถึงชุดโปรแกรมในเครื่องเราคือ library ซึ่งสามารถใช้ฟังก์ชัน .libPaths() เพื่อดูตำแหน่งของไลบรารีของเราได้ และใช้ฟังก์ชัน library() แสดงว่าชุดโปรแกรมนั้นได้บันทึกในไลบรารีของเราหรือยังรวมทั้งเป็นการเรียกใช้ชุดโปรแกรมนั้นด้วย

RStudio มีความสามารถในการวิเคราะห์นอกเหนือจากฟังก์ชันที่ติดตั้งมากับโปรแกรมบางครั้งการวิเคราะห์ทางสถิติขั้นสูงหรือการวิเคราะห์เฉพาะอย่าง ฟังก์ชันที่มีอยู่อาจไม่สามารถทำได้หรือทำได้ไม่ดี ไม่ครอบคลุมตามความต้องการของผู้ใช้ ซึ่งกรณีเช่นนี้ ผู้ใช้ RStudio สามารถดาวน์โหลดโมดูล (modules) สำหรับการวิเคราะห์ที่สอดคล้องกับความต้องการของเราได้ และนำมาติดตั้งใน RStudio ซึ่งปัจจุบันมีมากกว่า 5,5000 โมดูล โดยเรียกโมดูลเหล่านั้นว่า ชุดโปรแกรม (packages) โดยดาวน์โหลดได้จาก http://cran.r-project.org/web/packages ซึ่งได้จัดเตรียมเพื่อเพิ่มขีดความสามารถให้กับ RStudio และง่ายไปกว่านั้น หากคอมพิวเตอร์เราเชื่อมต่อกับอินเทอร์เน็ต เพียงแต่พิมพ์คำสั่ง install.packages (“ ”) โดยระบุชุดโปรแกรมที่ต้องการติดตั้ง หลังจากนั้นโปรแกรมจะติดตั้งให้อัตโนมัติ ซึ่งจะกล่าวเพิ่มเติมในหัวข้อถัดไป

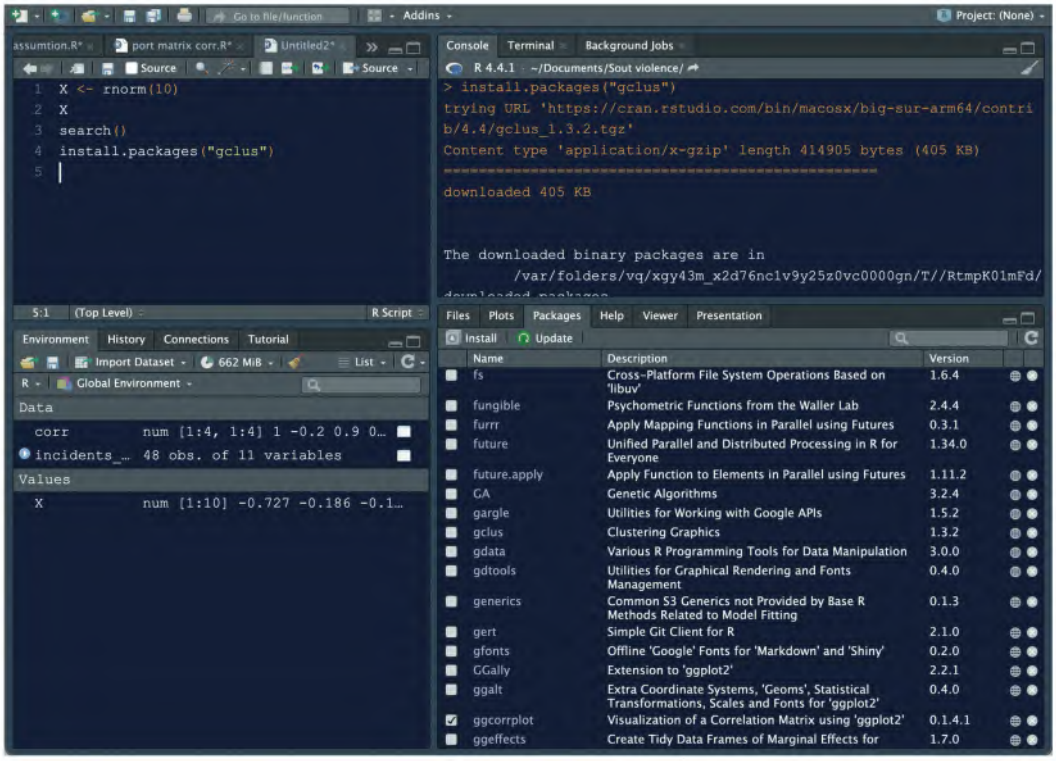
โปรแกรม RStudio จะมาพร้อมกับชุดโปรแกรมมาตรฐาน เช่น base, datasets, utils, grDevices, graphics, stats และ methods ส่วนชุดโปรแกรมอื่นสามารถดาวน์โหลดและติดตั้งเพิ่มเติมได้ และสามารถใช้คำสั่ง search() เพื่อตรวจสอบว่าชุดโปรแกรมนั้นได้ติดตั้งไว้แล้วหรือยัง



ภาพ 1-6 การใช้คำสั่ง search () และผลลัพธ์ที่ Console

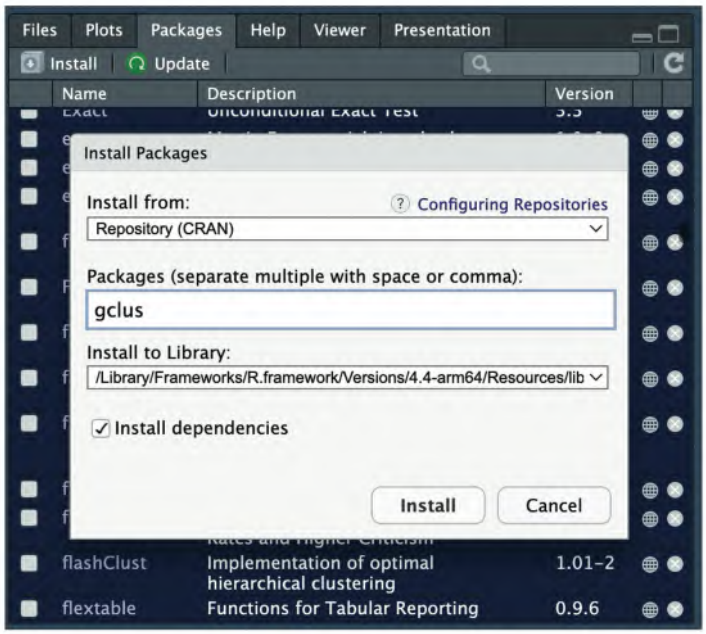
2.6.2 การติดตั้งและการเรียกใช้ชุดโปรแกรม

การติดตั้งชุดโปรแกรม (installing a package) การติดตั้งชุดโปรแกรมครั้งแรกใช้ฟังก์ชัน install.package() เช่นถ้าเราใช้คำสั่ง install.packages() โดยไม่กำหนดว่าจะติดตั้งชุดโปรแกรมใดผลการใช้ฟังก์ชันนี้จะนำเราไปยังเว็บไซต์ของ CRAN mirror แล้วเราเลือกไซต์ที่เราต้องการดาวน์โหลดและติดตั้งชุดโปรแกรม (ไซต์นี้จะเกี่ยวข้องกับประเทศ โดยปกติจะเลือกไซต์ที่ใกล้เรา) ถ้าเราต้องการดาวน์โหลดและติดตั้งเฉพาะโปรแกรม และเราทราบชื่อชุดโปรแกรม เช่น เราต้องการติดตั้งชุดโปรแกรม gclus เราสามารถใช้คำสั่ง install.packages("gclus") โปรแกรมจะดาวน์โหลด และติดตั้งให้เราโดยอัตโนมัติ เมื่อติดตั้งสำเร็จ โปรแกรมจะรายงานในส่วน Console พร้อมกับแสดงในส่วน Output pane ที่ Package tap ดังภาพ



ภาพ 1-7 การติดตั้งชุดโปรแกรม gclus ด้วยคำสั่ง install.package()

นอกจากใช้คำสั่ง install.packages () แล้วยังสามารถติดตั้งจากแผง Output โดยเลือกโดยใช้ Package tap และเลือก install tap หลังจากนั้นพิมพ์ชุดโปรแกรมที่ต้องการติดตั้งลงไปที่ช่อง Packages (separate with space or comma) กรณีที่ต้องการติดตั้งหลายชุดโปรแกรมพร้อมกันให้คั่นด้วยเครื่องหมายคอมม่า (,) หลังจากนั้นให้คลิกที่ Install



ภาพ 1-8 การติดตั้งชุดโปรแกรม gclus ด้วย install tap

ชุดโปรแกรม (packages) ที่นักวิจัยติดตั้งจะถูกเก็บไว้ใน Libraries การเรียกใช้จะใช้ฟังก์ชัน library () เช่น เมื่อนักวิจัยต้องการใช้ชุดปรแกรมแกรม ggplot2 เขียนคำสั่งได้ว่า library (ggplot2) ดูเพิ่มเติมในหัวข้อการเขียน R script

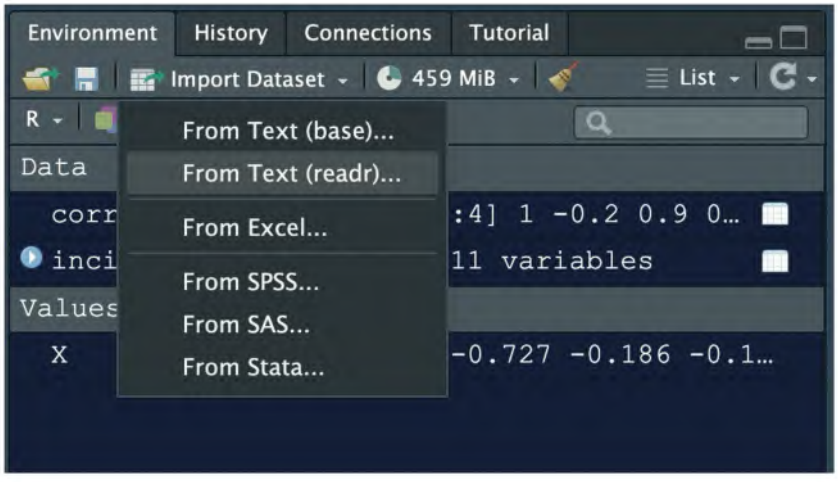
2.6.3 การ update ชุดโปรแกรม

เนื่องจากชุดโปรแกรมมีการพัฒนาอยู่ตลอดเวลา ทำให้เกิดการเปลี่ยนแปลงฟังก์ชันหรือคำสั่งในชุดโปรแกรม รวมทั้งเพิ่มเติมคุณสมบัติใหม่ ๆ (features) หรือยกเลิก หรือเปลี่ยนแปลงฟังก์ชัน (คำสั่งเฉพาะเพื่อให้โปรแกรมปฏิการ ตามที่นักวิจัยต้องการ โดยภายใต้ฟังก์ชันนั้นจะมีอัลกอลิทึม (algolithm) หรือโค้ดที่เขียนขึ้นภายใต้แนวคิดหรือหลักการอย่างหนึ่งอย่างใด ซึ่งผู้เขียนโค้ดมักจะอ้างอิงเอาไว้ที่สามารถตรวจสอบได้ นักวิจัยควรมีการ update ชุดโปรแกรมอยู่เสมอ โดยที่แผง Output ที่ Packages tap คลิกที่ Update tap หากชุดโปรแกรมใดมีการปรับปรุงและพัฒนาใหม่ เพิ่มเติมจะปรากฏรายชื่อชุดโปรแกรมขึ้นมา นักวิจัยสามารถเลือกชุดโปรแกรมที่ต้องการ update หรือถ้าต้องการ update ทั้งหมดให้เลือก Select All แล้วคลิก Install Updates

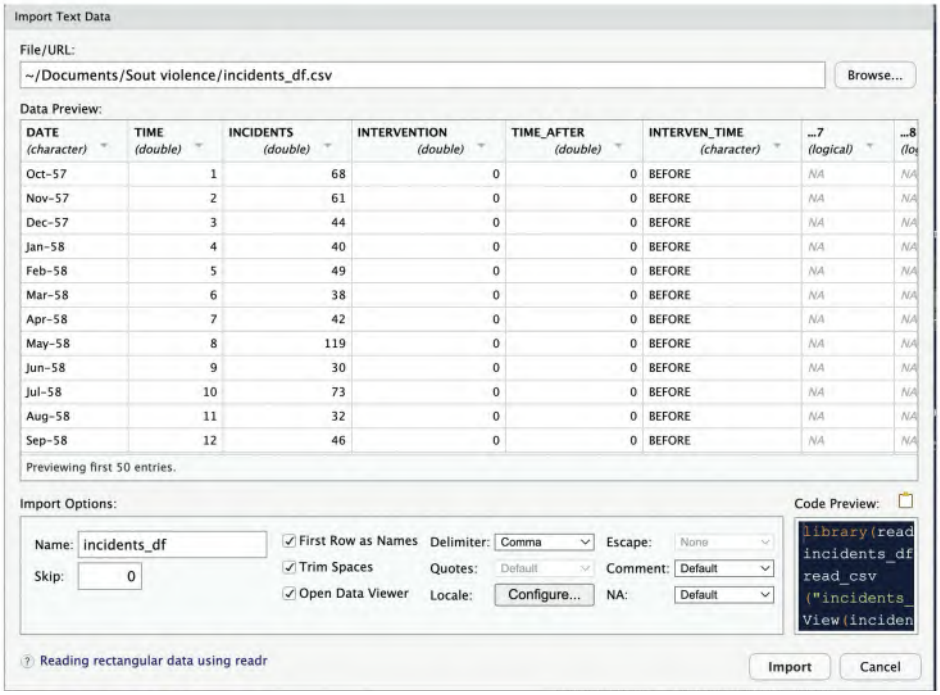
ภาพ 1-9 การ update ชุดโปรแกรม

2.6.4 ข้อมูลและการนำเข้าข้อมูล

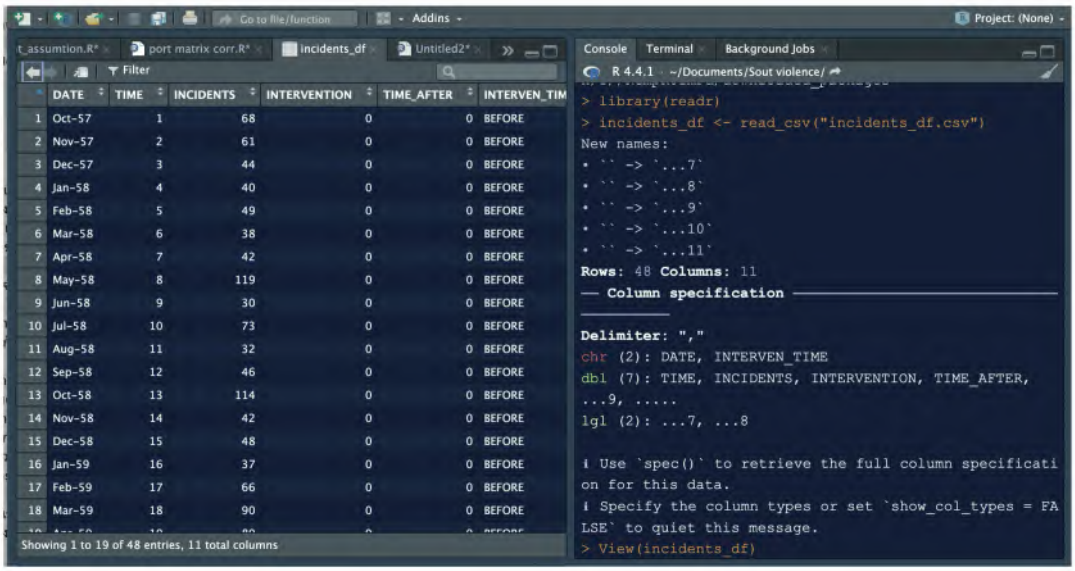
นักวิจัยสามารถนำเข้าข้อมูลได้สองวิธีคือการเขียนคำสั่งเพื่อนำเข้าข้อมูล แต่สำหรับมือใหม่ผู้เขียนขอแนะนำการนำเข้าข้อมูลโดยใช้เมนูคำสั่งที่มีอยู่แล้ว ซึ่งเป็นวิธีการนำเข้าข้อมูลเหมือนการนำไฟล์ต่าง ๆ เข้าในโปรแกรมที่เราต้องดำเนินการ (เช่น นำไฟล์ .docx เข้า MS Word หรือ ไฟล์ .xlsx เข้า MS Excel เป็นต้น) ตัวอย่างการนำเข้าข้อมูลบทที่ 4 ชื่อ ไฟล์ incidents_df.csv ไปที่ Environments pane เลือก Import Dataset ให้นักวิจัยเลือกประเภทข้อมูลให้ตรงกับข้อมูลที่จะนำเข้า ในตัวอย่างนี้จะเลือก From Text (readr) หลังจากนั้นไปที่ Browse... เลือกไฟล์ที่เตรียมไว้ (ในโฟลเดอร์ที่เป็นพื้นที่ทำงาน) หลังจากนั้นคลิกที่ Import ข้อมูลก็จะพร้อมสำหรับการจัดการหรือวิเคราะห์



ภาพ 1-10 การนำชุดข้อมูลเข้า

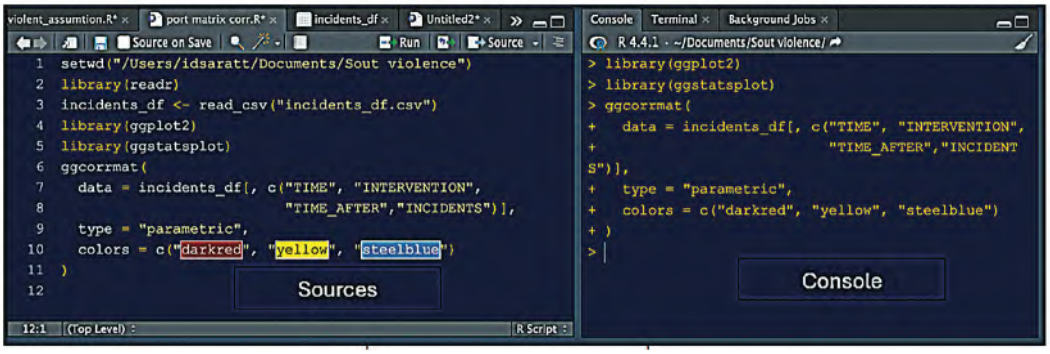


ภาพ 1-11 การนำชุดข้อมูลเข้าโดย Browse และรายละเอียดของชุดข้อมูล



ภาพ 1-12 การชุดข้อมูลเข้าและการแสดงข้อมูลในแผง Sources

จะเห็นว่าใน Console pane จะปรากฏคำสั่ง และจะมีเครื่องหมาย > อยู่หน้าทุกคำสั่งเสมอซึ่งเป็นเช่นเดียวกับ R นักวิจัยสามารถคัดลอกคำสั่งไว้ในแผง Sources (นักวิจัยสามารถ run คำสั่งในฝั่ง Console ได้แต่จะไม่สะดวก)

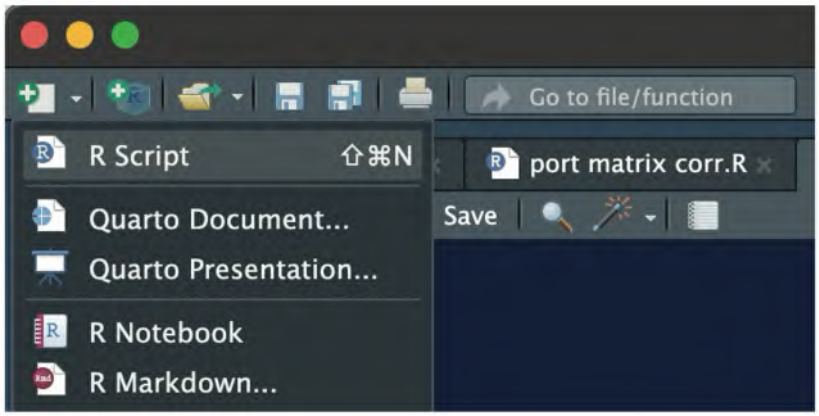


ภาพ 1-13 เทียบคำสั่งแผง Source กับ ผลการ run ที่แสดงในแผง Console

2.6.5 การเขียน การนำเข้าและผลลัพธ์ของ R Script

การเขียน R Script ก็คือการเขียนคำสั่ง หรือการเขียนโค้ด (coding) หรือชุดคำสั่ง เพื่อให้ R ดำเนินการตามที่นักวิจัยต้องการ แต่คำสั่งในฝั่งของ Sources pane จะเป็นหมายเลขบรรทัดที่คำสั่งนั้นอยู่ ในหนังสือเล่มนี้จะใช้รูปแบบคำสั่งในรูปแบบ Sources และแสดงผลลัพธ์ในแบบ Console pane

โดยปกติเมื่อเราเปิด R ขึ้นมา จะเริ่มต้นด้วยตอนปัจจุบันที่กำลังทำงาน (interactive session) โดยการป้อนข้อมูลผ่านแป้นพิมพ์ และแสดงผลลัพธ์บนหน้าจอ ซึ่งหมายความว่านักวิจัยต้องเขียนคำสั่งหรือชุดคำสั่ง โดยใน R จะเรียก Script ซึ่งในการเขียน Script ใหม่ สามารถเริ่มได้จาก File ->New File -> R Script หรือ กดที่ปุ่ม ตรงหัวมุมซ้ายของเมนู แล้วเลือก R script



ภาพ 1-14 การนำเข้า R Script

นักวิจัยจะได้ Source ที่ว่างเปล่าเพื่อรองรับการเขียน Script เช่น ผู้เขียนต้องการพล็อต กราฟเมทริกซ์สหสัมพันธ์ระหว่างตัวแปร โดยใช้ชุดข้อมูลจาก incidents_data.csv ซึ่งเก็บในไดเรกทอรี “/Users/idsaratt/Documents/Sout violence” สิ่งที่ผู้เขียนต้องทำคือนำเข้าข้อมูล เรียกใช้ชุดโปรแกรมในที่นี้เรียกใช้ readr package สำหรับอ่านไฟล์ .csv เรียกใช้ ggplot2 package และ ggstatplot package เพื่อสร้างข้อมูลภาพ (visual data) และคำนวณค่าสถิติ และกำหนดพื้นทำงานของโปรเจกต์นี้ด้วยคำสั่ง setwd ()

แต่นักวิจัยสามารถกำหนดกระบวนการของคำสั่งได้จาก script file หรือไฟล์ชุดคำสั่ง หรือไฟล์ที่บรรจุคำสั่งของ R และกำหนดผลลัพธ์ไปยังปลายทางต่าง ๆ ได้ตามที่เราต้องการ ตัวอย่างเช่น source (“myscript.R”)

การนำเข้า (input)

ใช้ฟังก์ชัน source (“filename”) เพื่อใช้ชุดคำสั่งสำหรับการประมวลผลในตอนปัจจุบัน ถ้าไม่ปรากฏเส้นทางของไฟล์คำสั่ง โปรแกรมก็จะเข้าใจว่าไฟล์คำสั่งนี้อยู่ในไดเรกทอรีปัจจุบัน ตัวอย่างคำสั่ง source(“myscript.R”) หมายถึงการประมวลผลคำสั่งต่าง ๆ ของ R ที่อยู่ในไฟล์ myscript.R

ผลลัพธ์ที่เป็นข้อความ (text output)

ใช้ฟังก์ชัน sink(“filename”) เพื่อกลับไปดูผลลัพธ์ในไฟล์ที่ชื่อ filename โดยปกติถ้าไฟล์นี้มีอยู่แล้ว ไฟล์เดิมจะถูกทับด้วยไฟล์ใหม่ นอกจากเราใช้คำสั่งเพิ่ม (option) append=TRUE เพื่อที่จะเพิ่มผลลัพธ์ข้อความเข้าไปในไฟล์ได้ไม่ทับอันเดิม แต่ถ้าเราใช้ฟังก์ชัน sink() โดยไม่มีออพชันใด ๆ ผลการวิเคราะห์จะปรากฏเฉพาะบนจอเท่านั้น

ผลลัพธ์ที่เป็นกราฟิก (graphic output)

การใช้ฟังก์ชัน sink() จะมีผลกับผลลัพธ์ที่เป็นข้อความเท่านั้น แต่ไม่มีผลต่อผลลัพธ์ที่เป็นกราฟิกเพื่อกำหนดรูปและผลลัพธ์ที่เป็นกราฟิก เราสามารถใช้ฟังก์ชันต่าง ๆ ตามที่เราต้องการให้กราฟิกนั้นเป็นไฟล์ประเภทใด โดยดูฟังก์ชันต่าง ๆ ตามตาราง

ตาราง 1-1 ฟังก์ชันการบันทึกผลลัพธ์กราฟิก

ฟังก์ชัน	ผลลัพธ์
bmp (“filename.bmp”)	ไฟล์ชนิด BMP
jpeg (“filename.jpeg”)	ไฟล์ชนิด JPEG
pdf (“filename.pdf”)	ไฟล์ชนิด PDF
png(“filename.png”)	ไฟล์ชนิด PNG
Postscript (“filename.ps”)	ไฟล์ชนิด PostScript file
svg (“filename.svg”)	ไฟล์ชนิด SVG
win.metafile (“filename.wmf”)	Window metafile

สมมติว่าเรามีไฟล์คำสั่ง R อยู่ 3 ไฟล์ ได้แก่ script1.R, script2.R และ script3.R

```
> Source(“script1.R”)
```

หมายถึงการนำชุดคำสั่ง R ที่อยู่ในไฟล์ script1.R มาใช้ประมวลผลในรอบปัจจุบัน (current session) และให้ปรากฏผลในจอภาพ (on the screen)

แต่ถ้าเรามีคำสั่ง เป็น

```
> sink(“myoutput”, append=TRUE, split=TRUE)
> pdf(“mygraphs.pdf”)
> source(“script2.R”)
```

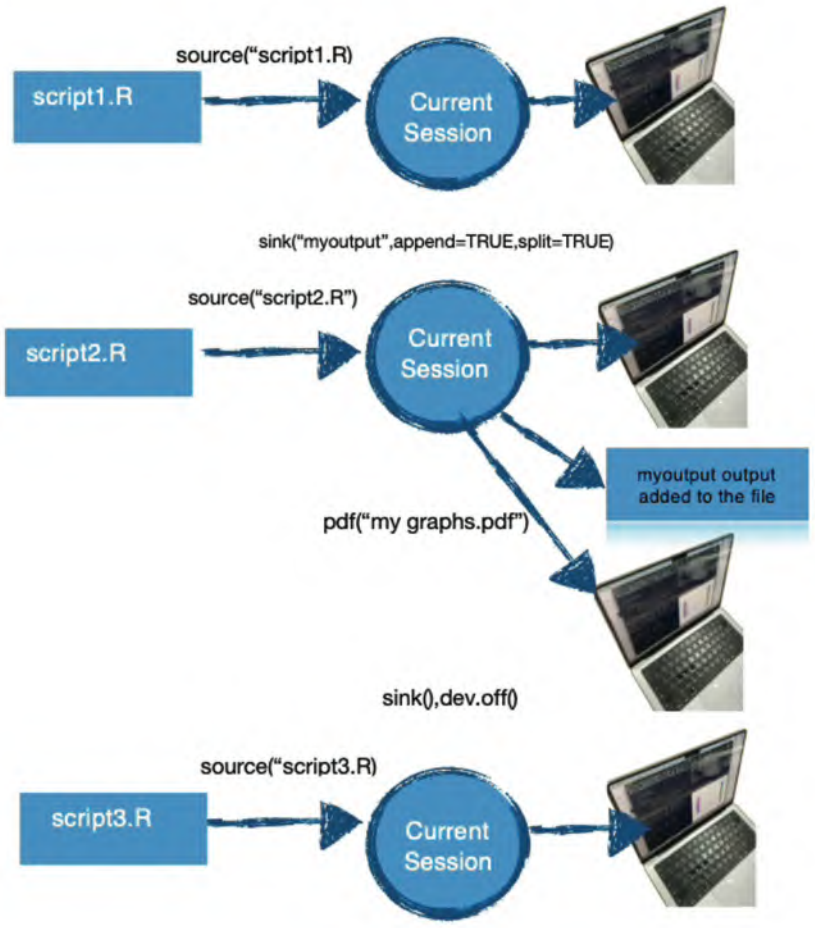
คำสั่ง R จาก script2.R จะถูกนำเสนอ และผลลัพธ์จะปรากฏอีกครั้งที่จอภาพ พร้อมกับผลการวิเคราะห์ที่เป็นข้อความเพิ่มเติมจากถูกบันทึกไว้ในไฟล์ myoutput และผลลัพธ์กราฟิก จะถูกบันทึกไว้ที่ไฟล์ mygraph.pdf

สุดท้ายถ้าเรามีคำสั่งเป็น

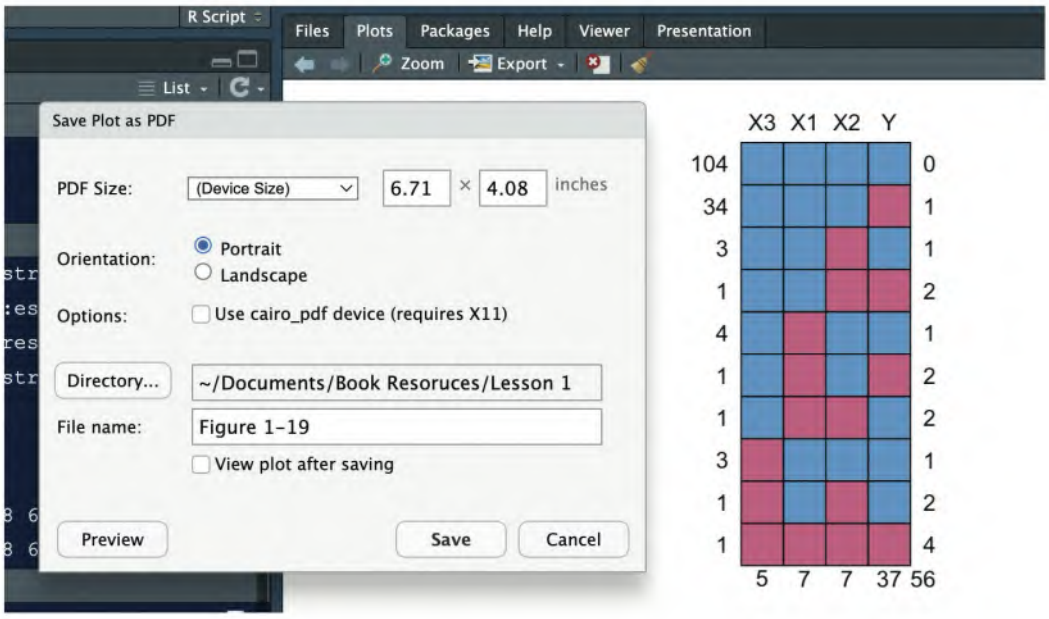
```
> sink ()
> dev.off ()
> source ("script3.R")
```

ชุดคำสั่งจาก script3.R จะถูกนำมาใช้ประมวลผล และผลลัพธ์ที่ได้จะปรากฏที่จอภาพ แต่ผลการวิเคราะห์ที่เป็นข้อความ และกราฟิกจะไม่ถูกบันทึกเป็นไฟล์ ลำดับของการใช้คำสั่งสามารถดูได้จากภาพ 1-15

สำหรับผู้เขียนส่วนใหญ่ใช้การแสดงผลแบบ on screen กล่าวคือ แสดงผลใน text (ตัวเลขตัวหนังสือต่าง ๆ) ในแผง Console และแสดงภาพ หรือกราฟิกในแผง Output เพราะสามารถเห็นแปรความ รวมทั้งกลับไปแก้ไขเพิ่มเติมคำสั่งเพื่อให้ได้ผลตามที่ต้องการได้ทันที แต่หากจะเก็บผลการประมวลเอาไว้ดูภายหลังจึงใช้คำสั่ง sink () ส่วนการเก็บรูปภาพกราฟิกข้อมูลภาพ สามารถสั่ง export จากแผง Output โดยไม่จำเป็นต้องเขียนโค้ดแต่ประการใด



ภาพ 1-15 การนำเข้าโดยฟังก์ชัน source () และผลลัพธ์โดยใช้ฟังก์ชัน sink ()



ภาพ 1-16 การบันทึกข้อมูลภาพจากแผง Output

ตัวอย่างต้องการบันทึกภาพเพื่อนำไปใช้จากภาพ 1-16 เลือก Export->เลือกประเภทภาพที่ต้องการบันทึก ผู้เขียนเลือก save as PDF หลังจากนั้นกำหนดขนาดภาพ ลักษณะทิศทางของภาพ (orientation) กำหนดพื้นที่ที่ต้องการจัดเก็บ (หากนักวิจัยกำหนดพื้นที่ทำงานไว้จัดเก็บในพื้นที่ดังกล่าวอัตโนมัติ) ตั้งชื่อตามที่ต้องการ จะดูภาพก่อนให้คลิก Preview หรือจะบันทึกโดยคลิกที่ Save ซึ่งไม่ต้องเขียนโค้ดใด ๆ

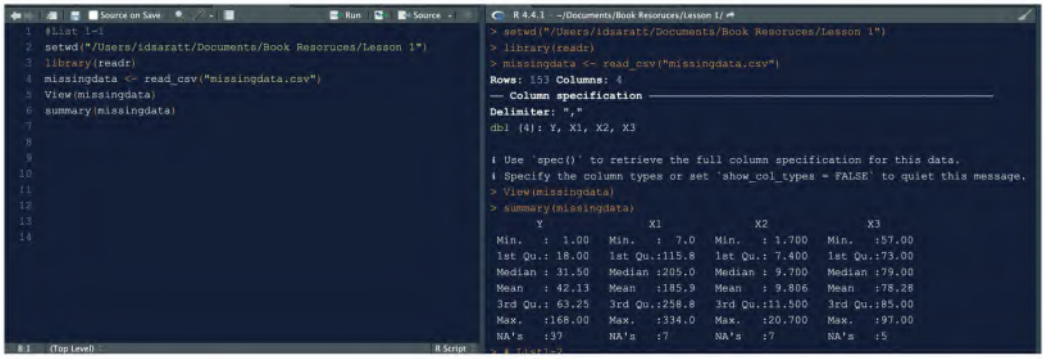
1.4.3 การใช้คำสั่งในหนังสือสำหรับ RStudio

เนื่องจากในหนังสือนี้ได้แสดงชุดคำสั่งต่าง ๆ สำหรับการวิเคราะห์ ซึ่งสามารถดำเนินการได้โดยดาวน์โหลดชุดคำสั่งจากที่ผู้เขียนได้เตรียมไว้ และนำเข้าใน RStuio โดยไปที่ File > Open File... เลือกชุดคำสั่งที่เตรียมไว้ (นามสกุล .R) อย่างไรก็ตามในหนังสือเล่มนี้จะใช้วิธีการที่ง่าย ไม่ซับซ้อนและพยายามใช้คำสั่ง หรือเขียนโค้ดเท่าที่จำเป็น แต่จะใช้การทำงานจากสภาพแวดล้อมที่ RStudio ได้จัดเตรียมไว้ให้

กรณีคัดลอกคำสั่งจากหนังสือไปวางในแผง Sources สามารถทำได้เช่นเดียวกัน แต่ต้องปรับเล็กน้อยโดยนำเครื่องหมาย > ที่อยู่หน้าคำสั่งออก ดูภาพ 1-17

```
> setwd("/Users/idsaratt/Documents/Book Resoruces/Lesson 1")
> library(readr)
> missingdata <- read_csv("missingdata.csv")
> View(missingdata)
> summary(missingdata)
```

```
       Y                X1               X2               X3        
 Min.   :  1.00   Min.   :  7.0   Min.   : 1.700   Min.   :57.00  
 1st Qu.: 18.00   1st Qu.:115.8   1st Qu.: 7.400   1st Qu.:73.00  
 Median : 31.50   Median :205.0   Median : 9.700   Median :79.00  
 Mean   : 42.13   Mean   :185.9   Mean   : 9.806   Mean   :78.28  
 3rd Qu.: 63.25   3rd Qu.:258.8   3rd Qu.:11.500   3rd Qu.:85.00  
 Max.   :168.00   Max.   :334.0   Max.   :20.700   Max.   :97.00  
 NA's   :37       NA's   :7       NA's   :7        NA's   :5      
```



ภาพ 1-17 รายการคำสั่ง และผลลัพธ์จากหนังสือ และการนำเขียนคำสั่งในแผง Sources

ด้านบนเป็นรูปแบบคำสั่ง และผลลัพธ์ในหนังสือ นักวิจัยสามารถคัดลอกคำสั่งมาวางในแผง Sorces ตามภาพด้านล่างแต่จะต้องนำเครื่องหมาย > ออก เนื่องจากเครื่องหมาย > ที่อยู่ในหนังสือติดมาจากแผง Console และตั้งใจคงไว้เพื่อไม่ให้เกิดความสับสนระหว่างคำสั่งกับผลลัพธ์ โดยจะสั่ง run ทีละคำสั่งหรือพร้อมกันทุกคำสั่งก็ได้ และจะแสดงผลในส่วน Console อย่างไรก็ตาม ผู้เขียนได้เตรียมชุดคำสั่งไว้ในแต่ละบท นักวิจัยสามารถนำเข้า R script ของแต่ละบท โดยกำหนดพื้นที่ทำงานให้ถูกต้องตามการจัดเก็บในเครื่องคอมพิวเตอร์ของตนเอง

จากคำสั่ง

บรรทัดที่ 1 # List 1-1 เป็นการตั้งชื่อคำสั่ง โดยเครื่องหมาย # อยู่ข้างหน้าจะไม่มีการประมวลผลใด ๆ

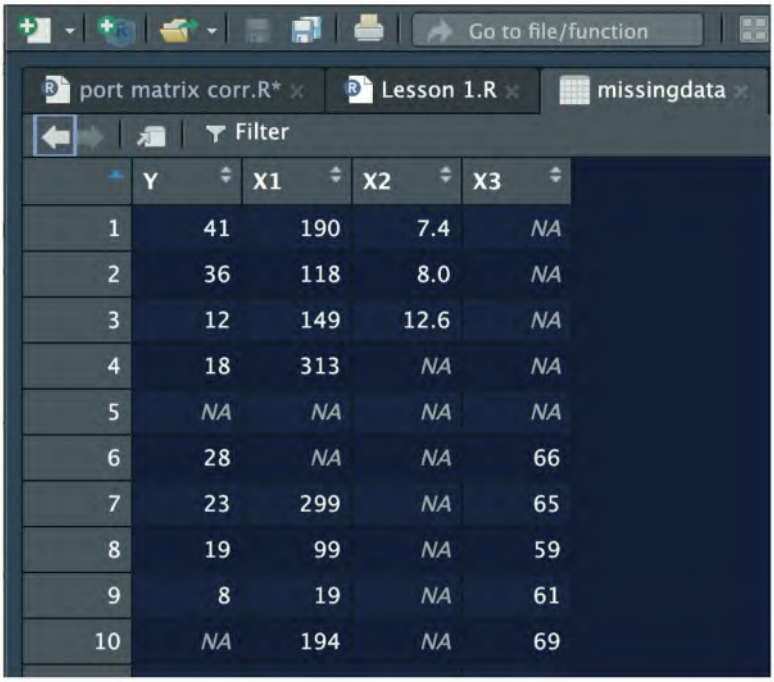
บรรทัดที่ 2 กำหนดพื้นที่ทำงาน (ไม่จำเป็นแต่ทำให้สะดวกในการทำงาน)

บรรทัดที่ 3 เรียกใช้ read package เพื่อนำเข้าไฟล์ข้อมูล .csv

บรรทัดที่ 4 นำเข้าข้อมูลไฟล์ missingdata.csv และเก็บไว้ในวัตถุชื่อว่า missingdata

บรรทัดที่ 5 แสดงข้อมูลในแผง Sources ดังภาพ 1-18

บรรทัดที่ 6 วิเคราะห์ค่าสถิติพื้นที่ฐานโดยสรุป จะแสดงผลที่แผง Console



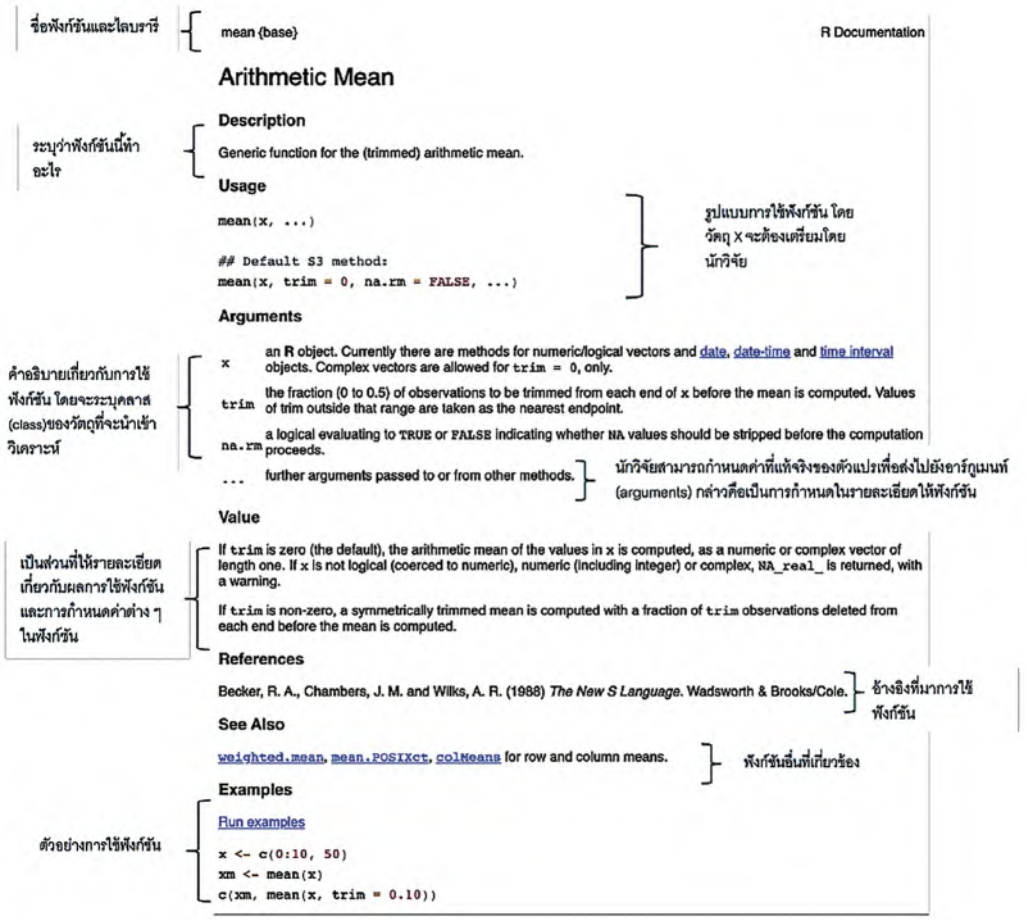
	Y	X1	X2	X3
1	41	190	7.4	NA
2	36	118	8.0	NA
3	12	149	12.6	NA
4	18	313	NA	NA
5	NA	NA	NA	NA
6	28	NA	NA	66
7	23	299	NA	65
8	19	99	NA	59
9	8	19	NA	61
10	NA	194	NA	69

ภาพ 1-18 แสดงข้อมูลจากคำสั่ง View ()

1.4.4 เอกสารช่วยเหลือ

ฟังก์ชัน ชุดข้อมูล และอ็อบเจ็กต์ในตัวอื่น ๆ ใน R จะถูกบันทึกไว้ในระบบช่วยเหลือ ค้นหาแท็บ “ช่วยเหลือ” ของ RStudio เพื่อดูเอกสารเกี่ยวกับแพ็คเกจ R และฟังก์ชันเฉพาะ ซึ่งอยู่ในบานหน้าต่างที่ประกอบด้วยไฟล์ พล็อต และแพ็คเกจ (โดยปกติจะอยู่ในบานหน้าต่างด้านล่างขวา) นอกจากนี้นักวิจัยยังสามารถพิมพ์ชื่อแพ็คเกจหรือฟังก์ชันลงในคอนโซล R หลังเครื่องหมายคำถามเพื่อเปิดหน้าช่วยเหลือที่เกี่ยวข้องได้อีกด้วย โดยไม่ใส่เครื่องหมายวงเล็บ ตัวอย่าง การใช้เอกสารช่วยเหลือเกี่ยวกับการฟังก์ชัน mean ดังภาพ 1-19

ภาพ 1-19 การใช้เอกสารช่วยเหลือเกี่ยวกับการฟังก์ชัน mean



ภาพ 1-20 รายละเอียดโครงสร้างของวิธีใช้จากเอกสารช่วยเหลือเกี่ยวกับการฟังก์ชัน mean

คุณภาพของหน้าวิธีใช้ของ R นั้นแตกต่างกันบ้าง โดยส่วนใหญ่มักจะสั้นและโครงสร้างเหมือนกันและจะเป็นประโยชน์หากทราบวิธีอ่าน ผู้เขียนได้อธิบายไว้ตามภาพ 1-20 จะให้ภาพรวมของสิ่งที่ต้องมองหา โปรดจำไว้ว่าฟังก์ชันจะรับอินพุต ดำเนินการ และส่งเอาต์พุต บางอย่างจะเข้าไป ดำเนินการจากนั้นจึงส่งเอาต์พุตออกมา นั่นหมายความว่านักวิจัยต้องการทราบว่าฟังก์ชันต้องการอะไร ทำอะไรและส่งคืนอะไร สิ่งที่ต้องการจะแสดงในส่วนการใช้งาน และอาร์กิวเมนต์ของหน้าวิธีใช้ ชื่อของอาร์กิวเมนต์ที่จำเป็นและไม่จำเป็น ทั้งหมดจะระบุตามชื่อและตามลำดับที่ฟังก์ชันต้องการ อาร์กิวเมนต์ บางตัวมีค่าเริ่มต้น ในกรณีของฟังก์ชัน mean () อาร์กิวเมนต์ na.rm จะถูกตั้งค่าเป็น FALSE ตามค่าเริ่มต้น ซึ่งจะแสดงในส่วนการใช้งาน หากอาร์กิวเมนต์ที่มีชื่อไม่มีค่าเริ่มต้น นักวิจัยจะต้องกำหนดค่าให้กับอาร์กิวเมนต์นั้น โดยอาร์กิวเมนต์อาจเป็นค่าตรรกะ ตัวเลข ชุดข้อมูล หรืออ็อบเจ็กต์อื่น ๆ ขึ้นอยู่กับอาร์กิวเมนต์

3.การเตรียมและตรวจสอบข้อมูล

โดยปกติแล้วข้อมูลที่ได้มามักเป็นข้อมูลที่ไม่สมบูรณ์ หรือมีข้อผิดพลาด (Jain et al.,2024) โดยข้อมูลผิดพลาด (missing data) อาจอยู่ในรูปของค่าที่สูงหรือต่ำผิดปกติ (outliers) หรือข้อมูลที่ไม่ถูกต้อง (wrong

data) บ่อยครั้งที่นักวิจัยมักจะทำให้เกิดความเสียหายกับผลวิจัยเนื่องจากไม่ได้มีการทำข้อมูลให้สะอาด คำว่าข้อมูลสะอาด (clean data) จะหมายถึงการตรวจคัดข้อมูล (screening data) เพื่อนำข้อมูลออกในกรณีที่เป็นข้อมูลที่อยู่นอกพิสัยของค่าข้อมูล หรือการเปลี่ยนสภาพข้อมูล (transform) เพื่อให้เป็นไปตามข้อตกลงเบื้องต้นของการวิเคราะห์ เมื่อนักวิจัยไม่ได้ทำความสะอาดข้อมูล ซึ่งหมายความว่าทำการวิเคราะห์ข้อมูลด้วยชุดข้อมูลที่สกปรก (dirty data sets) ในสารนิพนธ์วิทยานิพนธ์ หรือโครงการวิจัย การทำความสะอาดข้อมูลเปรียบเสมือนการทำความสะอาดงาน แต่การทำความสะอาดข้อมูลเป็นงานที่นักวิจัยไม่ค่อยอยากทำถึงแม้มันจะเป็นสิ่งสำคัญก็ตาม ในต่างประเทศมีค่าจ้างในการทำความสะอาดข้อมูลแพง พอ ๆ กับการจ้างวิเคราะห์ หรือการแปลความหมายจากการวิเคราะห์ หรือการเขียนผลการวิเคราะห์ (Pigott, 2001)

ข้อมูลผิดพลาดเป็นเรื่องที่เกิดขึ้นได้โดยปกติในการวิเคราะห์ทางสถิติ (Little & Rubin, 1987) อัตราของข้อมูลผิดปกตินั้น ถ้าน้อยกว่าร้อยละ 1 ถือว่าไม่มีปัญหาเป็นข้อมูลผิดพลาดเล็กน้อย ถ้าอยู่ระหว่างร้อยละ 1 – ร้อยละ 5 ก็ยังสามารถจัดการได้ แต่ถ้าอยู่ระหว่างร้อยละ 5 - ร้อยละ 15 จำเป็นต้องแสดงวิธีการจัดการข้อมูลให้เห็นว่าจัดการแก้ปัญหาด้วยวิธีใด แต่ถ้าเกินร้อยละ 15 ขึ้นไปจะมีผลต่อการตีความหมาย และเกิดความเสียหายต่อผลของการวิเคราะห์ข้อมูล (Luengo, García & Herrera, 2012) ในงานวิจัยเป็นข้อมูลที่ประกอบไปด้วยข้อมูลไม่ถูกต้อง (inaccurate) ข้อมูลไม่สมบูรณ์ (incomplete) และข้อมูลไม่สอดคล้อง (inconsistency) ข้อมูลเหล่านี้เป็นสาเหตุทำให้การวิเคราะห์ข้อมูลเกิดความคลาดเคลื่อน ส่งผลให้การสรุป และอภิปรายผลการวิจัยไม่ถูกต้อง มีโอกาสบิดเบือนจากความเป็นจริง เมื่อนำผลการวิจัยไปใช้อาจก่อให้เกิดผลเสียหาย ดังนั้น ก่อนการวิเคราะห์ข้อมูลจึงจำเป็นต้องมีการตรวจคัด (screening) และการทำสะอาดข้อมูล (cleaning data) ก่อนเสมอ

นอกจากการตรวจสอบความผิดพลาดของข้อมูลก่อนการวิเคราะห์แล้ว สิ่งสำคัญในการทดสอบทางสถิติคือการตรวจสอบข้อตกลงเบื้องต้น (assumption) ในการวิเคราะห์ของสถิติแต่ละตัวที่นักวิจัยเลือกใช้ เพราะสถิติแต่ละตัวมักจะมีข้อตกลงเบื้องต้นของการนำไปใช้ เนื่องจากจะทำให้สถิตินั้นมีอำนาจในการทดสอบสูง ซึ่งจะทำให้ได้ผลการวิเคราะห์ที่แม่นยำ และมีประสิทธิภาพ ทำให้ได้ผลการวิจัยที่น่าเชื่อถือ สถิติแต่ละตัว แต่ละประเภทจะมีข้อตกลงเบื้องต้นต่างกัน เช่น ในการศึกษาเพื่อทดสอบค่าเฉลี่ยประชากรของแต่ละทรีตเมนต์มีค่าเท่ากันหรือไม่ หรือทดสอบอิทธิพลของทรีตเมนต์มีนัยสำคัญหรือไม่นั้น มีข้อตกลงเบื้องต้นคือความคลาดเคลื่อนสุ่มจากประชากรเดียวกันที่มีการแจกแจงแบบปกติโดยมีค่าเฉลี่ยเป็นศูนย์ และมีความแปรปรวนเท่ากันในทุกทรีตเมนต์ และแต่ละตัวก็เป็นอิสระซึ่งกันและกัน หากพบว่าข้อมูลไม่เป็นไปตามข้อตกลงตามการทดสอบพาราเมตริก (parametric test) จะทำให้ประสิทธิภาพของการทดสอบลดลงนำไปสู่การสรุปผลการวิเคราะห์ที่ผิดพลาด และขาดความน่าเชื่อถือ ซึ่งวิธีการแก้ปัญหาในกรณีที่ข้อมูลมีการแจกแจงแบบไม่ปกติ และความแปรปรวนมีค่าไม่เท่ากันสามารถทำได้โดยการแปลงข้อมูล (data transformation) โดยใช้วิธีการทางสถิติ

จะเห็นได้ว่าการตรวจคัดจึงเป็นการตรวจหาความผิดพลาดที่เกิดจากการนำเข้าข้อมูล ซึ่งความผิดพลาดนี้จำเป็นที่จะมีการทำความสะอาดข้อมูลเพื่อทำให้ข้อมูลสมบูรณ์ มีความสอดคล้อง หรือมีความถูกต้องมากที่สุดด้วยค่าที่เป็นไปได้ก่อนนำไปสู่กระบวนการวิเคราะห์ข้อมูลต่อไป การทำความสะอาดข้อมูล ปรับปรุงข้อมูลที่เกิดการผิดพลาดให้มีความสมบูรณ์ ถูกต้อง และสอดคล้องกัน มีวัตถุประสงค์เพื่อปรับรูปแบบ (format) ของค่าในแต่ละตัวอย่างเหมาะสม เช่น ข้อความ (string) ตัวเลข (numeric) การค้นหาค่าตัวแปรที่นอกเหนือจากคู่มือรหัสกำหนดไว้ การรวมหรือตัดข้อมูลซ้ำกัน ขจัดความขัดแย้งโดยพิจารณาความสอดคล้องภายในตัวแปร หรือระหว่างระเบียน ค้นหา และจัดการแก้ไขค่าสุดโต่ง (outliers) หรือค่าผิดพลาด (missing values) สำหรับเทคนิคการทำข้อมูลให้สมบูรณ์มีด้วยกันหลายวิธี แต่การจัดการข้อมูลผิดพลาดจะต้องเป็นไปด้วยความระมัดระวัง และควรแก้ปัญหาโดยอาศัยความรู้ที่ถูกต้อง ซึ่งได้มีการนำเสนอผลการศึกษาเกี่ยวกับการแก้ปัญหาในเอกสารที่เกี่ยวข้องจำนวนมาก

(Acuna & Rodriguez, 2004; Batista & Monard, 2003) ตั้งแต่วิธีธรรมดาไปจนถึงการใช้โปรแกรมเข้ามาช่วย ในเนื้อหานี้จะกล่าวถึงลักษณะของข้อมูลที่ต้องมีการจัดการบางส่วนการแสดงวิธีการจัดการจากโปรแกรม R กับข้อมูลที่ผิดพลาด และการทดสอบข้อตกลงเบื้องต้นในบางประเด็นเท่านั้น

4. ค่าผิดพลาด

ค่าผิดพลาด (missing value) จะมีผลต่อผลการวิเคราะห์ข้อมูล Luengo et al. (2012) ได้ให้คำแนะนำว่า ค่าผิดพลาดจะมีผลต่อการวิเคราะห์ข้อมูล 3 ประการด้วยกัน คือ 1) ขาดประสิทธิภาพ (loss of efficiency) 2) มีความยุ่งยากต่อการวิเคราะห์ข้อมูล (complications in analysing the data) เนื่องจากในขณะที่มีกระบวนการวิเคราะห์ข้อมูล ไม่มีวิธีใดแยกได้ว่าข้อมูลใดเป็นค่าผิดพลาด แต่ในความเป็นจริงนักวิจัยต้องแก้ปัญหาค่าผิดพลาดก่อนทำการวิเคราะห์ข้อมูลจริง และ 3) ได้ผลการวิเคราะห์ที่ลำเอียงจากความแตกต่างระหว่างข้อมูลที่มีการสูญหายและข้อมูลที่สมบูรณ์ ในความเป็นจริงถึงแม้จะมีการแทนค่าผิดพลาดก็ไม่ได้ผลที่แท้จริงเหมือนข้อมูลที่มีความสมบูรณ์ ข้อมูลที่ผิดพลาดไปนั้นเป็นค่าต่าง ๆ ในการรวบรวมข้อมูลที่ไม่มีอยู่ด้วยสาเหตุปกติหรือผิดปกติ ไม่ถูกที่ ไม่สมจริง หรือไม่ได้ระบุเลย ค่าที่ขาดหายไปในข้อมูลทำให้เกิดความสับสนทั้งในการวิเคราะห์ข้อมูลและการยื่นคำอธิบายสำหรับข้อมูลใหม่ (Jain et al.,2024) อย่างไรก็ตามการวิเคราะห์ข้อมูลควรมีการตรวจสอบข้อมูลที่เป็นค่าผิดพลาดเสียก่อน ด้วยความสำคัญ และมีผลกระทบดังกล่าวจึงมีงานเขียนทางด้านสถิติว่าด้วยการวิเคราะห์กรณีค่าผิดพลาด (Acock, 2005; Jafrasteh et al., 2023; Jain et al., 2024; Ren et al., 2023)

4.1 ประเภทของค่าผิดพลาด

การพิจารณาประเภทของค่าผิดพลาดเป็นขั้นตอนที่สำคัญ ทั้งนี้ เพราะหากสามารถทราบถึงลักษณะของค่าผิดพลาดจะช่วยในการพิจารณาแนวทางสำหรับจัดการกับปัญหาความไม่สมบูรณ์ของข้อมูลได้อย่างเหมาะสมซึ่ง Salgado et al. (2016) ได้จำแนกค่าผิดพลาดออกเป็น 3 ประเภทด้วยกัน คือ Missing completely at random (MCAR), Missing at random (MAR) และ Not missing at random (NMAR) โดยในแต่ละประเภทมีรายละเอียดดังนี้ (Allison, 2002; Horton & Kleinman, 2007; Luengo et al., 2011; Salgado et al.,2016)

4.1.1 Missing completely at random (MCAR)

MCAR เป็นลักษณะของค่าผิดพลาดที่เกิดขึ้นอย่างสุ่มจากค่าสังเกตทั้งหมด นั่นคือข้อมูลที่สูญหายเป็นอิสระจากตัวแปรต่าง ๆ สามารถทำการตรวจสอบลักษณะของค่าผิดพลาดกลุ่มนี้โดยการแบ่งกลุ่มของค่าสังเกตเป็นกลุ่มข้อมูลปกติและค่าผิดพลาด ในกรณีนี้เมื่อทำการทดสอบจะไม่พบความแตกต่างอย่างมีนัยสำคัญระหว่างทั้งสองกลุ่มสำหรับตัวแปรต่าง ๆ ในชุดข้อมูล สำหรับสาเหตุที่ทำให้ข้อมูลเกิดการสูญหายมีอยู่หลากหลายเหตุผล อาจเกิดขึ้นเนื่องจากเครื่องมือเสีย อุปกรณ์เกิดข้อบกพร่อง สภาพอากาศเลวร้าย กลุ่มเป้าหมายที่ศึกษาล้มป่วย หรือการนำเข้าข้อมูลไม่ถูกต้อง สำหรับค่าผิดพลาดประเภทนี้จัดเป็นข้อมูลที่ก่อให้เกิดปัญหาน้อยที่สุด เพราะว่าค่าผิดพลาดไม่มีความเกี่ยวข้องต่อผลลัพธ์ของข้อมูล เพราะฉะนั้นสามารถเลือกทำการวิเคราะห์ข้อมูลในส่วนที่สมบูรณ์ได้ (Rubin, 1976) อย่างไรก็ตาม Little (1988) ได้คิดวิธีการทดสอบค่าผิดพลาดแบบ MCAR โดยใช้การทดสอบอัตราส่วนความเป็นไปได้ (likelihood ratio test) เพื่อทำการทดสอบข้อตกลงเบื้องต้นว่าด้วย MCRA โดยเรียกว่าโปรแกรม BMDPAM โดยเป็นส่วนหนึ่งของ BMDP (Dixon, 1983) หรือใช้ BaylorEdPsych package ทดสอบโดยมีสมติฐาน H_0: ข้อมูลมีค่าผิดพลาดแบบ MCAR ซึ่งมีลักษณะการแจกแจงแบบไคสแควร์ ถ้ายอมรับ H_0 แสดงว่าข้อมูลมีค่าผิดพลาดแบบ MCAR แต่ถ้าปฏิเสธ H_0 แสดงว่าข้อมูลมีค่าผิดพลาดแบบอื่นที่ไม่ใช่ MCRA

4.1.2 Missing at random (MAR)

MAR เป็นลักษณะของค่าผิดพลาดซึ่งไม่ได้เกิดขึ้นอย่างสุ่มจากค่าสังเกตทั้งหมด แต่เกิดขึ้นอย่างสุ่มภายในบางส่วน หรือบางกลุ่มของค่าสังเกต นั่นคือค่าของค่าผิดพลาดขึ้นอยู่กับตัวแปรตัวอื่น ๆ ในชุดข้อมูลซึ่งไม่ได้

เป็นตัวแปรที่เกิดค่าผิดพลาด ยกตัวอย่างเช่น หากพบว่าเฉพาะกลุ่มผู้ได้รับการศึกษาต่ำที่ไม่ให้ความร่วมมือในการตอบข้อคำถามเกี่ยวกับพฤติกรรมการใช้สารเสพติด ในลักษณะนี้สามารถกล่าวได้ว่าข้อมูลพฤติกรรมการใช้สารเสพติดมีค่าสูญหายแบบ MAR ทั้งนี้ เนื่องจากเป็นค่าสูญหายที่เกิดขึ้นเฉพาะในบางส่วนของตัวแปรระดับการศึกษา สำหรับค่าผิดพลาดประเภทนี้ยังไม่ส่งผลกระทบรุนแรงเท่ากับค่าผิดพลาดประเภทสุดท้าย

4.1.3 Not missing at random (NMAR)

NMAR เป็นลักษณะของค่าผิดพลาดซึ่งไม่ได้เกิดขึ้นอย่างสุ่ม โดยค่าของค่าผิดพลาดขึ้นอยู่กับค่าของข้อมูลสมบูรณ์ในตัวแปรเดียวกัน รวมถึงตัวแปรตัวอื่นด้วย เช่น หากค่าผิดพลาดของระดับรายได้ขึ้นอยู่ระดับรายได้ในแต่ละช่วงอายุ ค่าผิดพลาดที่เกิดขึ้นจัดอยู่ในประเภท NMAR หรือในบางกรณีค่าของค่าผิดพลาดอาจไม่ขึ้นอยู่กับตัวแปรใด ๆ ในชุดข้อมูลเลย แต่ขึ้นอยู่กับตัวแปรอื่นที่ไม่ได้ถูกเก็บรวบรวมไว้ในการศึกษาครั้งนั้น เช่น อัตราการลดพฤติกรรมก้าวร้าวขึ้นอยู่กับพฤติกรรมก้าวร้าวตอนเริ่มต้น แต่เนื่องจากตัวแปรพฤติกรรมก้าวร้าวตอนเริ่มต้นไม่ได้ถูกรวบรวมไว้ในฐานข้อมูลดังนั้น ค่าสูญหายของพฤติกรรมก้าวร้าวที่ลดลงจึงขึ้นอยู่กับตัวแปรภายนอกข้อมูลที่นักวิจัยมีอยู่

ลักษณะค่าผิดพลาดประเภทนี้จัดเป็นค่าผิดพลาดที่สามารถส่งผลกระทบอย่างรุนแรงในการวิเคราะห์ข้อมูลในทางปฏิบัติ ลักษณะของค่าผิดพลาดประเภท MCAR มักพบไม่บ่อยนัก ที่พบบ่อยครั้งมักเป็นค่าผิดพลาดประเภท MAR ดังนั้น วิธีการทางสถิติต่าง ๆ ที่พัฒนาขึ้นมาเพื่อแก้ปัญหาค่าผิดพลาด มักดำเนินการภายใต้ข้อสมมติของ MAR เป็นส่วนใหญ่ (Greiner et al., 1997)

4.2 วิธีการจัดการกับค่าผิดพลาด

การจัดการกับค่าผิดพลาดมีหลายวิธีการให้เลือกใช้ การพิจารณาเลือกใช้วิธีการใดขึ้นอยู่กับลักษณะของค่าผิดพลาดที่เกิดขึ้น หากเลือกวิธีการที่ไม่เหมาะสมมาใช้อาจเป็นการเพิ่มค่าความคลาดเคลื่อน และทำลายผลลัพธ์ที่ควรจะได้ สำหรับวิธีการจัดการกับค่าผิดพลาดมีอยู่ด้วยกัน 3 วิธีหลัก ๆ ได้แก่ 1) การลบค่าสูญหายออก (deletion) 2) การแทนค่าข้อมูล (imputation) และ 3) วิธีความเป็นไปได้สูงสุด (maximum likelihood) (Li et al., 2004; Farhangfar et al., 2008) เนื่องในปัจจุบันได้มีการศึกษาวิธีการจัดการกับค่าผิดพลาดเพิ่มขึ้นจำนวนมาก ผู้เขียนได้นำเสนอในบางวิธีมีรายละเอียด ดังนี้

4.2.1 การลบค่าผิดพลาด (deletion)

1) Listwise deletion: เป็นวิธีการจัดการกับค่าผิดพลาดที่ง่าย นั่นคือไม่สนใจค่าผิดพลาดที่เกิดขึ้น โดยจะทำการวิเคราะห์ข้อมูลจากข้อมูลเฉพาะส่วนที่สมบูรณ์ แนวทางนี้จะมีความเหมาะสมในกรณีที่ค่าผิดพลาดมีจำนวนน้อย ซึ่งวิธีการนี้มักถูกกำหนดให้ใช้เป็นหลัก (by default) สำหรับจัดการกับค่าผิดพลาดในโปรแกรมคอมพิวเตอร์ทางสถิติทั่ว ๆ ไป หากไม่เจาะจงเลือกใช้วิธีการอื่นในการจัดการกับค่าผิดพลาด ในการใช้ Listwise deletion สามารถใช้ได้ดีในกรณีค่าผิดพลาดแบบ NMAR และสามารถใช้ได้ในข้อมูลแบบ MCAR แต่อาจมีปัญหาเมื่อใช้กับ MAR ถ้าข้อมูลมีลักษณะแบบ MAR อาจทำให้มีการประมาณค่าที่ลำเอียง เช่น กรณีนักวิจัยต้องการประมาณค่าเฉลี่ยรายได้จากประชากร โดยร้อยละ 80 รายงานรายได้เป็นเพศหญิง และมีเพียงร้อยละ 60 ของเพศชายที่รายงาน (ลักษณะนี้เป็นการละเมิดแบบ MCAR) แต่ภายในแต่ละกลุ่มของเพศที่ผิดพลาดที่มีผลต่อรายได้ไม่ได้ขึ้นอยู่กับรายได้ (MAR) สมมติฐานในกลุ่มเพศชายเกี่ยวกับค่าเฉลี่ยคืออาจสูงกว่าเพศหญิง วิธีการ Listwise deletion จะทำให้ลดความลำเอียงในการประมาณค่าเฉลี่ยรายได้โดยรวมในประชากร (Allison, 2009) การใช้ Listwise deletion จะมีความแกร่งเมื่อมีการละเมิด MCAR (หรือ MAR) สำหรับตัวแปรทำนายในการวิเคราะห์ถดถอย โดยเฉพาะอย่างยิ่งเมื่อค่าสูญหายของตัวแปรทำนายไม่ได้ขึ้นกับตัวแปรตาม การใช้ Listwise deletion จะทำให้ได้ค่าการประมาณไม่ลำเอียงในค่าสัมประสิทธิ์ถดถอย (Little, 1992) และเทคนิคนี้ยังเหมาะกับการวิเคราะห์ถดถอยแบบอื่น ๆ เช่น linear, logistic, Poisson หรือ Cox regression (Allison, 2009)

2) Pairwise deletion: เป็นวิธีการจัดการกับค่าผิดพลาดสำหรับกรณีที่ทำการวิเคราะห์ความสัมพันธ์ระหว่างคู่ตัวแปร โดยจะทำการลบข้อมูลในส่วนที่เป็นค่าสูญหายของทั้งสองตัวแปรทิ้งและจะทำการวิเคราะห์ข้อมูลจากข้อมูลส่วนที่มีค่าสมบูรณ์ทั้งสองตัวแปร วิธี Pairwise deletion เป็นวิธีที่ได้รับความนิยมมาก โดยเฉพาะเมื่อนำมาใช้กับการวิเคราะห์โมเดลเชิงเส้น (linear model) แบบต่าง ๆ เช่น การวิเคราะห์ถดถอย หรือการวิเคราะห์โมเดลสมการโครงสร้าง (structural equation modeling) เหมาะกับการประมาณพารามิเตอร์ที่หลากหลาย เช่น ค่าเฉลี่ย ความแปรปรวน หรือ ความแปรปรวนร่วม รวมทั้งค่าสหสัมพันธ์ เมื่อข้อมูลมีลักษณะแบบ MCAR วิธีการ Pairwise deletion จะให้ค่าประมาณที่ไม่ลำเอียง (Kang, 2013) เช่นเดียวกับวิธี Listwise deletion ถ้าข้อมูลมีลักษณะแบบ MAR แต่ไม่เป็นแบบ MCAR การประมาณพารามิเตอร์อาจให้ค่าประมาณที่ลำเอียง วิธี Pairwise deletion โดยปกติแล้วจะยังคงให้ประสิทธิภาพมากกว่าวิธี Listwise deletion เนื่องจากจะได้ข้อมูลมากกว่าในการประมาณค่า อย่างไรก็ตามนักวิจัยอาจไม่สามารถวิเคราะห์ข้อมูลได้ เนื่องจากอาจเกิดปัญหาการไม่สามารถระบุความเป็นไปได้ในทางบวก (not a definite positive) และไม่สามารถ inverted เพื่อประมาณพารามิเตอร์ได้ ปัญหาดังกล่าวทำให้เป็นการยากที่จะได้ค่าความคลาดเคลื่อนมาตรฐานที่แม่นยำ เนื่องจากค่าความแปรปรวนร่วมหรือค่าสหสัมพันธ์จะถูกประมาณจากขนาดตัวอย่างที่แตกต่างกัน (Van Praag et al.,1985)

4.2.2 การแทนค่า (imputation)

1) Mean substitution: เป็นวิธีการแทนค่าผิดพลาดด้วยค่าเฉลี่ยของตัวแปรนั้นมาแทนค่าสูญหาย (Kantardzic, 2003) การแทนค่าที่เรียกว่า imputation คือการนำค่าเฉลี่ยของตัวอย่างนั้นมาแทนค่าสูญหาย ถ้าตัวอย่างในกลุ่มใดก็แทนค่าด้วยคะแนนเฉลี่ยของกลุ่มนั้น ซึ่งเป็นวิธีที่พัฒนามาจากการแทนค่าค่าผิดพลาดด้วยค่าเฉลี่ยของข้อมูลที่ทราบค่า ทั้งนี้ เนื่องจากข้อสมมติฐานที่ว่าค่าของค่าผิดพลาดควรจะต้องขึ้นอยู่กับลักษณะของหน่วยตัวอย่าง โดยลักษณะของหน่วยตัวอย่างที่ใกล้เคียงกันควรจะมีค่าข้อมูลที่สนใจคล้ายคลึงกัน แต่ผลที่ตามมาคือจะทำให้ค่าเบี่ยงเบนมาตรฐานลดลง วิธีนี้จะทำให้ได้ค่าประมาณที่ลำเอียง (Haitovsky, 1968)

2) Regression method: ทำการสร้างสมการถดถอยระหว่างตัวแปรใด ๆ ที่ต้องการจากข้อมูลที่สมบูรณ์ โดยกำหนดให้ตัวแปรตามเป็นตัวแปรที่มีข้อมูลไม่สมบูรณ์ จากนั้นใช้สมการถดถอยที่ได้ทำการประมาณค่าของข้อมูลที่ไม่สมบูรณ์ วิธีนี้เป็นวิธีการแทนค่าแบบหนึ่งแต่ให้ค่าประมาณค่าที่ดีกว่าแทนค่าด้วยค่าเฉลี่ย (Allison, 2002; Chen & Chang, 2024; Hernández-Herrera et al., 2024)

3) Hot deck imputation: เป็นวิธีการพิจารณาเลือกหน่วยตัวอย่างที่มีลักษณะคล้ายคลึงกันมากที่สุดกับหน่วยตัวอย่างที่เกิดค่าสูญหาย จากนั้นแทนค่าที่สูญหายด้วยค่าของหน่วยตัวอย่างที่คล้ายคลึงนั้น (Andridge & Little, 2010; Wang et al., 2022)

4) Multiple imputation (MI): เป็นวิธีการที่ผสมผสานระหว่างวิธีการ EM และ Raw maximum likelihood methods ร่วมกับความสามารถของนักวิจัยสมบัติ hot deck เพื่อทำการสร้างชุดจำลองของข้อมูลที่ได้ทำการแทนค่าค่าผิดพลาดด้วย imputed value แล้วขึ้นมาหลาย ๆ ชุด (ประมาณ 5-10 ชุด) จากนั้นจะทำการวิเคราะห์ข้อมูลจากชุดต่าง ๆ บันทึกผลการวิเคราะห์ที่ได้เหล่านี้จะถูกรวมเข้าด้วยกัน ปัญหาที่พบในการใช้วิธีการแทนค่า คือ ค่าความคลาดเคลื่อนมาตรฐานได้ค่าประมาณที่ต่ำกว่าความเป็นจริง ทำให้ได้ค่าสถิติที่มีค่าสูงและได้ค่านัยสำคัญทางสถิติที่ต่ำเกินไป (อ่านเพิ่มเติมได้จาก Madley-Dowd et al., 2019; Pedersen et al., 2017; Wang et al., 2022; Woods et al., 2024)

2.2.3 วิธีความเป็นไปได้สูงสุด (maximum likelihood)

1) Expectation Maximization (EM) approach: วิธีการนี้เป็นการอาศัยหลักของกระบวนการวนซ้ำ (iterative procedure) ระหว่าง 2 ขั้นตอน โดยขั้นตอนแรก เป็นขั้นตอนที่เรียกว่า Expectation (E) step ซึ่งจะทำการประมาณค่าคาดหวังจากฟังก์ชัน likelihood ภายใต้ข้อมูลที่สมบูรณ์ สำหรับขั้นตอนที่สอง เป็นขั้นตอนที่

เรียกว่า Maximization (M) step เพื่อทำการแทนค่าคาดหวังของค่าผิดพลาดด้วยค่าที่ได้จาก E step และทำการประมาณค่าคาดหวังจากฟังก์ชัน likelihood ในกรณีถ้าไม่เกิดค่าผิดพลาด โดยจะทำการทวนซ้ำ ระหว่าง 2 ขั้นตอนจนกว่าจะเกิดค่าที่ลู่เข้า (convergence) หรือค่าที่มีการเปลี่ยนแปลงน้อยมาก ใช้ค่านั้นแทนค่าค่าผิดพลาดที่เกิดขึ้น

2) Raw maximum likelihood methods: เป็นวิธีการที่อาศัยข้อมูลสมบูรณ์ในการสร้างค่า maximum likelihood ภายใต้ตัวแบบทางสถิติที่เหมาะสม ไม่ว่าจะเป็นโมเดลสมการโครงสร้าง โมเดลถดถอย ANOVA หรือ ANCOVA วิธีนี้เหมาะกับกรณีค่าผิดพลาดแบบ NMAR

ผลที่อาจเป็นไปได้ในการจัดการค่าผิดพลาดต่อการประมาณพารามิเตอร์ มีดังนี้ 1) กรณีที่ข้อมูลมีค่าสูญหายแบบ MCAR การลบตัวอย่างออกไปจะเป็นวิธีการที่ง่ายที่สุด และมีประมาณค่าอย่างไม่ลำเอียง 2) ถ้าข้อมูลเป็นแบบ MAR ใช้วิธีการทดแทนข้อมูล ยังเป็นวิธีการประมาณค่าแบบไม่ลำเอียง และ 3) ถ้าข้อมูลเป็นแบบ MNAR การประมาณค่าแบบไม่ลำเอียงเป็นสิ่งที่เป็นไปไม่ได้ แต่วิธีการแก้ไขแบบใหม่ ๆ ในปัจจุบัน จะช่วยลดการลำเอียงในการประมาณค่า

นักวิจัยสามารถตรวจสอบค่าผิดพลาด และจัดการได้ด้วยโปรแกรมทางสถิติ ตัวอย่างที่ใช้เป็นข้อมูลประกอบด้วยตัวแปร Y X1 X2 X3 ข้อมูลที่ได้จากตัวอย่าง 156 คน โดยตรวจสอบและจัดการด้วยโปรแกรม Rเมื่อตรวจสอบค่าสถิติพื้นฐานได้ค่า ดังนี้ (download ข้อมูลและคำสั่งจาก QR code)

รายการ 1-1 คำสั่งและผลลัพธ์การตรวจค่าสถิติพื้นฐาน

```
> setwd("/Users/idsaratt/Documents/Book Resoruces/Lesson 1")
> library(readr)
> missingdata <- read_csv("missingdata.csv")
> View(missingdata)
> summary(missingdata)
```

```
       Y                 X1                X2               X3
 Min.   :  1.00   Min.   :  7.0   Min.   :1.700   Min.   : 57.00
 1st Qu.: 18.00   1st Qu.: 115.8   1st Qu.: 7.400   1st Qu.: 73.00
 Median : 31.50   Median : 205.0   Median : 9.700   Median : 79.00
 Mean   : 42.13   Mean   : 185.9   Mean   : 9.806   Mean   : 78.28
 3rd Qu.: 63.25   3rd Qu.: 258.8   3rd Qu.:11.500   3rd Qu.: 85.00
 Max.   :168.00   Max.   : 334.0   Max.   :20.700   Max.   : 97.00
 NA's   : 37      NA's   : 7       NA's   : 7       NA's   : 5
```

จากข้อมูลจะพบว่าตัวแปร Y X1 X2 และ X3 มีค่าผิดพลาด (ดู NA's) โดยเฉพาะ Y มีค่าผิดพลาดถึง 37 ค่า X1 และ X2 ตัวแปรละ 7 ค่า ส่วน X3 มีจำนวน 5 ค่า รวมค่าผิดพลาดทั้งหมด 55 ค่า เห็นได้ว่าส่วนใหญ่ค่าผิดพลาดมาจากตัวแปร Y โดยทั่วไปถ้าพบว่า ค่าผิดพลาดเป็นแบบ MCAR มากเกินไปอาจเป็นปัญหาได้เช่นกัน โดยปกติเกณฑ์สูงสุดที่ปลอดภัย คือ ไม่เกินร้อยละ 5 ของจำนวนทั้งหมดสำหรับชุดข้อมูลขนาดใหญ่ หากข้อมูลที่ขาดหายไปสำหรับบางตัวอย่างเกินร้อยละ 5 อาจจะนำตัวอย่างนั้นออกจากการวิเคราะห์ ดังนั้น นักวิจัยจึงตรวจสอบนักวิจัยสมบัติ (คอลัมน์) และตัวอย่าง (แถว) ซึ่งข้อมูลมากกว่าร้อยละ 5 หายไปโดยใช้ฟังก์ชันง่าย ๆ

รายการ 1-2 คำสั่งและผลลัพธ์การตรวจสอบค่าผิดพลาดเบื้องต้น

```
> is.na(missingdata)
```

```
          Y     X1     X2     X3
[1,]  FALSE  FALSE  FALSE   TRUE
[2,]  FALSE  FALSE  FALSE   TRUE
[3,]  FALSE  FALSE  FALSE   TRUE
[4,]  FALSE  FALSE   TRUE   TRUE
[5,]   TRUE   TRUE   TRUE   TRUE
[6,]  FALSE   TRUE   TRUE  FALSE
[7,]  FALSE  FALSE   TRUE  FALSE
```

```
   [8,]  FALSE  FALSE   TRUE  FALSE
   [9,]  FALSE  FALSE   TRUE  FALSE
  [10,]   TRUE  FALSE   TRUE  FALSE
   . . .
 [142,] FALSE  FALSE  FALSE  FALSE
 [143,] FALSE  FALSE  FALSE  FALSE
 [144,] FALSE  FALSE  FALSE  FALSE
 [145,] FALSE  FALSE  FALSE  FALSE
 [146,] FALSE  FALSE  FALSE  FALSE
 [147,] FALSE  FALSE  FALSE  FALSE
 [148,] FALSE  FALSE  FALSE  FALSE
 [149,] FALSE  FALSE  FALSE  FALSE
 [150,]  TRUE  FALSE  FALSE  FALSE
 [151,] FALSE  FALSE  FALSE  FALSE
 [152,] FALSE  FALSE  FALSE  FALSE
 [153,] FALSE  FALSE  FALSE  FALSE
```

เมื่อใช้คำสั่งตามรายการจะเป็นการถามว่าค่าสังเกตใดที่มีค่าผิดพลาด ถ้า FALSE หมายถึง ไม่ใช่ค่าผิดพลาด แต่ถ้าตอบ TRUE หมายถึง เป็นค่าผิดพลาด เช่น ตัวแปร Y ค่าสังเกตที่มีค่าผิดพลาดค่าแรก คือค่าสังเกตที่ 5

รายการ 1-3 คำสั่งและผลลัพธ์การตรวจสอบร้อยละค่าผิดพลาด

```
> library(lattice)
> pMiss <- function(x){sum(is.na(x))/length(x)*100}
> apply(missingdat,2,pMiss)
```

```
       Y        X1        X2        X3
24.183007  4.575163  4.575163  3.267974
```

```
> apply(missingdata,1,pMiss)
```

```
  1]   25  25  25  50 100  50  25  25  25  50  25   0   0   0   0   0   0
 [18    0   0   0   0   0   0   0  25  25  50   0   0   0   0  25  25  25
[35]   25  25  25   0  25   0   0  25  25   0  25  25   0   0   0   0   0
[52]   25  25  25  25  25  25  25  25  25  25   0   0   0  25   0   0   0
[69]    0   0   0  25   0   0  25   0   0   0   0   0   0   0  25  25   0
[86]    0   0   0   0   0   0   0   0   0   0  25  25  25   0   0   0  25
[103]  25   0   0   0  25   0   0   0   0   0   0   0  25   0   0   0  25
[120]   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
[137]   0   0   0   0   0   0   0   0   0   0   0   0   0  25   0   0   0
```

จากข้อมูลจะพบว่า ตัวอย่างที่มีค่าผิดพลาดมากที่สุด คือ ตัวอย่างที่ 5 โดยมีค่าผิดพลาด ตัวแปรจาก 4 ตัวแปร คิดเป็นร้อยละ 100 รองลงมาเป็นตัวอย่างที่ 4 6 10 และ 27 มีค่าผิดพลาดร้อยละ 50 คือ มีค่าผิดพลาด 2 ใน 4 ตัวแปร ส่วนที่เหลือที่เกิดค่าผิดพลาดเกิดจากตัวแปรเดียว หรืออยู่ที่ร้อยละ 25 นักวิจัยทำให้ค่าผิดพลาดนี้มีค่าร้อยละลดลงได้โดยการเก็บข้อมูลเพิ่มเติม หรือทำให้ลดลงโดยหลักการทางสถิติดังได้กล่าวไปแล้ว ส่วนตัวอย่างอื่นที่ต่ำกว่าเกณฑ์ร้อยละ 5 นักวิจัยสามารถเก็บไว้เพื่อนำไปวิเคราะห์ข้อมูล แต่จากข้อมูลพบว่า บางตัวอย่างที่มีค่าผิดพลาดค่อนข้างมาก (ร้อยละ 50 ขึ้นไป) เช่น ตัวอย่างที่ 4 5 6 10 และ 28 ควรนำออกจากการวิเคราะห์ข้อมูลถ้าเป็นไปได้ และไม่มีผลกระทบต่อขนาดตัวอย่างในการวิเคราะห์มากนัก

นักวิจัยสามารถตรวจสอบแบบแผนของค่าผิดผลาดได้จาก mice package ดังนี้