

สรุปผลการวิจัย และข้อเสนอแนะ

การวิจัยครั้งนี้มีวัตถุประสงค์ เพื่อศึกษาเปรียบเทียบความน่าเชื่อถือของโปรแกรมสำเร็จรูปในงานวิเคราะห์ทางสถิติ จำนวน 5 โปรแกรม โดยได้แบ่งโปรแกรมทั้ง 5 โปรแกรมเป็น 2 ส่วน คือ โปรแกรมที่ต้องเขียนคำสั่งในการวิเคราะห์ข้อมูลขึ้นมาเอง (User programming) ประกอบด้วย โปรแกรม SAS รุ่น 9 และ R รุ่น 2.6.2 และโปรแกรมสำเร็จรูป (Package program) ประกอบด้วย โปรแกรม Microsoft Excel รุ่น 2007, SPSS รุ่น 15 และ Minitab รุ่น 15 โดยใช้ข้อมูลตามเกณฑ์ของ The National Institute of Standards and Technology (NIST) ซึ่งแบ่งระดับความยากของข้อมูลเป็น 3 ระดับคือ ระดับต่ำ ระดับปานกลาง และระดับสูง ผู้วิจัยทำการศึกษาการประมาณค่าทางสถิติ เมื่อให้โปรแกรมทั้ง 5 โปรแกรม ประมาณค่าทางสถิติประเภทเดียวกัน โดยกำหนดกำหนดหน่วยสำหรับวัดความน่าเชื่อถือสำหรับการวิเคราะห์ข้อมูล เพื่อพิจารณาว่าโปรแกรมใดให้ความน่าเชื่อถือสูงสุดในการประมาณค่าทางสถิติ ซึ่งการวิจัยศึกษาค่าประมาณของการวิเคราะห์ ดังนี้

- 1.การวิเคราะห์สถิติตัวแปรเดียว
- 2.การวิเคราะห์ความแปรปรวนจำแนกทางเดียว
- 3.การวิเคราะห์การถดถอย

5.1 สรุปผลการวิจัย

5.1.1การวิเคราะห์สถิติตัวแปรเดียว

จากการเปรียบเทียบความแม่นยำของการประมาณค่าเฉลี่ยของโปรแกรมทั้ง 5 ในการวิเคราะห์สถิติตัวแปรเดียว พบว่าข้อมูลที่มีความยากระดับต่ำ คือ ข้อมูลที่มีจำนวนหลักของค่าคงตัวหน้าทศนิยม หรือจำนวนค่าสังเกตน้อย ค่าเฉลี่ยของข้อมูลในการวิเคราะห์ด้วยโปรแกรมทั้ง 5 โปรแกรมให้ความแม่นยำสูงสุด สำหรับข้อมูลที่มีความยากระดับปานกลางและสูง ซึ่งเป็นข้อมูลที่มีจำนวนหลักของค่าคงตัวหน้าทศนิยม และจำนวนค่าสังเกตมากขึ้น ตามลำดับ พบว่าโปรแกรม SAS, R, SPSS และ Minitab ให้ความแม่นยำสูงสุด แม้ว่าจำนวนหลักของค่าคงตัวหน้าทศนิยม และจำนวนค่าสังเกตจะมีค่าสูงขึ้น

และจากการเปรียบเทียบความแม่นยำของส่วนเบี่ยงเบนมาตรฐานในการวิเคราะห์สถิติตัวแปรเดียวของข้อมูลที่มีความยากระดับต่ำ พบว่า ทุกชุดข้อมูลของทุกโปรแกรมจะให้ค่าความแม่นยำในระดับสูง สำหรับข้อมูลที่มีความยากระดับปานกลาง ในชุดข้อมูลที่มีจำนวนหลักของค่าคงตัวหน้าทศนิยมน้อย (NumAcc2) ให้ค่าความแม่นยำในระดับสูง ยกเว้นแต่โปรแกรม Excel ส่วนข้อมูลที่มีจำนวนหลักของค่าคงตัวหน้าทศนิยมมากขึ้น (NumAcc3) ให้ค่าความแม่นยำในระดับปานกลาง โดยทุกโปรแกรมให้ค่าใกล้เคียงกัน แต่โปรแกรม SPSS ให้ค่าสูงกว่าโปรแกรมอื่นเล็กน้อย และสำหรับข้อมูลที่มีความยากระดับสูงโปรแกรมทั้ง 5 ให้ค่าความแม่นยำในระดับต่ำและมีค่าใกล้เคียงกัน แต่โปรแกรม SAS ให้ค่าสูงกว่าโปรแกรมอื่นเล็กน้อย

จากการวิเคราะห์สถิติตัวแปรเดียว สรุปได้ว่า ค่าความแม่นยำของการประมาณค่าเฉลี่ยของโปรแกรมไม่ขึ้นอยู่กับระดับความยากของข้อมูล แต่ความแม่นยำของค่าส่วนเบี่ยงเบนมาตรฐานของโปรแกรมขึ้นอยู่กับระดับความยากของข้อมูล คือเมื่อข้อมูลมีความยากมากขึ้น คือจำนวนหลักของค่าคงตัวหน้าทศนิยมหรือจำนวนค่าสังเกตมีจำนวนมากขึ้น จะทำให้ความแม่นยำของโปรแกรมมีค่าต่ำลง อีกทั้งยังพบว่าค่าความแม่นยำของการประมาณค่าส่วนเบี่ยงเบนมาตรฐานมีค่าต่ำกว่าค่าความแม่นยำของการประมาณค่าเฉลี่ย ซึ่งสาเหตุที่ทำให้ความแม่นยำของค่าส่วนเบี่ยงเบนมาตรฐานมีค่าต่ำนั้นเนื่องมาจากสูตรที่ใช้ในการคำนวณส่วนเบี่ยงเบนมาตรฐานมีความซับซ้อนกว่าค่าเฉลี่ย ซึ่งอาจเกิดความผิดพลาดจากจากขั้นตอนการคำนวณทางคณิตศาสตร์ของโปรแกรมทั้ง 5 โปรแกรม ดังนั้นสำหรับการเลือกใช้โปรแกรมสำเร็จรูปในการวิเคราะห์สถิติตัวแปรเดียวทุกโปรแกรมมีความน่าเชื่อถือ และประสิทธิภาพไม่แตกต่างกันมากนัก เนื่องจากการวิเคราะห์สถิติตัวแปรเดียวเป็นการวิเคราะห์พื้นฐานที่ความสลับซับซ้อนในด้านสูตรและกระบวนการหาค่าผลลัพธ์ไม่ก่อให้เกิดความแตกต่างมากนัก

5.1.2 การวิเคราะห์ความแปรปรวนจำแนกทางเดียว

จากการเปรียบเทียบความแม่นยำของสถิติทดสอบเอฟ และค่าคลาดเคลื่อนกำลังสองเฉลี่ยของโปรแกรมทั้ง 5 ในวิเคราะห์ความแปรปรวนจำแนกทางเดียว พบว่าทุกชุดของข้อมูล ที่มีระดับยากต่ำและระดับยากปานกลาง โปรแกรม Minitab ให้ค่าความแม่นยำสูงสุดคือ 15

เมื่อพิจารณาค่าความแม่นยำของสถิติทดสอบเอฟ และค่าคลาดเคลื่อนกำลังสองเฉลี่ยของโปรแกรม สามารถสรุปได้ว่าความแม่นยำของโปรแกรมขึ้นอยู่กับจำนวนค่าสังเกต คือ เมื่อจำนวนค่าสังเกตมากขึ้นค่าความแม่นยำของค่าประมาณมีแนวโน้มต่ำลง นอกจากนี้ยังพบอีกว่าถ้าระดับ

ความยากของข้อมูลมีมากขึ้น ค่าความแม่นยำของค่าประมาณมีแนวโน้มต่ำลงด้วย กล่าวคือ จำนวนหลักของค่าคงตัวหน้าทศนิยมมีจำนวนมากขึ้นจะทำให้ค่าความแม่นยำต่ำมีค่าลง

จากการวิเคราะห์ความแปรปรวนจำแนกทางเดียวสรุปได้ว่าโปรแกรม SAS และ R มีความน่าเชื่อถือ เนื่องจากมีค่าความแม่นยำสูงกว่าโปรแกรมอื่น ยกเว้นโปรแกรม Minitab ซึ่งจากการวิเคราะห์พบว่าถึงแม้โปรแกรม Minitab จะมีความแม่นยำสูงสุด แต่ความแม่นยำที่มีค่าสูงเกินจริงนั้น อาจเกิดจากการที่โปรแกรม Minitab มีการแสดงผลที่จำกัด จึงทำให้ค่าความแม่นยำที่ได้สูงเกินจริง

5.1.3 การวิเคราะห์การถดถอย

ความแม่นยำของค่าสัมประสิทธิ์การถดถอยของแต่ละโปรแกรมของข้อมูลที่มีความยากระดับต่ำ ให้ความแม่นยำในการประมาณค่าใกล้เคียงกัน โดย b_1 ให้ความแม่นยำสูงสุดทุกโปรแกรม แต่ b_0 ไม่ได้ให้ค่าความแม่นยำสูงสุด เนื่องจากเป็นพารามิเตอร์ที่ไม่มีความแปรปรวนเพียงพอ และข้อมูลที่มีความยากระดับปานกลาง คือ สมการถดถอยผ่านจุดกำเนิด ค่าความแม่นยำมีค่าสูงสุดทุกโปรแกรม

เมื่อข้อมูลที่มีความยากระดับสูง (Wampler1, Wampler2, Wampler3, Wampler4 และ Wampler5) คือข้อมูลเกิดปัญหาวางระฆังเส้นตรง ค่าความแม่นยำขึ้นอยู่กับลักษณะของข้อมูล และโปรแกรมที่ใช้ในการวิเคราะห์ กล่าวคือ ชุดข้อมูล Wampler2 ให้ค่าความแม่นยำของสัมประสิทธิ์การถดถอยทุกตัวสูงกว่าข้อมูลชุดอื่นที่มีระดับความยากสูง เนื่องจากข้อมูลชุดนี้มีค่าคงตัวหน้าทศนิยมแค่ 1 และ 2 ตำแหน่ง ในขณะที่ชุดข้อมูล Wampler1 มีค่าคงตัวตั้งแต่ 1 ถึง 7 ตำแหน่ง แต่เมื่อข้อมูลมีการกระจายมากขึ้นในชุดข้อมูล Wampler3, Wampler4 และ Wampler5 ตามลำดับ ค่าความแม่นยำของค่าสัมประสิทธิ์ของการถดถอยมีแนวโน้มต่ำลง เมื่อเปรียบเทียบค่าความแม่นยำของแต่ละโปรแกรมจะพบว่า ข้อมูลมีการกระจายมากขึ้นโปรแกรม SPSS ค่าความแม่นยำต่ำกว่าโปรแกรมอื่นอย่างเห็นได้ชัด

ค่าความแม่นยำของสถิติทดสอบเอฟในการวิเคราะห์การถดถอย สำหรับความยากของข้อมูลทุกระดับ พบว่าโปรแกรม SAS, R และ Excel ให้ค่าความแม่นยำอยู่ในระดับสูง แต่โปรแกรม SPSS และ Minitab เมื่อข้อมูลที่มีความยากระดับปานกลาง จะให้ค่าความแม่นยำอยู่ในระดับสูง และเมื่อข้อมูลที่มีความยากระดับสูง จะให้ค่าความแม่นยำอยู่ในระดับต่ำ

เมื่อพิจารณาค่าความแม่นยำของค่าสัมประสิทธิ์ของการถดถอย และค่าความแม่นยำของสถิติทดสอบเอฟในการวิเคราะห์การถดถอยขึ้นอยู่กับความยากของข้อมูล โดยที่ข้อมูลที่มีความ

ยากระดับต่ำจะให้ค่าความแม่นยำ และความแม่นยำจะลดลงตามระดับความยากของข้อมูลที่สูงขึ้น โดยเฉพาะโปรแกรม SPSS ซึ่งจะเห็นได้อย่างชัดเจน

ค่าความแม่นยำของค่าสัมประสิทธิ์ของการถดถอยขึ้นอยู่กับตัวแบบการถดถอย กล่าวคือ ตัวแบบการถดถอยอย่างง่ายทั้งสมการพหุคูณที่ผ่านจุดกำเนิด และไม่ผ่านจุดกำเนิด จะมีค่าความแม่นยำสูงกว่าตัวแบบที่เป็นเชิงเส้นโค้งแบบพหุนาม แต่ในทางกลับกันค่าความแม่นยำของเอฟในการวิเคราะห์การถดถอย นั้นไม่ขึ้นอยู่กับตัวแบบ ในการเปรียบเทียบความน่าเชื่อถือของแต่ละโปรแกรม พบว่าโปรแกรม SPSS ในกรณีตัวแบบที่มีปัญหาภาวะร่วมเส้นตรง และเป็นตัวแบบเชิงเส้นโค้งแบบพหุนาม มีความน่าเชื่อถือค่อนข้างต่ำกว่าโปรแกรมอื่นๆ อย่างเห็นได้ชัด และในชุดข้อมูลที่มีการกระจายไม่มากนัก คือ ข้อมูลที่มีความยากระดับสูง (Wampler1, Wampler2 และ Wampler3) โปรแกรม SAS ให้ค่าความแม่นยำสูงกว่าโปรแกรมอื่น แต่ในชุดข้อมูลที่มีการกระจายมากขึ้น (Wampler4 และ Wampler5) โปรแกรม Minitab ให้ค่าความแม่นยำสูงกว่าโปรแกรมอื่น

5.2 การอภิปรายผล

(1) ผลลัพธ์ที่ได้จากการคำนวณค่าโดยใช้โปรแกรมที่ต้องเขียนคำสั่งในการวิเคราะห์ข้อมูลขึ้นมาเอง (User programming) ผู้ใช้สามารถเขียนคำสั่งให้นำผลลัพธ์ที่ได้จากการคำนวณค่าไปใช้ในการคำนวณต่อไปได้ แต่ผลลัพธ์ที่ได้จากโปรแกรมสำเร็จรูป (Package program) โดยส่วนมากผู้ใช้งานต้องทำการคัดลอกผลลัพธ์นั้น เพื่อนำค่านั้นไปคำนวณต่อไป ค่าที่คัดลอกออกมานั้น อาจไม่ใช่ค่าที่แท้จริง เพราะอาจเกิดความคลาดเคลื่อนของตัวเลขขึ้น

(2) สำหรับชุดข้อมูล Wampler1 และ Wampler2 ซึ่งมีค่ามาตรฐานของสถิติทดสอบเอฟเท่ากับค่าอนันต์ ทำให้ไม่สามารถใช้การคำนวณหาค่า LRE เพื่อเปรียบเทียบความแม่นยำของสถิติทดสอบเอฟของการวิเคราะห์การถดถอยของแต่ละโปรแกรมได้ แต่จากการศึกษาพบว่าโปรแกรม R และ Excel นั้นสามารถแสดงค่าสถิติทดสอบเอฟได้ในตำแหน่งที่ 32 และ 33 ตามลำดับ และโปรแกรม SPSS นั้นไม่สามารถแสดงค่าสถิติทดสอบเอฟ เมื่อพิจารณาการหาค่าสถิติทดสอบเอฟ ซึ่งหาได้จากค่ากำลังสองเฉลี่ยของทรีตเมนต์หารค่าคลาดเคลื่อนกำลังสองเฉลี่ย โดยใช้ค่าที่ได้จากตารางแสดงผลของโปรแกรม SPSS จะได้ค่าสถิติทดสอบเอฟของชุดข้อมูล Wampler1 และ Wampler2 เท่ากับ $2.8003E+32$ และ $4.1449E+32$ ตามลำดับ ซึ่งการที่ตาราง

แสดงผลของโปรแกรม SPSS ไม่แสดงผลดังกล่าวนี้พบว่าการแสดงผลมีการจำกัดอยู่เพียง 16 ตำแหน่งเท่านั้น

นั่นหมายความว่าโปรแกรม R, Excel และ SPSS มีกระบวนการคำนวณที่รองรับค่าที่มีความละเอียดได้มาก ซึ่งทำให้การแสดงผลคำนวณมีความถูกต้องแม่นยำมากขึ้น

(3) กรณีการวิเคราะห์ความแปรปรวนจำแนกทางเดียวที่ข้อมูลมีความยากระดับสูง ที่โปรแกรม Minitab ไม่สามารถแสดงค่าสถิติทดสอบเอฟนั้น อธิบายได้ดังนี้

ตาราง 5.1

แสดงค่าเฉลี่ยของแต่ละทรีตเมนต์ที่ระดับความยากสูง

จากชุดข้อมูล SmLs07, SmLs08 และ SmLs09

ชุดข้อมูล ค่าเฉลี่ย	SmLs07	SmLs08	SmLs09
$\bar{Y}_1$	1000000000000.4	1000000000000.4	1000000000000.4
$\bar{Y}_2$	1000000000000.3	1000000000000.3	1000000000000.3
$\bar{Y}_3$	1000000000000.5	1000000000000.5	1000000000000.5
$\bar{Y}_4$	1000000000000.3	1000000000000.3	1000000000000.3
$\bar{Y}_5$	1000000000000.5	1000000000000.5	1000000000000.5
$\bar{Y}_6$	1000000000000.3	1000000000000.3	1000000000000.3
$\bar{Y}_7$	1000000000000.5	1000000000000.5	1000000000000.5
$\bar{Y}_8$	1000000000000.3	1000000000000.3	1000000000000.3
$\bar{Y}_9$	1000000000000.5	1000000000000.5	1000000000000.5

ตาราง 5.2

แสดงส่วนเบี่ยงเบนมาตรฐานของแต่ละทรีตเมนต์ที่ระดับความยากสูง
ที่ได้จากชุดข้อมูล SmLs07, SmLs08 และ SmLs09

ชุดข้อมูล S	SmLs07	SmLs08	SmLs09
S_1	0.0999755859	0.0999755859	0.0999755859
S_2	0.1000321672	0.1000358473	0.1000365243
S_3	0.0999755859	0.0999755859	0.0999755859
S_4	0.1000321672	0.1000358473	0.1000365243
S_5	0.0999755859	0.0999755859	0.0999755859
S_6	0.1000321672	0.1000358473	0.1000365243
S_7	0.0999755859	0.0999755859	0.0999755859
S_8	0.1000321672	0.1000358473	0.1000365243
S_9	0.0999755859	0.0999755859	0.0999755859

สำหรับกรณีข้อมูลมีความยากระดับสูงพบว่าโปรแกรม Minitab ให้ค่าคลาดเคลื่อนกำลังสองเฉลี่ยเท่ากับ 0 จึงทำให้ค่าสถิติทดสอบเอฟ ในการวิเคราะห์ความแปรปรวนจำแนกทางเดียว ซึ่งเกิดจากสัดส่วนของค่ากำลังสองเฉลี่ยของทรีตเมนต์กับค่าคลาดเคลื่อนกำลังสองเฉลี่ย ไม่สามารถหาค่าได้ ซึ่งค่าคลาดเคลื่อนกำลังสองเฉลี่ยของที่ได้จากโปรแกรม Minitab ที่มีความแตกต่างจากค่ามาตรฐานนั้น เมื่อพิจารณาการคำนวณค่าคลาดเคลื่อนกำลังสองเฉลี่ยจาก

$$MS_E = \frac{SS_E}{N-t} \text{ และ } SS_E = \sum_{i=1}^t \sum_{j=1}^{r_i} (Y_{ij} - \bar{Y}_{i.})^2 \text{ จะเห็นได้ว่าค่าคลาดเคลื่อนกำลังสองเฉลี่ยจะมี}$$

ค่ามากถ้าความแตกต่างของค่าสังเกตกับค่าเฉลี่ยในทรีตเมนต์มีมาก ซึ่งจากตาราง 5.1 และ 5.2 แสดงให้เห็นว่าความความผันแปรในทรีตเมนต์หรือค่าคลาดเคลื่อนกำลังสองเฉลี่ยมีน้อยมาก กอปรกับความละเอียดในกระบวนการคำนวณของโปรแกรม Minitab ซึ่งค่อนข้างต่ำ จึงทำให้ค่าคลาดเคลื่อนกำลังสองเฉลี่ยมีค่าเท่ากับ 0 ซึ่งสังเกตได้จากค่าสถิติทดสอบเอฟที่ไม่สามารถคำนวณได้ในชุดข้อมูล SmLs07, SmLs08 และ SmLs09 จากโปรแกรม Minitab (ตารางที่ 4.7) และจากการวิเคราะห์ พบว่าโปรแกรม Minitab ไม่มีความทนทานต่อข้อมูลที่มีค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำ

5.3 ข้อเสนอแนะ

การวิจัยครั้งต่อไป ผู้วิจัยเห็นว่า

(1) ควรกำหนดตำแหน่งผลลัพธ์ให้มากกว่าตำแหน่งผลลัพธ์ ที่ต้องการเพื่อที่จะเพิ่มความแม่นยำให้มากขึ้น

(2) ควรศึกษาโปรแกรมทางสถิติในด้านอื่นๆ เพิ่มเติม เช่น โปรแกรม Stata, Winonlin, Matlab เป็นต้น เพราะในแต่ละหน่วยงานก็จะมีความชำนาญในแต่ละโปรแกรมต่างกัน

(3) ควรศึกษาว่าโปรแกรมในแต่ละรุ่น เช่นอาจเปรียบเทียบ โปรแกรม SPSS ตั้งแต่รุ่น 10 ถึง 15 เป็นต้น เพราะหากรุ่นสูงขึ้นไม่ได้ให้ความแม่นยำในการประมาณค่าสูงขึ้น มีเพียงแต่ให้ฟังก์ชันพิเศษมากขึ้น แต่หากผู้ใช้งาน ใช้เพียงฟังก์ชันพื้นฐานก็ไม่มี ความจำเป็นต้องใช้รุ่นสูงเพราะเนื่องจาก รุ่นที่สูงขึ้นมักสิ้นเปลืองทรัพยากรของเครื่องคอมพิวเตอร์ ทำให้การประมวลผลช้าลง

(4) ควรศึกษาการประมาณค่าทางสถิติอื่นๆ เช่น การวิเคราะห์การถดถอยไม่เชิงเส้น เป็นต้น