บทที่ 1
บทนำ

1. ความเป็นมาและความสำคัญของปัญหา

ในปัจจุบัน ข้อมูลสารสนเทศ (Information Data) มีความสำคัญมาก จึงมีความจำเป็นที่ต้องใช้ข้อมูลสารสนเทศที่มีคุณภาพ เพื่อนำไปวิเคราะห์ วางแผน หรือประกอบการตัดสินใจในการดำเนินกิจกรรมใด ๆให้เกิดประโยชน์สูงสุด ข้อมูลที่ใช้ควรมีความถูกต้องเชื่อถือได้ การวิจัยโดยทั่วไป ผู้วิจัยไม่สามารถศึกษาประชากรได้ทั้งหมด ทั้งนี้อาจเป็นเพราะต้องเสียค่าใช้จ่ายสูง สิ้นเปลืองเวลา แรงงาน และมีความยุ่งยากในการเก็บรวบรวมข้อมูล จึงใช้วิธีการเลือกสุ่มประชากรบางส่วน หรือที่เรียกว่ากลุ่มตัวอย่าง เพื่อใช้เป็นตัวแทนของประชากร กลุ่มตัวอย่างที่ได้นี้จะเป็นตัวแทนที่ดีของประชากรหรือไม่ขึ้นอยู่กับเทคนิคการสุ่มตัวอย่าง ผลที่ได้จากการศึกษาจากกลุ่มตัวอย่างสามารถอ้างอิงไปยังลักษณะเฉพาะของประชากรที่ศึกษา ซึ่งเรียกว่าพารามิเตอร์ (Parameter) บ่อยครั้งที่พบว่าข้อมูลที่ได้จากการเก็บรวบรวมหรือจากการทดลองมีค่าของข้อมูลบางค่าที่แตกต่างไปจากข้อมูลส่วนใหญ่ กล่าวคือ อาจมีบางค่าที่สูงเกินไปหรือต่ำเกินไปเมื่อเทียบกับข้อมูลตัวอื่น ๆซึ่งเรียกค่าเหล่านี้ว่า ค่าผิดปกติจากกลุ่ม (Outlier) ในกรณีที่ประชากรมีการแจกแจงแบบปกติหรือใกล้เคียง แต่มีค่าสังเกตบางค่าสูงมาก หรือมีค่าผิดปกติจากกลุ่ม ตัวอย่างที่สุ่มมาได้อาจมีหน่วยที่มีค่าผิดปกติจากกลุ่มรวมอยู่ด้วย ซึ่งหน่วยที่มีค่าผิดปกติจากกลุ่มนี้จะส่งผลกระทบต่อการประมาณค่าพารามิเตอร์ที่สนใจ ดังจะเห็นได้จากการประมาณค่าเฉลี่ยประชากรจากตัวอย่างสุ่มอย่างง่ายชนิดไม่ใส่คืน (Simple Random Sampling Without Replacement) ด้วยค่าเฉลี่ยเลขคณิตของตัวอย่าง ถ้าใช้วิธีการสุ่มที่อาศัยหลักความน่าจะเป็นแล้ว ตัวอย่างที่สุ่มได้อาจเป็นค่าสังเกตที่เป็นค่าผิดปกติจากกลุ่ม อาจจะเป็นค่าที่สูงมากหรือต่ำมาก ดังนั้นค่าเฉลี่ยเลขคณิตตัวอย่างที่คำนวณได้จะมีค่าสูงหรือต่ำกว่าค่าเฉลี่ยประชากรที่แท้จริง การแก้ปัญหานี้อาจทำได้โดยตัดข้อมูลที่มีค่าสูงสุดและค่าต่ำสุดออกข้างละ 5% เพื่อที่จะลดผลกระทบที่เกิดจากค่าผิดปกติจากกลุ่ม ซึ่งเรียกว่า Trimmed Mean การตัดค่าสูงสุดและต่ำสุดออกต้องแน่ใจว่าปริมาณค่าผิดปกติจากกลุ่มได้ถูกขจัดออกจากการคำนวณหมดแล้ว (ฉัตรศิริ ปิยะพิมลสิทธิ์, 2547) หรืออาจแก้ไขปัญหาโดยตัดค่าสังเกตที่ผิดปกติทิ้งไป ถ้าการกระทำดังกล่าวไม่กระทบต่อผลการวิจัย แต่ในบางกรณีไม่สามารถตัดค่าสังเกตเหล่านี้ทิ้งได้ เพราะจะทำให้การประมาณค่าเฉลี่ยประชากรคลาดเคลื่อนจากความเป็นจริงมาก และในทางตรงกันข้าม การนำค่าสังเกตดังกล่าวมาใช้ ก็จะมีผลให้ค่าประมาณค่าเฉลี่ยเลขคณิตที่ได้มากกว่าค่าที่แท้จริงของประชากรด้วย (อโนทัย ตรีวานิช, 2542) เมื่อเป็นเช่นนี้การวัดแนวโน้มเข้าสู่ส่วนกลางที่เหมาะสมน่าจะเป็นค่ามัธยฐาน และใช้ค่าเฉลี่ยเลขคณิตด้วยวิธี Trimmed Mean จะทำให้การวัดแนวโน้มเข้าสู่ส่วนกลางดูมีเหตุผลมากขึ้น (ฉัตรศิริ ปิยะพิมลสิทธิ์, 2547) ค่าสังเกตที่เป็นค่าผิดปติจากกลุ่มมีโอกาสพบได้ในการวิจัยในทุก ๆสาขา ส่วนผลกระทบที่เกิดขึ้นจากข้อมูลที่มีค่าผิดปกติจากกลุ่ม จะมีความรุนแรงมากน้อยเพียงใดขึ้นอยู่กับองค์ประกอบที่สำคัญคือ ปริมาณค่าผิดปกติจากกลุ่ม ระดับค่าผิดปกติจากกลุ่ม ตำแหน่งหรือตัวแปรที่พบค่าผิดปกติจากกลุ่ม ขนาดตัวอย่างที่ศึกษา และการเลือกใช้สถิติในการวิเคราะห์ที่เหมาะสม เดฟลิน และคณะ (Devlin et al., 1975) ได้สรุปถึงผลกระทบจากค่าผิดปกติจากกลุ่มต่อการประมาณค่าว่า การประมาณค่าพารามิเตอร์จากชุดข้อมูลที่มีค่าผิดปกติจากกลุ่มอยู่ด้วย ทำให้ค่าประมาณที่ได้มีความคลาดเคลื่อนจากความเป็นจริง โดยอาจเป็นค่าประมาณที่สูงหรือต่ำกว่าความจริง และส่งผลกระทบต่อการสรุปผลการมีนัยสำคัญทางสถิติในการทดสอบสมมติฐานอีกด้วย ข้อมูลที่มีค่าผิดปกติจากกลุ่มอาจมีสาเหตุมาจากบุคคล เครื่องมือวัดต่าง ๆหรือมาจากธรรมชาติของข้อมูลเอง การแก้ไขหรือ

การจัดการข้อมูลที่มีค่าผิดปกติจากกลุ่มสามารถทำได้หลายวิธีแต่ยังไม่มีสูตรสำเร็จรูป ทั้งนี้ขึ้นอยู่กับดุลยพินิจของนักวิจัยที่จะหาวิธีการแก้ไขหรือจัดการที่เหมาะสม เช่น ถ้าข้อมูลเกิดจากการบันทึกข้อมูลผิดพลาด วิธีแก้ไขอาจตัดข้อมูลที่มีค่าผิดปกติจากกลุ่มออกหรืออาจทำการปรับแก้ค่าให้เหมาะสม แต่ถ้าข้อมูลที่มีค่าผิดปกติจากกลุ่มได้รับการบันทึกมาอย่างถูกต้อง และผู้วิเคราะห์ไม่สามารถอธิบายถึงสาเหตุของการเกิดค่าผิดปกติจากกลุ่มแล้ว ควรนำค่าผิดปกติจากกลุ่มนั้นมาวิเคราะห์ด้วย เพื่อดูผลกระทบของค่าผิดปกติจากกลุ่มไปจากกลุ่ม ดังเช่นในการทดสอบสมมติฐานให้ทำทั้งในกรณีมีและไม่มีค่าผิดปกติจากกลุ่มรวมอยู่ด้วย ถ้าให้ผลในการวิเคราะห์ข้อมูลที่สอดคล้องกัน จะทำให้นักวิจัยสรุปด้วยความมั่นใจได้ว่าไม่มีผลกระทบจากค่าที่ผิดปกตินั้น แต่ถ้าผลการวิเคราะห์ไม่สอดคล้องกันนักวิจัยจะต้องสรุปผลการวิเคราะห์ทั้งสองกรณี (อรุณ จิรวัฒน์กุล, 2547)

ในการศึกษาวิจัยด้านวิทยาศาสตร์สุขภาพ มีการศึกษาจำนวนมากที่มีจุดมุ่งหมายเพื่อต้องการทราบความสัมพันธ์ระหว่างตัวแปรสองตัวว่ามีความสัมพันธ์กันหรือไม่และเป็นความสัมพันธ์ในทิศทางใด ซึ่งการหาความสัมพันธ์มีอยู่ด้วยกันหลายวิธี ขึ้นอยู่กับลักษณะการแจกแจงและมาตรวัดของตัวแปรนั้น ความสัมพันธ์ระหว่างตัวแปรสองตัวแปรจะมีค่ามากหรือน้อย สามารถสรุปได้จากการคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient) สัญลักษณ์ที่ใช้แทนค่าสัมประสิทธิ์สหสัมพันธ์ของประชากรคือ ρ (Rho) ค่าของ ρ จะมีค่าอยู่ระหว่าง -1ถึง +1 ถ้า ρ มีค่าใกล้ ± 1 แสดงว่าตัวแปรที่ศึกษาทั้งสองมีความสัมพันธ์กันมาก โดยความสัมพันธ์นี้พิจารณาเป็นความสัมพันธ์เชิงเส้นตรง แต่ถ้าค่า ρ มีค่าใกล้ 0 แสดงว่าตัวแปรที่ศึกษาทั้งสองมีความสัมพันธ์กันน้อย และถ้าค่า ρ มีค่าเท่ากับ 0 แสดงว่าตัวแปรที่ศึกษาทั้งสองไม่มีความสัมพันธ์กันเลย ส่วนเครื่องหมายบวกและลบ (+,-) จะบอกถึงทิศทางความสัมพันธ์ว่าตัวแปรมีความสัมพันธ์ในทิศทางเดียวกันหรือตรงกันข้าม ตามลำดับ ค่าสัมประสิทธิ์สหสัมพันธ์ได้รับผลกระทบจากข้อมูลที่มีค่าผิดปกติจากกลุ่มเช่นกัน โดยขนาดและทิศทางของความสัมพันธ์จะถูกปรับทิศทางไปตามข้อมูลที่ผิดปกติ Devlin (1975 อ้างถึงใน (Staudte and Sheather, 1990) ได้ศึกษาการวัดความสัมพันธ์ และการประมาณค่าโดยศึกษาจากตัวแปรสุ่มสองตัวแปรที่เป็นอิสระต่อกัน ในลักษณะความสัมพันธ์เชิงเส้นตรง พบว่าค่าที่ผิดปกติจากกลุ่มมีอิทธิพลอย่างมากต่อ ค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน โดยค่าที่ผิดปกติจากตัวแปรเพียง 2-3 คู่ จะทำให้เกิดความลำเอียง (Bias) อย่างมากต่อค่าสถิติที่ได้ Huber (1981) ได้สรุปว่าผลกระทบจากข้อมูลที่มีค่าห่างผิดปกติจากกลุ่มกับความสัมพันธ์ของตัวแปรที่เป็นอิสระต่อกัน มาจากประชากรชุดเดียวกัน มีลักษณะการแจกแจงที่เหมือนกัน (Independent and Identically Distributed ; i.i.d.) จะส่งผลให้เกิดความคลาดเคลื่อนต่อการทำนายหรือพยากรณ์รูปแบบความสัมพันธ์ ดังนั้นในการวิเคราะห์ข้อมูลเมื่อตรวจสอบพบค่าผิดปกติจากกลุ่มจะต้องมีความตระหนักต่อการเลือกตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์ที่เหมาะสมต่อการวิเคราะห์ข้อมูล การที่จะเลือกตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์ที่เหมาะสมมาใช้ในการวิเคราะห์จะพิจารณาจากข้อตกลงเบื้องต้นของสถิติทดสอบนั้นๆ วิธีประมาณค่าสัมประสิทธิ์สหสัมพันธ์ที่นิยมใช้มาก คือ วิธีประมาณค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน (Pearson Product-Moment Correlation) ซึ่งมีข้อตกลงเบื้องต้นดังนี้ ตัวแปรทั้งสองต้องเป็นตัวแปรแบบต่อเนื่อง อยู่ในมาตรวัดอัตราภาค (Interval Scale) ขึ้นไป และประชากรต้องมีการแจกแจงแบบปกติสองตัวแปร ดังนั้นถ้าตัวแปรอยู่ในมาตรวัดที่ต่ำกว่านี้จึงไม่ควรใช้วิธีประมาณค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน (จารุณี ไชยมูล, 2542) นอกจากนี้ Wilcox (1997) ได้เสนอวิธีการคำนวณค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างสองตัวแปร ที่เป็นข้อมูลแบบต่อเนื่องและมีความสัมพันธ์เชิงเส้นตรง คือ สหสัมพันธ์แบบถ่วงน้ำหนัก (Biweight Midcorrelation) โดยวิธีการถ่วงน้ำหนักจะเป็นการลดอิทธิพลของค่าผิดปกติจากกลุ่ม สหสัมพันธ์แบบถ่วงน้ำหนักจึงเป็นตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์อีกทางเลือกหนึ่งที่มีข้อตกลงเบื้องต้นในการใช้เหมือนสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน เมื่อตัวแปรอยู่ในมาตรวัดแบบอันดับ (Ordinal Scale) วิธีการประมาณค่าสัมประสิทธิ์สหสัมพันธ์ที่นิยมใช้มากคือ สหสัมพันธ์แบบสเปียร์แมน

(Spearman's Rank Correlation) (ทองดี แย้มสรวล, 2530) แต่ถ้าเป็นวิธีการหาความสัมพันธ์ร่วมของการจัดอันดับข้อมูล 2 เซต โดยที่ข้อมูลหรือตัวแปรทั้งสองอยู่ในมาตรวัดขั้นต่ำแบบเรียงอันดับ วิธีการหาสัมประสิทธิ์สหสัมพันธ์แบบเคนดอลล์จะมีความเหมาะสมเนื่องจาก ตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์แบบเคนดอลล์จะให้ค่าประมาณที่ไม่เอนเอียง (Un-bias Estimator) กับค่าพารามิเตอร์ของประชากร ในขณะที่ตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมนจะไม่นำไปใช้ในการประมาณค่าพารามิเตอร์ของประชากร (Siegel and Castellan,1988) และการหาค่าสัมประสิทธิ์สหสัมพันธ์แบบเคนดอลล์มีข้อตกลงเบื้องต้นในการใช้เหมือนกับวิธีการหาสัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมน นอกจากนี้ตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์แบบแบบสเปียร์แมนและสหสัมพันธ์แบบเคนดอล์ ยังสามารถใช้ได้กับข้อมูลที่เป็นแบบต่อเนื่องได้อีกด้วย(นิคม ถนอมเสียง, 2540) แต่งานวิจัยส่วนมากเมื่อตัวแปรอยู่ในมาตรวัดแบบอันตรภาคขึ้นไป มักจะใช้การประมาณค่าสัมประสิทธิ์สหสัมพันธ์โดยวิธีของเพียร์สัน โดยมองข้ามอิทธิพลของค่าผิดปกติจากกลุ่ม และการแจกแจงเป็นแบบปกติหรือไม่ ส่งผลให้การสรุปผลการวิเคราะห์มีความคลาดเคลื่อน

ดังนั้น การประมาณค่าสัมประสิทธิ์สหสัมพันธ์เมื่อพบข้อมูลจากตัวแปรใดตัวแปรหนึ่งหรือทั้งสองตัวแปรมีค่าผิดปกติจากกลุ่มอยู่ด้วย นับว่าเป็นปัญหาต่อการเลือกใช้สถิติที่จะนำมาวิเคราะห์สัมประสิทธิ์สหสัมพันธ์ เพราะนอกจากจะพิจารณาถึงเงื่อนไขและข้อตกลงเบื้องต้นของตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์แต่ละแบบแล้ว ยังต้องพิจารณาถึงอิทธิพลของค่าผิดปกติจากกลุ่มด้วย ดังนั้นการวิเคราะห์ข้อมูลในลักษณะนี้จึงต้องเลือกใช้ตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์ที่ไม่ไวต่ออิทธิพลของค่าผิดปกติจากกลุ่มและการละเมิดข้อตกลง ซึ่งเรียกว่าตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์ที่มีความแกร่งในการทดสอบ (Robustness of the Test) โดยจะให้ผลลัพธ์ในการวิเคราะห์ที่มีประสิทธิภาพ (Efficiency)และไม่เปลี่ยนแปลงง่าย (Consistency) (Freund, 1992 ; Huber,1981) จึงเป็นที่น่าศึกษาว่าวิธีการประมาณค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรสองตัวแปรที่เป็นข้อมูลแบบต่อเนื่อง และมีค่าผิดปกติจากกลุ่ม ได้แก่ ตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน สหสัมพันธ์แบบสเปียร์แมน สหสัมพันธ์แบบเคนดอลล์ และ สหสัมพันธ์แบบถ่วงน้ำหนัก ตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์วิธีใดจะเป็นตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์ที่แกร่ง (Robust Correlation) ภายใต้สถานการณ์ที่ข้อมูลมีค่าผิดปกติจากกลุ่มในจำนวนที่แตกต่างกัน

2. คำถามการวิจัย

ค่าผิดปกติจากกลุ่ม จะมีผลต่อความแกร่งของสัมประสิทธิ์สหสัมพันธ์อย่างไร

3. วัตถุประสงค์

เพื่อศึกษาความแกร่งของสัมประสิทธิ์สหสัมพันธ์เมื่อพบข้อมูลมีค่าผิดปกติจากกลุ่ม

4. ขอบเขตของการวิจัย

4.1 พิจารณาเฉพาะรูปแบบความสัมพันธ์เชิงเส้นตรง

4.2 พิจารณาเฉพาะสหสัมพันธ์อย่างง่าย

4.3 ตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์ ที่ใช้ในการทดสอบความแกร่งของสัมประสิทธิ์สหสัมพันธ์เมื่อมีค่าผิดปกติจากกลุ่มจะใช้ตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์ จำนวน 4 แบบ คือ ตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน สหสัมพันธ์แบบสเปียร์แมน สหสัมพันธ์แบบเคนดอลล์ และสหสัมพันธ์แบบถ่วงน้ำหนัก

5. คำจำกัดความ

5.1 ค่าผิดปกติจากกลุ่ม (Outlier) หมายถึงค่าที่สูงกว่าค่าของข้อมูลในชุดนั้น โดยข้อมูลที่มีค่ามากกว่า Q3 + 1.5(IQR) เป็นค่าที่สูงผิดปกติจากกลุ่ม

5.2 ความแกร่งของสัมประสิทธิ์สหสัมพันธ์ (Robust Correlation Coefficient) หมายถึงคุณสมบัติของตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์ที่ไม่ไวต่อการเปลี่ยนแปลงของปัจจัยอื่น ๆที่ไม่ใช่ปัจจัยของการทดสอบ เช่น ผลของค่าผิดปกติจากกลุ่ม การฝ่าฝืนข้อตกลงเบื้องต้นของการทดสอบนั้น โดยตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์ที่แกร่งจะต้องมีคุณสมบัติ 2 ประการ คือ

5.2.1 ความแกร่งในการทดสอบทางสถิติ (Robust of Statistical Test) ตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์ที่มีความแกร่งในการทดสอบทางสถิติจะพิจาณาจากประเด็นต่อไปนี้

5.2.1.1 ความสามารถในการควบคุมความคลาดเคลื่อนประเภทที่ 1(Type I Error) โดยพิจารณาจากความถี่สัมพัทธ์ของการปฏิเสธสมมติฐานหลัก เมื่อสมมติฐานหลักเป็นจริง (ρ=0) แล้วนำค่าความถี่สัมพัทธ์ไปเปรียบเทียบกับช่วงของระดับนัยสำคัญที่กำหนด ถ้าตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์แบบใดมีความถี่สัมพัทธ์ของการปฏิเสธสมมติฐานหลัก เมื่อสมมติฐานหลักเป็นจริง อยู่ภายในช่วง 0.036 ถึง 0.064 แสดงว่าตัวประมาณค่าสัมประสิทธิ์นั้นมีความแกร่งหรือควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ (Bradley, 1978)

5.2.1.2 อำนาจการทดสอบ (Power of Test) จะพิจารณาจากค่าความถี่สัมพัทธ์ของผลการทดสอบที่สรุปว่าปฏิเสธสมมติฐานหลักเมื่อสมมติฐานหลักไม่เป็นจริง ($\rho \neq 0$) ของตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์แต่ละแบบ และตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์นั้นจะต้องมีความสามารถในการควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ตามเกณฑ์ที่กำหนด ถ้าตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์ใดมีค่าความถี่สัมพัทธ์ของผลการทดสอบสูงกว่าตัวประมาณค่าแบบอื่น ๆ แสดงว่าตัวประมาณค่านั้นมีความแกร่งกว่าตัวประมาณค่าอื่น ๆที่เหลือ และถ้าตัวประมาณค่า 2 แบบมีค่าความถี่สัมพัทธ์เท่ากัน แสดงว่าตัวประมาณค่าทั้งสองมีความแกร่งไม่แตกต่างกัน เมื่อจำนวนครั้งที่ทำการทดลองเพิ่มขึ้นมาก ๆ ค่าความถี่สัมพัทธ์ของการปฏิเสธสมมติฐานหลัก .เมื่อสมมติฐานหลักไม่เป็นจริง จะมีค่าใกล้เคียงหรือเทียบเท่ากับค่าอำนาจของการทดสอบ

5.2.1.3 ค่าประสิทธิภาพสัมพัทธ์ (Relative Efficiency; R.E.) จะพิจารณาจากอัตราการเปลี่ยนแปลงของค่าความถี่สัมพัทธ์ของการปฏิเสธสมมติฐานหลัก เมื่อสมมติฐานหลักไม่เป็นจริง โดยมีค่าผิดปกติจากกลุ่มในจำนวน 5%, 10%, 20% และ 30% ของขนาดตัวอย่าง เปรียบเทียบกับเมื่อไม่มีค่าผิดปกติจากกลุ่ม เพื่อพิจารณาว่าการมีค่าผิดปกติจากกลุ่มในตัวอย่างในจำนวนที่แตกต่างกันจะส่งผลกระทบต่อตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์ในแต่ละแบบอย่างไร ถ้า R.E. = 1 หมายถึง การมีค่าผิดปกติจากกลุ่มในตัวอย่างไม่ส่งผลกระทบต่อตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์นั้น แสดงว่า ตัวประมาณค่านั้นมีความแกร่ง เมื่อค่า R.E. $\neq$ 1 ถ้าตัวประมาณค่าแบบใดที่มีค่า R.E. ผันแปรจาก 1 น้อยที่สุดในแต่ละสถานการณ์ จะเป็นตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์ที่ได้รับผลกระทบจากค่าผิดปกติจากกลุ่มน้อยที่สุด ซึ่งเป็นตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์มีความแกร่งที่สุดเมื่อเทียบกับตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์แบบอื่น

5.2.2 ความแกร่งในการประมาณค่า (Robust of Estimation) เกณฑ์การตัดสินว่าตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์แบบใดมีความแกร่งในการประมาณค่า จะพิจารณาจากค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Mean Square Error; MSE) ของตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์ทั้ง 4 แบบ วิธีการประมาณค่าสัมประสิทธิ์สหสัมพันธ์ที่มีความแกร่งในการประมาณค่า ค่า MSE จะเท่ากับ 0 แต่ถ้าค่า MSE ไม่เท่ากับ 0 ตัว

ประมาณค่าสัมประสิทธิ์สหสัมพันธ์แบบใดให้ค่า MSE ต่ำที่สุด จะถือว่าเป็นตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์ที่มีความแกร่งในการประมาณค่าเมื่อเทียบกับตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์แบบอื่น ๆ

6. ประโยชน์ที่คาดว่าจะได้รับ

6.1 นำผลการศึกษาที่ได้มาประยุกต์ใช้ในการประมาณค่าสัมประสิทธิ์สหสัมพันธ์ได้อย่างถูกต้องและเหมาะสม เมื่อมีหน่วยตัวอย่างบางหน่วยมีค่าผิดปกติจากกลุ่ม

6.2 ผลที่ได้จากการศึกษาเป็นพื้นฐานในการศึกษาด้านความแกร่งทางสถิติสำหรับผู้สนใจต่อไป