การเปรียบเทียบวิธีการสุ่มตัวอย่างเพื่อตรวจสอบการนำเข้าข้อมูล

COMPARISON OF SAMPLING METHODS IN VERIFICATION OF DATA ENTRY

จุฬาทิพย์ บุญมา / มาลินี เหล่าไพบูลย์ / ยุพา ถาวรพิทักษ์

มหาวิทยาลัยขอนแก่น

บทคัดย่อ

การนำเข้าข้อมูลเป็นขั้นตอนหนึ่งที่มีผลต่อคุณภาพของข้อมูล ซึ่งแต่ละวิธีจะมีลักษณะเฉพาะของการทำงานความซับซ้อนของกระบวนการทำงานและข้อจำกัดในการใช้งาน ซึ่งงานวิจัยนี้ศึกษาการสุ่มตัวอย่างเพื่อการตรวจสอบการนำเข้าข้อมูลด้วยวิธี Continuous Sampling Plan 1 (CSP-1), CSP-2, Systematic Continuous Sampling Plan 1 (SCSP-1), และ SCSP-2 โดยสร้างประชากรจำนวน 8 ชุด (สัดส่วนความผิดพลาดของการนำเข้าข้อมูล (p) = 0.0034, 0.0080, 0.0130, 0.0188, 0.0254, 0.0528, 0.0670, และ 0.0822) ที่มีรูปแบบของความไม่ถูกต้องเป็นแบบสุ่ม (Random error) ด้วยโปรแกรม Microsoft Office Excel 2007

ผลการศึกษา พบว่า เมื่อ p มีแนวโน้มเพิ่มขึ้น ค่าเปอร์เซ็นต์เฉลี่ยคุณภาพที่เพิ่มขึ้น (Percent Gain in Average Quality ; PGAQ) มีแนวโน้มเพิ่มขึ้นตามค่า p และที่ p=0.0034 และ i=5 มีค่า PGAQ และค่าเฉลี่ยของสัดส่วน record ที่ถูกตรวจสอบ (Average Fraction Inspection ; AFI) สูงกว่า i=10, 15, 20 และที่ p=0.0822 และ i=20 มีค่า PGAQ และค่า AFI สูงกว่า i=5, 10, 15 วิธีการสุ่ม CSP-2 พบว่า ทุกค่า p จะมีค่า PGAQ และค่า AFI ที่มีแนวโน้มคงที่ และที่ i=5 ให้ค่า PGAQ และค่า AFI สูงกว่า i=10, 15, และ 20 วิธีการสุ่ม SCSP-1 พบว่า ทุกค่า p จะมีค่า PGAQ และค่า AFI มีแนวโน้มคงที่ โดยที่ i=5 ให้ค่า PGAQ และค่า AFI สูงกว่า i=10, 15, และ 20 และวิธีการสุ่ม SCSP-2 พบว่า ทุกค่า p จะมีค่า PGAQ และค่า AFI มีแนวโน้มคงที่ โดยที่ i=5 ให้ค่า PGAQ และค่า AFI สูงกว่า i=10, 15, และ 20 โดยวิธีการที่ดีที่สุดค่า วิธี SCSP-1 และเป็นวิธีการที่ใช้งานง่ายและให้ประสิทธิภาพของข้อมูลหลังการสุ่มตรวจสอบที่ดีที่สุด ซึ่งจากผลการศึกษาจะเห็นได้ว่าค่า PGAQ และค่า AFI มีความสอดคล้องกันในลักษณะที่เมื่อจำนวนชุดที่ตรวจสอบมากโอกาสที่พบข้อมูลที่ไม่ถูกต้องมากขึ้น ทำให้ค่า PGAQ มีค่าสูงขึ้นจึงทำให้ประสิทธิภาพของวิธีการดีขึ้นตามไปด้วย

คำสำคัญ: การสุ่มตรวจสอบการนำเข้าข้อมูล, การสุ่มตรวจแบบต่อเนื่อง

ABSTRACT

Data entry is considered vital for the quality of the data. Each of the data entry's method is unique in the way of working, complication of the process and limitation. This research studied the sampling methods in verification of data entry by using Continuous Sampling Plan 1 (CSP-1), CSP-2, Systematic Continuous Sampling Plan 1 (SCSP-1), and SCSP-2 methods. There were 8 samples (p = 0.0034, 0.0080, 0.0130, 0.0188, 0.0254, 0.0528, 0.0670, and 0.0822) which were random error by Microsoft Office Excel 2007

The results of the study of CSP-1 sampling found that when the value of p tended to increase, the Percent Gain in Average Quality ; PGAQ also tended to rise according to the value of p. The value of PGAQ and the average value of record proportion of Average Fraction Inspection; AFI were higher when p=0.03345 and i=5. When i=10, 15, 20 and p=0.0822 and i=20, the value of PGAQ and AFI was higher than i=5, 10, 15. The results of CSP-2 sampling showed that every value of p, the value of PGAQ and AFI tended to be stable. When i=5, the value of PGAQ and AFI was higher than i=10, 15 and 20. The results of SCSP-1 found that, every value of PGAQ and AFI tented to be stable. I=5 caused the value of PGAQ and AFI to be higher than i=10, 15, and 20. In SCSP-2 sampling, it was found that in every value of p, the value of PGAQ and AFI tended to be stable. I-5 caused PGAQ and AFI to be higher than i=10, 15, and 20. The best method was SCSP-1 and it had the least conditions in using. Therefore, it was used easily and gave the most efficient data after random error. According to the study, it was found that the value of PGAQ and AFI were in accordant when there were a number of the investigated data sets. The opportunity to find errors was higher; consequently, the value of PGAQ was hither and gave better efficiency of the method.

Keywords: Sampling methods in verification of data, Continuous sampling plan

ความเป็นมาและความสำคัญของปัญหา

การนำเข้าข้อมูลเป็นขั้นตอนหนึ่งที่มีผลต่อคุณภาพของข้อมูลซึ่งสามารถแบ่งวิธีการนำเข้าข้อมูลได้ 3 วิธี คือ Single Data Entry (SDE), Double Data Entry (DDE) และ Data capture ซึ่งแต่ละวิธีจะมีลักษณะเฉพาะของการทำงานความซับซ้อนของกระบวนการทำงานและข้อจำกัดในการใช้งาน (บัณฑิต ถิ่นคำรพ, 2548) โดยวิธีการนำเข้าข้อมูลแบบ Data capture เป็นวิธีการนี้มีอุปกรณ์มากและมีค่าใช้จ่ายที่สูงมาก วิธีการนำเข้าข้อมูลแบบ DDE เป็นวิธีการนี้เป็นวิธีการที่นิยมใช้กันมาก (Dennis W. King และ Lashley. R, 2000) และไม่ต้องใช้อุปกรณ์ซับซ้อนในการนำเข้าข้อมูลใช้เวลาในการนำเข้าข้อมูลมากกว่าเมื่อเทียบกับวิธีการนำเข้าแบบ Data capture และมีค่าใช้จ่ายมาก และวิธีการนำเข้าข้อมูลแบบ SDE เป็นการนำวิธี Continuous Sampling Plan (CSP) ซึ่งเป็นวิธีการสุ่มตัวอย่างเพื่อตรวจสอบคุณภาพของผลิตภัณฑ์มาใช้ในการสุ่มข้อมูลเพื่อการตรวจสอบข้อมูล วิธีนี้จะทำให้ลดค่าใช้จ่ายในการนำเข้าข้อมูลลดลง เนื่องจากไม่ต้องตรวจสอบข้อมูลทุกทั้งหมด โดยวิธีนี้ยังมีข้อดีคือ (1) เป็นแผนการสุ่มตัวอย่างที่ง่ายพนักงานนำเข้าข้อมูลสามารถทำได้ด้วยตนเอง (2) เลือกแผนการสุ่มที่เหมาะสมเพื่อให้ข้อมูลมีคุณภาพในระดับที่ต้องการได้ (3) สามารถคำนวณค่าใช้จ่าย ระยะเวลาที่ใช้ในการตรวจสอบข้อมูลได้ (Dennis W. King และ Lashley. R, 2000, และ ยุทธ ไกยวรรณ์, 2546) ซึ่งในบทความนี้จะกล่าวถึงเฉพาะวิธีการนำเข้าแบบ SDE

CSP เป็นวิธีการสุ่มเพื่อตรวจสอบคุณภาพของสินค้าที่ผลิตในโรงงานอุตสาหกรรม เช่น น้ำอัดลม เบียร์ แก้วน้ำ ฯลฯ หลักการของวิธีการนี้คือ การตรวจสอบแบบ 100% (เป็นการตรวจสอบสินค้าต่อเนื่องกัน จำนวน i ชิ้น) สลับกับการตรวจสอบแบบอัตราส่วน f (โดยที่ f=1/r เป็นการสุ่มตรวจสอบสินค้า 1 ชิ้นจาก r ชิ้น) โดย CSP ที่เริ่มใช้ คือ CSP-1 ได้พัฒนาขึ้นมาโดย H. F. Dodge ในปี ค.ศ. 1943 ต่อมา Dodge and Miss M. N. Torrey ได้พัฒนาเป็น CSP-2 และ CSP-3 ตามลำดับ โดย CSP-1 มีค่าพารามิเตอร์ดังนี้ Average Outgoing Quality (AOQ) หมายถึง คุณภาพผลิตภัณฑ์ออกโดยเฉลี่ย คือค่าสัดส่วนผลิตภัณฑ์ไม่ถูกต้องโดยเฉลี่ยหลังจากการตรวจสอบ, Average Fraction Inspection (AFI) หมายถึง สัดส่วนผลิตภัณฑ์ที่ตรวจสอบโดยเฉลี่ย คือสัดส่วนของผลิตภัณฑ์ที่ถูกตรวจสอบซึ่งประกอบด้วยการตรวจสอบแบบ 100% และการตรวจสอบด้วยอัตราส่วน, Average Records Inspected (ARI) หมายถึง จำนวนผลิตภัณฑ์ที่ถูกตรวจสอบโดยเฉลี่ย คือจำนวนเฉลี่ยของผลิตภัณฑ์ทั้งหมดที่ถูกตรวจสอบ, Percent Gain in Average Quality (PGAQ) หมายถึง เปอร์เซ็นต์คุณภาพที่เพิ่มขึ้นโดยเฉลี่ย คือคุณภาพผลิตภัณฑ์ออกภายหลังการตรวจสอบที่มีคุณภาพที่ดีขึ้น ได้มีผู้ศึกษาการใช้ CSP-1 ในการตรวจสอบข้อมูลที่มีการ Simulate โดยสรุปได้ดังนี้

Dennis W. King และ Roderick Lashley (2000)ได้ทำการศึกษาประสิทธิภาพของวิธีการตรวจสอบข้อมูลที่นำเข้าข้อมูลแบบ SDE ด้วยการสุ่มตัวอย่างแบบ CSP-1 และทำการตรวจสอบข้อมูลด้วยวิธี Visual record verification check (VRVC) และ Roderick Lashley (2001) ได้ทำการศึกษาเพื่อเปรียบเทียบประสิทธิภาพของแผนการสุ่ม 2 แผนคือ CSP-1 i= 10 f=1/10 และ CSP-1 i= 5 f=1/5 กับชุดข้อมูลจำลองที่มี error 3 แบบ คือแบบที่1 Random Error Model แบบที่ 2 Non-random error แบบ Burst model แบบที่ 3 Non-random error แบบ Exponential model โดยแต่ละ error model สร้างชุดข้อมูลจำนวน 10000 ชุดข้อมูล (แต่ละชุดข้อมูลมีข้อมูลจำนวน 5000 ข้อมูล) และ Dennis W.King และ Melynda Hazelwood (2000) ทำการศึกษาวิธีการสุ่มตัวอย่างแบบบันทึกข้อมูล (Case Record Form ;CRF) เพื่อนำมาเปรียบเทียบกับ Source Document 3 วิธี คือ วิธีการสุ่มอย่างอย่างง่าย(SRS : Simple Random Sampling) วิธีการสุ่มตัวอย่าง SRSR (Simple Random Sampling with Rectified) และวิธีการสุ่ม CSP-1 ที่ i=40 f=1/10

จากการศึกษาการสุ่มตัวอย่างแบบ CSP-1 พบว่าจะเพิ่มคุณภาพข้อมูลหลังการตรวจสอบได้มากกว่าเมื่อเทียบกับการสุ่มตัวอย่างทั่วไป (เช่น SRS) และคุณภาพจะเพิ่มมากขึ้นเมื่อข้อมูลก่อนการตรวจสอบมีคุณภาพต่ำและ error เป็นแบบ non-random ดังนั้นผู้วิจัยจึงสนใจศึกษาการสุ่มตัวอย่างแบบ CSP-1 กับข้อมูลที่มีการนำเข้าจริง และสนใจที่จะทำการศึกษาการสุ่มตัวอย่างแบบ CSP-2 ด้วยเพราะวิธี CSP-2 เป็นวิธีการสุ่มที่มีข้อยืดหยุ่นมากกว่าวิธี CSP-1 และยังไม่มีการศึกษานำ CSP-2 มาใช้ในการตรวจสอบข้อมูล นอกจากนั้น ผู้วิจัยได้กำหนดวิธีการใหม่ขึ้นมาโดยพัฒนาจากวิธีการจาก CSP-1 และ CSP-2 ให้มีวิธีการสุ่มที่ง่ายขึ้น โดยสลับการตรวจแบบ 100% และ การตรวจแบบอัตราส่วน (f) แบบคงที่ วิธีดังกล่าว คือ วิธี Systematic Continuous Sampling Plan 1 (SCSP-1) ซึ่งมีการสลับการตรวจสอบแบบ 100% และการตรวจสอบแบบ f และวิธี Systematic Continuous Sampling Plan 2 (SCSP-2) มีการสลับการตรวจสอบแบบ 100% และการตรวจสอบแบบ f จำนวน 2 ครั้ง โดยการศึกษาครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบค่า PGAQ และ AFI ของวิธีการสุ่มตัวอย่างเพื่อการตรวตสอบการนำเข้าข้อมูล CSP-1, CSP-2, SCSP-1, และ SCSP-2 ผลการวิจัยครั้งนี้จะเป็นแนวทางให้นักชีวสถิติและนักวิจัยทางการแพทย์และทางสาธารณสุขเลือกใช้วิธีการสุ่มในการตรวจสอบข้อมูลเพื่อให้เหมาะสมกับคุณภาพการนำเข้าข้อมูล และทรัพยากรในการดำเนินการ

ขั้นตอนการดำเนินการวิจัย

2.1 สร้างประชากรจำนวน 8 ชุด (p = 0.0034, 0.0080, 0.0130, 0.0188, 0.0254, 0.0528, 0.0670, และ 0.0822) โดยแต่ละชุดสร้าง 10000 ประชากรย่อย โดย 1 ประชากรย่อยประกอบด้วย

ข้อมูล 5000 records ที่มีรูปแบบของความไม่ถูกต้องเป็นแบบสุ่ม (Random error) โดยใช้โปรแกรม Microsoft Office Excel 2007 ในการสร้างประชากร

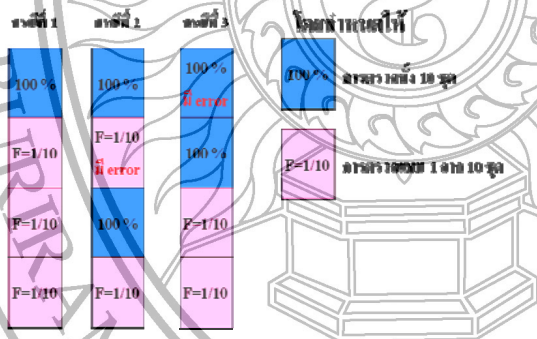
2.2 เขียนชุดคำสั่ง (Program) เพื่อตรวจสอบโดยใช้โปรแกรม Microsoft Office Excel 2007

2.2.1 CSP-1 (i, f=1/i)

(1) เริ่มต้นด้วยการตรวจสอบแบบ 100% จำนวน i ชุด ต่อเนื่องกัน หาผลรวมของ x กำหนดให้ x คือ ผลของความถูกต้องของข้อมูล มีค่าเท่ากับ 0 เมื่อ ข้อมูลถูกต้อง และมีค่าเท่ากับ 1 เมื่อข้อมูลไม่ถูกต้อง ถ้าผลรวมของ x มากกว่า 0 จะทำการตรวจสอบแบบ 100% จำนวน i ชุดถัดมา

(2) ถ้าผลรวมของ $x=0$ จะเปลี่ยนการตรวจสอบเป็นการตรวจสอบแบบอัตราส่วน 1/i โดยทำการสุ่มตัวอย่าง 1 ชุด จาก i ชุดถัดมา โดยใช้วิธีการสุ่มอย่างง่าย

ถ้าชุดที่สุ่มได้ มีค่า $x=0$ จะทำการสุ่มตรวจสอบแบบอัตราส่วน 1/i ไปเรื่อยๆ จนกระทั่ง พบ $x=1$ ให้กลับไปตรวจสอบแบบ 100% ตามวิธีข้อ (1) ทำเช่นนี้จนกระทั่งตรวจสอบครบทุกชุด ซึ่งแสดงตัวอย่างการสุ่มดังภาพ 1



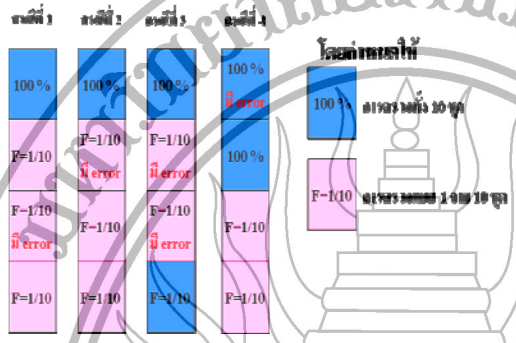
ภาพ 1 วิธีการสุ่มตรวจสอบแบบ CSP-1

2.2.2 CSP-2 (i, f=1/i)

(1) เริ่มต้นด้วยการตรวจสอบแบบ 100% จำนวน i ชุด ต่อเนื่องกัน ด้วยการหาผลรวมของ ถ้าผลรวมของ x มากกว่า 0 จะทำการตรวจสอบแบบ 100% จำนวน i ชุดถัดมาไปเรื่อยๆ จนกระทั่งพบผลรวมของ x เท่ากับ 0 จึงทำการสุ่มตรวจสอบแบบอัตราส่วน

(2) ถ้าผลรวมของ $x=0$ จะเปลี่ยนการตรวจสอบเป็นการตรวจสอบแบบอัตราส่วน 1/i โดยทำการสุ่มตัวอย่าง 1 ชุด จาก i ชุดถัดมา โดยใช้วิธีการสุ่มอย่างง่าย

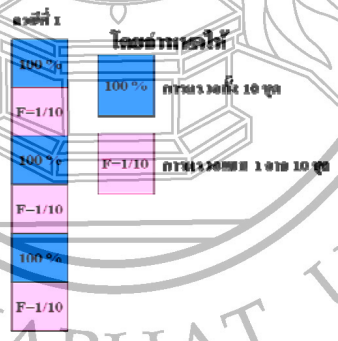
ถ้าชุดที่สุ่มได้ มีค่า x =0 จะทำการสุ่มตรวจสอบด้วยอัตราส่วน 1/i ไปเรื่อยๆ จนกระทั่ง พบ x =1 ติดต่อกัน 2 ครั้ง ให้กลับไปตรวจสอบแบบ 100% **ตามวิธีข้อ (1)** ทำเช่นนี้จนกระทั่งตรวจสอบครบทุกชุด ซึ่งแสดงตัวอย่างการสุ่มดังภาพ 2



ภาพ 2 วิธีการสุ่มตรวจสอบแบบ CSP-2

2.2.3 SCSP-1 (i, f=1/i)

เริ่มต้นด้วยการตรวจสอบแบบ 100% จำนวน i ชุด ต่อเนื่องกัน หาผลรวมของ x และใน i ชุดถัดมาทำการตรวจสอบแบบอัตราส่วน 1/i โดยทำการสุ่มตัวอย่าง 1 ชุด จาก i ชุด ด้วยใช้วิธีการสุ่มอย่างง่าย ทำเช่นนี้สลับกันไปเรื่อยๆ ซึ่งแสดงตัวอย่างการสุ่มดังภาพ 3

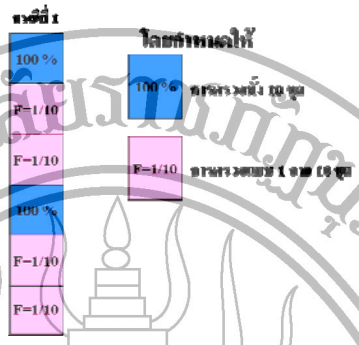


ภาพ 3 วิธีการสุ่มตรวจสอบแบบ SCSP-1

2.2.4 SCSP-2 (i, f=1/i)

เริ่มต้นด้วยการตรวจสอบแบบ 100% จำนวน i ชุด ต่อเนื่องกัน หาผลรวมของ x และใน i ชุดถัดมาทำการตรวจสอบแบบอัตราส่วน 1/i โดยทำการสุ่มตัวอย่าง 1 ชุด จาก i ชุด

ด้วยวิธีการสุ่มอย่างง่าย โดยวิธีการนี้จะทำการตรวจแบบอัตราส่วน 2 ครั้งติดกัน ทำเช่นนี้สลับกันไปเรื่อยๆ ซึ่งแสดงตัวอย่างการสุ่มดังภาพ 4



ภาพ 4 วิธีการสุ่มตรวจสอบแบบ SCSP-2

2.3 ทดลองใช้คำสั่งเพื่อตรวจสอบว่าแสดงผลได้ตามต้องการ และตรวจสอบการทำงานของชุดคำสั่งด้วยการเทียบผลกับการสุ่มด้วยมือ ทำการแก้ไขชุดคำสั่งให้ถูกต้อง

2.4 run program สุ่มตรวจสอบกับ 10000 ประชากรย่อย ของประชากรแต่ละชุด

2.5 คำนวณค่า AOQ, AFI, ARI, และ PGAQ ของประชากรแต่ละชุด

2.5.1 Percent Gain in Average Quality (PGAQ) คือ เปอร์เซ็นต์เฉลี่ยคุณภาพที่เพิ่มขึ้น โดยกำหนดให้

p = สัดส่วนความผิดพลาดของการนำเข้า

AOQ= ค่าเฉลี่ยของสัดส่วน record ที่ไม่ถูกต้องหลังการสุ่มตรวจสอบ

$$PGAQ = \left[1 - \frac{AOQ}{p}\right] \times 100$$

Average Outgoing Quality (AOQ) หมายถึง ค่าเฉลี่ยของสัดส่วน record ไม่ถูกต้องหลังการสุ่มตรวจสอบ

โดยกำหนดให้

A= จำนวน record ไม่ถูกต้องก่อนการสุ่มตรวจสอบ (เช่น p=0.0034 ค่า A เท่ากับ17)

B_j= จำนวน record ไม่ถูกต้องหลังการสุ่มตรวจสอบของประชากรย่อยลำดับที่ j

j= ลำดับประชากรย่อยซึ่งมีค่าตั้งแต่1ถึง T

T= จำนวนประชากรย่อยทั้งหมดในชุดข้อมูล

N= จำนวน record ในประชากรย่อย

OQ_j= สัดส่วน record ที่ไม่ถูกต้องหลังการสุ่มตรวจสอบของประชากรย่อยลำดับที่ j

$$OQ_j = \frac{A - B_j}{N}$$

$$AOQ = \frac{\sum_{j=1}^{T} OQ_j}{T}$$

2.5.2 Average Fraction Inspection (AFI) คือ ค่าเฉลี่ยของสัดส่วน record ที่ถูกสุ่มตรวจสอบ

$$AFI = \frac{ARI}{N}$$

Average Records Inspected (ARI) คือ ค่าเฉลี่ยของจำนวน record ที่ถูกสุ่มตรวจสอบ โดยกำหนดให้

O_j = จำนวน record ทั้งหมดที่ถูกสุ่มตรวจสอบของประชากรย่อยลำดับที่ j

$$ARI = \frac{\sum_{j=1}^{T} O_j}{T}$$

ผลการศึกษาและอภิปรายผล

จากการศึกษาซึ่งจะเห็นได้ว่าค่า PGAQ และค่า AFI มีความสอดคล้องกันในลักษณะที่เมื่อจำนวน record ที่ตรวจสอบมากขึ้นโอกาสที่พบ record ที่ไม่ถูกต้องมากขึ้นตามจึงทำให้ค่า PGAQ มีค่าสูงขึ้นตามไปด้วย

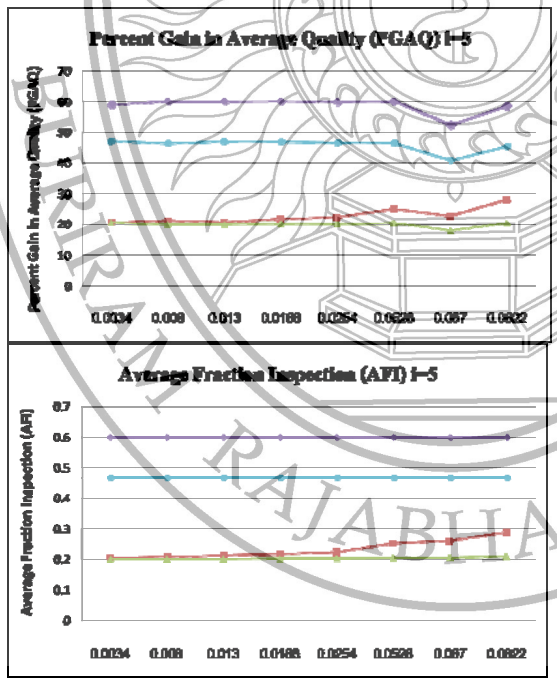
วิธีการสุ่ม CSP-1 พบว่าเมื่อ p มีแนวโน้มเพิ่มขึ้น ค่า PGAQ มีแนวโน้มเพิ่มขึ้นตามค่า p และที่ p=0.0034 และ i=5 มีค่า PGAQ สูงกว่า i=10, 15, 20 เพราะที่ i=5 เป็นการสุ่มตรวจสอบในช่วงแคบจึงทำให้จำนวนของ record ที่ถูกตรวจสอบมีจำนวนมากจึงทำให้มีโอกาสที่จะพบ record ที่นำเข้าไม่ถูกต้องมากขึ้นเพราะจำนวน record ที่นำเข้าไม่ถูกต้องมีจำนวนน้อย และที่ p=0.0822 และ i=20 มีค่า PGAQ สูงกว่า i=5, 10, 15 เพราะที่ i=20 เป็นการสุ่มตรวจสอบในกว้างจึง

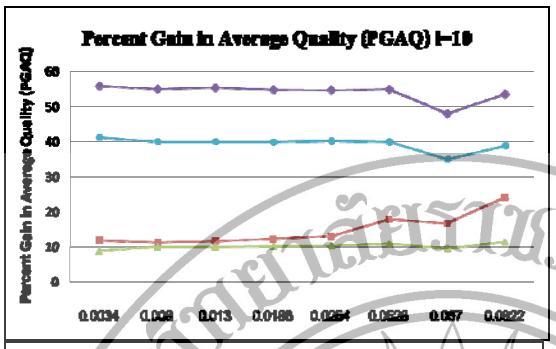
ทำให้ในการตรวจสอบแบบ 100% กว้างโอกาสที่จะพบ record ที่นำเข้าไม่ถูกต้องก็มีมากตามไปด้วยเพราะจำนวน record ที่นำเข้าไม่ถูกต้องมีจำนวนมาก

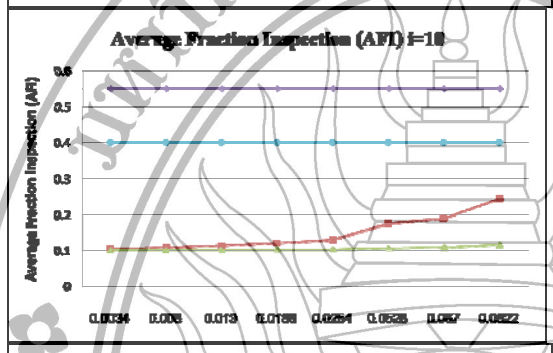
วิธีการสุ่ม CSP-2 พบว่า ทุกค่า p จะมีค่า PGAQ ที่มีแนวโน้มคงที่ และที่ i=5 ให้ค่า PGAQ สูงกว่า i=10, 15, และ 20 เพราะที่ i=5 เป็นการสุ่มตรวจสอบในช่วงแคบจึงทำให้จำนวนของ record ที่ถูกตรวจสอบมีจำนวนมากจึงทำให้มีโอกาสที่จะพบ record ที่นำเข้าไม่ถูกต้องมากขึ้น เพราะจำนวน record ที่นำเข้าไม่ถูกต้องมีจำนวนเพิ่มขึ้นตาม

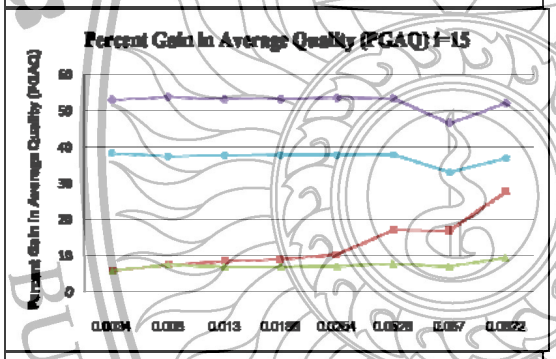
วิธีการสุ่ม SCSP-1 พบว่า ทุกค่า p จะมีค่า PGAQ มีแนวโน้มคงที่ โดยที่ i=5 ให้ค่า PGAQ สูงกว่า i=10, 15, และ 20 เพราะที่ i=5 เป็นการสุ่มตรวจสอบในช่วงแคบจึงทำให้จำนวนของ record ที่ถูกตรวจสอบมีจำนวนมากจึงทำให้มีโอกาสที่จะพบ record ที่นำเข้าไม่ถูกต้องมากขึ้น เพราะจำนวน record ที่นำเข้าไม่ถูกต้องมีจำนวนเพิ่มขึ้นตาม

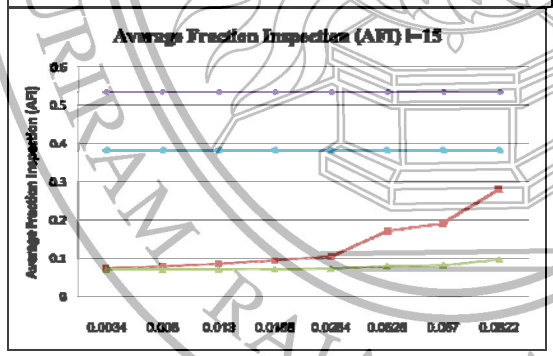
วิธีการสุ่ม SCSP-2 พบว่า ทุกค่า p จะมีค่า PGAQ มีแนวโน้มคงที่ โดยที่ i=5 ให้ค่า PGAQ สูงกว่า i=10, 15, และ 20 เพราะที่ i=5 เป็นการสุ่มตรวจสอบในช่วงแคบจึงทำให้จำนวนของ record ที่ถูกตรวจสอบมีจำนวนมากจึงทำให้มีโอกาสที่จะพบ record ที่นำเข้าไม่ถูกต้องมากขึ้น เพราะจำนวน record ที่นำเข้าไม่ถูกต้องมีจำนวนเพิ่มขึ้นตาม



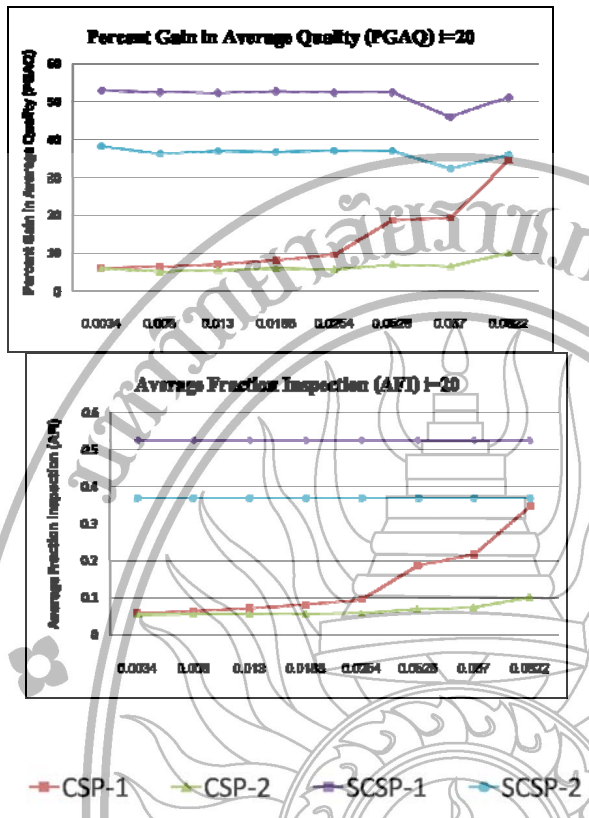








มหาวิทยาลัยราชภัฏบุรีรัมย์
BURIRAM RAJABHAT UNIVERSITY



ภาพ 5 ค่า PGAQ และ AFI ที่ระดับ i=5, 10, 15,และ 20

จากภาพที่แสดงการเปรียบเทียบค่า PGAQ ทุกวิธีการสุ่มตรวจสอบการนำเข้ามูล เมื่อ p มีแนวโน้นเพิ่มขึ้น ที่ระดับ i=5, 10, 15, และ 20 พบว่าวิธี SCSP-1 ให้ค่า PGAQ สูงกว่า วิธี SCSP-2, วิธี CSP-1, และวิธี CSP-2 ในทุกค่า i และเมื่อ i มีค่าเพิ่มขึ้นวิธี CSP-1 มีค่า PGAQ ที่แนวโน้มเพิ่มขึ้น และเมื่อ i=20 วิธี CSP-1 และวิธี SCSP-2 มีค่า PGAQ เท่ากัน

จากภาพที่แสดงการเปรียบเทียบค่า AFI ทุกวิธีการสุ่มตรวจสอบการนำเข้ามูล เมื่อ p มีแนวโน้นเพิ่มขึ้น ที่ระดับ i=5, 10, 15, และ 20 พบว่า วิธี SCSP-1 ให้ค่า AFI สูงกว่า วิธี SCSP-2, วิธี CSP-1, และวิธี CSP-2 ในทุกค่า i และเมื่อ i มีค่าเพิ่มขึ้นวิธี CSP-1 มีค่า AFI ที่แนวโน้มเพิ่มขึ้น และเมื่อ i=20 วิธี CSP-1 และวิธี SCSP-2 มีค่า AFI เท่ากัน

ซึ่งจากการศึกษาจะเห็นได้ว่าค่า PGAQ และค่า AFI มีความสอดคล้องกันในลักษณะที่เมื่อจำนวนชุดที่ตรวจสอบมากโอกาสที่พบข้อมูลที่ไม่ถูกต้องมากขึ้นจึงทำให้ค่า PGAQ มีค่าสูงขึ้นตามไปด้วย และในการเปรียบเทียบวิธีทั้ง 4 วิธี พบว่า วิธี SCSP-1 วิธี SCSP-2 วิธี CSP-1 และวิธี CSP-2 มีค่า PGAQ และ AFI เรียงลำดับจากมากไปน้อยตามลำดับ โดยวิธี SCSP-1 เป็นวิธีการที่มีค่า

PGAQ และค่า AFI ที่สูงที่สุด เพราะมีวิธีการตรวจสอบแบบ 100% และการสุ่มตรวจสอบแบบสัดส่วนสลับกันเป็นหนึ่งต่อหนึ่งครั้ง ซึ่งจะทำให้มีจำนวนของ record ที่ใช้ในการตรวจสอบมากโอกาสที่จะพบข้อมูลที่นำเข้าไม่ถูกต้องสูงกว่าทุกวิธี และวิธี SCSP-2 มีวิธีการตรวจสอบแบบ 100% และการสุ่มตรวจสอบแบบสัดส่วนสลับกันเป็นหนึ่งต่อสองครั้ง ซึ่งจะทำให้วิธี SCSP-2 เป็นวิธีการที่มีให้ค่า PGAQ และค่า AFI สูงรองจากวิธี SCSP-1 เพราะว่าวิธี SCSP-2 มีจำนวนของ record ที่ใช้ในการตรวจสอบน้อยกว่าวิธี SCSP-1 วิธี CSP-1 และวิธี CSP-2 เป็นวิธีที่มีค่า PGAQ และค่า AFI น้อยเพราะและวิธี CSP-1 มีค่า PGAQ และค่า AFI มากกว่าวิธี CSP-2 เพราะว่าวิธี CSP-2 มีเงื่อนไขในการสุ่มตรวจสอบที่พิจารณาผลการสุ่มตรวจสอบความถูกต้องของการนำเข้าข้อมูลในแต่ละขั้นตอนของการสุ่มตรวจสอบมากกว่าวิธี CSP-1 คือ เมื่อพบผลการสุ่มตรวจสอบความถูกต้องของการนำเข้าข้อมูลที่ไม่ถูกต้องนั้น วิธี CSP-1 จะทำการตรวจสอบแบบ 100% ทันทีในช่วงการตรวจสอบถัดไป ส่วนวิธี CSP-2 จะทำการสุ่มตรวจสอบแบบสัดส่วนอีกครั้งในช่วงการตรวจสอบถัดไป ทำให้โอกาสที่จะพบการนำเข้าข้อมูลที่ไม่ถูกต้องน้อยกว่าวิธี CSP-1 และเมื่อสัดส่วนความผิดพลาดของการนำเข้ามีค่าสูงขึ้น วิธี CSP-1 จะให้ค่า PGAQ และ AFI ที่สูงกว่าวิธี CSP-2 นั้นคือเมื่อสัดส่วนความผิดพลาดของการนำเข้ามีค่าสูงขึ้นวิธี CSP-1 สามารถสุ่มตรวจสอบความถูกต้องของการนำเข้าข้อมูลได้ดีกว่าวิธี CSP-2

สรุปผลและข้อเสนอแนะ

ซึ่งจากการจำลองจะเห็นได้ว่าค่า PGAQ และค่า AFI มีความสอดคล้องกันในลักษณะที่เมื่อจำนวนชุดที่ตรวจสอบมากโอกาสที่พบข้อมูลที่ไม่ถูกต้องมากขึ้นจึงทำให้ค่า PGAQ มีค่าสูงขึ้นตามไปด้วย และในการเปรียบเทียบวิธีทั้ง 4 วิธี พบว่า วิธี SCSP-1 วิธี SCSP-2 วิธี CSP-1 และวิธี CSP-2 มีค่า PGAQ และ AFI เรียงลำดับจากมากไปน้อยตามลำดับ โดยวิธี SCSP-1 เป็นวิธีการที่มีค่า PGAQ และค่า AFI ที่สูงที่สุด เพราะมีวิธีการตรวจสอบแบบ 100% และการสุ่มตรวจสอบแบบสัดส่วนสลับกันเป็นหนึ่งต่อหนึ่งครั้ง ซึ่งจะทำให้มีจำนวนของ record ที่ใช้ในการตรวจสอบมากโอกาสที่จะพบข้อมูลที่นำเข้าไม่ถูกต้องสูงกว่าทุกวิธี และวิธี SCSP-2 มีวิธีการตรวจสอบแบบ 100% และการสุ่มตรวจสอบแบบสัดส่วนสลับกันเป็นหนึ่งต่อสองครั้ง ซึ่งจะทำให้วิธี SCSP-2 เป็นวิธีการที่มีให้ค่า PGAQ และค่า AFI สูงรองจากวิธี SCSP-1 เพราะว่าวิธี SCSP-2 มีจำนวนของ record ที่ใช้ในการตรวจสอบน้อยกว่าวิธี SCSP-1 วิธี CSP-1 และวิธี CSP-2 เป็นวิธีที่มีค่า PGAQ และค่า AFI น้อยเพราะและวิธี CSP-1 มีค่า PGAQ และค่า AFI มากกว่าวิธี CSP-2 เพราะว่าวิธี CSP-2 มีเงื่อนไขในการสุ่มตรวจสอบที่พิจารณาผลการสุ่มตรวจสอบความถูกต้องของการนำเข้าข้อมูลในแต่ละขั้นตอนของการสุ่มตรวจสอบมากกว่าวิธี CSP-1 คือ เมื่อพบผลการสุ่มตรวจสอบความถูกต้อง

ของการนำเข้าข้อมูลที่ไม่ถูกต้องนั้น วิธี CSP-1 จะทำการตรวจสอบแบบ 100% ทันทีในช่วงการตรวจสอบถัดไป ส่วนวิธี CSP-2 จะทำการสุ่มตรวจสอบแบบสัดส่วนอีกครั้งในช่วงการตรวจสอบถัดไป ทำให้โอกาสที่จะพบการนำเข้าข้อมูลที่ไม่ถูกต้องน้อยกว่าวิธี CSP-1 และเมื่อสัดส่วนความผิดพลาดของการนำเข้ามีค่าสูงขึ้น วิธี CSP-1 จะให้ค่า PGAQ และ AFI ที่สูงกว่าวิธี CSP-2 นั้นคือเมื่อสัดส่วนความผิดพลาดของการนำเข้ามีค่าสูงขึ้นวิธี CSP-1 สามารถสุ่มตรวจสอบความถูกต้องของการนำเข้าข้อมูลได้ดีกว่าวิธี CSP-2

โดยวิธี SCSP-1 มีจุดเด่น คือ มีค่า PGAQ ที่สูงที่สุด และเป็นวิธีการที่มีเงื่อนไขในการสุ่มตรวจสอบน้อยที่สุดจึงทำให้วิธีการนี้สามารถใช้งานได้จริง สะดวกและง่ายต่อการใช้งาน ส่วนจุดด้อย คือ เมื่อมีค่า AFI ที่สูงซึ่งส่งผลให้มีจำนวนของ record ที่ใช้ในการสุ่มตรวจสอบก็มีจำนวนมากตามไปด้วย จึงอาจจะส่งผลทำให้มีค่าใช้จ่ายในการตรวจสอบความถูกต้องของการนำเข้าข้อมูล เช่น เงิน จำนวนคน และเวลาที่ใช้ในการทำงานจะเพิ่มมากขึ้นตามไปด้วย

การวิจัยครั้งนี้ ผู้วิจัยกำหนดให้มีสัดส่วนของข้อมูลที่นำเข้าไม่ถูกต้องที่ 8 ระดับเท่านั้น จึงควรมีการศึกษาการตรวจสอบความถูกต้องของข้อมูลที่มีสัดส่วนของข้อมูลที่นำเข้าไม่ถูกต้องที่ระดับข้อมูลที่มีคุณภาพต่ำกว่านี้ (คือ p มากกว่า 0.0822) และในการตรวจสอบแบบอัตราส่วน ผู้วิจัยได้ทำการศึกษาเพียงรูปแบบ f=1/i จึงควรมีการศึกษาในหลายๆรูป f=1/r

เอกสารอ้างอิง

จรณิต แก้วกังวาล. (2548). การตรวจสอบคุณภาพข้อมูล. **Data Management and Biostatistics Network Journal**; 1(2): 55-60.

จรรยา ภัทรอาชาชัย. (2548). DATA VALIDATION AND VERIFICATION. **Data Management and Biostatistics Network Journal**; 1(2): 35-40.

บัณฑิต ถิ่นคำรพ.(2548). ระบบการนำเข้าข้อมูล (Data Entry System). **Data Management and Biostatistics Network Journal**; 1(2): 41-53.

ประภัสสร เอื้อลลิตชูวงศ์.(2552). **อัตราความผิดพลาดของการนำเข้าข้อมูลด้วยวิธี Double Data Entry ที่มีมาตรฐานการปฏิบัติงานกำกับ**. วิทยานิพนธ์ปริญญาสาธารณสุขศาสตรมหาบัณฑิต ภาควิชาชีวสถิติและประชากรศาสตร์ บัณฑิตวิทยาลัย: มหาวิทยาลัยขอนแก่น

ยุทธ ไกยวรรณ์.(2546). **การควบคุมคุณภาพในงานอุตสาหกรรม (Industrial quality control).** กรุงเทพฯ: ชมรมเด็ก.

Dennis W. King. Quantifying gains in data quality for sampling plans used in clinical trial monitoring. **Drug Information Journal**.2003; 37: 135-141.

King DW, Lashley R.(2000). A quantifiable alternative to double data entry. **Cont Clin Trials**. 21: 94–102.

Roderick Lashley. Applying statistical sampling plans to data entry procedures to increase data quality. **STATKING Consulting Inc**.