นิพนธ์ต้นฉบับ

การเปรียบเทียบวิธีการสุ่มตัวอย่างเพื่อตรวจสอบการนำเข้าข้อมูล

จุฬาทิพย์ บุญมา **, มาลินี เหล่าไพบูลย์ ** และยุพา ถาวรพิทักษ์ ***

** นักศึกษาหลักสูตรสาธารณสุขศาสตรมหาบัณฑิต สาขาชีวสถิติ ** ภาควิชาชีวสถิติและประชากรศาสตร์ คณะสาธารณสุขศาสตร์ มหาวิทยาลัยขอนแก่น*

บทคัดย่อ

การนำเข้าข้อมูลเป็นขั้นตอนหนึ่งที่มีผลต่อคุณภาพของข้อมูล ซึ่งแต่ละวิธีจะมีลักษณะเฉพาะของการทำงานความซับซ้อนของกระบวนการทำงานและข้อจำกัดในการใช้งาน งานวิจัยนี้ศึกษาการสุ่มตัวอย่างเพื่อการตรวจสอบ(verify) การนำเข้าข้อมูล 4 วิธี ได้แก่ (1). Continuous Sampling Plan 1 (CSP-1) โดยการตรวจสอบ100% (การตรวจสอบข้อมูลต่อเนื่อง i ชุด) หากข้อมูลทั้ง i ชุด ถูกต้องจะสลับมาทำการตรวจสอบสัดส่วน คือ สุ่มข้อมูล 1 ชุดจาก i ชุด ถ้าข้อมูลชุดดังกล่าวถูกต้องจะทำการตรวจสอบแบบสัดส่วนไปเรื่อยๆ และจะสลับมาเป็นการตรวจสอบ 100% เมื่อการตรวจสอบสัดส่วนพบข้อมูลไม่ถูกต้อง (2). CSP-2 เหมือน CSP-1 แต่จะสลับมาตรวจสอบ 100% เมื่อพบข้อมูลไม่ถูกต้องจากการตรวจสอบสัดส่วน 2 ครั้งติดต่อกัน (3). Systematic Continuous Sampling Plan 1 (SCSP-1) ตรวจสอบ 100% 1 ครั้ง สลับกับตรวจสอบสัดส่วน 1 ครั้ง (4). SCSP-2 ตรวจสอบ 100% หนึ่งครั้ง สลับกับตรวจสอบสัดส่วน 2 ครั้ง กำหนดค่า i เป็น 5, 10, 15, 20 จำลองวิธีการสุ่มกับประชากรที่มีความผิดพลาดของการนำเข้าเป็นแบบสุ่ม(Random error) และมีสัดส่วนความผิดพลาด (p) 8 ระดับ คือ 0.0034, 0.0080, 0.0130, 0.0188, 0.0254, 0.0528, 0.0670, และ 0.0822 ด้วยโปรแกรม Microsoft Office Excel 2007

ผลการศึกษา พบว่า เมื่อ p มีแนวโน้มเพิ่มขึ้น ค่าเปอร์เซ็นต์เฉลี่ยคุณภาพที่เพิ่มขึ้น (Percent Gain in Average Quality ; PGAQ) มีแนวโน้มเพิ่มขึ้น ที่ 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, การสุ่มตรวจสอบแบบต่อเนื่อง*

Original Article

Comparison of Sampling Methods in Verification of Data Entry

Chulathip Boonma *, *Malinee Laopaiboon* ** *and Yupa Thavornpitak* **
* *Master Degree in Public Health Student, Biostatistics Program, Faculty of Public Health, Khon Kaen University*
** *Department of Biostatistics and Demography, Faculty of Public Health, Khon Kaen University*

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.

Keyword: *sampling methods in verification of data, continuous sampling plan*

บทนำ

การนำเข้าข้อมูลเป็นขั้นตอนหนึ่งที่มีผลต่อคุณภาพของข้อมูล วิธีการนำเข้าข้อมูลที่นิยมใช้คือการนำเข้าข้อมูลสองครั้ง(Double Data Entry: DDE) (Dennis W. King และ Lashley. R, 2000) วิธีการดังกล่าวนำเข้าข้อมูลสองครั้งโดยพนักงานต่างคน ตรวจสอบ(Verify)ข้อมูล โดยนำข้อมูลทั้งสองครั้งมาเปรียบเทียบกัน DDE เป็นวิธีที่มีค่าใช้จ่ายสูง ทางเลือกหนึ่งที่จะลดค่าใช้จ่ายคือนำเข้าข้อมูลครั้งเดียว (Single Data Entry : SDE) และสุ่มข้อมูลบางชุดเพื่อนำมาตรวจสอบ ได้มีการนำวิธี Continuous Sampling Plan (CSP-1) ซึ่งเป็นการสุ่มเพื่อตรวจสอบคุณภาพของผลิตภัณฑ์ในโรงงานอุตสาหกรรม มาใช้สุ่มชุดข้อมูลเพื่อนำมาตรวจสอบข้อมูลเนื่องจากมีข้อดีหลายประการ คือ (1) เป็นวิธีการที่ง่ายพนักงานนำเข้าข้อมูลสามารถทำได้ด้วยตนเอง (2) เลือกแผนการสุ่มที่เหมาะสมเพื่อให้ข้อมูลมีคุณภาพในระดับที่ต้องการได้ (3) คำนวณค่าใช้จ่าย ระยะเวลาที่ใช้ในการตรวจสอบข้อมูลได้ (Dennis W. King และ Lashley. R, 2000) H. F. Dodge ในปี ค.ศ. 1943 ได้เริ่มนำวิธี CSP-1 มาใช้ตรวจสอบคุณภาพของสินค้าที่ผลิตในโรงงานอุตสาหกรรม วิธีการสุ่มเป็น การตรวจสอบ 100% (ตรวจสอบสินค้าต่อเนื่องกัน จำนวน i ชิ้น) สลับกับการตรวจสอบสัดส่วน f (f=1/r คือสุ่มตรวจสอบสินค้า 1 ชิ้นจาก r ชิ้น) เงื่อนไขของการสลับคือถ้าผลการตรวจสอบ 100% ไม่พบสินค้าที่บกพร่องให้สลับมาตรวจสอบสัดส่วน ถ้าผลการตรวจสอบสัดส่วนไม่พบสินค้าที่บกพร่อง ให้ทำการตรวจสอบสัดส่วนต่อไปเรื่อยๆ จนกระทั่งพบสินค้าที่บกพร่องจึงจะสลับมาตรวจสอบ 100% ถ้าผลการตรวจสอบ 100% พบสินค้าที่บกพร่องจะทำการตรวจสอบ 100% ต่อไปเรื่อยๆ จนกระทั่งไม่พบสินค้าที่บกพร่องจึงสลับมาตรวจสอบสัดส่วน ต่อมา Dodge and Miss M. N. Torrey ได้ปรับปรุงวิธีการสุ่มเป็น CSP-2 ที่มีความเคร่งครัดน้อยกว่า คืออนุโลมให้เปลี่ยนกลับมาตรวจสอบ 100% ได้ถ้าพบข้อมูลที่ไม่ถูกต้องติดต่อกันสองครั้งจากการตรวจสอบสัดส่วน CSP-1 ถูกนำมาใช้ในการสุ่มตรวจสอบความถูกต้องของการนำเข้าข้อมูล Dennis W. King และ Roderick Lashley (2000)ได้ทำการศึกษาประสิทธิภาพของ CSP-1 พบว่า ที่เมื่อข้อมูลมีความผิดพลาดมากขึ้น ประสิทธิภาพของ CSP-1 (พิจารณาจากร้อยละคุณภาพที่เพิ่มขึ้น) เพิ่มมากขึ้น Roderick Lashley (2001) ได้ใช้ CSP-1 กับประชากรจำลองที่มีรูปแบบความผิดพลาดแบบสุ่ม(Random Error Model)และแบบไม่สุ่ม(Non-random error) พบว่า ประสิทธิภาพของ CSP-1 กรณี Non-random error มากกว่าrandom error และ CSP-1 i=5 f=1/5 มีประสิทธิภาพมากกว่า CSP-1 i=10 f=1/10 Dennis W.King และ Melynda Hazelwood (2000) ได้ศึกษาประสิทธิภาพของวิธีการสุ่ม 3 วิธีคือ (1).วิธีการสุ่มอย่างอย่างง่าย(SRS : Simple Random Sampling) (2).วิธี SRSR (Simple Random Sampling with Rectified) คล้ายกับการสุ่ม SRS แต่หากพบข้อมูลไม่ถกต้องเกิน 25 ชุด จะตรวจสอบข้อมูลทั้งหมดพบว่าและ (3). CSP-1 i=40 f=1/10 ผลการศึกษาพบว่า ที่สัดส่วนความผิดพลาดต่ำ (0.005, 0.010, 0.015) ทั้งสามวิธีมีประสิทธิภาพไม่แตกต่างกัน เมื่อ สัดส่วนความผิดพลาดมีค่าเพิ่มขึ้นเป็น 0.020 และ 0.025 SRSC มีประสิทธิภาพสูงสุด รองลงมาเป็น CSP-1 การศึกษาดังกล่าวข้างต้นเป็นการศึกษาในเชิงทฤษฎีหรือใช้กับข้อมูลประชากรจำลอง ยังไม่มีการศึกษาใดที่ใช้ข้อมูลนำเข้าจริง ดังนั้นผู้วิจัยจึงสนใจทำการศึกษาประสิทธิภาพของการสุ่ม CSP-1 กับข้อมูลจากการศึกษาอัตราความผิดพลาดของการนำเข้าข้อมูลด้วยวิธี Double Data Entry ที่มีมาตรฐานการปฏิบัติงานกำกับของประภัสสร เอื้อลลิตชูวงศ์ (2552) และได้เพิ่มวิธีการสุ่มอีก 3 วิธีดังนี้ (1). CSP-2 เนื่องจากเป็นวิธีที่ยืดหยุ่นมากกว่า CSP-1 ประกอบกับยังไม่เคยมีการศึกษามาก่อน (2). Systematic Continuous Sampling Plan 1 (SCSP-1) เป็นวิธีการสุ่มที่ผู้วิจัยได้ปรับปรุงให้ง่ายในทางปฏิบัติด้วยการสลับการตรวจสอบเป็นแบบคงที่คือ การตรวจสอบ 100% หนึ่งครั้ง สลับกับการตรวจสอบสัดส่วนหนึ่งครั้งโดยไม่ขึ้นอยู่กับผลการตรวจสอบก่อนหน้า (3).วิธี Systematic Continuous Sampling Plan 2 (SCSP-2) คล้ายกับ SCSP-1 เป็นการตรวจสอบ 100% หนึ่งครั้ง สลับกับการตรวจสอบสัดส่วน 2 ครั้ง ผลการวิจัยครั้งนี้จะเป็นแนวทางให้นักชีวสถิติและนักวิจัยทางการแพทย์

และทางสาธารณสุขเลือกใช้วิธีการสุ่มตรวจสอบข้อมูลให้เหมาะสมกับคุณภาพการนำเข้าข้อมูล และทรัพยากรในการดำเนินการ

วัตถุประสงค์การวิจัย

เพื่อเปรียบเทียบประสิทธิภาพของวิธีการสุ่มตัวอย่างในการตรวจสอบการนำเข้าข้อมูล 4 วิธี ได้แก่ CSP-1, CSP-2, SCSP-1 และ SCSP-2

วิธีการดำเนินการวิจัย

การวิจัยครั้งนี้เป็นการจำลองเพื่อศึกษาวิธีการสุ่ม 4 วิธี ได้แก่ CSP-1, CSP-2, SCSP-1 และ SCSP-2 ซึ่งแต่ละวิธีการสุ่มได้กำหนดค่า i และ f เป็น 4 ชุด ดังนี้

1) ชุดที่ 1 กำหนดให้ i =5, f=1/5
2) ชุดที่ 2 กำหนดให้ i =10, f=1/10
3) ชุดที่ 3 กำหนดให้ i =15, f=1/15
4) ชุดที่ 4 กำหนดให้ i =20, f=1/20

ดังนั้นในการศึกษาครั้งนี้จะมีทั้งหมด 16 แผนการสุ่มและประชากรที่ศึกษาคือ ข้อมูลที่นำเข้าครั้งเดียว (Single data entry : SDE) โดยใช้โปรแกรมสำเร็จรูป EpiData ที่มีมาตรฐานการปฏิบัติงาน (Standard Operating Procedure)กำกับจากการศึกษาของประภัสสร(2552) โดยประชากรจำนวน 8 ชุดจะมีสัดส่วนความผิดพลาดของการนำเข้า (p) เท่ากับ 0.0034, 0.0080, 0.0130, 0.0188, 0.0254, 0.0528, 0.0670, และ 0.0822 และประชากรแต่ละชุดจะมีทั้งหมด 5,000 เรคคอร์ด และทำการสุ่มซ้ำ จำนวน 10,000 ครั้งด้วยโปรแกรม Microsoft Office Excel 2007 โดยมีรายละเอียดของวิธีการสุ่มดังนี้

1) ***วิธี CSP-1*** ทำการตรวจสอบ 100% นั่นคือจะตรวจสอบ record ที่อยู่ต่อเนื่องกัน i ชุด และหากพบว่ามีบาง record ไม่ถูกต้อง จะทำการตรวจสอบ 100% ใน i record ถัดมา แต่ถ้าข้อมูล i record ทั้งหมดถูกต้อง จะสลับมาทำการตรวจสอบสัดส่วน f เมื่อ f = 1/ i คือ การสุ่มตรวจสอบ 1 record จาก i record ถัดมา ถ้า record ที่สุ่มได้ถูกต้อง จะทำการตรวจสอบสัดส่วน f ไปเรื่อยๆ จนกระทั่งพบ record ที่ไม่ถูกต้อง จึงจะสลับมาทำการตรวจสอบ 100% การสลับระหว่างการตรวจสอบ 100% และการตรวจสอบสัดส่วน f ขึ้นอยู่กับผลของการตรวจสอบก่อนหน้า หากข้อมูลนำเข้ามีคุณภาพต่ำ หรือมีสัดส่วนความผิดพลาดสูง โอกาสที่จะมีการตรวจสอบ 100% ก็จะมีมากและทำให้จำนวน record ที่ถูกตรวจสอบมีมากเช่นเดียวกัน

2) *CSP-2* เป็นวิธีคล้ายกับวิธี CSP-1 แต่ถ้าการตรวจสอบสัดส่วนพบ record ที่ไม่ถูกต้อง จะอนุโลมให้ทำการตรวจสอบสัดส่วนต่อไปได้ จนกว่าจะ record ที่ไม่ถูกต้องติดต่อกัน 2 ครั้ง จึงจะเปลี่ยนมาทำการตรวจสอบ 100%

3) *SCSP-1* เป็นวิธีคล้ายกับวิธี CSP-1 แต่จะทำการตรวจสอบ 100% หนึ่งครั้งสลับกับการตรวจสอบสัดส่วน f หนึ่งครั้งไปเรื่อยๆ ซึ่งวิธีนี้ทำได้ง่ายเพราะการสลับการตรวจสอบเป็นแบบคงที่ ไม่ว่าผลการตรวจสอบจะพบ record ที่ผิดหรือไม่ก็ตาม

4) *SCSP-2* เป็นวิธีคล้ายกับวิธี SCSP-1 แต่จะทำการตรวจสอบ 100% หนึ่งครั้งสลับกับการตรวจสอบสัดส่วน f สองครั้ง จำนวน record ที่ตรวจสอบโดยวิธี SCSP-2 จะน้อยกว่าวิธี SCSP-1

โดยทางปฏิบัติเมื่อมีการตรวจสอบพบว่า record ใดมีการนำเข้าที่ไม่ถูกต้อง record นั้น จะได้รับการแก้ไขใหม่ให้ถูกต้อง ซึ่งในการวิจัยครั้งนี้ก็เช่นเดียวกัน record ที่ถูกตรวจสอบพบว่าไม่ถูกต้อง จะถูกเปลี่ยนเป็น record ที่ถูกต้อง ทำให้สัดส่วนความผิดพลาดหลังการสุ่มมีค่าลดลง หรือ คุณภาพหลังการสุ่มเพิ่มขึ้น โดยประสิทธิภาพของการสุ่มจะพิจารณาจากค่าร้อยละเฉลี่ยของคุณภาพที่เพิ่มขึ้น (Percent Gain in Average Quality : PGAQ)

$$PGAQ = \left[\frac{p - \bar{p}_A}{p}\right] \times 100$$

เมื่อ $$\bar{p}_A = \frac{\sum_{j=1}^{k} p_{A,j}}{k}$$

p คือ สัดส่วนความผิดพลาดของการนำเข้า

$\bar{p}_A$ คือ ค่าเฉลี่ยสัดส่วนความผิดพลาดหลังการสุ่ม

$p_{A,j}$ คือ ค่าสัดส่วนความผิดพลาดหลังการสุ่มซ้ำครั้งที่ j

k คือ จำนวนสุ่มซ้ำ มีค่า 10,000

และตัวแปรวัดประสิทธิภาพของการสุ่มอีกตัวแปร ได้แก่ ค่าใช้จ่ายสำหรับการตรวจสอบข้อมูล ซึ่งพิจารณาจากค่าเฉลี่ยของสัดส่วน record ที่ถูกสุ่มตรวจสอบ (Average Fraction Inspection : AFI) ซึ่งถ้าค่า AFI มาก ค่าใช้จ่ายที่เกิดขึ้นก็จะมากเช่นเดียวกัน

$$AFI = \frac{R\bar{I}}{N}$$

เมื่อ $$R\bar{I} = \frac{\sum_{j=1}^{k} RI_j}{k}$$

$R\bar{I}$ คือ ค่าเฉลี่ยจำนวน record ที่ตรวจสอบ

RI_j คือ จำนวน record ที่ตรวจสอบของการสุ่มซ้ำที่ j

N คือ จำนวน record ทั้งหมด มีค่า 5,000

ผลการวิจัย

การเปรียบเทียบ PGAQ

พบว่า ที่ i = 5 และ f = 1/5 ทุกค่า p วิธี SCSP-1 จะมีค่า PGAQ สูงกว่าวิธีอื่น นั่นคือ มีค่าระหว่างร้อยละ 52 - 60 รองลงมาได้แก่ วิธี SCSP-2 มีค่าระหว่างร้อยละ 40 - 47 ส่วนวิธี CSP-1 และวิธี CSP-2 จะมีค่า PGAQ ต่ำสุดและใกล้เคียงกัน คือ ร้อยละ 20 - 22 และเมื่อพิจารณาการปรับค่าระดับของ p พบว่า ที่ p ระดับต่ำ (0.0034 ถึง 0.0254) เมื่อเพิ่มขึ้นเป็น 0.0528 0.0670 และ 0.0822 วิธี CSP-1 จะมีค่า PGAQ สูงกว่าวิธี CSP-2 (ดังตารางที่ 1)

ที่ i = 10 และ f = 1/10 ทุกค่า p วิธี SCSP-1 จะมีค่า PGAQ สูงกว่าวิธีอื่น นั่นคือ มีค่าระหว่างร้อยละ 48 - 55 รองลงมาได้แก่ วิธี SCSP-2 มีค่าระหว่างร้อยละ 34 - 41 ส่วนวิธี CSP-1 และวิธี CSP-2 จะมีค่า PGAQ ต่ำสุดและใกล้เคียงกัน คือ ร้อยละ 10 - 12 และเมื่อพิจารณาการปรับค่าระดับของ p พบว่า ที่ p ระดับต่ำ (0.0034 ถึง 0.0254) เมื่อเพิ่มขึ้นเป็น 0.0528 0.0670 และ 0.0822 ค่า PGAQ ของวิธี CSP-1 จะมีแนวโน้มเพิ่มสูงขึ้นมากกว่าวิธี CSP-2 (ดังตารางที่ 2)

ที่ i = 15 และ f = 1/15 ทุกค่า p วิธี SCSP-1 จะมีค่า PGAQ สูงกว่าวิธีอื่น นั่นคือ มีค่าระหว่างร้อยละ 46 - 53 รองลงมาได้แก่ วิธี SCSP-2 มีค่าระหว่างร้อยละ 32 - 38 ส่วนวิธี CSP-1 และวิธี CSP-2 จะมีค่า PGAQ ต่ำสุดและใกล้เคียงกัน คือ ร้อยละ 5 - 10 และเมื่อพิจารณาการปรับค่าระดับของ p พบว่า ที่ p ระดับต่ำ (0.0034 ถึง 0.0254) เมื่อเพิ่มขึ้นเป็น 0.0528 0.0670 และ 0.0822 ค่า PGAQ ของวิธี CSP-1 จะมีแนวโน้มเพิ่มสูงขึ้นมากกว่าวิธี CSP-2 (ดังตารางที่ 3)

ที่ i = 20 และ f = 1/20 ทุกค่า p วิธี SCSP-1 จะมีค่า PGAQ สูงกว่าวิธีอื่น นั่นคือ มีค่าระหว่างร้อยละ 45 - 53 รองลงมาได้แก่ วิธี SCSP-2 มีค่าระหว่างร้อยละ 32 - 38 ส่วนวิธี CSP-1 และวิธี CSP-2 จะมีค่า PGAQ ต่ำสุดและใกล้เคียงกัน คือ ร้อยละ 5 - 9 และเมื่อพิจารณาการปรับค่าระดับของ p พบว่า ที่ p ระดับต่ำ (0.0034 ถึง 0.0254) เมื่อเพิ่มขึ้นเป็น 0.0528 0.0670 และ 0.0822 ค่า PGAQ ของวิธี CSP-1 จะมีแนวโน้มเพิ่มขึ้นมากกว่าวิธี CSP-2 (ดังตารางที่ 4)

และทุกวิธีการสุ่ม เมื่อ i เพิ่มมากขึ้น ค่า PGAQ จะมีค่าลดลง โดยวิธี CSP-1 และวิธี CSP-2 จะให้ค่า PGAQ และ AFI ลดลงมากกว่าวิธี SCSP-1 และวิธี SCSP-2 โดยเฉพาะที่ระดับ i = 5 และ f = 1/5 และ i = 10 และ f = 1/10 พบว่า ค่า PGAQ ของวิธี CSP-1 และวิธี CSP-2 จะมีค่าลดลงจากเดิมประมาณร้อยละ 20 เป็นร้อยละ 10 ในขณะที่ค่า PGAQ ของวิธี SCSP-1 ลดลงจากเดิมประมาณร้อยละ 60 เป็นร้อยละ 55 และวิธี SCSP-2 ลดลงจากเดิมประมาณร้อยละ 46 เป็นร้อยละ 40

การเปรียบเทียบ AFI

ที่ I = 5 และ f = 1/5 (ดังตารางที่ 1) ทุกค่า p วิธี SCSP-1 จะมีค่า AFI สูงสุดและคงที่เท่ากับ 0.60 รองลงมาได้แก่ วิธี SCSP-2 มีค่าคงที่เท่ากับ 0.46 และวิธี CSP-2 จะมีค่า AFI ต่ำสุดประมาณ 0.20 และพิจารณาที่ระดับค่า p ต่ำ (0.0034 ถึง 0.0254) พบว่า วิธี CSP-1 จะมีค่า AFI ใกล้เคียงกับวิธี CSP-1 และเมื่อ p เพิ่มขึ้น พบว่า วิธี CSP-1 จะมีค่า AFI มากกว่าวิธี CSP-2 นั่นคือ

มีค่าระหว่าง 0.25-0.28 และเมื่อเพิ่มค่า i เป็น i = 10 และ f=1/10, i = 15 และ f = 1/15, i = 20 และ f = 1/20 รูปแบบของ AFI จะเหมือนกับค่า AFI ที่ระดับ i = 5 และ f = 1/5 นั่นคือ วิธี SCSP-1 จะให้ค่า AFI สูงที่สุด รองลงมาเป็นวิธี SCSP-2 ส่วนวิธี CSP-1 และวิธี CSP-2 พบว่าจะมีค่า AFI ใกล้เคียงกันที่ระดับ p ต่ำ และเมื่อระดับ p เพิ่มมากขึ้น วิธี CSP-1 จะให้ค่า AFI สูงกว่าวิธี CSP-2 (ดังตารางที่ 3, 4 และ 5)

บทสรุปและอภิปรายผล

จากผลการศึกษาวิธีการสุ่มตัวอย่างเพื่อตรวจสอบการนำเข้าข้อมูลที่มีค่า PGAQ และ AFI เรียงลำดับจากมากไปน้อย สรุปได้ดังนี้ วิธี SCSP-1, วิธี SCSP-2, วิธี CSP-1 และวิธี CSP-2 โดยวิธี SCSP-1 จะมีค่า PGAQ และค่า AFI สูงสุดและมีค่าคงที่ไม่ว่าข้อมูลจะมีคุณภาพมากหรือน้อย และวิธี SCSP-2 จะให้ผลเช่นเดียวกับวิธี SCSP-1 แต่ให้ค่า PGAQ และค่า AFI น้อยกว่าวิธี SCSP-1 และวิธี CSP-1 และวิธี CSP-2 ให้ผลใกล้เคียงกันเมื่อข้อมูลมีคุณภาพสูง และวิธี CSP-1 มีค่า PGAQ และค่า AFI สูงกว่าวิธี CSP-2 เมื่อข้อมูลมีคุณภาพลดลง

ดังนั้นจึงพบว่า ค่า PGAQ และค่า AFI จะมีความสอดคล้องกัน เมื่อจำนวน record ที่ต้องการตรวจสอบมากโอกาสที่พบข้อมูลที่ไม่ถูกต้องก็มากขึ้นตามไปด้วย จึงทำให้ค่า PGAQ มีค่าสูงขึ้น และเมื่อเปรียบเทียบวิธีการสุ่ม 4 วิธี พบว่า วิธีที่มีค่า PGAQ และ AFI สูงสุดคือ SCSP-1 รองลงมาเป็นวิธี SCSP-2 วิธีที่เหลือ CSP-1 และ CSP-2 จะมีค่าใกล้เคียงเมื่อข้อมูลมีคุณภาพสูง และวิธี CSP-1 จะดีกว่าเมื่อข้อมูลมีคุณภาพต่ำ โดยจำนวน record ที่ตรวจสอบด้วยวิธี SCSP-1 และ SCSP-2 จะมีค่าคงที่ไม่ขึ้นอยู่กับคุณภาพข้อมูล ทั้งนี้เนื่องจากเงื่อนไขของการสลับการตรวจสอบระหว่างการตรวจ 100% และการตรวจแบบสัดส่วนไม่ขึ้นกับผลการตรวจสอบ ส่วนวิธี SCSP-1 จะมีค่า PGAQ และ AFI มากกว่าวิธี SCSP-2 เสมอ เนื่องจากวิธี SCSP-2 จะตรวจสอบ 100% หนึ่งครั้งสลับกับการตรวจสัดส่วนสองครั้ง ขณะที่วิธี SCSP-1 จะสลับการตรวจสอบทั้งสองแบบอย่างละหนึ่งครั้งเท่านั้น สำหรับวิธี CSP-1 และ CSP-2 คุณภาพของข้อมูลมีผลต่อจำนวน record ที่ตรวจสอบ นั่นคือ ในการศึกษานี้ใช้ข้อมูลที่มีคุณภาพสูง เนื่องจากเป็นการนำเข้าข้อมูลที่มีมาตรฐานการปฏิบัติงานกำกับ ดังนั้นโอกาสที่จะพบ record ที่ไม่ถูกต้องจึงมีน้อย ทำให้การตรวจสอบแบบสัดส่วนมีมากกว่าการตรวจแบบ 100% ทำให้ค่า PGAQ และ AFI ของวิธี CSP-1 และ CSP-2 ต่ำกว่าวิธี SCSP-1 และ SCSP-2 และเมื่อคุณภาพของข้อมูลลดลง (p เพิ่มขึ้น) แสดงว่า จำนวน record ที่ไม่ถูกต้องมีมากขึ้น โอกาสที่จะสลับมาตรวจสอบ 100% จึงมีมากขึ้น ดังนั้นวิธี CSP-1 จึงมีโอกาสที่จะสลับมาตรวจสอบ 100% มากกว่าวิธี CSP-2 เนื่องจากหากการตรวจแบบสัดส่วนพบ record ที่ไม่ถูกต้อง ในวิธี CSP-1 จะกำหนดให้สลับมาตรวจสอบแบบ 100% ทันที ในขณะที่วิธี CSP-2 อนุโลมให้ตรวจแบบสัดส่วนต่อไปได้อีก จนกระทั่งพบ record ที่ไม่ถูกต้องติดต่อกันสองครั้งจึงจะสลับมาตรวจแบบ 100% ดังนั้นจึงทำให้ค่า PGAQ ของวิธี CSP-1 จึงมีแนวโน้มเพิ่มมากขึ้นเมื่อข้อมูลมีคุณภาพต่ำ เช่นเดียวกับการศึกษาของ Dennis W. King และ Roderick Lashley (2000) และเมื่อข้อมูลมีคุณภาพสูง โอกาสที่จะพบ record ที่ไม่ถูกต้องมีน้อย จึงทำให้การตรวจสอบแบบสัดส่วนมีมากกว่าการตรวจแบบ 100% ดังนั้นเมื่อ i เพิ่มขึ้น จำนวน record ที่ถูกตรวจสอบจึงน้อยลงและค่า PGAQ ก็ลดลงด้วยเช่นกัน และเนื่องจากวิธี SCSP-1 และวิธี SCSP-2 มีการสลับการตรวจสอบ 100% และแบบสัดส่วนคงที่ ทำให้ เมื่อ i เพิ่มขึ้น ค่า PGAQ จึงลดลงน้อยวิธีการสุ่มแบบ CSP-1 และแบบ CSP-2

ข้อเสนอแนะจากการวิจัย

จากผลการศึกษา พบว่า วิธีการสุ่มแบบ SCSP-1 มีจุดเด่น ได้แก่ การมีค่า PGAQ ที่สูงสุด และเป็นวิธีการที่มีเงื่อนไขในการสุ่มตรวจสอบน้อยที่สุด ดังนั้นจึงทำให้วิธีการสุ่มนี้มีความเป็นได้ในทางปฏิบัติมากสุด ขณะเดียวกันก็พบว่า ยังมีจุดด้อยเกี่ยวกับค่า AFI ที่มีค่าสูง ทั้งนี้เนื่องจากจำนวน record ที่ใช้ในการสุ่มตรวจสอบมีจำนวนมาก จึงทำให้ค่าใช้จ่ายในการตรวจสอบความถูกต้องของการนำเข้าข้อมูลมากด้วยเช่นกัน และข้อเสนอแนะในการทำ

วิจัยครั้งต่อไป เนื่องจากในการศึกษาครั้งนี้ใช้ข้อมูลด้วยวิธี Double Data Entry ที่มีมาตรฐานการปฏิบัติงานกำกับ ซึ่งถือเป็นข้อมูลที่มีคุณภาพสูง ดังนั้นหากมีการศึกษาในครั้งต่อไป จึงควรตรวจสอบความถูกต้องของข้อมูลที่มีสัดส่วนของข้อมูลนำเข้าไม่ถูกต้อง หรือข้อมูลที่มีคุณภาพต่ำ เพื่อนำผลที่ได้มาเปรียบเทียบและพัฒนาวิธีการสุ่มตัวอย่างที่ดีต่อไป

กิตติกรรมประกาศ

ขอขอบคุณผู้ช่วยศาสตราจารย์ ดร. จรณิต แก้วกังวาล รองศาสตราจารย์ชัชวาลย์ เรืองประพันธ์ คณาจารย์ภาควิชาชีวสถิติและประชากรศาสตร์ คณะสาธารณสุขศาสตร์ มหาวิทยาลัยขอนแก่นทุกท่าน ที่กรุณาให้ข้อเสนอแนะที่เป็นประโยชน์ต่อการศึกษาครั้งนี้ ตลอดจนช่วยตรวจสอบแก้ไขการศึกษาให้มีความสมบูรณ์ยิ่งขึ้น

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: ซมรมเด็ก.

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.**

ตารางที่ 1 แสดงค่า PGAQ และ AFI วิธีการสุ่มแบบ CSP-1, CSP-2, SCSP-1 และ SCSP-2 เมื่อ i = 5 และ f = 1/5

p	PGAQ				AFI			
	CSP-1	CSP-2	SCSP-1	SCSP-2	CSP-1	CSP-2	SCSP-1	SCSP-2
0.0034	20.59	20.59	58.82	47.06	0.2036	0.2008	0.6000	0.4672
0.0080	21.25	20.00	60.00	46.25	0.2074	0.2009	0.6000	0.4672
0.0130	20.77	20.00	60.00	46.92	0.2118	0.2010	0.6000	0.4672
0.0188	21.81	20.21	60.10	46.81	0.2170	0.2012	0.6000	0.4672
0.0254	22.44	20.08	59.84	46.46	0.2233	0.2015	0.6000	0.4672
0.0528	25.19	20.45	60.04	46.40	0.2526	0.2037	0.6000	0.4672
0.0670	22.84	17.91	52.39	40.75	0.2596	0.2045	0.6000	0.4672
0.0822	28.10	20.32	58.39	45.38	0.2879	0.2085	0.6000	0.4672

ตารางที่ 2 แสดงค่า PGAQ และ AFI วิธีการสุ่มแบบ CSP-1, CSP-2, SCSP-1 และ SCSP-2 เมื่อ i = 10 และ f = 1/10

p	PGAQ				AFI			
	CSP-1	CSP-2	SCSP-1	SCSP-2	CSP-1	CSP-2	SCSP-1	SCSP-2
0.0034	11.76	8.82	55.88	41.18	0.1050	0.1019	0.5500	0.4006
0.0080	11.25	10.00	55.00	40.00	0.1097	0.1021	0.5500	0.4006
0.0130	11.54	10.00	55.38	40.00	0.1151	0.1022	0.5500	0.4006
0.0188	12.23	10.11	54.79	39.89	0.1220	0.1026	0.5500	0.4006
0.0254	12.99	10.24	54.72	40.16	0.1307	0.1030	0.5500	0.4006
0.0528	17.80	10.80	54.92	39.96	0.1775	0.1072	0.5500	0.4006
0.0670	16.72	9.55	48.06	34.93	0.1901	0.1087	0.5500	0.4006
0.0822	24.09	11.31	53.53	38.93	0.2446	0.1166	0.5500	0.4006

ตารางที่ 3 แสดงค่า PGAQ และ AFI วิธีการสุ่มแบบ CSP-1, CSP-2, SCSP-1 และ SCSP-2 เมื่อ i = 15 และ f = 1/15

p	PGAQ				AFI			
	CSP-1	CSP-2	SCSP-1	SCSP-2	CSP-1	CSP-2	SCSP-1	SCSP-2
0.0034	5.88	5.88	52.94	38.24	0.0731	0.0700	0.5344	0.3804
0.0080	7.50	7.50	53.75	37.50	0.0782	0.0700	0.5344	0.3804
0.0130	8.46	6.92	53.08	37.69	0.0847	0.0704	0.5344	0.3804
0.0188	9.04	6.91	53.19	37.77	0.0932	0.0709	0.5344	0.3804
0.0254	10.23	7.09	53.54	37.80	0.1043	0.0717	0.5344	0.3804
0.0528	17.05	7.77	53.41	37.88	0.1710	0.0785	0.5344	0.3804
0.0670	16.87	7.01	46.57	32.99	0.1908	0.0812	0.5344	0.3804
0.0822	27.62	9.37	51.95	36.86	0.2802	0.0959	0.5344	0.3804

ตารางที่ 4 แสดงค่า PGAQ และ AFI วิธีการสุ่มแบบ CSP-1, CSP-2, SCSP-1 และ SCSP-2 เมื่อ i = 20 และ f = 1/20

p	PGAQ				AFI			
	CSP-1	CSP-2	SCSP-1	SCSP-2	CSP-1	CSP-2	SCSP-1	SCSP-2
0.0034	5.88	5.88	52.94	38.23	0.0575	0.0540	0.5250	0.3692
0.0080	6.25	5.00	52.50	36.25	0.0631	0.0545	0.5250	0.3692
0.0130	6.92	5.38	52.31	36.92	0.0706	0.0551	0.5250	0.3692
0.0188	7.98	5.85	52.66	36.70	0.0805	0.0560	0.5250	0.3692
0.0254	9.45	5.51	52.36	37.00	0.0944	0.0572	0.5250	0.3692
0.0528	18.56	6.82	52.46	36.93	0.1864	0.0683	0.5250	0.3692
0.0670	19.25	6.42	45.82	32.24	0.2164	0.0731	0.5250	0.3692
0.0822	34.55	9.85	51.09	35.89	0.3489	0.1003	0.5250	0.3692