

ชื่อเรื่อง : การพัฒนาวิธีการจัดการข้อมูลสูญหายแบบอีพีเอสเอสอีและ
การตรวจสอบความแม่นยำและอำนาจการทดสอบเปรียบเทียบ
กับวิธีอีเอ็มและลิสท์ไวส์ : เทคนิคอนติ คาร์โล

ผู้วิจัย : นายเชาว์ อินโย

ที่ปรึกษา : รองศาสตราจารย์ ดร.รัตนะ บัวสนธิ์

ประเภทสารนิพนธ์ : วิทยานิพนธ์ กศ.ด.(วิจัยและประเมินผลการศึกษา)
มหาวิทยาลัยนเรศวร, 2547

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อพัฒนาวิธีการจัดการข้อมูลสูญหายแบบอีพีเอสเอสอีและตรวจสอบความแม่นยำและอำนาจการทดสอบที่ได้จากวิธีการจัดการข้อมูลสูญหายแบบอีพีเอสเอสอีกับแบบอีเอ็มและแบบลิสท์ไวส์ ตามวิธีการสุ่มตัวอย่างแบบแบ่งชั้น แบบกลุ่ม และแบบหลายขั้นตอน ที่ความสัมพันธ์ระหว่างตัวแปรระดับต่ำ ($r=.30$) ปานกลาง ($r=.50$) และสูง ($r=.70$) และจำนวนข้อมูลสูญหาย 5% 10% 20% และ 30% และศึกษาปฏิสัมพันธ์ระหว่างวิธีการสุ่มตัวอย่าง วิธีการจัดการข้อมูลสูญหาย จำนวนข้อมูลสูญหาย และความสัมพันธ์ระหว่างตัวแปร ที่มีต่อความแม่นยำของค่าเฉลี่ยเลขคณิต ความแปรปรวน และสัมประสิทธิ์สหสัมพันธ์ ข้อมูลที่ใช้ศึกษามีลักษณะการแจกแจงแบบปกติสองตัวแปรและใช้เทคนิคอนติคาร์โล จิมูเลชัน จำลองการทดลองด้วยเครื่องคอมพิวเตอร์โดยการทำซ้ำจำนวน 1,000 ครั้ง

ผลการวิจัยพบว่า

1. วิธีการจัดการข้อมูลสูญหายโดยการแทนค่าแบบอีพีเอสเอสอีได้ค่าความแม่นยำของค่าเฉลี่ยเลขคณิต ไม่แตกต่างจากวิธีอีเอ็ม อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และเป็นค่าที่สูงที่สุด เมื่อใช้วิธีการสุ่มตัวอย่างแบบแบ่งชั้น แบบกลุ่มและแบบหลายขั้นตอน จำนวนข้อมูลสูญหายอยู่ในระดับสูงที่สุดคือเท่ากับ 30% ความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับสูง ($r=.70$)
2. วิธีการจัดการข้อมูลสูญหายโดยการตัดออกแบบลิสท์ไวส์ได้ค่าความแม่นยำของความแปรปรวนแตกต่างจากวิธีอื่น ๆ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และเป็นค่าที่สูงที่สุด เมื่อใช้วิธีการสุ่มตัวอย่างแบบแบ่งชั้น แบบกลุ่มและแบบหลายขั้นตอน จำนวนข้อมูลสูญหาย 10% 20% และ 30% ความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับต่ำ ($r=.30$) ปานกลาง ($r=.50$) และสูง ($r=.70$)

3. วิธีการจัดการข้อมูลสูญหายแบบลิสต์ไวส์ ได้ค่าความแม่นยำของสัมประสิทธิ์สหสัมพันธ์แตกต่างจากวิธีการจัดการข้อมูลสูญหายแบบอื่น ๆ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และเป็นค่าที่สูงที่สุด เมื่อใช้วิธีการสุ่มตัวอย่างแบบแบ่งชั้น แบบกลุ่ม และแบบหลายขั้นตอน จำนวนข้อมูลสูญหายอยู่ในระดับสูงที่สุดคือเท่ากับ 30% ความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับสูง ($r=.70$)

4. วิธีการจัดการข้อมูลสูญหายโดยการแทนค่าแบบอีพีเอสเอสอี ได้ค่าอำนาจการทดสอบ เมื่อใช้การทดสอบความสัมพันธ์สูงกว่าวิธีการจัดการข้อมูลสูญหายโดยการตัดออกแบบลิสต์ไวส์และการแทนค่าแบบอีเอ็ม เมื่อใช้วิธีการสุ่มตัวอย่างแบบแบ่งชั้น จำนวนข้อมูลสูญหายทุกระดับ (5%, 10%, 20% และ 30%) ความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับต่ำ ($r=.30$)

5. ปฏิสัมพันธ์ระหว่างวิธีการสุ่มตัวอย่าง วิธีการจัดการข้อมูลสูญหาย จำนวนข้อมูลสูญหาย และความสัมพันธ์ระหว่างตัวแปรที่มีต่อความแม่นยำของค่าเฉลี่ยเลขคณิต ความแปรปรวน และสัมประสิทธิ์สหสัมพันธ์ ไม่มีนัยสำคัญทางสถิติที่ระดับ .01

6. ปฏิสัมพันธ์สามทางระหว่าง 1). วิธีการสุ่มตัวอย่างกับวิธีการจัดการข้อมูลสูญหาย และจำนวนข้อมูลสูญหาย 2). วิธีการสุ่มตัวอย่างกับวิธีการจัดการข้อมูลสูญหายและความสัมพันธ์ระหว่างตัวแปร 3). วิธีการจัดการข้อมูลสูญหายกับจำนวนข้อมูลสูญหายและความสัมพันธ์ระหว่างตัวแปร ที่มีต่อความแม่นยำของความแปรปรวน มีนัยสำคัญทางสถิติที่ระดับ .01 และปฏิสัมพันธ์สามทางระหว่างระหว่างวิธีการสุ่มตัวอย่างกับวิธีการจัดการข้อมูลสูญหายและจำนวนข้อมูลสูญหาย ที่มีต่อความแม่นยำของสัมประสิทธิ์สหสัมพันธ์ มีนัยสำคัญทางสถิติที่ระดับ .01

Title : THE DEVELOPMENT OF ESTIMATED PARAMETER AND
THE SMALLEST STANDARD ERROR (EPSSE) MISSING
DATA TREATMENT METHOD AND DETECTING OF
ACCURACY AND STATISTICAL POWER COMPARISON
WITH EM ALGORITHM AND LISTWISE METHOD : A
MONTE CARLO STUDY

Author : Mr.Chao Inyai

Adviser : Asso. Prof. Dr. Rattana Buosonte

Type of Degree : Doctor of Philosophy Degree in Educational Research and
Evaluation (Ph.d. in Educational Research and Evaluation)
Naresuan University, 2004

Abstract

The purposes of this study were to develop the EPSSE missing data treatment method and to compare its accuracy and statistical power with EM algorithm and Listwise deletion missing data treatment method. The comparisons were made under the three sampling methods (stratified random sampling, cluster random sampling and multi-stage random sampling), and the three level of correlation of variables (low $r=.30$, medium $r=.50$ and high $r=.70$) and the four level of number of missing data (5%, 10%, 20% and 30%). A study of interactions between sampling method, missing data treatment method, number of missing data and correlation of variables effected on the accuracy of mean, variance and coefficient of correlation were also made. The character of data was bivariate normal distribution. This study simulated the experiment with computer using Monte Carlo Simulation technique by iterative 1,000 times.

The finding were as follows :

1. EPSSE missing data treatment method provided the accuracy of mean is not statistical significant different at .05 level from EM algorithm missing data treatment method and the highest values when using the stratified random sampling, the cluster

random sampling and the multi-stage random sampling , number of missing data is 30% and correlation of variables is high ($r=.70$).

2. Listwise deletion missing data treatment method provided the accuracy of variance is statistical significant different at .05 level from other method and the highest values when using the stratified random sampling , the cluster random sampling , number of missing data is 10% , 20% and 30% and correlation of variables is low ($r=.30$) , medium ($r=.50$) and high ($r=.70$).

3. Listwise missing data treatment method provided the accuracy of coefficient of correlation is statistical significant different at .05 level from other method and the high values when using the stratified random sampling , the cluster random sampling and the multi-stage random sampling , number of missing data is 30% and correlation of variables is high ($r=.70$).

4. EPSSE missing data treatment method provided the statistical power when the test of correlation higher than listwise deletion and EM algorithm when using the stratified random sampling , all level of number of missing data (5% , 10% , 20% and 30%) and correlation of variables is low ($r=.30$).

5. Four way interaction between sampling method , missing data treatment method , number of missing data and correlation of variables effected on the accuracy of mean , variance and coefficient of correlation are not statistical significant at .01 level.

6. Three way interaction between 1). Sampling method with missing data treatment method and number of missing data. 2). Sampling method with missing data treatment method and correlation of variables. 3). Missing data treatment method with number of missing data and correlation of variables effected on the accuracy of variance are statistical significant at .01 level. Three way interaction between sampling method with missing data treatment method and number of missing data effected on the accuracy of coefficient of correlation is statistical significant at .01 level.