

การเป็นอิสระต่อกันและมาจากรูปแบบความน่าจะเป็นเดียวกัน เป็นคุณสมบัติซึ่งเป็นข้อสมมติเบื้องต้นของข้อมูลในการสร้างตัวสถิติทดสอบเพียร์สันไคกำลังสองและตัวสถิติทดสอบอัตราส่วนควรจะเป็นไคกำลังสองสำหรับทดสอบความเป็นอิสระ ซึ่งในกรณีที่ข้อมูลได้มาจากการสุ่มตัวอย่างอย่างง่ายแบบไม่ใส่คืนจะเป็นข้อมูลที่ขาดคุณสมบัติดังกล่าวนี้ ข้อสรุปที่ได้จากการทดสอบจึงย่อมมีความคลาดเคลื่อนหรือข้อผิดพลาด

ในการวิจัยครั้งนี้จึงมีวัตถุประสงค์เพื่อตรวจสอบความเชื่อถือได้ของการนำตัวสถิติทั้ง 2 มาทำการทดสอบความเป็นอิสระสำหรับข้อมูลจากการสำรวจตัวอย่าง ทั้งจากการสุ่มตัวอย่างอย่างง่ายแบบใส่คืนและแบบไม่ใส่คืน การศึกษาดำเนินการโดยการจำลองข้อมูล โดยสถานการณ์ที่ศึกษา คือ ข้อมูลในตารางการแจกแจงขนาด 2×2 กำหนดขนาดตัวอย่างเท่ากับ 50 , 100 , 150 และ 200 ระดับนัยสำคัญที่กำหนดเท่ากับ 0.01 และ 0.05 มีการทำซ้ำจำนวน 1000 ครั้งในแต่ละสถานการณ์ ผลการวิจัยพบว่า กรณีที่ข้อมูลได้จากการสุ่มตัวอย่างอย่างง่ายแบบใส่คืน มีค่าความผิดพลาดประเภทที่ 1 น้อยกว่าและใกล้เคียงกับระดับนัยสำคัญที่กำหนดมากกว่ากรณีแบบไม่ใส่คืน โดยที่ค่าความผิดพลาดประเภทที่ 1 ของทั้ง 2 กรณีจะมีค่าลดลงและ ใกล้เคียงกับระดับนัยสำคัญมากขึ้นเมื่อขนาดตัวอย่างเพิ่มมากขึ้น และเมื่อขนาดประชากรเพิ่มมากขึ้น ค่าความผิดพลาดประเภทที่ 1 กรณีที่ข้อมูลได้จากการสุ่มตัวอย่างอย่างง่ายแบบไม่ใส่คืนจะมีค่าเข้าใกล้ค่าความผิดพลาดประเภทที่ 1 กรณีที่ข้อมูลได้จากการสุ่มตัวอย่างอย่างง่ายแบบใส่คืน

ดังนั้นสรุปได้ว่าการขาดคุณสมบัติการเป็นอิสระต่อกันและมาจากรูปแบบความน่าจะเป็นเดียวกันของข้อมูลที่ได้จากการสุ่มตัวอย่างอย่างง่ายแบบไม่ใส่คืน ทำให้ผลการทดสอบความเป็นอิสระเกิดความผิดพลาดสูง ซึ่งจะนำไปสู่ข้อสรุปและการตัดสินใจที่ไม่ถูกต้อง การประยุกต์ใช้ระเบียบวิธีทางสถิติในลักษณะเช่นนี้ จึงต้องทำความเข้าใจในรากฐานทฤษฎีการอนุมานทางสถิติ และทฤษฎีการสำรวจตัวอย่างให้ถ่องแท้ และจึงจำเป็นอย่างยั้งที่จะต้องมีความเข้าใจในทฤษฎีความน่าจะเป็นในระดับที่ดี

Independent and identically distributed are the fundamental properties of data in constructing Pearson's chi-square, and likelihood ratio for testing independence between two factors. In fact, when data were collected through a sample survey using simple random sampling without replacement scheme, they are lack of these properties. Consequently, thus obtained conclusion may be unreliable.

The main objective of this study is to investigate the reliability of these two test statistics when apply to survey data from both with and without replacement sampling schemes. The investigation is based on simulation for two by two contingency table with the sample size of 50 , 100 , 150 and 200 of 0.01 and 0.05 level of significance under 1000 repetitions for each scenario.

It was found that type one error of these two test statistics are lower and closed to the specified level of significance for those cases of data from sampling with replacement comparing to those of without replacement scheme. Obviously, when sample size is increased, type one error of these two sampling scheme are decreased. Most importantly, when population size is increased, type one error of those without replacement is closed to the cases of with replacement.

It can be concluded that when data are not independent and identically distributed, as in the case of without replacement scheme, type one error of test statistics is higher than the case of with replacement. This therefore is resulting in wrong conclusion and decision. Indeed, the application of statistical method needs clear understanding in foundation of theories of statistical inference, and sample survey, as well as theory of probability.