

บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

ในการทำวิจัยเชิงปริมาณจำเป็นต้องอาศัยวิธีการทางสถิติเพื่อช่วยในการวิเคราะห์ข้อมูล ตัดสินใจ และสรุปผล โดยเฉพาะอย่างยิ่งการสรุปอ้างอิงไปสู่ประชากรที่อาศัยข้อมูลจากกลุ่มตัวอย่าง ถ้าผู้วิจัยเลือกใช้สถิติทดสอบได้อย่างถูกต้องและเหมาะสม จะทำให้ผลสรุปของการวิจัยมีความถูกต้องและสมเหตุสมผล ลักษณะของข้อมูลหรือตัวแปรที่ศึกษาเป็นส่วนสำคัญที่ผู้วิจัยใช้ในการตัดสินใจว่าจะเลือกใช้สถิติใดในการวิเคราะห์ข้อมูล ลักษณะของข้อมูลจำแนกได้เป็น 4 ระดับ คือ มาตรฐานนามบัญญัติ (nominal scale) มาตรฐานจัดอันดับ (ordinal scale) มาตรฐานมาตรา (interval scale) มาตรฐานอัตราส่วน (ratio scale) ซึ่งในมาตราแต่ละระดับจะใช้สถิติเชิงบรรยาย และทดสอบที่แตกต่างกัน

สำหรับข้อมูลที่มีลักษณะเป็นมาตรฐานนามบัญญัติและมาตรฐานจัดอันดับ สามารถนำข้อมูลที่ได้มาจัดให้อยู่ในรูปของตารางแจกแจงความถี่ ค่าที่ปรากฏในตารางจะเป็นความถี่ของค่าสังเกตที่เกิดขึ้นรวมกันได้ (observed frequency) เรียกตารางที่มีลักษณะแบบนี้ว่าตารางการณัจร (contingency table) ข้อมูลที่นำเสนอในรูปตารางการณัจรนี้จะเรียกว่าข้อมูลจำนวนนับ (counted data) หรือข้อมูลจำแนกประเภท (categorical data) การทดสอบสมมติฐานข้อมูลที่มีลักษณะเช่นนี้มักจะทำให้ทดสอบความเท่ากันของสัดส่วนของประชากร เรียกว่าการทดสอบความเป็นเอกพันธ์ของสัดส่วน (test of homogeneity of proportion) ทดสอบความเท่ากันของการแจกแจงของประชากร เรียกว่าการทดสอบความเป็นเอกพันธ์ของการแจกแจง (test of homogeneity of distribution) ทดสอบความแตกต่างของค่าที่ได้จากการคาดหวัง (expected values) ภายใต้ทฤษฎีใดทฤษฎีหนึ่งกับค่าสังเกตได้ (observed values) เรียกว่าการทดสอบภาวะสารูปสนิทธิ (goodness of fit test) และทดสอบความเป็นอิสระระหว่างตัวแปร เรียกว่าการทดสอบความเป็นอิสระ (test of independence)

การทดสอบความเป็นอิสระ (test of independence) เป็นการทดสอบหนึ่งที่นิยมใช้ในการวิจัยทางด้านศึกษาศาสตร์และด้านสังคมศาสตร์เพื่อศึกษาว่าตัวแปรจำแนกประเภท 2 ตัว (categorical variable) เป็นอิสระต่อกันหรือไม่ โดยตั้งสมมติฐานหลักว่าตัวแปรจำแนกประเภททั้งสองเป็นอิสระต่อกัน ($H_0 : p_{ij} = p_i \cdot p_j$) ซึ่งถ้าผลการทดสอบปฏิเสธสมมติฐานหลัก แสดงว่าตัวแปรทั้งสองมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติ แต่ถ้าผลการทดสอบยอมรับสมมติฐาน

หลัก แสดงว่า ตัวแปรทั้งสองเป็นอิสระต่อกัน นักสถิติบางท่านเรียกชื่อการทดสอบแบบนี้ว่า การทดสอบความสัมพันธ์ระหว่างตัวแปร 2 ตัว (test of association) สำหรับการทดสอบสมมติฐานของการทดสอบความเป็นอิสระของตัวแปร จะใช้สถิติทดสอบไคสแควร์ (chi-square statistic) ซึ่งสถิติทดสอบไคสแควร์จะมีการแจกแจงแบบไคสแควร์ (chi-square distribution) คือมีการแจกแจงแบบเบ้ขวา (positively skewed distribution) ลักษณะของการแจกแจงแบบไคสแควร์จะขึ้นอยู่กับองศาอิสระ (degree of freedom) เมื่อองศาอิสระเพิ่มขึ้นจะทำให้การแจกแจงแบบไคสแควร์เข้าใกล้การแจกแจงแบบปกติ (normal distribution) องศาอิสระของการทดสอบไคสแควร์ที่ใช้ในการทดสอบความเป็นอิสระจะเท่ากับ $(r - 1)(c - 1)$ (Howell, 1997, 1999; Agresti, 1990, 1996) สถิติที่นิยมใช้ในการทดสอบคือ เพียร์สันไคสแควร์ (Pearson's chi-square statistic) และอัตราส่วนภาวะน่าจะเป็น (Likelihood-ratio chi-square statistic)

เพียร์สันไคสแควร์เป็นสถิติที่ใช้ทดสอบความเป็นอิสระระหว่างตัวแปรแบบจำแนกประเภท 2 ตัวที่จัดอยู่ในรูปตารางการถัว 2 ทาง (Glass & Stanley, 1970) ค่าของเพียร์สันไคสแควร์จะขึ้นอยู่กับค่าความถี่ที่สังเกตได้และค่าความถี่ที่คาดหวัง เพียร์สันไคสแควร์จะให้ค่าน้อยที่สุดเป็น 0 เมื่อความถี่ที่สังเกตได้มีค่าเท่ากับค่าความถี่ที่คาดหวัง เมื่อกลุ่มตัวอย่างมีขนาดเพิ่มขึ้นจะทำให้เพียร์สันไคสแควร์มีการแจกแจงที่เข้าใกล้การแจกแจงแบบไคสแควร์ (Agresti, 1990, 1992) เมื่อพิจารณาค่าของเพียร์สันไคสแควร์พบว่าเกิดจากผลรวมของผลต่างระหว่างความถี่ของค่าที่สังเกตได้กับความถี่ของค่าที่คาดหวังยกกำลังสองหารด้วยความถี่ที่คาดหวัง จึงอาจกล่าวได้ว่า เมื่อจำนวนแถวบนหรือแถวตั้งของตารางการถัวเพิ่มขึ้นอาจทำให้ค่าไคสแควร์สูงขึ้นด้วย ทั้งที่ระดับความสัมพันธ์ระหว่างตัวแปรยังเหมือนเดิม (วันทิพย์ เดชชูไชย, 2529) เพียร์สันไคสแควร์นิยมใช้ทดสอบความเป็นอิสระเมื่อกลุ่มตัวอย่างมีขนาดใหญ่ ค่าความถี่ที่คาดหวังในแต่ละเซลล์ไม่ควรต่ำกว่า 5 ($E_{ij} \geq 5; i = 1, 2, 3, \dots, r$ และ $j = 1, 2, 3, \dots, c$) หรือค่าความถี่ที่คาดหวังน้อยกว่า 5 ไม่เกิน 20% ของจำนวนเซลล์ทั้งหมด และไม่มีเซลล์ใดที่มีค่าความถี่ที่คาดหวังน้อยกว่า 1 (กัลยา วานิชย์บัญชา, 2544)

อัตราส่วนภาวะน่าจะเป็นเป็นสถิติอีกตัวหนึ่งของสถิติทดสอบไคสแควร์ที่ใช้ในการทดสอบสมมติฐานของการทดสอบความเป็นอิสระ อัตราส่วนภาวะน่าจะเป็นจะให้ค่าน้อยที่สุดเป็น 0 เมื่อความถี่ที่สังเกตได้มีค่าเท่ากับค่าความถี่ที่คาดหวัง และเมื่อกลุ่มตัวอย่างมีขนาดใหญ่ อัตราส่วนภาวะน่าจะเป็นมีการแจกแจงแบบไคสแควร์ (Wilks 1935, 1938) เช่นเดียวกับเพียร์สันไคสแควร์ Agresti (1990, 1996) กล่าวว่า โดยทั่วไปเมื่อกลุ่มตัวอย่างมีขนาดใหญ่เพียร์สันไคสแควร์และอัตราส่วนภาวะน่าจะเป็นจะให้ผลสรุปของการทดสอบเหมือนกัน แต่สำหรับกลุ่มตัวอย่างที่มีขนาดเล็ก พบว่า เพียร์สันไคสแควร์มีการแจกแจงใกล้เคียงการแจกแจงแบบไคสแควร์ได้ดีกว่า อัตราส่วนภาวะน่าจะเป็นนอกจากจะใช้ในการทดสอบความเป็นอิสระแล้วยังสามารถนำไปใช้ในการวิเคราะห์

Loglinear Model สำหรับตารางการถ้อยหลายทางได้อีกด้วย (Howell, 1997) สำหรับการทดสอบสมมติฐานของเพียร์สันไคสแควร์และอัตราส่วนภาวะน่าจะเป็นใช้เกณฑ์ในการตัดสินใจเช่นเดียวกัน กล่าวคือจะปฏิเสธสมมติฐานหลัก ถ้าค่าเพียร์สันไคสแควร์หรืออัตราส่วนภาวะน่าจะเป็นที่ได้จากการคำนวณมีค่ามากกว่าไคสแควร์จากตารางการแจกแจงแบบไคสแควร์ที่องศาอิสระเท่ากับ $(r - 1)(c - 1)$ ณ ระดับนัยสำคัญที่กำหนด หรือจะปฏิเสธสมมติฐานหลัก ถ้า P value ของเพียร์สันไคสแควร์หรืออัตราส่วนภาวะน่าจะเป็นที่คำนวณได้มีค่าน้อยกว่าระดับนัยสำคัญที่กำหนด

จากที่กล่าวมาข้างต้นสถิติทดสอบไคสแควร์มีข้อจำกัดในเรื่องของค่าความถี่ที่คาดหวังและขนาดของกลุ่มตัวอย่าง Yate (1934) เสนอว่า ควรปรับแก้ค่าของเพียร์สันไคสแควร์ด้วยการเอา 0.5 ลบออกจากผลต่างของค่าความถี่ที่สังเกตได้กับค่าความถี่ที่คาดหวังก่อนที่จะยกกำลังสองในการคำนวณตามปกติ เรียกสถิติตัวนี้ว่าการปรับแก้ของเยทส์ (Yate's continuity correction) เพื่อเป็นเกียรติแก่ Dr. Frank Yate ในการทดสอบสมมติฐานของสถิติทดสอบการปรับแก้ของเยทส์ ใช้เกณฑ์ในการตัดสินใจเช่นเดียวกับเพียร์สันไคสแควร์ และอัตราส่วนภาวะน่าจะเป็น Howell (1997) กล่าวว่า มีบทความที่ตีพิมพ์เกี่ยวกับการใช้สถิติทดสอบการปรับแก้ของเยทส์อย่างแพร่หลายซึ่งมีทั้งบทความที่เห็นด้วยและไม่เห็นด้วย โดย Howell แนะนำว่าการใช้การปรับแก้ของเยทส์ในการทดสอบ ผู้วิจัยไม่สามารถที่จะใช้การปรับแก้ของเยทส์ได้ในทุกกรณี แต่การใช้การปรับแก้ของเยทส์จะขึ้นอยู่กับกรณีที่ผู้วิจัยทำการศึกษามากกว่าว่าผู้วิจัยมีความเข้าใจและพิถีพิถันในใช้สถิติทดสอบมากน้อยเพียงใด โดยทั่วไปการปรับแก้ของเยทส์จะใช้เมื่อตารางการถ้อยมีขนาด 2×2 และกลุ่มตัวอย่างที่มีขนาดเล็กเท่านั้น เช่นเดียวกับ Fisher's exact test ที่ใช้ทดสอบความเป็นอิสระในกรณีที่มีตัวแปรจำแนกประเภททั้ง 2 ตัวมีค่าเพียง 2 ค่า ซึ่งมีองศาอิสระเท่ากับ 1 (ตารางการถ้อยขนาด 2×2) Fisher's exact test มีรูปแบบการแจกแจงแบบไฮเปอร์จีอเมตริก (hypergeometric distribution) สำหรับการทดสอบสมมติฐานหลักของการทดสอบความเป็นอิสระโดยการใช้ Fisher's exact test จะนำ P value ที่คำนวณได้ไปเปรียบเทียบกับระดับนัยสำคัญที่กำหนด โดยจะปฏิเสธสมมติฐานหลัก ถ้าค่า P value ที่คำนวณได้มีค่าน้อยกว่าระดับนัยสำคัญที่กำหนด ในอดีต Fisher's exact test สามารถใช้ตารางสำเร็จได้เฉพาะกลุ่มตัวอย่างที่มีขนาดไม่เกิน 15 เท่านั้น (Marascuilo และ Mcsweeney, 1977) หากกลุ่มตัวอย่างมีขนาดมากกว่านี้ก็ต้องคำนวณจากสูตรซึ่งคำนวณได้ยากและซับซ้อน ค่าที่ได้จากการคำนวณของ Fisher's exact test จะเป็น exact probability value ซึ่งจะแตกต่างจากค่าที่คำนวณได้จากไคสแควร์เพราะค่าที่คำนวณได้จากไคสแควร์นั้นจะเป็นค่าประมาณของ true probability value โดยค่าประมาณที่ได้จากไคสแควร์อาจจะไม่มีความแม่นยำนักในกรณีที่ค่าความถี่ที่คาดหวังของเซลล์ใดเซลล์หนึ่งในตารางการถ้อยมีค่าน้อยกว่า 5 โดยทั่วไปแล้วการทดสอบความเป็นอิสระจะนิยมใช้สถิติทดสอบไคสแควร์มากกว่า Fisher's exact test เพราะคำนวณง่าย

สูตรไม่มีความยุ่งยากและซับซ้อน แต่ในปัจจุบันได้มีการพัฒนาโปรแกรมคอมพิวเตอร์เพื่อช่วยในการคำนวณค่าสถิติได้ง่ายขึ้น เช่น โปรแกรม SPSS (Statistical Package for Social Science) ซึ่งสามารถคำนวณค่า Fisher's exact test และการปรับแก้ของเยทส์ในกรณีที่กลุ่มตัวอย่างมีขนาดใหญ่ได้ โดยอาศัยโปรแกรมคอมพิวเตอร์ในการคำนวณซึ่งไม่ยุ่งยากดังที่ต้องอาศัยจากการคำนวณมืออีกต่อไป

Cochran (1954) ได้เสนอหลักในการพิจารณาการใช้สถิติทดสอบความเป็นอิสระของตารางการถัวดังนี้

1. สำหรับตารางการถัวขนาด 2×2 ถ้าจำนวนความถี่ทั้งหมดน้อยกว่า 20 ให้ใช้สถิติทดสอบ Fisher's exact test หรือถ้าจำนวนความถี่อยู่ระหว่าง 20 – 40 และค่าความถี่ที่คาดหวังในเซลล์ใดเซลล์หนึ่งน้อยกว่า 5 ให้ใช้การปรับแก้ของเยทส์ (Yate's continuity correction)

2. สำหรับตารางการถัวขนาด $r \times c$ ถ้าความถี่ที่คาดหวังที่น้อยกว่า 5 น้อยกว่า 20% ของจำนวนเซลล์ทั้งหมดก็สามารถใช้สถิติทดสอบไคสแควร์ในการทดสอบได้

จากที่กล่าวข้างต้นในการทดสอบความเป็นอิสระสำหรับข้อมูลที่อยู่ในรูปตารางการถัว สถิติทดสอบที่นิยมใช้ได้แก่ เพียร์สันไคสแควร์ อัตราส่วนภาวะน่าจะเป็น การปรับแก้ของเยทส์ และ Fisher's exact test ซึ่งทั้งเพียร์สันไคสแควร์ และอัตราส่วนภาวะน่าจะเป็น เป็นสถิติที่ใช้ในการทดสอบไคสแควร์ที่มีข้อจำกัดในการใช้ ในหนังสือสถิติมาตรฐานที่ใช้อยู่ทั่วไป แนะนำว่า การใช้ไคสแควร์ในการทดสอบความเป็นอิสระ ค่าความถี่ที่คาดหวังควรเป็น 5 หรือมากกว่า (Marascuilo และ Mesweeney 1977; Yamane 1973) โดยยึดหลักของการแจกแจงแบบทวินาม (Binomial Distribution) และให้เกณฑ์ในการตัดสินใจไว้ว่าถ้าความถี่ที่คาดหวังเป็น 5 หรือมากกว่าแล้วจะสามารถประมาณค่าของการแจกแจงแบบทวินามได้ด้วยการแจกแจงปกติ จึงอาจสรุปได้ว่าค่าความถี่ที่คาดหวังเป็น 5 หรือมากกว่าสอดคล้องกับข้อตกลงเบื้องต้นของการแจกแจงปกติ แต่ในบางกรณีที่ผู้วิจัยทำการทดสอบความเป็นอิสระแล้วพบว่าเซลล์ที่มีค่าความถี่ที่คาดหวังน้อยกว่า 5 ผู้วิจัยควรจะเลือกใช้สถิติทดสอบตัวใดในการวิเคราะห์ข้อมูลเพื่อให้ได้ผลการวิจัยที่ถูกต้องและสมเหตุสมผล และในการทดสอบความเป็นอิสระกรณีเช่นนี้ผู้วิจัยจะยังสามารถใช้สถิติทดสอบเพียร์สันไคสแควร์และอัตราส่วนภาวะน่าจะเป็นในการทดสอบได้หรือไม่ เมื่อสถิติทดสอบทั้ง 2 ตัว มีข้อตกลงเบื้องต้นว่า ค่าความถี่ที่คาดหวังต้องมีค่ามากกว่าหรือเท่ากับ 5 นอกจากสถิติทดสอบไคสแควร์แล้ว สถิติทดสอบการปรับแก้ของเยทส์และ Fisher's exact test จะสามารถนำมาทดสอบในกรณีเช่นนี้ได้หรือไม่

จากการศึกษาของ Roscoe and Byare (1971) ที่ทำการศึกษาเกี่ยวกับความสัมพันธ์ระหว่างกลุ่มตัวอย่างกับการทดสอบไคสแควร์ โดยใช้กลุ่มตัวอย่างขนาด 10, 15, 20, 30, 50 และ 100 สำหรับตารางการถัวขนาด 2×2 ถึง 5×5 ผลการทดลองพบว่า การทดสอบไคสแควร์

มีความแกร่ง (robust) อยู่ในระดับสูง และมีการแจกแจงใกล้เคียงการแจกแจงแบบโคสแคอร์เมื่อกลุ่มตัวอย่างมีขนาดเล็ก และจากการศึกษาของวิชิตร์ เสรีอรุโณ (2531) เรื่องความแม่นยำในการใช้การทดสอบโคสแคอร์สำหรับทดสอบความเป็นเอกพันธ์ของสัดส่วนเมื่อความถี่ที่คาดหวังมีขนาดเล็กพบว่า สถิติทดสอบโคสแคอร์สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ที่ระดับนัยสำคัญ .05 เมื่อใช้ทดสอบความเป็นเอกพันธ์ของสัดส่วนประชากร 2 กลุ่มที่ความถี่ที่คาดหวังมีค่าระหว่าง 1 ถึง 5 จำนวน 1 เซลล์ และสามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ที่ระดับนัยสำคัญ .05 เมื่อใช้ทดสอบความเป็นเอกพันธ์ของสัดส่วนประชากร 3 กลุ่มที่ความถี่ที่คาดหวังมีค่าระหว่าง 1 ถึง 5 จำนวนไม่เกิน 2 เซลล์ ซึ่งสอดคล้องกับงานวิจัยของ Camilli and Hopkins (1978) ที่ว่า สำหรับการทดสอบความเป็นเอกพันธ์และการทดสอบความเป็นอิสระเมื่อค่าความถี่ที่คาดหวังน้อยกว่า 5 สถิติทดสอบเพียร์สันโคสแคอร์สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ดีกว่า การปรับแก้ของเยทส์เมื่อมีจำนวนเซลล์ที่มีค่าความถี่ที่คาดหวังค่าเป็น 1 หรือ 2 ไม่เกิน 2 เซลล์

จากการศึกษางานวิจัยที่เกี่ยวข้อง ผู้วิจัยพบว่างานวิจัยที่ผ่านมาจะทำการศึกษาเฉพาะสถิติทดสอบโคสแคอร์เท่านั้น ยังไม่มีการศึกษาเปรียบเทียบค่าความคลาดเคลื่อนประเภทที่ 1 ซึ่งถือว่ามีความสำคัญต่อการวิจัยสำหรับการทดสอบความเป็นอิสระในกลุ่มสถิติทดสอบเพียร์สันโคสแคอร์ (Pearson's chi-square statistic) อัตราส่วนภาวะน่าจะเป็น (Likelihood-ratio chi-square statistic) การปรับแก้ของเยทส์ (Yate's continuity correction) และ Fisher's exact test จากเหตุผลดังกล่าวข้างต้น ผู้วิจัยจึงมีความสนใจที่จะศึกษาว่าในการทดสอบความเป็นอิสระสำหรับข้อมูลที่จัดอยู่ในรูปตารางการถัวจร เมื่อความถี่ที่คาดหวังมีค่าน้อยกว่า 5 ไม่เกิน 25% และ 50% ของจำนวนเซลล์ทั้งหมด สถิติทดสอบที่ใช้ในการทดสอบความเป็นอิสระตัวใดที่จะสามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ตามที่กำหนด และสถิติทดสอบแต่ละตัวที่ใช้ในการศึกษามีความเหมาะสมกับข้อมูลที่อยู่ในรูปของตารางการถัวจรสถานการณ์ใด ในการวิจัยครั้งนี้กำหนดขนาดของกลุ่มตัวอย่างให้มีขนาดเล็ก กลาง ใหญ่ สำหรับตารางการถัวจรขนาดเล็ก (2×2) และ ตารางการถัวจรขนาดกลาง (3×4) โดยกำหนดกลุ่มตัวอย่างขนาด 25 50 และ 100 สำหรับตารางการถัวจรขนาด 2×2 สถิติทดสอบที่ใช้ในการศึกษาได้แก่ สถิติทดสอบเพียร์สันโคสแคอร์ อัตราส่วนภาวะน่าจะเป็น การปรับแก้ของเยทส์ และ Fisher's exact test กำหนดกลุ่มตัวอย่างขนาด 80 150 และ 300 สำหรับตารางการถัวจรขนาด 3×4 ซึ่งเป็นตารางการถัวจรขนาด $r \times c$ สถิติทดสอบที่ใช้ในการศึกษาได้แก่ เพียร์สันโคสแคอร์ และอัตราส่วนภาวะน่าจะเป็น

นอกจากการเลือกใช้สถิติทดสอบจะดูจากข้อตกลงเบื้องต้นเป็นหลักแล้วค่าความคลาดเคลื่อนประเภทที่ 1 ก็เป็นอีกสิ่งหนึ่งที่จะช่วยให้ผู้วิจัยตัดสินใจได้ว่าควรจะใช้สถิติตัวนั้น ๆ ในการ

ทดสอบหรือไม่ ในการวิจัยครั้งนี้สามารถศึกษาได้โดยวิธีการจำลอง (simulation) โดยใช้เทคนิคมอนติคาร์โล (monte carlo technique) ซึ่งเป็นวิธีที่สามารถนำมาดำเนินการเพื่อหาผลสรุปในการแก้ปัญหาได้ในปัจจุบัน นอกจากนี้แล้วการวิจัยครั้งนี้ผู้วิจัยยังได้ทำการตรวจสอบความสอดคล้องผลที่ได้จากข้อมูลจำลอง และข้อมูลจริงด้วย

วัตถุประสงค์

เปรียบเทียบค่าความคลาดเคลื่อนประเภทที่ 1 ของสถิติทดสอบ เพียร์สันไคสแควร์ อัตราส่วนภาวะน่าจะเป็น การปรับแก้ของเยทส์ และ Fisher's exact test ที่ใช้ในการทดสอบความเป็นอิสระ ในสถานการณ์ ดังนี้

1. เมื่อความถี่ที่คาดหวังแต่ละเซลล์น้อยกว่า 5 หรือ $E_{ij} < 5$ ไม่เกิน 25% ของจำนวนเซลล์ทั้งหมด
2. เมื่อความถี่ที่คาดหวังแต่ละเซลล์น้อยกว่า 5 หรือ $E_{ij} < 5$ ไม่เกิน 50% ของจำนวนเซลล์ทั้งหมด

ขอบเขตของการวิจัย

1. การวิจัยครั้งนี้ได้ศึกษาตัวแปรต่อไปนี้
 - 1.1 ตัวแปรตาม (dependent variable) คือ อัตราความคลาดเคลื่อนประเภทที่ 1 ของสถิติทดสอบความเป็นอิสระ
 - 1.2 ตัวแปรอิสระ (independent variable) คือ สถิติทดสอบความเป็นอิสระ ได้แก่ เพียร์สันไคสแควร์ อัตราส่วนภาวะน่าจะเป็น การปรับแก้ของเยทส์ และ Fisher's exact test
2. อัตราความคลาดเคลื่อนประเภทที่ 1 จากสถิติทดสอบความเป็นอิสระของการวิจัยนี้ พิจารณาที่สมมติฐาน $H_0 : p_{ij} = p_i \cdot p_j$ เป็นจริง
3. การวิจัยครั้งนี้ศึกษาจากประชากรที่มีลักษณะการแจกแจงแบบพหุนาม (multinomial distribution)
4. การวิจัยครั้งนี้ศึกษาจากกลุ่มตัวอย่างและขนาดของตารางการถัวจร โดยกำหนดกลุ่มตัวอย่างที่ศึกษามีขนาด 25 50 และ 100 สำหรับตารางการถัวจรขนาด 2×2 และกลุ่มตัวอย่างขนาด 80 150 และ 300 สำหรับตารางการถัวจรขนาด 3×4

5. ในการทดสอบความเป็นอิสระของตารางการถัวขนาด 2×2 ใช้สถิติในการทดสอบ 4 ตัว ได้แก่ เพียร์สันไคสแควร์ อัตราส่วนภาวะน่าจะเป็น การปรับแก้ของเยทส์ และ Fisher's exact test สำหรับตารางการถัวขนาด 3×4 ใช้สถิติในการทดสอบ 2 ตัว ได้แก่ เพียร์สันไคสแควร์ และอัตราส่วนภาวะน่าจะเป็น
6. การวิจัยครั้งนี้ศึกษาเฉพาะค่าความถี่ที่คาดหวังของเซลล์ที่มีค่าน้อยกว่า 5 ไม่เกิน 25% และ 50% ของจำนวนเซลล์ทั้งหมด
7. การวิจัยครั้งนี้กำหนดระดับนัยสำคัญ 2 ระดับ คือ ณ ระดับนัยสำคัญ .05 และ .01
8. แต่ละกรณีของการวิจัยครั้งนี้จะเรียกชุดเลขสุ่มซ้ำจำนวน 10,000 ครั้ง ทำการจำลองข้อมูล โดยอาศัยเทคนิคมอนติคาร์โล (monte carlo simulation technique)

นิยามศัพท์ที่ใช้เฉพาะในการวิจัย

ตัวแปรจำแนกประเภท หมายถึง ตัวแปรที่บอกคุณลักษณะแตกต่างกันเป็นประเภท ๆ ซึ่งได้จากการวัดระดับนามบัญญัติและจัดอันดับ

ระดับนัยสำคัญ หมายถึง โอกาสที่ยอมให้เกิดความคลาดเคลื่อนประเภทที่ 1 แทนด้วยสัญลักษณ์ α

ความคลาดเคลื่อนประเภทที่ 1 หมายถึง ความผิดพลาดที่เกิดจากการปฏิเสธสมมติฐานศูนย์เมื่อสมมติฐานศูนย์นั้นเป็นจริง

อัตราความคลาดเคลื่อนประเภทที่ 1 หมายถึง จำนวนครั้งของการปฏิเสธสมมติฐานหลักหารด้วยจำนวนครั้งที่เกิดค่าความถี่ที่คาดหวังที่เป็นไปตามสถานการณ์ที่กำหนดในการทดสอบสมมติฐานที่เกิดขึ้นจริงในการจำลองข้อมูล

อัตราความคลาดเคลื่อนที่ระบุ หมายถึง อัตราความคลาดเคลื่อนประเภทที่ 1 ที่ควบคุมด้วยระดับนัยสำคัญ α ซึ่งผู้วิจัยเป็นผู้กำหนด

ความถี่ที่สังเกตได้ หมายถึง ค่าของข้อมูลที่เก็บรวบรวมได้ในแถวอนที่ i และแถวตั้งที่ j

ความถี่ที่คาดหวัง หมายถึง ค่าของข้อมูลที่ควรจะเป็นตามสมมติฐาน หรือ ทฤษฎีที่ตั้งไว้

$E_{ij} < 5$ ไม่เกิน 25% ของจำนวนเซลล์ทั้งหมด หมายถึง ค่าความถี่ที่คาดหวังที่น้อยกว่า 5 จำนวน 1 เซลล์ในตารางการถัวขนาด 2×2 และจำนวนไม่เกิน 3 เซลล์ในตารางการถัวขนาด 3×4

$E_{ij} < 5$ ไม่เกิน 50% ของจำนวนเซลล์ทั้งหมด หมายถึง ค่าความถี่ที่คาดหวังที่น้อยกว่า 5 จำนวน 2 เซลล์ในตารางการถัวขนาด 2×2 และจำนวน 4 ถึง 6 เซลล์ในตารางการถัวขนาด 3×4

ค่าสัดส่วนส่วนรวมของแถวอน หมายถึง ผลรวมของค่าสัดส่วนในแถวตั้งที่ j ของแต่ละแถวอนที่ i

ค่าสัดส่วนส่วนรวมของแถวตั้ง หมายถึง ผลรวมของค่าสัดส่วนในแถวอนที่ i ของแต่ละแถวตั้งที่ j

ประโยชน์ที่คาดว่าจะได้รับ

การวิจัยครั้งนี้ให้ประโยชน์แก่นักวิจัยได้มีผลสรุปและหลักฐานในการใช้สถิติทดสอบความเป็นอิสระและมีเกณฑ์ในการตัดสินใจว่าเมื่อค่าความถี่ที่คาดหวังน้อยกว่า 5 ไม่เกิน 25% และ 50% ของจำนวนเซลล์ทั้งหมดแล้ว ควรจะเลือกใช้สถิติทดสอบความเป็นอิสระตัวใดที่จะทำได้ผลของการตัดสินใจมีประสิทธิภาพและเหมาะสมกับสถานการณ์

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย