หอสมุดกลาง จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

การวิจัยเชิงทดลอง เป็นการวิจัยที่มีจุดมุ่งหมายหลักเพื่อศึกษาหาความสัมพันธ์ในเชิงเหตุและผล (Cause and Effect Relationship) ระหว่างตัวแปรอิสระ หรือในการทดลองเรียกว่า ตัวแปรทดลอง กับตัวแปรตามซึ่งเป็นผลการทดลอง โดยผู้วิจัยเป็นผู้จัดกระทำแปรเปลี่ยนค่าของตัวแปรอิสระหรือตัวแปรทดลองให้กับพลวิจัย (Subject) ในกลุ่มทดลอง แล้ววัดค่าของตัวแปรตามหรือผลของการทดลองว่ามีการเปลี่ยนแปลงอย่างไร ในปริมาณเท่าใด ซึ่งผลการวิจัยมักนำไปเปรียบเทียบกับกลุ่มควบคุม ซึ่งได้รับการจัดกระทำแตกต่างจากกลุ่มทดลอง แต่ทั้งนี้ผู้วิจัยจะต้องควบคุมสภาพของการทดลองให้ปราศจากอิทธิพลของตัวแปรแทรกซ้อนที่อาจจะส่งผลต่อตัวแปรตามที่ต้องการศึกษา เพื่อให้สามารถสรุปผลการทดลองได้ด้วยความมั่นใจว่า ผลการทดลองนั้นสืบเนื่องมาจากตัวแปรอิสระหรือตัวแปรทดลองและผลนี้ ไม่ได้เกิดจากตัวแปรแทรกซ้อนอื่น ซึ่งระดับของการควบคุมตัวแปรแทรกซ้อนในการวิจัยเชิงทดลอง พอแบ่งได้เป็น 2 ระดับ (อุทุมพร จามรมาน 2527: 118) คือการวิจัยในห้องปฏิบัติการ (Laboratory Experiment) และการวิจัยกึ่งทดลอง (Quasi-Experiment) สำหรับการวิจัยเชิงทดลองในห้องปฏิบัติการนั้น ผู้วิจัยสามารถควบคุมระดับของตัวแปรแทรกซ้อนที่อาจส่งผลต่อตัวแปรตามได้มากที่สุด ส่วนการวิจัยกึ่งทดลอง เป็นการทดลองในสถานการณ์ธรรมชาติตามสภาพที่เป็นจริง ระดับการควบคุมตัวแปรแทรกซ้อนทำได้จำกัดหรือทำได้เป็นบางส่วน

การวิจัยเชิงทดลองใช้กันมากในวิทยาศาสตร์แขนงต่าง ๆ ในงานวิจัยทางด้านจิตวิทยาก็นำมาใช้กันอย่างแพร่หลายเช่นกัน โดยเฉพาะการวิจัยทางพฤติกรรมศาสตร์ ซึ่งผู้วิจัยต้องการศึกษาหาความสัมพันธ์ในเชิงเหตุและผลของพฤติกรรมที่เกิดขึ้น เพื่อจะอธิบายว่าสภาพการณ์ใดหรือพฤติกรรมใดบ้างที่ทำให้เกิดผลลัพธ์เช่นนั้น และในทางกลับกันก็อาจทดลองหาความสัมพันธ์ว่าถ้าเราให้สภาพการณ์นั้น ๆ แล้วจะเกิดผลอะไร เท่าใดและอย่างไรบ้าง เนื่องจากในการจัดพลวิจัยเข้ากลุ่มทดลองและกลุ่มควบคุม ผู้วิจัยจะใช้หลักการสุ่ม (Randomization) คือให้

สมาชิกทุกตัวจากกลุ่มประชากรมีโอกาสเท่ากันที่จะถูกคัดเลือกมาเป็นกลุ่มตัวอย่าง ในรูปแบบงานวิจัยบางแบบ สมาชิกในกลุ่มตัวอย่างก็เลือกอย่างสุ่มเข้ากลุ่มทดลองและกลุ่มควบคุมในลักษณะ Matched pair (Winer 1971: 517) ซึ่ง เคอร์ลิงเจอร์ (Kerlinger 1964: 303-304) กล่าวว่าวิธีการ Randomization จะทำให้กลุ่มทดลองและกลุ่มควบคุมมีสภาพการณ์ทั่ว ๆ ไปในตอนเริ่มต้นก่อนที่จะได้รับทรีทเมนท์ (Treatment) ไม่แตกต่างกัน กลุ่มตัวอย่างที่ทำการศึกษาก็ใช้บุคคลหรือสัตว์เป็นตัวรับการทดลอง วิธีการศึกษาก็อาจทำได้ทั้งแบบการศึกษาระยะสั้นและแบบการศึกษาระยะยาว สำหรับการศึกษาระยะยาว (Longitudinal Study) นั้น ใช้เวลาทดลองศึกษากับกลุ่มตัวอย่างเป็นระยะเวลานาน เพื่อให้สามารถสรุปผลการวิจัยที่แน่ชัดและแม่นยำว่าผลที่เกิดขึ้นนั้น เนื่องมาจากตัวแปรอิสระหรือตัวแปรทดลองหรือไม่เพียงใด เนื่องจากการทดลองใช้เวลานาน ดังนั้นผู้ที่ได้รับการคัดเลือกเข้าศึกษาอาจไม่ได้รับทรีทเมนท์ครบสมบูรณ์ตามที่ผู้วิจัยวางแผนไว้ นั่นคือมีข้อมูลบางส่วนหายไป (Missing Data)

การขาดหายไปของข้อมูลหรือการสูญหายไปของข้อมูล เกิดจากการที่ผู้วิจัยไม่สามารถเก็บรวบรวมข้อมูลของพลวิจัยได้ครบถ้วนเมื่อทำการศึกษาเสร็จสิ้นแล้ว ข้อมูลที่สูญหายอาจเกิดขึ้นได้ในหลายกรณีต่อไปนี้ เช่น พลวิจัยเกิดอาการเจ็บป่วยหรือตายในระหว่างที่ผู้วิจัยทำการศึกษา หรือเกิดจากการไม่ตอบสนองของพลวิจัยต่อการวัดหรือการเก็บรวบรวมข้อมูล หรือเกิดจากการย้ายสถานที่อยู่ไม่สามารถติดตามผลได้ การสูญหายของข้อมูลมักเกิดขึ้นเสมอและมีโอกาสไม่มากนักที่ผู้วิจัยจะเก็บรวบรวมข้อมูลได้ครบถ้วน โดยเฉพาะอย่างยิ่งการวิจัยสนาม (field research) ซึ่งเป็นการวิจัยแบบกึ่งทดลองที่ผู้วิจัยเข้าไปทดลองศึกษาในสภาพการณ์จริง ๆ การควบคุมสภาพการณ์ต่าง ๆ กระทำได้ไม่มากนัก จึงจัดเป็นปัญหาอันดับแรกของความคลาดเคลื่อนในการเก็บรวบรวมข้อมูล (Zarkovich 1966: 8)

สำหรับการวิจัยเชิงทดลองผู้วิจัยมักจะกำหนดจำนวนพลวิจัยไว้ให้เท่ากันในแต่ละระดับของการทดลอง ทั้งนี้เนื่องจากใช้หลักวิธี Random assignment และ Matched pair เมื่อมีการสูญหายของพลวิจัยขณะทำการทดลองก็จะทำให้กลุ่มตัวอย่างมีขนาดไม่เท่ากัน ซึ่งถ้าไม่ได้วางแผนการวิเคราะห์ข้อมูลในกรณีนี้ไว้ด้วยแล้ว จะเป็นข้อเสียหายอย่างรุนแรงต่องานวิจัย เพราะจะไปเปลี่ยนแปลงรูปแบบการคำนวณให้แตกต่างไปจากแผนการวิเคราะห์ข้อมูลที่กำหนดไว้ และจะมีผลกระทบอย่างมากต่อการฝ่าฝืนข้อตกลงสำหรับผลรวมกำลังสองในการวิเคราะห์ความแปรปรวน (Keppel 1982: 99-100) ทุก ๆ หน่วยตัวอย่างในการทดลองทั้งข้อมูลที่สูญหาย

กับที่ยังคงอยู่ มีความสำคัญไม่น้อยกว่ากัน การสูญหายของข้อมูลจึงมีผลทำให้ค่าความแปรปรวนเปลี่ยนแปลงไปจากเดิม และยิ่งถ้ากลุ่มตัวอย่างที่มีขนาดเล็กอยู่แล้วค่าของข้อมูลที่สังเกตได้แต่ละค่าจะมีผลต่อค่าความแปรปรวนอย่างเห็นได้ชัด เมื่อมีข้อมูลสูญหายไปจากกลุ่มนี้อีกจะทำให้กลุ่มตัวอย่างเล็กลงยิ่งขึ้น ความถูกต้องแม่นยำในการวิเคราะห์ข้อมูลย่อมลดน้อยลง (Morrison 1976: 120)

และประการสำคัญอย่างมาก คือ ปัญหาของอำนาจของการทดสอบ จะลดน้อยลงเมื่อการทดลองนั้นมีข้อมูลที่ลดลงกว่าเดิม

การสูญหายของข้อมูลจากกลุ่มตัวอย่างโดยเฉพาะเมื่อกลุ่มตัวอย่างมีขนาดเล็กแล้วจึงเป็นเรื่องที่ผู้วิจัยละเลยไม่ได้ จำเป็นต้องจัดกระทำเพื่อแก้ปัญหาการสูญหายของข้อมูลที่เกิดขึ้นนี้ ซึ่ง ลี (Lee 1975: 276-277) ได้เสนอวิธีแก้ปัญหาไว้ 4 ประการดังต่อไปนี้

1. กำหนดให้มีการเลือกพลวิจัยไว้เกินสักสองหรือสามหน่วยในแต่ละกลุ่มของการวิจัยนั้น ๆ เมื่อมีการสูญหายของข้อมูลเกิดขึ้นก็สามารถนำข้อมูลจากพลวิจัยที่เลือกเกินไว้แทนเข้าไปซึ่งก็ต้องมีข้อตกลงไว้ก่อนว่า ข้อมูลของพลวิจัยที่แทนเข้าไปมีคุณสมบัติไม่แตกต่างจากข้อมูลของพลวิจัยที่สูญหายไป

2. กำจัดหรือตัดข้อมูลในกลุ่มที่มากกว่า เพื่อให้เกิดรูปแบบกลุ่มตัวอย่างที่เท่ากัน (Balance Design) อาจใช้การจับฉลากข้อมูลในกลุ่มตัวอย่างที่มากกว่าออกให้เหลือจำนวนเท่ากับกลุ่มที่มีการสูญหายของข้อมูล หรือจัดกระทำด้วยวิธีอื่นเพื่อลดข้อมูลให้เหลือเท่ากัน ซึ่งวิธีการนี้ ทาแบชนิค (Tabachnick 1983: 70) กล่าวว่าผู้วิจัยส่วนใหญ่ไม่นิยมปฏิบัติกันเนื่องจากจะขาดความเชื่อมั่นในผลการวิเคราะห์ข้อมูล

3. วิเคราะห์ข้อมูลโดยไม่ใส่ใจส่วนที่สูญหาย คือการที่ผู้วิจัยนำข้อมูลที่เก็บรวบรวมมาได้ แล้ววิเคราะห์ผลเลย โดยไม่คำนึงถึงข้อมูลส่วนที่ขาดหายไป ซึ่งผลของการวิเคราะห์ข้อมูลโดยวิธีนี้จะทำให้เกิดความไม่แน่ใจแก่ผู้วิจัยและผู้ที่จะนำผลงานวิจัยไปใช้ เมื่อมีการสูญหายของข้อมูลเกิดขึ้นหลายจำนวนและเป็นการแสดงว่าผู้วิจัยไม่ได้ให้ความสำคัญกับข้อมูลที่สูญหายไปเลย ทั้งที่ข้อมูลที่สูญหายนั้นอาจมีความสำคัญเท่ากันหรือมากกว่าข้อมูลที่ยังคงอยู่เสียอีก (Babbie 1986: 372-373)

4. ประมาณค่าแทนข้อมูลที่สูญหาย โดยผู้วิจัยจะกำหนดค่าแทนข้อมูลที่สูญหายด้วยวิธีการที่เหมาะสม เพื่อให้ได้ข้อมูลครบถ้วนในการวิเคราะห์ข้อมูล ซึ่งประโยชน์ของการประมาณค่าแทนข้อมูลที่สูญหายนั้น ได้มีผู้ศึกษาและเสนอแนะข้อความรู้ที่สำคัญไว้ดังต่อไปนี้

คอคแครน และ คอกซ์ (Cochran and Cox 1957: 80-82) กล่าวว่า เมื่อสิ่งที่สังเกตขาดหายไป ในการที่จะกระทำผลลัพธ์ให้ถูกต้องผู้วิจัยสามารถเลือกใช้รูปแบบสมการทางคณิตศาสตร์ หรือวิธีการอื่นคำนวณค่าขึ้นแทนข้อมูลที่สูญหายได้ โดยใช้ข้อมูลที่มีอยู่ทั้งหมดเป็นหลักในการคำนวณ

ซาร์โควิช (Zarkovich 1966: 148) ได้ศึกษาพบว่าถ้าใช้วิธีการประมาณค่าแทนข้อมูลที่สูญหายไปสามารถป้องกันการลดลงของความแม่นตรง (precision) เนื่องจากความแม่นตรงจะลดลงเมื่อขนาดของกลุ่มตัวอย่างเล็กลง และการแทนค่าข้อมูลที่สูญหายจะมีผลต่อค่าของความแปรปรวนมากกว่าการตัดข้อมูลที่สูญหายออกจากการวิเคราะห์ข้อมูล กล่าวคือจะทำให้ได้ค่าความแปรปรวนที่มีความถูกต้องแม่นยำสูงขึ้น

กลีสันและสเตลิน (Gleason and Staelin 1975: 230-231) กล่าวว่า กรณีที่มีข้อมูลสูญหาย ถ้าใช้วิธีประมาณค่าแทนข้อมูลที่สูญหายจะทำให้ได้รูปแบบของผลลัพธ์ที่ไม่กำกวม คือการได้มาของข้อมูลจากวิธีประมาณค่ามีความชัดเจนในกระบวนการดำเนินการ สามารถนำไปวิเคราะห์ประเมินค่าได้และอำนาจการวิเคราะห์โดยวิธีนี้จะทำให้คุณสมบัติของข้อตกลงเบื้องต้นเกี่ยวกับการแจกแจงของประชากรมีความถูกต้อง

มอร์ริสัน (Morrison 1976: 120) กล่าวว่า เมื่อมีข้อมูลในบางวาระตกหล่นหรือสูญหายนอกจากจะทำให้ขนาดของกลุ่มตัวอย่างเล็กลงแล้ว ยังจะส่งผลไปถึงความถูกต้องแม่นยำในการคำนวณด้วย และผู้วิจัยก็ไม่สามารถวิเคราะห์ข้อมูลตามรูปแบบที่ต้องการได้ ปัญหานี้ผู้วิจัยสามารถใช้วิธีประมาณค่าแทนข้อมูลที่สูญหาย เพื่อที่จะทดสอบสมมุติฐานในเรื่องนั้นและสามารถวิเคราะห์ข้อมูลตามรูปแบบที่กำหนดไว้ได้ ทั้งจะทำให้ได้ค่าทดสอบสูงสุด (maximal) ซึ่งก็สอดคล้องกับคำกล่าวของ มาร์ซูโล (Marascuilo 1983: 65-66) ที่ว่าถ้าผู้วิจัยต้องการให้ได้ข้อมูลครบถ้วนเมื่อมีข้อมูลสูญหาย ในทางปฏิบัติผู้วิจัยสามารถใช้วิธีประมาณค่าขึ้นแทนข้อมูลที่สูญหายได้

แบบบี (Babbie 1986: 372-373) ได้ศึกษาพบว่า การประมาณค่าแทนข้อมูลที่สูญหายจะอนุรักษ์ผลการวิเคราะห์ข้อมูล (conservative) ได้มากกว่าการตัดข้อมูลในส่วนที่สมบูรณ์ ทั้งเพื่อให้เหลือข้อมูลในแต่ละกลุ่มการทดลองเท่ากัน หรือวิเคราะห์ข้อมูลไปโดยไม่ใส่ใจข้อมูลส่วนที่สูญหาย

จากเหตุผลและข้อเสนอแนะดังกล่าวสามารถสรุปได้ว่า การประมาณค่าแทนข้อมูลที่สูญหาย เป็นวิธีการแก้ปัญหาการสูญหายของข้อมูลที่เป็นไปได้ และเหมาะสมกว่าอีกทั้ง 3 วิธี คือ การเลือกพลวิจัยไว้เกิน การตัดข้อมูลส่วนที่มากกว่าในกลุ่มอื่น ๆ ออก และการวิเคราะห์ข้อมูลโดยไม่ใส่ใจข้อมูลส่วนที่สูญหาย ซึ่งวิธีการประมาณค่าแทนข้อมูลที่สูญหายที่ใช้กันอยู่ในปัจจุบันมีอยู่ด้วยกันหลายวิธี ผู้วิจัยขอนำเสนอเฉพาะวิธีที่นิยมใช้กัน สะดวกต่อการปฏิบัติและมีงานวิจัยสนับสนุนดังต่อไปนี้ คือ (1) วิธีประมาณค่าข้อมูลที่สูญหายโดยใช้ค่าเฉลี่ยจากกลุ่มตัวอย่าง (2) วิธีประมาณค่าข้อมูลที่สูญหายโดยใช้สมการถดถอย และ (3) วิธีประมาณค่าข้อมูลที่สูญหายโดยใช้ค่าเฉลี่ยระหว่างค่าเฉลี่ยจากกลุ่มตัวอย่างและสมการถดถอย

วิธีประมาณค่าข้อมูลที่สูญหายโดยใช้ค่าเฉลี่ยจากกลุ่มตัวอย่าง เคปเพล (Keppel 1982: 100-101) กล่าวว่าเป็นวิธีที่สะดวกและง่ายต่อการปฏิบัติ มีผู้นิยมใช้มากที่สุดเพราะไม่มีขั้นตอนยุ่งยากซับซ้อนในการคำนวณ

ทาแบชนิค (Tabachnick 1983: 71-72) ได้ทำการศึกษาและเสนอว่า วิธีใช้ค่าเฉลี่ยจากกลุ่มตัวอย่างจะเป็นค่าประมาณที่ดีเกี่ยวกับข้อมูลที่สูญหายไปและโดยปกติผู้วิจัยมักแทนข้อมูลที่สูญหายด้วยค่าเฉลี่ยนั้นเป็นวิธีปฏิบัติในลักษณะอนุรักษ์นิยม (Conservative) เพราะค่าเฉลี่ยนี้จะมีการแจกแจงเหมือนข้อมูลส่วนที่เหลืออยู่ทั้งหมด ซึ่งผู้วิจัยก็ไม่ต้องไปคาดคะเนการแจกแจงของส่วนที่หายไป

อะฟิฟีและอิลาสฮอฟฟ์ (Afifi and Elashoff 1966: 599-600) ได้ศึกษาวิธีประมาณค่าข้อมูลที่สูญหายด้วยสมการถดถอยและใช้ชื่อว่า The First Order Regression ได้เสนอแนะว่า ถ้ากลุ่มตัวอย่างมีขนาดเล็กและจำนวนข้อมูลที่สูญหายมีไม่มากนัก เมื่อเปรียบเทียบกับข้อมูลที่ยังคงเหลืออยู่ที่มีจำนวนมากกว่าหรือใกล้เคียงกับข้อมูลที่สูญหาย การคำนวณค่าแทนข้อมูลที่สูญหายด้วยวิธีการ The First Order Regression จะเป็นที่ยอมรับและนิยมใช้กันอย่างแพร่หลาย ซึ่งอะฟิฟีและอิลาสฮอฟฟ์ (1967: 16) ได้ศึกษาต่อพบว่า วิธีประมาณค่าข้อมูลที่สูญหายโดยใช้ค่าเฉลี่ยระหว่างค่าเฉลี่ยจากกลุ่มตัวอย่างและสมการถดถอยนั้น จะมีค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Mean Square Error) น้อยกว่าวิธีใช้สมการถดถอยและวิธีใช้ค่าเฉลี่ยจากกลุ่มตัวอย่าง

หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

มาร์ชูไล (Marascuilo 1983: 65) ได้ให้แนวความคิดเกี่ยวกับวิธีการประมาณค่าคะแนนที่สูญหายไว้หลายวิธีคือ วิธีหาค่าเฉลี่ยจากพลวิจัยอื่นที่มีคุณสมบัติคล้ายคลึงกัน วิธีหาค่าเฉลี่ยจากข้อมูลที่มีอยู่ทั้งหมดและวิธีใช้สมการถดถอย พบว่าแต่ละวิธีประมาณค่าจะประมาณค่าของคะแนนที่สูญหายได้แตกต่างกัน ดังนั้นจึงจำเป็นต้องพิจารณาเลือกใช้ให้เหมาะสม

จากผลการศึกษาและข้อเสนอแนะดังกล่าวพบว่า ยังไม่สามารถหาหลักฐานและข้อสรุปที่ชัดเจนเกี่ยวกับการเลือกใช้วิธีประมาณค่าข้อมูลทั้งสามวิธี เมื่อมีข้อมูลสูญหายไปจากกลุ่มตัวอย่างโดยเฉพาะอย่างยิ่งเมื่อกลุ่มตัวอย่างนั้นมีขนาดเล็ก และทีค่าสัมประสิทธิ์สหสัมพันธ์ของข้อมูลในระดับที่แตกต่างกันแล้ว วิธีใดจะสามารถประมาณค่าข้อมูลได้ถูกต้องแม่นยำ (Accuracy) มากกว่ากัน โดยศึกษาด้วยการซิมูเลท (Simulate) ซึ่งจะทำให้ได้ผลสรุปที่เด่นชัดภายใต้สภาวะของสถานการณ์ที่คล้ายกับการทดลอง คือสามารถที่จะระบุหรือจำกัดขนาดของกลุ่มตัวอย่าง ค่าเฉลี่ย ความแปรปรวน และลักษณะการแจกแจงของประชากรได้ด้วยเครื่องคอมพิวเตอร์ โดยใช้เทคนิควิธีที่เรียกว่ามอนติคาร์โล (Monte Carlo Method) ซึ่งเป็นวิธีที่สามารถดำเนินการได้ในปัจจุบัน

วัตถุประสงค์ของการวิจัย

วัตถุประสงค์ของการวิจัยครั้งนี้เพื่อเปรียบเทียบความแม่นยำ (Accuracy) ของวิธีประมาณค่าข้อมูลที่สูญหายสามวิธีคือ

1. วิธีประมาณค่าข้อมูลที่สูญหายโดยใช้ค่าเฉลี่ยจากกลุ่มตัวอย่าง
2. วิธีประมาณค่าข้อมูลที่สูญหายโดยใช้สมการถดถอย
3. วิธีประมาณค่าข้อมูลที่สูญหายโดยใช้ค่าเฉลี่ยระหว่างค่าเฉลี่ยจากกลุ่มตัวอย่างและสมการถดถอย

ข้อตกลงเบื้องต้น

1. การวิจัยครั้งนี้ถือว่า ค่าเฉลี่ยของผลต่างระหว่างค่าของข้อมูลที่สูญหายกับค่าที่ประมาณได้จากวิธีประมาณค่า และค่าเฉลี่ยกำลังสองของผลต่างระหว่างค่าของข้อมูลที่สูญหายกับค่าที่ประมาณได้จากวิธีประมาณค่าเป็นดัชนีที่สำคัญในการเปรียบเทียบความแม่นยำของวิธีประมาณค่าข้อมูล

2. การวิจัยนี้เป็นการวิจัยที่ใช้เครื่องคอมพิวเตอร์ในการซิมูเลท (Simulate) ซึ่งผู้วิจัยได้เลือกใช้โปรแกรมย่อยสับรูทีน (Subroutine) ที่พัฒนาขึ้นมาจนเป็นที่ยอมรับในการสร้างข้อมูลตามลักษณะประชากรที่ต้องการวิจัย

3. การวิจัยครั้งนี้ใช้ข้อมูลในลักษณะที่มีการแจกแจงแบบปกติสองตัวแปร (Bivariate Normal Distribution)

ขอบเขตของการวิจัย

1. การวิจัยครั้งนี้ศึกษาตัวแปรดังต่อไปนี้

1.1 ตัวแปรตาม (Dependent Variable) คือ

1.1.1 ค่าเฉลี่ยของผลต่างระหว่างค่าของข้อมูลที่สูญหายกับค่าที่ประมาณได้จากวิธีประมาณค่า

1.1.2 ค่าเฉลี่ยกำลังสองของผลต่างระหว่างค่าของข้อมูลที่สูญหายกับค่าที่ประมาณได้จากวิธีประมาณค่า

1.2 ตัวแปรอิสระ (Independent Variable) คือ

1.2.1 วิธีประมาณค่าข้อมูลที่สูญหายสามวิธี คือ

1.2.1.1 วิธีใช้ค่าเฉลี่ยจากกลุ่มตัวอย่าง

1.2.1.2 วิธีใช้สมการถดถอย

1.2.1.3 วิธีใช้ค่าเฉลี่ยระหว่างค่าเฉลี่ยจากกลุ่มตัวอย่างและสมการถดถอย

1.2.2 ขนาดของกลุ่มตัวอย่างมี 3 ระดับ คือ 5, 10 และ 15

1.2.3 ตำแหน่งของข้อมูลที่สูญหาย แบ่งออกเป็น 3 เขต คือ

1.2.3.1 เขตที่ 1 ค่าของข้อมูลที่สูญหายมากกว่า $\bar{X}$ + 1 S.D.

1.2.3.2 เขตที่ 2 ค่าของข้อมูลที่สูญหายตั้งแต่ $\bar{X}$ - 1 S.D. ถึง $\bar{X}$ + 1 S.D.

1.2.3.3 เขตที่ 3 ค่าของข้อมูลที่สูญหายน้อยกว่า $\bar{X}$ - 1 S.D.

1.2.4 จำนวนข้อมูลที่สูญหายจากกลุ่มตัวอย่างมี 2 ระดับ คือ หายไปครั้งละ 1 ค่า และหายไปครั้งละ 2 ค่าจากกลุ่มตัวอย่าง

2. กำหนดค่าพารามิเตอร์ $\mu = 500$ และ $\sigma^2 = 100$

3. ศึกษาเฉพาะค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรควบคุมกับตัวแปรตามของประชากร (ρ) ที่ระดับ 0.2, 0.4 และ 0.6

4. ศึกษาเฉพาะขนาดของกลุ่มตัวอย่างที่ระดับ 5, 10 และ 15

5. ในแต่ละกรณีของการวิจัยนี้ จะทำการทดลองซ้ำเป็นจำนวน 4,000 ครั้ง

<u>คำจำกัดความ</u>

ข้อมูลที่สูญหาย (Missing Data) หมายถึง การที่พลวิจัยหน่วยนั้นไม่มีข้อมูลของการได้รับการทดลองที่ครบถ้วน โดยที่การวิจัยนี้จะกำหนดข้อมูลให้ครบแล้วสุ่มให้หายไป

วิธีประมาณค่าข้อมูลที่สูญหายสามวิธี หมายถึง

1. วิธีประมาณค่าข้อมูลที่สูญหายโดยใช้ค่าเฉลี่ยจากกลุ่มตัวอย่าง (Estimated by Sampling mean)

เป็นวิธีประมาณค่าของข้อมูลที่สูญหาย (X) ด้วยค่าเฉลี่ยจากกลุ่มตัวอย่างที่มีอยู่แทนด้วย $\bar{X}_{n_x}$

โดย $$\bar{X}_{n_x} = \sum_{i=1}^{n_x} X_i / N_x$$

เมื่อ $\bar{X}_{n_x}$ คือค่าเฉลี่ยจากกลุ่มตัวอย่างที่มีค่าของ X ทั้งหมด

N_x คือ จำนวนหน่วยตัวอย่างที่มีค่าของ X ทั้งหมด

2. วิธีประมาณค่าข้อมูลที่สูญหายโดยใช้สมการถดถอย (Estimated by Regression Equation)

เป็นวิธีประมาณค่าที่คำนวณการถดถอยของ X ในเทอม Y ที่มีความสัมพันธ์กัน จากข้อมูลที่มีทั้งค่าของตัวแปรตาม (X) และตัวแปรควบคุม (Y) โดยวิธีกำลังสองน้อยที่สุด และประมาณค่า X ที่สูญหายไปโดยใช้สมการถดถอยที่ได้นี้

คือ $X_i^{LS} = d_o^{LS} + d_1^{LS} Y_i$

โดยที่ $d_1^{LS} = \dfrac{\sum_{i=1}^{n_c} X_i Y_i - n_c \bar{Y}_c \bar{X}_c}{\sum_{i=1}^{n_c} Y^2 - n_c \bar{Y}_c^2}$; $\bar{X}_c = \sum_{i=1}^{n_c} X_i/n_c$

$d_o^{LS} = \bar{X}_c - d_1^{LS} \bar{Y}_c$; $\bar{Y}_c = \sum_{i=1}^{n_c} Y_i/n_c$

และ n_c คือ จำนวนหน่วยตัวอย่างที่มีทั้งค่า X และค่า Y

3. วิธีประมาณค่าข้อมูลที่สูญหายโดยใช้ค่าเฉลี่ยระหว่างค่าเฉลี่ยจากกลุ่มตัวอย่างและสมการถดถอย (Estimated by mean of Sampling mean and Regression Equation)

เป็นวิธีประมาณค่าข้อมูลที่สูญหายโดยใช้ค่าเฉลี่ยระหว่างค่าเฉลี่ยจากกลุ่มตัวอย่างและค่าที่ได้จากสมการถดถอย มีขั้นตอนดำเนินการดังต่อไปนี้

ก. หาค่าเฉลี่ยของข้อมูลที่มีอยู่จากกลุ่มตัวอย่าง จะได้ค่า $\bar{X}_{n_x}$

ข. แทนข้อมูล X_i ที่สูญหายด้วย $\bar{X}_{n_x}$ แล้วนำข้อมูลที่ได้ทั้งหมดคำนวณการถดถอยของ X ในเทอม Y โดยวิธีกำลังสองน้อยที่สุด จะได้สมการถดถอย คือ

$$X_i^{LS} = d_o^{LS} + d_1^{LS} Y_i$$

ค. คำนวณหาค่า X_i ที่สูญหายโดยใช้สมการถดถอยที่ได้จากข้อ ข. จะได้ค่า $\hat{X}_i^{LS}$

ง. หาค่าของ $(\bar{X}_{n_x} + \hat{X}_i^{LS})/2$ ซึ่งเป็นค่าที่ประมาณได้จากวิธีประมาณค่านี้

ความแม่นยำ (Accuracy) หมายถึง ความแตกต่างของค่าประมาณกับค่าของข้อมูลที่สูญหาย โดยใช้เกณฑ์ตัดสินความแม่นยำที่มากที่สุดของวิธีประมาณค่าทั้ง 3 วิธี ดังต่อไปนี้

1. มีค่าเฉลี่ยของผลต่างระหว่างค่าของข้อมูลที่สูญหาย กับค่าที่ประมาณได้จากวิธีประมาณค่ามีค่าน้อย (เข้าใกล้ศูนย์) และ

2. มีค่าเฉลี่ยกำลังสองของผลต่างระหว่างค่าของข้อมูลที่สูญหายกับค่าที่ประมาณได้จากวิธีประมาณค่าน้อยที่สุด และวิธีที่ให้ค่าเฉลี่ยกำลังสองของผลต่างน้อยที่สุดจะเป็นวิธีที่ดีที่สุดใน 3 วิธีนี้

ค่าเฉลี่ยของผลต่างระหว่างค่าของข้อมูลที่สูญหายกับค่าที่ประมาณได้จากวิธีประมาณค่าหรือค่าเฉลี่ยของผลต่าง (Mean of Difference: MD) หมายถึง ค่าที่ได้จาก

$$\sum_{j=1}^{N} (X_j^M - X_j^{(J)})/N$$

เมื่อ $j = 1, 2, 3 \ldots, N$

X_j^M คือ ค่าของข้อมูลที่สูญหายครั้งที่ j

$X_j^{(J)}$ คือ ค่าของข้อมูลที่ประมาณได้ครั้งที่ j จากวิธีประมาณค่าวิธีที่ J

$J = 1, 2, 3$

ค่าเฉลี่ยกำลังสองของผลต่างระหว่างค่าของข้อมูลที่สูญหายกับค่าที่ประมาณได้จากวิธีประมาณค่า หรือค่าเฉลี่ยกำลังสองของผลต่าง (Mean of the Square of Difference: MSD) หมายถึง ค่าที่ได้จาก

$$\sum_{j=1}^{N} (X_j^M - X_j^{(J)})^2/N$$

ตำแหน่งของข้อมูลที่สูญหาย หมายถึง ขอบเขตของข้อมูลที่สูญหายมีค่าตกอยู่ในช่วงเขตใดเขตหนึ่งต่อไปนี้

1. เขตที่ 1 ค่าของข้อมูลที่สูญหายมากกว่า $\bar{X}$ + 1 S.D. หรือ
2. เขตที่ 2 ค่าของข้อมูลที่สูญหายมีค่าตั้งแต่ $\bar{X}$ - 1 S.D. ถึง $\bar{X}$ + 1 S.D. หรือ
3. เขตที่ 3 ค่าของข้อมูลที่สูญหายน้อยกว่า $\bar{X}$ - 1 S.D.

เมื่อ $\bar{X}$ คือ ค่าเฉลี่ยเลขคณิตของกลุ่มตัวอย่างที่ยังคงเหลืออยู่

S.D. คือ ค่าส่วนเบี่ยงเบนมาตรฐานของกลุ่มตัวอย่างที่ยังคงเหลืออยู่

ประโยชน์ของการวิจัย

การวิจัยครั้งนี้ให้ประโยชน์ในการหาหลักฐานและผลสรุป ที่จะแนะนำผู้ใช้วิธีประมาณค่าแทนข้อมูลที่สูญหาย ให้สามารถมีเกณฑ์ในการตัดสินใจที่มีประสิทธิภาพว่า ถ้ามีการสูญหายของข้อมูลเกิดขึ้นควรจะเลือกใช้วิธีประมาณค่าวิธีใดที่จะสามารถประมาณค่าแทนข้อมูลที่สูญหายได้ใกล้เคียงความเป็นจริง

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย