บทที่ ๑

หอสมุดกลาง สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

๑.๑ บทนำ

ปัจจุบันนี้การวิจัยในสาขาต่าง ๆ โดยเฉพาะวิทยาศาสตร์และสังคมศาสตร์ จำเป็นจะต้องอาศัยความรู้ทางด้านสถิติ โดยเฉพาะอย่างยิ่งในขั้นตอนที่เกี่ยวกับการวิเคราะห์และสรุปผลของข้อมูล โดยแท้จริงแล้วนักวิจัยไม่อาจจะวางแผนและดำเนินการวิจัยได้ หากไม่มีความรู้ทางด้านสถิติตั้งแต่แรกเริ่ม นักวิจัยจะต้องคิดคำนึงถึงการวิเคราะห์ข้อมูลในทุกขั้นตอนของกระบวนการวิจัย

อย่างไรก็ตาม แม้ว่านักวิจัยจะต้องมีความรู้ทางสถิติ แต่ก็ไม่ได้หมายความว่าจะต้องเป็นผู้เชี่ยวชาญทางด้านสถิติถึงขั้นเป็นนักสถิติ เพียงแต่หมายความว่าอย่างน้อยที่สุด นักวิจัยควรรู้วิธีการทางสถิติที่จำเป็นและที่จะนำมาใช้ในการวิเคราะห์ กล่าวคือ รู้ถึงวัตถุประสงค์ของวิธีการสถิติแต่ละวิธี วิธีการคำนวณทางสถิติของวิธีนั้น ข้อจำกัดของแต่ละวิธี และความเหมาะสมของแต่ละวิธีที่จะนำมาใช้กับข้อมูล ตลอดจนความสอดคล้องระหว่างสถิติที่ใช้กับเป้าหมายของการวิเคราะห์

เมื่อพิจารณาในแง่ของการวิเคราะห์และสรุปข้อมูล เราอาจแบ่งแยกวิธีการทางสถิติได้ ๒ ประเภทตามลักษณะของข้อมูล คือ

- ประเภทที่ไม่ใช้พารามิเตอร์[๑] หรือ ไม่มีรูปแบบการแจกแจงของข้อมูล (nonparametric Statistics or distribution-free methods)

- ประเภทที่ใช้พารามิเตอร์[๒] (parameter Statistics) คือทราบรูปแบบการแจกแจงของข้อมูล

---

[๑]สุชาติ ประสิทธิ์รัฐสินธ์, ธวัชชัย อาทรธุระสุข และ พิสิฐ ศุกรียพงศ์. "สถิติสำหรับการวิจัยทางสังคมศาสตร์," สถาบันบัณฑิตพัฒนบริหารศาสตร์, มิถุนายน ๒๕๑๑, หน้า ๑-๔.

[๒]เรื่องเดียวกัน.

สถิติประเภทหลังมีใช้กันอย่างกว้างขวางในการวิจัย แต่การจะใช้วิธีใดนั้นขึ้นกับข้อมูลที่จะวิเคราะห์ว่ามีลักษณะอย่างไร เหมาะสมกับประเภทของสถิติที่จะนำมาใช้หรือไม่

การใช้สถิติประเภทจำกัดรูปแบบการแจกแจงเพื่อการทดสอบทางสถิติ หรือการอ้างอิงจำเป็นจะต้องทราบข้อสมมติฐานบางอย่างเกี่ยวกับรูปแบบของการแจกแจง หรือลักษณะของประชากรที่มีอยู่จริง ทั้งนี้เพราะหลักเกณฑ์ต่าง ๆ เกี่ยวกับวิธีทางสถิติเหล่านี้ ได้สร้างขึ้นมาจากข้อสมมติฐานเกี่ยวกับการแจกแจงของประชากร เช่น การแจกแจงในรูปแบบปกติ (Normal distribution) แบบสม่ำเสมอ (Uniform distribution) หรือรูปแบบอื่น ๆ ที่ระบุ ในบางครั้งยังมีการกำหนดเงื่อนไขของการวิเคราะห์ เช่น กำหนดว่าจะต้องมีภาวะเอกพันธ์ในด้านความแปรปรวน (Homogeneity of variances)

๑.๒ <u>ที่มาของปัญหา</u>

ในกรณีที่ทราบการแจกแจงของข้อมูล จะสามารถทำการวิเคราะห์เชิงสถิติได้ง่าย เนื่องจากการวิเคราะห์เชิงสถิติแบบจำกัดรูปแบบการแจกแจงได้รับการพัฒนาเป็นอย่างดีแล้ว

ในทางปฏิบัตินักวิจัยส่วนใหญ่มักจะไม่ทดสอบว่าการแจกแจงของข้อมูลที่ศึกษามีลักษณะอย่างไร ซึ่งนับเป็นผลเสีย เพราะถ้าคุณสมบัติของประชากรที่ศึกษามีการแจกแจงแบบหนึ่ง ความถูกต้องและความหมายของผลที่ได้จากการวิเคราะห์ โดยใช้วิธีการทางสถิติประเภทจำกัดรูปแบบการแจกแจง ย่อมขึ้นอยู่กับความถูกต้องของข้อกำหนดในตอนแรก ข้อกำหนดนั้นอาจจะเป็นจริงหรือไม่ก็ได้ ในกรณีที่ไม่เป็นจริงจะทำให้ผลที่ได้จากการวิเคราะห์ไม่ถูกต้อง

การแจกแจงของข้อมูลจะเป็นชนิดใด สามารถบอกได้ถ้าตัวเลขที่ศึกษาเป็นตัวเลขของทั้งประชากร แต่โดยปกตินักวิจัยมักจะไม่ได้ศึกษาข้อมูลของประชากรทั้งหมด ทั้งนี้เพราะขนาดของประชากรที่สนใจมักมีขนาดใหญ่มาก ในแต่ละโครงการวิจัย เวลา ทุนทรัพย์ และจำนวนพนักงานที่ใช้เก็บข้อมูลมีจำกัด ในทางปฏิบัติจึงมีการสุ่มตัวอย่างจากประชากรที่สนใจ

มาเพียงบางส่วน และทำการศึกษาข้อมูลจากตัวอย่างเพื่ออนุมานคุณสมบัติของประชากร จากคุณสมบัติของตัวอย่างที่ได้ ความถูกต้องและเชื่อถือได้ของการคาดคะเนคุณสมบัติของประชากรขึ้นอยู่กับการวิเคราะห์ทางสถิติ ซึ่งขึ้นอยู่กับข้อกำหนดเกี่ยวกับการแจกแจงของข้อมูล รวมทั้งข้อกำหนดอื่น ๆ ดังนั้น ปัญหาสำคัญคือทำอย่างไรจึงจะทราบว่าข้อมูลของประชากรมีการแจกแจงอย่างไรในเมื่อ ข้อมูลที่มีอยู่เป็นข้อมูลจากตัวอย่างเท่านั้น

มีการทดสอบทางสถิติหลายวิธีเพื่อศึกษาว่าข้อมูลจากตัวอย่างที่ได้จะมาจากประชากรที่มีการแจกแจงคล้ายคลึงกับการแจกแจงมาตรฐานชนิดใด การทดสอบที่ใช้กันอยู่คือ

- การทดสอบแบบไคสแควร์ (Chi-Square test)[๑]
- การทดสอบแบบโคลโมโกรอฟ (Kolmogorov test)[๒]
- การทดสอบแบบโคลโมโกรอฟ-สเมย์นอฟ (Kolmogorov-Smirnov test)[๓]
- การทดสอบแบบลิลลายฟอรัส (lillieforus test)[๔]

ในการทดสอบแต่ละวิธีจะต้องทราบฟังค์ชั่นการแจกแจงของข้อมูลตามทฤษฎี เพื่อใช้คำนวณค่าสถิติที่ใช้ในการทดสอบ ข้อมูลที่ได้จะผ่านการทดสอบได้หรือไม่นั้นขึ้นอยู่กับค่าสถิติทดสอบที่สร้างขึ้น ปัญหาที่ตามมาคือจะหาค่าของพารามิเตอร์ซึ่งอยู่ในฟังก์ชั่นของการแจกแจงได้อย่างไร โดยปกติผู้ทดสอบจะประมาณค่าพารามิเตอร์จากข้อมูลตัวอย่างที่ได้รับนั่นเอง เช่น ถ้าฟังค์ชั่นการแจกแจงเป็นฟังค์ชั่นของค่าเฉลี่ยจะคำนวณค่าเฉลี่ยจากตัวอย่างมาใช้แทนที่ค่าเฉลี่ยของประชากรในฟังค์ชั่น หรือถ้าฟังค์ชั่นการแจกแจงขึ้นอยู่กับฟังค์ชั่นของค่าเฉลี่ยและความแปรปรวน ก็จะคำนวณค่าทั้งสองจากตัวอย่างมาแทนที่พารามิเตอร์ที่ต้องการ เมื่อกำหนดค่าพารามิเตอร์แล้วจึงหาค่าสถิติทดสอบ

---

[๑]HOGG and CRAIG, "Introduction to Mathematical Statistics," Macmillan Publishing Co., Inc. New York, pp. 308. 1968.

[๒]W. Feller, "On the Kolmogorov-Smirnov limit theorems for empirical distribution," Annals of Math. Stat., 19, pp.177-189. 1965.

[๓]Jaroslov Hajek, "Non Parametric Statistics," San Francisco, Holden-Day, 1968.

[๔]Ibid.

กำหนดเขตของการยอมรับและปฏิเสธ (acceptance and rejection region) สมมติฐาน โดยนำค่าสถิติทดสอบจากตารางมาเป็นตัวกำหนดแล้ว-พิจารณาดูว่าค่าสถิติทดสอบที่คำนวณได้ตกอยู่ในเขตใด ถ้าค่าสถิติทดสอบตกอยู่ในเขตยอมรับจึงจะสรุปว่าข้อมูลชุดนี้มีการแจกแจงตามสมมติฐานที่ตั้งไว้ แต่ถ้าตกอยู่ในเขตที่จะปฏิเสธสมมติฐานจะสรุปว่าข้อมูลชุดนี้ไม่มีการแจกแจงตามสมมติฐานที่ตั้งไว้

สำหรับในการวิจัยนี้ใช้วิธีการทดสอบแบบไคสแควร์ เป็นที่ทราบกันโดยทั่วไปว่าถ้าค่าสถิติทดสอบไคสแควร์ต่ำลงมาก-ๆ จนเข้าใกล้ศูนย์ แสดงว่าข้อมูลที่ศึกษามีการแจกแจงเข้าใกล้กับฟังก์ชั่นการแจกแจงที่เรากำหนดมากขึ้น

อย่างไรก็ดียังไม่มีหลักฐานยืนยันได้แน่ชัดว่า ค่าของตัวสถิติทดสอบไคสแควร์ที่คำนวณได้นั้น เป็นค่าที่ดีที่สุด ค่าที่ดีที่สุดของตัวสถิติทดสอบไคสแควร์ในการวิจัยนี้หมายความว่าเป็นค่าของตัวสถิติทดสอบไคสแควร์ที่ต่ำที่สุดในบรรดาค่าของตัวสถิติทดสอบไคสแควร์ค่าอื่น ๆ ที่ใช้ทดสอบกับข้อมูลชุดเดียวกัน และค่าของตัวสถิติทดสอบเหล่านี้ต่างกัน ก็เพราะใช้ตัวประมาณค่าของพารามิเตอร์ที่ต่างค่ากัน ทั้งนี้เพราะในการทดสอบการแจกแจงของข้อมูลโดยทั่วไป ถ้าผลการทดสอบปฏิเสธสมมติฐานที่ตั้งเอาไว้ในตอนแรก เราจะหาค่าของตัวประมาณค่าของพารามิเตอร์ใหม่ แล้วทำการทดสอบการแจกแจงของข้อมูลตามวิธีการเดิม โดยได้ค่าของสถิติทดสอบไคสแควร์ค่าใหม่ และจะทำเช่นนี้ไปเรื่อย ๆ จนกว่าผลการทดสอบยอมรับสมมติฐานใดสมมติฐานหนึ่ง ซึ่งเรียกวิธีการอย่างนี้ว่า Cut And Try จะเห็นว่าการยอมรับหรือการปฏิเสธสมมติฐานที่ตั้งไว้ ขึ้นอยู่กับค่าของตัวสถิติทดสอบไคสแควร์ที่ได้จากการกำหนดค่าประมาณของตัวพารามิเตอร์สำหรับข้อมูลที่ทำการวิจัย ถ้าค่าของตัวสถิติทดสอบไคสแควร์ที่ได้ เป็นค่าที่ต่ำที่สุดซึ่งได้จากการกำหนดค่าประมาณของพารามิเตอร์ และไม่มีค่าประมาณของพารามิเตอร์ค่าอื่น ๆ ที่ทำให้ค่าของตัวสถิติทดสอบไคสแควร์ต่ำกว่านี้อีกแล้วจะทำให้มีความมั่นใจในผลการทดสอบที่ยอมรับหรือปฏิเสธสมมติฐานที่ตั้งไว้

ดังนั้นการวิจัยนี้จึงทำการทดลองว่าจะมีเครื่องมือที่ดีพอหรือไม่ที่จะทำให้เราแน่ใจได้ว่า ค่าของตัวสถิติทดสอบไคสแควร์ที่ได้จากการกำหนดค่าประมาณของพารามิเตอร์แบบใดเป็นค่าที่ต่ำที่สุด และค่าประมาณของพารามิเตอร์ที่ได้จากการทำให้ค่าของตัวสถิติทดสอบไคสแควร์ต่ำที่สุดนั้น ยังคงเป็นค่าประมาณของพารามิเตอร์ที่ดีพอสำหรับนำไปใช้งานในขั้นต่อ ๆ ไปทางสถิติได้หรือไม่

๑.๓ แนวความคิดการกำหนดค่าพารามิเตอร์และประวัติ

ผู้วิจัยได้แนวความคิดที่ว่าจะกำหนดค่าพารามิเตอร์ เพื่อทำให้ตัวสถิติทดสอบไคสแควร์มีค่าต่ำสุดได้อย่างไร จากวิทยานิพนธ์เรื่องการหาค่าอุดมะของพารามิเตอร์ของฟังค์ชั่นการแจกแจงสำหรับข้อมูลที่กำหนดให้[๑] (Optimal Determination of Density Function Parameters for a Given Set of Data) เนื้อหาหลักของวิทยานิพนธ์ที่กล่าวถึงนี้แบ่งออกเป็น ๒ ตอน คือ

ตอนที่ ๑ กรณีที่ข้อมูลมีพารามิเตอร์เพียงตัวเดียว ข้อมูลตัวอย่างที่ใช้เป็นข้อมูลที่เก็บจากจำนวนครั้งของการบำรุงรักษาซ่อมแซมคอมพิวเตอร์ ซึ่งในวิทยานิพนธ์ฉบับดังกล่าวเชื่อว่าลักษณะการแจกแจงของข้อมูลดังกล่าวมีความคล้ายคลึงกับการแจกแจงแบบเอ็กโพเน็นเชียล (Exponential Distribution) และวิธีการค้นหาค่าพารามิเตอร์ที่ดีที่สุด คือ วิธีการค้นหาตัวแปรเดียวของพาวเวลล์โดยไม่ใช้อนุพันธ์ (Powell's univariate search without derivative)

ตอนที่ ๒ กรณีที่ข้อมูลมีพารามิเตอร์ ๒ ตัว ข้อมูลตัวอย่างที่ใช้เป็นข้อมูลที่เก็บจากช่วงเวลาที่คนไข้ใช้ในการรักษาในตึกหมายเลขหนึ่งในโรงพยาบาลวชิระ ซึ่งในวิทยานิพนธ์ฉบับดังกล่าวเชื่อว่า ลักษณะการแจกแจงของข้อมูลมีความคล้ายคลึงกับการแจกแจงแบบแกมม่า (Gamma Distribution) และวิธีการค้นหาค่าพารามิเตอร์ที่ดีที่สุด คือ วิธีทิศทางสังยุคของ

---

[๑]วิภาวรรณ รุ่งเรือง, "การหาค่าอุดมะของพารามิเตอร์ของฟังค์ชั่นการแจกแจงสำหรับข้อมูลที่กำหนดให้" (วิทยานิพนธ์ปริญญามหาบัณฑิต สาขาสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์, ๒๕๑๙).

พาวเวลล์ (Powell's Conjugate Direction Method) ร่วมกับวิธีที่ได้กล่าวไว้ในตอนที่ ๑ ซึ่งอยู่ในหน้า ๔ ของวิทยานิพนธ์ฉบับดังกล่าว

วิทยานิพนธ์ฉบับดังกล่าวไม่ได้พิสูจน์ผลตัวประมาณค่าพารามิเตอร์ตัวใหม่ที่ได้จากการใช้เทคนิคการหาค่าอุตมะ (Optimization technique) ทั้ง ๒ วิธี ว่าเป็นตัวประมาณค่าที่ดีกว่าตัวประมาณค่าพารามิเตอร์เดิม โดยเป็นตัวแทนที่ดีพอของพารามิเตอร์หรือไม่ เพียงแต่ได้อ้างอิงไว้ในวิทยานิพนธ์ดังกล่าวว่า ตัวประมาณค่าพารามิเตอร์ตัวใหม่ที่ได้จากการใช้เทคนิคการหาค่าอุตมะทั้ง ๒ วิธี ทำให้ค่าสถิติทดสอบที่ใช้ในการทดสอบการแจกแจงมีค่าต่ำกว่าค่าสถิติทดสอบ ที่ได้จากตัวประมาณค่าพารามิเตอร์ที่ได้จากข้อมูลโดยตรง (Conventional mean and variance) ซึ่งมีอยู่ในหน้า ๕๓ - ๕๖, ๗๔ - ๘๒ ของวิทยานิพนธ์ฉบับดังกล่าว ทำให้ผู้วิจัยตั้งข้อสงสัยว่าวิธีการเทคนิคการหาค่าอุตมะ (optimization technique) ซึ่งเป็นส่วนหนึ่งของงานโปรแกรมเชิงเส้น (linear programming) จะเป็นเครื่องมือที่ดีพอสำหรับก้าวใหม่ทางสถิติหรือไม่ เรามีความมั่นใจมากน้อยเพียงใดว่า ตัวประมาณค่าพารามิเตอร์ตัวใหม่ที่ได้จากการใช้เทคนิคการหาค่าอุตมะที่ทำให้ตัวสถิติทดสอบต่ำที่สุดนั้น เป็นตัวแทนของพารามิเตอร์ของประชากรที่เราต้องการศึกษาที่ดีพอแล้ว และสามารถนำเอาตัวประมาณค่าพารามิเตอร์ตัวใหม่นี้ไปใช้ในการแก้ปัญหาทางสถิติในขั้นต่อไปได้อย่างมีประสิทธิภาพ

ดังนั้นผู้วิจัยจึงได้สร้างขั้นตอนการวิจัยที่เน้นหนักไปในทางทฤษฎีทางสถิติ เพื่อที่จะสามารถดูผลการวิจัยว่าค่าพารามิเตอร์ตัวใหม่ที่ได้จากการใช้เทคนิคการหาค่าอุตมะที่ทำให้สถิติทดสอบไคสแควร์ต่ำสุดนั้น เป็นตัวประมาณค่าของพารามิเตอร์ที่ดีพอที่จะนำไปใช้ในขั้นตอนที่จะศึกษาต่อไปด้วย ความน่าจะเป็นมากน้อยเพียงใด ซึ่งจะได้กล่าวถึงโดยละเอียดต่อไปในบทที่ ๔ ผู้วิจัยได้เสนอโครงการวิทยานิพนธ์ของผู้วิจัยเอง ต่อบัณฑิตจุฬาลงกรณ์มหาวิทยาลัย และได้รับอนุมัติเมื่อวันที่ ๑ สิงหาคม ๒๕๒๒ โดยกำหนดโครงการวิทยานิพนธ์เดิมไว้ดังนี้

เรื่อง การเปรียบเทียบวิธีการทดสอบการแจกแจงของข้อมูลแบบไคสแควร์ และโคลโมโกรอฟ-สเมอร์นอฟ

(A comparison of methods for testing the distribution of data between chi-square and kolmogorov-Smirnov tests)

สำหรับในหัวข้อ ๑.๓.๑ และ ๑.๓.๒ จะได้กล่าวถึงขั้นตอนของการวิจัยเดิมและวัตถุประสงค์เดิมของการวิจัย

๑.๓.๑ ขั้นตอนเดิมของการวิจัยในข้อเสนอของวิทยานิพนธ์ดังกล่าวมีดังนี้

(ก) สร้างเลขสุ่มที่มีการแจกแจงเข้ากับการแจกแจงมาตรฐานที่ใช้กันโดยทั่ว ๆ ไป คือ

- การแจกแจงแบบปกติ (Normal Distribution)
- การแจกแจงแบบยูนิฟอร์ม (Uniform Distribution)

(ข) จากประชากรที่สร้างขึ้นมาตามข้อ (ก) สุ่มตัวอย่างจากประชากรขึ้นมาจำนวนหนึ่งหลาย ๆ ตัวอย่าง หาตัวประมาณค่าของพารามิเตอร์คือค่าเฉลี่ยและค่าความแปรปรวนจากข้อมูลที่ศึกษา ทำการทดสอบว่าตัวอย่างใหม่ที่สุ่มมามีการแจกแจงเข้ากับการแจกแจงมาตรฐานตามข้อ (ก) หรือไม่ โดยใช้ตัวสถิติทดสอบคือ

- การทดสอบไคสแควร์ (Chi-square test)
- การทดสอบโคลโมโกรอฟสเมอร์นอฟ (Kolmogorov-Smirnov test)

(ค) ประมาณค่าพารามิเตอร์ตัวใหม่ โดยอาศัยตัวประมาณค่าพารามิเตอร์เดิม (Conventional mean and variance) ที่ได้ในข้อ (ข) โดยให้ได้ค่าสถิติทดสอบค่าใหม่ดีกว่าค่าสถิติทดสอบเดิมในข้อ (ข) โดยใช้วิธีการเทคนิคการหาค่าอุตมะ (ปรับค่าสถิติทดสอบเดิมให้มีค่าเหมาะสมที่สุดจนได้ค่าสถิติทดสอบไคสแควร์ตัวใหม่มีค่าต่ำที่สุดในบรรดาค่าสถิติทดสอบที่ได้จากตัวประมาณค่าพารามิเตอร์อื่น ๆ) เทคนิคการหาค่าอุตมะที่ทำให้ค่าสถิติทดสอบไคสแควร์ต่ำที่สุด ได้แก่

- วิธีทิศทางสังยุคของพาวเวลล์ (Powell Conjugate Direction Method)
- วิธีเพนนัลตี้ฟังก์ชั่น (Penalty function Method)

(ง) เปรียบเทียบค่าสถิติทดสอบที่ได้จาก ๒ วิธีในข้อ (ข) ว่าดีกว่ากันเพียงใดและเปรียบเทียบตัวประมาณค่าพารามิเตอร์ที่คำนวณได้ใหม่โดยใช้วิธีในข้อ (ค) นี้

ใกล้เคียงกับค่าพารามิเตอร์ของประชากรในข้อ (ก) หรือไม่เพียงใด โดยการใช้ t-test ซึ่งจะได้กล่าวถึงโดยละเอียดต่อไปในหัวข้อ ๒.๓

๑.๓.๒ วัตถุประสงค์ของการวิจัย

(ก) เพื่อทดสอบวิธีการหาค่าประมาณของค่าพารามิเตอร์ของประชากร เพื่อให้ค่าของตัวสถิติทดสอบมีค่าต่ำสุด ดังในหัวข้อ ๑.๓.๑ ข้อ (ค)

(ข) เพื่อทำการเปรียบเทียบวิธีการทดสอบการแจกแจงของข้อมูลระหว่างการทดสอบแบบไคสแควร์และโคลโมโกรอฟ-สเมอร์นอฟ เมื่อการประมาณค่าพารามิเตอร์ได้มาจากข้อมูลโดยตรงเทียบกับการประมาณค่าพารามิเตอร์ ใช้วิธีทิศทางสังยุคของพาวเวลล์และเพนนัลตี้ฟังค์ชั่น

๑.๓.๓ ขั้นตอนของงานวิจัย

ขั้นตอนเดิมของการวิจัยที่ได้เสนอไว้แล้วนี้ ตั้งอยู่บนรากฐานที่ว่า การทดสอบการแจกแจงมาตรฐานของตัวอย่างที่สุ่มขึ้นมา โดยใช้วิธีการทดสอบแบบไคสแควร์ และการค้นหาค่าอุตมะของค่าไคสแควร์ดังกล่าวใช้วิธีการทิศทางสังยุคของพาวเวลล์ และการค้นหาตัวแปรเดียวของพาวเวลล์โดยไม่ต้องอาศัยอนุพันธ์ ใช้ได้ผลกับเลขสุ่มที่มีลักษณะการแจกแจงมาตรฐานของข้อมูลคล้ายคลึงกับการแจกแจงแบบแกมม่าและการแจกแจงแบบเอ็กโพเนนเชียล ซึ่งพอจะสรุปสมมติฐานที่ได้ตั้งไว้ในขณะนั้นว่า

- การใช้โคลโมโกรอฟ-สเมอร์นอฟทดสอบ และไคสแควร์ทดสอบ เมื่อประมาณค่าพารามิเตอร์ด้วยวิธีทิศทางสังยุคของพาวเวลล์ (Powell Conjugate direction) และวิธีการเพนนัลตี้ฟังค์ชั่น (Penalty function) จะให้ผลผิดพลาดน้อยกว่าการใช้โคลโมโกรอฟ-สเมอร์นอฟ และไคสแควร์ทดสอบ

- การใช้โคลโมโกรอฟ-สเมอร์นอฟ ในการทดสอบการแจกแจงของข้อมูลจะให้ผลผิดพลาดน้อยกว่าการใช้ไคสแควร์ทดสอบ เมื่อใช้วิธีทิศทางสังยุคของพาวเวลล์ และวิธีการเพนนัลตี้ฟังค์ชั่น ในการประมาณค่าพารามิเตอร์ในฟังค์ชั่นการแจกแจงของข้อมูล

เนื่องจากผู้วิจัยสำเร็จการศึกษาระดับปริญญาตรีทางสถิติ ไม่มีความรู้ทางด้านการหาค่าอุตมะ (Optimization technique) มาก่อน จึงมีความจำเป็นอย่างมากที่ต้องใช้เวลาในการศึกษาทำความเข้าใจในเรื่องนี้ด้วยตัวเอง โดยการค้นคว้าหาความรู้จากหนังสือบ้าง วารสารบ้าง และสอบถามผู้มีความรู้ในด้านนี้โดยตรง โดยการเริ่มเรียนรู้ในขั้นพื้นฐานและทำแบบฝึกหัด เพื่อเป็นแนวความรู้ในการทำความเข้าใจในวิธีการของพาวเวลล์ และเพนนัลตี้-ฟังค์ชั่น อย่างไรก็ดีภายหลังจากที่ได้ดำเนินการวิจัยไปแล้วเป็นเวลา ๑๕ เดือน พบว่าแซงวิลล์ (Zangwill)[๑] ได้เขียนบทความเกี่ยวกับวิธีการของพาวเวลล์ว่า มีข้อผิดพลาดเกิดขึ้นใช้ไม่ได้ผลกับฟังค์ชั่นที่หาอนุพันธ์ไม่ได้บางกรณี ซึ่งข้อผิดพลาดดังกล่าวมิได้มีปรากฏในวิทยานิพนธ์ของ น.ส. วิภาวรรณ รุ่งเรือง[๒] ซึ่งรายละเอียดข้อผิดพลาดของพาวเวลล์จะได้กล่าวโดยละเอียดในบทที่สอง ในหัวข้อ ๒.๗

๑.๓.๔ ปัญหาที่เกิดขึ้น

ก่อนที่จะได้กล่าวถึงวัตถุประสงค์ใหม่ของการวิจัย จะขอกล่าวสรุปถึงปัญหาที่เกิดขึ้นเสียก่อน (โปรดดูรูปที่ ๑ ประกอบ)

(ก) เป็นที่ทราบกันดีว่าวิธีการค้นหาค่าอุตมะ (Optimization Technique) ในงานโปรแกรมเชิงเส้น (Linear programming) หรือในงานโปรแกรมไม่ใช่เชิงเส้น (Nonlinear programming) มีเทคนิคการค้นหามากมายหลายวิธี แต่ละวิธีก็มีเงื่อนไขในการนำมาใช้แตกต่างกันไป การที่จะเลือกเอาวิธีการอันไหนมาใช้ขึ้นอยู่กับความเหมาะสมของปัญหา วิธีที่ใช้กันแพร่หลายได้แก่ วิธีการแกรเดียนท์ (Gradiant Method) เป็นวิธีการที่ต้องอาศัยอนุพันธ์ (derivative) ของฟังคชั่น ซึ่งหมายความว่าฟังค์ชั่นที่จะใช้วิธีการนี้ค้นหาค่าอุตมะต้องเป็นฟังค์ชั่นที่หาอนุพันธ์ได้

---

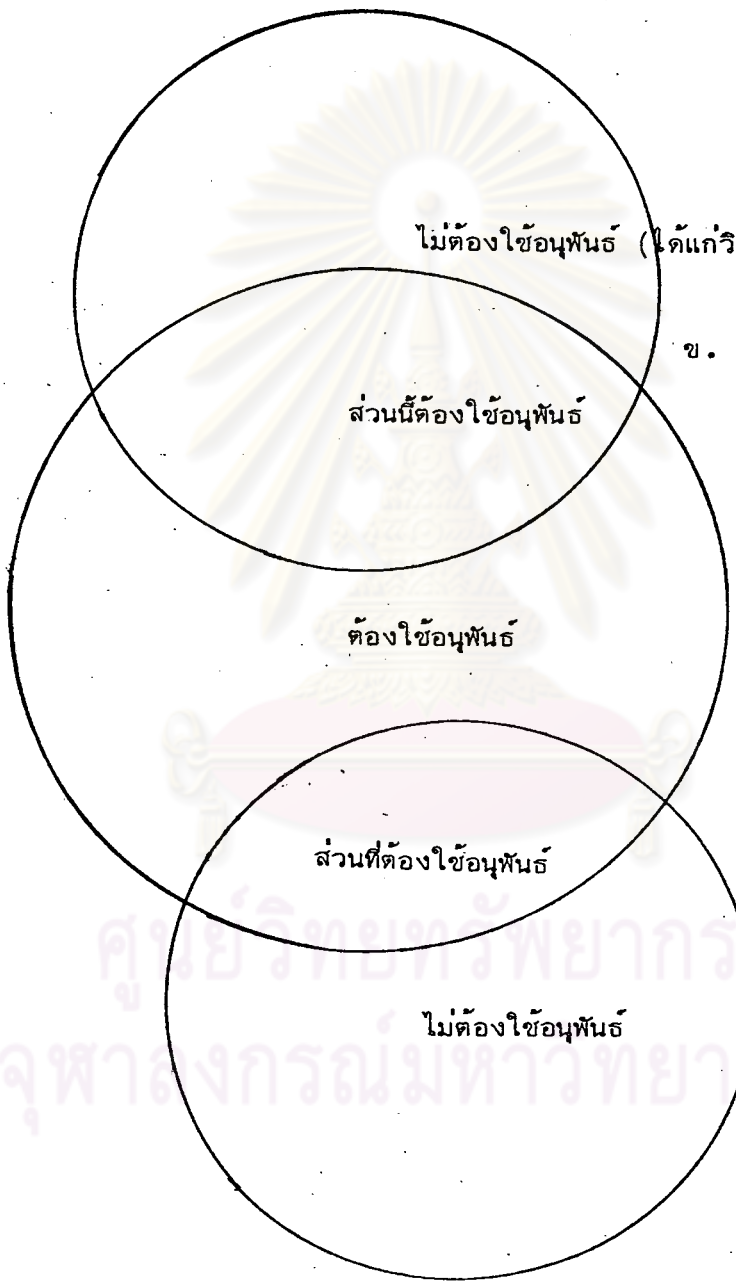
[๑]Zangwill, W.I. "Minimizing a Function Without Calculating Derivatives", The Computer Journal, 10, 293-296 Nov. 1967.

[๒]วิภาวรรณ รุ่งเรือง "การหาค่าอุตมะของพารามิเตอร์ของฟังค์ชั่นการแจกแจงสำหรับข้อมูลที่กำหนดให้" (วิทยานิพนธ์ คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์ ๒๕๑๙).

รูปที่ ๑.๑

ผังแสดงขอบเขตคุณสมบัติของวิธีการค้นหาค่าอุดมะ (Optimization Technique)

ประเภท Multidimensional - squential search methods



(ข) วิธีทิศทางสังยุค (Conjugate Direction) เป็นวิธีค้นหาที่มีประสิทธิภาพดีอีกวิธีหนึ่ง ซึ่งมีคุณสมบัติของ Quadratic Convergence และโดยปกติแล้วอนุพันธ์ของฟังค์ชั่นที่มีอยู่ต้องหาค่าได้ก่อน วิธีการของพาวเวลล์ (Powell's Method) เป็นวิธีทิศทางสังยุคประเภทหนึ่งซึ่งไม่ต้องการอนุพันธ์ของฟังค์ชั่น

อย่างไรก็ดีแซงวิลล์ (Zangwill) พบว่าวิธีการของพาวเวลล์ (Powell) ไม่ถูกต้องในทุกกรณีที่ฟังค์ชั่นหาอนุพันธ์ไม่ได้ (โปรดอ่านรายละเอียดในหัวข้อ ๒.๖)

(ค) เนื่องจากไม่ต้องการเสียเวลาศึกษาบทความของแซงวิลล์อีก จึงมีความจำเป็นต้องหาวิธีใหม่ที่รู้จักกันเป็นที่แพร่หลายแล้ว ได้แก่ วิธีการค้นหาตัวแปรเดียว (Univariate Search) ดังเป็นที่ทราบกันดีแล้วว่า ฟังค์ชั่นของการแจกแจงแบบปกตินั้น ไม่สามารถหาอนุพันธ์ของฟังค์ชั่นไคสแควร์ได้ ดังนั้นจึงมีความจำเป็นต้องอาศัยการค้นหาตัวแปรเดียวแบบไม่ต้องอาศัยอนุพันธ์ของฟังค์ชั่น

นอกจากนี้ขอบเขตของหัวข้อวิทยานิพนธ์ที่ได้เสนอไปแล้วนี้กว้างขวางเกินกว่าที่จะทำได้สำเร็จเรียบร้อยภายในระยะเวลาอันสมควร ดังนั้นผู้วิจัยจึงมีความจำเป็นต้องเปลี่ยนแปลงหัวข้อวิทยานิพนธ์และขอบเขตของการวิจัยเสียใหม่ให้เหมาะสมกับสถานการณ์และเวลาที่มีอยู่

๑.๔ วัตถุประสงค์ของการวิจัย

๑.๔.๑ ศึกษาวิธีการทดสอบการแจกแจงของข้อมูลว่าเข้ากับลักษณะการแจกแจงมาตรฐานชนิดใดเมื่อ

(ก) ตัวสถิติทดสอบการแจกแจงของข้อมูลใช้ตัวประมาณค่าพารามิเตอร์ที่ได้มาจากข้อมูลโดยตรง

(ข) ตัวสถิติทดสอบการแจกแจงของข้อมูลใช้ตัวประมาณค่าพารามิเตอร์ที่ได้มาจากการใช้เทคนิคการหาค่าอุตมะ

๑.๔.๒ ศึกษาเปรียบเทียบตัวประมาณค่าพารามิเตอร์ที่ได้จากวิธีการใช้เทคนิคการหาค่าอุตมะกับตัวประมาณค่าพารามิเตอร์ที่ได้จากการใช้สูตรโดยตรง

๑.๕ ขั้นตอนใหม่และขอบเขตใหม่ของการวิจัย

๑.๕.๑ สร้างเลขสุ่มที่มีการแจกแจงเข้ากับการแจกแจงมาตรฐานที่ใช้กันโดยทั่ว ๆ ไป การแจกแจงมาตรฐานที่จะใช้ในการวิจัยคือ การแจกแจงแบบปกติ (Normal Distribution) ซึ่งมีค่าเฉลี่ยเป็น ๔๐.๕ และค่าความเบี่ยงเบนมาตรฐานเป็น ๗.๕

๑.๕.๒ จากเลขสุ่มที่ถูกสร้างขึ้นมาในข้อ ๑.๕.๑ สุ่มตัวอย่างขึ้นมาจำนวนหนึ่งหลาย ๆ ชุดแล้วทำการทดสอบว่าตัวอย่างข้อมูลใหม่ที่สุ่มมานี้มีการแจกแจงเข้ากับการแจกแจงมาตรฐานตามข้อ ๑.๕.๑ หรือไม่ และกี่เปอร์เซ็นต์ของจำนวนชุดทั้งหมดที่ถูกเลือกขึ้นมามีการแจกแจงมาตรฐานตามข้อ ๑.๕.๑ ตัวสถิติทดสอบที่ใช้ในการวัดการแจกแจงของข้อมูลในการวิจัยนี้ คือ ไคสแควร์ซึ่งใช้ตัวประมาณค่าพารามิเตอร์ที่มาจากค่าเฉลี่ยและค่าความเบี่ยงเบนมาตรฐานข้อมูลตัวอย่างโดยตรง

๑.๕.๓ หาค่าประมาณพารามิเตอร์ตัวใหม่โดยใช้ตัวอย่างข้อมูลเดิมในข้อ ๑.๕.๒ โดยมีจุดมุ่งหมายว่าค่าประมาณพารามิเตอร์ตัวใหม่นี้ต้องทำให้ตัวสถิติทดสอบไคสแควร์ต่ำที่สุด นั่นคือต่ำกว่าตัวสถิติทดสอบไคสแควร์ที่ได้ในข้อ ๑.๕.๒ วิธีการหาค่าอุตมะที่ทำให้ตัวสถิติทดสอบไคสแควร์ต่ำที่สุด จะใช้วิธีการค้นหาตัวแปรเดียวโดยไม่ต้องอาศัยอนุพันธ์ (Univariate Search without derivative) โดยการอาศัยค่าประมาณของพารามิเตอร์ที่ได้มาจาก ๑.๕.๒ เป็นจุดเริ่มต้นในการค้นหาค่าประมาณของพารามิเตอร์ตัวใหม่ที่ทำให้ค่าสถิติทดสอบไคสแควร์ต่ำสุด

๑.๕.๔ เปรียบเทียบค่าสถิติทดสอบไคสแควร์ที่ได้จาก ๑.๕.๓ ว่าต่ำกว่าค่าสถิติทดสอบไคสแควร์ที่ได้จากข้อ ๑.๕.๒ หรือไม่เพียงใด ซึ่งเป็นที่ทราบกันดีอยู่แล้วว่าค่าสถิติทดสอบไคสแควร์ต่ำย่อมดีกว่าค่าสถิติทดสอบไคสแควร์ที่สูง ๆ และในการใช้เทคนิคการหาค่าอุตมะของค่าสถิติทดสอบ

ไคสแควร์ที่สูง ๆ และในการใช้เทคนิคการหาค่าอุตมะของค่าสถิติทดสอบไคสแควร์นี้ทำให้เปอร์-เซนต์ของจำนวนตัวอย่างที่ทดสอบการแจกแจงของข้อมูลโดยการใช้ค่าสถิติทดสอบไคสแควร์ในข้อ ๑.๕.๒ ที่แตกต่างไปจากการแจกแจงมาตรฐานในข้อ ๑.๕.๑ คือ การแจกแจงแบบปกติลดลงจากเดิมมากน้อยเพียงใด

๑.๕.๕ เปรียบเทียบค่าประมาณของพารามิเตอร์ที่ได้จากการใช้สูตรโดยตรงจากข้อมูล (Conventional mean and variance) กับค่าประมาณของพารามิเตอร์ที่ใช้เทคนิคการหาค่าอุตมะ (Optimization technique) ว่ามีความแตกต่างกันอย่างมีนัยสำคัญหรือไม่เพียงใด โดยอาศัยการทดสอบแบบสติวเด้นท์ที (t-test)

๑.๕.๖ เปรียบเทียบค่าประมาณของพารามิเตอร์ที่ได้จากข้อ ๑.๕.๓ ว่ามีความแตกต่างกันอย่างมีนัยสำคัญหรือไม่เพียงใดกับค่าพารามิเตอร์ที่มีอยู่ในข้อ ๑.๕.๑ โดยอาศัยการทดสอบแบบสติวเด้นท์ที (t-test)

๑.๖ <u>ความสำคัญหรือประโยชน์ที่คาดว่าจะได้รับจากการวิจัยมีดังนี้</u>

(ก) วิทยานิพนธ์ฉบับนี้จะเสนอขั้นตอนการคำนวณค่าที่เหมาะสม (Optimal) ของตัวประมาณค่าพารามิเตอร์ ($\bar{X}^*$, S^*) ซึ่งจะให้ค่าสถิติทดสอบไคสแควร์ต่ำสุดในการทดสอบภาวะสารูปสนิทดีสำหรับฟังค์ชั่นการแจกแจงแบบปกติ

(ข) ตัวประมาณค่าของพารามิเตอร์ที่ได้จากการใช้เทคนิคการค้นหาค่าอุตมะ ด้วยวิธีการค้นหาตัวแปรเดียว (Univariate Search) ของค่าสถิติทดสอบไคสแควร์ (minimum chi-square) นี้ สามารถนำเอาไปใช้ในการแก้ปัญหาสถิติในขั้นต่อไปได้ กล่าวคือ ผู้วิจัยเชื่อว่าการนำตัวประมาณค่าของพารามิเตอร์ใหม่นี้ไปใช้งาน ย่อมจะให้ผลเชิงสถิติดีกว่าตัวประมาณค่าของพารามิเตอร์ที่หาได้โดยอาศัยวิธีมาตรฐานทั่วไป (Conventional Methods)