บทที่ 3

วิธีดำเนินการวิจัย

การทำวิจัยครั้งนี้ต้องการศึกษาเปรียบเทียบอำนาจการทดสอบ และความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ของตัวสถิติ 3 วิธี คือวิธีจีแบร์รี่ วิธีของคุ๊ก วิธีของแอนดูร์และเพรตจิบอน ในการตรวจสอบค่าผิดปกติและอิทธิพลของค่าสังเกต เมื่อความความคลาดเคลื่อน (ε) มีการแจกแจงแบบปกติปลอมปน โดยการสร้างจากการแจกแจงแบบปกติที่มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวน เท่ากับ 1 สำหรับค่าผิดปกติสร้างจากการแจกแจงปกติ ที่มีค่าเฉลี่ยและค่าความแปรปรวน แตกต่างจากค่าสังเกตอื่น โดยกำหนด สเกลแฟคเตอร์ 3 ระดับ คือ 3 4 และ 5 โลเคชั่นแฟคเตอร์ 3 ระดับ คือ 4 6 15 จำนวนตัวแปรอิสระเป็น 2 4 6 8 10 และขนาดตัวอย่างเป็น 20 30 50 70 ในการศึกษาค่าอำนาจการทดสอบ และความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ใช้วิธีมอนติคาร์โล ซึ่งเป็นเทคนิคอย่างหนึ่งที่ใช้แก้ปัญหาในการคำนวณทางคณิตศาสตร์ โดยเฉพาะกรณีที่ไม่สามารถหาคำตอบโดยวิธีทางทฤษฎีได้

เนื่องจากวิธีมอนติคาร์โลเป็นเทคนิคที่ใช้ในการวิจัยครั้งนี้ ดังนั้นในตอนแรกของบทจะกล่าวถึงวิธีมอนติคาร์โลก่อน แล้วจึงแสดงรายละเอียดของแผนการทดลอง ขั้นตอนการวิจัยและโปรแกรมที่ใช้ในการวิจัยตามลำดับ ซึ่งรายละเอียดต่าง ๆ เป็นดังนี้

3.1 วิธีมอนติคาร์โล (Monte Carlo Method)

เทคนิคที่ใช้สำหรับแก้ปัญหาในการคำนวณทางคณิตศาสตร์ นั้นมีอยู่หลายวิธี วิธีมอนติคาร์โลเป็นวิธีหนึ่งที่จะใช้แก้ปัญหาได้ และเป็นวิธีที่นิยมใช้กันอย่างแพร่หลายในปัจจุบัน แฮมเมอร์เลย์ และแฮนสคอมบ์ (Hammersley and Handscomb 1964:2) กล่าวว่า วิธีมอนติคาร์โลเป็นสาขาหนึ่งของคณิตศาสตร์เชิงทดลอง ซึ่งหลักการของวิธีมอนติคาร์โลนั้นจะใช้ตัวเลขสุ่ม (Random number) มาช่วยในการหาคำตอบของปัญหาที่ต้องการศึกษา

ในการวิจัยครั้งนี้ จะใช้เทคนิคมอนติคาร์โลดังกล่าว ในการสร้างข้อมูลที่มีสภาพการแจกแจงตามที่ต้องการ ซึ่งขั้นตอนของวิธีมอนติคาร์โลที่ใช้กันอยู่ในปัจจุบัน แบ่งได้เป็น 3 ขั้นตอนดังนี้

3.1.1 การสร้างตัวเลขสุ่ม การใช้ตัวเลขสุ่มเป็นสิ่งที่สำคัญมากในวิธีมอนติคาร์โล ทั้งนี้เพราะว่า หลักการของวิธีมอนติคาร์โลนั้นจะใช้ตัวเลขสุ่มมาช่วย ในการหาคำตอบของปัญหา ลักษณะของตัวเลขสุ่มจะมีการแจกแจงยูนิฟอร์ม ในช่วง (0,1) สำหรับวิธีการสร้างตัวเลขสุ่ม มีผู้เสนอไว้หลายวิธี แต่วิธีที่ดีนั้นลักษณะของตัวเลขสุ่ม ที่เกิดขึ้นจะต้องมีการแจกแจงยูนิฟอร์มในช่วง (0,1) และเป็นอิสระกัน

3.1.2 การประยุกต์ปัญหาที่ต้องการศึกษามาใช้กับตัวเลขสุ่ม ซึ่งขั้นตอนนี้ขึ้นอยู่กับลักษณะปัญหาที่ต้องการศึกษา บางปัญหาอาจจะไม่ใช้ตัวเลขสุ่มโดยตรง แต่อาจจะมีขั้นตอนอื่นอีกหลายขั้นตอน ซึ่งขั้นตอนเหล่านี้บางขั้นตอนที่ต้องใช้ตัวเลขสุ่ม

3.1.3 การทดลองกระทำ เมื่อประยุกต์ปัญหาให้ใช้กับตัวเลขสุ่มได้แล้ว ขั้นตอนต่อไปก็คือ การทดลองโดยใช้กระบวนการของการสุ่ม (Random Process) มากระทำในลักษณะที่ซ้ำ ๆ กัน เพื่อหาคำตอบของปัญหาที่ต้องการศึกษา

3.2 แผนการทดลอง

ในการวิจัยครั้งนี้ กำหนดสถานการณ์ต่าง ๆ สำหรับศึกษาเปรียบเทียบ อำนาจของการทดสอบและความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ของวิธีต่าง ๆ ดังกล่าว โดยสุ่มตัวอย่างจากประชากรที่มีการแจกแจงแบบเดียวกัน ยกเว้นค่าผิดปกติ ที่มีการแจกแจงแบบปกติที่มีสเกลแฟคเตอร์ 3 ระดับ คือ 3 4 5 และโลเคชั่นแฟคเตอร์ 3 ระดับคือ 4 6 และ 15 แต่ละระดับกำหนดให้ มีจำนวนค่าผิดปกติเป็น 1 และ 2 ทุกกรณี รายละเอียดของสเกลแฟคเตอร์และโลเคชั่นแฟคเตอร์เป็นดังนี้

ตารางที่ 3.1 แสดงค่าสเกลแฟคเตอร์ และโลเคชั่นแฟคเตอร์ที่ใช้ในการวิจัยครั้งนี้

(0,1)	(0,3)	(0,4)	(0,5)
(0,1)	(4,1)	(6,1)	(15,1)

โดยที่มีจำนวนตัวแปรอิสระเป็น 2 4 6 8 10 และขนาดตัวอย่างเป็น 20 30 50 70 ในแต่ละการแจกแจง

3.3 ขั้นตอนในการวิจัย

ขั้นตอนในการวิจัยแบ่งออกเป็น 4 ขั้นตอน คือ

1. การสร้างโปรแกรมย่อย สำหรับสร้างการแจกแจงของความคลาดเคลื่อน (ε) ตามที่กำหนด

2. การสร้างข้อมูล (X, Y) ให้มีความสัมพันธ์กันเชิงเส้นตรง

3. การคำนวณค่าสถิติทั้ง 3 วิธี คือ วิธีของจีแบร์รี่ วิธีของคุ๊ก วิธีของแอนดรูและเพรสลิบอน

4. การหาค่าอำนาจของการทดสอบและความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 สำหรับการทดสอบสมมติฐานเกี่ยวกับค่าผิดปกติ

ซึ่งรายละเอียดสำหรับแต่ละขั้นตอนเป็นดังนี้

3.3.1 การสร้างโปรแกรมย่อยสำหรับสร้างการแจกแจงของความคลาดเคลื่อน (ε) ตามที่กำหนด

การสร้างลักษณะการแจกแจงของความคลาดเคลื่อนทุกรูปแบบ ตามที่กำหนดไว้ในแผนการทดลองนั้น ใช้โปรแกรมภาษาฟอร์แทรนโฟ (Fortran IV) โดยใช้กับเครื่อง IBM 370/3031 ซึ่งการสร้างลักษณะการแจกแจงแบบปกติ จะต้องใช้กับตัวเลขสุ่ม ซึ่งมีการแจกแจงแบบยูนิฟอร์มในช่วง (0,1) เป็นพื้นฐานในการสร้าง สำหรับโปรแกรมที่ใช้ตัวเลขสุ่มในการวิจัยครั้งนี้ ใช้วิธีไวท์และสมิดท์ (White and Schmidt 1975:421) เสนอไว้ ซึ่งรายละเอียดแสดงไว้ในภาคผนวก ก. ส่วนรายละเอียดในการสร้างการแจกแจงแบบปกติปลอมปนเป็นดังนี้

สร้างการแจกแจงแบบปกติ โดยใช้วิธีของ Gauss ซึ่งมีค่าเฉลี่ยเป็น 0 และ ความแปรปรวนเป็น 1 ส่วนค่าเฉลี่ยและความแปรปรวนอื่น ที่ใช้สร้างค่าผิดปกติ จะใช้การแปลงข้อมูลในรูป X = SMEAN + (SIGMA) X โดย SMEAN และ $(SIGMA)^2$ คือค่าเฉลี่ยและความแปรปรวนที่ต้องการ สำหรับรายละเอียดแสดงไว้ในภาคผนวก ก การใช้โปรแกรมย่อยนี้ใช้คำสั่ง CALL NORMAL (SMEAN, SIGMA, X) ค่า SMEAN, $(SIGMA)^2$ เป็นค่าพารามิเตอร์ ซึ่งกำหนดค่าเฉลี่ย และความแปรปรวน ซึ่งจะถูกส่งมาจากโปรแกรมหลัก ผลลัพธ์ คือ ค่า X ซึ่งเป็นตัวแปรสุ่มที่มีการแจกแจงปกติ ที่มีค่าเฉลี่ย เป็น SMEAN และความแปรปรวนเป็น $(SIGMA)^2$

3.3.2 การสร้างข้อมูล (X,Y) ที่มีความสัมพันธ์กันเชิงเส้นตรง

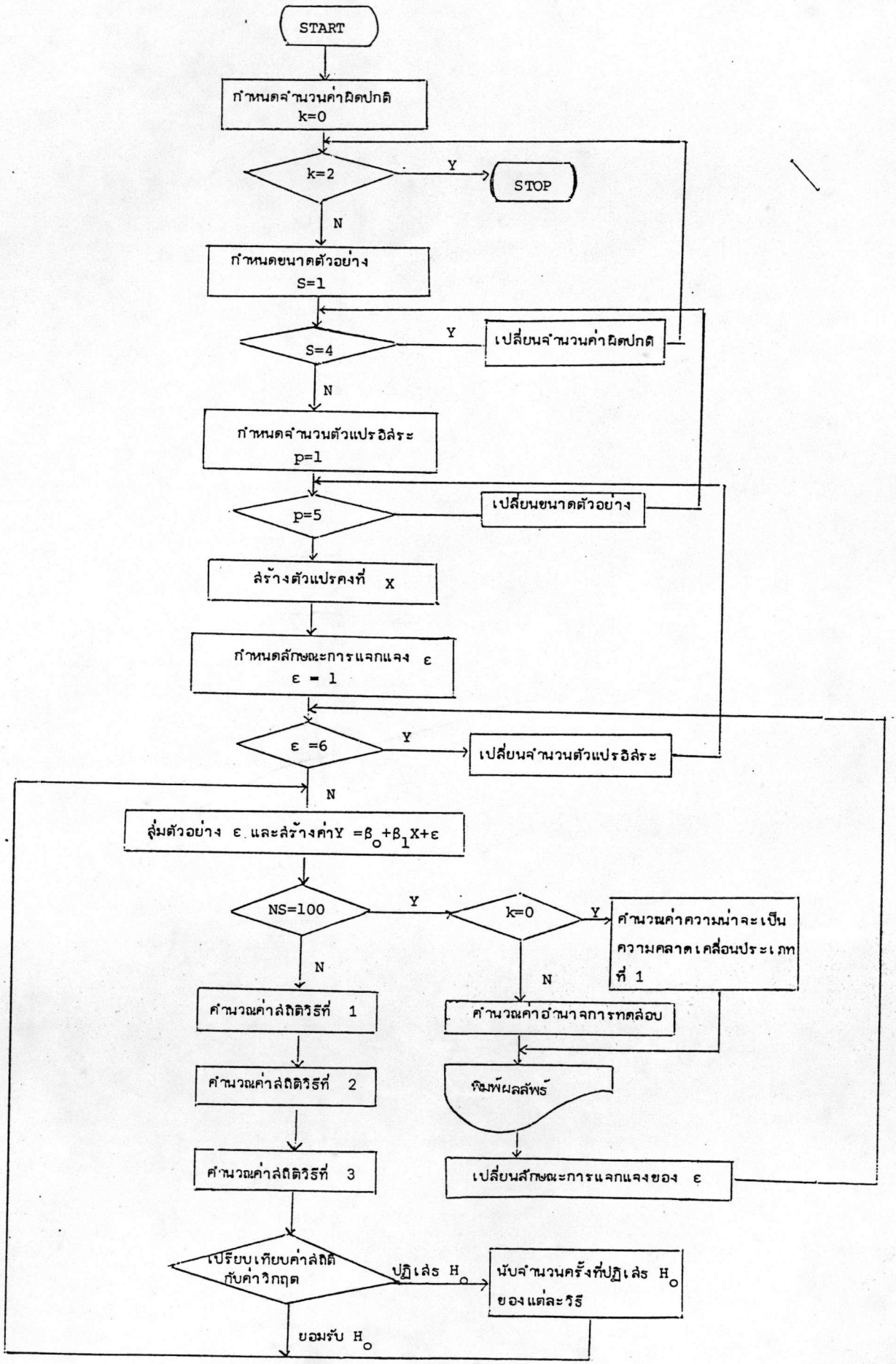
การสร้างข้อมูล (X,Y) ที่มีความสัมพันธ์กันเชิงเส้นตรงนั้น ในการวิจัยครั้งนี้จะทำการสร้างค่า X ซึ่งเป็นค่าคงที่ก่อน แล้วจึงสร้างค่า y ที่มีความสัมพันธ์เชิงเส้นตรงกับ X ตามรูปแบบดังนี้คือ $\underset{\sim}{y} = \beta_o + \beta X + \underset{\sim}{\varepsilon}$ เมื่อ β_o , β เป็นพารามิเตอร์ที่ถูกกำหนดขึ้นมา และ $\underset{\sim}{\varepsilon}$ เป็นความคลาดเคลื่อนที่มีรูปแบบการแจกแจงเป็นแบบปกติ ดังที่กล่าวมาแล้วใน 3.1.1 สำหรับการสร้าง ค่า X จากการแจกแจงแบบปกติที่มีค่าเฉลี่ยเป็น 0 และความแปรปรวนเป็น 1 ในการสร้างข้อมูลนั้นจะเริ่มจากการกำหนดขนาดของตัวอย่าง จำนวนค่าผิดปกติ และจำนวนตัวแปรอิสระ พารามิเตอร์ β_o , β ในการวิจัยนี้ได้สร้างค่าผิดปกติ 1 ค่า และ 2 ค่า โดยการสร้างจากลักษณะการแจกแจงของ $\underset{\sim}{\varepsilon}$ เป็น 2 ลักษณะคือ กรณีที่ค่าผิดปกติเท่ากับ 1 ค่า จะทำการสร้างโดยให้ ε_1 , ε_2 , ..., ε_{n-1} มาจากการแจกแจงแบบปกติที่มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 1 ε_n มาจากการแจกแจงแบบปกติที่มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ σ^2 (ในกรณีที่ค่าผิดปกติมาจากสเกลแฟคเตอร์) และจากการแจกแจงแบบปกติ ที่มีค่าเฉลี่ยเท่ากับ μ และความแปรปรวนเป็น 1 (ในกรณีที่ค่าผิดปกติมาจากโลเคชั่นแฟคเตอร์) ในทำนองเดียวกันกรณีที่ค่าผิดปกติมีจำนวน 2 ค่า ε_1 , ε_2, ... ε_{n-2} จะสร้างจากการแจกแจงแบบปกติ ที่มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 1 ε_{n-1} , ε_n สร้างมาจากการแจกแจงแบบปกติที่มีค่าเฉลี่ยเป็น μ และความแปรปรวนเท่ากับ 1 ตามที่กำหนด (ในกรณีที่เป็นโลเคชั่นแฟคเตอร์และสเกลแฟคเตอร์) ตามลำดับ แล้วจึงสร้างค่า $\underset{\sim}{y}$ ตามรูปแบบความสัมพันธ์ดังกล่าว

3.3.3 การหาค่าอำนาจการทดสอบ และความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 สำหรับการทดสอบสมมติฐานเกี่ยวกับค่าผิดปกติ

เมื่อสร้างข้อมูล (X ,Y) ที่มีความสัมพันธ์เชิงเส้นตรงตามรูปแบบที่ต้องการได้แล้วขั้นต่อไปคือ การทดลอง เพื่อหาอำนาจการทดสอบ และความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 ในการทดลอง จะแบ่งออกเป็น 4 กรณี คือ กรณีที่ค่าผิดปกติมีค่าเท่ากับ 1 (k = 1) และค่าผิดปกติมีค่าเท่ากับ 2 (k = 2) ซึ่งค่าผิดปกติสร้างมาจากสเกลแฟคเตอร์และโลเคชั่นแฟคเตอร์ มีขั้นตอนการคำนวณคือ การคำนวณค่าสถิติทั้ง สามวิธี คือ ประมาณค่าสัมประสิทธิ์การถดถอย ความคลาดเคลื่อน และความคลาดเคลื่อนมาตรฐาน เลือกค่าความคลาดเคลื่อนมาตรฐานที่มีค่าสูงสุด หารด้วยผลต่างของขนาดตัวอย่างกับจำนวนตัวแปรอิสระ จะได้ตัวสถิติทดสอบวิธีของจีแบร์รี่ ส่วนวิธีคุ๊ก จะนำค่าความคลาดเคลื่อนมาตรฐานที่มีค่าสูงสุด ยกกับผลต่างของขนาดตัวอย่าง กับจำนวนตัวแปรอิสระ หารด้วยผลต่าง ของขนาดตัวอย่าง จำนวนตัวแปรอิสระ และค่าความคลาดเคลื่อนมาตรฐานที่มีค่าสูงสุดกำลังสอง $(t_i^2 (n-p-1)/(n-p-t_i^2))$ในการคำนวณค่าสถิติ ของวิธีแอนดูร์และเพรสจิบอน จุดสำคัญ คือ การคำนวณค่าความน่าจะเป็นแน่นอน (Exact probability) แล้วนำมาเปรียบเทียบกับระดับนัยสำคัญ 0.05 0.025 0.005 โดยการคำนวณหาขอบบนของค่าวิกฤต คือเปลี่ยนเมตริกซ์ของตัวแปรอิสระที่บวกตัวแปรตาม 1 สดมภ์ (X^*) และเมตริกซ์ของตัวแปรอิสระ (X) เป็นเมตริกซ์ฉายาตั้งฉาก (Q) จากนั้นคำนวณค่าแฮทเมตริกซ์ (QQ^T) เลือกสมาชิกแนวทะแยงของเมตริกซ์ทั้งสองที่มีค่าต่ำสุด (R_o^k , R_i^k (X)) นำมาหารกัน เป็นค่าขอบบนของค่าวิกฤต เมื่อคำนวณค่าสถิติทั้ง 3 วิธีแล้ว นำค่าสถิติ วิธีของจีแบร์รี่ และวิธีของคุ๊ก มาเปรียบเทียบกับค่าวิกฤต วิธีของแอนดูร์และเพรสจิบอน มาเปรียบเทียบกับระดับนัยสำคัญ เพื่อตัดสินใจว่าจะปฏิเสธ หรือยอมรับสมมติฐานว่า ในกรณีที่ค่าผิดปกติ มี 1 ค่า จะทำการตรวจเพื่อแสดงว่าค่าผิดปกติมีเพียง 1 ค่า โดยมีเงื่อนไขว่า ถ้าค่าสถิติค่าที่หนึ่งมากกว่าค่าวิกฤต ค่าสถิติค่าที่สอง น้อยกว่าค่าวิกฤต แสดงว่าค่าผิดปกติมีเพียง 1 ค่า ในกรณีที่ค่าผิดปกติมี 2 ค่า จะทำการตรวจเพื่อแสดงว่า ค่าผิดปกติมี 2 ค่า โดยมีเงื่อนไข ถ้าค่าสถิติค่าที่สองมีค่ามากกว่าค่าวิกฤต และค่าสถิติค่าที่สามน้อยกว่าค่าวิกฤต แสดงว่าค่าผิดปกติมี 2 ค่า นับจำนวนครั้งที่ทำการตรวจสอบ ที่ตรงตามเงื่อนไข จากนั้นย้อนกลับไปทำการสุ่ม

ตัวอย่างชุดใหม่ จนกระทั่งครบ 100 ครั้ง แล้วคำนวณค่าความน่าจะเป็นความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบจากนั้นก็จะเปลี่ยน สเกลแฟคเตอร์ และโลเคชั่นแฟคเตอร์ เมื่อครบค่าตามต้องการแล้ว ทำการเปลี่ยนจำนวนตัวแปรอิสระ และขนาดตัวอย่างตามลำดับ จนครบรูปแบบที่ต้องการศึกษา ในแต่ละจำนวนตัวแปรอิสระ และขนาดตัวอย่าง ซึ่งสรุปเป็นผังงานได้ดังนี้

รูปที่ 3.1 แสดงผังงานสำหรับการหาค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบ ของสถิติทดสอบ 3 วิธี ในการทดสอบค่าผิดปกติ



3.4 โปรแกรมที่ใช้ในการวิจัย

โปรแกรมที่ใช้ในการวิจัยทั้งหมดเขียนด้วย ภาษาฟอร์แทรนโฟ ซึ่งแสดงรายละเอียดไว้ในภาคผนวก โดยลักษณะการทำงานของโปรแกรม เป็นดังนี้

ตารางที่ 3.2 แสดงลักษณะการทำงานของโปรแกรมทั้งหมดที่ใช้ในการวิจัย

ลำดับที่ของโปรแกรม	ลักษณะการทำงาน	โปรแกรมย่อยที่เรียกใช้
1	การคำนวณค่าอำนาจการทดสอบ หรือความน่าจะเป็นความคลาดเคลื่อนประเภทที่ 1 ในกรณีที่มีจำนวนค่าผิดปกติเป็น 1 และ 2 ของตัวสถิติทั้ง 3 วิธี เมื่อ $\beta_o=10$, $\beta_i=1$,i=1,2,..., p ขนาดตัวอย่างเป็น 20 30 50 70 โดยศึกษา เมื่อ ε มีการแจกแจงเป็นสแกลคอนทามิเนตและโลเคชั่นคอนทามิเนต	โปรแกรมสร้างตัวเลขสุ่ม โปรแกรมการแจกแจงปกติ โปรแกรมการประมาณค่าสัมประสิทธิถดถอย โปรแกรมแฮทเมตริกซ์ โปรแกรมการคำนวณค่าความคลาดเคลื่อนมาตรฐาน โปรแกรมค่าสูงสุด โปรแกรมค่าต่ำสุด โปรแกรมความน่าจะเป็นแน่นอน
2	การคำนวณค่าความน่าจะเป็นความคลาดเคลื่อนประเภทที่ 1 ในกรณีที่มีจำนวนค่าผิดปกติเป็น 1 และ 2 ของตัวสถิติทั้ง 3 วิธี เมื่อ $\beta_o=10$, $\beta_i=1.0$, i=1,2,...,p ขนาดตัวอย่างเป็น 20 30 50 70 เมื่อ ε มีการแจกแจงเป็นแบบปกติ	เหมือนโปรแกรมในข้อ 1