

แนวคิดและทฤษฎีประกอบการเลือกวิธีการวิเคราะห์ทางสถิติ

การทำงานวิจัยในด้านต่างๆ นักวิจัยมักประสบปัญหาในหลายๆ ด้าน ปัญหาหนึ่งที่นักวิจัยจำนวนมากประสบในการทำงานวิจัยคือ ความไม่แน่ใจในการเลือกใช้วิธีการวิเคราะห์ข้อมูล เนื่องจากวิธีการวิเคราะห์ข้อมูลแต่ละวิธีจะเหมาะสมกับวัตถุประสงค์และลักษณะข้อมูลแต่ละประเภท การเลือกใช้วิธีการวิเคราะห์ข้อมูลให้เหมาะสมกับวัตถุประสงค์และลักษณะข้อมูลของงานวิจัยนั้นๆ จึงเป็นสิ่งที่จำเป็นมาก ถ้าหากนักวิจัยเลือกใช้วิธีการวิเคราะห์ข้อมูลที่ไม่ถูกต้องจะมีผลทำให้ผลสรุปของงานวิจัยผิดพลาดได้ ซึ่งจะส่งผลกระทบต่อคุณภาพของงานวิจัย

ดังนั้น นักวิจัยจึงมีความจำเป็นที่จะต้องรู้ว่าวัตถุประสงค์และลักษณะข้อมูลที่นักวิจัยมีอยู่ ควรที่จะเลือกใช้วิธีการวิเคราะห์ทางสถิติด้วยวิธีใด ก่อนที่จะทำการวิจัย จากเหตุผลดังกล่าว จึงมีแนวความคิดที่จะพัฒนาระบบสนับสนุนการตัดสินใจเลือกวิธีการวิเคราะห์ทางสถิติ เพื่อสนองความต้องการของนักวิจัยหรือผู้ที่สนใจ โดยแบ่งกลุ่มผู้ใช้ออกเป็น

1. กลุ่มผู้ใช้ที่จะทำการวิจัย โดยมีวัตถุประสงค์ของการวิจัยแล้ว แต่ยังไม่ได้กำหนดประเภทของการวิเคราะห์ทางสถิติ *

ระบบที่เป็นประโยชน์สำหรับผู้ใช้กลุ่มนี้คือ ระบบที่เชื่อมโยงวัตถุประสงค์ของงานวิจัยกับประเภทของการวิเคราะห์ทางสถิติ ดังนั้นระบบที่พัฒนาขึ้นสำหรับผู้ใช้กลุ่มนี้จึงประกอบด้วย ส่วนที่แสดงวัตถุประสงค์ของการวิจัย ตัวอย่างเช่น ผู้ใช้ต้องการเปรียบเทียบวิธีการใหม่กับวิธีการเก่า ผู้ใช้ต้องการเปรียบเทียบผลก่อนการทดลองกับหลังการทดลอง ผู้ใช้ต้องการศึกษาว่าตัวแปรหรือปัจจัยใดบ้างที่มีความสัมพันธ์กัน มากน้อยแค่ไหน และเป็นไปในทิศทางใด เป็นต้น เพื่อให้ผู้ใช้เลือกว่าวัตถุประสงค์ที่ผู้ใช้มีอยู่ตรงกับกรณีใด โดยระบบจะช่วยเสนอแนะว่าวัตถุประสงค์ของงานวิจัยของผู้ใช้เข้าข่ายประเภทของการวิเคราะห์ทางสถิติประเภทใดและควรที่จะใช้ระบบสนับสนุนการตัดสินใจเลือกวิธีการวิเคราะห์ทางสถิติ ทางด้านใด และพัฒนาโดยใคร

* ประเภทของการวิเคราะห์ทางสถิติ หมายถึง ประเภทปัญหาของการวิเคราะห์ เช่น การทดสอบสมมติฐาน การวิเคราะห์ความแปรปรวนและการวิเคราะห์ความถดถอย เป็นต้น

2. กลุ่มผู้ใช้ที่กำหนดประเภทของการวิเคราะห์ทางสถิติแล้ว แต่ยังไม่ได้กำหนดวิธีการวิเคราะห์ทางสถิติ *

ระบบที่เป็นประโยชน์สำหรับผู้ใช้งานคือ ระบบจะเสนอแนะให้กับผู้ใช้ว่ามีระบบสนับสนุนการตัดสินใจเลือกวิธีการวิเคราะห์ทางสถิติในด้านใดบ้างที่ได้ทำการพัฒนาไว้ และพัฒนาโดยใคร ดังนั้นระบบที่พัฒนาขึ้นสำหรับผู้ใช้งานนี้จึงประกอบด้วย ส่วนที่แสดงประเภทของการวิเคราะห์ทางสถิติ เพื่อให้ผู้ใช้เลือกประเภทของการวิเคราะห์ทางสถิติให้ตรงกับที่ผู้ใช้กำหนดไว้แล้ว และมีตัวอย่างประกอบการเลือกประเภทของการวิเคราะห์ทางสถิติแต่ละประเภทด้วย

2.1 แนวคิดและทฤษฎีประกอบการเลือกวิธีการวิเคราะห์ทางสถิติในด้านต่างๆ

การพัฒนาระบบสนับสนุนการตัดสินใจเลือกวิธีการวิเคราะห์ทางสถิติในด้านต่างๆ ** มีแนวคิดและทฤษฎีประกอบการเลือกวิธีการวิเคราะห์ทางสถิติ ดังต่อไปนี้

2.1.1 สำหรับกลุ่มผู้ใช้ที่จะทำการวิจัย โดยมีวัตถุประสงค์ของการวิจัยแล้ว แต่ยังไม่ได้กำหนดประเภทของการวิเคราะห์ทางสถิติ ในขั้นตอนแรกได้นำ วัตถุประสงค์ของงานวิจัย ในด้านต่างๆ มาเป็นเกณฑ์ในการกำหนดประเภทของการวิเคราะห์ทางสถิติ เนื่องจากวัตถุประสงค์ของงานวิจัยสามารถบอถึงประเภทของการวิเคราะห์ทางสถิติที่เหมาะสมได้ ***

จากข้อกำหนดข้างต้นสามารถพัฒนาระบบโดยใช้คำเฉพาะ (Keywords) ที่ได้จากการเก็บรวบรวมวัตถุประสงค์ของงานวิจัยในด้านต่างๆ มาเป็นแนวทางในการกำหนดทางเลือก ดังต่อไปนี้

* วิธีการวิเคราะห์ทางสถิติ ตัวอย่างเช่น วิธีการทดสอบไคสแควร์ วิธีการกำลังสองน้อยที่สุด เป็นต้น

** พัฒนาโดย น.ส. ชลธิชา ศรีนาคา น.ส. สุภาเพ็ญ คุณแสง น.ส. สายัน เกื้อสกุล

*** วัตถุประสงค์ของงานวิจัยสามารถบอถึงประเภทของการวิเคราะห์ทางสถิติได้ เพราะจากการศึกษาและรวบรวมวัตถุประสงค์ของงานวิจัยในสาขาต่างๆ พบว่าวัตถุประสงค์และการเลือกใช้ประเภทของการวิเคราะห์ทางสถิติของงานวิจัยนั้นๆ มีความเหมาะสมเมื่อเทียบกับข้อกำหนดและทฤษฎีบทของวิธีการวิเคราะห์ทางสถิติ

ก. ประเภทของการวิเคราะห์ทางสถิติด้านการประมาณค่า ใช้คำเฉพาะแสดง
วัตถุประสงค์ ดังต่อไปนี้คือ

1. ประมาณค่าต่างๆ เช่น ค่าเฉลี่ย ค่าสัดส่วน ค่าความแปรปรวน
2. ประมาณค่าความแตกต่างของค่าเฉลี่ย ค่าสัดส่วนและค่าความแปรปรวน

ข. ประเภทของการวิเคราะห์ทางสถิติด้านการทดสอบสมมติฐาน ใช้คำเฉพาะ
แสดงวัตถุประสงค์ ดังต่อไปนี้คือ

1. เปรียบเทียบวิธีการใหม่กับวิธีการเก่า
2. เปรียบเทียบผลก่อนการทดลองกับหลังการทดลอง
3. เปรียบเทียบวิธีการ 2 วิธีการว่าแตกต่างกันหรือไม่
4. เปรียบเทียบวิธีการมากกว่า 2 วิธีการขึ้นไปว่าแตกต่างกันหรือไม่
5. ทดสอบว่าค่าของประชากรจะเท่ากับค่าที่กำหนดไว้หรือไม่
6. ทดสอบว่าค่าของประชากรหนึ่งจะเท่ากับอีกประชากรหนึ่งหรือไม่
7. ทดสอบว่าค่าของประชากรมากกว่า 2 ประชากรจะเท่ากันหรือไม่
8. ทดสอบค่าเฉลี่ย ค่าสัดส่วน ค่าความแปรปรวนของข้อมูล
9. ทดสอบว่าค่าของประชากรมีรูปร่าง ลักษณะหรือการแจกแจงแบบใด

ค. ประเภทของการวิเคราะห์ทางสถิติด้านการวางแผนการทดลอง และการ
วิเคราะห์ความแปรปรวน ใช้คำเฉพาะแสดงวัตถุประสงค์ ดังต่อไปนี้คือ เปรียบเทียบวิธีการ
มากกว่า 2 วิธีการขึ้นไปว่าแตกต่างกันหรือไม่

ง. ประเภทของการวิเคราะห์ทางสถิติด้านการหาและการทดสอบความสัมพันธ์
การทดสอบความเป็นอิสระระหว่างตัวแปร ใช้คำเฉพาะแสดงวัตถุประสงค์ ดังต่อไปนี้คือ

1. ศึกษาว่าตัวแปรหรือปัจจัยใดบ้างที่มีความสัมพันธ์กัน สัมพันธ์กันมาก
น้อยแค่ไหนและเป็นไปในทิศทางใด
2. ศึกษาว่าตัวแปรหรือปัจจัยแต่ละตัว เป็นอิสระต่อกันหรือไม่
3. ศึกษาความสัมพันธ์ของตัวแปรหรือปัจจัยระหว่างกลุ่มหนึ่งกับอีกกลุ่มหนึ่ง
4. ศึกษาว่าเมื่อตัวแปร หรือปัจจัยตัวหนึ่งมีค่า เปลี่ยน ไปอีกตัวจะ เปลี่ยนไป
อย่างไร

5. ศึกษาว่าตัวแปรหรือปัจจัยตัวหนึ่งมีค่าเปลี่ยนแปลงไปอย่างไร เมื่อตัวแปรอื่นๆ มีค่าเปลี่ยนแปลงไป
6. ศึกษาตัวแปรหรือปัจจัยตัวหนึ่งขึ้นอยู่กับตัวแปรอีกตัวหนึ่งหรืออีกหลายๆ ตัวหรือไม่

จ. ประเภทของการวิเคราะห์ทางสถิติด้านการวิเคราะห์ความถดถอย ใช้คำเฉพาะแสดงวัตถุประสงค์ ดังต่อไปนี้คือ

1. ศึกษาว่าตัวแปรหรือปัจจัยใดบ้างที่มีความสัมพันธ์กัน สัมพันธ์กันมากน้อยแค่ไหนและ เป็นไปในทิศทางใด
2. ศึกษาว่าเมื่อตัวแปรหรือปัจจัยตัวหนึ่งมีค่าเปลี่ยนแปลงไปอีกตัวจะเปลี่ยนแปลงไปอย่างไร
3. ศึกษาว่าตัวแปรหรือปัจจัยตัวหนึ่งมีค่าเปลี่ยนแปลงไปอย่างไรเมื่อตัวแปรอื่น ๆ มีค่าเปลี่ยนแปลงไป
4. ศึกษาตัวแปรหรือปัจจัยตัวหนึ่งขึ้นอยู่กับตัวแปรอีกตัวหนึ่งหรืออีกหลายๆ ตัวหรือไม่
5. ศึกษารูปแบบความสัมพันธ์ระหว่างตัวแปรตัวหนึ่งหรือชุดหนึ่งกับตัวแปรอีกตัวหนึ่งหรืออีกชุดหนึ่ง

ฉ. ประเภทของการวิเคราะห์ทางสถิติด้านการพยากรณ์ ใช้คำเฉพาะแสดงวัตถุประสงค์ ดังต่อไปนี้คือ นำข้อมูลในอดีตมาใช้คาดการณ์ค่าในอนาคต

ช. ประเภทของการวิเคราะห์ทางสถิติด้านการหาสมการ รูปแบบหรือตัวแบบ ใช้คำเฉพาะแสดงวัตถุประสงค์ ดังต่อไปนี้คือ นำข้อมูลมาหาสมการ รูปแบบหรือตัวแบบ

ซ. ประเภทของการวิเคราะห์ทางสถิติ ด้านการทดสอบรูปแบบของประชากร ใช้คำเฉพาะแสดงวัตถุประสงค์ ดังต่อไปนี้คือ ทดสอบว่าค่าของประชากรมีรูปร่าง ลักษณะหรือการแจกแจงแบบใด

ฅ. ประเภทของการวิเคราะห์ทางสถิติ ด้านการจัดกลุ่ม แบ่งกลุ่ม จำแนกกลุ่มตัวแปร ใช้คำเฉพาะแสดงวัตถุประสงค์ ดังต่อไปนี้คือ นำข้อมูลมาจัดกลุ่ม แบ่งกลุ่ม จำแนกกลุ่ม

ญ. ประเภทของการวิเคราะห์ทางสถิติด้านการวิเคราะห์องค์ประกอบ ใช้คำเฉพาะแสดงวัตถุประสงค์ ดังต่อไปนี้คือ จัดองค์ประกอบของตัวแปร

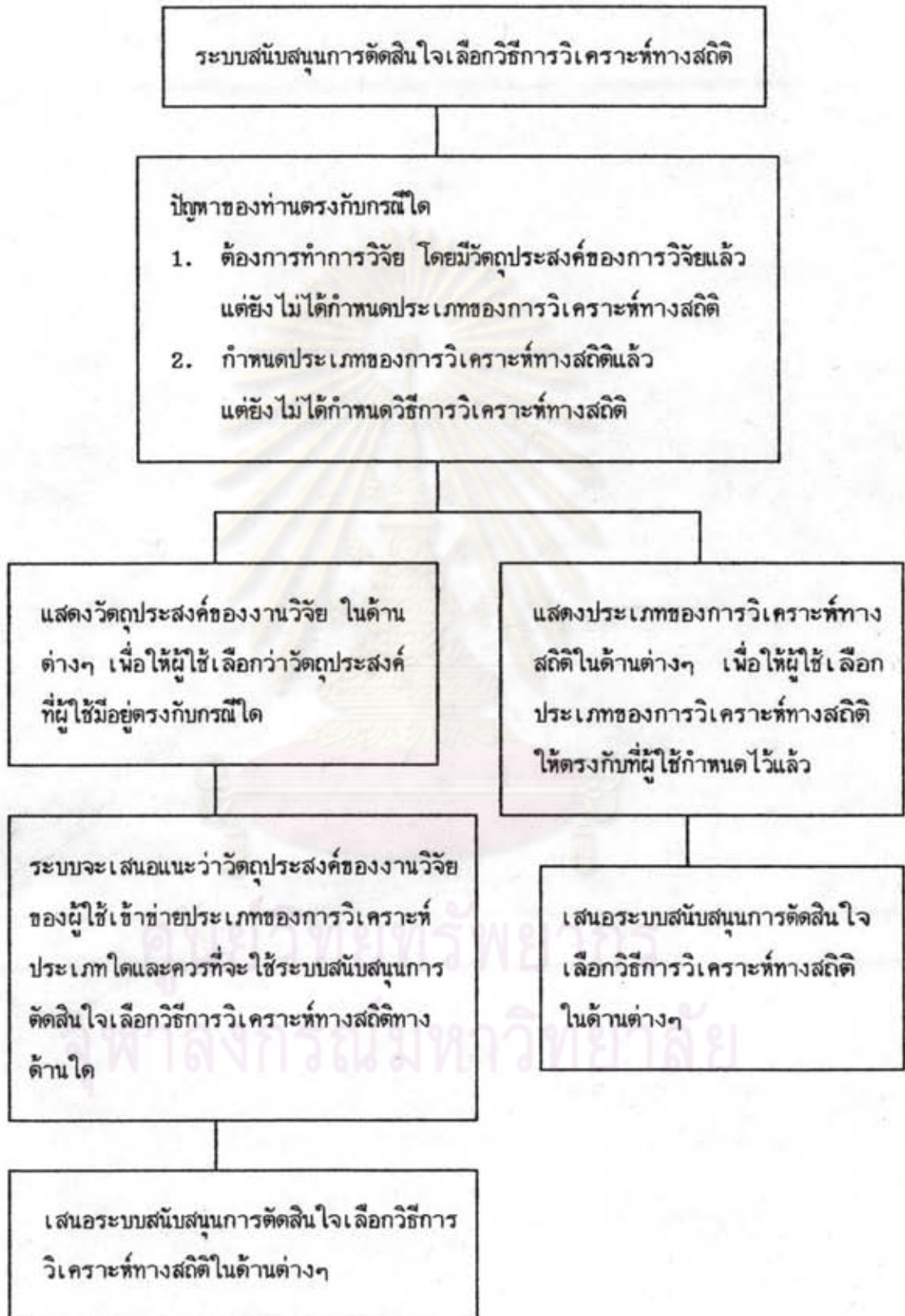
ดังนั้นจึงใช้คำเฉพาะดังกล่าวนี้ มาเป็นเกณฑ์ในการพิจารณาแบ่งประเภทของการวิเคราะห์ทางสถิติ

2.1.2 สำหรับกลุ่มผู้ใช้ที่กำหนดประเภทของการวิเคราะห์ทางสถิติแล้ว แต่ยังไม่ได้กำหนดวิธีการวิเคราะห์ทางสถิติ ในขั้นตอนแรกแสดง ประเภทของการวิเคราะห์ทางสถิติ ในด้านต่างๆ เพื่อให้ผู้ใช้เลือกประเภทของการวิเคราะห์ทางสถิติให้ตรงกับที่ผู้ใช้กำหนดไว้ แล้วระบบจะเสนอแนะระบบสนับสนุนการตัดสินใจเลือกวิธีการวิเคราะห์ทางสถิติในด้านต่างๆ ที่ได้ทำการพัฒนาไว้ เพื่อให้ผู้ใช้สามารถที่จะ ไปใช้ระบบสนับสนุนการตัดสินใจเลือกวิธีการวิเคราะห์ทางสถิติในด้านที่ตรงกับที่ผู้ใช้ต้องการ ซึ่งระบบที่ได้พัฒนาขึ้นนี้ได้เสนอประเภทของการวิเคราะห์ทางสถิติในด้านต่างๆ ดังต่อไปนี้

- ก. ประเภทของการวิเคราะห์ทางสถิติด้านการประมาณค่า
- ข. ประเภทของการวิเคราะห์ทางสถิติด้านการทดสอบสมมติฐาน
- ค. ประเภทของการวิเคราะห์ทางสถิติด้านการวางแผนการทดลองและการวิเคราะห์ความแปรปรวน
- ง. ประเภทของการวิเคราะห์ทางสถิติด้านการหาและการทดสอบความสัมพันธ์ การทดสอบความเป็นอิสระระหว่างตัวแปร
- จ. ประเภทของการวิเคราะห์ทางสถิติด้านการวิเคราะห์ความถดถอย
- ฉ. ประเภทของการวิเคราะห์ทางสถิติด้านการพยากรณ์
- ช. ประเภทของการวิเคราะห์ทางสถิติด้านการหาสมการ รูปแบบหรือตัวแบบ
- ซ. ประเภทของการวิเคราะห์ทางสถิติด้านการทดสอบรูปแบบของประชากร
- ณ. ประเภทของการวิเคราะห์ทางสถิติด้านการจัดกลุ่ม แบ่งกลุ่ม จำแนกกลุ่มตัวแปร
- ญ. ประเภทของการวิเคราะห์ทางสถิติด้านการวิเคราะห์องค์ประกอบ

ดังนั้นจึงนำประเภทของการวิเคราะห์ทางสถิติดังกล่าวนี้มาแสดง เพื่อให้ผู้ใช้เลือกประเภทของการวิเคราะห์ทางสถิติให้ตรงกับที่ผู้ใช้กำหนดไว้แล้ว

จากที่กล่าวมาข้างต้นสามารถแสดง เป็นผังงาน ได้ดังต่อไปนี้



ดังนั้นผู้ใช้ทั้งสองกลุ่มจะสามารถที่จะใช้ระบบสนับสนุนการตัดสินใจเลือกวิธีการวิเคราะห์ทางสถิติต่าง ๆ ต่อไปได้ ซึ่งวิทยานิพนธ์ฉบับนี้จะกล่าวถึงเฉพาะการพัฒนาระบบทางด้าน การหาความสัมพันธ์และการวิเคราะห์ความถดถอยเท่านั้น ส่วนทางด้าน การทดสอบสมมติฐานและการวางแผนการทดลองและการวิเคราะห์ความแปรปรวนอยู่ในวิทยานิพนธ์ของ น.ส.สาธิต เกื้อสกุล และ น.ส.สุภาเพ็ญ คุณแสง ตามลำดับ

2.2 แนวคิดและทฤษฎีประกอบการเลือกวิธีการวิเคราะห์ทางสถิติ ด้านการหาความสัมพันธ์และการวิเคราะห์ความถดถอย

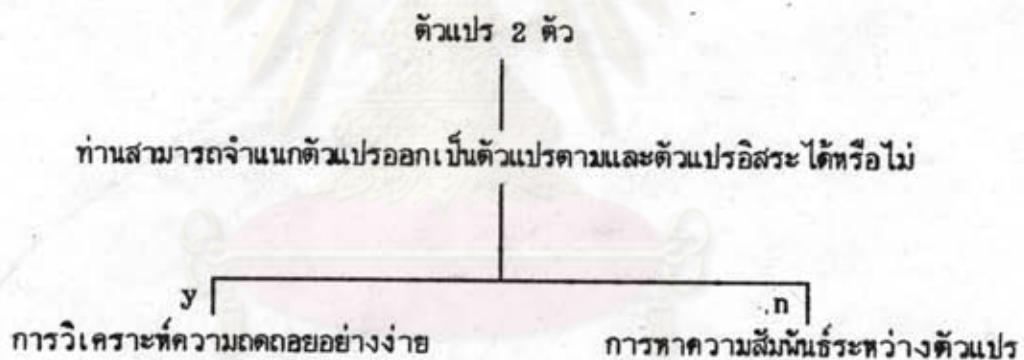
การพัฒนาระบบสนับสนุนการตัดสินใจเลือกวิธีการวิเคราะห์ทางสถิติ ด้านการหาความสัมพันธ์และการวิเคราะห์ความถดถอย มีแนวคิดและทฤษฎีประกอบการเลือกวิธีการวิเคราะห์ทางสถิติ ดังต่อไปนี้

ขั้นตอนแรกได้นำ จำนวนตัวแปร มาเป็นเกณฑ์ในการพิจารณาแบ่งวิธีการวิเคราะห์ทางสถิติอย่างกว้างๆ เนื่องจากการหาความสัมพันธ์ระหว่างตัวแปรและการวิเคราะห์ความถดถอยนั้นจะต้องมีจำนวนตัวแปรในการวิเคราะห์ข้อมูลตั้งแต่ 2 ตัวขึ้นไป สำหรับกรณีจำนวนตัวแปร 2 ตัว จะใช้วิธีการวิเคราะห์ทางสถิติที่แตกต่างจากกรณีที่มีจำนวนตัวแปรมากกว่า 2 ตัว ตัวอย่างเช่น การวิเคราะห์ความถดถอย กรณีจำนวนตัวแปร 2 ตัว เป็นการวิเคราะห์ความถดถอยอย่างง่าย (Simple Regression) ส่วนกรณีที่มีจำนวนตัวแปรมากกว่า 2 ตัว เป็นการวิเคราะห์ความถดถอยพหุ (Multiple Regression) เป็นต้น

ดังนั้นจึงใช้จำนวนตัวแปรมาเป็นเกณฑ์ในการพิจารณาแบ่งวิธีการวิเคราะห์ทางสถิติในขั้นตอนแรก โดยแบ่งออกเป็นกรณีจำนวนตัวแปร 2 ตัว และกรณีจำนวนตัวแปรมากกว่า 2 ตัว ดังต่อไปนี้

2.2.1 กรณีจำนวนตัวแปร 2 ตัว

ขั้นที่ 1 เนื่องจากการวิเคราะห์ความถดถอยอย่างง่าย เป็นเทคนิคทางสถิติที่ใช้ในการวิเคราะห์หาความสัมพันธ์ระหว่างตัวแปร ซึ่งตัวแปรตัวหนึ่งเป็นตัวแปรตามและตัวแปรอีกตัวหนึ่งเป็นตัวแปรอิสระ ส่วนการหาความสัมพันธ์ระหว่างตัวแปรเป็นการศึกษาถึงความสัมพันธ์ระหว่างตัวแปรที่สนใจว่ามีความสัมพันธ์กันหรือไม่ มากน้อยเพียงใดและความสัมพันธ์เป็นไปในทิศทางใด¹ ดังนั้นจึงนำ การจำแนกตัวแปรออกเป็นตัวแปรตามและตัวแปรอิสระ มาเป็นเกณฑ์ในการพิจารณาแบ่งวิธีการวิเคราะห์ทางสถิติ เพราะการวิเคราะห์ความถดถอยสามารถที่จะจำแนกตัวแปรออกเป็นตัวแปรตามและตัวแปรอิสระได้ ส่วนการหาความสัมพันธ์ระหว่างตัวแปรไม่สามารถจำแนกตัวแปรออกเป็นตัวแปรตามและตัวแปรอิสระได้ ซึ่งแสดงเป็นผังงานได้ดังต่อไปนี้



ทางด้านกรวิเคราะห์ความถดถอยอย่างง่ายได้ทำการพัฒนาต่อไปในขั้นที่ 2 ส่วนทางด้านกรหาความสัมพันธ์นั้นได้ทำการพัฒนาต่อไปในขั้นที่ 3 ดังต่อไปนี้

¹ Draper, N.R. and H. Smith, Applied Regression Analysis, 1981, pp. 1-7, 33-35.

ขั้นที่ 2 ในขั้นตอนนี้ได้นำ ลักษณะข้อมูล ที่เป็นเชิงปริมาณหรือเชิงคุณภาพมาเป็นเกณฑ์ในการพิจารณาแบ่งวิธีการวิเคราะห์ทางสถิติ เพราะลักษณะข้อมูลที่แตกต่างกันจะใช้วิธีการวิเคราะห์ทางสถิติที่แตกต่างกันออกไป

จากข้อกำหนดข้างต้นสามารถพัฒนาระบบให้มีแนวทางเลือกดังต่อไปนี้

ก. ลักษณะข้อมูลที่ใช้ในการวิเคราะห์ความถดถอยอย่างง่าย (Simple Regression) ตัวแปรตามเป็นข้อมูลเชิงปริมาณหรือข้อมูลเชิงคุณภาพ ตัวแปรอิสระเป็นข้อมูลเชิงปริมาณ²

ข. ลักษณะข้อมูลที่ใช้ในการวิเคราะห์ความแปรปรวน (ANOVA) ตัวแปรตามเป็นข้อมูลเชิงปริมาณ ตัวแปรอิสระเป็นข้อมูลเชิงคุณภาพ³

ค. ลักษณะข้อมูลที่ใช้ในการวิเคราะห์ความถดถอยทวิ (Binary Regression) ตัวแปรตามเป็นข้อมูลเชิงคุณภาพที่มีค่า 0 หรือ 1 ตัวแปรอิสระเป็นข้อมูลเชิงปริมาณ⁴

ง. ลักษณะข้อมูลที่ใช้ในการทดสอบไคสแควร์ (χ^2 - test) และแบบจำลองลอกเชิงเส้น (Log-linear Model) ตัวแปรตามและตัวแปรอิสระเป็นข้อมูลเชิงคุณภาพ⁵

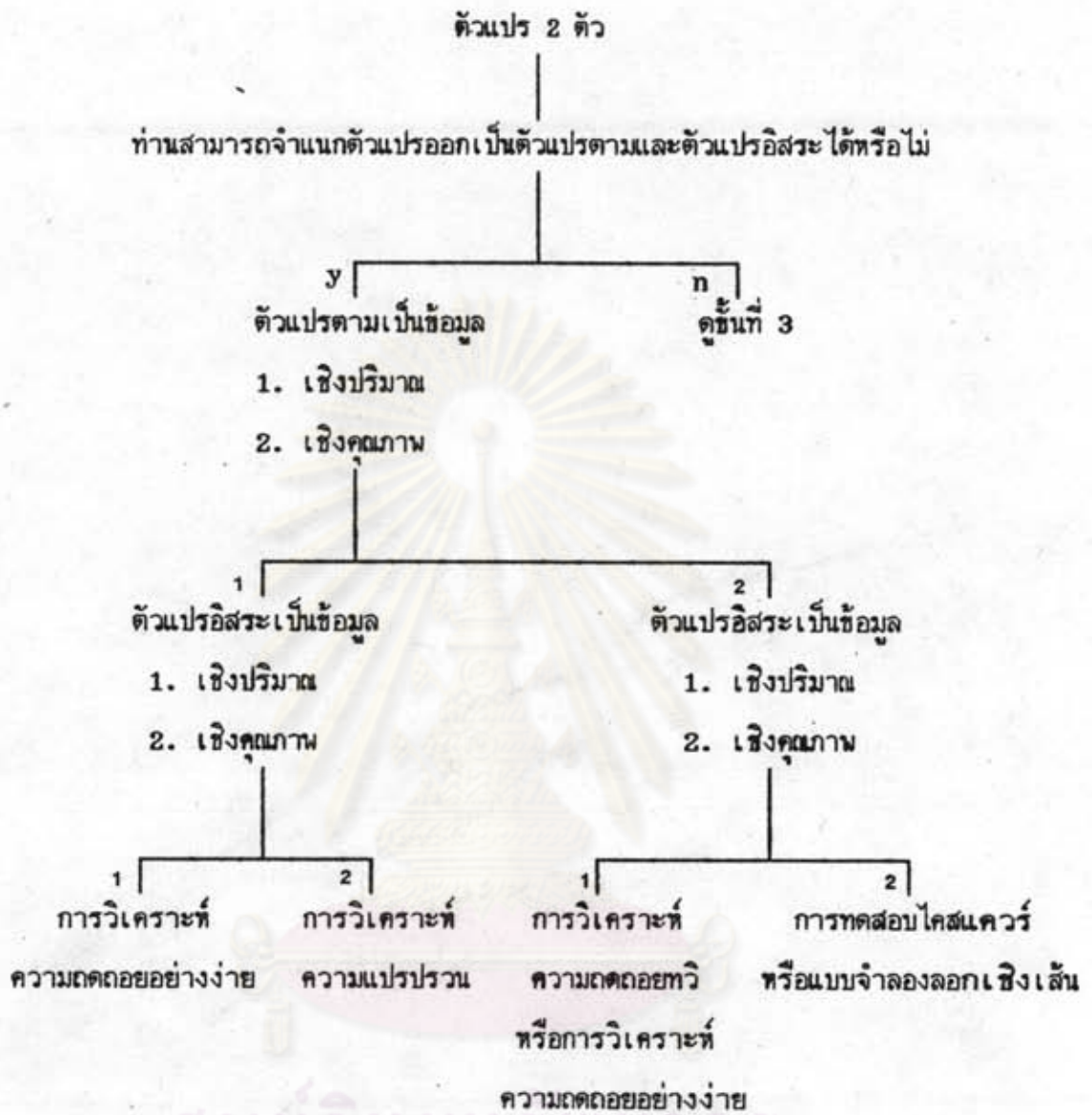
ดังนั้นจึงใช้ข้อกำหนดนี้มาเป็นเกณฑ์ในการพิจารณาแบ่งวิธีการวิเคราะห์ทางสถิติ ดังแสดงเป็นผังงานได้ดังต่อไปนี้

² Draper, N.R. and H. Smith, Applied Regression Analysis, 1981, pp. 1-7.

³ Cochran and Cox, Experimental Designs, 1957, pp. 1-14.

⁴ Judge, G.G., et. al. Introduction to the Theory and Practice of Econometrics, 1982, pp. 517-528.

⁵ Dillon, W.R. and Goldstein, M., Multivariate Analysis, 1984, pp. 302-336.



สำหรับวิทยานิพนธ์ฉบับนี้ ได้ทำการพัฒนาในรายละเอียดเฉพาะ ทางด้านการวิเคราะห์ ความตดถอยอย่างง่าย ดังแสดงในหัวข้อ 2.2.1.1 ส่วนทางด้าน การวิเคราะห์ความแปรปรวน การวิเคราะห์ความตดถอยทวิ การทดสอบไคสแควร์และแบบจำลองลอกเชิงเส้น มิได้ทำการพัฒนาไป ในรายละเอียด จึงได้เสนอเพียงกรณีศึกษาของการวิเคราะห์ดังกล่าวนี้เท่านั้น

2.2.1.1 แนวคิดและทฤษฎีด้านการวิเคราะห์ความถดถอยอย่างง่าย

การวิเคราะห์ความถดถอยอย่างง่าย (Simple Regression) เป็นเทคนิคทางสถิติที่ใช้ในการวิเคราะห์หาความสัมพันธ์ระหว่างตัวแปร 2 ตัว ซึ่งตัวแปรตัวหนึ่งเป็นตัวแปรตามและตัวแปรอีกตัวหนึ่งเป็นตัวแปรอิสระ โดย

ตัวแปรตามคือ ตัวแปรที่เราสนใจศึกษา ใช้สัญลักษณ์ y

ตัวแปรอิสระคือ ตัวแปรที่มีอิทธิพลต่อการเปลี่ยนแปลงของตัวแปรที่เราสนใจศึกษา ใช้สัญลักษณ์ x

สามารถเขียนรูปแบบฟังก์ชันได้เป็น $y = f(x) + \varepsilon$

ตัวแบบความถดถอยเชิงเส้นอย่างง่าย เขียนได้เป็น

$$y_i = \beta_0 + \beta_1 x_{i1} + \varepsilon_i \quad ; i = 1, 2, 3, \dots, n$$

หรือเขียนในรูปเมทริกซ์ได้เป็น $\tilde{y} = \tilde{x}\tilde{\beta} + \tilde{\varepsilon}$

$$\begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} \\ 1 & x_{21} \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & x_{n1} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \cdot \\ \varepsilon_n \end{bmatrix}$$

เมื่อ \tilde{y} คือเวกเตอร์ของตัวแปรตาม y ที่มีขนาด $n \times 1$

$\tilde{\beta}$ คือเวกเตอร์ของสัมประสิทธิ์ของความถดถอย ที่มีขนาด 2×1

\tilde{x} คือเมทริกซ์ของตัวแปรอิสระ ที่มีขนาด $n \times 2$

$\tilde{\varepsilon}$ คือเวกเตอร์ของความคลาดเคลื่อน ที่มีขนาด $n \times 1$

เนื่องจากการกำหนดความสัมพันธ์ $y = f(x) + \varepsilon$ ซึ่งนักวิจัยจำเป็นต้องกำหนดตัวแปรอิสระให้ครบถ้วน ในทางปฏิบัตินักวิจัยไม่สามารถกระทำได้นอกจากนี้ยังมีความผิดพลาดอื่นๆ ร่วมส่งผลกระทบต่ออยู่ด้วย ทำให้ค่าของ y ผิดพลาดไปจากความเป็นจริง ด้วยเหตุนี้จึงจัดให้ความผิดพลาดจากแหล่งต่างๆ รวมอยู่ในค่าความคลาดเคลื่อน ε_i และควบคุมด้วยข้อตกลงของสมการถดถอย^๖ ดังนี้

1. ε_i เป็นตัวแปรเชิงสุ่ม

2. $E(\varepsilon_i) = 0$; $i = 1, 2, 3, \dots, n$

ค่าคาดหวังของค่าความคลาดเคลื่อนมีค่าเท่ากับ 0

3. $E(\varepsilon_i \varepsilon_j) = 0$; $i \neq j$

ค่าความคลาดเคลื่อนในต่างวาระกันเป็นอิสระต่อกัน

4. $E(\varepsilon_i \varepsilon_j) = \sigma^2$; $i = j$

ค่าความคลาดเคลื่อนมีความแปรปรวนคงที่เท่ากับ σ^2

5. $\varepsilon_i \sim N(0, \sigma^2)$

ค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติมีค่าเฉลี่ยเท่ากับ 0 ความแปรปรวนเท่ากับ σ^2 .

ขั้นตอนที่สำคัญของการวิเคราะห์ความถดถอยคือ การประมาณค่า ส.ป.ส. ความถดถอย โดยวิธีใดวิธีหนึ่งซึ่งเห็นว่าเป็นวิธีที่เหมาะสมกับลักษณะข้อมูล โดยปกติมักเลือกใช้วิธีกำลังสองน้อยที่สุด (Ordinary Least Square : OLS) ซึ่งเป็นวิธีที่ทำให้ผลรวมกำลังสองของความคลาดเคลื่อน (Sum Square Error : SSE) มีค่าต่ำที่สุด และทฤษฎีของ Gauss - Markov กล่าวไว้ว่า "การประมาณค่า ส.ป.ส. ความถดถอยด้วยวิธี OLS จะได้ตัวประมาณค่าที่มีคุณสมบัติเป็นตัวประมาณที่ไม่เอนเอียงเชิงเส้นที่ดีที่สุด (Best Linear Unbiased Estimator (BLUE))"⁷

^๖ Draper, N.R. and H. Smith, Applied Regression Analysis, 1981, p. 17.

⁷ Wonnacott, R.J. and Wonnacott, T.H., Regression, 1981, pp. 31-32.

ในทางปฏิบัติบ่อยครั้งที่พบว่าข้อมูลที่นำมาวิเคราะห์นั้นมีความผิดปกติไม่เป็นไปตามข้อตกลงของสมการถดถอย ก่อให้เกิดปัญหาต่างๆ ดังต่อไปนี้

1. ปัญหาความคลาดเคลื่อนมิได้มีการแจกแจงแบบปกติ (Nonnormal Error) คือสถานการณ์ที่ตัวแปรสุ่ม มีการแจกแจงแบบอื่นๆ ที่มีใช้การแจกแจงแบบปกติ ผลกระทบที่ตามมาคือ $\hat{\beta} = (x'x)^{-1} x'y$ ที่ประมาณได้ด้วยวิธี OLS ยังคงมีความผิดปกติเป็นตัวประมาณที่ไม่เอนเอียงเชิงเส้นที่ดีที่สุดอยู่เช่นเดิม แต่ $\hat{\beta}$ และ \hat{y} ที่ประมาณได้จะมิได้มีการแจกแจงแบบปกติ รวมทั้งไม่อาจสร้างเขตความเชื่อมั่นซึ่งอ้างอิงการแจกแจงแบบปกติได้^๘

2. ปัญหาความคลาดเคลื่อนมีอัตสหสัมพันธ์ (Autocorrelation) คือสถานการณ์ที่เกิดความสัมพันธ์ระหว่าง ε_i กับ ε_j กล่าวคือ $E(\varepsilon_i \varepsilon_j) \neq 0 ; i \neq j$ ผลกระทบที่ตามมาคือ $\hat{\beta}$ ที่ประมาณได้ด้วยวิธี OLS ยังคงเป็นตัวประมาณค่าที่ไม่เอนเอียง แต่จะมีประสิทธิภาพต่ำกว่าวิธีอื่นๆ และโดยเฉพาะอย่างยิ่งถ้าข้อมูลเกิดปัญหาอัตสหสัมพันธ์ทางบวก (Positive Autocorrelation) ค่าประมาณของ σ^2 จะมีค่าต่ำกว่าความเป็นจริง ทำให้ $V(\hat{\beta})$ ต่ำกว่าความเป็นจริงไปด้วย เนื่องจากการประมาณค่า β โดยวิธี OLS นั้น $V(\hat{\beta}) = \sigma^2 (x'x)^{-1}$ ซึ่ง $V(\hat{\beta})$ มีค่าต่ำจึงเป็นตัวเลขที่ลวงตา หากนักวิจัยไม่ระมัดระวังอาจเข้าใจผิดคิดว่างานวิจัยนั้นๆ มีคุณภาพสูง ทั้งๆ ที่ความจริงมิได้เป็นเช่นนั้น^๙

3. ปัญหาความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ (Heteroscedasticity) กล่าวคือ $E(\varepsilon_i \varepsilon_j) \neq \sigma^2 ; i = j$ เมื่อความคลาดเคลื่อนมีความแปรปรวนไม่คงที่จะก่อให้เกิดผลกระทบคือเราไม่อาจประมาณค่า ส.ป.ส. ความถดถอยด้วยวิธี OLS ได้ ตลอดจนการตรวจสอบนัยสำคัญของ ส.ป.ส. ความถดถอย แต่ถ้ายังคงเห็นใช้วิธี OLS ก็จะมีพบว่า $\hat{\beta}$ ที่ประมาณได้ยังคงเป็นตัวประมาณค่าที่ไม่เอนเอียง แต่จะไม่มีประสิทธิภาพ ค่าของ $V(\hat{\beta})$ จะสูง

^๘ Peter J. Huber, Robust Statistics, 1981, pp. 127-152.

^๙ Huang, D.S., Regression and Econometric Methods, 1970, pp. 127-147. or Johnston, J., Econometric Methods, 1963, pp. 208-266. or Judge, G.G., et. al. Introduction to the Theory and Practice of Econometrics, 1982, pp. 434-465.

เกินไป กล่าวคือ $\hat{\beta}$ จะขาดคุณสมบัติของตัวประมาณที่ไม่เอนเอียงเชิงเส้นที่ตีที่สุด อันจะมีผลทำให้ช่วงเชื่อมั่นของ $\hat{\beta}$ กว้างเกินไป รวมทั้งมีผลทำให้ค่า t-test มีค่าต่ำกว่าความเป็นจริง ซึ่งก่อให้เกิดการยอมรับสมมติฐาน $H_0 : \beta_j = 0$ ได้ง่าย และอาจจะมีผลทำให้ต้องตัดตัวแปรอิสระตัวนั้นทิ้งไป ทั้งที่ตัวแปรอิสระตัวนั้นอาจมีความสำคัญเป็นอย่างยิ่งก็ได้¹⁰

แต่เนื่องจากปัญหาต่างๆ ที่เกิดขึ้นทำให้เกิดผลกระทบต่อคุณภาพของงานวิจัย การประมาณค่า ส.ป.ส. ความถดถอยด้วยวิธี OLS ไม่อาจให้คำตอบที่ถูกต้องได้ นักวิจัยจำเป็นต้องใช้เทคนิคการประมาณค่า ส.ป.ส. ความถดถอยด้วยวิธีอื่นๆ ที่เหมาะสมที่มีผลทำให้ข้อตกลงของสมการถดถอยกลับเป็นจริงตามเดิม และถ้ายังคงเฝ้าใช้วิธี OLS อยู่ สมการถดถอยที่ได้จะมีคุณภาพต่ำ ทั้งในแง่โครงสร้าง การพยากรณ์ และคุณสมบัติที่พึงปรารถนา สำหรับตัวประมาณค่าที่ตี ดังนั้นภาระกิจที่สำคัญของนักวิจัยที่จะละเลยไม่ได้ นอกเหนือไปจากการประมาณค่า ส.ป.ส. ความถดถอยคือ การตรวจสอบข้อตกลงของสมการถดถอย ดังต่อไปนี้

1. การตรวจสอบข้อมูลมีการแจกแจงแบบปกติหรือไม่
2. การตรวจสอบข้อมูลมีปัญหาอัตตสหสัมพันธ์หรือไม่
3. การตรวจสอบข้อมูลมีปัญหาความแปรปรวนไม่คงที่หรือไม่

1. การตรวจสอบข้อมูลมีการแจกแจงแบบปกติหรือไม่ มีอยู่หลายวิธี แต่วิธีที่นำมาใช้ในการพัฒนาระบบครั้งนี้ เป็นวิธีการทดสอบข้อมูลทีนิยมใช้และมีประสิทธิภาพในการทดสอบข้อมูล ซึ่งวิธีทดสอบดังต่อไปนี้

- 1.1 วิธีการวัดค่าความเอนเอียงและความโด่ง (Skewness and Kurtosis)
- 1.2 วิธีการทดสอบโคลโมโกรอฟและสไมร์นอฟ (Kolmogorov - Smirnov)
- 1.3 วิธีการทดสอบชาปีโรและวิลค์ (Shapiro - Wilk)

การตรวจสอบข้อมูลมีการแจกแจงแบบปกติหรือไม่ด้วยวิธีการทดสอบต่างๆ ถ้าพบว่า

¹⁰ Huang, D.S., Regression and Econometric Methods, 1970, pp. 127-147. or Johnston, J., Econometric Methods, 1963, pp. 208-241. or Judge, G.G., et. al. Introduction to the Theory and Practice of Econometrics, 1982, pp. 409-429.

ข้อมูลมีการแจกแจงแบบปกติ ควรประมาณค่า ส.ป.ส. ความถดถอยด้วย วิธี OLS เพราะการประมาณค่า ส.ป.ส. ความถดถอยด้วยวิธี OLS จะได้ตัวประมาณค่า $\hat{\beta}$ ที่มีคุณสมบัติเป็นตัวประมาณที่ไม่เอนเอียงเชิงเส้นที่ดีที่สุด เมื่อข้อมูลมีคุณสมบัติเป็นไปตามข้อตกลงของสมการถดถอย แต่ถ้าข้อมูลมีการแจกแจงแบบทางยาวกว่าการแจกแจงแบบปกติ ควรประมาณค่า ส.ป.ส. ความถดถอยด้วย วิธี M - Estimates เพราะการประมาณค่า ส.ป.ส. ความถดถอยด้วยวิธี M - Estimates จะได้ตัวประมาณค่าที่มีคุณสมบัติดังต่อไปนี้

1. เป็นตัวประมาณที่ Consistency
2. ตัวประมาณมีการแจกแจงแบบ Asymptotic Normal
3. ตัวประมาณอาจจะไม่มีคุณสมบัติไม่เอนเอียง แต่มีความยืดหยุ่นในการใช้งานได้ดีกว่าตัวประมาณที่ได้จากวิธี OLS ในกรณีที่ข้อมูลมีค่าผิดปกติมาก ๆ ¹¹

และถ้าพบว่าข้อมูลมีการแจกแจงแบบเบ้สามารถประมาณค่า ส.ป.ส. ความถดถอยได้ทั้ง วิธี OLS และ วิธี M - Estimates เพราะการประมาณค่า ส.ป.ส. ความถดถอยด้วยวิธี OLS และ วิธี M - Estimates จะได้ตัวประมาณที่มีประสิทธิภาพใกล้เคียงกัน

2. การตรวจสอบข้อมูลมีปัญหาอัตโนมัติหรือไม่มีอยู่หลายวิธี แต่วิธีที่นำมาใช้ในการพัฒนาระบบครั้งนี้ เป็นวิธีการทดสอบข้อมูลที่นิยมใช้และมีประสิทธิภาพในการทดสอบข้อมูล ซึ่งมีวิธีการทดสอบดังต่อไปนี้

- 2.1 วิธีการทดสอบเบเนนบลูตและเวบบ์ (Berenblut - Webb)
- 2.2 วิธีการทดสอบเดอว์บินและวัตสัน (Durbin - Watson)

การตรวจสอบข้อมูลมีปัญหาอัตโนมัติหรือไม่มีอยู่ด้วยวิธีการทดสอบต่างๆ ถ้าพบว่าข้อมูลไม่มีปัญหาอัตโนมัติควรประมาณค่า ส.ป.ส. ความถดถอยด้วย วิธี OLS เพราะการประมาณค่า ส.ป.ส. ความถดถอยด้วยวิธี OLS จะได้ตัวประมาณค่า $\hat{\beta}$ ที่มีคุณสมบัติเป็นตัวประมาณที่ไม่เอนเอียงเชิงเส้นที่ดีที่สุด เมื่อข้อมูลมีคุณสมบัติเป็นไปตามข้อตกลงของสมการถดถอย แต่ถ้าข้อมูลมีปัญหาอัตโนมัติควรประมาณค่า ส.ป.ส. ความถดถอยด้วยวิธี Generalized Least Squares (GLS) เพราะการประมาณค่า ส.ป.ส. ความถดถอย

¹¹ Peter J. Huber, Robust Statistics, 1981, pp. 127-152.

ด้วยวิธี GLS จะได้ตัวประมาณค่า $\hat{\beta} = (x'\psi^{-1}x)^{-1} x'\psi^{-1}y$ เมื่อ ψ คือ Symmetric Positive Definite Matrix (เมตริกซ์ที่ให้ค่า Characteristic Value เป็นบวกทุกค่า) ที่มีคุณสมบัติเป็นตัวประมาณที่ไม่เอนเอียงเชิงเส้นที่ดีที่สุด¹²

3. การตรวจสอบข้อมูลมีปัญหาคงที่หรือไม่นั้นมีอยู่หลายวิธี แต่วิธีที่นำมาใช้ในการพัฒนาระบบครั้งนี้ เป็นวิธีการทดสอบข้อมูลที่ยอมรับใช้และมีประสิทธิภาพในการทดสอบข้อมูล ซึ่งมีวิธีการทดสอบดังต่อไปนี้

- 3.1 วิธีการทดสอบโกลด์ฟิลด์และควอนท์ (Goldfeld - Quandt)
- 3.2 วิธีการทดสอบสโรเตอร์ (Sroeter)
- 3.3 วิธีการทดสอบบรูสและพาแกน (Breusch - Pagan)
- 3.4 วิธีแบมเซต (BAMSET)

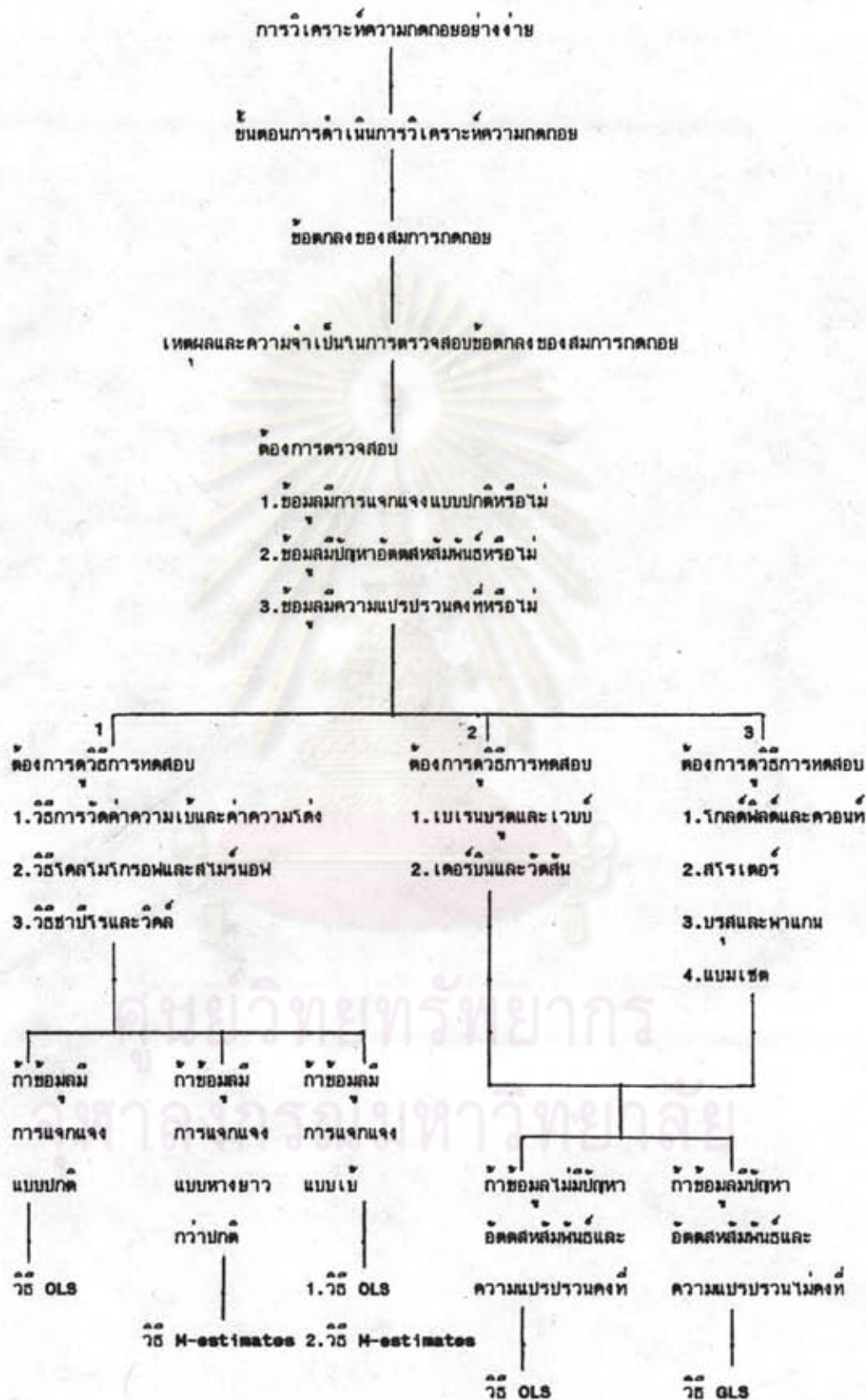
การตรวจสอบข้อมูลมีปัญหาคงที่หรือไม่นั้นด้วยวิธีการทดสอบต่างๆ ถ้าพบว่า ข้อมูลมีความแปรปรวนคงที่ ควรประมาณค่า ส.ป.ส. ความถดถอยด้วย วิธี OLS เพราะการประมาณค่า ส.ป.ส. ความถดถอยด้วยวิธี OLS จะได้ตัวประมาณค่า $\hat{\beta}$ ที่มีคุณสมบัติเป็นตัวประมาณที่ไม่เอนเอียงเชิงเส้นที่ดีที่สุด เมื่อข้อมูลมีคุณสมบัติเป็นไปตามข้อตกลงของสมการถดถอย แต่ถ้าข้อมูลมีปัญหาคงที่หรือไม่นั้น ควรประมาณค่า ส.ป.ส. ความถดถอยด้วยวิธี Generalized Least Squares (GLS) เพราะการประมาณค่า ส.ป.ส. ความถดถอยด้วยวิธี GLS จะได้ตัวประมาณค่า $\hat{\beta}$ ที่มีคุณสมบัติเป็นตัวประมาณที่ไม่เอนเอียงเชิงเส้นที่ดีที่สุด¹³

จากแนวคิดและทฤษฎีด้านการวิเคราะห์ความถดถอยอย่างง่ายที่กล่าวในหัวข้อ

2.2.1.1 แสดงเป็นผังงานได้ดังต่อไปนี้

¹² Huang, D.S., Regression and Econometric Methods, 1970, pp. 127-147. or Johnston, J., Econometric Methods, 1963, pp. 208-266. or Judge, G.G., et. al. Introduction to the Theory and Practice of Econometrics, 1982, pp. 409-429, 434-465.

¹³ Ibid., pp. 409-429, 434-465.



ขั้นที่ 3 ขั้นตอนนี้เป็นการศึกษาความสัมพันธ์ระหว่างตัวแปร ซึ่งไม่สามารถจำแนกตัวแปรออกเป็นตัวแปรตามและตัวแปรอิสระได้ โดยได้นำ ประเภทของตัวแปร มาเป็นเกณฑ์ในการพิจารณาแบ่งวิธีการวิเคราะห์ทางสถิติ เพราะประเภทของตัวแปรที่แตกต่างกันจะใช้วิธีการวิเคราะห์ทางสถิติที่แตกต่างกันออกไป โดยประเภทของตัวแปรแบ่งออกได้เป็น

1. ตัวแปรแบบต่อเนื่อง
2. ตัวแปรแบบต่อเนื่องแต่จัดแบ่งเป็นกลุ่ม
3. ตัวแปรแบบแบ่งเป็นกลุ่ม
4. ตัวแปรแบบจัดอันดับ

จากข้อกำหนดข้างต้นสามารถพัฒนาระบบให้มีแนวทาง เลือกดังต่อไปนี้

- ก. วิธี ส.ป.ส. สหสัมพันธ์แบบเพียร์สัน เป็นวิธีการวัดความสัมพันธ์ระหว่างตัวแปร 2 ตัว ซึ่งมีข้อตกลงเบื้องต้น ดังต่อไปนี้
1. ตัวแปรทั้งสองเป็นข้อมูลแบบต่อเนื่อง
 2. ความสัมพันธ์ระหว่างตัวแปรทั้งสองเป็นแบบเส้นตรง ¹⁴
- ข. วิธี ส.ป.ส. สหสัมพันธ์แบบไบซีเรียล เป็นวิธีการวัดความสัมพันธ์ระหว่างตัวแปร 2 ตัว ซึ่งมีข้อตกลงเบื้องต้น ดังต่อไปนี้
1. ตัวแปรตัวหนึ่งเป็นตัวแปรแบบต่อเนื่อง ส่วนตัวแปรอีกตัวหนึ่งเป็นตัวแปรแบบต่อเนื่องแต่แบ่งเป็นกลุ่ม
 2. ตัวแปรแบบต่อเนื่องไม่จำเป็นต้องมีการแจกแจงแบบปกติ เพียงแต่ให้มีการแจกแจงแบบโค้งเตี้ยและค่อนข้างสมมาตร
 3. ตัวแปรแบบต่อเนื่องแต่แบ่งเป็นกลุ่มจะต้องมีการแจกแจงแบบปกติ ¹⁵

¹⁴ Thorndike, Robert M. Correlational Procedure for Research, 1976, pp. 44-48.

¹⁵ Ibid., 1976, pp. 82-86.

ค. วิธี ส.ป.ส. สหสัมพันธ์แบบพอยต์ไบซีเรียล เป็นวิธีการวัดความสัมพันธ์ระหว่างตัวแปร 2 ตัว ซึ่งมีข้อตกลงเบื้องต้น ดังต่อไปนี้

1. ตัวแปรตัวหนึ่งเป็นตัวแปรแบบต่อเนื่อง ส่วนตัวแปรอีกตัวหนึ่งเป็นตัวแปรแบบเป็นกลุ่ม
2. ตัวแปรแบบต่อเนื่องไม่จำเป็นต้องมีการแจกแจงแบบปกติ เพียงแต่ให้มีการแจกแจงแบบโค้งเตี้ยและค่อนข้างสมมาตร ¹⁶

ง. วิธี ส.ป.ส. สหสัมพันธ์เตตระคอร์ริค เป็นวิธีการวัดความสัมพันธ์ระหว่างตัวแปร 2 ตัว ซึ่งมีข้อตกลงเบื้องต้นดังต่อไปนี้

ตัวแปรทั้งสองตัวเป็นตัวแปรแบบต่อเนื่องแต่แบ่งเป็นกลุ่ม ซึ่งโดยธรรมชาติแล้วไม่ได้มีลักษณะเป็นกลุ่มแต่ถูกแบ่งให้เป็นกลุ่ม ¹⁷

จ. วิธี ส.ป.ส. สหสัมพันธ์พี เป็นวิธีการวัดความสัมพันธ์ระหว่างตัวแปร 2 ตัว ซึ่งมีข้อตกลงเบื้องต้นดังต่อไปนี้

ตัวแปรทั้งสองตัวเป็นตัวแปรแบบเป็นกลุ่ม ¹⁸

ฉ. วิธี ส.ป.ส. สหสัมพันธ์อันดับที่ของสเปียร์แมนและวิธี ส.ป.ส. สหสัมพันธ์อันดับที่ของเคนคอลล เป็นวิธีการวัดความสัมพันธ์ระหว่างตัวแปร 2 ตัว ซึ่งมีข้อตกลงเบื้องต้นดังต่อไปนี้

ตัวแปรทั้งสองตัวเป็นตัวแปรแบบจัดอันดับ ¹⁹

ดังนั้นจึงใช้ข้อตกลงเบื้องต้นดังกล่าวนี้มาเป็นเกณฑ์ในการพิจารณาแบ่งวิธีการวิเคราะห์ทางสถิติ ดังแสดงเป็นผังงานได้ดังต่อไปนี้

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

¹⁶ Thorndike, Robert M. Correlational Procedure for Research, 1976, pp. 77-79.

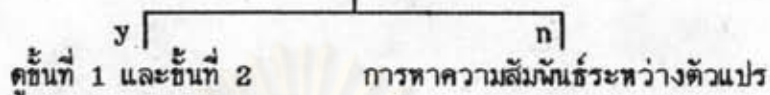
¹⁷ Ibid., 1976, pp. 86-89.

¹⁸ Ibid., 1976, pp. 79-81.

¹⁹ Ibid., 1976, pp. 74-77, 89-91.

ตัวแปร 2 ตัว

ทำสามารถจำแนกตัวแปรออกเป็นตัวแปรตามและตัวแปรอิสระได้หรือไม่



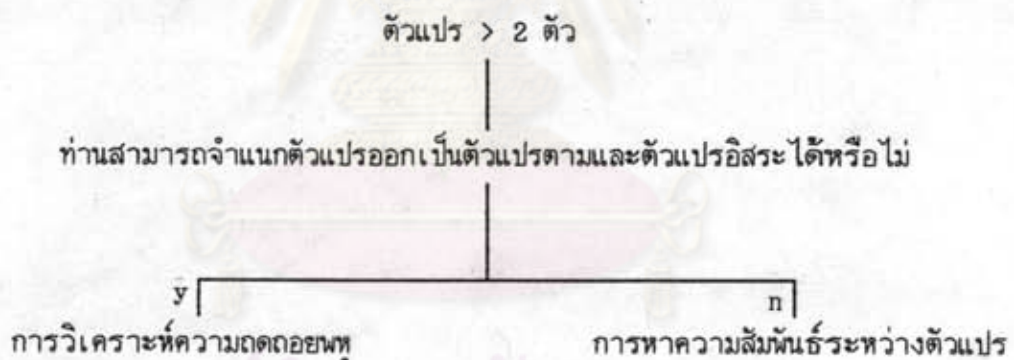
ประเภทของตัวแปร

	ตัวที่ 1	ตัวที่ 2
1.	ต่อเนื่อง	ต่อเนื่อง
2.	ต่อเนื่อง	ต่อเนื่องแต่นแบ่งเป็นกลุ่ม
3.	ต่อเนื่อง	เป็นกลุ่ม
4.	ต่อเนื่องแต่นแบ่งเป็นกลุ่ม	ต่อเนื่องแต่นแบ่งเป็นกลุ่ม
5.	เป็นกลุ่ม	เป็นกลุ่ม
6.	จัดอันดับ	จัดอันดับ

1	2	3	4	5	6
วิธี ส.ป.ส.	วิธี ส.ป.ส.	วิธี ส.ป.ส.	วิธี ส.ป.ส.	วิธี ส.ป.ส.	วิธี ส.ป.ส.
สหสัมพันธ์	สหสัมพันธ์	สหสัมพันธ์	สหสัมพันธ์	สหสัมพันธ์	สหสัมพันธ์
เพียร์สัน	ไบซีเรียล	พอยต์ไบซีเรียล	เตตระคอริก	ฟี	อันดับที่ของ
					สเปียร์แมน
					หรือ วิธี ส.ป.ส.
					สหสัมพันธ์
					อันดับที่ของ
					เคนคอลล์

2.2.2 กรณีจำนวนตัวแปรมากกว่า 2 ตัว

ขั้นที่ 1 เนื่องจากเป็นการพิจารณาการหาความสัมพันธ์ระหว่างตัวแปรและการวิเคราะห์ความถดถอย จึงใช้ข้อกำหนดที่ว่า การวิเคราะห์ความถดถอยพหุเป็นเทคนิคทางสถิติที่ใช้ในการวิเคราะห์หารูปแบบความสัมพันธ์ระหว่างตัวแปร ซึ่งตัวแปรตัวหนึ่งเป็นตัวแปรตามและตัวแปรอื่นๆ เป็นตัวแปรอิสระ ส่วนการหาความสัมพันธ์ระหว่างตัวแปรเป็นการศึกษาถึงความสัมพันธ์ระหว่างตัวแปรที่สนใจว่ามีความสัมพันธ์กันหรือไม่ มากน้อยเพียงใดและความสัมพันธ์เป็นไปในทิศทางใด²⁰ ดังนั้นจึงนำ การจำแนกตัวแปรออกเป็นตัวแปรตามและตัวแปรอิสระ มาเป็นเกณฑ์ในการพิจารณาแบ่งวิธีการวิเคราะห์ทางสถิติ เพราะการวิเคราะห์ความถดถอยสามารถที่จะจำแนกตัวแปรออกเป็นตัวแปรตามและตัวแปรอิสระได้ ส่วนการหาความสัมพันธ์ระหว่างตัวแปรไม่สามารถจำแนกตัวแปรออกเป็นตัวแปรตามและตัวแปรอิสระได้ ซึ่งแสดงเป็นผังงานได้ดังต่อไปนี้



อย่างไรก็ตามมีข้อสังเกตสำคัญประการหนึ่งว่ายังมีประเภทของการวิเคราะห์อีกหลายประเภทที่ไม่สามารถจำแนกตัวแปรออกเป็นตัวแปรตามและตัวแปรอิสระได้ คือ การวิเคราะห์จำแนกประเภท

²⁰ Draper, N.R. and H. Smith, Applied Regression Analysis, 1981, pp. 1-7, 33-35.

(Discriminant Analysis) การวิเคราะห์จัดกลุ่ม (Cluster Analysis) การวิเคราะห์ปัจจัย (Factor Analysis) และการวิเคราะห์แคนอนนิคัล (Canonical Analysis) ซึ่งได้นำมาผนวกไว้ในแนวทางเลือกเดียวกันกับการหาความสัมพันธ์ระหว่างตัวแปร ทางด้านการวิเคราะห์ความถดถอยพหุได้ทำการพัฒนาต่อไปในขั้นที่ 2 ส่วนทางด้านการศึกษาหาความสัมพันธ์ระหว่างตัวแปร การวิเคราะห์จำแนกประเภท การวิเคราะห์จัดกลุ่ม การวิเคราะห์ปัจจัย และการวิเคราะห์แคนอนนิคัลได้ทำการพัฒนาต่อไปในขั้นที่ 4 ดังต่อไปนี้

ขั้นที่ 2 ในขั้นตอนนี้ได้นำ จำนวนตัวแปรตาม มาเป็นเกณฑ์ในการพิจารณาแบ่งวิธีการวิเคราะห์ทางสถิติ เพราะการวิเคราะห์ความถดถอยพหุมีจำนวนตัวแปรมากกว่า 2 ตัว ซึ่งอาจจะมีตัวแปรตามมากกว่า 1 ตัวก็ได้ ดังนั้นจึงนำหลักเกณฑ์นี้มาเป็นเกณฑ์ในการพิจารณาแบ่งวิธีการวิเคราะห์ทางสถิติ โดยจำนวนตัวแปรตามแบ่งออกได้เป็นจำนวนตัวแปรตาม 1 ตัว และจำนวนตัวแปรตามมากกว่า 1 ตัว

ซึ่งกรณีจำนวนตัวแปรตาม 1 ตัวได้ทำการพัฒนาต่อไปในขั้นที่ 3 ส่วนกรณีจำนวนตัวแปรตามมากกว่า 1 ตัวจะเป็นเรื่องเกี่ยวกับ Multivariate Analysis²¹ ซึ่งเป็นเรื่องที่สลับซับซ้อนเนื่องจากมีจำนวนตัวแปรเข้ามาเกี่ยวข้องหลายตัว และมีได้ทำการพัฒนาลงไปในรายละเอียด ผู้ที่สนใจสามารถขอคำแนะนำเพิ่มเติมได้จากนักสถิติหรือผู้ที่มีความรู้ทางด้านนี้ได้

ขั้นที่ 3 ในขั้นตอนนี้ได้นำ ลักษณะข้อมูล ที่เป็นเชิงปริมาณหรือเชิงคุณภาพ มาเป็นเกณฑ์ในการพิจารณาแบ่งวิธีการวิเคราะห์ทางสถิติ เพราะลักษณะข้อมูลที่แตกต่างกันจะใช้วิธีการวิเคราะห์ทางสถิติที่แตกต่างกันออกไป

จากข้อกำหนดข้างต้นนี้สามารถพัฒนาระบบให้มีแนวทางเลือกดังต่อไปนี้

²¹ Dillon, W.R. and Goldstein, M., Multivariate Analysis, 1984, pp. 1-22.

ก. ลักษณะข้อมูลที่ใช้ในการวิเคราะห์ความถดถอยพหุ (Multiple Regression) ตัวแปรตามเป็นข้อมูลเชิงปริมาณหรือข้อมูลเชิงคุณภาพ ตัวแปรอิสระเป็นข้อมูลเชิงปริมาณหรือข้อมูลเชิงปริมาณปนกับข้อมูลเชิงคุณภาพ ²²

ข. ลักษณะข้อมูลที่ใช้ในการวิเคราะห์ความแปรปรวน (ANOVA) ตัวแปรตามเป็นข้อมูลเชิงปริมาณ ตัวแปรอิสระเป็นข้อมูลเชิงคุณภาพ ²³

ค. ลักษณะข้อมูลที่ใช้ในการวิเคราะห์จำแนกพหุ (Multiple Classification) ตัวแปรตามเป็นข้อมูลเชิงปริมาณ ตัวแปรอิสระเป็นข้อมูลเชิงคุณภาพหรือข้อมูลเชิงปริมาณปนกับข้อมูลเชิงคุณภาพ ²⁴

ง. ลักษณะข้อมูลที่ใช้ในการวิเคราะห์ความถดถอยทวิ (Binary Regression) ตัวแปรตามเป็นข้อมูลเชิงคุณภาพที่มีค่า 0 หรือ 1 ตัวแปรอิสระเป็นข้อมูลเชิงปริมาณหรือข้อมูลเชิงปริมาณปนกับข้อมูลเชิงคุณภาพ ²⁵

จ. ลักษณะข้อมูลที่ใช้ในการทดสอบไคสแควร์ (χ^2 - test) และแบบจำลองลอกเชิงเส้น (Log - linear Model) ตัวแปรตามและตัวแปรอิสระเป็นข้อมูลเชิงคุณภาพ ²⁶

ดังนั้นจึงใช้ข้อกำหนดนี้มาเป็นเกณฑ์ในการพิจารณาแบ่งวิธีการวิเคราะห์ทางสถิติ แต่ทางด้านการวิเคราะห์ความแปรปรวน การวิเคราะห์จำแนกพหุ การวิเคราะห์ความถดถอยทวิ การทดสอบไคสแควร์ และแบบจำลองลอกเชิงเส้น มิได้ทำการพัฒนาลงไปในรายละเอียดจึงได้เสนอเพียงกรณีศึกษาของการวิเคราะห์ดังกล่าวนี้เท่านั้น ส่วนทางด้านการวิเคราะห์ความถดถอยพหุทำการพัฒนาลงไปในรายละเอียด ดังแสดงในหัวข้อ 2.2.2.1

²² Draper, N.R. and H. Smith, Applied Regression Analysis, 1981, pp. 1-7.

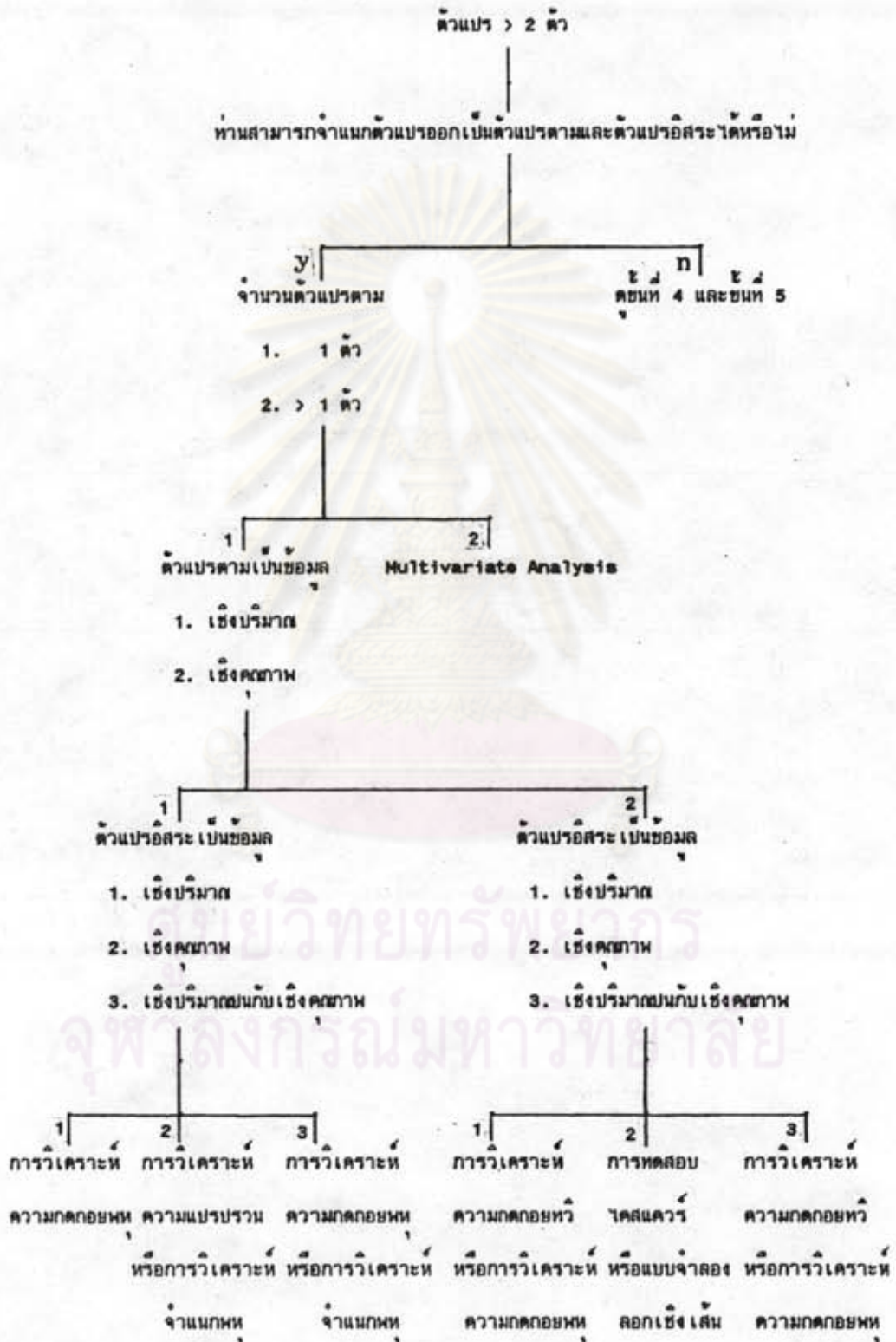
²³ Cochran and Cox, Experimental Designs, 1957, pp. 1-14.

²⁴ Andrews, Frank M., et. al. Multiple Classification Analysis, 1973.

²⁵ Judge, G.G., et. al. Introduction to the Theory and Practice of Econometrics, 1982, pp. 302-336.

²⁶ Dillon, W.R. and Goldstein, M., Multivariate Analysis, 1984, pp. 302-336.

จากข้อกำหนดข้างต้น แสดง เป็นผังงานได้ดังต่อไปนี้



2.2.2.1 แนวคิดและทฤษฎีด้านการวิเคราะห์ความถดถอยพหุ

การวิเคราะห์ความถดถอยพหุ (Multiple Regression) เป็นเทคนิคทางสถิติที่ใช้ในการวิเคราะห์หาความสัมพันธ์ระหว่างตัวแปรมากกว่า 2 ตัว ซึ่งตัวแปรตัวหนึ่งเป็นตัวแปรตามและตัวแปรอื่นๆเป็นตัวแปรอิสระ โดย

ตัวแปรตามคือ ตัวแปรที่เราสนใจศึกษา ใช้สัญลักษณ์ y

ตัวแปรอิสระคือ ตัวแปรที่มีอิทธิพลต่อการเปลี่ยนแปลงของตัวแปรที่เราสนใจศึกษา ใช้สัญลักษณ์ x

สามารถเขียนรูปแบบฟังก์ชันได้เป็น $y = f(x's) + \varepsilon$

ตัวแบบความถดถอยเชิงเส้นพหุ เขียนได้เป็น

$$y_i = \beta_0 + \beta_j x_{ij} + \varepsilon_i \quad ; i = 1, 2, 3, \dots, n$$

$$; j = 1, 2, 3, \dots, k$$

หรือเขียนในรูปเมทริกซ์ได้เป็น $\underset{\sim}{y} = \underset{\sim}{x}\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$

$$\begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \cdot \\ \cdot \\ \cdot \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_n \end{bmatrix}$$

เมื่อ $\underset{\sim}{y}$ คือเวกเตอร์ของตัวแปรตาม y ที่มีขนาด $n \times 1$

$\underset{\sim}{\beta}$ คือเวกเตอร์ของสัมประสิทธิ์ของความถดถอย ที่มีขนาด $(k+1) \times 1$

$\underset{\sim}{x}$ คือเมทริกซ์ของตัวแปรอิสระ ที่มีขนาด $n \times (k+1)$

$\underset{\sim}{\varepsilon}$ คือเวกเตอร์ของความคลาดเคลื่อน ที่มีขนาด $n \times 1$

เนื่องจากการกำหนดความสัมพันธ์ $y = f(x's) + \varepsilon$ ซึ่งนักวิจัยจำเป็นต้องกำหนดตัวแปรอิสระให้ครบถ้วน ในทางปฏิบัตินักวิจัยไม่สามารถกระทำได้นอกจากนี้ยังมีความผิดพลาดอื่นๆ ร่วมส่งผลกระทบต่ออยู่ด้วย ทำให้ค่าของ y ผิดพลาดไปจากความเป็นจริง ด้วยเหตุนี้จึงจัดให้ความผิดพลาดจากแหล่งต่างๆ รวมอยู่ในค่าความคลาดเคลื่อน ε_i และควบคุมด้วยข้อตกลงของสมการถดถอย²⁷ ดังนี้

1. ε_i เป็นตัวแปรเชิงสุ่ม

2. $E(\varepsilon_i) = 0$; $i = 1, 2, 3, \dots, n$

ค่าคาดหวังของค่าความคลาดเคลื่อนมีค่าเท่ากับ 0

3. $E(\varepsilon_i \varepsilon_j) = 0$; $i \neq j$

ค่าความคลาดเคลื่อนในต่างวาระกันเป็นอิสระต่อกัน

4. $E(\varepsilon_i \varepsilon_j) = \sigma^2$; $i = j$

ค่าความคลาดเคลื่อนมีความแปรปรวนคงที่เท่ากับ σ^2

5. $\varepsilon_i \sim N(0, \sigma^2)$

ค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติมีค่าเฉลี่ยเท่ากับ 0 ความแปรปรวนเท่ากับ σ^2

6. ตัวแปรอิสระต้องเป็นอิสระซึ่งกันและกัน

ขั้นตอนที่สำคัญของการวิเคราะห์ความถดถอยคือ การประมาณค่า ส.ป.ส. ความถดถอย โดยวิธีใดวิธีหนึ่งที่เห็นว่าเป็นวิธีที่เหมาะสมกับลักษณะข้อมูล โดยปกติมักเลือกใช้วิธีกำลังสองน้อยที่สุด (Ordinary Least Square : OLS) ซึ่งเป็นวิธีที่ทำให้ผลรวมกำลังสองของความคลาดเคลื่อน (Sum Square Error : SSE) มีค่าต่ำที่สุด และทฤษฎีของ Gauss - Markov กล่าวไว้ว่า "การประมาณค่า ส.ป.ส. ความถดถอยด้วยวิธี OLS จะได้ตัวประมาณค่าที่มีคุณสมบัติเป็นตัวประมาณที่ไม่เอนเอียงเชิงเส้นที่ดีที่สุด (Best Linear Unbiased Estimator (BLUE))"²⁸

²⁷ Draper, N.R. and H. Smith, Applied Regression Analysis, 1981, p. 17.

²⁸ Wonnacott, R.J. and Wonnacott, T.H., Regression, 1981, pp. 31-32.

ในทางปฏิบัติบ่อยครั้งที่พบว่าข้อมูลที่นำมาวิเคราะห์นั้นมีความสัมพันธ์ไม่เป็นไปตามข้อตกลงของสมการถดถอย ก่อให้เกิดปัญหาต่างๆ ดังต่อไปนี้

1. ปัญหาความคลาดเคลื่อนมิได้มีการแจกแจงแบบปกติ (Nonnormal Error) คือสถานการณ์ที่ตัวแปรสุ่ม ϵ_i มีการแจกแจงแบบอื่นๆ ที่มิใช่การแจกแจงแบบปกติ ผลกระทบที่ตามมาทำนองเดียวกันกับปัญหาความคลาดเคลื่อนมิได้มีการแจกแจงแบบปกติ ที่กล่าวไว้แล้วในหัวข้อ 2.2.1.1 ²⁹

2. ปัญหาความคลาดเคลื่อนมีอัตราสหสัมพันธ์ (Autocorrelation) คือสถานการณ์ที่เกิดความสัมพันธ์ระหว่าง ϵ_i กับ ϵ_j กล่าวคือ $E(\epsilon_i \epsilon_j) \neq 0 ; i \neq j$ ผลกระทบที่ตามมาทำนองเดียวกันกับปัญหาความคลาดเคลื่อนมีอัตราสหสัมพันธ์ ที่กล่าวไว้แล้วในหัวข้อ 2.2.1.1

3. ปัญหาความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ (Heteroscedasticity) กล่าวคือ $E(\epsilon_i \epsilon_j) \neq \sigma^2 ; i = j$ ผลกระทบที่ตามมาทำนองเดียวกันกับปัญหาความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ ที่กล่าวไว้แล้วในหัวข้อ 2.2.1.1 ³⁰

4. ปัญหาตัวแปรอิสระมีความสัมพันธ์กันสูง (Multicollinearity) ผลกระทบที่ตามมาคือ นักวิจัยจะประสบความยุ่งยากในการจำแนกอิทธิพลของตัวแปรอิสระออกมาให้เห็นอย่างชัดเจนว่าตัวแปรอิสระตัวใดมีอิทธิพลต่อตัวแปรตามมากน้อยเพียงใด และตัวประมาณค่า $\hat{\beta}$ จะเปลี่ยนแปลงไปเมื่อเพิ่มหรือลดขนาดตัวอย่างเพียงเล็กน้อย รวมทั้งเครื่องหมายหรือขนาดของตัวประมาณค่า $\hat{\beta}$ อาจผิดไปจากความเป็นจริง ³¹

²⁹ Peter J. Huber, Robust Statistics, 1981, pp. 127-152.

³⁰ Huang, D.S., Regression and Econometric Methods, 1970, pp. 127-147. or Johnston, J., Econometric Methods, 1963, pp. 208-266. or Judge, G.G., et. al. Introduction to the Theory and Practice of Econometrics, 1982, pp. 409-429, 434-465.

³¹ Chatterjee, S., and Price, B., Regression Analysis By Example, 1977, pp. 181-192. or Huang, D.S., Regression and Econometric Methods, 1970, pp. 149-157. or Judge, G.G., et. al. Introduction to the Theory and Practice of Econometrics, 1982, pp. 610-629.

แต่เนื่องจากปัญหาต่างๆ ที่เกิดขึ้นทำให้เกิดผลกระทบต่อคุณภาพของงานวิจัย การประมาณค่า ส.ป.ส. ความถดถอยด้วยวิธี OLS ไม่อาจให้คำตอบที่ถูกต้องได้ นักวิจัยจำเป็นต้องใช้เทคนิคการประมาณค่า ส.ป.ส. ความถดถอยด้วยวิธีอื่นๆ ที่เหมาะสมที่มีผลทำให้ข้อตกลงของสมการถดถอยกลับเป็นจริงตามเดิม และถ้ายังคงเน้นใช้วิธี OLS อยู่ สมการถดถอยที่ได้จะมีคุณภาพต่ำ ทั้งในแง่โครงสร้าง การพยากรณ์ และคุณสมบัติที่พึงปรารถนา สำหรับตัวประมาณค่าที่ดี ดังนั้นภาระกิจที่สำคัญของนักวิจัยที่จะละเลยไม่ได้ นอกเหนือไปจากการประมาณค่า ส.ป.ส. ความถดถอยคือ การตรวจสอบข้อตกลงของสมการถดถอย ดังต่อไปนี้

1. การตรวจสอบข้อมูลมีการแจกแจงแบบปกติหรือไม่
2. การตรวจสอบข้อมูลมีปัญหาอัตโนมัติสัมพันธ์หรือไม่
3. การตรวจสอบข้อมูลมีปัญหาความแปรปรวนไม่คงที่หรือไม่
4. การตรวจสอบตัวแปรอิสระมีความสัมพันธ์กันสูงหรือไม่

1. การตรวจสอบข้อมูลมีการแจกแจงแบบปกติหรือไม่ มีอยู่หลายวิธี แต่วิธีที่นำมาใช้ในการพัฒนาระบบครั้งนี้ เป็นวิธีการทดสอบข้อมูลที่นิยมใช้และมีประสิทธิภาพในการทดสอบข้อมูล ซึ่งมีวิธีการทดสอบดังต่อไปนี้

- 1.1 วิธีการวัดค่าความเบ้และค่าความโด่ง (Skewness and Kurtosis)
- 1.2 วิธีการทดสอบโคลโมโกรอฟและสไมร์นอฟ (Kolmogorov - Smirnov)
- 1.3 วิธีการทดสอบชาปีโรและวิลค์ (Shapiro - Wilk)

การตรวจสอบข้อมูลมีการแจกแจงแบบปกติหรือไม่ด้วยวิธีการทดสอบต่างๆ ถ้าพบว่าข้อมูลมีการแจกแจงแบบปกติ ควรประมาณค่า ส.ป.ส. ความถดถอยด้วย วิธี OLS เพราะการประมาณค่า ส.ป.ส. ความถดถอยด้วยวิธี OLS จะได้ตัวประมาณค่า $\hat{\beta}$ ที่มีคุณสมบัติเป็นตัวประมาณที่ไม่เอนเอียงเชิงเส้นที่ดีที่สุด เมื่อข้อมูลมีคุณสมบัติเป็นไปตามข้อตกลงของสมการถดถอย แต่ถ้าข้อมูลมีการแจกแจงแบบทางยาวกว่าการแจกแจงแบบปกติ ควรประมาณค่า ส.ป.ส. ความถดถอยด้วย วิธี M - Estimates เพราะการประมาณค่า ส.ป.ส. ความถดถอยด้วยวิธี M - Estimates จะได้ตัวประมาณค่าที่มีคุณสมบัติดังต่อไปนี้

1. เป็นตัวประมาณที่ Consistency
2. ตัวประมาณมีการแจกแจงแบบ Asymptotic Normal
3. ตัวประมาณอาจจะไม่มีคุณสมบัติไม่เอนเอียง แต่มีความยืดหยุ่นในการใช้งาน

ได้ดีกว่าตัวประมาณที่ได้จากวิธี OLS ในกรณีที่ข้อมูลมีค่าผิดปกติมาก ๆ ³²

และถ้าพบว่าข้อมูลมีการแจกแจงแบบเบ้สามารถประมาณค่า ส.ป.ส. ความถดถอยได้ ทั้ง วิธี OLS และวิธี M - Estimates เพราะการประมาณค่า ส.ป.ส. ความถดถอยด้วยวิธี OLS และ วิธี M - Estimates จะได้ตัวประมาณที่มีประสิทธิภาพใกล้เคียงกัน

2. การตรวจสอบข้อมูลมีปัญหาออคตสสัมพันธ์หรือไม่ มีอยู่หลายวิธี แต่วิธีที่นำมาใช้ในการพัฒนาระบบครั้งนี้ เป็นวิธีการทดสอบข้อมูลทีนิยมใช้และมีประสิทธิภาพในการทดสอบข้อมูล ซึ่งมีวิธีการทดสอบดังต่อไปนี้

2.1 วิธีการทดสอบเบเรเนบรูตและเวบบ์ (Berenblut - Webb)

2.2 วิธีการทดสอบเดอร์บินและวัตสัน (Durbin - Watson)

การตรวจสอบข้อมูลมีปัญหาออคตสสัมพันธ์หรือไม่ด้วยวิธีการทดสอบต่างๆ ถ้าพบว่าข้อมูลไม่มีปัญหาออคตสสัมพันธ์ควรประมาณค่า ส.ป.ส. ความถดถอยด้วย วิธี OLS เพราะการประมาณค่า ส.ป.ส. ความถดถอยด้วยวิธี OLS จะได้ตัวประมาณค่า $\hat{\beta}$ ที่มีคุณสมบัติเป็นตัวประมาณที่ไม่เอนเอียงเชิงเส้นที่ดีที่สุด เมื่อข้อมูลมีคุณสมบัติเป็นไปตามข้อตกลงของสมการถดถอย แต่ถ้าข้อมูลมีปัญหาออคตสสัมพันธ์ควรประมาณค่า ส.ป.ส. ความถดถอยด้วย วิธี Generalized Least Squares (GLS) เพราะการประมาณค่า ส.ป.ส. ความถดถอยด้วยวิธี GLS จะได้ตัวประมาณค่า $\hat{\beta}^*$ ที่มีคุณสมบัติเป็นตัวประมาณที่ไม่เอนเอียงเชิงเส้นที่ดีที่สุด ³³

3. การตรวจสอบข้อมูลมีปัญหาความแปรปรวนไม่คงที่หรือไม่ มีอยู่หลายวิธี แต่วิธีที่นำมาใช้ในการพัฒนาระบบครั้งนี้ เป็นวิธีการทดสอบข้อมูลทีนิยมใช้และมีประสิทธิภาพในการทดสอบข้อมูล ซึ่งมีวิธีการทดสอบดังต่อไปนี้

³² Peter J. Huber, Robust Statistics, 1981, pp. 127-152.

³³ Huang, D.S., Regression and Econometric Methods, 1970, pp. 127-147. or Johnston, J., Econometric Methods, 1963, pp. 208-266. or Judge, G.G., et. al. Introduction to the Theory and Practice of Econometrics, 1982, pp. 434-465.

- 3.1 วิธีการทดสอบโกลด์ฟิลด์และควอนท์ (Goldfeld - Quandt)
- 3.2 วิธีการทดสอบสโรเตอร์ (Sroeter)
- 3.3 วิธีการทดสอบบรูสและพาแกน (Breusch - Pagan)
- 3.4 วิธีแบมเซต (BAMSET)

การตรวจสอบข้อมูลมีปัญหาความแปรปรวนไม่คงที่หรือไม่ด้วยวิธีการทดสอบต่างๆ ถ้าพบว่า ข้อมูลมีความแปรปรวนคงที่ ควรประมาณค่า ส.ป.ส. ความถดถอยด้วย วิธี OLS เพราะการประมาณค่า ส.ป.ส. ความถดถอยด้วยวิธี OLS จะได้ตัวประมาณค่า $\hat{\beta}$ ที่มีคุณสมบัติเป็นตัวประมาณที่ไม่เอนเอียงเชิงเส้นที่ดีที่สุด เมื่อข้อมูลมีคุณสมบัติเป็นไปตามข้อตกลงของสมการถดถอย แต่ถ้าข้อมูลมีปัญหาความแปรปรวนไม่คงที่ ควรประมาณค่า ส.ป.ส. ความถดถอยด้วยวิธี Generalized Least Squares (GLS) เพราะการประมาณค่า ส.ป.ส. ความถดถอยด้วยวิธี GLS จะได้ตัวประมาณค่า $\hat{\beta}^*$ ที่มีคุณสมบัติเป็นตัวประมาณที่ไม่เอนเอียงเชิงเส้นที่ดีที่สุด³⁴

4. การตรวจสอบตัวแปรอิสระมีความสัมพันธ์กันสูงหรือไม่ มีอยู่หลายวิธี แต่วิธีที่นำมาใช้ในการพัฒนาระบบครั้งนี้ เป็นวิธีการทดสอบข้อมูลกันนิยมใช้และมีประสิทธิภาพในการทดสอบข้อมูล ซึ่งมีวิธีการทดสอบดังต่อไปนี้

- 4.1 วิธีการวัดค่าสหสัมพันธ์เชิงเส้นระหว่างตัวแปรอิสระ
- 4.2 วิธีฟาเซียล รีเกรสชัน (Partial Regression)
- 4.3 วิธีออกซีลารี รีเกรสชัน (Auxiliary Regression)

การตรวจสอบตัวแปรอิสระมีความสัมพันธ์กันสูงหรือไม่ด้วยวิธีการทดสอบต่างๆ ถ้าพบว่า ตัวแปรอิสระมิได้มีความสัมพันธ์กันสูง ควรประมาณค่า ส.ป.ส. ความถดถอยด้วย วิธี OLS เพราะการประมาณค่า ส.ป.ส. ความถดถอยด้วยวิธี OLS จะได้ตัวประมาณค่า $\hat{\beta}$ ที่มีคุณสมบัติเป็นตัวประมาณที่ไม่เอนเอียงเชิงเส้นที่ดีที่สุด เมื่อข้อมูลมีคุณสมบัติเป็นไปตามข้อตกลง

³⁴ Huang, D.S., Regression and Econometric Methods, 1970, pp. 127-147. or Johnston, J., Econometric Methods, 1963, pp. 208-241. or Judge, G.G., et. al. Introduction to the Theory and Practice of Econometrics, 1982, pp. 409-429.

ของสมการถดถอย แต่ถ้าตัวแปรอิสระมีความสัมพันธ์กันสูงควรประมาณค่า ส.ป.ส. ความถดถอย ด้วย วิธี Ridge Regression ³⁵

หลักการโดยทั่วไปของ วิธี Ridge Regression คือเมื่อพบว่า $|x'x|$ มีค่าใกล้ 0 ซึ่งจะมีผลให้ $(x'x)^{-1}$ มีแนวโน้มที่จะไม่ปรากฏค่าและตัวประมาณค่าของ β คือ $\hat{\beta}$ มีค่าสูงผิดความจริง ให้อาศัยการปรับรูปเมตริกซ์ $x'x$ โดยการนำค่าคงที่ c ที่มากกว่าศูนย์มาบวกกับสมาชิกบนแนวเส้นทแยงมุม (Main Diagonal) ของเมตริกซ์ $x'x$ ซึ่งจะทำให้ Characteristic Root มีค่าสูงขึ้นและแก้ไขปัญหา Multicollinearity ได้

Ridge Regression นอกจากจะเป็นวิธีที่ใช้แก้ไขปัญหา Multicollinearity แล้วยังสามารถแก้ไขจุดอ่อนบางประการของวิธี OLS อีกด้วย กล่าวคือเวกเตอร์ของตัวประมาณค่าของ $\hat{\beta}$ คือ $[\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k]'$ จะมีความยาวเฉลี่ยเท่ากับความยาวของ β

การประมาณค่า ส.ป.ส. ความถดถอยด้วย วิธี Ridge Regression จะได้ ตัวประมาณ $\hat{\beta}^{**} = (x'x + cI_k)^{-1} x'y$; $c > 0$ ซึ่งมีคุณสมบัติดังต่อไปนี้

1. $\hat{\beta}^{**}$ เป็นเวกเตอร์ที่มีความยาวน้อยกว่าเวกเตอร์อื่นๆ
2. SS(Residual) จะมีค่าสูงขึ้นตามค่าของ c
3. $\hat{\beta}^{**}$ มีความยาวน้อยกว่า $\hat{\beta}$ เมื่อ $c \neq 0$ และเมื่อ c มีค่าสูงขึ้น $\hat{\beta}^{**}$ จะลดความยาวลง และเมื่อ $c \rightarrow \infty$ แล้ว $\hat{\beta}^{**}, \hat{\beta}^{**} \rightarrow 0$
4. สามารถหาค่าของ c ; $c > 0$ ที่มีผลให้ $MSE(\hat{\beta}^{**}) < MSE(\hat{\beta})$ ได้เสมอ เมื่อ MSE คือความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error)
5. $\hat{\beta}^{**}$ สามารถที่จะเขียนเป็น Linear Transformation ของตัวประมาณที่ได้

จากวิธี OLS ดังนี้

$$\begin{aligned}\hat{\beta}^{**} &= (x'x + cI_k)^{-1} x'y \\ &= (I_k + c(x'x)^{-1})^{-1} \hat{\beta}\end{aligned}$$

³⁵ Chatterjee, S., and Price, B., Regression Analysis By Example, 1977, pp. 181-192. or Huang, D.S., Regression and Econometric Methods, 1970, pp. 149-157. or Judge, G.G., et. al. Introduction to the Theory and Practice of Econometrics, 1982, pp. 610-629.

ให้ $z = (I_k + c(x'x)^{-1})^{-1}$ จะได้ว่า $\hat{\beta}^{**} = z\hat{\beta}$ ถ้า $c = 0$ จะพบว่า $\hat{\beta}^{**} = \hat{\beta}$ หรือวิธี OLS เป็นกรณีเฉพาะของวิธี Ridge Regression

6. $\hat{\beta}^{**}$ เป็นตัวประมาณที่เอนเอียงของ β (Biased Estimator)

$$\begin{aligned} E(\hat{\beta}^{**}) &= z E(\hat{\beta}) \\ &= z\beta \end{aligned}$$

จากแนวคิดและทฤษฎีด้านการวิเคราะห์ความถดถอยที่กล่าวในหัวข้อ 2.2.2.1 แสดงเป็นผังงานได้ดังต่อไปนี้



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ขั้นที่ 4 ในขั้นตอนนี้ได้นำ วัตถุประสงค์ของการวิเคราะห์ มาเป็นเกณฑ์ในการพิจารณาแบ่งวิธีการวิเคราะห์ทางสถิติ เพราะวัตถุประสงค์ของการวิเคราะห์จะเป็นเกณฑ์ที่ทำให้เรารู้ว่าจะใช้วิธีการวิเคราะห์ทางสถิติวิธีใด การที่ไม่ได้นำลักษณะข้อมูลมาเป็นเกณฑ์ในการพิจารณาแบ่งวิธีการวิเคราะห์ทางสถิติเหมือนกับขั้นตอนอื่นๆ เนื่องจากวิธีการวิเคราะห์ทางสถิติหลายๆ วิธีในแนวทางเลือกนี้ใช้ลักษณะข้อมูลที่คล้ายคลึงกัน ทำให้ไม่สามารถนำลักษณะข้อมูลมาใช้เป็นเกณฑ์ในการพิจารณาแบ่งวิธีการวิเคราะห์ทางสถิติให้ชัดเจนลงไปได้

จากข้อกำหนดข้างต้น สามารถพัฒนาระบบให้มีแนวทางเลือกดังต่อไปนี้

ก. การวิเคราะห์จำแนกประเภท (Discriminant Analysis) มีวัตถุประสงค์ที่จะจำแนกหน่วยวิเคราะห์ออกเป็นกลุ่ม ซึ่งนักวิจัยรู้ว่าหน่วยใดเป็นสมาชิกของกลุ่มใดอยู่แล้วแต่ต้องการจะใช้ข้อมูลที่มีอยู่มาศึกษาว่าตัวแปรต่างๆ ที่คิดว่ามีคุณสมบัติที่จะจำแนกหน่วยวิเคราะห์ออกไปตามกลุ่มต่างๆ นั้นสามารถจำแนกได้ดีเพียงใด ตัวแปรใดไม่มีความสำคัญหรือไม่มีประโยชน์ต่อการจำแนกและตัวแปรที่จำแนกประเภทได้ดีแต่ละตัวมีอิทธิพลแตกต่างกันอย่างไร ³⁶

ข. การวิเคราะห์จัดกลุ่ม (Cluster Analysis) มีวัตถุประสงค์ที่จะจัดกลุ่มตัวแปร ตัวแปรใดที่มีความสัมพันธ์กันสูง ในทางบวกจะถูกจัดให้อยู่ในกลุ่มเดียวกัน ³⁷

ค. การวิเคราะห์ปัจจัย (Factor Analysis) มีวัตถุประสงค์ที่จะวัดขนาดของตัวแปรเชิงเปรียบเทียบ ตัวอย่างเช่น การวัดขนาดของเต่า ซึ่งตัวแปรที่มีอิทธิพลต่อขนาดของเต่า เช่น ความกว้าง ความยาว และความสูงของเต่า โดยผลสรุปที่ได้จากการวิเคราะห์จะทำให้เรารู้ว่าตัวแปรใดจะมีอิทธิพลต่อขนาดของเต่ามากที่สุด โดยบอกเป็นขนาดเชิงเปรียบเทียบ เช่น ความยาวมีอิทธิพลต่อขนาดของเต่ามากกว่าความกว้าง 1.6 เท่า เป็นต้น หรือมีวัตถุประสงค์ที่จะรวมตัวแปรที่มีความสัมพันธ์กันสูงหลายๆ ตัวแปรให้เป็นปัจจัย ³⁸

จุฬาลงกรณ์มหาวิทยาลัย

³⁶ Dillon, W.R. and Goldstein, M., Multivariate Analysis, 1984, pp. 360-364.

³⁷ Ibid., 1984, pp. 157-161.

³⁸ Ibid., 1984, pp. 53-60.

ทางด้านกรวิเคราะห์จำแนกประเภท การวิเคราะห์จัดกลุ่ม การวิเคราะห์ปัจจัย และการวิเคราะห์แคนอนนิคัล มิได้ทำการพัฒนาไปในรายละเอียด ซึ่งผู้ที่สนใจสามารถขอคำแนะนำเพิ่มเติมได้จากนักสถิติหรือผู้ที่มีความรู้ทางด้านนี้

ขั้นที่ 5 ในขั้นตอนนี้ได้นำ ประเภทของตัวแปร มาเป็นเกณฑ์ในการพิจารณาแบ่งวิธีการวิเคราะห์ทางสถิติ เพราะประเภทของตัวแปรที่แตกต่างกันจะใช้วิธีการวิเคราะห์ทางสถิติที่แตกต่างกันออกไป โดยประเภทของตัวแปรแบ่งออกได้เป็น

1. ตัวแปรแบบต่อเนื่อง
2. ตัวแปรแบบจัดอันดับ

โดยตัวแปรแบบต่อเนื่อง ได้นำ วัตถุประสงค์ของการวิเคราะห์ มาเป็นเกณฑ์ในการพิจารณาแบ่งวิธีการวิเคราะห์ทางสถิติอีกขั้นตอนหนึ่งด้วย

จากข้อกำหนดข้างต้นสามารถพัฒนาระบบให้มีแนวทางเลือกดังต่อไปนี้

ก. วิธี ส.ป.ส. สหสัมพันธ์บางส่วน เป็นวิธีการวัดความสัมพันธ์ระหว่างตัวแปร 2 ตัว โดยจัดอิทธิพลของตัวแปรอื่นๆออกไป ซึ่งมีข้อตกลงเบื้องต้นดังต่อไปนี้

1. จำนวนตัวแปรมากกว่า 2 ตัว
2. ประเภทของตัวแปรเป็นตัวแปรแบบต่อเนื่อง ⁴⁰

ข. วิธี ส.ป.ส. สหสัมพันธ์พหุคูณ เป็นวิธีการวัดความสัมพันธ์ระหว่างตัวแปรหลายๆตัว ซึ่งมีข้อตกลงเบื้องต้นดังต่อไปนี้

1. จำนวนตัวแปรมากกว่า 2 ตัว
2. ประเภทของตัวแปรเป็นตัวแปรแบบต่อเนื่อง ⁴¹

ค. วิธี ส.ป.ส. สหสัมพันธ์ความสอดคล้องของเคนดอลล์ เป็นวิธีการวัดความ

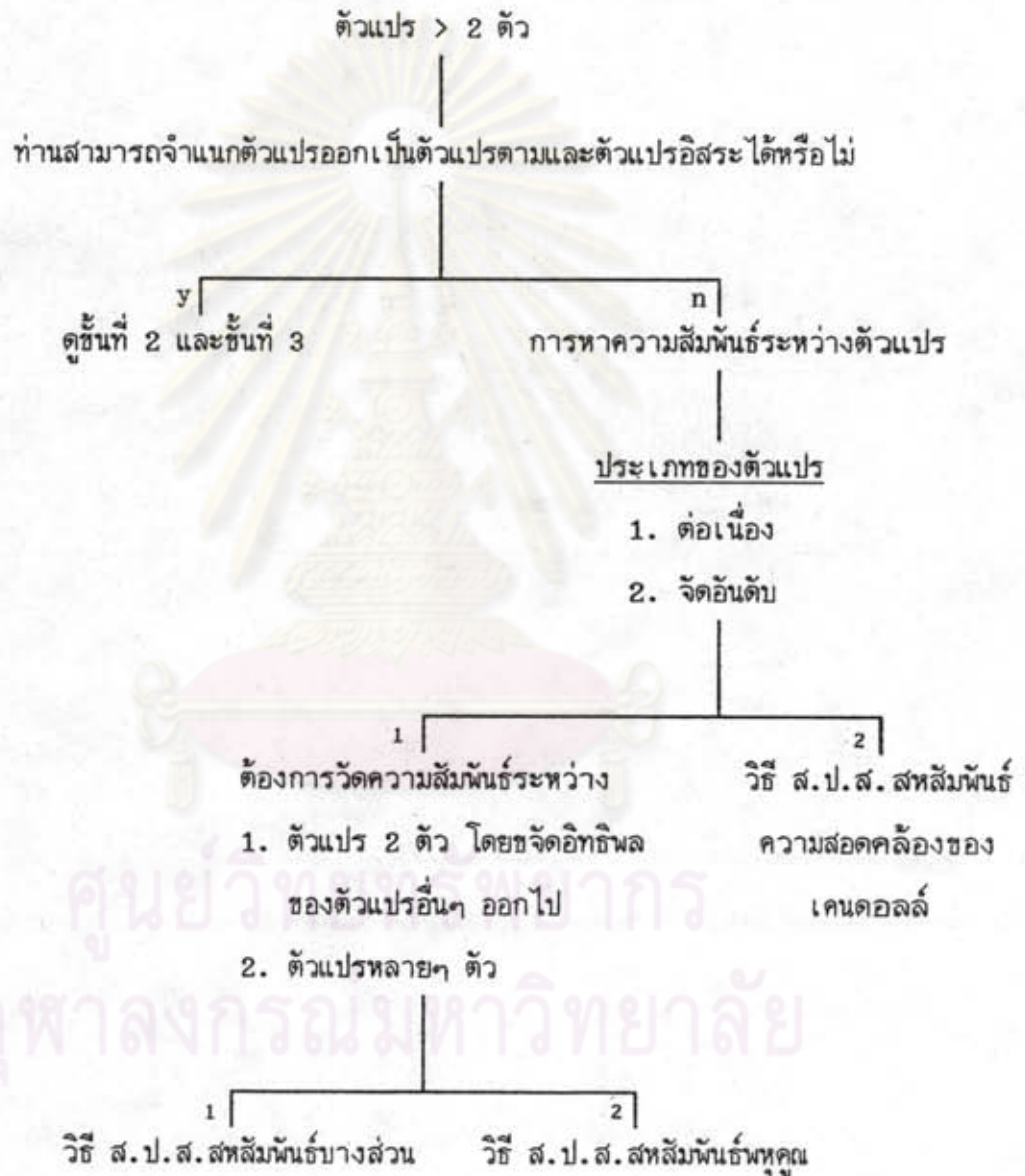
⁴⁰ Thorndike, Robert M. Correlational Procedure for Research, 1976, pp. 130-136.

⁴¹ Ibid., 1976, pp. 139-167.

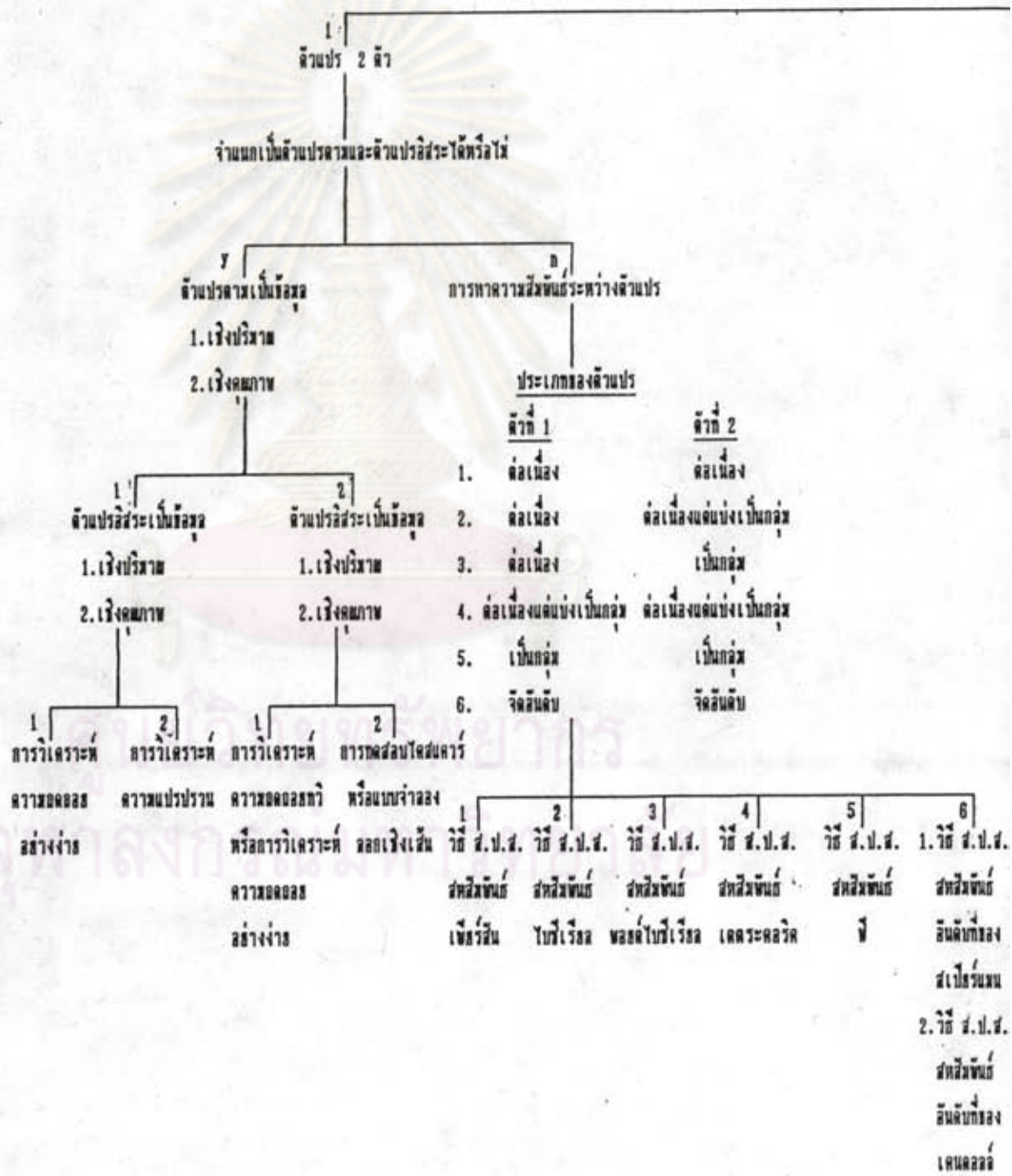
สัมพันธ์ระหว่างตัวแปรหลายๆ ตัว ซึ่งมีข้อตกลงเบื้องต้นดังต่อไปนี้

1. จำนวนตัวแปรมากกว่า 2 ตัว
2. ประเภทของตัวแปรเป็นตัวแปรแบบจัดอันดับ⁴²

จากข้อกำหนดข้างต้น สามารถแสดงเป็นผังงานได้ดังต่อไปนี้



⁴² William L. Hays, Statistics for the Social Sciences, 1973, pp. 801-803.



การวิเคราะห์ความสอดคล้องอย่างง่าย

ขั้นตอนการคำนวณการวิเคราะห์ความสอดคล้อง

ข้อตกลงของสมการถดถอย

เหตุผลและควมจำเป็นในการตรวจสอบข้อตกลงของสมการถดถอย

ผลของการตรวจสอบ

1. ตรวจสอบการแจกแจงปกติหรือไม
2. ตรวจสอบค่าคาดหวังสหสัมพันธ์หรือไม
3. ตรวจสอบค่าความแปรปรวนคงที่หรือไม

