



โปรแกรมสำเร็จรูปทางสถิติ
สำหรับการวิเคราะห์การถดถอยเชิงเส้น

นายเชษฐพงษ์ เลิศวิทยาพนธ์
นายพรเทพ อุโฆษอารามิก
นายไพศาล ตีวรรณกิจ

ปัญหาพิเศษนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรบัณฑิต
ภาควิชาสถิติประยุกต์
คณะวิทยาศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ปีการศึกษา 2541

ปพ.
๗๗๒ ป
๒๕๔๑

เลขหมู่.....
เลขทะเบียน..... 32852
วัน, เดือน, ปี 14 ส.ย. 2542

Statistical Program for Linear Regression Analysis

Mr.Chettapong Lertwidthayapon
Mr.Pornthap U-kodarramik
Mr.Paisal Deewannakij

A special problem submitted in Partial Fulfillment of the Requirement
for
the Degree of Bachelor of Science

Department of Applied Statistics
Faculty of Science

King Monkut's Institute of Technology Ladkrabang
1998

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หน้าอนุมัติ

หัวข้อโครงการพิเศษ โปรแกรมสำเร็จรูปทางสถิติสำหรับการวิเคราะห์การถดถอย
เชิงเส้น

โดย นายเชษฐพงษ์ เลิศวิทยาพนธ์
นายพรเทพ อุโฆษอารามิก
นายไพศาล ตีวรรณกิจ

ภาควิชา สถิติประยุกต์

อาจารย์ที่ปรึกษา อ.วราพร เหลือสินทรัพย์

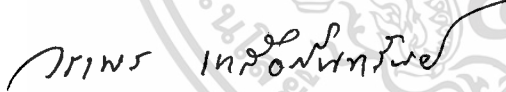
ภาควิชาสถิติประยุกต์ คณะวิทยาศาสตร์สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
อนุมัติให้นำปัญหาพิเศษฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร วิทยาศาสตร์บัณฑิต



(ผศ.วรารัตน์ เรืองรัตนเมธี)

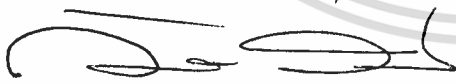
หัวหน้าภาควิชา

คณะกรรมการสอบโครงการพิเศษ



(อาจารย์วราพร เหลือสินทรัพย์)

ประธานกรรมการ



(ผศ.วีรศักดิ์ สุรพัฒน์)

กรรมการ



(อาจารย์ชลชาติ ตันติวานิช)

กรรมการ

ลิขสิทธิ์ของภาควิชา สถิติประยุกต์ คณะวิทยาศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อโครงการพิเศษ โปรแกรมสำเร็จรูปทางสถิติสำหรับการวิเคราะห์การถดถอยเชิงเส้น
โดย นายเชษฐพงษ์ เลิศวิทยพานิช
 นายพรเทพ อุโฆษอารามิก
 นายไพศาล ตีวรรณกิจ
ภาควิชา สถิติประยุกต์
อาจารย์ที่ปรึกษา อ.วราพร เหลือสินทรัพย์
ปีการศึกษา 2541

บทคัดย่อ

จุดประสงค์ของงานวิจัยนี้ เพื่อสร้างโปรแกรมสำเร็จรูปทางสถิติสำหรับการถดถอยเชิงเส้นอย่างง่าย ที่สามารถให้บุคคลทั่วไปใช้โปรแกรมและวิเคราะห์ข้อมูลได้โดยไม่ต้องมีความรู้ทางด้านสถิติมากนัก โปรแกรมที่จัดสร้างขึ้นนี้ จะเน้นความสำคัญของผลลัพธ์ที่ออกมาให้อยู่ในรูปที่สามารถเข้าใจได้ง่าย สรุปผลในการปฏิเสธหรือยอมรับสมมติฐานนั้นๆ อย่างทันที ในส่วนของการใช้งาน ก็เป็นเมนูตัวเลือกที่ไม่ซับซ้อนแสดงค่าสถิติและกราฟได้อย่างแม่นยำ และอำนวยความสะดวกในการใช้งานแก่บุคคลทั่วไปที่สนใจให้มากที่สุด จัดสร้างตรงตามเนื้อหาวิชา ซึ่งก็เหมาะกับนักศึกษาที่เริ่มเรียน สามารถใช้ประกอบการเรียนการสอนได้เป็นอย่างดี โดยการจัดสร้างโปรแกรมนี้ได้ใช้โปรแกรม BORLAND DELPHI Version 4.0 เป็นเครื่องมือในการพัฒนาโปรแกรม ซึ่งได้พัฒนาเพื่อให้ใช้ได้กับเครื่องคอมพิวเตอร์ตั้งแต่วรุ่น 80486 ขึ้นไป และดำเนินงานภายใต้ระบบปฏิบัติการ WINDOWS 95 ขึ้นไป

Special Project Title **Statistical Program for Linear Regression Analysis**

Name **Mr.Chettapong Lertwidthayapon**
Mr.Pornthap U-kodarramik
Mr.Paisal Deewannakij

Special Project Advisor **Waraporn Lursinsap**

Department **Applied Statistics**

Academic Year **1998**

Abstract

The purpose of this special project is to develop a program for Linear regression analysis. In order to help the people who have not much knowledge about statistics can analysis easily. This program gives the most interesting at output that understand easily, summary by themselves, provides accurate statistical value and user friendly interface. The Linear Regression Analysis Program developed from BORLAND DELPHI Version 4.0. The base computer system is IBM-PC Compatible 80486 up and Windows 95 up.

กิตติกรรมประกาศ

ในการทำวิจัยวิชาปัญหาพิเศษนี้จนสำเร็จลงได้ด้วยดี เพราะได้รับความกรุณาให้ความช่วยเหลือจากอาจารย์ที่ปรึกษา อาจารย์วรภาพ เหลือสินทรัพย์ ผู้ให้คำปรึกษาปัญหาต่างๆ สละเวลาอันมีค่าให้คำแนะนำและตรวจสอบแก้ไขข้อบกพร่องอย่างละเอียดยิ่ง ให้กำลังใจทั้งให้หนังสืออ้างอิงที่เป็นแหล่งข้อมูลที่ช่วยให้ทำงานได้รวดเร็วขึ้นมาก จนสำเร็จลุล่วงไปด้วยดีทุกประการ

กราบขอบพระคุณคณะกรรมการควบคุมปัญหาพิเศษ ผศ.วิรัชศักดิ์ สุรพัฒน์ และอาจารย์ดลชาติ ตันติวานิช และ ผู้ให้ประสบการณ์ที่ไม่อาจหาได้จากแหล่งความรู้ทั่วไป ให้คำแนะนำในการปรับปรุงแก้ไขตัวโปรแกรม ให้มีความสมบูรณ์ยิ่งขึ้นตลอดระยะเวลาในการทำปัญหาพิเศษครั้งนี้ และขอบพระคุณคณาจารย์ภาควิชาสถิติประยุกต์ทุกท่านเป็นอย่างสูง ที่ได้ประสิทธิ์ประสาทวิชาความรู้ จนสามารถประยุกต์ใช้เป็นงานที่มีคุณค่าขึ้นมาได้

สุดท้ายนี้ ต้องกราบเท้าขอบพระคุณคุณพ่อและคุณแม่ ที่มอบทุกสิ่งทุกอย่าง กำลังกาย กำลังใจ และกำลังทรัพย์ ให้สามารถมีชีวิตที่จะทำงานจนลุล่วงได้ตลอดไป

นายเชษฐพงษ์ เลิศวิทยาพนธ์

นายพรเทพ อูโฆษอารามิก

นายไพศาล ติวรรณกิจ

สารบัญ

	หน้า
บทคัดย่อปัญหาพิเศษภาษาไทย	ก
บทคัดย่อปัญหาพิเศษภาษาอังกฤษ	ข
กิตติกรรมประกาศ	ค
สารบัญ	ง
บทที่ 1 บทนำ	1
1.1 ความเป็นมาและความสำคัญของปัญหา	1
1.2 วัตถุประสงค์ในการศึกษา	2
1.3 ขอบเขตของการวิจัย	2
1.4 ขั้นตอนการดำเนินงาน	2
1.5 ประโยชน์ที่ได้รับ	3
1.6 อุปกรณ์ที่ใช้ในการวิจัย	3
บทที่ 2 ทฤษฎีและรายงานการวิจัยที่เกี่ยวข้อง	4
2.1 รายงานการวิจัยที่เกี่ยวข้อง	4
2.2 ทฤษฎีและหลักเกณฑ์ที่เกี่ยวข้อง	4
2.2.1 การวิเคราะห์การถดถอย	4
2.2.2 การถดถอยแบบเชิงเดียว	6
2.2.3 การประมาณค่าพารามิเตอร์โดยวิธีกำลังสองน้อยที่สุด	8
2.2.4 ความคลาดเคลื่อนมาตรฐานของการประมาณ	12
2.2.5 สัมประสิทธิ์ของตัวกำหนด	13
2.2.6 สัมประสิทธิ์ของตัวกำหนดที่ปรับค่าแล้ว	14
2.2.7 ความแปรปรวนและความคลาดเคลื่อนมาตรฐานของค่า b	14
2.2.8 ความแปรปรวนและความคลาดเคลื่อนมาตรฐานของค่า a	15
2.2.9 ความแปรปรวนและความคลาดเคลื่อนมาตรฐานของ ค่าประมาณ Y	15
2.2.10 การทดสอบสมมติฐาน	16
2.2.11 การทดสอบความแปรปรวน	17
2.2.12 การวิเคราะห์สหสัมพันธ์แบบเส้นตรงเชิงเดียว	17

2.2.13 การตรวจสอบและแก้ไขกรณีข้อมูลที่ไม่น่าเชื่อถือ	
ข้อกำหนดสำหรับการวิเคราะห์การถดถอยเชิงเส้นอย่างง่าย 19	
2.2.14 การวิเคราะห์สมการถดถอยเชิงเส้นที่ตัวแปรอิสระหนึ่งค่า	
มีค่าสังเกตหลายค่า	25
บทที่ 3 วิธีการดำเนินงาน	29
3.1 ศึกษาวิธีในการหาค่าการวิเคราะห์การถดถอย	29
3.2 ศึกษาโปรแกรมที่ใช้ในการวิจัย	29
3.3 การเก็บรวบรวมข้อมูล	29
3.4 การพัฒนาโปรแกรม	30
3.5 การตรวจสอบโปรแกรม	30
3.6 สรุปผลการวิจัย	30
บทที่ 4 ผลการวิจัยและการวิจารณ์	31
4.1 การใช้งานในส่วนต่าง ๆ ของโปรแกรม	31
4.2 ความถูกต้องของค่าที่ได้จากการคำนวณ	44
4.3 ส่วนการแสดงผล Output	44
บทที่ 5 บทสรุป	55
ภาคผนวก	56
บรรณานุกรม	60

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในปัจจุบันนี้วิชาสถิติทางด้านสมการถดถอยเชิงเส้น ได้มีบทบาทต่อการทำงานของหน่วยงานและองค์กรต่าง ๆ แต่ด้วยความที่วิชานี้มีความหลากหลายและยากต่อการเข้าใจของคนทั่วไปที่ไม่ได้ทำการศึกษาโดยตรง จึงได้มีการนำคอมพิวเตอร์เข้ามาช่วยอำนวยความสะดวกในการวิเคราะห์ทางสถิติ โปรแกรมทางสถิติต่าง ๆ จึงได้เกิดขึ้นเพื่อช่วยลดความซับซ้อนของตัวสถิติ การสรุปผลจึงทำได้ง่ายขึ้น

ในการเรียนการสอนวิชา Regression ของภาควิชาสถิติประยุกต์ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ได้บรรจุเนื้อหาในการคำนวณทางสถิติ บางส่วนที่ไม่มีอยู่ในโปรแกรมทางสถิติที่ใช้อยู่ในสถาบันฯ เช่น การตรวจสอบค่าความแปรปรวนในกลุ่มตัวอย่างโดยใช้วิธี Goldfeld - Quandt Test และคณะผู้จัดทำคิดว่าการคำนวณทางสถิติเหล่านั้นมีประโยชน์ต่อการวิเคราะห์ผล จึงได้คิดที่จะพัฒนาโปรแกรมที่ตรงกับหลักสูตรการเรียนการสอนขึ้น ด้วยเหตุผลสองประการคือ

1. เนื่องด้วยโปรแกรมที่ได้พัฒนานี้เป็นไปตามหลักสูตรการเรียนการสอน จึงสามารถใช้โปรแกรมนี้ในการเรียนการสอนของสถาบันได้อย่างมีประสิทธิภาพ
2. โปรแกรมที่พัฒนาจะมีตัวสถิติใหม่ๆมากขึ้น ทำให้ผู้ใช้สามารถเลือกใช้ตัวสถิติได้มากกว่าเดิม

ในการพัฒนาโปรแกรมในปัจจุบัน ต่างก็คำนึงถึงความยืดหยุ่นของโปรแกรมเป็นสำคัญ และพบว่า Dynamic Link Library Programming (DLL Programming) นั้นได้มีบทบาทต่อการพัฒนาโปรแกรมในปัจจุบันเป็นอย่างมาก ทั้งนี้จะเห็นได้จากโปรแกรมคอมพิวเตอร์ที่นิยมใช้กันในปัจจุบันต่างก็ใช้เทคนิค DLL Programming ตัวอย่างเช่น ในโปรแกรม Windows 95, Windows NT ต่างก็ใช้เทคนิคการเขียนโปรแกรม DLL และเนื่องจากข้อดีของการพัฒนาโปรแกรมโดยใช้ DLL ทางคณะผู้จัดทำจึงได้ทำการพัฒนาโปรแกรมโดยใช้ DLL Programming เพื่อช่วยต่อการพัฒนาและแก้ไขโปรแกรมต่อไป

1.2 วัตถุประสงค์ในการศึกษา

วัตถุประสงค์ในการศึกษามีดังนี้

1. เพื่อสร้างโปรแกรมสำเร็จรูปทางสถิติที่คำนวณ เกี่ยวกับการวิเคราะห์การถดถอยเชิงเส้น
2. เพื่อนำโปรแกรมสำเร็จรูปที่สร้างขึ้นมาวิเคราะห์ข้อมูลในเชิงปริมาณ

1.3 ขอบเขตของการวิจัย

เนื่องจากเวลาในการสร้างโปรแกรมมีอย่างจำกัด โปรแกรมที่คณะผู้จัดทำจะทำนี้จะครอบคลุมเนื้อหาที่เกี่ยวกับ Simple Linear Regression เท่านั้น ซึ่งจะประกอบไปด้วย

1. การประมาณค่าพารามิเตอร์
2. การทดสอบค่าพารามิเตอร์
3. การทดสอบ Assumption ของ ϵ_i

1.4 ขั้นตอนการดำเนินงาน

ในการสร้างโปรแกรมสำเร็จรูปมีขั้นตอนดังนี้

1. กำหนดวัตถุประสงค์ของการศึกษาและกำหนดขอบเขตของการศึกษา
2. ศึกษาค้นคว้าเพิ่มเติมและดูความเป็นไปได้ในการพัฒนาโปรแกรม
3. ศึกษาทฤษฎีทางสถิติที่เกี่ยวข้องกับ Linear Regression
4. ออกแบบโปรแกรมคอมพิวเตอร์ให้เหมาะสม
5. เขียนโปรแกรมคอมพิวเตอร์
6. ตรวจสอบความถูกต้องของโปรแกรมและแก้ไขเพื่อเพิ่มประสิทธิภาพของโปรแกรม
7. สรุปผล วิเคราะห์ปัญหาที่เกิดขึ้นและแนวทางในการแก้ไขปัญหา

1.5 ประโยชน์ที่ได้รับ

ประโยชน์ที่จะได้รับเมื่อสร้างโปรแกรมสำเร็จรูปมีดังนี้

1. สามารถใช้โปรแกรมคอมพิวเตอร์ในการวิเคราะห์คำนวณ ค่าทางสถิติได้มีประสิทธิภาพยิ่งขึ้น
2. สามารถพัฒนาโปรแกรมได้อย่างมีประสิทธิภาพต่อไป โดยไม่จำเป็นต้องเขียนโปรแกรมใหม่ทั้งหมด
3. สามารถ Link กับโปรแกรมอื่นได้ เช่น Microsoft Word

1.6 อุปกรณ์ที่ใช้ในการวิจัย

ในการสร้างโปรแกรมสำเร็จรูปมีอุปกรณ์ที่ใช้สำหรับวิจัยดังนี้

1. คอมพิวเตอร์รุ่น Pentium 166 ขึ้นไป
2. โปรแกรม Borland Delphi 4.0 Client/Server
3. เครื่องพิมพ์ (Printer)

บทที่ 2

ทฤษฎีและรายงานการวิจัยที่เกี่ยวข้อง

ทฤษฎีที่เกี่ยวข้อง รวมทั้งบทความและรายงานการวิจัยที่เกี่ยวข้องในการสร้างโปรแกรม มีรายละเอียดดังนี้

2.1 รายงานการวิจัยที่เกี่ยวข้อง

จินตนา ชนะกุล และคณะ (2540) ได้สร้างโปรแกรมพยากรณ์ค่าอนุกรมเวลาด้วยวิธีทำให้เรียบ โดยใช้โปรแกรม Delphi ภายใต้โปรแกรมควบคุม Windows 95 โดยมีเนื้อหาในการใช้งานดังนี้

1. Simple Average
2. Single Moving Average
3. Double Moving Average
4. Single Exponential Smoothing
5. Double Exponential Smoothing

ซึ่งโปรแกรมที่ถูกพัฒนาขึ้นมานี้เป็นโปรแกรมที่แก้ไขส่วนบกพร่องของ โปรแกรม Forecast Plus ที่มีอยู่เดิม จากที่สามารถใช้ได้กับคอมพิวเตอร์เฉพาะรุ่น 286 เท่านั้น

2.2 ทฤษฎีและหลักเกณฑ์ที่เกี่ยวข้อง

2.2.1 การวิเคราะห์การถดถอย (Regression Analysis)

การถดถอยเป็นวิธีการทางสถิติอย่างหนึ่ง ที่ใช้ในการตรวจหาหรือตรวจสอบลักษณะของความสัมพันธ์ระหว่างข้อมูลตั้งแต่ 2 ชุดขึ้นไป ซึ่งอาศัยความรู้ ประสบการณ์ และการทดลองที่แล้วๆ มาเป็นหลัก โดยมีวัตถุประสงค์ว่าจะนำเอาลักษณะความสัมพันธ์อันนี้ ไปใช้ให้เป็นประโยชน์ในอนาคต การวิเคราะห์การถดถอยเป็นเทคนิคในการวิเคราะห์ความสัมพันธ์ของตัวแปรตั้งแต่ 2 ตัวแปรขึ้นไป จากความสัมพันธ์ของตัวแปรทำให้สามารถประมาณค่าของตัวแปรหนึ่งจากตัวแปรอื่นๆ ในการวิเคราะห์การถดถอย จะต้องกำหนดลักษณะของตัวแปรซึ่งแบ่งออกเป็น 2 ชนิดคือ

1. ตัวแปรอิสระ (Independent or Predictor Variables) เป็นตัวแปรที่ส่งผลกระทบต่อตัวแปรตาม หรือเป็นปัจจัยในการกำหนดความคุมตัวแปรตาม ตัวแปรอิสระนี้อาจเป็นตัวแปรไม่ต่อเนื่อง หรือต่อเนื่อง ซึ่งอาจมีมาตรวัดแบบช่วง (Interval Scale) มาตรวัดแบบอัตราส่วน (Ratio Scale) หรือมีโค้ด (Code) เป็น 0, 1 ก็ได้ ตัวแปรอิสระนี้จะใช้สัญลักษณ์ X
2. ตัวแปรตาม (Dependent or Response Variable) เป็นตัวแปรที่ขึ้นอยู่กับตัวแปรอิสระ ซึ่งพบว่าถ้าค่าของตัวแปรอิสระเปลี่ยนแปลงไป จะมีผลทำให้ตัวแปรตามมีค่าเปลี่ยนแปลงไปด้วย ตัวแปรตามจะต้องเป็นตัวแปรแบบต่อเนื่อง ซึ่งอาจมีมาตรวัดแบบช่วง (Interval Scale) หรือมาตรวัดแบบอัตราส่วน (Ratio Scale) ตัวแปรตามนี้จะใช้สัญลักษณ์ Y

ในการวิเคราะห์การถดถอยแบ่งออกเป็น 3 ประเภทดังนี้

1. การถดถอยเชิงเดียว (Simple Regression)

เป็นการศึกษาความสัมพันธ์ระหว่างตัวแปร 2 ชุดเท่านั้น โดยตัวแปรชุดหนึ่งเป็นปัจจัยตาม และอีกชุดหนึ่งเป็นปัจจัยอิสระ หรือจะกล่าวง่ายๆ ได้อีกอย่างหนึ่ง คือเป็นการหาความสัมพันธ์ของข้อมูลที่พิจารณา ให้มีตัวแปรอิสระเพียงตัวเดียวเท่านั้น

2. การถดถอยเชิงซ้อน (Multiple Regression)

เป็นการศึกษาหาความสัมพันธ์ระหว่างตัวแปรตั้งแต่ 3 ชุดขึ้นไป โดยตัวแปรชุดหนึ่งเป็นปัจจัยตาม ส่วนตัวแปรอื่นๆ จะเป็นปัจจัยอิสระ หรือมีตัวแปรอิสระตั้งแต่ 2 ตัวขึ้นไป

3. การถดถอยแบบเส้นโค้ง (Curvilinear Regression)

ในการศึกษาความสัมพันธ์ของตัวแปรไม่ว่าจะเป็นแบบการถดถอยเชิงเดียวหรือการถดถอยเชิงซ้อนก็ตาม อาจมีความสัมพันธ์ได้ทั้งในรูปเส้นโค้งและเส้นตรง ถ้าความสัมพันธ์ที่หาได้เป็นรูปเส้นตรงก็เรียกว่าเป็นการถดถอยแบบเส้นตรงเชิงเดียว (Simple Linear Regression) หรือการถดถอยแบบเส้นตรงเชิงซ้อน (Multiple Linear Regression) แต่ลักษณะของข้อมูลที่ศึกษาบางประเภทอาจจะไม่เป็นเส้นตรงก็ต้องศึกษาในรูปของเส้นโค้ง ซึ่งเรียกว่าการถดถอยแบบเส้นโค้ง (Non Linear Regression) หรือ (Curvilinear Regression) ดังนั้นการศึกษการถดถอยแบบเส้นโค้งจึงมีการศึกษาทั้งที่เป็นแบบการถดถอยเชิงเดียว และแบบเชิงซ้อน ซึ่งในการจัดทำปัญหาพิเศษนี้ จะพิจารณาเพียงการถดถอยแบบเส้นตรงเชิงเดียวเท่านั้น

2.2.2 การถดถอยแบบเชิงเดียว (Simple Linear Regression)

ในการวิเคราะห์ความสัมพันธ์ของตัวแปรนั้น สิ่งที่น่ามาใช้เป็นหลักในการศึกษาก็คือ ข้อมูลที่แสดงความสัมพันธ์เกี่ยวกับตัวแปรนั้นๆ นั่นเอง ข้อมูลเหล่านี้อาจได้มาจากการทดลองเก็บรวบรวมหรือสอบถามก็เป็นได้ เมื่อได้ข้อมูลแล้วจำเป็นต้องนำข้อมูลมาพิจารณาอย่างคร่าวๆ เสียก่อนว่าตัวแปรเหล่านี้จะมีความสัมพันธ์กันหรือไม่ และสัมพันธ์กันในรูปใด เส้นตรงหรือเส้นโค้ง ในการพิจารณาความสัมพันธ์ของตัวแปรอย่างง่าย ๆ นั้น นิยมนำข้อมูลมา Plot ในแผ่นกราฟ โดยให้แกนตั้งแทนด้วยค่าตัวแปรตาม และแกนนอนแทนด้วยค่าตัวแปรอิสระ ซึ่งเรียกว่าแผนภาพกระจาย (Scatter Plot Diagram) จุดที่ Plot ได้บนแผนภาพกระจาย ถ้าลากเส้นผ่านจุดเหล่านี้ก็จะได้เส้นซึ่งถือว่าเป็นตัวแทนที่แสดงลักษณะหรือแนวโน้มของความสัมพันธ์เรียกว่า Regression Line ซึ่งจะเป็นเส้นตรงหรือเส้นโค้งหรือไม่นั้นย่อมแล้วแต่ลักษณะจุดที่ Plot ได้บนแผนภาพกระจายเมื่อกำหนด α เป็นค่าตัดแกน Y และ β เป็น Slope สามารถเขียนสมการเส้นตรงได้ว่า

$$\mu_{Y.X} = \alpha + \beta X_i$$

เมื่อ $\mu_{Y.X}$ คือค่า Y บนเส้นสมการถดถอยและเรียกสมการนี้ว่า สมการถดถอยแบบเส้นตรงเชิงเดียว (Simple Linear Regression Equation)

2.2.2.1 กรณีที่ข้อมูลเป็นประชากร (Population Data)

ค่า Y ต่างๆ ที่เปลี่ยนแปลงไปตามค่า X แต่ละค่าเรียกว่าประชากรย่อย (Sub-Population) เมื่อนำค่า Y ในแต่ละประชากรย่อยมาหาค่าเฉลี่ยแล้ว เรียกค่านั้นว่าค่าคาดหวังของ Y ที่กำหนดให้ X คงที่ (Expected Value of Y for Given X) เขียนสัญลักษณ์ว่า $E(Y/X)$ จะเห็นว่าเส้นตรงที่ลากผ่าน $E(Y/X)$ นี้เป็นเส้นตรงที่แสดงลักษณะความสัมพันธ์ระหว่างค่าเฉลี่ยของ Y ในแต่ละประชากรย่อยกับค่า X นั่นเอง และเรียกเส้นนี้ว่าเส้นการถดถอย สมการเส้นการถดถอยจึงเขียนได้ว่า

$$\mu_{Y.X} = \alpha + \beta X_i$$

เมื่อ $\mu_{Y.X}$ คือค่า Y บนเส้นสมการถดถอยโดยมีแบบหุ่น (Model) คือ

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

เมื่อ ε_i คือค่าความคลาดเคลื่อน

ลักษณะสำคัญบางประการของค่าคลาดเคลื่อน

1. ค่าคลาดเคลื่อนในแต่ละประชากรย่อยจะมีความแปรปรวน (Variance) คงที่ โดยให้ $\sigma^2 =$ ความแปรปรวนของค่า ε_i
2. ค่าเฉลี่ยของค่าคลาดเคลื่อนในแต่ละประชากรย่อย มีค่าเท่ากับศูนย์
3. การกระจายของค่าคลาดเคลื่อนในแต่ละประชากรย่อยมีลักษณะเป็นโค้งปกติ (Normal Distribution)
4. ค่าคลาดเคลื่อน เป็นค่าที่รวมเอาอิทธิพลตัวแปรอิสระอื่นๆ ที่มีต่อค่า Y_i อันนอกเหนือไปจากตัวแปรอิสระที่เราพิจารณาอยู่ในสมการ

2.2.2.2 กรณีที่ข้อมูลเป็นตัวอย่าง (Sample Data)

เนื่องจากในทางปฏิบัติเราไม่สามารถจะเก็บข้อมูลของค่า Y ได้ทุกค่า ในแต่ละประชากรย่อย เพราะปกติจะทราบค่า Y ได้เพียงบางค่าในแต่ละค่าของ X ซึ่งถือว่าเป็นข้อมูลตัวอย่าง ดังนั้นการศึกษาความสัมพันธ์ระหว่าง X และ Y เมื่อทราบข้อมูลเป็นเพียงตัวอย่าง จึงไม่ใช่ความสัมพันธ์ระหว่าง X และ Y อย่างแท้จริง เป็นเพียงความสัมพันธ์ที่คาดคะเนเท่านั้น เรียกว่าการวิเคราะห์การถดถอยแบบเส้นตรงเชิงเดียวของข้อมูลตัวอย่าง (Sample Simple Linear Regression Analysis) ให้สมการการถดถอยที่ลากผ่านค่า Y ของข้อมูลตัวอย่างเป็น

$$\hat{Y}_i = a + bX_i$$

เรียกสมการนี้ว่า Sample Simple Linear Regression Equation ซึ่งแบบหุ่่นของความสัมพันธ์ระหว่าง Y_i กับ สมการดังกล่าวคือ

$$Y_i = a + bX_i + e_i$$

เมื่อ e_i คือ Error Term และเรียกแบบพจน์นี้ว่า Sample Simple Linear Regression Model โดย

- a คือ ค่าที่เส้นการถดถอยตัดแกน Y (Y-Intercept)
- b คือ ค่าความชัน เรียกว่า สัมประสิทธิ์ของการถดถอย
(Coefficient of Regression)

เมื่อเปรียบเทียบเส้นการถดถอยของ Y บนค่า X ที่แท้จริง ซึ่งมีสมการเป็น

$$\mu_{Y.X} = \alpha + \beta X_i \dots\dots\dots(1)$$

กับเส้นสมการถดถอยของ Y บนค่า X ที่ได้จากข้อมูลตัวอย่างซึ่งสมการเป็น

$$\hat{Y}_i = a + bX_i \dots\dots\dots(2)$$

แล้วจะเห็นว่าเส้นการถดถอยของสมการ (2) เป็นเพียงใช้คาดคะเนเส้นการถดถอยของสมการ (1) เท่านั้นเอง ดังนั้นในวิธีการทางสถิติจะต้องพยายามหาสมการ (2) สำหรับคาดคะเนสมการ (1) ที่ดีที่สุดซึ่งจะต้องมีคุณสมบัติดังนี้

1. เป็นตัวคาดคะเนที่ไม่มีความคลาดเคลื่อน (Unbiased Estimator)
2. ความแปรปรวน (Variance) ของตัวคาดคะเน (Estimator) จะต้องต่ำที่สุด โดยจะเป็นตามคุณสมบัติดังกล่าวจะต้องใช้วิธี The Method of Least Squares ดังแสดงในหัวข้อ 2.2.3 ต่อไป

2.2.3 การประมาณค่าพารามิเตอร์โดยวิธีกำลังสองน้อยที่สุด (Estimating The Model Parameters : The Method of Least Squares)

จากตัวแบบการถดถอยเชิงเส้นตรงอย่างง่าย

$$Y_i = \alpha + \beta X_i + \varepsilon_i \quad : i = 1, 2, 3, \dots, n$$

ข้อสมมติ (Assumption) $\varepsilon_i \sim N(0, \sigma^2)$

การที่จะหาค่า α และ β ได้ถูกต้องมากที่สุดก็คือจะต้องมีค่า ε_i น้อยที่สุด นั่นคือให้ผลรวมกำลังสองของความแตกต่างระหว่างค่าสังเกต Y_i กับเส้นตรง ($\mu_{Y.X}$) มีค่าน้อยที่สุด ซึ่งสามารถทำได้โดยวิธีกำลังสองน้อยที่สุด

$$\varepsilon_i = Y_i - \alpha - \beta X_i$$

$$\sum_{i=1}^n (\varepsilon_i)^2 = \sum_{i=1}^n (Y_i - \alpha - \beta X_i)^2$$

การทำให้ $\sum_{i=1}^n \varepsilon_i^2$ มีค่าน้อยที่สุดโดยการทำอนุพันธ์บางส่วน (Partial Derivatives) เทียบกับ α และ β ตามลำดับ แล้วให้มีค่าเท่ากับ 0

$$\frac{\partial \sum_{i=1}^n (\varepsilon_i)^2}{\partial \alpha} = 2 \sum_{i=1}^n (Y_i - \alpha - \beta X_i)(-1) = 0 \quad (1)$$

$$\frac{\partial \sum_{i=1}^n (\varepsilon_i)^2}{\partial \beta} = 2 \sum_{i=1}^n (Y_i - \alpha - \beta X_i)(-X_i) = 0 \quad (2)$$

แทนค่า a และ b เป็นตัวประมาณค่าของ α และ β ใน (1) และ (2)

จาก (1)

$$2 \sum_{i=1}^n (Y_i - a - bX_i)(-1) = 0$$

$$\sum_{i=1}^n (-Y_i + a + bX_i) = 0$$

$$-\sum_{i=1}^n Y_i + \sum_{i=1}^n a + \sum_{i=1}^n bX_i = 0$$

$$na + b \sum_{i=1}^n X_i = \sum_{i=1}^n Y_i \quad (3)$$

จาก (2)

$$2 \sum_{i=1}^n (Y_i - a - bX_i)(-X_i) = 0$$

$$\sum_{i=1}^n (-X_i Y_i + aX_i + bX_i^2) = 0$$

$$- \sum_{i=1}^n X_i Y_i + a \sum_{i=1}^n X_i + b \sum_{i=1}^n X_i^2 = 0$$

$$a \sum_{i=1}^n X_i + b \sum_{i=1}^n X_i^2 = \sum_{i=1}^n X_i Y_i \quad (4)$$

สมการ (3) และ (4) เรียกว่า สมการปกติ (Normal Equations) สามารถแก้หาค่า a และ b

$$na + b \sum_{i=1}^n X_i = \sum_{i=1}^n Y_i$$

$$na = \sum_{i=1}^n Y_i - b \sum_{i=1}^n X_i$$

$$a = \frac{\sum_{i=1}^n Y_i}{n} - \frac{b \sum_{i=1}^n X_i}{n}$$

$$a = \bar{Y} - b\bar{X}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จาก (4)

$$a \sum_{i=1}^n X_i + b \sum_{i=1}^n X_i^2 = \sum_{i=1}^n X_i Y_i$$

$$(\bar{Y} - b\bar{X}) \sum_{i=1}^n X_i + b \sum_{i=1}^n X_i^2 = \sum_{i=1}^n X_i Y_i$$

$$(\bar{Y} - b\bar{X})n\bar{X} + b \sum_{i=1}^n X_i^2 = \sum_{i=1}^n X_i Y_i$$

$$n\bar{X}\bar{Y} - nb\bar{X}^2 + b \sum_{i=1}^n X_i^2 = \sum_{i=1}^n X_i Y_i$$

$$b \left(\sum_{i=1}^n X_i^2 - n\bar{X}^2 \right) = \sum_{i=1}^n X_i Y_i - n\bar{X}\bar{Y}$$

$$b = \frac{\sum_{i=1}^n X_i Y_i - n\bar{X}\bar{Y}}{\sum_{i=1}^n X_i^2 - n\bar{X}^2} = \frac{SS_{XY}}{SS_{XX}}$$

คุณสมบัติกำลังสองน้อยที่สุดของการถดถอย

(Least Squares Regression Properties)

1. ผลรวมของค่าความคลาดเคลื่อนมีค่าเท่ากับ 0

$$\sum_{i=1}^n \varepsilon_i = \sum_{i=1}^n (Y_i - \hat{Y}_i) = 0$$

2. ผลรวมกำลังสองของค่าความคลาดเคลื่อนมีค่าต่ำสุด

$$\sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \text{ค่าต่ำสุด}$$

3. เส้นการถดถอยอย่างง่าย (Simple Regression Line) จะลากผ่านจุดตัดของค่าเฉลี่ย Y (\bar{Y}) และค่าเฉลี่ย X (\bar{X})
4. a และ b เป็นตัวประมาณค่าที่ดีที่สุดสำหรับ α และ β ตามลำดับ เพราะไม่มีการเอนเอียง (Unbiased Estimator)

การตีความหมายของ ค่า a และ b

1. ในกรณีค่าของ a คือจุดตัดบนแกน Y เมื่อ $X = 0$

ถ้า $a = 0$ จะได้ $Y = 0$ และ $X = 0$

$a > 0$ จะได้ $Y > 0$

$a < 0$ จะได้ $Y < 0$

นอกจากนี้ค่าของ a อาจจะไม่มีความหมายก็ได้ ถ้าค่าตัวแปรอิสระของข้อมูลที่ศึกษาไม่มีค่าเป็น 0 หรือขึ้นอยู่กับความหมายของข้อมูลที่สนใจ

2. ในกรณีของ b คือ เป็น slope ของเส้นการถดถอย ถ้าตัวแปรอิสระมีค่าเปลี่ยนแปลงไป 1 หน่วยจะมีผลทำให้ตัวแปรตามเปลี่ยนแปลงไปเท่ากับค่า b

2.2.4 ความคลาดเคลื่อนมาตรฐานของการประมาณ

(Standard Error of Estimate : $s_{Y.X}$)

การใช้เส้นการถดถอยนำไปคาดคะเนค่า Y ในอนาคตนั้นจะถูกต้องแค่ไหนขึ้นอยู่กับข้อมูลที่รวบรวมมาได้นั้น จะกระจัดกระจายไปจากเส้นการถดถอยที่คำนวณได้มากน้อยเพียงใด กล่าวคือถ้าข้อมูลกระจัดกระจายห่างจากเส้นการถดถอยมาก การคาดคะเนจะมีโอกาสที่จะแตกต่างไปจาก Y ที่

$e_i = Y_i - \hat{Y}_i$ และ $\frac{\sum e_i^2}{d.f.} = \frac{\sum (Y_i - \hat{Y}_i)^2}{n-2}$ ซึ่งเรียกว่าความแปรปรวนของการประมาณ (Variance

of Estimate) มาตรการที่วัดการกระจายของข้อมูลที่รวบรวมมาได้นี้จะวัดอยู่ในรูปของความแตกต่างหรืออยู่ในรูปของค่า e นั้นเอง

จึงได้เป็นสูตรว่า

$$s_{Y.X} = \sqrt{\frac{\sum (Y_i - \hat{Y}_i)^2}{n-2}}$$

$s_{Y.X}$ เป็นความคลาดเคลื่อนมาตรฐาน (Standard Error of Estimate)

$s_{Y.X}^2$ เป็นความแปรปรวนของค่า Y บน X

2.2.5 สัมประสิทธิ์ของตัวกำหนด (Coefficient of Determination : R^2)

สมการที่ดีที่สุดที่สามารถนำไปคาดคะเนค่า Y ในอนาคตโดยมีโอกาที่จะถูกต้องตรงกับค่า Y ที่แท้จริงได้มากนั้น จะต้องสามารถอธิบายการกระจายที่เกิดขึ้นในค่า Y ได้มากที่สุด นั่นคือผลรวมกำลังสองของการกระจายที่จะอธิบายได้ จะต้องมามีค่าใกล้เคียงกับผลรวมกำลังสองของการกระจายทั้งหมด และถ้าผลรวมกำลังสองของการกระจายที่อธิบายได้ เท่ากับผลรวมกำลังสองของการกระจายทั้งหมดแล้ว สมการถดถอยจะสามารถอธิบายการกระจายของ Y ทั้งหมดได้ถึง 100% ถ้านำสมการไปคาดคะเนค่า Y ในอนาคตจะได้ค่าถูกต้องกับความจริง 100% เช่นกัน จึงมีการวัดอัตราส่วนระหว่างผลรวมกำลังสองของการกระจายที่อธิบายได้กับผลรวมกำลังสองของการกระจายทั้งหมด ซึ่งเรียกว่า สัมประสิทธิ์ของตัวกำหนด (Coefficient of Determination , R^2)

$$R^2 = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2}$$

คุณสมบัติของสัมประสิทธิ์ของตัวกำหนดมีดังนี้

1. มีค่าเป็นบวกเสมอ
2. มีค่าตั้งแต่ 0-1 หรือ 0% ถึง 100%

ประโยชน์ของสัมประสิทธิ์ของตัวกำหนด คือ

1. เป็นเครื่องวัดความใกล้ชิดระหว่างเส้นการถดถอยกับค่า Y คือ

$R^2 = 1$ แสดงว่าค่า Y จะอยู่บนเส้นการถดถอยทุกจุด

$R^2 > 0$ แสดงว่าค่า Y จะอยู่ใกล้ชิดกับเส้นการถดถอย

$R^2 = 0$ แสดงว่าค่า Y จะกระจายห่างกับเส้นการถดถอยมากจนหาแนวโน้มที่ถูกต้องไม่ได้

2. เป็นเครื่องวัดความเหมาะสมของเส้นการถดถอยว่า จะแสดงแนวโน้มของข้อมูลได้มากน้อยเพียงใด คือ

R^2 ใกล้ 1 แสดงว่าแนวโน้มของข้อมูลจะมีลักษณะแนวเส้นตรงมากที่สุด

R^2 ใกล้ 0 แสดงว่าแนวโน้มของข้อมูลจะมีลักษณะไม่เป็นเส้นตรงคืออาจโค้งก็ได้

3. ในทางปฏิบัติใช้เป็นเครื่องแสดงอิทธิพลของตัวแปรอิสระกับ ตัวแปรตามว่าการกระจายของค่า Y นั้นสามารถอธิบายได้จากเส้นการถดถอยที่เปอร์เซ็นต์

2.2.6 สัมประสิทธิ์ของตัวกำหนดที่ปรับค่าแล้ว

(Adjusted Coefficient of Determination : R^{-2})

ในกรณีข้อมูลที่น่ามาศึกษามีจำนวน n น้อยๆ การใช้สัมประสิทธิ์ของตัวกำหนด เป็นเครื่องชี้ว่าสมการการถดถอยอธิบายการกระจายของค่า Y ทั้งหมดได้เท่านั้นเท่านั้นเปอร์เซ็นต์นั้น อาจจะมีข้อผิดพลาดและมีความคลาดเคลื่อน ถ้าเรานำสมการการถดถอยนี้ไปทดสอบนัยสำคัญ (Significance) เช่นในกรณี n น้อย ถึงแม้คำนวณ ได้ 90% เมื่อทดสอบสมการแล้วอาจไม่มีนัยสำคัญก็ได้ เพื่อขจัดปัญหาอันนี้ จึงต้องนำมาปรับเสียใหม่ เรียกว่า Adjusted Coefficient of Determination ซึ่งมีสูตรดังนี้

$$R^{-2} = 1 - \frac{\sum (Y_i - \hat{Y}_i)^2}{\sum (Y_i - \bar{Y})^2} \cdot \frac{n-1}{n-2}$$

2.2.7 ความแปรปรวนและความคลาดเคลื่อนมาตรฐานของค่า b (S_b^2, S_b)

ค่า S_b^2 และ S_b นี้เป็นค่าสถิติที่สำคัญอันหนึ่ง เพราะเป็นค่าที่ใช้เป็นหลักประกันว่าสมการ

$\hat{Y}_i = a + bX_i$ นั้นจะเป็นตัวแทนที่ดีของสมการการถดถอย $\mu_{Y.X} = \alpha + \beta x_i$ ได้มากน้อยเพียงใด กล่าวคือ ถ้าค่านี้มีค่าน้อยเส้นสมการการถดถอยตัวอย่าง ย่อมจะมีแนวโน้มใกล้เคียงกันกับเส้นสมการถดถอยประชากรมาก จึงหวังได้ว่าสมการการถดถอยตัวอย่างที่มีค่าดังกล่าวนี้ น้อย จะเป็นตัวคาดคะเนที่ดีของประชากรนั้นได้ จากสูตร

$$S_b^2 = \frac{S_{Y.X}^2}{\sum X_i^2} \text{ เป็นค่าความแปรปรวนของ } b$$

$$S_b = \frac{S_{Y.X}}{\sqrt{\sum X_i^2}} \text{ เป็นค่าความคลาดเคลื่อนมาตรฐานของ } b$$

2.2.8 ความแปรปรวนและความคลาดเคลื่อนมาตรฐานของค่า a (S_a^2, S_a)

ค่า s_a^2 นี้เป็นเครื่องมือที่ใช้วัดว่าค่า a นั้นจะกระจายอยู่ห่างจากค่าพารามิเตอร์มากน้อยเพียงใด กล่าวคือ ถ้ามีค่าน้อยย่อมแสดงว่าค่า a มีการกระจายใกล้ชิด กับค่าพารามิเตอร์มาก ค่า a ก็จะเป็นตัวคาดคะเนที่ดีของ α

$$S_a^2 = S_{Y.X}^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum x_i^2} \right) \text{ เป็นความแปรปรวนของค่า } a$$

$$S_a = S_{Y.X} \sqrt{\left(\frac{1}{n} + \frac{\bar{x}^2}{\sum x_i^2} \right)} \text{ เป็นความคลาดเคลื่อนของค่า } a$$

2.2.9 ความแปรปรวนและความคลาดเคลื่อนมาตรฐานของค่าประมาณ Y

$$(S_{\hat{Y}}^2, S_{\hat{Y}})$$

ความแปรปรวนและความคลาดเคลื่อนมาตรฐานของค่าประมาณ Y แบ่งออกเป็น 2 ค่าด้วยกันซึ่งย่อมาแล้วแต่การนำค่านี้ไปคาดคะเนอะไรคือ

1. ความแปรปรวนและความคลาดเคลื่อนมาตรฐานของค่าประมาณ Y บนเส้นการถดถอย ซึ่งความคลาดเคลื่อนอันนี้เกิดจากความแตกต่างระหว่างค่าประมาณ Y กับ ค่าเฉลี่ยของแต่ละค่า Y นั่นเอง ดังนั้นหากต้องการคาดคะเน ค่าเฉลี่ย ณ จุด X ใดๆ ก็ตามโดยใช้ค่าประมาณ Y เป็นตัวคาดคะเนจะต้องใช้ค่านี้เป็นตัววัดความคลาดเคลื่อน

$$S_{\hat{Y}}^2 = S_{Y.X}^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum x_i^2} \right) \text{ เป็นความแปรปรวนของ } \hat{Y}$$

$$S_{\hat{Y}} = S_{Y.X} \sqrt{\left(\frac{1}{n} + \frac{\bar{x}^2}{\sum x_i^2} \right)} \text{ เป็นความคลาดเคลื่อนมาตรฐานของ } \hat{Y}$$

2. ความแปรปรวนและความคลาดเคลื่อนมาตรฐานของค่าประมาณ Y บนจุดใดจุดหนึ่งของค่า Y ซึ่งความคลาดเคลื่อนอันนี้เกิดจากการที่นำค่าประมาณไปคาดคะเนค่า Y ให้สังเกตอย่างหนึ่งว่าความคลาดเคลื่อนของค่าประมาณ Y ตามข้อ 1 นั้นเกิดจากการที่นำค่าประมาณ Y ไปคาดคะเนค่าเฉลี่ยของ Y ที่ค่าใดค่าหนึ่งของ X แต่ความคลาดเคลื่อนข้อ 2

เกิดขึ้นเมื่อนำค่าประมาณ \hat{Y} ไป คัดคะแนนแต่ละค่าของ Y ดังนั้นความคลาดเคลื่อนกรณีนี้ จึงต้องรวมเอาความคลาดเคลื่อนของเส้นสมการถดถอยเข้าไปด้วย

$$s_{\hat{Y}-\hat{Y}}^2 = s_{Y.X}^2 \left(1 + \frac{1}{n} + \frac{x_i^2}{\sum x_i^2} \right) \text{ เป็นความแปรปรวนของ } \hat{Y}$$

$$s_{\hat{Y}-\hat{Y}} = s_{Y.X} \sqrt{\left(1 + \frac{1}{n} + \frac{x_i^2}{\sum x_i^2} \right)} \text{ เป็นความคลาดเคลื่อนมาตรฐานของ } \hat{Y}$$

2.2.10 การทดสอบสมมติฐาน (Test for Hypothesis)

การทดสอบสมมติฐานจะใช้การกระจายแบบ t เป็นหลักในการทดสอบแทนที่จะใช้การกระจายปกติมาตรฐาน อย่างไรก็ตามการทดสอบสมมติฐานนิยมทดสอบเพียง 2 ค่าคือ ค่าพารามิเตอร์ของ a และ b แต่ในทางปฏิบัตินั้นการทดสอบค่าพารามิเตอร์ของ b ถือว่าสำคัญที่สุด

2.2.10.1 การทดสอบค่าพารามิเตอร์ของ a

$$H_0 : \alpha = \alpha_0$$

$$H_a : \alpha \neq \alpha_0$$

คำนวณค่า

$$t_c = \frac{a - \alpha_0}{s_a}$$

จะปฏิเสธสมมติฐานหลักก็ต่อเมื่อ $t_c > t_{\frac{\alpha}{2}}$ หรือ $t_c < -t_{\frac{\alpha}{2}}$ ที่ $d.f. = n-2$

หมายเหตุ

ถ้า $\alpha_0 = 0$ และผลการทดสอบปรากฏว่าปฏิเสธสมมติฐานแย้ง จะได้

สมการถดถอยคือ $\hat{Y}_i = \beta X_i$

2.2.10.2 การทดสอบค่าพารามิเตอร์ b

$$H_0 : \beta = \beta_0$$

$$H_a : \beta \neq \beta_0$$

คำนวณค่า

$$t_c = \frac{b - \beta_0}{s_b}$$

จะปฏิเสธสมมติฐานหลักก็ต่อเมื่อ $t_c > t_{\frac{\alpha}{2}}$ หรือ $t_c < -t_{\frac{\alpha}{2}}$ ที่ d.f. = n-2

2.2.11 การทดสอบความแปรปรวน (Analysis of Variance)

เป็นวิธีการอย่างหนึ่งสำหรับทดสอบว่าค่าพารามิเตอร์ของ $b = 0$ หรือไม่นอกเหนือจากการใช้ t-test ถ้ายอมรับสมมติฐานหลัก แสดงว่าในด้านข้อมูลของประชากร X ย่อมไม่มีอิทธิพลต่อ Y มีขั้นตอนดังนี้

$$H_0 : \beta = 0 \quad (\text{X ไม่มีอิทธิพลต่อ Y})$$

$$H_0 : \beta \neq 0$$

คำนวณค่า

$$F_c = \frac{MSR}{MSE}$$

$$MSR = b \left(\sum X_i Y_i - \frac{\sum X_i \sum Y_i}{n} \right)$$

$$MSE = \left(\sum Y_i^2 - \frac{(\sum Y_i)^2}{n} \right) \frac{b \left(\sum X_i Y_i - \frac{\sum X_i \sum Y_i}{n} \right)}{n-2}$$

จะปฏิเสธสมมติฐานหลักก็ต่อเมื่อ $F_c > F_{\alpha, 1, n-1}$

2.2.12 การวิเคราะห์สหสัมพันธ์แบบเส้นตรงเชิงเดียว

(Simple Linear Correlation Analysis)

1. การวิเคราะห์สหสัมพันธ์แบบเส้นตรงเชิงเดียว มุ่งสนใจเฉพาะค่าสัมประสิทธิ์ของสหสัมพันธ์แบบเส้นตรงเชิงเดียว (Coefficient of Correlation) ใช้สัญลักษณ์ว่า r โดยศึกษาถึงการวัดหรือตรวจสอบหา ความมากน้อยของความสัมพันธ์ระหว่างตัวแปร 2 ชุดโดยไม่ต้องทราบว่าจะใครควรเป็นเหตุ อะไรควรเป็นผล
2. การวัดหรือตรวจสอบว่า สมการการถดถอยที่ศึกษานั้นเป็นสมการที่เหมาะสมกับข้อมูลหรือไม่ สูตรที่ใช้จำเป็นต้องกำหนดไว้ว่าข้อมูลที่นำมา วิเคราะห์มีการกระจายเป็นแบบ Bivariate Normal Distribution ซึ่งคิดโดยนาย Pearson ดังนี้คือ

$$r = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2}}$$

คุณสมบัติคือ r จะมีค่าตั้งแต่ -1 ถึง 1 ถ้ามีค่าเป็น $+$ หมายความว่า ตัวแปร 2 ตัวนั้นมีความสัมพันธ์ไปในทางเดียวกัน แต่ถ้าค่า r เป็น $-$ ย่อมหมายถึงตัวแปรทั้ง 2 มีความสัมพันธ์กันไปในด้านตรงข้าม

2.2.12.1 การทดสอบสมมติฐานของค่าพารามิเตอร์ของ r

1. การทดสอบค่า $\rho = 0$ โดยใช้การกระจายแบบ t พบว่าการกระจายของค่า r จะมีลักษณะใกล้เคียงกับการกระจายแบบ t โดยมีค่าเฉลี่ยของ

$$r = \rho = 0 \text{ และค่าความแปรปรวนของ } r = \sqrt{\frac{1-r^2}{n-2}}$$

$$H_0 : \rho = 0$$

$$H_a : \rho \neq 0$$

คำนวณค่า

$$t_c = \frac{r}{\sqrt{1-r^2/n-2}}$$

เมื่อ

$$r = \frac{\sum X_i Y_i - \sum X_i \sum Y_i / n}{\sqrt{[\sum X_i^2 - (\sum X_i)^2 / n][\sum Y_i^2 - (\sum Y_i)^2 / n]}}$$

จะปฏิเสธสมมติฐานหลักก็ต่อเมื่อ $t_c > t_{\alpha/2}$ หรือ $t_c < -t_{\alpha/2}$ ที่ d.f. = $n-2$

2. การทดสอบค่า $\rho \neq 0$

พบว่าการกระจายของค่า r ไม่มีลักษณะการกระจาย t จึงต้องแปลงให้เป็นค่า Z โดย

$$H_0 : \rho = \rho_0$$

$$H_a : \rho \neq \rho_0$$

คำนวณค่า

$$Z_c = \frac{Z_r - Z_{\rho}}{\sigma_r}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมื่อ

$$Z_r = \frac{1}{2} \ln \frac{1+r}{1-r}$$

$$Z_\rho = \frac{1}{2} \ln \frac{1+\rho_0}{1-\rho_0}$$

$$\sigma_r = \frac{1}{\sqrt{n-3}}$$

จะปฏิเสธสมมติฐานหลักก็ต่อเมื่อ

$$Z_c > Z_{\frac{\alpha}{2}} \text{ หรือ } Z_c < -Z_{\frac{\alpha}{2}}$$

2.2.13 การตรวจสอบและแก้ไขกรณีที่ข้อมูลไม่เป็นไปตามข้อกำหนดสำหรับการวิเคราะห์ความถดถอยเชิงเส้นอย่างง่าย (Detection and Correction of Model Violation for Simple Linear Regression Analysis)

ในการหาสมการถดถอยบางครั้งพบว่าสมการที่ได้นั้นอาจไม่ถูกต้อง และสิ่งที่ทำให้การวิเคราะห์สมการถดถอยผิดพลาดได้ คือ

2.2.13.1 ข้อมูลนอกรอบ (Outliers)

ซึ่งสาเหตุอาจเกิดจากการบันทึกข้อมูลผิดพลาดมีค่ามากหรือน้อยผิดปกติ และบ่อยครั้งพบว่าค่าคงเหลือ (Residuals) มีค่าสูง นอกจากข้อมูลนอกรอบที่มีอิทธิพลต่อสมการถดถอยแล้วยังพบว่าค่าสังเกตบางค่าที่มีผลกระทบเปลี่ยนแปลงสมการถดถอยได้ แต่ค่าคงเหลือมีค่าไม่สูง ค่าสังเกตนี้เรียกว่า Influential Point

การตรวจสอบหาข้อมูลนอกรอบและ Influential Point อาจทำได้โดย

1. พิจารณาค่าคงเหลือ (Residuals = $Y_i - \hat{Y}_i$) ถ้าพบว่ามีค่าคงเหลือ ค่าใดแตกต่างจากค่าคงเหลืออื่น ๆ แสดงว่าข้อมูลนี้มีข้อมูลนอกรอบเกิดขึ้น
2. พิจารณาจากแผนภาพการกระจาย (Scatter Diagram) ระหว่างค่าคงเหลือกับค่า \hat{Y}_i หรือ ระหว่างค่าคงเหลือกับค่า X_i ซึ่งแผนภาพนี้ยังช่วยชี้ให้เห็นถึงความแปรปรวนค่าคงที่หรือไม่ได้อย่างคร่าว ๆ เท่านั้น เพราะที่จริงแล้วควรพิสูจน์ข้อมูลว่ามีค่าความแปรปรวนค่าคงที่หรือไม่ โดยวิธี Goldfeld - Quandt Test และจะได้กล่าวต่อไปอีกครั้ง

3. ค่ามาตรฐานของค่าคงเหลือ (Standardized Residuals : Z_i)

$$Z_i = \frac{\hat{e}}{\sqrt{MSE}} \sim N(0,1)$$

MSE = Mean Square Error

\hat{e} = Residual

ค่า Z_i ที่คำนวณได้มากกว่า 2 หรือน้อยกว่า -2 แสดงว่าเป็นข้อมูลนอกกรอบ ด้วยความเชื่อมั่น 95%

4. Studentized residuals (r_i)

$$r_i = \frac{\hat{e}}{\sqrt{MSE(1-h_i)}} \sim t_{(n-2)} ; 0 \leq h_i \leq 1$$

MSE = Mean Square Error

\hat{e} = Residual

$|r_i|$ มากกว่า 2 (หรืออาจใช้มากกว่า 3 แล้วแต่ผู้วิเคราะห์) แสดงว่า มี outlier h_i คือค่า leverage ของค่าสังเกตที่ i

5. Leverage (h_i)

ในกรณีรูปแบบการถดถอยเชิงเส้นอย่างง่ายมี ตัวแปรอิสระ 1 ตัว จะได้ค่า Leverage คือ

$$h_i = \frac{1}{n} + \frac{(X_i - \bar{X})^2}{(n-1)S_x^2} ; 0 \leq h_i \leq 1$$

$$\text{โดย } S_x^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

Hoaglin และ Welsch แนะนำว่า ถ้าค่าของ $h_i > 2 \frac{(k+1)}{n}$ แสดงว่าค่าสังเกตที่ i เป็นข้อมูลนอกกรอบ เมื่อ k = จำนวนตัวแปรอิสระ

2.2.13.2 ปัญหาที่พบว่าค่าสังเกตตัวแปรอิสระ และตัวแปรตามมีความสัมพันธ์ไม่ใช่เชิงเส้น

ซึ่งก็คือการกำหนดตัวแบบผิดพลาดไป ดังนั้นจึงต้องตรวจสอบตัวแบบก่อน โดยการพลอตระหว่างค่าคงเหลือกับตัวแปรอิสระ ซึ่งบางครั้งรูปแบบของข้อมูลที่ไม่เป็นเส้นตรง ก็สามารถที่จะแปลงให้เป็นรูปแบบเส้นตรงได้มีหลายรูปแบบ เช่น รูปแบบกำลัง รูปแบบเอ็กซ์โปเนนเชียล รูปแบบไฮเพอร์โบล่า รูปแบบเอ็กซ์โปเนนเชียลผกผัน เป็นต้น ซึ่งการวิเคราะห์การถดถอยจากรูปแบบเส้นตรงที่แปลงแล้วจะเป็นทำนองเดียวกันกับการวิเคราะห์การถดถอยเชิงเส้นตรงแบบง่าย เมื่อได้สมการถดถอยจากรูปแบบเส้นตรงที่แปลงแล้วจะต้องแปลงสมการถดถอยดังกล่าว ให้เป็นสมการถดถอยสำหรับรูปแบบไม่เป็นเส้นตรง แต่บางกรณีจะมีรูปแบบการถดถอยแบบไม่เป็นเส้นตรงที่แปลงเป็นเส้นตรงไม่ได้ เช่น รูปแบบการถดถอยแบบเอ็กซ์โปเนนเชียลตัดแปลงแบบ Gompertz แบบ Logistic เป็นต้น หรือรูปแบบการถดถอยแบบไม่เป็นเส้นตรงที่ไม่ประสงค์จะแปลงให้เส้นตรงนั้น การแก้สมการปกติทำได้ยากซึ่งจะต้องใช้วิธีการวิเคราะห์ทางคณิตศาสตร์ (Numerical Analysis) ทำการแก้สมการต่อไป

2.2.13.3 การตรวจสอบความแปรปรวนไม่คงที่ (Detection of Heteroscedasticity)

ถ้าความแปรปรวนของค่าความคลาดเคลื่อนไม่คงที่จะทำให้ละเมิดข้อสมมติในตัวแบบจำลองถดถอย ทำให้การทดสอบสมมติฐาน และการหาค่าความเชื่อมั่นผิดพลาด การตรวจสอบค่าความแปรปรวนคงที่ สามารถทำได้ดังนี้

1. ตรวจสอบจากแผนภาพการกระจายของค่าคงเหลือกับค่าพยากรณ์ (\hat{Y})

- ในกรณีที่ เป็น Poisson นั้นจะพบว่า Y มีค่าเฉลี่ยและค่าความแปรปรวนเท่ากัน คือ λ จะเปลี่ยนแปลงค่า Y เป็น \sqrt{Y} หรือ $\sqrt{Y+1}$

- ส่วนถ้า Y มีการแจกแจงแบบทวินาม (Binomial) ซึ่งมีค่าเฉลี่ยเป็น np และความแปรปรวนเป็น npq ดังนั้นจะแปลงข้อมูลค่า Y เป็น $\sin^{-1} \sqrt{Y}$

- ในกรณีที่ เป็น Multiplicative นั้นจะเปลี่ยนค่า Y เป็น $\text{Log}(Y)$ ซึ่งจะพบบ่อยมาก สำหรับข้อมูลทางด้านบริหารหรือเศรษฐศาสตร์ เช่น เกี่ยวกับราคา, เงินเดือน เป็นต้น

2. ตรวจสอบด้วยวิธีของโกลด์เฟลด์-ควอนต์ (Goldfeld-Quandt Test) (Doran Howard E., 1989) ถ้าในตัวแบบมีตัวแปรอิสระหลายตัว ให้ยึดตัวแปรอิสระที่มี

ความแปรปรวนมากเป็นหลักในการเรียงลำดับจากน้อยไปมาก แล้วดำเนินขั้นตอนดังนี้

1. แบ่งข้อมูลออกเป็น 2 ส่วนเท่า ๆ กัน ถ้าจำนวนข้อมูลเป็นเลขคี่ ให้ตัดข้อมูลชุดกลางออก 1 ตัว
2. สมมติฐานที่ทดสอบ คือ

$$H_0 : \sigma_1^2 = \sigma_2^2 \text{ (Homoscedasticity)}$$

$$H_a : \sigma_2^2 > \sigma_1^2 \text{ (Heteroscedasticity)}$$

3. คำนวณตัวสถิติโดยใช้ F-Ratio

$$F_{\text{calc}} = \frac{SSE_2 / (L - P)}{SSE_1 / (L - P)} = \frac{SSE_2}{SSE_1}$$

เมื่อ L = จำนวนข้อมูลที่อยู่ในแต่ละกลุ่ม

P = จำนวนพารามิเตอร์

SSE₁ = Sum Square of Error ที่มีค่าน้อย

SSE₂ = Sum Square of Error ที่มีค่ามาก

ปฏิเสธสมมติฐานหลัก (H_0) ถ้าค่า F_{calc} มีค่ามากกว่าค่า

$F_{\alpha, (L-P, L-P)}$ ที่เปิดตาราง

คำแนะนำ

ถ้าใช้ข้อมูลที่ไม่มีการเปลี่ยนแปลงข้อมูล (Untransformed Data) ทดสอบปรากฏว่าปฏิเสธ H_0 ให้เปลี่ยนแปลงข้อมูล (Transformed Data) แล้วทดสอบโดยวิธีนี้ใหม่

หมายเหตุ

เพื่อความสะดวกในการพัฒนาโปรแกรมในครั้งนี จะใช้เฉพาะการทดสอบสมมติฐานตั้งข้างบนนี้

2.2.13.4 มีการกระจายแบบไม่ปกติ

ซึ่งเป็นการละเมิดข้อสมมติ ที่ค่าคงเหลือจะต้องมีการกระจายปกติที่อาจมีผลต่อการทดสอบทางสถิติ เช่น การทดสอบ F และ T นอกจากนี้ยังมีผลต่อค่าช่วงความเชื่อมั่นและช่วงการพยากรณ์ได้ วิธีตรวจสอบการกระจายแบบไม่ปกตินี้ทำได้ดังนี้

1. การทดสอบโดยการวัดค่า Skewness และ Kurtosis (Peakedness – Flatness)

โดยปกติแล้วถ้าตัวแปรสุ่มใดมีการแจกแจงแบบปกติ (มีโค้งเป็นรูปประฆังคว่ำ) ตัวแปรสุ่มนั้นจะมีค่า ความเบ้ (Skewness) เท่ากับ 0 และค่าความโด่ง (Kurtosis) เท่ากับ 3 ซึ่งเรานิยมใช้ค่าทั้งสองนี้เป็นมาตรฐานเพื่อเปรียบเทียบดูว่าตัวแปรสุ่มที่สนใจมีการแจกแจงเป็นแบบปกติหรือไม่ สัมประสิทธิ์ทั้งสองดังกล่าว นิยามได้ดังนี้

Coefficient of Skewness : $\hat{\gamma}_1$

$$\hat{\gamma}_1 = \frac{\frac{1}{n} \sum (e_i - \bar{e})^3}{s_e^3}$$

ถ้า $\hat{\gamma}_1 > 0$ แสดงว่า โค้งเบ้ขวา

ถ้า $\hat{\gamma}_1 < 0$ แสดงว่า โค้งเบ้ซ้าย

ถ้า $\hat{\gamma}_1 = 0$ แสดงว่า โค้งสมมาตร

Coefficient of Kurtosis : $\hat{\gamma}_2$

$$\hat{\gamma}_2 = \frac{\frac{1}{n} \sum (e_i - \bar{e})^4}{s_e^4}$$

ถ้า $\hat{\gamma}_2 > 3$ แสดงว่า โค้งแบนกว่าปกติ

ถ้า $\hat{\gamma}_2 < 3$ แสดงว่า โค้งโด่งกว่าปกติ

ถ้า $\hat{\gamma}_2 = 3$ แสดงว่าโค้งสมมาตรคล้ายโค้งปกติ

2. การทดสอบของ Kolmogorov-Smirnov One Sample Test

การทดสอบนี้จะใช้เมื่อตัวแปรที่สนใจมีการแจกแจงแบบต่อเนื่อง และมีอำนาจการทดสอบมากกว่าการทดสอบไคสแควร์ เพราะการทดสอบไคสแควร์จำเป็นต้องรวมกลุ่มด้วยกันเมื่อ $E_i < 5$ ทำให้สูญเสียข้อมูลไป และการทดสอบนี้ใช้ได้ดีเมื่อ n เล็ก โดยมีสมมติฐานที่ตั้งคือ

H_0 : ประชากรมีการแจกแจงแบบปกติ

H_1 : ประชากรมีการแจกแจงแบบไม่ปกติ

ให้ $S_N(X)$ คือ ความถี่สะสมของข้อมูลที่สังเกตได้ N จำนวน ในรูปสัดส่วน

ให้ $S_N(X) = k/N$ เมื่อ k = จำนวนของค่าสังเกตที่เท่ากับหรือน้อยกว่า X

$F_0(X)$ เป็นความถี่ที่คาดหวังในรูปสัดส่วน หรือค่าสัดส่วนที่คาดหวังว่าจะมีคะแนนเท่ากับหรือน้อยกว่า X ภายใต้สมมติฐาน H_0

ให้ $D = \text{Maximum } |F_0(X) - S_N(X)|$ ค่า D นี้สามารถหาการแจกแจงตัวอย่างได้ และได้มีตารางสำเร็จรูปแสดงค่า D ภายใต้พื้นที่ต่างๆ

โดยที่ จะปฏิเสธ H_0 เมื่อ $D_{\text{compute}} \geq D_{\text{table}}$

2.2.12.5 ความเป็นอิสระของค่าความคลาดเคลื่อน

ปกติค่าความคลาดเคลื่อนจะต้องไม่มีความสัมพันธ์กัน แต่ถ้าพบว่าค่าความคลาดเคลื่อนมีความสัมพันธ์กัน เรียกว่า ออโธคอร์เรชัน (Autocorrelation)

การทดสอบ ออโธคอร์เรชัน โดยดูบินวัตสัน (Durbin Watson Test) J. Durbin และ G.S. Watson เป็นผู้คิดค้นวิธีนี้โดยให้ ρ เป็น สัมประสิทธิ์ที่แท้จริงของ ออโธคอร์เรชัน ระหว่าง ϵ_i และ ϵ_j และค่าสถิติ d ของ Durbin Watson คือ

$$d_{\text{calc}} = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2}$$

โดยให้ข้อสมมติ (Assumption) คือค่าคงเหลือจะมีการแจกแจงแบบปกติ ช่วงของค่าสถิติ d คือ $0 \leq d \leq 4$ และ ให้

1. ถ้าค่าคงเหลือไม่มีความสัมพันธ์กัน ค่าสถิติ $d \approx 2$
2. ถ้าค่าคงเหลือมีความสัมพันธ์กันในทิศทางบวก (Positive Autocorrelation) ค่าสถิติ $d < 2$ และถ้ามีความสัมพันธ์มากในทิศทางบวก ค่าสถิติ $d \approx 0$
3. ถ้าค่าคงเหลือมีความสัมพันธ์กันในทิศทางลบ (Negative Autocorrelation) ค่าสถิติ $d > 2$ และถ้ามีความสัมพันธ์มากในทิศทางลบ ค่าสถิติ $d \approx 4$

2.2.14 การวิเคราะห์สมการถดถอยเชิงเส้น ที่ตัวแปรอิสระ 1 ค่า มีค่าสังเกตหลายค่า

การวิเคราะห์สมการถดถอยเชิงเส้นอย่างง่ายในกรณีที่ตัวแปรอิสระ 1 ค่า ได้ค่าสังเกตหลายค่า จะทดสอบสมการถดถอยเชิงเส้นด้วย เอฟ (F Test for Lack of Fit) ข้อสมมติฐาน

1. ค่าสังเกต Y ทุกค่าจะต้องเป็นอิสระ
2. ค่าสังเกต Y_i จะมีการแจกแจงแบบปกติ
3. ค่าสังเกต Y_i จะมีค่าความแปรปรวนคงที่เท่ากับ σ^2

2.2.14.1 การทดสอบความเหมาะสมของการถดถอยเชิงเส้น (Testing The Adequacy of The Simple Linear Regression Model)

$$\text{Simple Linear Regression Model : } Y_i = \beta_0 + \beta_1 + e_i$$

ให้ k = จำนวนระดับของตัวแปรอิสระ X (Replication)

n_i = จำนวนค่าสังเกต (Y) ที่ระดับ i ของ X (Replicate)

$$n = \text{จำนวนค่าสังเกตทั้งหมด} = \sum_{j=1}^k n_j$$

Y_{ij} เป็นตำแหน่งที่ j ของค่าสังเกตที่ระดับ i ของ X

เมื่อ $i = 1, 2, \dots, k$ และ $j = 1, 2, \dots, n_i$

Y_i เป็นผลรวมของค่าสังเกต Y ที่ระดับ i ของ $X = \sum_{j=1}^{n_i} Y_{ij}$ •

\bar{Y}_i เป็นค่าเฉลี่ย Y ที่ระดับ i ของ $X = Y_i / n_i$

$Y_{..}$ เป็นผลรวมของค่าสังเกต Y ทั้งหมด (Grand Total)

$$Y_{..} = \sum_{i=1}^k \sum_{j=1}^{n_i} Y_{ij} = \sum_{i=1}^k Y_i$$

$\bar{Y}_{..}$ เป็นค่าเฉลี่ย Y ของค่าสังเกต Y ทั้งหมด (Overall Mean) และ Sum Square Equation :

$$\sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{..})^2 = \sum_{k=1}^k n_i (\hat{Y}_i - \bar{Y}_{..})^2 + \sum_{i=1}^k n_i (Y_i - \hat{Y}_i)^2 + \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2$$

นั่นคือ $SS_{yy} = SSR + SSLOF + SSPE$

ซึ่ง $SSE = SSLOF + SSPE$

$SSLOF = \text{Lack of Fit Sum of Squares}$

$SSPE = \text{Pure Error Sum of Squares}$

ตารางที่ 4.1 การวิเคราะห์ความแปรปรวน (Anova Table)

Source	d.f.	Sum of Squares	Mean Square	F
Total	n	$\sum_{i=1}^k \sum_{j=1}^{n_i} Y_{ij}^2$		
C.F.(Due to β_0)	1	$\frac{Y_{..}^2}{n} = n\bar{Y}_{..}^2$		
Total (Corr.)	n-1	$SS_{yy} = \sum_{i=1}^k \sum_{j=1}^{n_i} Y_{ij}^2 - n\bar{Y}_{..}^2$		
REG(Due to β_1)	1	$SSR = \hat{\beta}_1^2 SS_{xx} = \hat{\beta}_1 SS_{xy}$	MSR	$\frac{MSR}{MSE}$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.1 (ต่อ)

Source	d.f.	Sum of Squares	Mean Square	F
Error	n-2	$SSE = \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \hat{Y}_i)^2 = SS_{YY} - SSR$	MSE	
Lack of fit	k-2	SSLOF=SSE-SSPE	MSLOF	$\frac{MSLOF}{MSPE}$
Pure Error	n-k	$SSPE = \sum_{i=1}^k SS_i = \sum_{i=1}^k \sum_{j=1}^{n_i} Y_{ij}^2 - \sum_{i=1}^k \frac{Y_i^2}{n_i}$	MSPE	
At X ₁	n ₁ -1	$SS_1 = \sum_{j=1}^{n_1} Y_{1j}^2 - \frac{Y_{1.}^2}{n_1}$	S ₁ ²	
At X ₂	n ₂ -1	$SS_2 = \sum_{j=1}^{n_2} Y_{2j}^2 - \frac{Y_{2.}^2}{n_2}$	S ₂ ²	
⋮	⋮	⋮	⋮	⋮
At X _k	n _k -1	$SS_k = \sum_{j=1}^{n_k} Y_{kj}^2 - \frac{Y_{k.}^2}{n_k}$	S _k ²	

การทดสอบลำดับเอฟ (Ordered F-Tests)

1. การทดสอบความเหมาะสมของเส้นถดถอย (Test for Lack of Fit)

$$H_0 : \mu_i = \beta_0 + \beta_1 X_i$$

$$H_a : \mu_i \neq \beta_0 + \beta_1 X_i$$

$$F_{\text{calc}} = \frac{MSLOF}{MSPE}$$

ปฏิเสธ H_0 ถ้า $F_{\text{cal}} > F_{\alpha(k-2, n-k)}$ แสดงว่าสมการถดถอยนั้นไม่ใช่เส้นตรง
ยอมรับ H_0 ถ้า $F_{\text{cal}} < F_{\alpha(k-2, n-k)}$ แสดงว่าสมการถดถอยเชิงเส้นตรงนี้ สามารถชี้
ถึงความสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตาม

2. การทดสอบความสัมพันธ์ของตัวแปรอิสระและตัวแปรตาม (Test for Regression Relationship)

$$H_0 : \beta_1 = 0$$

$$H_a : \beta_1 \neq 0$$

$$F_{\text{calc}} = \frac{\text{MSR}}{\text{MSE}}$$

ปฏิเสธ H_0 ถ้า $F_{\text{calc}} > F_{\alpha(1, n-2)}$ แสดงว่าตัวแปรอิสระและตัวแปรตามมีความสัมพันธ์กัน

ยอมรับ H_0 ถ้า $F_{\text{calc}} < F_{\alpha(1, n-2)}$ แสดงว่าตัวแปรอิสระและตัวแปรตามไม่มีความสัมพันธ์กัน



บทที่ 3

วิธีการดำเนินงาน

3.1 ศึกษาวิธีการหาค่าการวิเคราะห์การถดถอย .

ศึกษาวิธีการหาค่าในสมการวิเคราะห์การถดถอย และสูตรคำนวณในการวิเคราะห์การถดถอยแต่ละวิธีโดยละเอียด และทำการดัดแปลงสูตรให้สามารถคำนวณค่าได้เร็วขึ้น และศึกษาความเป็นไปได้ในการนำคอมพิวเตอร์มาช่วยทำการคำนวณ

3.2 ศึกษาโปรแกรมที่ใช้ในการวิจัย

ศึกษาโปรแกรม SPSS ที่ใช้คำนวณการวิเคราะห์การถดถอย ว่ารูปแบบการอ่านข้อมูลมีค่าของข้อมูลดิบ ค่าทางสถิติ กราฟ และการทำงานอย่างไร เพื่อให้สามารถเข้าใจและเป็นแนวทางในการพัฒนาโปรแกรม ต่อจากนั้นจึงทำการศึกษาโปรแกรม Borland Delphi Version 4.0 ว่ามีวิธีการในการเขียนโปรแกรมอย่างไร เพื่อนำมาเขียนโปรแกรม

3.3 การเก็บรวบรวมข้อมูล

เก็บรวบรวมข้อมูลที่ใช้ในการวิเคราะห์การถดถอย เพื่อนำมาตรวจสอบโปรแกรมที่จะพัฒนาขึ้นมาใหม่

3.4 การพัฒนาโปรแกรม

แบ่งเป็นขั้นตอนดังนี้

- กำหนดความต้องการของโปรแกรมที่จะทำการพัฒนาทั้งหมด ว่าต้องการให้มีขอบเขตเท่าใดมีรูปแบบอย่างไร
- การออกแบบโปรแกรม เป็นการออกแบบที่เป็นลักษณะเมนูให้เลือกใช้ ซึ่งง่ายต่อการใช้งาน ในส่วนของการใส่ข้อมูลก็สามารถใส่ข้อมูลได้โดยง่าย สามารถแก้ไขข้อมูลได้ แสดงการคำนวณค่าต่าง ๆ และแสดงกราฟได้
- เขียนโปรแกรม

3.5 ตรวจสอบโปรแกรม

เนื่องจากโปรแกรมที่พัฒนาขึ้นมาใหม่นั้นอาจมีข้อผิดพลาดในการคำนวณได้ ดังนั้นจึงต้องมีการตรวจสอบค่าที่คำนวณได้

ในการตรวจสอบโปรแกรม จะทำโดยนำค่าการวิเคราะห์การถดถอยที่คำนวณจากโปรแกรมสำเร็จรูปทางสถิติที่พัฒนาขึ้นมาใหม่ มาเปรียบเทียบกับค่าพยากรณ์ที่คำนวณได้จากโปรแกรม SPSS ซึ่งจะพิจารณาค่าความคลาดเคลื่อนที่ได้จากโปรแกรมสำเร็จรูปทางสถิติ ที่พัฒนาขึ้นมาใหม่ ว่าน้อยกว่าหรือเท่ากับค่าความคลาดเคลื่อนที่ได้จากโปรแกรม SPSS

3.6 สรุปผลงานวิจัย

พิจารณาจากการตรวจสอบโปรแกรมที่ได้มีการพัฒนาขึ้นมาแล้ว ว่ามีความเหมาะสมถูกต้องในการใช้งานแค่ไหน ให้ค่าในการพยากรณ์และค่าทางสถิติ แสดงกราฟต่าง ๆ พร้อมทั้งให้ค่าตรวจสอบตัวแบบว่าเป็นในลักษณะใด

บทที่ 4

ผลการวิจัยและการวิจารณ์

4.1 การใช้งานในส่วนต่างๆ ของโปรแกรม

การวิเคราะห์ข้อมูลและแสดงผลด้วยโปรแกรมที่จัดสร้างขึ้นนี้ ต้องมีการกำหนดค่าต่างๆ เช่น ตัวแปรอิสระ ตัวแปรตาม ค่านัยสำคัญ และ ตัวพารามิเตอร์ที่จะทำการทดสอบ ฯลฯ ซึ่งสามารถทำได้โดยง่ายจากการเลือกหัวข้อและใส่ข้อมูลที่ต้องการ ก็จะได้ผลลัพธ์ออกมา โดยไม่ต้องใช้เวลาในการศึกษามากนัก วิธีการและตัวเลือกต่างๆ จะได้แสดงดังต่อไปนี้

ส่วนประกอบของโปรแกรม ประกอบไปด้วย 5 ส่วน ดังนี้



รูปที่ 4.1 ส่วนต่างๆ ของโปรแกรม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ส่วนที่ 1 เป็นส่วนของ เมนูต่าง ๆ ของโปรแกรม แบ่งเป็น 3 เมนูหลักได้แก่

1.1 File เมนู เป็นเมนูที่ประกอบด้วยคำสั่งจัดการ File

New สร้าง File ใหม่

Open เปิด File เก่าที่มีอยู่

Save บันทึก File

Save as บันทึก File ในชื่อใหม่

Import เปิด File ใน Format อื่นขึ้นมาทำงานในโปรแกรม

Export Save File ของโปรแกรม ให้เป็น File Format อื่น ๆ

Exit ออกจากโปรแกรม

1.2 Simple linear regression เป็นเมนูส่วนที่เก็บตัวสถิติ

1.2.1 Simple linear regression model ทดสอบสมการถดถอยเชิงเส้น

1.2.2 Tests for Assumptions

ทดสอบ Assumption ของ Error ซึ่งมีการทดสอบ 3 แบบ

ก. *Normal* ทดสอบความเป็นปกติของ Error มี 2 วิธีด้วยกันคือ

1. Kolmogorov Smirnov Test ใช้วิธีของโคลโมโกรอฟ

สมอร์นอฟ

2. Skewness-Kurtosis ใช้วิธีความเบ้-ความโด่ง

ข. *Mean* ทดสอบค่าเฉลี่ยของ Error

ค. *Homoscedasticity* ทดสอบความแปรปรวนของ Error ซึ่งมี 2 วิธีด้วยกันคือ

1. Scatter Plot Diagram แผนภาพการกระจายของ Error

2. Goldfeld-Quandt Test วิธีการใช้การทดสอบของ Goldfeld-Quandt

1.2.3 Outlier หาข้อมูลที่ผิดปกติ โดยมีวิธีในการหา 4 แบบด้วยกันคือ

- Scatter Plot Diagram แผนภาพการกระจายของข้อมูล

- Standardized Residual วิธี Standardized

- Studentized Residual วิธี Studentized

- Leverage วิธี Leverage

1.2.4 Test for Hypothesis ซึ่งในโปรแกรมนี้มี 2 การทดสอบด้วยกันคือ

- Tests for Regression Coefficients

- Correlation Coefficient Test

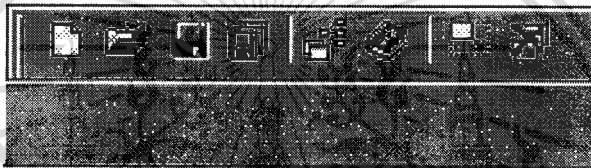
1.2.5 Lack of Fit Test การทดสอบความเหมาะสมของสมการถดถอย

1.2.6 Durbin - Watson Test การทดสอบความเป็นอิสระของค่า Error

1.3 Option ประกอบไปด้วย 2 ส่วนคือ

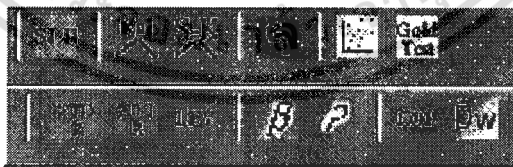
- Setting แก๊ไขตัวเลือกต่าง ๆ
- Help ส่วนช่วยเหลือมีรายละเอียดดังนี้
 - เกี่ยวกับโปรแกรม
 - คณะผู้จัดทำ
 - เนื้อหาวิชาการวิเคราะห์การถดถอย
 - วิธีใช้โปรแกรม

ส่วนที่ 2 เป็นส่วนของ Shortcut ของโปรแกรม



- | | |
|----------------|--------------------------------|
| = ปุ่ม New | = ปุ่ม Option |
| = ปุ่ม Open | = ปุ่ม Help |
| = ปุ่ม Save | = ปุ่มย่อหน้าต่างแสดงผลทั้งหมด |
| = ปุ่ม Save As | = ปุ่มปิดหน้าต่างแสดงผลทั้งหมด |

ส่วนที่ 3 แสดงส่วนของ Shortcut ตัวสถิติต่าง ๆ ในโปรแกรมที่มีอยู่



- | | |
|----------------------------------|-------------------------------------|
| = Simple Linear Regression Model | = Kolmogorov – Smirnov Test |
| = Skewness Kurtosis | = Error Mean Test |
| = Scatter Plot Diagram | = Goldfeld Quandt Test |
| = Standardized Residual | = Studentized Residual |
| = Leverage | = Tests for Regression Coefficients |
| = Correlation Coefficient Test | = Lack of Fit Test |
| = Durbin – Watson Test | |

ส่วนที่ 4 เป็น หน้าต่างไว้สำหรับ Key ข้อมูล

จากรูปที่ 4 จะเห็นว่า ด้านบนของตารางจะเป็น ชื่อตัวแปรต่าง ๆ ส่วนทางด้านซ้ายของตารางจะเป็น เลขที่แถว

ส่วนที่ 5 เป็นส่วนแสดงสถานะต่าง ๆ ของ โปรแกรม และข้อมูล

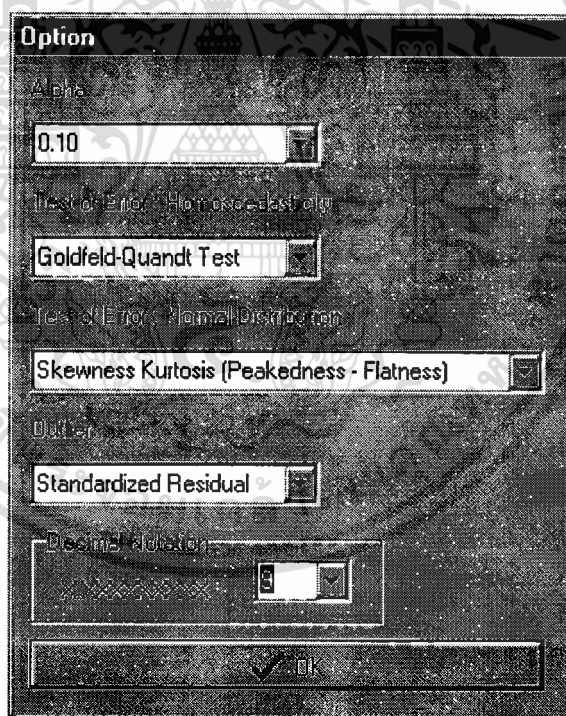
Sum คือ ผลรวมของทุกค่าที่อยู่ใน Column ณ Cell ปัจจุบัน

Bar คือ ค่าเฉลี่ยของค่าทุกค่าที่อยู่ใน Column ณ Cell ปัจจุบัน

N คือ จำนวนของข้อมูลที่อยู่ใน Column ณ Cell ปัจจุบัน

รายละเอียดของหน้าต่างที่มีอยู่ในตัวโปรแกรม มีดังนี้

หน้าต่าง Option



รูปที่ 4.2 Option ต่าง ๆ ที่สามารถ Set ได้ในตัวโปรแกรม

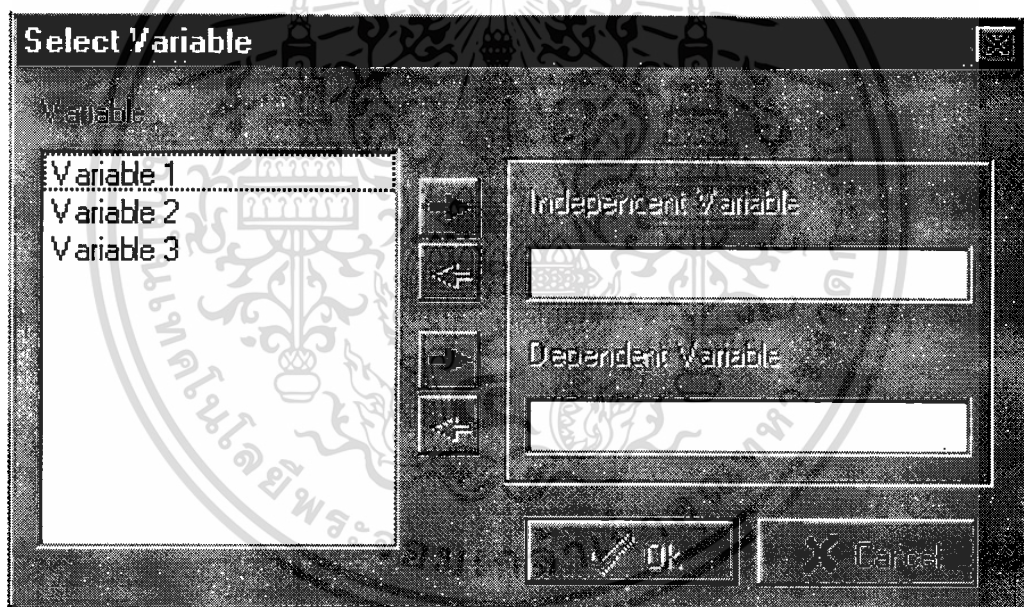
1. Alpha - ในตัวโปรแกรมสามารถ Set ค่า Alpha ที่ต้องการได้ซึ่งตัวสถิติที่อยู่ภายในโปรแกรมทั้งหมดทำงานที่ Alpha ที่ Set ไว้ ซึ่งค่า Alpha ที่สามารถ Set ได้ มีอยู่ 3 ค่าคือ 0.01 , 0.05 และ 0.10

2. *Test of Error : Homoscedasticity* - ตัวสถิตินี้ไว้สำหรับเป็น Default ในการตรวจค่าความแปรปรวนของ Error
3. *Test of Error : Normal Distribution* - ตัวสถิตินี้ไว้สำหรับเป็น Default ในการตรวจการแจกแจงแบบปกติของ Error
4. *Outlier* - ตัวสถิตินี้ไว้สำหรับ เป็น Default ในการตรวจข้อมูลนอกรอบ
5. *Decimal Notation* - จำนวนทศนิยมที่ใช้ในการแสดงผลทั้งหมด

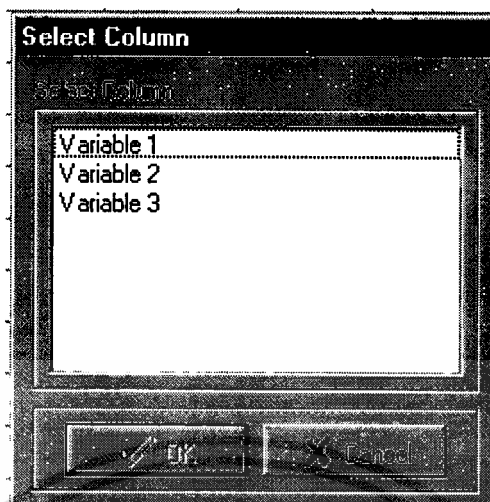
หน้าต่างเลือกตัวแปรเข้าทำการคำนวณ

ในโปรแกรมมีหน้าต่างในการเลือกตัวแปรเข้าไปทำการคำนวณอยู่ 2 แบบคือ

1. การเลือกตัวแปรเข้าทำการคำนวณ 2 ตัวแปร (ดังรูปที่ 4.3)
2. การเลือกตัวแปรเข้าทำการคำนวณ 1 ตัวแปร (ดังรูปที่ 4.4)



รูปที่ 4.3 หน้าต่างในการเลือกตัวแปรเข้าไปทำการคำนวณ 2 ตัวแปร



รูปที่ 4.4 หน้าต่างในการเลือกตัวแปรเข้าทำการคำนวณ 1 ตัวแปร

หน้าต่างทั้ง 2 แบบมีไว้เพื่อเลือกตัวแปรเข้าไปทำการคำนวณ โดยจะทำการคำนวณตามตัวสถิติที่ตั้งไว้ตอนต้น โดยมีขั้นตอนการใช้งานดังนี้

1. เลือกตัวแปรที่ต้องการทำการคำนวณ
2. กด OK เพื่อเป็นการยืนยันและทำการคำนวณต่อไป
3. กด Cancel เพื่อเป็นการยกเลิกการเลือกตัวแปร

หน้าต่างแก้ไขในหน้าต่างป้อนข้อมูล

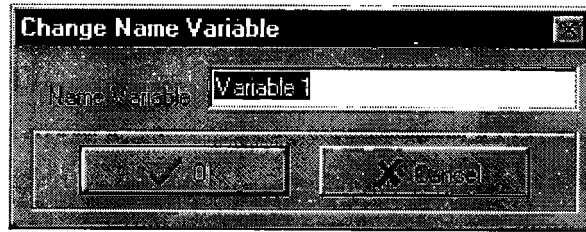
หน้าต่างนี้จะปรากฏเมื่อ Click ขวาที่หน้าต่างป้อนข้อมูล จะปรากฏหน้าต่างที่มีหน้าที่เพื่อให้ความสะดวกแก่ผู้ใช้ในการแก้ไขข้อมูลที่อยู่ในตาราง ดังรูปที่ 4.5

Variable 1	Variable 2	Variable 3	Variable 4	Variable 5	Variable 6
1	48				
2	44				
3	46				
4	49				
5	50				

รูปที่ 4.5 หน้าต่างแก้ไขในหน้าต่างป้อนข้อมูล

1. Change Name ไว้สำหรับแก้ไขชื่อตัวแปรของข้อมูลใน Column นั้นๆ ในโปรแกรม
2. Transform Data ไว้สำหรับ Transform ข้อมูลในกรณีข้อมูลไม่เป็นไปตาม Assumption
3. Copy ไว้สำหรับ Copy ข้อมูลใน Column ที่ต้องการ
4. Cut ไว้สำหรับตัดข้อมูลใน Column ที่ต้องการเก็บไว้
5. Paste ไว้สำหรับวางข้อมูลที่ทำการ Copy หรือ การ Cut ไว้ใน Column โดยสามารถเลือกได้ว่าจะวางทับหรือว่าจะวางต่อท้าย
6. Delete Row ไว้สำหรับลบ Row ที่ไม่ต้องการทิ้ง
7. Insert Row ไว้สำหรับแทรก Row ในตำแหน่งที่ต้องการ

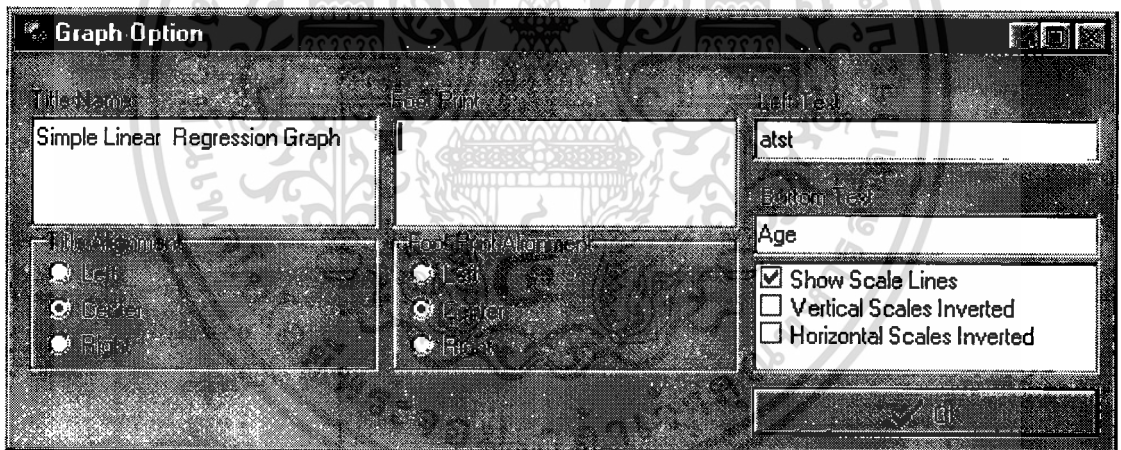
หน้าต่างเปลี่ยนชื่อตัวแปร



รูปที่ 4.6 หน้าต่างในการเปลี่ยนชื่อตัวแปร

จะแสดงขึ้นเมื่อมีการ Click ขวาที่หน้าต่างป้อนข้อมูลแล้วเลือก Change name หรือ เมื่อโปรแกรมต้องการให้ผู้ใช้ตั้งชื่อของ Column ที่จะเกิดขึ้นใหม่

หน้าต่าง Set Option ในผลลัพธ์ที่เป็น Graph



รูปที่ 4.7 หน้าต่างในการปรับแต่ง Graph

Title Name = ใส่หัวข้อบนตาราง

Foot Print = ใส่หัวข้อท้ายตาราง

Left Text = ใส่ชื่อตัวแปรด้านซ้าย

Bottom Text = ใส่ชื่อตัวแปรด้านขวา

Foot Alignment = เลือกการจัดวาง Foot Print

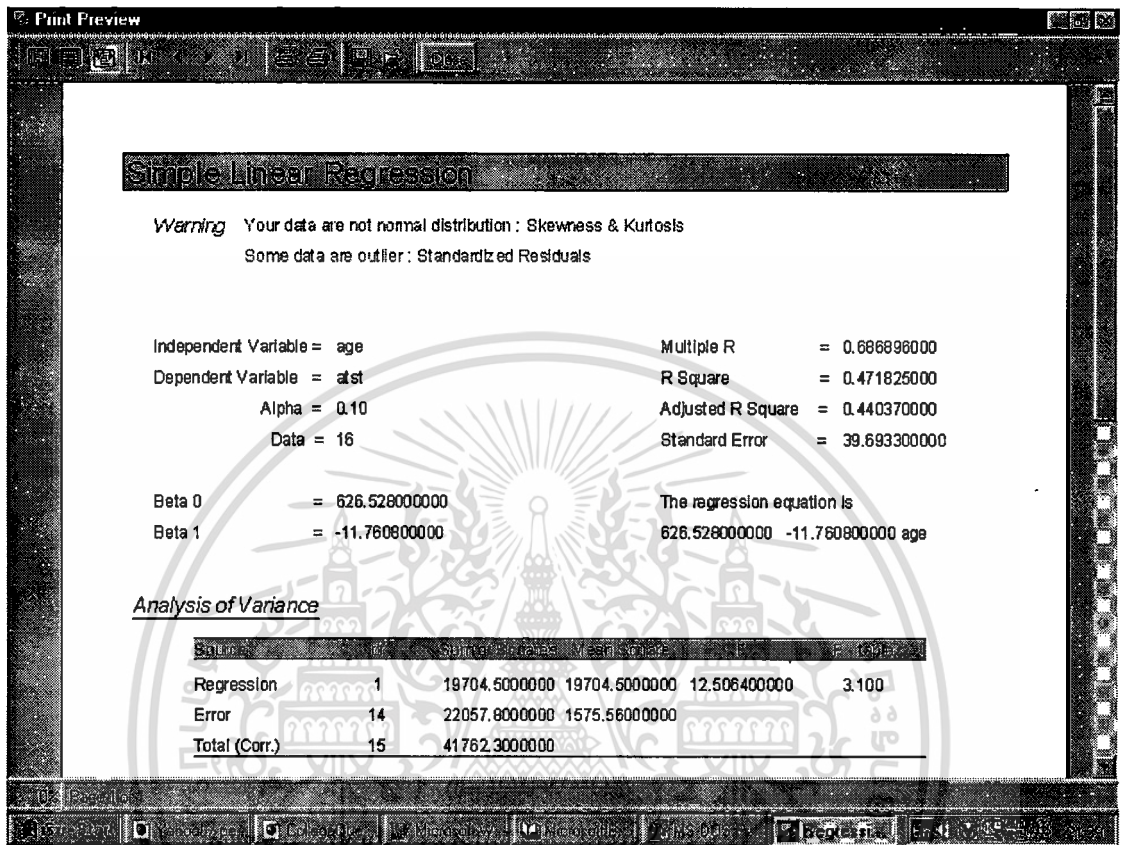
Title Alignment = เลือกการจัดวาง Title Name

Show Scale Lines = แสดง Scale ใน Graph

Vertical Scales Inverted = ทำการ Inverted Scales ในแนวตั้ง

Horizontal Scales Inverted = ทำการ Inverted Scales ในแนวนอน

หน้าต่างผลลัพธ์



รูปที่ 4.8 หน้าต่างแสดงผลลัพธ์ในลักษณะเป็นข้อมูลที่ผ่านการคำนวณ

ปุ่มในการ แสดงรายงานในจอภาพ แบ่งออกเป็น 3 ลักษณะด้วยกันคือ

- ปุ่มแสดงรายงานทั้งหน้า
- ปุ่มแสดงรายงานพอดีหน้า
- ปุ่มแสดงรายงานเต็มหน้าตามลำดับ

ปุ่มเปลี่ยนหน้ารายงาน (ในกรณีที่มีหน้ารายงานเกิด 1 หน้า) แบ่งออกเป็น 4 ปุ่ม

- ปุ่มไปหน้าเริ่มต้น
- ปุ่มไปหน้าก่อนหน้าปัจจุบัน
- ปุ่มไปหน้าถัดจากหน้าปัจจุบัน
- ปุ่มไปหน้าสุดท้าย

ปุ่มเกี่ยวกับการพิมพ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- ปุ่มติดตั้งและแก้ไขการพิมพ์
- ปุ่มสั่งให้พิมพ์

- ปุ่มในการเปิด File และ Save File
 - ปุ่ม Save File
 - ปุ่ม เปิด File

- ปุ่มปิดหน้าต่างแสดงผล

หน้าต่างผลลัพธ์ในลักษณะ Graph

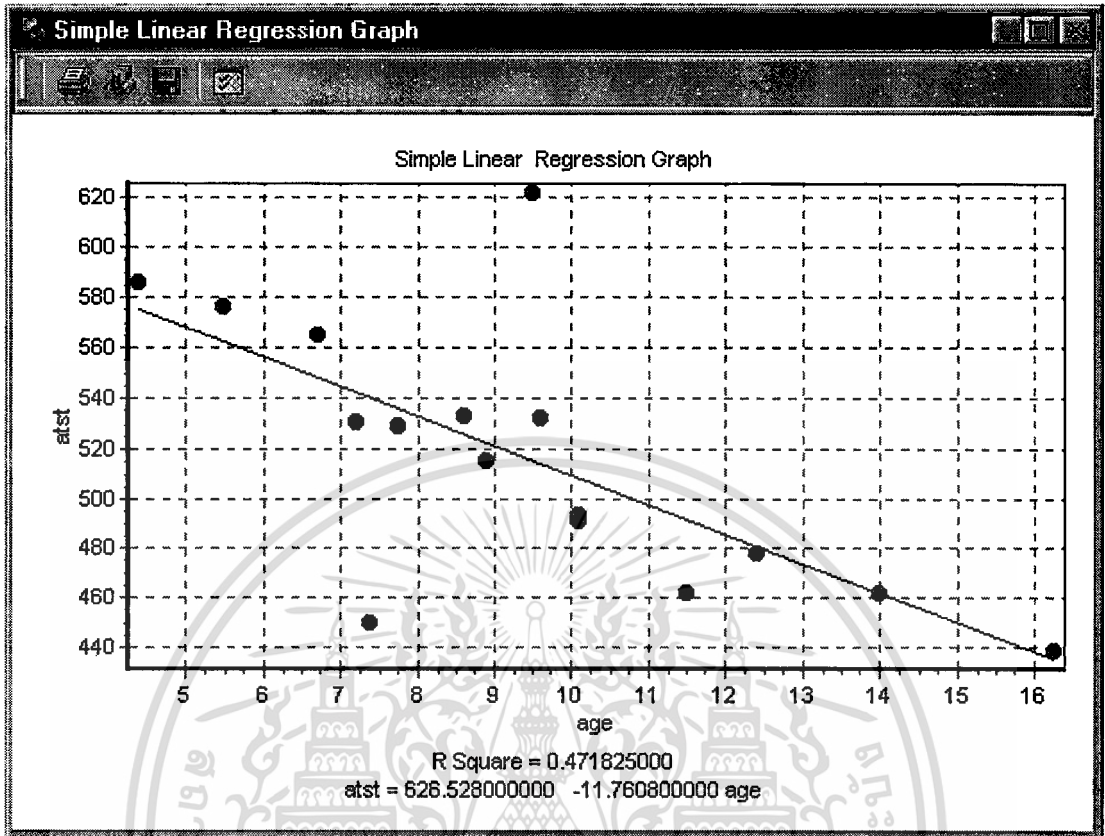
แบ่งออกได้เป็น 2 ลักษณะคือ ในหน้าต่างผลลัพธ์ Graph นั้นผู้ใช้สามารถที่จะเลือก Zoom In หรือ Zoom Out ข้อมูลในตำแหน่งที่ต้องการได้ (เมื่อผู้ใช้ต้องการพิจารณา หรือ พิมพ์ข้อมูลบางตัว) โดยมีวิธีการ Zoom ดังนี้

Zoom In

- คลิก Mouse ด้านขวา (ค้างไว้) เป็นจุดเริ่มต้นในการ Zoom
- ลาก Mouse จากซ้ายบนมาด้านขวาล่างให้ครอบคลุมตัวข้อมูลที่ต้องการ
- ปล่อย Mouse ด้านขวา โปรแกรมจะทำการ Zoom ส่วนที่ต้องการ

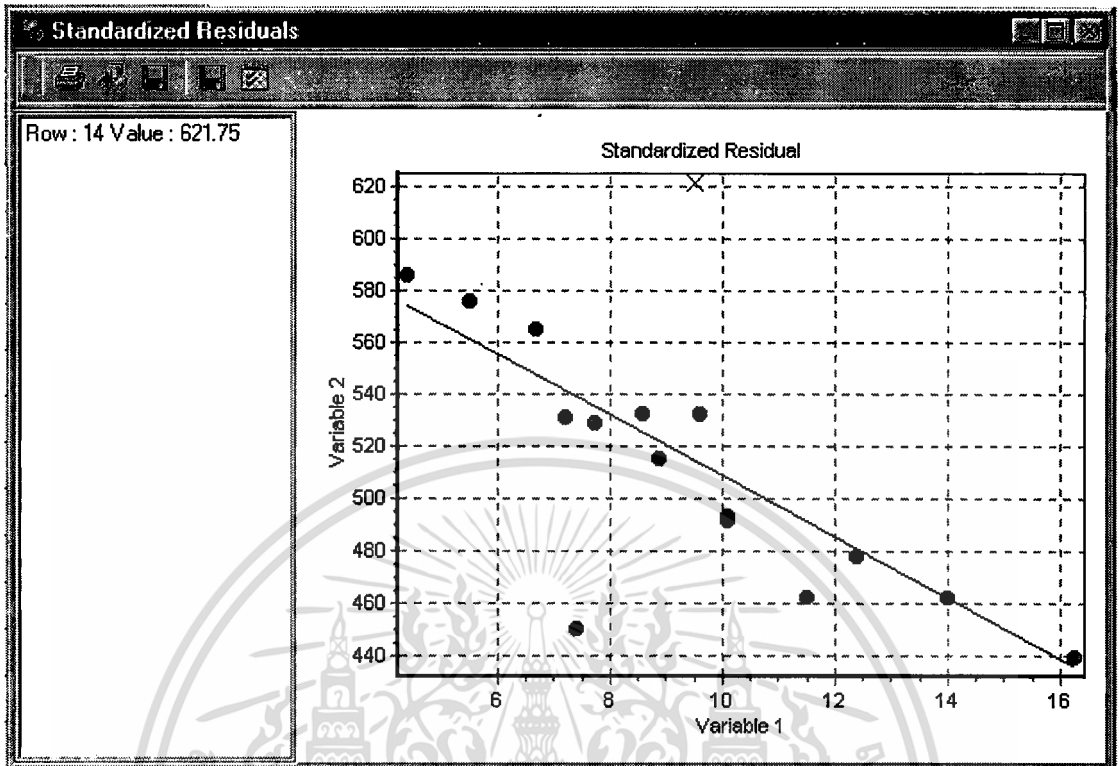
Zoom Out

- คลิก Mouse ด้านขวา (ค้างไว้)
- ลาก Mouse มาทางด้านซ้ายบน แล้วทำการปล่อยโปรแกรมจะทำการ Zoom out ออกมาอยู่ในลักษณะก่อน Zoom



รูปที่ 4.9 หน้าต่างที่แสดงผลในลักษณะ Graph

-  = ปุ่มพิมพ์ Graph ออกทางเครื่องพิมพ์
-  = ปุ่ม Copy Graph ลงไปสู่ Clipboard เมื่อต้องการนำภาพ ไปใช้ในโปรแกรมอื่น
-  = ปุ่ม Save Graph ในรูปแบบ Bitmap File (BMP)
-  = ปุ่ม ตั้ง/แก้ไข Graph



รูปที่ 4.10 หน้าต่างแสดงผลในลักษณะ Graph ที่แสดงข้อมูล outlier
(X เป็นข้อมูล Outlier)

-  = ปุ่มพิมพ์ Graph ออกทางเครื่องพิมพ์
-  = ปุ่ม Copy Graph ลงไปสู่ Clipboard เมื่อต้องการนำภาพไปใช้ในโปรแกรมอื่นๆ
-  = ปุ่ม Save Graph ในรูปแบบ Bitmap File (BMP)
-  = ปุ่ม Save ข้อมูล Outlier ไปเป็น File
-  = ปุ่มตั้ง/แก้ไข Graph

หน้าต่างที่อยู่ทางซ้ายของ Graph เป็นหน้าต่างแสดงข้อมูลที่เกิด Outlier ซึ่งจะบอกถึงค่าที่เกิด Outlier และตำแหน่งของค่านั้นว่าเป็นข้อมูลตัวที่เท่าไร ซึ่งผู้ใช้สามารถที่จะดูข้อมูลที่เกิด Outlier ได้โดยการ Click ไปที่ข้อมูลที่ปรากฏขึ้นในหน้าต่างทางซ้ายของ Graph

การใช้งานโปรแกรม

การใช้งานโปรแกรมสามารถทำได้ดังนี้

1. การป้อนข้อมูล

ทำได้โดยการเลือก Column ของข้อมูลที่ต้องการจะป้อน แล้วทำการป้อนข้อมูล (โปรแกรมจะไม่อนุญาตให้ป้อนค่าอื่น ๆ นอกจากค่าที่เป็นตัวเลข) เมื่อป้อนข้อมูลเสร็จแล้ว ให้ทำการกด Enter ตำแหน่งของการป้อนก็จะเลื่อนลงมาอัตโนมัติ (โปรแกรมจะไม่อนุญาตให้ทำการป้อนข้อมูลโดยมีการเว้นช่องว่างโดยไม่จำเป็น) ในการนำข้อมูลไปทำการคำนวณทุกครั้งโปรแกรมจะมีการตรวจสอบข้อมูลว่าข้อมูลสามารถนำไปทำการคำนวณได้หรือไม่ ถ้าเกิดข้อผิดพลาดขึ้นโปรแกรมก็จะแจ้งเตือนต่อผู้ใช้ และให้ผู้ใช้เลือกว่าจะทำการแก้ไขข้อมูลที่ผิดพลาดหรือไม่

2. การนำข้อมูลมาคำนวณ

ในการนำข้อมูลมาทำการคำนวณนั้น สามารถทำได้ดังนี้

เลือกตัวสถิติ หรือการคำนวณที่ต้องการ โดยเลือกได้จาก 2 แห่งคือ จาก Menu หรือ จาก Shortcut ที่เตรียมไว้ให้ (เมื่อต้องการทราบว่า Shortcut ใดไว้สำหรับทำอะไรให้เลื่อน Mouse ไปในตำแหน่งของ Shortcut ที่ต้องการทราบ รอสักพักจะมีข้อความ Help ขึ้นมา) โปรแกรมจะให้ผู้ใช้เลือกตัวแปรที่ต้องการจะนำมาคำนวณ ให้ผู้ใช้ใส่ตัวแปรที่ต้องการคำนวณเข้าไป และกดปุ่ม OK แล้วโปรแกรมจะทำการแสดงรายงานต่าง ๆ พร้อมทั้งค่าเตือน , ข้อผิดพลาด และวิธีแก้ไขแก้ไข

3. การบันทึกข้อมูล

เมื่อผู้ใช้ต้องการบันทึกข้อมูลเพื่อที่จะนำไปใช้ต่อไปในภายหลังสามารถทำการบันทึกข้อมูลได้โดยการเลือกคำสั่ง Save ข้อมูลที่มีอยู่ใน Menu File หรือ Shortcut ที่เตรียมไว้ให้ (ถ้าข้อมูลที่ต้องการบันทึกนั้น ไม่เคยทำการบันทึกมาก่อน โปรแกรมจะให้ผู้ใช้ทำการใส่ชื่อของ File ที่ต้องการจะบันทึกไว้)

ในบางกรณีปุ่ม Save ข้อมูลจะไม่ทำงานอันเนื่องมาจากยังไม่ได้มีการแก้ไขข้อมูลที่เปิดขึ้นมาใช้งาน โดยสามารถดูได้จากหน้าต่างที่ป้อนข้อมูล เมื่อมีการแก้ไขข้อมูล ด้านล่างของหน้าต่างที่ป้อนข้อมูลจะปรากฏคำว่า Modify ขึ้น

4. การเปิดข้อมูลที่มีอยู่แล้ว

ทำการเลือก File ข้อมูลที่ต้องการจะเปิดได้โดยเลือกคำสั่ง Load ข้อมูลที่มีอยู่ใน Menu File หรือ Shortcut ที่เตรียมไว้ให้โดยทางโปรแกรม จะให้ผู้ใช้เลือกว่าต้องการเปิด File ข้อมูลไหน

5. การ Save as ข้อมูล

การ Save as ข้อมูลเป็นการบันทึกข้อมูลเดิมที่มีอยู่ในชื่อใหม่ สามารถทำได้โดยเลือกคำสั่ง Save as ข้อมูลที่มีอยู่หรือ Shortcut ที่เตรียมไว้ให้

4.2 ความถูกต้องของค่าที่ได้จากการคำนวณ

โปรแกรมที่ได้สร้างขึ้นมีความแม่นยำ ถูกต้องเมื่อเปรียบเทียบค่าที่คำนวณได้กับ โปรแกรมมาตรฐานอื่น เช่น SPSS แม้จะเป็นทศนิยมในพจน์ท้าย ๆ และโปรแกรมที่สร้างขึ้นนี้ สามารถแสดงได้ละเอียดถึง 9 หลัก แม้ในส่วนของกราฟก็มีความชัดเจน ซึ่งถึงข้อมูลนอกกรอบทันที

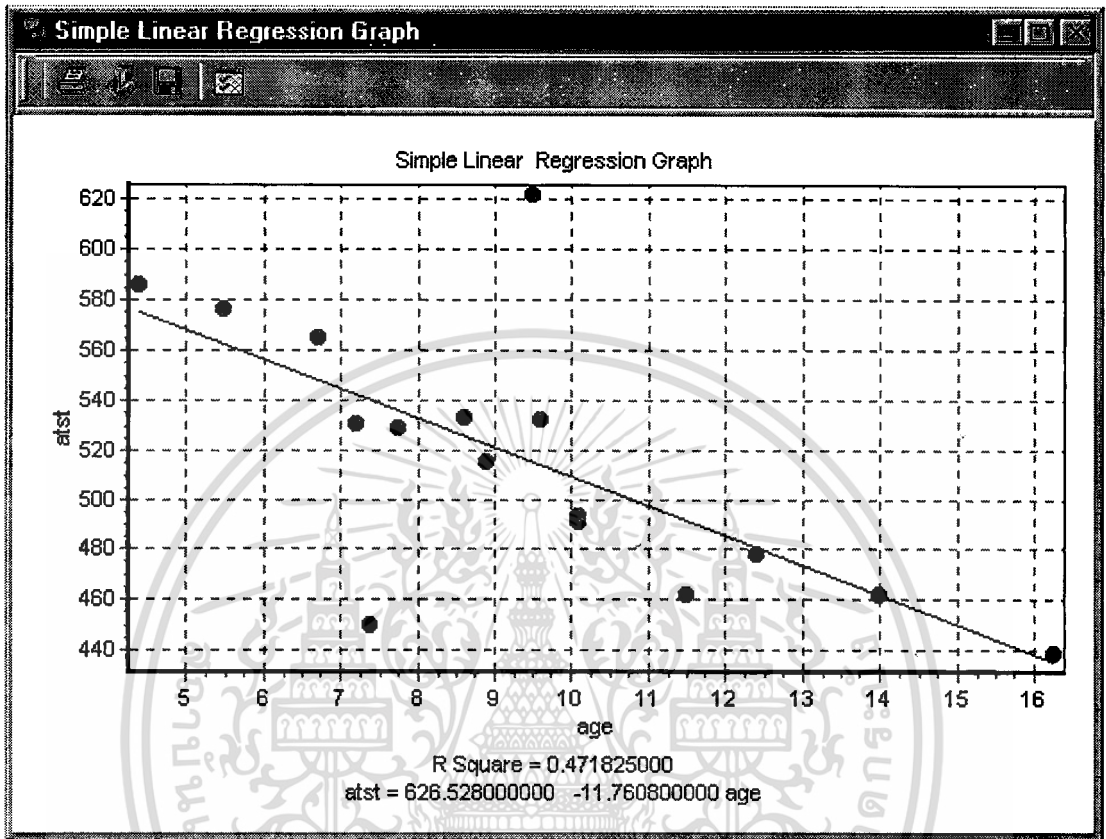
4.3 ส่วนการแสดงผล (Output)

จากการที่ได้ใช้โปรแกรมสำเร็จรูปทางสถิติที่จัดทำขึ้นนี้ มาตรวจสอบพบว่าผลลัพธ์ของการคำนวณที่ได้มีความละเอียดถูกต้อง และมีลักษณะของหน้าจอ Output สำหรับการแสดงผลที่ดูง่าย เสนอและสรุปผลได้อย่างเป็นขั้นตอนตรงตามหลักการ ทางคณะผู้จัดทำได้นำข้อมูลตัวอย่าง ซึ่งเป็นการศึกษาความสัมพันธ์ระหว่างอายุ และค่าเฉลี่ยชั่วโมงการนอนของเด็ก 16 คน ปรากฏผลดังตารางข้างล่างนี้

อายุ (ปี)	ค่าเฉลี่ยชั่วโมงการนอน (นาที)
4.40	586.00
14.00	461.75
10.10	491.10
6.70	565.00
11.50	462.00
9.60	532.10
12.40	477.60
8.90	515.20
10.10	493.00
7.75	528.30
5.50	575.90
8.60	532.50
7.20	530.50
9.50	621.75
16.25	438.80
7.40	450.00

ให้ อายุ = AGE และค่าเฉลี่ยชั่วโมงการนอน = ATST จะได้ output ของโปรแกรม
ที่ได้สร้างขึ้นดังนี้

Output ของโปรแกรมที่ได้สร้างขึ้น



Simple Linear Regression

Warning Your data are not normal distribution : Skewness & Kurtosis
 Some data are outlier : Standardized Residuals

Independent Variable = age
 Dependent Variable = atst
 Alpha = 0.10
 Data = 16

Multiple R = 0.686896000
 R Square = 0.471825000
 Adjusted R Square = 0.440370000
 Standard Error = 39.693300000

Beta 0 = 626.528000000
 Beta 1 = -11.760800000

The regression equation is
 626.528000000 - 11.760800000 age

Analysis of Variance

Source	df	Sum of Squares	Mean Square	F	Sig.
Regression	1	19704.5000000	19704.5000000	12.506400000	3.100
Error	14	22057.8000000	1575.560000000		
Total (Corr.)	15	41762.3000000			

Summary There is relationship between atst and age

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Test for Hypothesis

Parameters	Parameter Estimate	Standard Error	t-Statistic	p-Value	90% Confidence Interval	95% Confidence Interval
Beta 0	626.528000000	32.698900000	19.160500000	1.761	593.829000000	659.227000000
Beta 1	-11.760800000	3.325610000	-3.536430000		-15.086400000	-8.435190000

Summary *Beta0 is not equal to 0*

Beta1 is not equal to 0

Correlation Coefficient Test

0.471825000 0.235642000 2.002300000 1.761

Summary *The Correlation is significant*

รูปที่ 4.11 กราฟที่แสดงมาพร้อมกันกับผลลัพธ์ด้านบนเมื่อเลือกเมนู Simple Linear Regression Model

สามารถดูผลสรุปตามสมมติฐานได้ทันทีจากตัวหนังสือสีฟ้า และมีการเตือนในกรณีที่พบว่าข้อมูลมีลักษณะที่ผิดปกติด้วยตัวหนังสือสีแดง

Tests for Regression Coefficients

Independent Variable =	age	Beta 0	=	626.528000000
Dependent Variable =	atst	Beta 1	=	-11.760800000
Alpha =	0.10	Standard Error of Beta 0	=	32.698900000
Data =	16	Standard Error of Beta 1	=	3.325610000

Hypothesis for Beta 0

$$H_0: \beta_0 = 0.000000000$$

$$H_1: \beta_0 \neq 0.000000000$$

$$T = 19.160500000$$

$$T\text{-table} = 1.761000000$$

Summary Reject H_0 : ($T < -T\text{-table}$) or ($T > T\text{-table}$)

Hypothesis for Beta 1

$$H_0: \beta_1 = 0.000000000$$

$$H_1: \beta_1 \neq 0.000000000$$

$$T = -3.536430000$$

$$T\text{-table} = 1.761000000$$

Summary Reject H_0 : ($T < -T\text{-table}$) or ($T > T\text{-table}$)

รูปที่ 4.12 ผลลัพธ์เมื่อเลือกเมนู Parameter Test

-เพื่อต้องการทดสอบว่าค่าพารามิเตอร์มีค่าตามสมมติฐานที่ตั้งไว้หรือไม่

Correlation Coefficient Test

Independent Variable = age
 Dependent Variable = atst
 Alpha = 0.10

Data = 16
 R Square = 0.471825000
 Standard Error of R Square = 0.235642000

$$H_0: \rho = 0.000000000$$

$$H_1: \rho \neq 0.000000000$$

$$T = 2.914990000$$

$$T\text{-table} = 1.345$$

Summary *Reject H0 : (T < -T-table) Or (T > T-table)*

รูปที่ 4.13 ผลลัพธ์เมื่อเลือกเมนู Correlation Coefficient Test

เพื่อทดสอบว่าค่า Correlation Coefficient มีค่าตามที่ตั้งสมมติฐานไว้หรือไม่
 และเป็นการทดสอบความสัมพันธ์ของตัวแปรอิสระและตัวแปรตามกรณีที่ตั้งสมมติฐานว่า
 $\rho = 0$

Test for Normal Distribution Kolmogorov Test

Variable = age
 Alpha = 0.10

$$H_0: F(x) = \text{Normal}$$

$$H_1: F(x) \neq \text{Normal}$$

$$T = 2.937500000$$

$$T\text{-table} = 0.404000000$$

Summary *Reject H0 : T > T-table : Data are not normal distribution*

รูปที่ 4.14 ผลลัพธ์เมื่อเลือกเมนู Kolmogorov Test

Skewness, Kurtosis

Independent Variable = age

Dependent Variable = atst

Skewness statistical value = -0.001014680

Kurtosis statistical value = 0.000307591

*Suggestion : If Skewness statistical value is zero , data are normal distribution ,
If Kurtosis statistical value is three , data are normal distribution.*

รูปที่ 4.15 ผลลัพธ์เมื่อเลือกเมนู Skewness and Kurtosis

Error Mean Test

Independent Variable = age

Dependent Variable = atst

Alpha = 0.10

Hypothesis for Error Mean

$$H_0 : \mu_x = 0$$

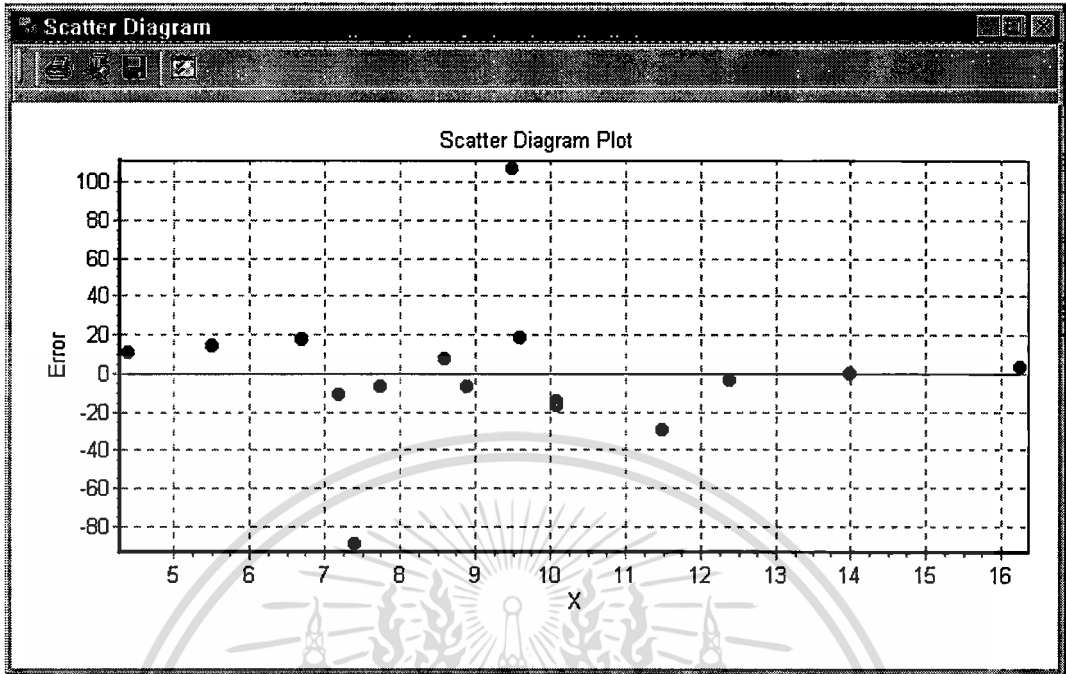
$$H_1 : \mu_x \neq 0$$

$$T = 0.000000000$$

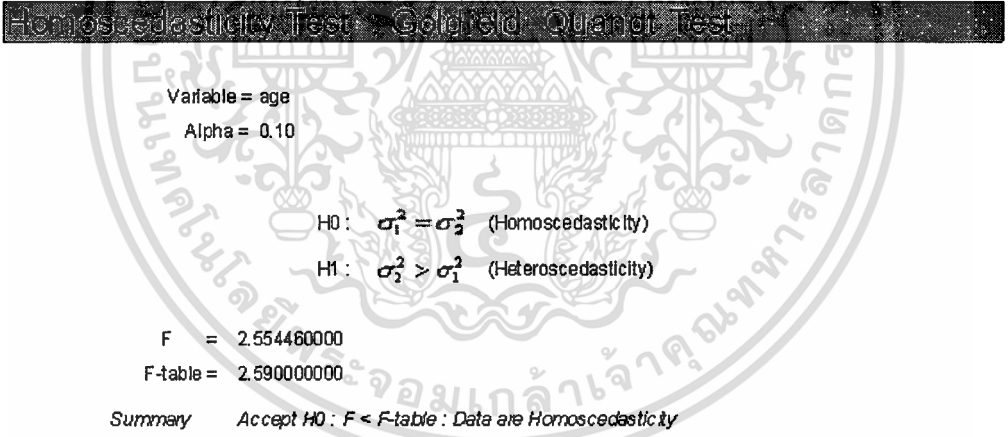
$$T\text{-table} = 1.753$$

Summary Accept H_0 : Mean = 0

รูปที่ 4.16 ผลลัพธ์เมื่อเลือกเมนู Error Test



รูปที่ 4.17 ผลลัพธ์ของแผนภาพการกระจายระหว่างค่า Error และ ค่าตัวแปรอิสระ X



รูปที่ 4.18 ผลลัพธ์เมื่อเลือกเมนู Goldfeld Quandt Test เพื่อการพิจารณาและแสดงว่าค่า Error มีความแปรปรวนคงที่ (DORAN HOWARD E. , 1989)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Lack of Fit Test

Independent Variable = age
 Dependent Variable = atst
 Alpha = 0.10

Analysis of Variance

Source	df	Sum of Squares	Mean Square	F	Table
Total (corr.)	15	41762.3000000			
REG.	1	19704.5000000	19704.5000000	12.506300000	3.100
Error	14	22057.8000000	1575.56000000		
Lack of fit	13	22056.0000000	1696.62000000	939.956000000	60.900
Pure Error	1	1.805000000	1.805000000		

Lack of Fit Test

$$H_0: \mu = \beta_0 + \beta_1 X_1$$

$$H_1: \mu \neq \beta_0 + \beta_1 X_1$$

$$F = 939.956000000$$

$$F\text{-Table} = 60.900$$

Summary Reject H_0 : $F > F\text{-Table}$. Regression equation is not linear
 Data must be transformed before using regression analysis

Test for regression relationship

$$H_0: \beta_1 = 0$$

$$H_1: \beta_1 \neq 0$$

$$F = 12.506300000$$

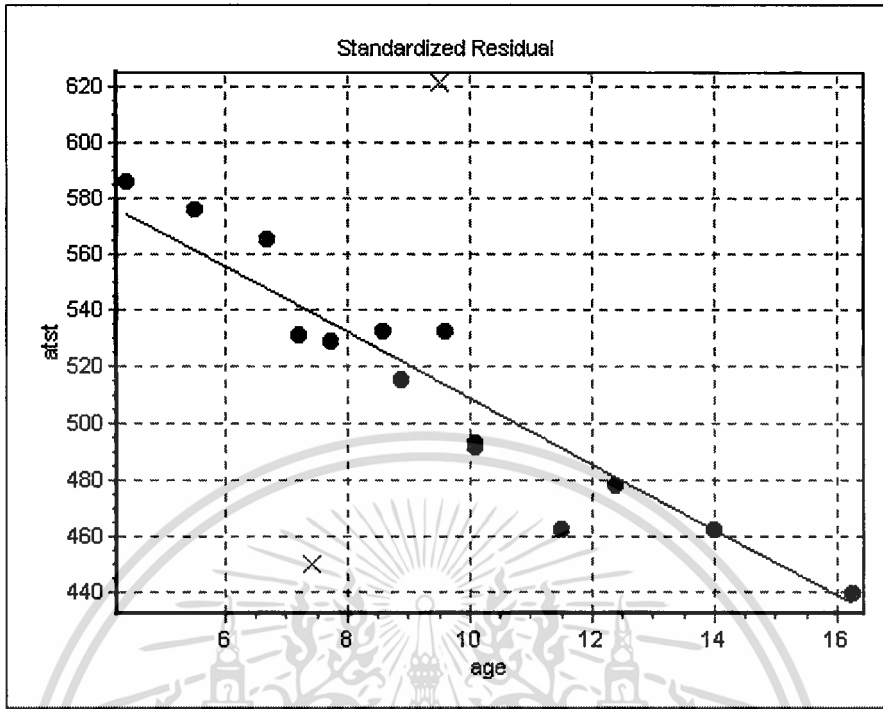
$$F\text{-Table} = 3.100$$

Summary Reject H_0 : $F > F\text{-Table}$. There is relationship between atst and age

รูปที่ 4.19 ผลลัพธ์เมื่อเลือกเมนู Lack of Fit Test

เพื่อทำการวิเคราะห์กรณีตัวแปรอิสระ 1 ค่า มีค่าสังเกตหลายค่า และมีการเสนอแนะให้การแปลงข้อมูล (ตัวหนังสือสีแดง) กรณีที่พบว่าสมการมีลักษณะที่ไม่เป็นเส้นตรง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



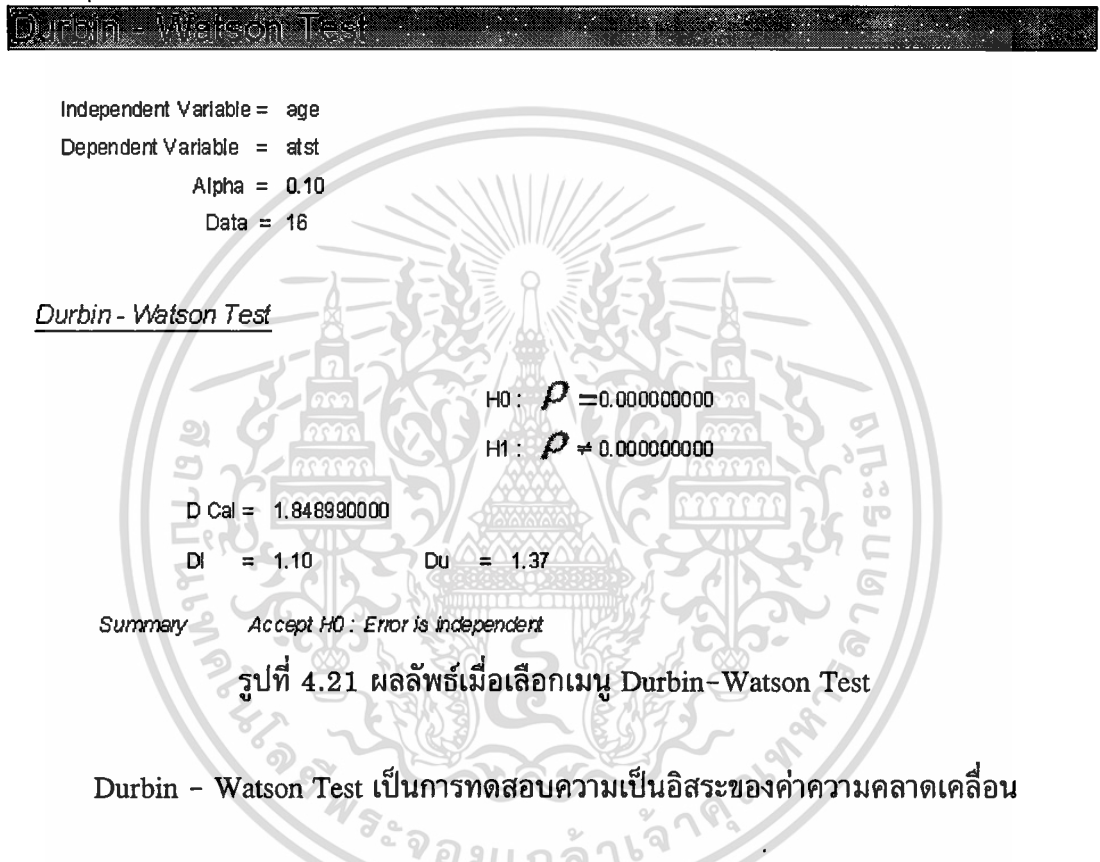
The screenshot shows a software window titled "Simple linear regression - Output". The main area contains a data table with columns for "age", "atst", and several "Variable" columns. The "Standardiz..." dialog box is open on the right, showing "Row: 14 Value: 621.75" and "Row: 16 Value: 450". The taskbar at the bottom shows several open applications including "Microsoft Word", "Microsoft Excel", "Adobe Photoshop", and "Regression".

	age	atst	Variable 3	Variable 4	Variable 5	Variable 6	Variable 7	Variable 8	Variable 9
1	4.4	586	0.282660000						
2	14	461.75	-0.003186510						
3	10.1	491.1	-0.419307000						
4	6.7	565	0.435076000						
5	11.5	462	-0.737619000						
6	9.6	532.1	0.465468000						
7	12.4	477.6	-0.077942300						
8	8.9	515.2	-0.167702000						
9	10.1	493	-0.371440000						
10	7.75	528.3	-0.178407000						
11	5.5	575.9	0.354131000						
12	8.6	532.5	0.179253000						
13	7.2	530.5	-0.285943000						
14	9.5	621.75	2.694410000						

รูปที่ 4.20 กราฟและผลลัพธ์เมื่อเลือกเมนู Outliers

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 4.20 จะมีการแสดงข้อมูลนอกรอบเป็นกากบาทสีแดงให้ทราบได้ทันที และเมื่อทำการ Click เลือกที่ข้อมูลนั้น โปรแกรมก็จะไปที่ตัวข้อมูลใน Data Sheet ให้แก้ไขได้อย่างสะดวก หรือจากรูปที่สองเป็นตัวอย่างการแสดงค่า Standardized Residual ไว้ใน Column ที่ 3 และแสดงข้อมูลนอกรอบใน Frame ข้าง ๆ อย่างชัดเจนว่าเป็นค่าใด



รูปที่ 4.21 ผลลัพธ์เมื่อเลือกเมนู Durbin-Watson Test

Durbin - Watson Test เป็นการทดสอบความเป็นอิสระของค่าความคลาดเคลื่อน

โปรแกรมที่ได้จัดทำขึ้นนี้จะบอกอย่างชัดเจนทั้งในรูปกราฟ และตารางข้อมูล และยังตามแก้ข้อมูลนั้นได้ง่ายอีกด้วย ส่วนการสรุปสมมติฐานตามนัยสำคัญก็ดูผลและตัดสินใจจากผลที่ได้จากการคำนวณ จะพบว่าโปรแกรมที่ได้จัดทำขึ้นจะอำนวยความสะดวกให้กับผู้ใช้ สรุปออกมาให้เห็นในรูปที่ดูง่าย สังเกตได้ชัดเจน เช่น มีการขึ้นเตือนทันทีสำหรับข้อมูลที่นอกรอบ หรือมีลักษณะที่ผิดปกติ และสรุปผลบอกได้เลยว่าจะยอมรับหรือปฏิเสธสมมติฐานนั้น ๆ

บทที่ 5

บทสรุป

5.1 บทสรุป

ในการสร้างโปรแกรม Statistical Simple Linear Regression Analysis Program ได้สร้างตรงตามวัตถุประสงค์คือ เพื่อสร้างโปรแกรมสำเร็จรูปทางสถิติที่คำนวณเกี่ยวกับการวิเคราะห์การถดถอยเชิงเส้น และ เพื่อนำโปรแกรมสำเร็จรูปที่สร้างขึ้นมาวิเคราะห์ข้อมูลในเชิงปริมาณ ซึ่งโปรแกรมที่สร้างขึ้นนี้สามารถวิเคราะห์หาค่าพารามิเตอร์ ทดสอบสมมติฐานที่เกี่ยวกับพารามิเตอร์ การทดสอบความเหมาะสมของสมการถดถอย ทดสอบ Assumption ของค่าความคลาดเคลื่อนต่าง ๆ และการหาข้อมูลนอกกรอบโดยใช้วิธีต่าง ๆ ผู้จัดทำได้สร้างโปรแกรมนี้ขึ้นซึ่งมีจุดเด่นในส่วนแสดงผลและดูง่าย สรุปผลอย่างเป็นขั้นเป็นตอนตรงตามเนื้อหา สะดวกในการนำผลลัพธ์ที่ได้มาประยุกต์ใช้งานกับโปรแกรมอื่น ๆ สามารถนำผลลัพธ์ที่ได้มาวิเคราะห์ข้อมูลเชิงปริมาณได้อย่างมีประสิทธิภาพ

5.2 ข้อเสนอแนะในการศึกษาครั้งต่อไป

จากการสร้างโปรแกรมในครั้งนี้ ผู้จัดทำได้พบข้อสังเกตบางประการ ที่จะขอเสนอแนะไว้สำหรับ ผู้ที่จะทำการศึกษาหรือพัฒนาต่อไป ดังนี้

1. ถ้าได้มีการเพิ่มเติมในเรื่องของ Multimedia ก็จะทำให้โปรแกรมมีความน่าใช้ และ สะดวกมากยิ่งขึ้น
2. ควรได้มีการพัฒนาตัวสถิติเพิ่มเติม และเพิ่มขอบเขตของโครงงานให้กว้างยิ่งขึ้น ก็จะเป็นประโยชน์ในทางที่ใช้วิเคราะห์ข้อมูลได้หลากหลายมากขึ้น
3. ถ้ามีส่วนของการช่วยสอน (Tutorial) และส่วนของการช่วยเหลือ (Help) ที่มีลักษณะการนำเสนอเป็น Multimedia ก็จะทำให้โปรแกรมมีความสมบูรณ์ ทั้งในด้านการใช้งานและด้านการศึกษา

ภาคผนวก

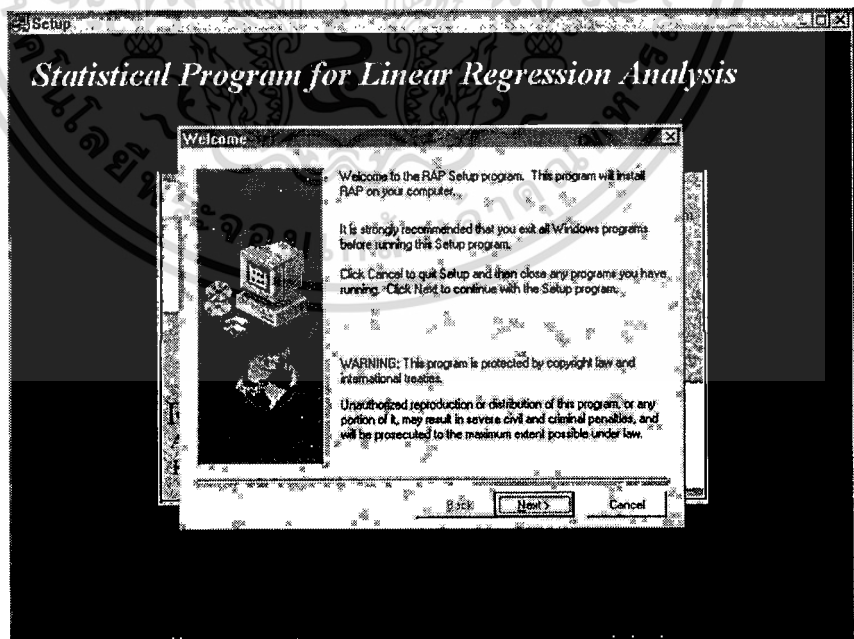
การติดตั้งโปรแกรม RAP

ระบบที่ต้องการ

- IBM PC Compatible รุ่น 486 ขึ้นไป
- หน่วยความจำ RAM 4 MB ขึ้นไป
- พื้นที่ว่างที่เหลือนบน Harddisk อย่างน้อย 24 MB ขึ้นไป (20 MB สำหรับพื้นที่ Swap File ของ Windows)
- ระบบปฏิบัติการ Windows 95 ขึ้นไป

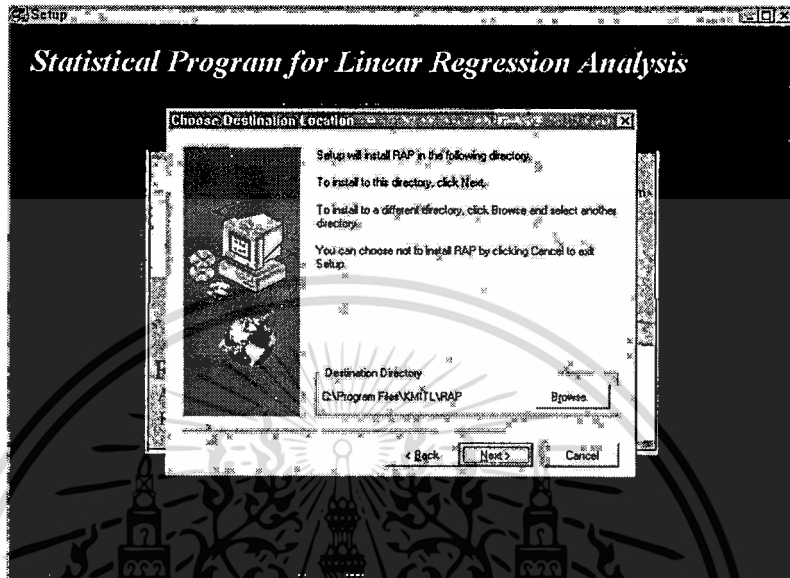
การติดตั้ง

ให้ทำการเลือก Run จาก Taskbar ของ Start Menu แล้วพิมพ์ชื่อเต็มๆที่ไฟล์ติดตั้งอยู่ (เช่น A:\Setup.exe) กดปุ่ม Enter แล้วจะขึ้นหน้าจอดังต่อไปนี้



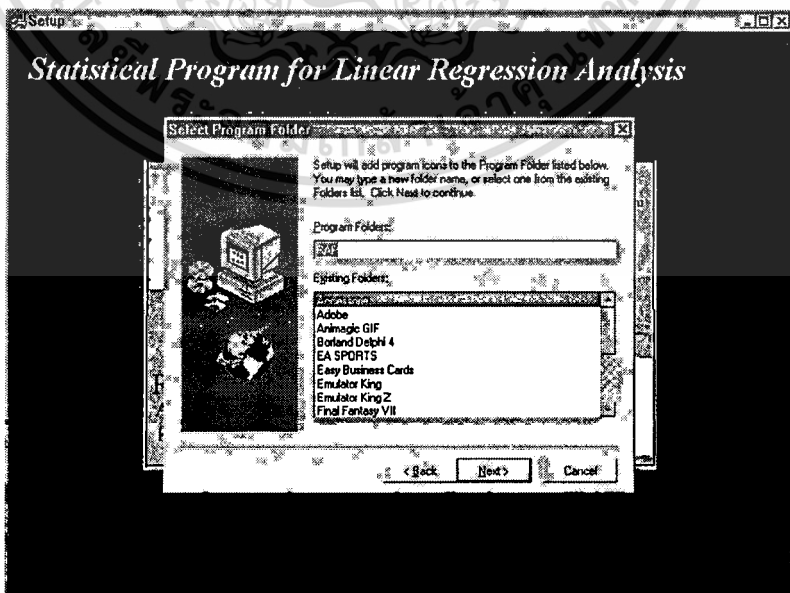
รูปที่ 1 การเริ่มการติดตั้งโปรแกรม RAP

เมื่อทำการเรียกโปรแกรม Setup แล้วที่จอภาพจะขึ้นดังรูปที่ 1 สามารถเลือกตำแหน่งที่เก็บของโปรแกรมว่าจะให้อยู่ในตำแหน่งใดบน Harddisk ซึ่งจะปรากฏดังรูปที่ 2



รูปที่ 2 ทำการเลือก Directory ที่ทำการจัดเก็บไฟล์ใน Harddisk

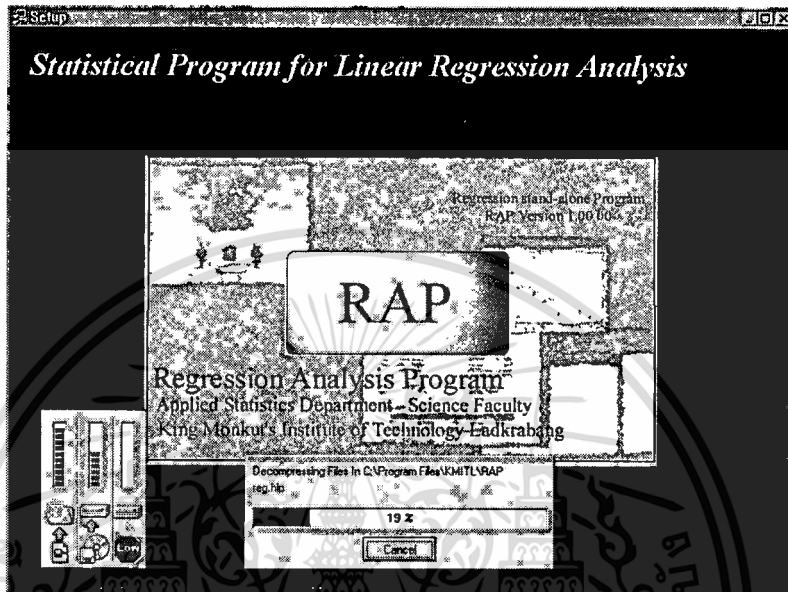
เมื่อทำการเลือกตำแหน่งที่จะจัดเก็บลงบน Harddisk แล้วก็จะให้ทำการเลือกว่าจะให้โปรแกรม RAP นั้นอยู่ใน Folder ไต ในเมนูของโปรแกรม ดังรูปที่ 3



รูปที่ 3 แสดงการนำโปรแกรมมาไว้ใน Folder ที่อยู่ในเมนูของโปรแกรม

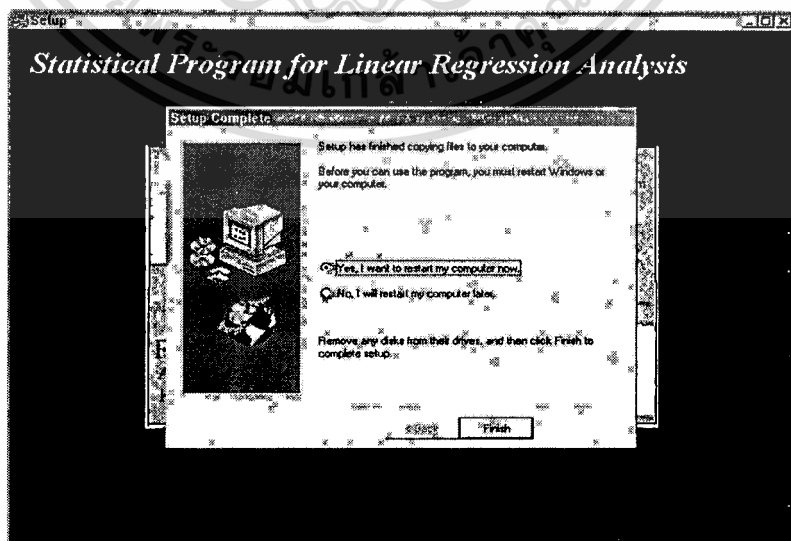
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมื่อได้ทำการเลือก Folder ที่จะให้โปรแกรม RAP อยู่แล้ว ตัวโปรแกรมก็จะเริ่มทำการติดตั้งตัวเองลงสู่ Harddisk ดังรูปที่ 4



รูปที่ 4 ขณะทำการติดตั้งตัวโปรแกรมไฟล์มายัง Harddisk

เมื่อทำการติดตั้งตัวโปรแกรมลงสู่ Harddisk เป็นที่เรียบร้อยแล้วเครื่องก็จะขึ้นตัวเลือกว่าการที่ต้องการที่จะ Restart เครื่องหรือไม่ ดังรูปที่ 5



รูปที่ 5 เมื่อทำการติดตั้งเสร็จจะทำการถามว่าต้องการ restart เครื่องหรือไม่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ไม่ว่าจะเลือกตัวเลือกใดก็ตามเมื่อ Click ที่ปุ่ม Finish ก็จะทำให้การติดตั้งโปรแกรม RAP เสร็จสมบูรณ์ พร้อมทั้งจะใช้งานได้ทันที

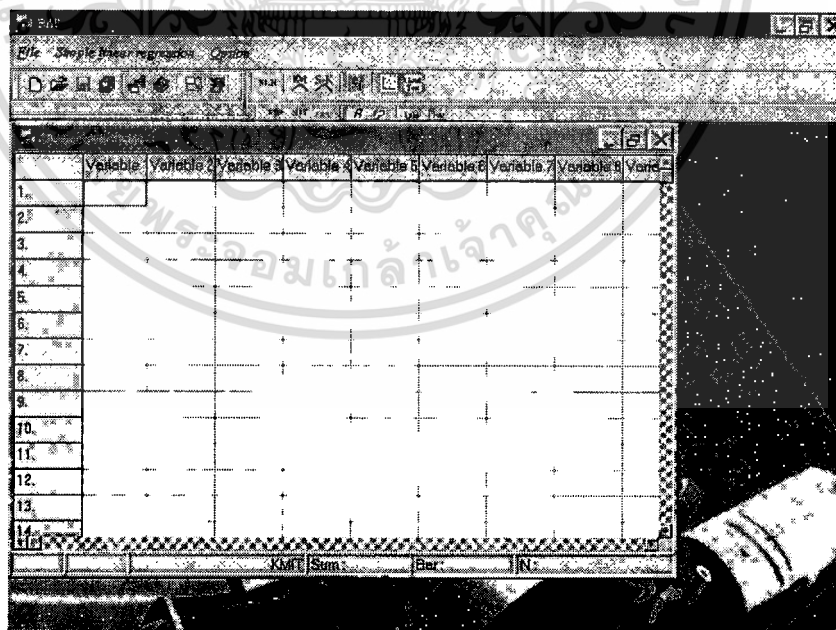
การเรียกใช้โปรแกรม

ให้ทำการเลือกจาก เมนูโปรแกรมแล้วเลือก Folder RAP แล้ว Click ที่โปรแกรม RAP ดังรูป



รูปที่ 6 การเรียกใช้โปรแกรม

จะปรากฏหน้าจอตั้งที่พร้อมจะทำงาน ดังรูปที่ 7



รูปที่ 7 โปรแกรม RAP

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บรรณานุกรม

- ทรงศิริ แต่สมบัติ. การวิเคราะห์การถดถอย. (กรุงเทพมหานคร, สำนักพิมพ์มหาวิทยาลัย
เกษตรศาสตร์, 2541).
- ทัศนีย์ ชังเทศ และ สมภพ ถาวรยิ่ง. การวิเคราะห์การถดถอยและสหสัมพันธ์.
(กรุงเทพมหานคร, สำนักพิมพ์จุฬาลงกรณ์มหาวิทยาลัย, 2530).
- มนตรี พิริยะกุล. เทคนิคการวิเคราะห์สมการถดถอยเล่ม 2. (กรุงเทพมหานคร,
สำนักพิมพ์มหาวิทยาลัยรามคำแหง, 2528).
- วราพร เหลือสินทรัพย์. เอกสารประกอบการสอนการวิเคราะห์การถดถอยเชิงเส้น.
(กรุงเทพมหานคร, สำนักพิมพ์สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง,
2540).
- สุพล ดุรงค์วัฒนา. การวิเคราะห์เชิงสถิติ การวิเคราะห์การถดถอย. (กรุงเทพมหานคร,
สำนักพิมพ์มหาวิทยาลัยธรรมศาสตร์, 2530๗).
- อุมาพร จันทศร. เอกสารประกอบคำสอนวิชา สถิติไม่ใช้พารามิเตอร์. (กรุงเทพมหานคร,
สำนักพิมพ์สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง, 2540).
- Doran Howard E. **Applied Regression Analysis In Economics**. (USA, Marcel Dekker Inc.,
1989).
- Draper Mark And Smith Harry. **Applied Regression Analysis 2nd**. (USA, Wiley
Interscience, 1981).
- Lewis-Beck Michael S. **Regression Analysis Vol 2**. (USA, 1980).
- Microsoft Corporation . **Microsoft Win32 Programmer's Reference Vol 5**, (USA,
Microsoft Press, 1993).
- Miller Todd And Powell David, **Special Edition Using Delphi 3**, (USA, Que
Corporation, 1997).
- Mood Alexander M., Graybill Franklin A. And Boes Duane C. **Introduction To The
Theory Of Statistics 3rd**. (USA, McGraw-Hill International Editions, 1974).
- Thurrgtt Paul, Brent Gary, Bagdazian Richard And Tendon Steve.
Delphi 3 Superbible. (USA, The Waite Group Inc., 1997).