ตัวแบบพยากรณ์ดัชนีคุณภาพอากาศในเขตกรุงเทพมหานคร

FORECASTING MODELS OF AIR QUALITY INDEX IN BANGKOK

เกศกานต์ บริบูรณ์

KATEKARN BORIBOON

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิชาคณิตประยุกต์

คณะวิทยาศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ. 2551

KMITL-2008-SC-M-050-051

สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

ตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศในเขตกรุงเทพมหานคร

FORECASTING MODELS OF AIR QUALITY INDEX IN BANGKOK

เกศกาญจน์ บริบูรณ์

KATEKARN BORIBOON

เลขหมู่.....
เลขทะเบียน.....87863
วัน,เดือน,ปี.....19 มี.ค. 2552

.b.....
.i.....

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิชาสถิติประยุกต์

คณะวิทยาศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ. 2551

KMITL-2008-SC-M-050-051

FORECASTING MODELS OF AIR QUALITY INDEX IN BANGKOK

KATEKARN BORIBOON

A THESIS SUBMITTED IN PARTIAL FULFILLMENT

OF THE REQUIREMENT FOR THE DEGREE OF

MASTER OF SCIENCE PROGRAM IN APPLIED STATISTICS

FACULTY OF SCIENCE

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

2008

KMITL-2008-SC-M-050-051

COPYRIGHT 2008

FACULTY OF SCIENCE

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

คณะวิทยาศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ใบรับรองวิทยานิพนธ์

หัวข้อวิทยานิพนธ์ ตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศในเขตกรุงเทพมหานคร

Forecasting Models of Air Quality Index in Bangkok

นักศึกษา นางสาวเกศกาญจน์ บริบูรณ์

รหัสประจำตัว 48068202

ปริญญา วิทยาศาสตรมหาบัณฑิต

สาขาวิชา สถิติประยุกต์

อาจารย์ที่ปรึกษาวิทยานิพนธ์ ผศ.ดร.สมศรี บัณฑิตวิไล

คณะกรรมการสอบวิทยานิพนธ์		ลายมือชื่อ
ดร.น้อมจิต	กิตติโชติพาณิชย์	
ผศ.ดร.สมศรี	บัณฑิตวิไล	
ผศ.ดร.จุฑาธิป	ตัณสถิตย์	
รศ.สุมิตรา	เรืองพีระกุล	

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

วัน / เดือน/ ปี ที่สอบ 15 ตุลาคม 2551 เวลา 13.00 น. เป็นต้นไป

สถานที่สอบ ณ อาคารจุฬาภรณวลัยลักษณ์ 1 ห้อง 115

คณะวิทยาศาสตร์รับรองแล้ว

(รองศาสตราจารย์ ดร.ธีรวัฒน์ มงคลอัศวรัตน์)

คณบดีคณะวิทยาศาสตร์

วันที่ 30 **เดือน** ตุลาคม **พ.ศ.** 2551

หัวข้อวิทยานิพนธ์	การพยากรณ์ดัชนีคุณภาพอากาศในเขตกรุงเทพมหานคร
นักศึกษา	นางสาวเกศกาญจน์ บริบูรณ์
รหัสประจำตัว	48068202
ปริญญา	วิทยาศาสตรมหาบัณฑิต
สาขาวิชา	สถิติประยุกต์
พ.ศ.	2551
อาจารย์ผู้ควบคุมวิทยานิพนธ์	ผศ. ดร.สมศรี บัณฑิตวิไล

บทคัดย่อ

วิทยานิพนธ์ฉบับนี้ทำการศึกษาหาตัวแบบการพยากรณ์ที่เหมาะสมสำหรับดัชนีคุณภาพอากาศในเขตกรุงเทพมหานครทั้ง 7 สถานี ได้แก่ (1) สถานีการเคหะชุมชนแห่งชาติ (2) สถานีห้วยขวาง (3) สถานียานนาวา (4) สถานีบางขุนเทียน (5) สถานีอินทรพิทักษ์ (6) สถานีลาดพร้าว (7) สถานีดินแดง ข้อมูลที่ใช้ในการพยากรณ์ได้มาจากกรมควบคุมมลพิษ และกรมอุตุนิยมวิทยา ตัวแปรอิสระที่ใช้มีทั้งหมด 19 ตัวแปร ได้แก่ (1) ก๊าซโอโซน (O_3) เฉลี่ย 1 ชั่วโมง (2) ก๊าซไนโตรเจนไดออกไซด์ (NO_2) เฉลี่ย 1 ชั่วโมง (3) ก๊าซคาร์บอนมอนอกไซด์ (CO) เฉลี่ย 1 ชั่วโมง (4) ก๊าซคาร์บอนมอนอกไซด์ (CO) เฉลี่ย 8 ชั่วโมง (5) ก๊าซซัลเฟอร์ไดออกไซด์ (SO_2) เฉลี่ย 24 ชั่วโมง (6) ฝุ่นละอองขนาดเล็กกว่า 10 ไมครอน (PM_{10}) เฉลี่ย 24 ชั่วโมง (7) การระเหยของน้ำในแต่ละวัน (8) ข้อมูลเมฆ (9) อุณหภูมิตุ้มเปียก (10) ทัศนวิสัยโดยเฉลี่ย (11) อุณหภูมิสูงสุดในแต่ละวัน (12) อุณหภูมิต่ำสุดในแต่ละวัน (13) อุณหภูมิตุ้มแห้ง (14) ความยาวนานของแสงแดดในแต่ละวัน (15) ความชื้นสัมพัทธ์เฉลี่ยในแต่ละวัน (16) ปริมาณน้ำฝนในแต่ละวัน (17) ความกดอากาศเฉลี่ยในแต่ละวัน (18) ทิศทางลมในแต่ละวัน (19) ความเร็วลมสูงสุดในแต่ละวัน ซึ่งเป็นข้อมูลรายวัน ตั้งแต่วันที่ 1 มกราคม พ.ศ. 2549 ถึง 31 ธันวาคม พ.ศ. 2549 งานวิจัยนี้ ใช้การวิเคราะห์ปัจจัยด้วยวิธีการวิเคราะห์องค์ประกอบหลักหมุนแกนด้วยวิธีวาริแมกซ์ในการจับกลุ่มตัวแปรอิสระ และใช้การวิเคราะห์การถดถอยเชิงเส้นพหุคูณ โดยมีวิธีการคัดเลือกตัวแปร 3 วิธี ได้แก่ วิธีการเพิ่มตัวแปรอิสระ วิธีการลดตัวแปรอิสระ และวิธีการเพิ่มตัวแปรอิสระทีละขั้นตอน

ผลการวิจัยพบว่า สถานีการเคหะชุมชนแห่งชาติ สถานีห้วยขวาง สถานียานนาวา สถานีอินทรพิทักษ์ สถานีลาดพร้าว และสถานีดินแดงประกอบด้วยปัจจัยร่วม 5 ปัจจัย โดยมีสัมประสิทธิ์การตัดสินใจเชิงพหุ เท่ากับ 0.511, 0.695, 0.716, 0.790, 0.817 และ 0.789 ตามลำดับ และสถานีบางขุนเทียน ประกอบด้วยปัจจัยร่วม 4 ปัจจัย โดยมีสัมประสิทธิ์การตัดสินใจเชิงพหุ เท่ากับ 0.837

Thesis Title	Forecasting Models of Air Quality Index in Bangkok
Student	Miss Katekarn Boriboon
Student ID	48068202
Degree	Master of Science
Programme	Applied Statistics
Year	2008
Thesis Advisor	Asst.Prof.Dr. Somsri Banditvilai

ABSTRACT

The objective of this research is to find the appropriate forecasting models of air quality index for 7 stations in Bangkok : (1) National Housing Authority Station (2) Huay-Khwang Station (3) Yannawa Station (4) Bang-Khunthien Station (5) Intrapituk Station (6) Lad-Phrao Station (7) Din-Daeng Station. The data in this study are collected from Pollution Control Department and Thai Meterological Department. There are 19 independent variables : (1) average ozone per hour (2) average nitrogendioxide per hour (3) average carbonmonoxide per hour (4) average carbonmonoxide per 8 hours (5) average sulfurdioxide per day (6) suspended particulate matter (7) daily water evaporation (8) cloud data (9) wet-bulb temperature (10) mean visibility (11) daily maximum temperature (12) daily minimum temperature (13) dry-bulb temperature (14) daily sunshine duration (15) daily mean relative humidity (16) daily rainfall amount (17) daily mean pressure (18) wind direction (19) wind speed. All data are in daily basis starting from 1 January 2006 to 31 December 2006. This research employed Factor analysis method with Principal component analysis rotated by Varimax method in grouping independent variables. Then, using Multiple Linear Regression Analysis with Forward Selection Procedure, Backward Elimination Procedure and Stepwise Regression Procedure in selecting factors into regression equation.

The result of this research reveals that National Housing Authority Station, Huay-Khwang Station, Yannawa Station, Intrapituk Station, Lad-Phrao Station and Din-Daeng Station are composed of 5 factors with the multiple coefficient of determination 0.511, 0.695, 0.716,

0.790, 0.817 and 0.789 respectively and Bang-Khunthien Station is composed of 4 factors with the multiple coefficient of determination 0.837.

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้อย่างดีเนื่องจากได้รับความอนุเคราะห์จากอาจารย์ที่ปรึกษาวิทยานิพนธ์ ผศ.ดร.สมศรี บัณฑิตวิไล ที่กรุณาให้คำแนะนำ ให้คำปรึกษา ตรวจทานแก้ไขข้อบกพร่องต่าง ๆ เป็นอย่างดี ข้าพเจ้ารู้สึกซาบซึ้งในความอนุเคราะห์จากท่านอาจารย์ และขอกราบขอบพระคุณเป็นอย่างสูง

ขอกราบขอบพระคุณ ผศ.ดร.น้อมจิต กิตติโชติพาณิชย์ ประธานกรรมการสอบวิทยานิพนธ์ ผศ.ดร.จุฑาธิป ตัณสถิตย์ และรศ.สุมิตรา เรืองพีระกุล คณะกรรมการสอบวิทยานิพนธ์ที่ได้กรุณาให้คำแนะนำ ตลอดจนชี้แนะจนวิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยดี

ขอกราบขอบพระคุณคณาจารย์ภาควิชาสถิติประยุกต์ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ทุก ๆ ท่านที่ได้ประสิทธิ์ประสาทวิชาความรู้ให้กับข้าพเจ้า

ขอขอบพระคุณกรมควบคุมมลพิษ กระทรวงทรัพยากรและสิ่งแวดล้อม และกรมอุตุนิยมวิทยา กระทรวงเทคโนโลยีสารสนเทศและการสื่อสาร ที่ให้ความอนุเคราะห์ข้อมูล และให้คำปรึกษาในการทำวิทยานิพนธ์ฉบับนี้

ขอขอบคุณนางสาวพัชราภรณ์ จันต๊ะวงศ์ เพื่อนที่คอยชี้แนะให้คำปรึกษาตลอดจนกำลังใจจนทำให้วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปด้วยดี และขอบคุณเพื่อน ๆ ทุกคนที่ให้ความช่วยเหลือและเป็นกำลังใจในการทำวิทยานิพนธ์นี้ตลอดมา

สุดท้ายนี้ข้าพเจ้าขอกราบขอบพระคุณบิดา มารดา และคุณป้า ของข้าพเจ้าที่คอยเป็นกำลังใจ และให้การสนับสนุนในทุก ๆ เรื่องด้วยดีเสมอมา

คุณความดีอันใดที่บังเกิดจากวิทยานิพนธ์นี้ขอมอบทั้งหมดให้กับคุณย่า คุณปู่ คุณลุง และคุณป้าผู้ล่วงลับอันเป็นที่รักยิ่ง

เกศกาญจน์ บริบูรณ์

สารบัญ

หน้า

บทคัดย่อภาษาไทย.....I

บทคัดย่อภาษาอังกฤษ.....II

กิตติกรรมประกาศ.....IV

สารบัญ.....V

สารบัญตาราง.....VII

สารบัญรูป.....XI

บทที่ 1 บทนำ.....1

1.1 ความเป็นมาและความสำคัญของปัญหา.....1

1.2 วัตถุประสงค์ของการวิจัย.....2

1.3 ขอบเขตของการวิจัย.....2

1.4 นิยามศัพท์.....4

1.5 ประโยชน์ที่คาดว่าจะได้รับ.....9

บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง.....10

2.1 ทฤษฎีที่ใช้ในการวิเคราะห์.....10

2.1.1 การวิเคราะห์ปัจจัย.....10

2.1.2 การวิเคราะห์การถดถอยเชิงเส้นพหุคูณ.....22

2.2 งานวิจัยที่เกี่ยวข้อง.....40

บทที่ 3 วิธีการดำเนินการวิจัย.....43

3.1 การเก็บรวบรวมข้อมูล.....43

3.2 การวิเคราะห์ข้อมูลเพื่อสร้างตัวแบบ.....44

สารบัญ (ต่อ)

หน้า

บทที่ 4 ผลการวิเคราะห์ข้อมูล....48

4.1 ผลการวิเคราะห์ข้อมูลในการสร้างตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีการเคหะชุมชนแห่งชาติ....48

4.2 ผลการวิเคราะห์ข้อมูลในการสร้างตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีห้วยขวาง....63

4.3 ผลการวิเคราะห์ข้อมูลในการสร้างตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานียานนาวา....73

4.4 ผลการวิเคราะห์ข้อมูลในการสร้างตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีบางขุนเทียน....82

4.5 ผลการวิเคราะห์ข้อมูลในการสร้างตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีอินทรพิทักษ์....91

4.6 ผลการวิเคราะห์ข้อมูลในการสร้างตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีลาดพร้าว....100

4.7 ผลการวิเคราะห์ข้อมูลในการสร้างตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีดินแดง....109

บทที่ 5 สรุปผลการวิเคราะห์และข้อเสนอแนะ....125

5.1 สรุปผลการวิจัย....125

5.2 ข้อเสนอแนะ....129

บรรณานุกรม....131

ภาคผนวก....134

ประวัติผู้เขียน....135

สารบัญตาราง

ตารางที่ หน้า

1.1 เกณฑ์ของดัชนีคุณภาพอากาศสำหรับประเทศไทย.....5
1.2 ค่าความเข้มข้นของสารมลพิษทางอากาศที่เทียบกับค่าดัชนีคุณภาพอากาศ.....6
2.1 การวิเคราะห์ความแปรปรวนของการวิเคราะห์การถดถอยเชิงพหุคูณ.....32
4.1 แสดงผลการวิเคราะห์ปัจจัยของตัวแปรอิสระสำหรับการพยากรณ์ดัชนีคุณภาพอากาศ ของสถานีการเคหะชุมชนแห่งชาติ.....49
4.2 แสดงผลการหมุนแกนปัจจัยด้วยวิธีวาริแมกซ์ของสถานีการเคหะชุมชนแห่งชาติ.....50
4.3 แสดงผลการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยโดยวิธีการเพิ่มตัวแปรอิสระ ของตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีการเคหะชุมชนแห่งชาติ.....53
4.4 แสดงผลการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยโดยวิธีการลดตัวแปรอิสระ ของตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีการเคหะชุมชนแห่งชาติ.....54
4.5 แสดงการทดสอบค่าความคลาดเคลื่อนมีการแจกแจงปกติของสถานีการเคหะ ชุมชนแห่งชาติ.....55
4.6 แสดงค่า Durbin-Watson ที่ได้จากการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยด้วยวิธี การลดตัวแปรอิสระของสถานีการเคหะชุมชนแห่งชาติ.....56
4.7 แสดงผลการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยโดยวิธีการเพิ่มตัวแปรอิสระ ของตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีการเคหะชุมชนแห่งชาติ หลังจากการแปลงข้อมูลตัวแปรตาม.....59
4.8 แสดงการทดสอบค่าความคลาดเคลื่อนมีการแจกแจงปกติของสถานีการเคหะ ชุมชนแห่งชาติ หลังจากการแปลงข้อมูลตัวแปรตาม.....61
4.9 แสดงค่า Durbin-Watson ที่ได้จากการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยด้วยวิธี การเพิ่มตัวแปรอิสระทีละขั้นตอนของสถานีการเคหะชุมชนแห่งชาติ หลังจากการแปลง ข้อมูลตัวแปรตาม.....61
4.10 แสดงผลการวิเคราะห์ปัจจัยของตัวแปรอิสระสำหรับการพยากรณ์ดัชนีคุณภาพอากาศ ของสถานีห้วยขวาง.....64
4.11 แสดงผลการหมุนแกนปัจจัยด้วยวิธีวาริแมกซ์ของสถานีห้วยขวาง.....65

สารบัญตาราง (ต่อ)

ตารางที่ หน้า

4.12 แสดงผลการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยโดยวิธีการเพิ่มตัวแปรอิสระของตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีห้วยขวาง....68
4.13 แสดงผลการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยโดยวิธีการลดตัวแปรอิสระของตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีห้วยขวาง....69
4.14 แสดงการทดสอบค่าความคลาดเคลื่อนมีการแจกแจงปกติของสถานีห้วยขวาง....71
4.15 แสดงค่า Durbin-Watson ที่ได้จากการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยด้วยวิธีการลดตัวแปรอิสระของสถานีห้วยขวาง....71
4.16 แสดงผลการวิเคราะห์ปัจจัยของตัวแปรอิสระสำหรับการพยากรณ์ดัชนีคุณภาพอากาศของสถานียานนาวา....74
4.17 แสดงผลการหมุนแกนปัจจัยด้วยวิธีวาริแมกซ์ของสถานียานนาวา....75
4.18 แสดงผลการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยโดยวิธีการเพิ่มตัวแปรอิสระของตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานียานนาวา....78
4.19 แสดงการทดสอบค่าความคลาดเคลื่อนมีการแจกแจงปกติของสถานียานนาวา....80
4.20 แสดงค่า Durbin-Watson ที่ได้จากการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยด้วยวิธีการเพิ่มตัวแปรอิสระทีละขั้นตอนของสถานียานนาวา....80
4.21 แสดงผลการวิเคราะห์ปัจจัยของตัวแปรอิสระสำหรับการพยากรณ์ดัชนีคุณภาพอากาศของสถานีบางขุนเทียน....83
4.22 แสดงผลการหมุนแกนปัจจัยด้วยวิธีวาริแมกซ์ของสถานีบางขุนเทียน....84
4.23 แสดงผลการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยโดยวิธีการเพิ่มตัวแปรอิสระของตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีบางขุนเทียน....86
4.24 แสดงการทดสอบค่าความคลาดเคลื่อนมีการแจกแจงปกติของสถานีบางขุนเทียน....88
4.25 แสดงค่า Durbin-Watson ที่ได้จากการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยด้วยวิธีการเพิ่มตัวแปรอิสระทีละขั้นตอนของสถานีบางขุนเทียน....89
4.26 แสดงผลการวิเคราะห์ปัจจัยของตัวแปรอิสระสำหรับการพยากรณ์ดัชนีคุณภาพอากาศของสถานีอินทรพิทักษ์....92

สารบัญตาราง (ต่อ)

ตารางที่ หน้า

4.27 แสดงผลการหมุนแกนปัจจัยด้วยวิธีวาริแมกซ์ของสถานีอินทรพิทักษ์.....93
4.28 แสดงผลการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยโดยวิธีการเพิ่มตัวแปรอิสระ ของตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีอินทรพิทักษ์..... 95
4.29 แสดงการทดสอบค่าความคลาดเคลื่อนมีการแจกแจงปกติของสถานีอินทรพิทักษ์.....96
4.30 แสดงค่า Durbin-Watson ที่ได้จากการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยด้วยวิธี การเพิ่มตัวแปรอิสระทีละขั้นตอนของสถานีอินทรพิทักษ์.....98
4.31 แสดงผลการวิเคราะห์ปัจจัยของตัวแปรอิสระสำหรับการพยากรณ์ดัชนีคุณภาพอากาศ ของสถานีลาดพร้าว.....101
4.32 แสดงผลการหมุนแกนปัจจัยด้วยวิธีวาริแมกซ์ของสถานีลาดพร้าว.....102
4.33 แสดงผลการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยโดยวิธีการเพิ่มตัวแปรอิสระ ของตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีลาดพร้าว.....105
4.34 แสดงการทดสอบค่าความคลาดเคลื่อนมีการแจกแจงปกติของสถานีลาดพร้าว.....107
4.35 แสดงค่า Durbin-Watson ที่ได้จากการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยด้วยวิธี การเพิ่มตัวแปรอิสระทีละขั้นตอนของสถานีลาดพร้าว.....107
4.36 แสดงผลการวิเคราะห์ปัจจัยของตัวแปรอิสระสำหรับการพยากรณ์ดัชนีคุณภาพอากาศ ของสถานีดินแดง.....110
4.37 แสดงผลการหมุนแกนปัจจัยด้วยวิธีวาริแมกซ์ของสถานีดินแดง.....111
4.38 แสดงผลการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยโดยวิธีการเพิ่มตัวแปรอิสระ ของตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีดินแดง.....114
4.39 แสดงการทดสอบค่าความคลาดเคลื่อนมีการแจกแจงปกติของสถานีดินแดง.....116
4.40 แสดงค่า Durbin-Watson ที่ได้จากการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยด้วยวิธี การเพิ่มตัวแปรอิสระทีละขั้นตอนของสถานีดินแดง.....116
4.41 แสดงผลการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยโดยวิธีการเพิ่มตัวแปรอิสระ ของตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีดินแดง หลังจากการแปลง ข้อมูลตัวแปรตาม.....119

สารบัญตาราง (ต่อ)

ตารางที่ หน้า

4.42 แสดงผลการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยโดยวิธีการลดตัวแปรอิสระของตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีดินแดง หลังจากการแปลงข้อมูลตัวแปรตาม.....120
4.43 แสดงการทดสอบค่าความคลาดเคลื่อนมีการแจกแจงปกติของสถานีดินแดง หลังการแปลงข้อมูลตัวแปรตาม.....122
4.44 แสดงค่า Durbin-Watson ที่ได้จากการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยด้วยวิธีการลดตัวแปรอิสระของสถานีดินแดง หลังจากการแปลงข้อมูลตัวแปรตาม.....122
ก-1 ข้อมูลที่ใช้ในการวิเคราะห์ดัชนีคุณภาพอากาศของสถานีการเคหะชุมชนแห่งชาติ.....133
ก-2 ข้อมูลที่ใช้ในการวิเคราะห์ดัชนีคุณภาพอากาศของสถานีห้วยขวาง.....133
ก-3 ข้อมูลที่ใช้ในการวิเคราะห์ดัชนีคุณภาพอากาศของสถานียานนาวา.....133
ก-4 ข้อมูลที่ใช้ในการวิเคราะห์ดัชนีคุณภาพอากาศของสถานีบางขุนเทียน.....133
ก-5 ข้อมูลที่ใช้ในการวิเคราะห์ดัชนีคุณภาพอากาศของสถานีอินทรพิทักษ์.....133
ก-6 ข้อมูลที่ใช้ในการวิเคราะห์ดัชนีคุณภาพอากาศของสถานีลาดพร้าว.....133
ก-7 ข้อมูลที่ใช้ในการวิเคราะห์ดัชนีคุณภาพอากาศของสถานีดินแดง.....133

สารบัญรูป

รูปที่ หน้า

1.1 แสดงสถานที่ตั้งของสถานีตรวจวัดคุณภาพอากาศทั้ง 7 สถานี และสถานีตรวจวัดสภาพอากาศ กรมอุตุนิยมวิทยา.....3

2.1 แสดงลักษณะของค่าความคลาดเคลื่อน กรณีที่ค่าความแปรปรวน (ก) คงที่ (ข) เพิ่มขึ้น และ (ค) ลดลง.....29

4.1 กราฟ Normal Probability Plot ของค่าเศษเหลือของสมการถดถอยสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีการเคหะชุมชนแห่งชาติ.....55

4.2 ความสัมพันธ์ระหว่าง $E(e)$ กับ $E(\hat{Y})$ ของสถานีการเคหะชุมชนแห่งชาติ.....57

4.3 แสดง Histogram ค่าเศษเหลือของสถานีการเคหะชุมชนแห่งชาติ.....58

4.4 กราฟ Normal Probability Plot ของค่าเศษเหลือของสมการถดถอยสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีการเคหะชุมชนแห่งชาติ หลังจากการแปลงข้อมูลตัวแปรตาม.....60

4.5 ความสัมพันธ์ระหว่าง $E(e)$ กับ $E(\hat{Y})$ ของสถานีการเคหะชุมชนแห่งชาติหลังการแปลงข้อมูลตัวแปรตาม.....62

4.6 กราฟ Normal Probability Plot ของค่าเศษเหลือของสมการถดถอยสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีห้วยขวาง.....70

4.7 ความสัมพันธ์ระหว่าง $E(e)$ กับ $E(\hat{Y})$ ของสถานีห้วยขวาง.....72

4.8 กราฟ Normal Probability Plot ของค่าเศษเหลือของสมการถดถอยสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานียานนาวา.....79

4.9 ความสัมพันธ์ระหว่าง $E(e)$ กับ $E(\hat{Y})$ ของสถานียานนาวา.....81

4.10 กราฟ Normal Probability Plot ของค่าเศษเหลือของสมการถดถอยสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีบางขุนเทียน.....88

4.11 ความสัมพันธ์ระหว่าง $E(e)$ กับ $E(\hat{Y})$ ของสถานีบางขุนเทียน.....90

4.12 กราฟ Normal Probability Plot ของค่าเศษเหลือของสมการถดถอยสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีอินทรพิทักษ์.....97

สารบัญรูป (ต่อ)

รูปที่ หน้า

4.13 ความสัมพันธ์ระหว่าง $E(e)$ กับ $E(\hat{Y})$ ของสถานีอินทรพิทักษ์.....99
4.14 กราฟ Normal Probability Plot ของค่าเศษเหลือของสมการถดถอยสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีลาดพร้าว.....106
4.15 ความสัมพันธ์ระหว่าง $E(e)$ กับ $E(\hat{Y})$ ของสถานีลาดพร้าว.....108
4.16 กราฟ Normal Probability Plot ของค่าเศษเหลือของสมการถดถอยสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีดินแดง.....115
4.17 ความสัมพันธ์ระหว่าง $E(e)$ กับ $E(\hat{Y})$ ของสถานีดินแดง.....117
4.18 แสดง Histogram ค่าเศษเหลือของสถานีดินแดง.....118
4.19 กราฟ Normal Probability Plot ของค่าเศษเหลือของสมการถดถอยสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีดินแดง หลังจากการแปลงข้อมูลตัวแปรตาม.....121
4.20 ความสัมพันธ์ระหว่าง $E(e)$ กับ $E(\hat{Y})$ ของสถานีดินแดง หลังจากการแปลงข้อมูลตัวแปรตาม.....123

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในปัจจุบันปัญหาสิ่งแวดล้อมเป็นพิษมีหลายปัญหา ปัญหาที่สำคัญและเป็นปัญหาที่รุนแรง คือ ปัญหามลพิษทางอากาศ ซึ่งมักเกิดกับเมืองใหญ่ โดยเฉพาะอย่างยิ่งกรุงเทพมหานคร (จันทนี เกียรติโพธา. 2542 : 1) เนื่องจากกรุงเทพมหานครเป็นเมืองที่มีขนาดใหญ่เป็นที่อยู่ของประชากรจำนวนมาก อีกทั้งยังเป็นศูนย์รวมของประเทศทั้งการเมือง เศรษฐกิจ และสังคม มีความก้าวหน้าทางวิชาการและเทคโนโลยีสูง การขยายตัวอย่างรวดเร็วของกิจการอุตสาหกรรม การคมนาคมขนส่ง และการจราจร (ณัฐวสา เจริญ 2541 : 1) จากสภาพดังกล่าวทำให้กรุงเทพมหานครมีการพัฒนาเมืองอย่างไม่สิ้นสุด ได้แก่ การก่อสร้างตึกระฟ้าเพื่อใช้เป็นสำนักงานของเอกชนและที่อยู่อาศัย การพัฒนาระบบคมนาคม การขยาย และการปรับปรุงระบบสาธารณูปโภคของภาครัฐ สิ่งเหล่านี้ก่อให้เกิดมลพิษทางอากาศอย่างหลีกเลี่ยงไม่ได้ (จันทนี เกียรติโพธา. 2542 : 1)

สาเหตุสำคัญของมลพิษทางอากาศในกรุงเทพมหานครมาจากการคมนาคมขนส่ง ดังนั้น กรุงเทพมหานครจึงได้ชื่อว่าเป็นเมืองที่มีรถติดมากที่สุดในโลก เนื่องจากจำนวนยานพาหนะที่เพิ่มขึ้นอย่างรวดเร็ว แต่อัตราการเพิ่มพื้นที่จราจรและยานพาหนะไม่ได้สัดส่วน โดยอัตราการเพิ่มของยานพาหนะมีปริมาณสูงกว่าถนน ทำให้เกิดปัญหาการจราจรติดขัด มีผลทำให้เกิดการเผาไหม้ไม่สมบูรณ์และเกิดการระบายสารพิษต่าง ๆ จากท่อไอเสียรถยนต์ นอกจากนี้ยังรวมไปถึงการก่อสร้างอาคารบ้านเรือน การก่อสร้างปรับปรุงถนน และพื้นผิวการจราจร ตลอดจนโรงงานอุตสาหกรรมและสถานประกอบการต่าง ๆ สารมลพิษเหล่านี้ได้แก่ ฝุ่นละออง ก๊าซคาร์บอนมอนอกไซด์ ก๊าซซัลเฟอร์ไดออกไซด์ และก๊าซไนโตรเจนไดออกไซด์ (กรมควบคุมมลพิษ. 2544 : 3)

มลพิษทางอากาศทำให้เกิดผลกระทบต่อมนุษย์ทั้งทางร่างกายและจิตใจ การสูดดมอากาศที่มีมลพิษเข้าสู่ร่างกายทุกวันจะทำให้สุขภาพเสื่อมโทรม ร่างกายอ่อนแอ เจ็บป่วยด้วยโรคระบบทางเดินหายใจแบบเรื้อรังและเฉียบพลันได้ ทำให้เกิดความสิ้นเปลืองในการรักษาพยาบาล จึงนับได้ว่าเป็นการสูญเสียทางเศรษฐกิจ และกระทบต่อการพัฒนาประเทศโดยรวม ดังนั้นจึงมีการเฝ้าระวังโดยการตรวจวัดคุณภาพของอากาศจากกรมควบคุมมลพิษ กระทรวงทรัพยากรธรรมชาติและสิ่งแวดล้อม สำหรับการตรวจวัดสารมลพิษทางอากาศ มี 5 ประเภท ได้แก่ ก๊าซโอโซน (O_3) เฉลี่ย 1 ชั่วโมง ก๊าซไนโตรเจนไดออกไซด์ (NO_2) เฉลี่ย 1 ชั่วโมง ก๊าซคาร์บอนมอนอกไซด์ (CO) เฉลี่ย 1

ชั่วโมง และ 8 ชั่วโมง ก๊าซซัลเฟอร์ไดออกไซด์ (SO_2) เฉลี่ย 24 ชั่วโมง และฝุ่นละอองขนาดเล็กกว่า 10 ไมครอน (PM_{10}) เฉลี่ย 24 ชั่วโมง ทำการตรวจวัดในบริเวณต่าง ๆ ทั่วกรุงเทพมหานครทั้งหมด 7 สถานี นั่นคือ สถานีการเคหะชุมชนแห่งชาติ สถานีห้วยขวาง สถานียานนาวา สถานีบางขุนเทียน สถานีอินทรพิทักษ์ สถานีลาดพร้าว และสถานีดินแดง

นอกจากนี้สภาพอากาศ อุณหภูมิ ความเร็วลม ความกดอากาศ ฯลฯ ก็เป็นส่วนหนึ่งที่ส่งผลกระทบต่อการกระจายตัวของสารพิษในอากาศได้เช่นกัน ผู้วิจัยจึงนำเข้ามาร่วมวิเคราะห์เพื่อหาความสัมพันธ์ที่เกิดขึ้น

1.2 วัตถุประสงค์ของการวิจัย

เพื่อหาตัวแบบที่เหมาะสมในการพยากรณ์ดัชนีคุณภาพอากาศของกรุงเทพมหานครทั้ง 7 สถานี อันได้แก่ สถานีการเคหะชุมชนแห่งชาติ สถานีห้วยขวาง สถานียานนาวา สถานีบางขุนเทียน สถานีอินทรพิทักษ์ สถานีลาดพร้าว และสถานีดินแดง โดยใช้การวิเคราะห์ปัจจัย (Factor Analysis) และการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ (Multiple Linear Regression Analysis)

1.3 ขอบเขตของการวิจัย

การศึกษาวิจัยครั้งนี้ มุ่งเน้นในการหาตัวแบบที่เหมาะสมในการพยากรณ์ดัชนีคุณภาพอากาศของกรุงเทพมหานครทั้ง 7 สถานี ได้แก่

1. สถานีการเคหะชุมชนแห่งชาติ จุดตรวจวัดอยู่ที่สำนักงานการเคหะชุมชนคลองจั่น ถนนสุขาภิบาล 1 เขตบางกะปิ

2. สถานีห้วยขวาง จุดตรวจวัดอยู่ที่สนามกีฬาการเคหะชุมชนห้วยขวาง ถนนประชาสงเคราะห์ เขตห้วยขวาง

3. สถานียานนาวา จุดตรวจวัดอยู่ที่โรงเรียนนนทรีวิทยา ถนนนางลิ้นจี่ เขตยานนาวา

4. สถานีบางขุนเทียน จุดตรวจวัดอยู่ที่โรงเรียนสิงหราชพิทยาคม ถนนเอกชัย เขตบางขุนเทียน

5. สถานีอินทรพิทักษ์ จุดตรวจวัดอยู่ที่สถานีการไฟฟ้าย่อยธนบุรี ถนนอินทรพิทักษ์ เขตธนบุรี

6. สถานีลาดพร้าว จุดตรวจวัดอยู่ที่ที่พักตำรวจจราจรโชคชัย 4 ถนนลาดพร้าว เขตลาดพร้าว

7. สถานีดินแดง จุดตรวจวัดอยู่ที่การเคหะชุมชนดินแดง ถนนดินแดง เขตดินแดง

รูปที่ 1.1 แสดงสถานที่ตั้งของสถานีตรวจวัดคุณภาพอากาศทั้ง 7 สถานี และสถานีตรวจวัดสภาพอากาศ กรมอุตุนิยมวิทยา

ซึ่งจะวิเคราะห์โดยใช้การวิเคราะห์การถดถอยเชิงเส้นพหุคูณ (Multiple Linear Regression Analysis) ใช้วิธีการคัดเลือกตัวแปรทั้ง 3 วิธี คือ วิธีการเพิ่มตัวแปรอิสระ (Forward Selection Procedure) วิธีการลดตัวแปรอิสระ (Backward Elimination Procedure) และวิธีการเพิ่มตัวแปรอิสระทีละขั้นตอน (Stepwise Regression Procedure) พิจารณาจากวิธีการที่ได้ค่า R^2 สูงที่สุดเป็นตัวแบบที่ดีที่สุด เนื่องจากตัวแปรอิสระที่ใช้มีจำนวนมาก และตัวแปรอิสระบางตัวอาจมีความสัมพันธ์กัน จึงใช้การวิเคราะห์ปัจจัย (Factor Analysis) ด้วยวิธีวิเคราะห์องค์ประกอบหลัก (Principal component analysis) มาช่วยในการจับกลุ่มตัวแปรอิสระที่มีความสัมพันธ์กัน ซึ่งมีตัวแปรทั้งหมดดังนี้

ตัวแปรตาม คือ ดัชนีคุณภาพอากาศทั้ง 7 สถานี ได้แก่

1. สถานีการเคหะชุมชนแห่งชาติ (เขตบางกะปิ)
2. สถานีห้วยขวาง (เขตห้วยขวาง)
3. สถานียานนาวา (เขตยานนาวา)
4. สถานีบางขุนเทียน (เขตบางขุนเทียน)
5. สถานีอินทรพิทักษ์ (เขตธนบุรี)
6. สถานีลาดพร้าว (เขตลาดพร้าว)
7. สถานีดินแดง (เขตดินแดง)

ตัวแปรอิสระ คือ ค่าที่วัดมลพิษทางอากาศ และค่าที่วัดสภาพอากาศ มีทั้งหมด 19 ตัว ได้แก่

1. ก๊าซโอโซน (O_3) เฉลี่ย 1 ชั่วโมง มีหน่วยเป็นส่วนในพันล้านส่วน (X_1)
2. ก๊าซไนโตรเจนไดออกไซด์ (NO_2) เฉลี่ย 1 ชั่วโมง มีหน่วยเป็นส่วนในพันล้านส่วน (X_2)
3. ก๊าซคาร์บอนมอนอกไซด์ (CO) เฉลี่ย 1 ชั่วโมง มีหน่วยเป็นส่วนในล้านส่วน (X_3)
4. ก๊าซคาร์บอนมอนอกไซด์ (CO) เฉลี่ย 8 ชั่วโมง มีหน่วยเป็นส่วนในล้านส่วน (X_4)
5. ก๊าซซัลเฟอร์ไดออกไซด์ (SO_2) เฉลี่ย 24 ชั่วโมง มีหน่วยเป็นส่วนในพันล้านส่วน (X_5)
6. ฝุ่นละอองขนาดเล็กกว่า 10 ไมครอน (PM_{10}) เฉลี่ย 24 ชั่วโมง มีหน่วยเป็นไมโครกรัมต่อลูกบาศก์เมตร (X_6)
7. การระเหยของน้ำในแต่ละวัน มีหน่วยเป็นมิลลิเมตร (X_7)
8. ข้อมูลเมฆ มีหน่วยเป็นเดคะ (X_8)
9. อุณหภูมิตุ้มเปียก มีหน่วยเป็นเซลเซียส (X_9)
10. ทัศนวิสัยโดยเฉลี่ย มีหน่วยเป็นกิโลเมตร (X_{10})
11. อุณหภูมิสูงสุดในแต่ละวัน มีหน่วยเป็นเซลเซียส (X_{11})
12. อุณหภูมิต่ำสุดในแต่ละวัน มีหน่วยเป็นเซลเซียส (X_{12})
13. อุณหภูมิตุ้มแห้ง มีหน่วยเป็นเซลเซียส (X_{13})
14. ความยาวนานของแสงแดดในแต่ละวัน มีหน่วยเป็นชั่วโมง (X_{14})
15. ความชื้นสัมพัทธ์เฉลี่ยในแต่ละวัน มีหน่วยเป็นเปอร์เซนต์ (X_{15})
16. ปริมาณน้ำฝนในแต่ละวัน มีหน่วยเป็นมิลลิเมตร (X_{16})
17. ความกดอากาศเฉลี่ยในแต่ละวัน มีหน่วยเป็นเฮกโตปาสคาล (X_{17})
18. ทิศทางลมในแต่ละวัน มีหน่วยเป็นมุมองศา (X_{18})
19. ความเร็วลมสูงสุดในแต่ละวัน มีหน่วยเป็นนอต (X_{19})

1.4 นิยามศัพท์

มลพิษทางอากาศ หมายถึง ภาวะอากาศที่มีสารเจือปนอยู่ในปริมาณที่สูงกว่าระดับปกติ เป็นเวลานานพอที่จะทำให้เกิดอันตรายแก่มนุษย์ สัตว์ พืช หรือทรัพย์สินต่าง ๆ อาจเกิดขึ้นเองตามธรรมชาติ เช่น ฝุ่นละอองจากลมพายุ ภูเขาไฟระเบิด แผ่นดินไหว ไฟไหม้ป่า กรณีที่เกิดจากการกระทำของมนุษย์ เช่น มลพิษจากท่อไอเสียรถยนต์ โรงงานอุตสาหกรรม (www.pcd.go.th)

ดัชนีคุณภาพอากาศ (AQI) เป็นการรายงานข้อมูลคุณภาพอากาศในรูปแบบหนึ่งที่เผยแพร่ประชาสัมพันธ์ให้สาธารณชนได้ทราบถึงสถานการณ์มลพิษทางอากาศในแต่ละพื้นที่ว่าอยู่ในระดับ

ใด มีผลกระทบต่อสุขภาพอนามัยหรือไม่ โดยดัชนีคุณภาพอากาศที่ใช้อยู่ในประเทศไทย แบ่งเป็น 5 ระดับ คือ ตั้งแต่ 0 ถึง มากกว่า 300 ซึ่งแต่ละระดับใช้สีเป็นสัญลักษณ์เปรียบเทียบระดับของผลกระทบต่อสุขภาพอนามัย ดังตารางที่ 1.1 (www.pcd.go.th)

ตารางที่ 1.1 เกณฑ์ของดัชนีคุณภาพอากาศสำหรับประเทศไทย

AQI	ความหมาย	สีที่ใช้	แนวทางการป้องกันผลกระทบ
0-50	คุณภาพดี	ฟ้า	ไม่มีผลกระทบต่อสุขภาพ
51-100	คุณภาพปานกลาง	เขียว	ไม่มีผลกระทบต่อสุขภาพ
101-200	มีผลกระทบต่อสุขภาพ	เหลือง	ผู้ป่วยโรคระบบทางเดินหายใจ ควรหลีกเลี่ยงการออกกำลังกายภายนอกอาคาร บุคคลทั่วไป โดยเฉพาะเด็กและผู้สูงอายุ ไม่ควรทำกิจกรรมภายนอกอาคารเป็นเวลานาน
201-300	มีผลกระทบต่อสุขภาพมาก	ส้ม	ผู้ป่วยโรคระบบทางเดินหายใจ ควรหลีกเลี่ยงกิจกรรมภายนอกอาคาร บุคคลทั่วไป โดยเฉพาะเด็กและผู้สูงอายุ ควรจำกัดการทำกิจกรรมภายนอกอาคาร
มากกว่า 300	อันตราย	แดง	ผู้ป่วยโรคระบบทางเดินหายใจ ควรอยู่ภายในอาคาร บุคคลทั่วไป ควรหลีกเลี่ยงการทำกิจกรรมภายนอกอาคาร

<u>การคำนวณดัชนีคุณภาพอากาศรายวัน</u>

คำนวณจากค่าความเข้มข้นของสารมลพิษทางอากาศจากข้อมูลผลการตรวจวัดคุณภาพอากาศ โดยแต่ละระดับของค่าความเข้มข้นของสารมลพิษทางอากาศเทียบเท่ากับค่าดัชนีคุณภาพอากาศที่ระดับต่าง ๆ (ตารางที่ 1.2) และมีสูตรการคำนวณดังนี้

$$I_i = \frac{I_{ij+1}-I_{ij}}{X_{ij+1}-X_{ij}}\left(X_i-X_{ij}\right) + I_{ij} \qquad (1.1)$$

โดยที่ X_i = ความเข้มข้นของสารมลพิษทางอากาศจากผลการตรวจวัด

X_{ij} = ความเข้มข้นของสารมลพิษทางอากาศที่เป็นค่าต่ำสุดของช่วงพิสัยที่มีค่า X_i นั้น

X_{ij+1} = ความเข้มข้นของสารมลพิษทางอากาศที่เป็นค่าสูงสุดของช่วงพิสัยที่มีค่า X_i นั้น

I_i = ค่าดัชนีย่อยคุณภาพอากาศ

I_{ij} = ค่าดัชนีย่อยคุณภาพอากาศที่เป็นค่าต่ำสุดของช่วงพิสัยที่มีค่า I_i นั้น

I_{ij+1} = ค่าดัชนีย่อยคุณภาพอากาศที่เป็นค่าสูงสุดของช่วงพิสัยที่มีค่า I_i นั้น

AQI = ค่าดัชนีคุณภาพอากาศ

เมื่อ i = ประเภทของสารมลพิษทางอากาศ ; $i = 1, 2, \dots, 6$

j = ช่วงในตารางที่ 1.2 ที่สารมลพิษทางอากาศตกอยู่ ; $j = 1, 2, \dots, 6$

$$AQI = Max(I_1, I_2, \dots, I_6) \quad (1.2)$$

ค่าดัชนีคุณภาพอากาศ (AQI) ในแต่ละวันจะมาจากค่าดัชนีย่อยคุณภาพอากาศที่มากที่สุดในวันนั้น

ตารางที่ 1.2 ค่าความเข้มข้นของสารมลพิษทางอากาศที่เทียบกับค่าดัชนีคุณภาพอากาศ

AQI	PM_{10} (24 ชม.)	O_3 (1 ชม.)		SO_2 (24 ชม.)		NO_2 (1 ชม.)		CO (8 ชม.)	
	มคก./ลบม.	มคก./ลบม.	ppb	มคก./ลบม.	ppb	มคก./ลบม.	ppb	มคก./ลบม.	ppb
50	40	100	51	65	25	160	85	5.13	4.48
100	120	200	100	300	120	320	170	10.26	9.00
200	350	400	203	800	305	1130	600	17.00	14.84
300	420	800	405	1600	610	2260	1202	34.00	29.69
400	500	1000	509	2100	802	3000	1594	46.00	40.17
500	600	1200	611	2620	1000	3750	1993	57.50	50.21

ก๊าซโอโซน (O_3) หมายถึง สารมลพิษทางอากาศชนิดหนึ่งที่เกิดจากปฏิกิริยาของก๊าซออกไซด์ของไนโตรเจนและไฮโดรคาร์บอน ซึ่งเกิดจากโรงงานอุตสาหกรรม โรงไฟฟ้า และยานพาหนะต่าง ๆ โดยมีแสงแดดเป็นตัวเร่งปฏิกิริยา มีหน่วยเป็นส่วนในพันล้านส่วน (ppb, 1/1,000,000,000) (กรมควบคุมมลพิษ. 2544)

ก๊าซไนโตรเจนไดออกไซด์ (NO_2) หมายถึง ก๊าซที่เกิดจากการทำปฏิกิริยาทางเคมีของไนโตรเจนกับออกซิเจนในระหว่างการเผาไหม้เชื้อเพลิงประเภทต่าง ๆ ที่อุณหภูมิสูง ๆ มีหน่วยเป็นส่วนในพันล้านส่วน (ppb, 1/1,000,000,000) (จันทนี เกียรติโพธา. 2542 : 25)

ก๊าซคาร์บอนมอนอกไซด์ (CO) หมายถึง เป็นก๊าซไม่มีสี ไม่มีรส ไม่มีกลิ่น เบากว่าอากาศทั่วไปเล็กน้อย เป็นก๊าซที่เกิดจากการเผาไหม้ไม่สมบูรณ์ของคาร์บอนในเชื้อเพลิง มีหน่วยเป็นส่วนในล้านส่วน (ppm, 1/1,000,000) (กรมควบคุมมลพิษ. 2544)

ก๊าซซัลเฟอร์ไดออกไซด์ (SO_2) หมายถึง ก๊าซไม่มีสี ไม่ไวไฟ เกิดจากการเผาไหม้ของถ่านหินและน้ำมัน ซึ่งมีกำมะถัน (S) เป็นองค์ประกอบ มีหน่วยเป็นส่วนในพันล้านส่วน (ppb, 1/1,000,000,000) (กรมควบคุมมลพิษ. 2544)

ฝุ่นละออง (Particulate Matter) หมายถึง สารที่มีความหลากหลายทางด้านกายภาพ และองค์ประกอบอาจมีสภาพเป็นของแข็งหรือของเหลวก็ได้ ฝุ่นละอองที่มีอยู่ในบรรยากาศรอบ ๆ ตัวเรามีขนาดตั้งแต่ 0.002 ไมครอน ซึ่งเป็นกลุ่มของโมเลกุลที่มองด้วยตาเปล่าไม่เห็นต้องใช้กล้องจุลทรรศน์แบบอิเล็กตรอน ไปจนถึงฝุ่นขนาดใหญ่กว่า 500 ไมครอน ซึ่งเป็นฝุ่นทรายขนาดใหญ่มองเห็นได้ด้วยตาเปล่า โดยฝุ่นที่มองเห็นได้ด้วยตาเปล่ามีขนาด 50 ไมครอนขึ้นไป มีหน่วยเป็นมิลลิกรัมต่อลูกบาศก์เมตร (www.pcd.go.th)

ฝุ่นละอองขนาดเล็กกว่า 10 ไมครอน (Suspended Particulate Matter : SPM หรือ PM_{10}) หมายถึง ฝุ่นละอองที่มีขนาดเส้นผ่านศูนย์กลางต่ำกว่า 10 ไมครอน มีหน่วยเป็นไมโครกรัมต่อลูกบาศก์เมตร (www.pcd.go.th)

การระเหยของน้ำ (Evaporation) เป็นการวัดการระเหยซึ่งประกอบด้วยถังหรือถาดแบน ๆ ค่อนข้างลึกและกว้างใส่น้ำตั้งไว้กลางแจ้ง ค่าของการระเหยของน้ำวัดได้จากระดับน้ำในถังที่ลดลงไป มีหน่วยเป็นมิลลิเมตร (www.tmd.go.th)

เมฆ (Cloud) คือ ไอน้ำที่กลั่นตัวรวมกันเข้าเป็นกลุ่มก้อน อาจมีสภาพเป็นอนุภาคเล็ก ๆ ของน้ำหรือน้ำแข็ง หรือทั้งสองอย่างปนกันลอยอยู่ในอากาศ สามารถมองเห็นได้ด้วยตาเปล่า ในกลุ่มของผสมนี้อาจมีอนุภาคใหญ่ ๆ ของน้ำแข็งปนอยู่ด้วย หรืออาจมีอนุภาคที่ไม่มีน้ำหรืออนุภาคที่เป็นของแข็ง ตัวอย่างเช่น ก๊าซ ผงฝุ่น หรือควัน ฯลฯ ซึ่งเกิดจากโรงงานอุตสาหกรรมปนอยู่ด้วย มีหน่วยเป็นเดคะ (DECA) (www.tmd.go.th)

อุณหภูมิตุ้มเปียก (Wet-bulb temperature) เป็นอุณหภูมิที่อ่านได้จากเทอร์โมมิเตอร์ซึ่งกระเปาะของมันถูกหุ้มไว้ด้วยผ้าเปียกน้ำ ซึ่งติดตั้งอยู่ในที่ ๆ มีอากาศถ่ายเทตามธรรมชาติ และอยู่ในร่มเงาไม่ถูกรังสีจากดวงอาทิตย์โดยตรง มีหน่วยเป็นเซลเซียส (www.tmd.go.th)

ทัศนวิสัย (Visibility) คือ ระยะทางไกลที่สุดในทิศทางที่กำหนดไว้ ซึ่งผู้ตรวจสามารถมองเห็นวัตถุที่มีขนาดพอสมควรได้ด้วยตาเปล่าและบอกได้ว่าวัตถุนั้นเป็นอะไร โดยวัดค่าทัศนวิสัยออกมาเป็นระยะทาง มีหน่วยเป็นกิโลเมตร (www.tmd.go.th)

อุณหภูมิสูงสุดประจำวัน (Daily maximum temperature) คือ ค่าของอุณหภูมิสูงสุดที่ตรวจวัดได้ในช่วงเวลา 24 ชั่วโมงต่อเนื่องกัน มีหน่วยเป็นเซลเซียส (www.tmd.go.th)

อุณหภูมิต่ำสุดประจำวัน (Daily minimum temperature) คือ ค่าของอุณหภูมิต่ำสุดที่ตรวจวัดได้ในช่วงเวลา 24 ชั่วโมงต่อเนื่องกัน มีหน่วยเป็นเซลเซียส (www.tmd.go.th)

อุณหภูมิตุ้มแห้ง (Dry-bulb temperature) คือ อุณหภูมิที่อ่านได้จากเทอร์โมมิเตอร์ตุ้มแห้งหรือเทอร์โมมิเตอร์ธรรมดา ซึ่งติดตั้งอยู่ในที่ ๆ มีอากาศถ่ายเทตามธรรมชาติ และอยู่ในร่มเงาไม่ถูกรังสีจากดวงอาทิตย์โดยตรง มีหน่วยเป็นเซลเซียส (www.tmd.go.th)

ความชื้นสัมพัทธ์ (Relative humidity) คือ อัตราส่วนระหว่างมวลไอน้ำที่มีอยู่จริงในอากาศขณะนั้นกับมวลไอน้ำอิ่มตัวในอากาศที่อุณหภูมิเดียวกัน และความดันบรรยากาศเดียวกัน ความชื้นสัมพัทธ์นิยมคิดเป็นร้อยละ (www.tmd.go.th)

ปริมาณน้ำฝน (Rainfall amount) คือ ปริมาณน้ำฝนที่ตกสะสมในช่วงเวลาที่กำหนด วัดเป็นความสูง มีหน่วยเป็นมิลลิเมตร (มม.) (www.tmd.go.th)

ทิศทางลม (Wind direction) หมายถึง ทิศซึ่งลมพัดเข้า มีหน่วยเป็นมุมองศา (www.tmd.go.th)

ความเร็วลม (Wind speed) คือ ระยะทางที่มวลอากาศเคลื่อนที่ไป ในหนึ่งหน่วยเวลา มีหน่วยเป็นนอต (knot) (www.tmd.go.th)

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1. ได้ตัวแบบที่เหมาะสมในการพยากรณ์ดัชนีคุณภาพอากาศในเขตกรุงเทพมหานคร
2. สามารถใช้เป็นแนวทางในการควบคุมมลพิษทางอากาศในเขตกรุงเทพมหานคร

บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 ทฤษฎีที่ใช้ในการวิเคราะห์

1. วิธีการวิเคราะห์ปัจจัย (Factor Analysis)
2. วิธีการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ (Multiple linear Regression Analysis)

2.1.1 การวิเคราะห์ปัจจัย (Factor Analysis)

การวิเคราะห์ปัจจัยเป็นวิธีการวิเคราะห์ตัวแปรหลายตัว (Multivariate Statistical Technique) ที่มีวัตถุประสงค์ในการลดปริมาณตัวแปรให้มีจำนวนน้อยลง (variable reduction) โดยอาศัยโครงสร้างและแบบแผนของความสัมพันธ์ที่อยู่ในข้อมูล หรือระหว่างตัวแปร ในการที่ตัวแปรหรือข้อมูลต่าง ๆ มีความสัมพันธ์กัน สามารถสังเกตได้จากการจับกลุ่มกันของตัวแปร หรือค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรนั้นมีความสัมพันธ์สูง วิธีการใช้การวิเคราะห์ปัจจัยมีจุดมุ่งหมายที่สำคัญอยู่ 3 ประการ (สุชาติ ประสิทธิ์รัฐสินธุ์. 2540 : 331) คือ

1. เพื่อแสวงหาปัจจัยที่อธิบายถึงความสัมพันธ์ร่วมกันระหว่างตัวแปรต่าง ๆ โดยที่จำนวนของปัจจัยที่หาได้นั้นจะต้องมีจำนวนน้อยกว่าจำนวนตัวแปร เรียกว่า Exploratory factor analysis

2. เพื่อพิสูจน์ สนับสนุน ตรวจสอบสมมติฐานเกี่ยวกับความสัมพันธ์ของข้อมูล หรือตัวแปรว่ามีปัจจัยกี่ปัจจัย อะไรบ้าง

3. เพื่อนำเอาความสัมพันธ์ระหว่างปัจจัย และตัวแปรเหล่านี้ไปใช้สร้างคะแนนปัจจัย (factor score) คะแนนที่ได้นี้เปรียบเสมือนค่าของตัวแปรใหม่ที่ประกอบด้วยตัวแปรเดิมหลาย ๆ ตัว เรียกว่า ตัวแปรส่วนผสม (composite variable)

การวิเคราะห์ปัจจัยแบบ Exploratory (สุชาติ ประสิทธิ์รัฐสินธุ์. 2540 : 331-333) มีวิธีการวิเคราะห์ที่แบ่งออกเป็น 2 ประเภท คือ

1. การวิเคราะห์ปัจจัยแบบ Component Analysis เป็นการวิเคราะห์ที่ไม่มีการจำแนกปัจจัยร่วม (common factor) จากปัจจัยเฉพาะ (unique factor) ซึ่งมีวิธีการสกัดปัจจัย คือ วิธีองค์ประกอบหลัก (Principal component analysis)

2. การวิเคราะห์ปัจจัยแบบ Common factor analysis เป็นวิธีการวิเคราะห์ที่มีการจำแนกปัจจัยร่วม (common factor) จากปัจจัยเฉพาะ (unique factor) ซึ่งมีวิธีการสกัดปัจจัยหลายวิธี เช่น วิธีกำลังสองน้อยที่สุด (Generalized least squares) วิธีความเป็นไปได้สูงสุด (Maximum likelihood) วิธีกำลังสองน้อยที่สุดไม่ปรับน้ำหนัก (Unweight least squares) เป็นต้น

ในงานวิจัยนี้ใช้การวิเคราะห์ปัจจัยแบบ Component Analysis เนื่องจากการวิเคราะห์ปัจจัยแบบ Component Analysis สามารถอธิบายความสัมพันธ์ระหว่างตัวแปร โดยที่ความแปรปรวนทั้งหมดของแต่ละตัวแปรจะถูกแบ่งให้กับปัจจัยทั้งหมด ซึ่งเป็นการรวมกันระหว่างความแปรปรวนร่วมกันกับตัวแปรอื่น ๆ และความแปรปรวนเฉพาะ

ลักษณะของตัวแปรที่นำมาใช้ในการวิเคราะห์ปัจจัยนั้น ตัวแปรที่จะใช้จะต้องมีระดับการวัดเป็นระดับช่วง (Interval Scale) หรือระดับอัตราส่วน (Ratio Scale) ถ้าข้อมูลเป็นกลุ่ม Nominal Scale หรือ Categorical Scale จะต้องแปลงข้อมูลนั้นให้มีค่าเป็น 0 หรือ 1 (สุชาติ ประสิทธิ์รัฐสินธุ์. 2533 : 3)

ผลการวิเคราะห์ปัจจัยในบางครั้งอาจพบความซ้ำซ้อนของปัจจัย ในกรณีที่ตัวแปรใดตัวแปรหนึ่งมีน้ำหนักปัจจัย (factor loading) ใกล้เคียงกันมากกว่าหนึ่งปัจจัยจะทำให้ยากแก่การตีความหมาย หรือการระบุว่าตัวแปรตัวนั้นเป็นสมาชิกของปัจจัยใด ดังนั้นการเปลี่ยนตำแหน่งของแกนหรือมิติของปัจจัยบ้างเล็กน้อย จะทำให้ตัวแปรบางตัวซึ่งแต่เดิมเป็นสมาชิกของหลายปัจจัยกลายเป็นสมาชิกของปัจจัยใดปัจจัยหนึ่งอย่างเด่นชัดขึ้นมากกว่าเดิม ซึ่งการเปลี่ยนตำแหน่งดังกล่าวอาจทำได้โดยการหมุนแกนปัจจัย

สำหรับวิธีการหมุนแกนปัจจัยในงานวิจัยนี้ใช้การหมุนแกนแบบมุมฉาก (Orthogonal Rotation) โดยที่แกนของปัจจัยจะคงความเป็นมุมฉากซึ่งกันและกันตลอดเวลาที่หมุน การหมุนแกนปัจจัยแบบนี้จะทำให้ปัจจัยแต่ละปัจจัยเป็นอิสระต่อกัน (สุชาติ ประสิทธิ์รัฐสินธุ์. 2540 : 337-338) ส่วนวิธีการหมุนแกนใช้วิธีวาริแมกซ์ (Varimax) ซึ่งวิธีนี้เป็นวิธีที่นิยมใช้กันมากที่สุด และข้อดีของวิธีการนี้ คือ พยายามที่จะลดจำนวนตัวแปรที่มีน้ำหนักปัจจัยมากบนแต่ละปัจจัยให้เหลือน้อยที่สุด ซึ่งจะช่วยให้เราตีความหมายของปัจจัยได้ง่ายขึ้น

ในการวิเคราะห์ปัจจัยนั้น เป้าหมายของการวิเคราะห์ไม่เพียงแต่ต้องการหาปัจจัยเท่านั้น แต่ยังครอบคลุมไปถึงการสร้างคะแนนปัจจัยด้วย ซึ่งคะแนนปัจจัยเป็นคะแนนที่ได้จากค่าของตัวแปรต่าง ๆ หลายตัวที่รวมกลุ่มกันและอยู่ในปัจจัยเดียวกัน คะแนนปัจจัยจึงเปรียบเสมือนค่าของตัวแปรส่วนผสม โดยมีการให้ค่าน้ำหนักของตัวแปรแต่ละตัวตามน้ำหนักของตัวแปรนั้น วิธีการสร้างคะแนนปัจจัยในงานวิจัยนี้ใช้วิธีวิเคราะห์การถดถอย (Regression Method) วิธีนี้เป็นการสร้าง

คะแนนปัจจัยที่มีค่าความแปรปรวนเท่ากับความสัมพันธ์พหุ ระหว่างค่าคะแนนที่ประมาณได้และค่าของปัจจัยจริง ซึ่งจะเห็นได้จากเมตริกซ์ความแปรปรวนร่วมระหว่างปัจจัย และคะแนนปัจจัยที่ประมาณได้

2.1.1.1 ตัวแบบปัจจัย

1. ตัวแบบปัจจัยของประชากร

$$Z_j = a_{j1}F_1 + a_{j2}F_2 + \ldots + a_{jm}F_m + u_jY_j \qquad (2.1)$$

เมื่อ Z_j = คะแนนมาตรฐานของตัวแปรที่ j ของประชากร

F_p = ปัจจัยร่วม (common factor) ที่ p

Y_j = ปัจจัยเฉพาะ (unique factor) ของตัวแปรที่ j

a_{jp} = น้ำหนักปัจจัย (factor loading) ของตัวแปรที่ j บนปัจจัยร่วมที่ p

u_j = สัมประสิทธิ์ของปัจจัยเฉพาะของตัวแปรที่ j

$j = 1, 2, \ldots, n$

$p = 1, 2, \ldots, m$

ตัวแบบปัจจัยประชากรเขียนในรูปเมตริกซ์ได้ ดังนี้

$$\mathbf{Z} = \mathbf{AF} + \mathbf{UY} \qquad (2.2)$$

เมื่อ

$$\mathbf{Z} = \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_n \end{bmatrix}_{n\times 1} \qquad \mathbf{F} = \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_m \end{bmatrix}_{m\times 1} \qquad \mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}_{n\times 1}$$

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix}_{n\times m} \qquad \mathbf{U} = \begin{bmatrix} u_1 & 0 & \cdots & 0 \\ 0 & u_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & u_n \end{bmatrix}_{n\times n}$$

เมื่อ $\mathbf{Z}$ = เมตริกซ์ของคะแนนมาตรฐานของตัวแปร

$\mathbf{F}$ = เมตริกซ์ของปัจจัยร่วม (common factor)

$\mathbf{Y}$ = ปัจจัยเฉพาะ (unique factor)

$\mathbf{A}$ = เมตริกซ์ของน้ำหนักปัจจัย (factor loading)

$\mathbf{U}$ = เมตริกซ์ของสัมประสิทธิ์ของปัจจัยเฉพาะ

ข้อสมมติ (Assumptions) (Alvin C. Rencher. 2002 : 410)

1. $E(\mathbf{F}) = \mathbf{0},\ Var(\mathbf{F}) = \mathbf{1}$
2. $E(\mathbf{Y}) = \mathbf{0},\ Var(\mathbf{Y}) = \mathbf{\Psi}$ เมื่อ $\mathbf{\Psi}$ คือ diagonal matrix
3. $Cov(F_p, F_q) = Cov(F_p, Y_k) = Cov(Y_j, Y_k) = 0 \ ; p \neq q$ และ $j \neq k$
4. $Corr(F_i, F_j) = Corr(e_i, e_j) = Corr(F_i, e_i) = 0 \ ; i \neq j$

โดยที่ $j, k = 1, 2, \ldots, n$

$p, q = 1, 2, \ldots, m$

องค์ประกอบความแปรปรวนของตัวแปรประชากร อาจเขียนได้ดังนี้

$$\Sigma = \mathbf{AA}' + \mathbf{U}^2 \tag{2.3}$$

โดย

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{bmatrix}$$

เมื่อ σ_j^2 = ความแปรปรวนของตัวแปรที่ j

σ_{jk} = ความแปรปรวนร่วมระหว่างตัวแปรที่ j กับตัวแปรที่ k เมื่อ $j \neq k$

$j, k = 1, 2, \ldots, n$

2. ตัวแบบปัจจัยของตัวอย่าง

$$\hat{Z}_j = \hat{a}_{j1}F_1 + \hat{a}_{j2}F_2 + \ldots + \hat{a}_{jm}F_m + \hat{u}_j Y_j \tag{2.4}$$

เมื่อ $\hat{Z}_j$ = คะแนนมาตรฐานของตัวแปรที่ j ของตัวอย่าง

F_p = ปัจจัยร่วม (common factor) ที่ p

Y_j = ปัจจัยเฉพาะ (unique factor) ของตัวแปรที่ j

$\hat{a}_{jp}$ = ค่าประมาณของน้ำหนักปัจจัย (factor loading) ของตัวแปรที่ j บนปัจจัยร่วมที่ p

$\hat{u}_j$ = ค่าประมาณสัมประสิทธิ์ของปัจจัยเฉพาะ ของตัวแปรที่ j

$j = 1, 2, \ldots, n$

$p = 1, 2, \ldots, m$

ให้ S_j^2 = ความแปรปรวนของตัวแปรที่ j ของตัวอย่าง

เนื่องจากตัวแปรแต่ละตัวอยู่ในรูปคะแนนมาตรฐาน ความแปรปรวนของแต่ละตัวจึงเท่ากับ 1 และจากข้อกำหนดที่ว่าปัจจัยร่วมแต่ละตัวเป็นอิสระต่อกัน และปัจจัยร่วมกับปัจจัยเฉพาะเป็นอิสระต่อกัน ดังนั้นจะได้องค์ประกอบของความแปรปรวนของแต่ละตัวแปรของตัวอย่าง (Alvin C. Rencher. 2002 : 413) ดังนี้

$$S_j^2 = 1 = \hat{a}_{j1}^2 + \hat{a}_{j2}^2 + \ldots + \hat{a}_{jm}^2 + \hat{u}_j^2 \qquad (2.5)$$

ซึ่งในส่วนแรก $\hat{a}_{j1}^2 + \hat{a}_{j2}^2 + \ldots + \hat{a}_{jm}^2$ จะถูกเรียกว่าเป็นอัตราส่วนร่วม (communality) ของตัวอย่างที่ j

ให้ h_j^2 = อัตราส่วนร่วมของตัวแปรที่ j

$$h_j^2 = \hat{a}_{j1}^2 + \hat{a}_{j2}^2 + \ldots + \hat{a}_{jm}^2 \qquad (2.6)$$

ในส่วนที่สองของ u_j^2 จะถูกเรียกว่าเป็นความแปรปรวนเฉพาะ (specific variance) ของตัวแปรที่ j

ความแปรปรวนของตัวแปรทั้งหมดที่อธิบายได้ด้วยปัจจัยร่วมแต่ละตัวสามารถคำนวณได้ดังนี้ คือ

ให้ V_p = ความแปรปรวนของตัวแปรทั้งหมดที่อธิบายได้ด้วยปัจจัยร่วมที่ p

$$= \sum_{j=1}^{n} a_{jp}$$

ความแปรปรวนของตัวแปรรวมทั้งหมดที่อธิบายได้ด้วยปัจจัยร่วม m ตัว สามารถคำนวณได้ดังนี้ คือ

ให้ V = ผลรวมของความแปรปรวนที่อธิบายได้ด้วยปัจจัยร่วม m ตัว

$$= \sum_{p=1}^{m} V_p$$

หลักการของการวิเคราะห์ปัจจัย คือ การพยายามหาปัจจัยร่วมที่สามารถอธิบายความแปรปรวนรวมของตัวแปรทั้งหมดให้ได้มากที่สุด และอธิบายความสัมพันธ์ระหว่างตัวแปรได้เป็นอย่างดี ซึ่งปัจจัยร่วมที่ได้จะต้องน้อยกว่าจำนวนตัวแปรเดิม

ในการประมาณค่าน้ำหนักปัจจัย และการหาปัจจัยร่วมที่สามารถอธิบายความแปรปรวนรวมของตัวแปรทั้งหมดให้ได้มากที่สุด และอธิบายความสัมพันธ์ระหว่างตัวแปรได้เป็นอย่างดีนั้น มีวิธีการที่เรียกว่า "การสกัดปัจจัย" ซึ่งการสกัดปัจจัยมีหลายวิธีด้วยกัน แต่วิธีที่นิยมใช้และเป็นที่ยอมรับโดยทั่วไปใช้วิธีองค์ประกอบหลัก เป็นวิธีที่มีวัตถุประสงค์ที่จะนำรายละเอียดของตัวแปรที่มีจำนวนตัวแปรมาก ๆ มาไว้ในปัจจัยที่มีเพียงไม่กี่ปัจจัย โดยพิจารณาจากรายละเอียดทั้งหมดจากแต่ละตัวแปร ดังนั้นผู้วิจัยจึงใช้การสกัดปัจจัยโดยวิธีองค์ประกอบหลัก

2.1.1.2 วิธีพิจารณาความเหมาะสมของข้อมูลที่ใช้ในการวิเคราะห์ปัจจัย

1) หาค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปรอิสระ ซึ่งเป็นค่าที่ใช้ดูความสัมพันธ์ของตัวแปรอิสระ ถ้าค่า $|r|$ มีค่ามากกว่า 0.5 แสดงว่าตัวแปรอิสระมีความสัมพันธ์กันสูง

2) หาค่า KMO (Kaiser-Meyer-Olkin) ซึ่งเป็นค่าที่ใช้วัดความเหมาะสมของข้อมูลที่จะนำมาวิเคราะห์ปัจจัยโดยมีสูตร

$$KMO = \frac{\sum r_j^2}{\sum r_j^2 + \sum (\text{partial correlation})^2} \qquad (2.7)$$

เมื่อ r_j = ค่าสัมประสิทธิ์สหสัมพันธ์

โดยที่ $0 \leq KMO \leq 1$

ถ้าค่า KMO มีค่าน้อยกว่า 0.5 แสดงว่าวิธีวิเคราะห์ปัจจัยไม่เหมาะสมกับข้อมูลที่มีอยู่ แต่ถ้าค่า KMO มีค่ามากกว่าหรือเท่ากับ 0.5 แสดงว่าวิธีวิเคราะห์ปัจจัยเหมาะสมกับข้อมูลที่มีอยู่ (Johnson J.D. 1992 : 131)

2.1.1.3 วิธีวิเคราะห์องค์ประกอบหลัก (Principal component analysis)

วิธีวิเคราะห์องค์ประกอบหลักอาศัยหลักความสัมพันธ์เชิงเส้นระหว่างตัวแปรที่ใช้เป็นข้อมูล องค์ประกอบหลักแรกสามารถอธิบายความแปรปรวนของตัวแปรได้มากที่สุด จากนั้น

หาองค์ประกอบหลักที่สอง ที่สามารถอธิบายความแปรปรวนของตัวแปรได้มากเป็นอันดับสอง โดยที่ไม่มีความสัมพันธ์กับองค์ประกอบหลักแรก ทำเช่นนี้เรื่อยไปจนได้องค์ประกอบหลัก ที่สามารถอธิบายความแปรปรวนของทุกตัวแปรได้ครบถ้วน ซึ่งองค์ประกอบหลักหลัง ๆ จะอธิบายความแปรปรวนได้น้อยลงตามลำดับ และทุกองค์ประกอบหลักไม่สัมพันธ์กัน สำหรับการประมาณค่าน้ำหนักปัจจัยของวิธีองค์ประกอบหลัก มีรายละเอียดดังนี้

สมมติ $\mathbf{X}' = (X_1, X_2, \ldots, X_r)$ เป็น r-dimentional random vector ที่มีค่าเฉลี่ย $\boldsymbol{\mu}$ และ covariance matrix $\sum$ ต้องการหาตัวแปรชุดใหม่ $(F_1, F_2, \ldots, F_p)$ ที่ไม่มีความสัมพันธ์ต่อกัน และ $\mathrm{Var}(F_1) \geq \mathrm{Var}(F_2) \geq \ldots \geq \mathrm{Var}(F_p)$ โดยแต่ละ F_j เป็นผลบวกเชิงเส้น (linear combination) ของ X

$$F_j = \gamma_{1j}X_1 + \gamma_{2j}X_2 + \ldots + \gamma_{rj}X_r = \gamma_j'\mathbf{X} \tag{2.8}$$

เมื่อ $\gamma_j' = (\gamma_{1j}, \gamma_{2j}, \ldots, \gamma_{rj})$ เป็นเวกเตอร์ของค่าคงที่

ดังนั้น องค์ประกอบแรก (F_1) ได้จากการเลือก γ_1 ที่ทำให้ F_1 มีความแปรปรวนมากที่สุดในบรรดาผลบวกเชิงเส้น (linear combination) ทั้งหมดของ X ภายใต้ข้อบังคับว่า $\gamma_1'\gamma_1 = 1$ (normalization constraint) นั่นคือ หา γ_1 ที่

$$\max_{\gamma_1} \mathrm{Var}(F_1) = \max_{\gamma_1} \gamma_1'\sum\gamma_1$$

ภายใต้เงื่อนไข $\gamma_1'\gamma_1 = 1$

โดยการใช้วิธีของ Lagrange multipliers จะต้องสอดคล้องกับ p สมการเชิงเส้นพร้อมกัน

$$(\sum - \lambda_1\mathbf{I})\gamma_1 = \mathbf{0} \tag{2.9}$$

เมื่อ λ_1 เป็น Lagrange multipliers

ถ้าจะให้คำตอบสำหรับ γ_1 ไม่ใช่ null vector $(\sum - \lambda_1\mathbf{I})$ ต้องเป็น singular matrix ดังนั้น ต้องเลือก λ_1 ที่ทำให้

สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

$$|\Sigma - \lambda_1 \mathbf{I}| = 0 \qquad (2.10)$$

ซึ่งแสดงว่า λ_1 เป็น Eigen Value ตัวหนึ่งของ Σ

$$\mathrm{Var}(\gamma_j' x) = \gamma_1' \Sigma \gamma_1 = \lambda_1$$

λ_1 ต้องเป็น Eigen Value ตัวที่ใหญ่ที่สุดของ Σ และ γ_1 ต้องเป็น Eigen Vector ของ Σ กับ Eigen Value ที่ใหญ่ที่สุด

องค์ประกอบแรก คือ $F_1 = \gamma_1' \mathbf{X}$

องค์ประกอบตัวที่สอง คือ $F_2 = \gamma_2' \mathbf{X}$

ซึ่งได้จากการเลือก γ_2 ที่

$$\max_{\gamma_2} \mathrm{Var}(F_2) = \max_{\gamma_2} \gamma_2' \Sigma \gamma_2$$

ภายใต้เงื่อนไข $\gamma_2' \gamma_2 = 1$ และ $\gamma_2' \gamma_1 = 0$

ด้วยวิธีการของ Lagrange multipliers เช่นเดิม γ_2 ที่ $\mathrm{Var}(F_2)$ มากที่สุด คือ γ_2 ที่ทำให้

$$(\Sigma - \lambda_2 \mathbf{I})\gamma_2 = \mathbf{0} \qquad (2.11)$$

เมื่อ λ_2 เป็น Lagrange multipliers

λ_2 ต้องเป็น Eigen Value ตัวที่ใหญ่เป็นลำดับที่สองรองจาก λ_1 ของ Σ และ γ_2 ต้องเป็น Eigen Vector ของ Σ ที่ตรงกับ Eigen Value ที่ใหญ่เป็นลำดับที่สอง

ดังนั้น ถ้า $\lambda_1 > \lambda_2 > \ldots > \lambda_p \geq 0$ เป็น Eigen Values ของ Σ ที่มีค่าต่างกัน และ $\gamma_1, \gamma_2, \ldots, \gamma_p$ เป็น Eigen Vectors ที่ตรงกับ Eigen Values แต่ละตัว ในทำนองเดียวกัน γ_j สำหรับองค์ประกอบตัวที่ j ก็คือ Eigen Vector ที่ตรงกับ Eigen Value ตัวที่ใหญ่เป็นลำดับที่ j (λ_j) ให้เป็น P เป็น $p \times p$ Orthogonal matrix ของ Eigen Vector (Alvin C. Rencher. 2002 : 453)

$$P = [\gamma_1, \gamma_2, \ldots, \gamma_p]$$

$$\mathbf{F} = \mathbf{P'X}$$

87863

$$\mathrm{Var}(\mathbf{F}) = \mathbf{P}'\Sigma\mathbf{P} = \mathbf{\Lambda}$$

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_p \end{bmatrix}$$

$$\begin{aligned} \mathrm{trace}(\mathbf{\Lambda}) &= \mathrm{trace}(\mathbf{P}'\Sigma\mathbf{P}) \\ &= \mathrm{trace}(\Sigma\mathbf{P}'\mathbf{P}) \qquad \because \mathbf{P}'\mathbf{P} = \mathbf{I} \\ &= \mathrm{trace}(\Sigma) \end{aligned}$$

$$\therefore \sum_{j=1}^{p}\mathrm{Var}(F_j) = \sum_{j=1}^{p}\lambda_j = \sum_{i=1}^{r}\sum_{j=1}^{p}\mathrm{Var}(\gamma_{ij}X_i) \qquad (2.12)$$

แสดงว่า ความแปรปรวนทั้งหมดของตัวแปรเดิมจะเท่ากับผลรวมของความแปรปรวนขององค์ประกอบทุกตัว

ดังนั้น สัดส่วนความแปรปรวนรวมทั้งหมดที่อธิบายได้ด้วยองค์ประกอบตัวที่ j $= \dfrac{\lambda_j}{\sum_{j=1}^{p}\lambda_j}$

ค่าประมาณน้ำหนักปัจจัย คือ $\dfrac{\hat{\gamma}_{ij}\sqrt{\hat{\lambda}_j}}{\sqrt{\sum_{i=1}^{r}S_{ii}}}$ (2.13)

เมื่อ S_{ii} เป็นความแปรปรวนของ X_i

2.1.1.4 การหมุนแกนปัจจัย

ผลจากการสกัดปัจจัยจะพบว่า ถ้าตัวแปรตัวนั้นมีน้ำหนักปัจจัยใกล้เคียงกันบนปัจจัยมากกว่าหนึ่งปัจจัย ก็จะทำให้ยากต่อการอ่านและการแปลความหมาย หรือยากต่อการระบุว่าตัวแปรนั้นเป็นสมาชิกของปัจจัยใด ซึ่งปัจจัยที่ได้จะมีความหมายชัดเจนก็ต่อเมื่อประกอบด้วยตัวแปรที่มีความสัมพันธ์กันมากที่สุด และมีน้ำหนักมากต่อปัจจัยใดปัจจัยหนึ่งเป็นพิเศษ วิธีการที่จะทำให้ปัจจัยที่ได้มีความหมายชัดเจน คือ การหมุนแกนปัจจัยซึ่งทำให้ตัวแปรบางตัวซึ่งแต่เดิมเป็นสมาชิกของหลายปัจจัย กลายเป็นสมาชิกของปัจจัยร่วมใดปัจจัยร่วมหนึ่งอย่างเด่นชัดมากขึ้น การที่ตัวแปรจะเป็นสมาชิกของปัจจัยร่วมใด จะต้องพิจารณาจากน้ำหนักปัจจัย (factor loading) ของตัวแปรนั้น

การศึกษาวิจัยในครั้งนี้ ผู้วิจัยใช้วิธีหมุนแกนปัจจัยแบบมุมฉากด้วยวิธีวาริแมกซ์ (Varimax) ซึ่งวิธีการหมุนแบบนี้จะทำให้ได้ปัจจัยที่มีโครงสร้างง่าย ซึ่งจะทำให้การแปลความหมายของปัจจัยชัดเจนยิ่งขึ้น (สุชาติ ประสิทธิ์รัฐสินธุ์. 2540 : 355)

การหมุนแกนปัจจัยด้วยวิธีวาริแมกซ์ เป็นการหมุนแกนปัจจัยแบบมุมฉาก โดยมีหลักการว่า น้ำหนักปัจจัยของตัวแปรหนึ่ง ๆ บนปัจจัยใดปัจจัยหนึ่ง เมื่อทำการหมุนแกนปัจจัยที่ทำมุมเท่ากับ φ จะต้องทำให้ความแปรปรวนรวมที่อธิบายได้ด้วยปัจจัยทั้งหมด (V) มีค่าสูงสุด

เมื่อ $$V = n\sum_{p=1}^{m}\sum_{j=1}^{n}\left(\frac{b_{jp}}{h_j}\right)^4 - \sum_{p=1}^{m}\left(\sum_{j=1}^{n}\frac{b_{jp}^2}{h_j^2}\right)^2$$

โดยที่ b_{jp} = น้ำหนักปัจจัยหลังการหมุนแกนปัจจัยของตัวแปรที่ j บนปัจจัยร่วมที่ p

h_j^2 = อัตราส่วนร่วมของตัวแปรที่ j

h_j = รากที่สองของอัตราส่วนร่วมของตัวแปรที่ j

2.1.1.5 การหาค่าประมาณน้ำหนักปัจจัยหลังการหมุนแกน

การหาค่าประมาณน้ำหนักปัจจัยหลังการหมุนแกนปัจจัย $\left(b_{jp}\right)$ พิจารณารูปแบบเมตริกซ์

$$\mathbf{B} = \mathbf{AT} \qquad (2.14)$$

เมื่อ $\mathbf{B}$ = เมตริกซ์น้ำหนักปัจจัยหลังการหมุนแกนปัจจัย ขนาด $n \times 2$

$\mathbf{A}$ = เมตริกซ์น้ำหนักปัจจัยก่อนการหมุนแกนปัจจัย ขนาด $n \times 2$

$\mathbf{T}$ = เมตริกซ์น้ำหนักปัจจัยที่ใช้ในการหมุนแกนปัจจัย ขนาด 2×2

โดยที่

$$\mathbf{B} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ \vdots & \vdots \\ b_{n1} & b_{n2} \end{bmatrix} \qquad \mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ \vdots & \vdots \\ a_{n1} & a_{n2} \end{bmatrix} \qquad \mathbf{T} = \begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix}$$

การหมุนแกนปัจจัย จะทำการหมุนแกนปัจจัยทีละคู่จนครบทุกคู่ เมื่อมีปัจจัยร่วมทั้งหมด m ปัจจัย ดังนั้นจะต้องหมุนแกนปัจจัยทั้งหมด $\frac{m(m-1)}{2}$ ครั้ง

น้ำหนักปัจจัยหลังการหมุนแกนปัจจัยด้วยวิธีวาริแมกซ์ จะทำให้สามารถแปลความหมายของปัจจัยร่วมที่ได้ชัดเจนขึ้น

2.1.1.6 การสร้างคะแนนปัจจัย (Factor Scores)

ในการวิเคราะห์ปัจจัย เป้าหมายของการวิเคราะห์ไม่เพียงแต่ต้องการหาโครงสร้างหรือแบบแผนของปัจจัยเท่านั้น แต่ยังครอบคลุมไปถึงการสร้างคะแนนปัจจัยด้วย คะแนนปัจจัยสามารถนำไปทำการศึกษาต่อไปได้ สำหรับการสร้างคะแนนปัจจัยนั้น สุชาติ ประสิทธิ์รัฐสินธุ์ (2540 : 358) กล่าวว่า การสร้างคะแนนปัจจัยด้วยวิธีวิเคราะห์การถดถอย (Regression Method) จะทำให้ได้คะแนนปัจจัยที่มีความสัมพันธ์กับปัจจัยสูงกว่าวิธีอื่น ๆ ดังนั้น หลังจากได้ทำการวิเคราะห์ปัจจัยแล้ว ผู้วิจัยจึงสร้างคะแนนปัจจัยด้วยวิธีวิเคราะห์การถดถอย

หลังจากทำการเลือกวิธีการสกัดปัจจัยที่เหมาะสมแล้ว ผู้วิจัยจะทำการสร้างคะแนนปัจจัยด้วยวิธีวิเคราะห์การถดถอย

สมการถดถอยสำหรับการประมาณปัจจัยร่วม (F_p) แต่ละตัว โดยตัวแปร (Z_j) ทั้งหมด n ตัว (Alvin C. Rencher. 2002 : 439) สามารถเขียนได้ดังนี้

$$\hat{F}_p = \beta_{p1}Z_1 + \beta_{p2}Z_2 + ... + \beta_{pn}Z_n \qquad (2.15)$$

เมื่อ $\hat{F}_p$ = ค่าประมาณของคะแนนปัจจัยร่วมที่ p

Z_j = คะแนนมาตรฐานของตัวแปรที่ j

β_{pj} = สัมประสิทธิ์การถดถอยเชิงเส้นของปัจจัยร่วมที่ p บนตัวแปรที่ j

$p = 1, 2, ..., m$

$j = 1, 2, ..., n$

รูปแบบการถดถอยเชิงเส้นสำหรับการประมาณ จะเขียนในรูปเมตริกซ์ได้ดังนี้

$$\mathbf{F} = \mathbf{BZ} + \mathbf{E} \qquad (2.16)$$

โดยที่

$$\mathbf{F} = \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_m \end{bmatrix}_{m\times 1} \qquad \mathbf{Z} = \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_n \end{bmatrix}_{n\times 1} \qquad E = \begin{bmatrix} E_1 \\ E_2 \\ \vdots \\ E_m \end{bmatrix}_{m\times 1}$$

$$\mathbf{B} = \begin{bmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1n} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{m1} & \beta_{m2} & \cdots & \beta_{mn} \end{bmatrix}_{m\times n}$$

เมื่อนำตัวแปร **Z** และ **F** มาเขียนรวมกันได้เมตริกซ์ ดังนี้

$$\begin{bmatrix} \mathbf{Z} \\ \mathbf{F} \end{bmatrix}$$

เมื่อนำ Transpose ของ $\begin{bmatrix} \mathbf{Z} \\ \mathbf{F} \end{bmatrix}$ มาคูณ จะได้ดังนี้

$$\begin{bmatrix} \mathbf{Z} \\ \mathbf{F} \end{bmatrix}\begin{bmatrix} \mathbf{Z}' & \mathbf{F}' \end{bmatrix} = \begin{bmatrix} \mathbf{ZZ}' & \mathbf{ZF}' \\ \mathbf{FZ}' & \mathbf{FF}' \end{bmatrix} \tag{2.17}$$

เมื่อ $\mathbf{ZZ}'$ = เมตริกซ์สัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรของตัวอย่าง (R)

$\mathbf{ZF}'$ = เมตริกซ์สัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรมาตรฐานกับปัจจัยร่วม

$\mathbf{FZ}'$ = เมตริกซ์สัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรมาตรฐานกับปัจจัยร่วม

$\mathbf{FF}'$ = เมตริกซ์สัมประสิทธิ์สหสัมพันธ์ระหว่างปัจจัยร่วม = $\mathrm{I}(\text{identity matrix})$

$r_{z_j F_p}$ = สัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรมาตรฐานที่ j กับปัจจัยร่วมที่ p

$$r_{z_j F_p} = a_{j1} r_{F_p F_1} + a_{j2} r_{F_p F_2} + \ldots + a_{jm} r_{F_p F_m} \tag{2.18}$$

จากข้อกำหนดที่ว่าปัจจัยร่วมแต่ละตัวเป็นอิสระต่อกัน ดังนั้น $r_{F_p F_m} = 0$

จะได้ว่า $r_{z_j F_p} = a_{jp}$

เขียนในรูปเมตริกซ์ $\mathbf{ZF'} = \mathbf{A}$ (2.19)

เมื่อ $\mathbf{A}$ = เมตริกซ์น้ำหนักปัจจัยร่วม

จาก (2.16) ถ้า $\mathbf{E}$ และ $\mathbf{Z}$ เป็นอิสะต่อกัน เมื่อนำ $\mathbf{Z'}$ คูณทั้งสองข้างจะได้

$$\mathbf{FZ'} = \mathbf{BZZ'} \tag{2.20}$$

นำ $(\mathbf{ZZ'})^{-1}$ คูณทั้งสองข้างจะได้

$$\mathbf{FZ'}(\mathbf{ZZ'})^{-1} = \mathbf{B} \tag{2.21}$$

แทนค่า $\mathbf{ZZ'}$ ด้วย $\mathbf{R}$ และแทน $\mathbf{FZ'}$ ด้วย $\mathbf{A'}$ ได้ดังนี้

$$\mathbf{B} = \mathbf{A'R^{-1}} \tag{2.22}$$

แทนค่า $\mathbf{B}$ จาก (2.22) ลงใน (2.15) จะได้ค่าประมาณคะแนนปัจจัยร่วม m ปัจจัย ดังนี้

$$\hat{\mathbf{F}} = \mathbf{A'R^{-1}Z} \tag{2.23}$$

2.1.2 การวิเคราะห์การถดถอยเชิงเส้นพหุคูณ (Multiple Linear Regression Analysis)

การวิเคราะห์การถดถอยเชิงเส้นพหุคูณ เป็นวิธีการวิเคราะห์ทางสถิติที่เกี่ยวข้องกับการสร้างตัวแบบทางคณิตศาสตร์ เพื่อแสดงความสัมพันธ์ระหว่างตัวแปรมากกว่า 2 ตัวขึ้นไป โดยมีตัวแปรตัวหนึ่งที่เรียกว่า ตัวแปรตาม และตัวแปรตัวอื่น ๆ เรียกว่า ตัวแปรอิสระ ซึ่งการสร้างตัวแบบความสัมพันธ์ดังกล่าว เรียกว่า สมการการถดถอย ซึ่งมีวัตถุประสงค์เพื่อการพยากรณ์ โดยตัวแบบการถดถอยเชิงเส้นพหุคูณ จะเขียนได้เป็น

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + \varepsilon_i \tag{2.24}$$

เมื่อ $i = 1, 2, \ldots, n$ และ $j = 1, 2, \ldots, k$

โดยที่ Y_i คือ ค่าสังเกตที่ i ของตัวแปรตามของประชากร

X_{ji} คือ ค่าสังเกตที่ i ของตัวแปรอิสระที่ j ของประชากร

β_0 คือ จุดตัดแกน Y เมื่อกำหนดให้ $X_{1i} = X_{2i} = \ldots = X_{ki} = 0$

β_j คือ ค่าสัมประสิทธิ์การถดถอยบางส่วน (Partial Regression Coefficient) ซึ่งเป็นค่าที่แสดงถึงการเปลี่ยนแปลงของตัวแปรตาม Y เมื่อตัวแปรอิสระ X_{ji} เปลี่ยนไป 1 หน่วย โดยกำหนดให้ตัวแปรอิสระอื่น ๆ คงที่

ε_i คือ ค่าความคลาดเคลื่อนที่ i

2.1.2.1 การประมาณค่าพารามิเตอร์ในตัวแบบการถดถอย

จะประมาณค่า Y_i ด้วย $\hat{Y}_i$ ซึ่งหาได้จากสมการ

$$\hat{Y}_i = b_0 + b_1X_{1i} + b_2X_{2i} + \ldots + b_kX_{ki} \qquad (2.25)$$

เมื่อ $i = 1, 2, \ldots, n$ และ $j = 1, 2, \ldots, k$

และสมการนี้ เรียกว่า สมการถดถอยเชิงเส้นพหุคูณของตัวอย่างสุ่ม โดยที่ $\hat{Y}_i$ เป็นค่าประมาณของ Y_i และ $b_0, b_1, b_2, \ldots, b_k$ เป็นค่าประมาณของ $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$ ตามลำดับ

การหาตัวประมาณ $b_0, b_1, b_2, \ldots, b_k$ ของพารามิเตอร์ $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$ จะหาได้โดยวิธีกำลังสองน้อยที่สุด (Least Square Method) ซึ่งเป็นวิธีที่ทำให้ผลบวกของค่าคลาดเคลื่อนกำลังสอง $\sum_{i=1}^{n} e_i^2$ มีค่าต่ำสุด หรือ $\sum\left(Y_i - \hat{Y}_i\right)^2$ มีค่าน้อยที่สุด ซึ่งตัวประมาณค่า $b_0, b_1, b_2, \ldots, b_k$ ที่หาได้ด้วยวิธีนี้จะเป็นตัวประมาณค่าที่ดีที่สุดของพารามิเตอร์ $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$ นั่นคือ เป็นตัวประมาณค่าที่มีคุณสมบัติเป็นตัวประมาณค่าที่ไม่เอนเอียงและมีความแปรปรวนต่ำสุด (Best Linear Unbiased Estimator : BLUE) โดยค่าเศษเหลือ (Residuals) เขียนได้ในรูป

$$e_i = Y_i - \hat{Y}_i$$

$$e_i = Y_i - \left(b_0 + b_1X_{1i} + \ldots + b_kX_{ki}\right)$$

$$\text{ให้ } \varphi = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n}\left(Y_i - b_0 - b_1X_{1i} - \ldots - b_kX_{ki}\right)^2$$

ซึ่งเป็นค่าผลบวกกำลังสองของค่าที่เบี่ยงเบนไปจากสมการถดถอยที่แท้จริง โดยตัวประมาณค่า $b_0, b_1, b_2, \ldots, b_k$ นี้จะทำให้ φ มีค่าต่ำสุด ซึ่งจะหาได้ดังนี้

$$\frac{\partial \sum e_i^2}{\partial b_0} = -2\sum\left(Y_i - b_0 - b_1X_{1i} - \ldots - b_kX_{ki}\right) = 0$$

$$\frac{\partial \sum e_i^2}{\partial b_1} = -2\sum X_{1i}\left(Y_i - b_0 - b_1X_{1i} - \ldots - b_kX_{ki}\right) = 0$$

$$\vdots$$

$$\frac{\partial \sum e_i^2}{\partial b_k} = -2\sum X_{ki}\left(Y_i - b_0 - b_1X_{1i} - \ldots - b_kX_{ki}\right) = 0$$

นั่นคือ ถ้ามีตัวแปรอิสระ k ตัว ในสมการ จะได้สมการปกติ (Normal Equation) จำนวน k+1 สมการ ดังต่อไปนี้

$$nb_0 + b_1\sum_{i=1}^{n}X_{1i} + b_2\sum_{i=1}^{n}X_{2i} + \ldots + b_k\sum_{i=1}^{n}X_{ki} = \sum_{i=1}^{n}Y_i$$

$$b_0\sum_{i=1}^{n}X_{1i} + b_1\sum_{i=1}^{n}X_{1i}^2 + b_2\sum_{i=1}^{n}X_{1i}X_{2i} + \ldots + b_k\sum_{i=1}^{n}X_{1i}X_{ki} = \sum_{i=1}^{n}X_{1i}Y_i$$

$$b_0\sum_{i=1}^{n}X_{2i} + b_1\sum_{i=1}^{n}X_{2i}X_{1i} + b_2\sum_{i=1}^{n}X_{2i}^2 + \ldots + b_k\sum_{i=1}^{n}X_{2i}X_{ki} = \sum_{i=1}^{n}X_{2i}Y_i$$

$$\vdots$$

$$b_0\sum_{i=1}^{n}X_{ki} + b_1\sum_{i=1}^{n}X_{ki}X_{1i} + b_2\sum_{i=1}^{n}X_{ki}X_{2i} + \ldots + b_k\sum_{i=1}^{n}X_{ki}^2 = \sum_{i=1}^{n}X_{ki}Y_i$$

โดยทั่วไปในการวิเคราะห์การถดถอยเชิงเส้นพหุคูณจะใช้เมตริกซ์เป็นเครื่องมือ ซึ่งทำให้การวิเคราะห์สะดวกมากขึ้น เมื่อกำหนด

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \qquad \mathbf{X} = \begin{bmatrix} 1 & X_{11} & X_{21} & \cdots & X_{k1} \\ 1 & X_{12} & X_{22} & \cdots & X_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1n} & X_{2n} & \cdots & X_{kn} \end{bmatrix}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \qquad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \qquad \mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_k \end{bmatrix}$$

โดยที่ $\mathbf{Y}$ เป็นเวกเตอร์ขนาด $n\times1$ ของตัวแปรสุ่ม หรือเวกเตอร์ขนาด n ของตัวแปรสุ่ม

$\mathbf{X}$ เป็นเวกเตอร์ขนาด $n\times(k+1)$ ของตัวแปรอิสระ

$\boldsymbol{\beta}$ เป็นเวกเตอร์ขนาด $(k+1)\times1$ ของพารามิเตอร์

$\boldsymbol{\varepsilon}$ เป็นเวกเตอร์ขนาด $n\times1$ ของตัวแปรสุ่มค่าคลาดเคลื่อน

$\mathbf{b}$ เป็นเวกเตอร์ขนาด $(k+1)\times1$ ของตัวประมาณพารามิเตอร์

ดังนั้น ด้วยการถดถอยเชิงเส้นพหุคูณ สามารถเขียนในรูปเมตริกซ์ ได้ดังนี้

$$\mathbf{Y} = \mathbf{X\beta} + \boldsymbol{\varepsilon} \tag{2.26}$$

และข้อกำหนดของ $\boldsymbol{\varepsilon}$ สามารถเขียนได้ดังนี้

$$\boldsymbol{\varepsilon} \sim N_n\left(\mathbf{0}, \sigma^2\mathbf{I}\right)$$

ซึ่งหมายความว่า $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ เป็นอิสระต่อกัน และต่างมีการแจกแจงปกติที่มีค่าเฉลี่ย 0 และความแปรปรวน σ^2 และสมการปกติในเทอมของเมตริกซ์ จะเขียนได้ดังนี้

$$\mathbf{X'Xb} = \mathbf{X'Y} \tag{2.27}$$

การแก้สมการหาค่า $\mathbf{b}$ จะสมมติว่าหาเมตริกซ์ผกผันของเมตริกซ์ $\mathbf{X'X}$ ได้ ซึ่งจะเป็นจริงโดยทั่วไป ในทางปฏิบัติ เพราะฉะนั้น ตัวประมาณแบบกำลังสองน้อยที่สุด คือ

$$\mathbf{b} = \left(\mathbf{X'X}\right)^{-1}\mathbf{X'Y} \tag{2.28}$$

2.1.2.2 ข้อสมมติของความคลาดเคลื่อนมีดังนี้

1. ε_i มีการแจกแจงแบบปกติ (Normal distribution)
2. ε_i มีค่าคาดหมาย (Expected Value) เป็น 0 นั่นคือ $E(\varepsilon_i) = 0$
3. ε_i มีความแปรปรวนคงที่ นั่นคือ $V(\varepsilon_i) = \sigma^2$
4. ค่า ε_i และ ε_j สำหรับ $i \neq j$ ต้องไม่มีความสัมพันธ์กันหรือเป็นอิสระต่อกัน

นั่นคือ $COV(\varepsilon_i,\varepsilon_j) = 0$ สำหรับบางค่า $i \neq j$

5. ตัวแปรอิสระต้องเป็นอิสระกัน

2.1.2.3 ปัญหาของการวิเคราะห์การถดถอย

ในการวิเคราะห์การถดถอยนั้น เราจะต้องทำตามข้อสมมติของตัวแบบที่กำหนดไว้ล่วงหน้า หากข้อสมมติของตัวแบบเป็นจริงการสรุปผลการวิเคราะห์จะทำได้อย่างถูกต้อง แต่ถ้าข้อสมมติไม่เป็นจริง ผลการวิเคราะห์อาจเกิดความผิดพลาดได้ ดังนั้นหากพบว่าข้อสมมติของตัวแบบการถดถอยไม่เป็นจริง ควรมีการแปลงข้อมูล (Transform) หรือหาตัวแบบการถดถอยอื่นที่เหมาะสมกว่า เพื่อให้ข้อมูลเป็นไปตามข้อสมมติของตัวแบบ (วิรัชช พานิชวงศ์. 2547 : 85) ซึ่งหากข้อสมมติข้อใดข้อหนึ่งไม่เป็นจริงแล้ว จะมีผลทำให้ตัวประมาณที่ได้ไม่มีคุณสมบัติเป็นตัวประมาณที่ดี วิธีการตรวจสอบแบบง่ายและได้ผลสำหรับการตรวจสอบความเหมาะสมของตัวแบบการถดถอย คือ การตรวจสอบโดยใช้ค่าเศษเหลือ (Residuals : e_i) เพราะ ค่าเศษเหลือเป็นตัวประมาณของ ε_i ที่ได้จากสมการถดถอยที่สร้างขึ้น ซึ่ง $e_i = Y_i - \hat{Y}_i$ ได้แก่

1. การตรวจสอบค่าความคลาดเคลื่อนว่ามีการแจกแจงแบบปกติหรือไม่ ข้อสมมติของตัวแบบการถดถอยกำหนดว่าค่าความคลาดเคลื่อนจะต้องมีการแจกแจงแบบปกติ จากการวิเคราะห์การถดถอย หากพบว่าค่าความคลาดเคลื่อนไม่มีการแจกแจงแบบปกติ จะส่งผลให้การประมาณแบบช่วง และการทดสอบสมมติฐานที่เกี่ยวข้องกับพารามิเตอร์ในตัวแบบไม่ถูกต้อง ผลสรุปที่ได้จะผิดพลาด การทดสอบสามารถทำได้หลายวิธี ดังนี้

1) การตรวจสอบโดยการพล็อต NPP (Normal Probability Plot) เป็นการพล็อตค่าตัวเศษเหลือ (e_i) โดยทำการเรียงลำดับค่า e_i จากน้อยไปหามาก จะได้ค่า $e_{(i)}$ เป็นค่าตัวเศษเหลือเรียงลำดับที่ i (Ordered Observation) ของค่าตัวเศษเหลือ ในการพล็อตกราฟจะกำหนดให้ $e_{(i)}$ เป็นแกนตั้ง ส่วนแกนนอนเป็นค่าคาดหวังของ $e_{(i)}$ หรือ $E(e_{(i)})$ ซึ่งหาได้จากสมการ (Neter. 1990 : 111)

$$E(e_{(i)}) = \sqrt{MSE} \times Z\left(\frac{i-0.375}{n+0.25}\right) \qquad (2.29)$$

เมื่อ $\sqrt{MSE}$ เป็นค่าประมาณของ σ^2 เมื่อ $MSE = \dfrac{\sum e_i^2}{n-2}$

$Z(A)$ เป็นเปอร์เซนต์ไทล์ที่ A ของการแจกแจงแบบปกติมาตรฐาน

n เป็นจำนวนค่าความคลาดเคลื่อน

การพิจารณาว่าค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติหรือไม่ จะพิจารณาจากจุด $\left(E\left(e_{(i)}\right), e_{(i)}\right)$ ที่ได้จากการพล็อตในแผนภาพการกระจาย (Scatter) คือ ถ้าจุด $\left(E\left(e_{(i)}\right), e_{(i)}\right)$ อยู่ในแนวเส้นตรง จะสรุปได้ว่าค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติ แต่ถ้าจุด $\left(E\left(e_{(i)}\right), e_{(i)}\right)$ ไม่อยู่ในแนวเส้นตรง แสดงว่าค่าความคลาดเคลื่อนไม่มีการแจกแจงแบปกติ

2) การทดสอบของ Kolmorogov-Smirnov (K-S) มีหลักเกณฑ์ในการทดสอบการแจกแจงข้อมูลของตัวอย่าง คือ การเปรียบเทียบความน่าจะเป็นสะสมของตัวอย่าง (S(x)) กับ ความน่าจะเป็นสะสมภายใต้สมมติฐานหลัก H_0 (F(x)) โดยมีข้อจำกัดว่าจะต้องทราบค่าพารามิเตอร์ของการแจกแจงที่ต้องการทดสอบ ซึ่งสำหรับการแจกแจงแบบปกติจะต้องทราบค่าเฉลี่ย และค่าความแปรปรวนของประชากร สมมติฐานของการทดสอบ คือ

H_0 : ค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติด้วย $\mu = \mu_0$ และ $\sigma^2 = \sigma_0^2$

H_1 : ค่าความคลาดเคลื่อนไม่มีการแจกแจงแบบปกติด้วย $\mu = \mu_0$ และ $\sigma^2 = \sigma_0^2$

สถิติทดสอบ คือ

$$D = \text{Max}\left|F(x)-S(x)\right| \tag{2.30}$$

โดยที่ $F(x) = P(X \le x) = P\left(z < \frac{x-\mu}{\sigma}\right)$

$S(x) = P(X \le x) =$ (จำนวนข้อมูลตัวอย่างที่ $X < x$) / n

เมื่อ S(x) คือ ค่าความน่าจะเป็นสะสมของตัวอย่าง

F(x) คือ ค่าความน่าจะเป็นสะสมภายใต้สมมติฐานหลัก

ค่าวิกฤต D หาได้จากตาราง Kolmogorov-Smirnov และจะปฏิเสธสมมติฐานหลักเมื่อ $D > D_\alpha$

3) การทดสอบของ Lillifors (อุมาพร จันทศร. 2542 : 257) Lillifors ได้ปรับปรุงการทดสอบของ Kolmogorov-Smirnov ในกรณีที่ต้องการทดสอบเกี่ยวกับการแจกแจงแบบปกติที่ไม่ได้ระบุค่าเฉลี่ยและค่าความแปรปรวน การทดสอบของ Lillifors จะเหมือนกับการทดสอบของ Kolmogorov-Smirnov เกือบทุกประการ ยกเว้นการใช้คะแนนมาตรฐาน (Normalized value) แทนคะแนนดิบ

2. การตรวจสอบความเป็นอิสระของค่าความคลาดเคลื่อน ในงานวิจัยนี้ใช้การทดสอบของ Durbin-Watson เป็นการทดสอบความสัมพันธ์ของ e_i กับ e_{i-1} สถิติทดสอบ คือ

$$d = \frac{\sum_{i=2}^{n}(e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2} \qquad (2.31)$$

สำหรับค่าวิกฤตของ Durbin-Watson จะขึ้นกับขนาดตัวอย่าง (n) และจำนวนตัวแปรอิสระในสมการความถดถอย (k) สรุปคร่าว ๆ ได้ดังนี้ (กัลยา วาณิชย์บัญชา. 2548 : 331)

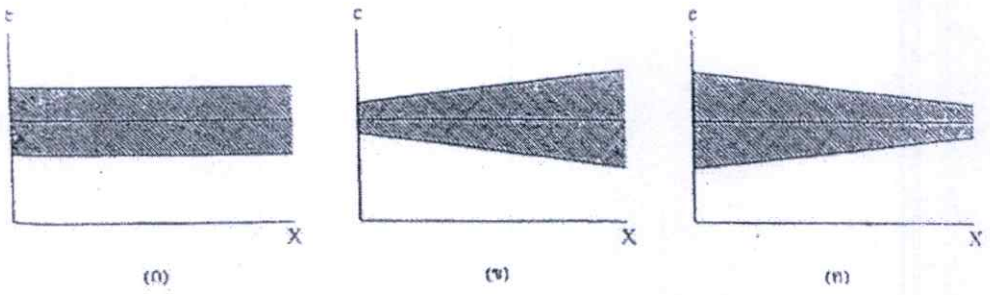
ถ้าค่า Durbin-Watson มีค่าใกล้ 2 (นั่นคือ มีค่าในช่วง 1.5 – 2.5) จะสรุปว่า e_i กับ e_j เป็นอิสระกัน

ถ้าค่า Durbin-Watson < 1.5 แสดงว่า ความสัมพันธ์ของ e_i กับ e_j อยู่ในทิศทางบวก และถ้าค่า Durbin-Watson มีค่าใกล้ 0 แสดงว่า e_i กับ e_j มีความสัมพันธ์กันมาก

ถ้าค่า Durbin-Watson > 2.5 แสดงว่า ความสัมพันธ์ของ e_i กับ e_j อยู่ในทิศทางลบ และถ้าค่า Durbin-Watson มีค่าใกล้ 4 แสดงว่า e_i กับ e_j มีความสัมพันธ์กันมาก

หรือจะพิจารณาค่า p-value ของสถิติทดสอบ Durbin-Watson ถ้าค่า p-value น้อยกว่าระดับนัยสำคัญที่กำหนด จะปฏิเสธสมมติฐาน H_0 หรือสรุปได้ว่า e_i กับ e_j มีความสัมพันธ์กัน

3. การตรวจสอบความคงที่ของความแปรปรวนของค่าคลาดเคลื่อน (Heteroscedasticity) การที่ความแปรปรวนของค่าความคลาดเคลื่อนไม่คงที่ นั่นคือ $V(\varepsilon_i) \neq \sigma^2$ ซึ่งจะมีผลทำให้การหาช่วงความเชื่อมั่นและการทดสอบสมมติฐานทำได้ไม่ถูกต้อง ดังนั้น ก่อนการวิเคราะห์การถดถอยควรตรวจสอบความแปรปรวนของค่าความคลาดเคลื่อนก่อนว่าคงที่หรือไม่ ซึ่งการทดสอบความคงที่ของความแปรปรวนของค่าความคลาดเคลื่อน ทำได้โดยการพล็อตค่ามาตรฐานของตัวเศษเหลือ (e_i) กับค่าประมาณ $\hat{Y}_i$ ถ้าพบว่าจุดต่าง ๆ ในแผนภาพการกระจายเป็นแกนขนาน ดังแสดงในรูปที่ 3.1(ก) จะสรุปได้ว่าความแปรปรวนคงที่ แต่ถ้าพบเป็นรูปทรงอื่น ๆ ดังแสดงในรูป 3.1(ข) และ 3.1(ค) จะสรุปได้ว่าความแปรปรวนไม่คงที่ (ทรงศิริ แต้สมบัติ. 2548 : 194)



รูปที่ 2.1 แสดงลักษณะของค่าความคลาดเคลื่อน กรณีที่ค่าความแปรปรวน

(ก) คงที่ (ข) เพิ่มขึ้น และ (ค) ลดลง

4. การตรวจสอบความเป็นอิสระกันของตัวแปรอิสระ สำหรับตัวแบบการถดถอยที่ดี ตามข้อสมมติของตัวแบบการถดถอย ตัวแปรอิสระทุกตัวจะต้องเป็นอิสระกัน การที่ตัวแปรอิสระมีความสัมพันธ์กัน เรียกว่า เกิดสหสัมพันธ์ร่วม (multi-collinearity) การตรวจสอบว่าตัวแปรอิสระมีสหสัมพันธ์ร่วมหรือไม่นั้น ได้จากการพิจารณาค่า VIF (Variance Inflation Factor) ดังนี้

$$(VIF)_j = \frac{1}{1 - R_j^2} \qquad \text{สำหรับ } j = 1, 2, \ldots, k \qquad (2.32)$$

โดย R_j^2 เป็นค่าสัมประสิทธิ์การตัดสินใจที่ใช้วัดส่วนของความผันแปรรวมของ X_j ที่อธิบายได้ด้วยตัวแปรอิสระอื่น ๆ ที่ไม่รวม X_j

ค่า VIF มีค่าอยู่ระหว่าง 0 ถึง ∞ ถ้าค่า VIF มีค่ามาก หมายความว่า ตัวแปรอิสระ X_j มีความสัมพันธ์กับตัวแปรอิสระอื่น ๆ มาก โดยปกติจะใช้เกณฑ์ว่า เมื่อ $(VIF)_j$ มีค่าไม่เกิน 10 แสดงว่าตัวแปรอิสระจะไม่มีความสัมพันธ์กัน (ทรงศิริ แต้สมบัติ. 2548 : 207)

2.1.2.4 แนวทางการแก้ปัญหา

เมื่อตรวจพบว่าค่าความคลาดเคลื่อนไม่มีการแจกแจงแบบปกติ และความแปรปรวนของความคลาดเคลื่อนไม่คงที่ สามารถแก้ไขได้โดยทำการแปลงข้อมูลของตัวแปรตามด้วยวิธี Box – Cox Transformation (Neter. 1990 : 134-135) Box และ Cox ได้เสนอวิธีเพื่อช่วยในการแก้ปัญหาดังกล่าว โดยการแปลงตามสมการ ดังนี้

$$Y' = Y^{\lambda} \qquad (2.33)$$

เมื่อ λ คือ ค่าพารามิเตอร์ที่กำหนดจากข้อมูล ซึ่งมีเงื่อนไขในการกำหนด λ คือ พยายามกำหนด λ ที่ทำให้ค่า sum square of error (SSE) มีค่าน้อยที่สุดในการนำ Y′ ไปวิเคราะห์การถดถอยเชิงเส้น โดยที่ Y′ จะแตกต่างกันขึ้นอยู่กับการกำหนดค่า λ ดังนี้

$$\lambda = 2, \quad Y' = Y^2$$
$$\lambda = 0.5, \quad Y' = \sqrt{Y}$$
$$\lambda = 0, \quad Y' = \log_e Y$$
$$\lambda = -0.5, \quad Y' = \frac{1}{\sqrt{Y}}$$
$$\lambda = -1, \quad Y' = \frac{1}{Y}$$

2.1.2.5 การทดสอบความแปรปรวน

การทดสอบสมมติฐานนี้จะใช้การวิเคราะห์ความแปรปรวน (Analysis of Variance : ANOVA) เป็นเครื่องมือในการทดสอบว่าตัวแปรอิสระมีอิทธิพลต่อตัวแปรตาม Y หรือไม่ โดยการวิเคราะห์ความแปรปรวน มีแนวความคิดพื้นฐานในการทดสอบ คือ เปรียบเทียบค่าความแปรปรวนที่ใช้ในการอธิบายได้ด้วยสมการการถดถอยเชิงเส้นพหุคูณ กับค่าความแปรปรวนที่อธิบายไม่ได้ด้วยสมการการถดถอยเชิงเส้นพหุคูณซึ่งความสัมพันธ์ระหว่างค่าทั้งสองสรุปได้ดังนี้

$$Y_i - \bar{Y} = \left(\hat{Y}_i - \bar{Y}\right) + \left(Y_i - \hat{Y}_i\right)$$

ยกกำลังทั้งสองข้างจะได้ $\left(Y_i - \bar{Y}\right)^2 = \left[\left(\hat{Y}_i - \bar{Y}\right) + \left(Y_i - \hat{Y}_i\right)\right]^2$

$$\sum\left(Y_i - \bar{Y}\right)^2 = \sum\left[\left(\hat{Y}_i - \bar{Y}\right) + \left(Y_i - \hat{Y}_i\right)\right]^2$$

$$= \sum\left(\hat{Y}_i - \bar{Y}\right)^2 + \sum\left(Y_i - \hat{Y}_i\right)^2 + 2\sum\left(\hat{Y}_i - \bar{Y}\right)\left(Y_i - \hat{Y}_i\right)$$

แต่ $\sum\left(\hat{Y}_i - \bar{Y}\right)\left(Y_i - \hat{Y}_i\right) = \sum\left(b_0 + b_1X_{1i} + \ldots + b_kX_{ki} - \bar{Y}\right)\left(Y_i - b_0 - b_1X_{1i} - \ldots - b_kX_{ki}\right)$

$$= b_0\sum\left(Y_i - b_0 - b_1X_{1i} - \ldots - b_kX_{ki}\right) + b_1\sum X_{1i}\left(Y_i - b_0 - b_1X_{1i} - \ldots - b_kX_{ki}\right) + \ldots$$

$$+ b_k\sum X_{ki}\left(Y_i - b_0 - b_1X_{1i} - \ldots - b_kX_{ki}\right) - \bar{Y}\sum\left(Y_i - b_0 - b_1X_{1i} - \ldots - b_kX_{ki}\right)$$

ซึ่ง $\sum\left(Y_i - b_0 - b_1X_{1i} - \ldots - b_kX_{ki}\right) = 0 \quad \left[\because \sum Y_i = nb_0 + b_1\sum X_{1i} + \ldots + b_k\sum X_{ki}\right]$

และ

$$\sum X_{1i}\left(Y_i - b_0 - b_1X_{1i} - \ldots - b_kX_{ki}\right) = 0$$
$$\vdots$$
$$\sum X_{ki}\left(Y_i - b_0 - b_1X_{1i} - \ldots - b_kX_{ki}\right) = 0$$

เนื่องจาก

$$\sum X_{1i}Y_i = b_0\sum X_{1i} + b_1\sum X_{1i}^2 + \ldots + b_k\sum X_{1i}X_{ki}$$
$$\vdots$$
$$\sum X_{ki}Y_i = b_0\sum X_{ki} + b_1\sum X_{1i}X_{ki} + \ldots + b_k\sum X_{ki}^2$$

ดังนั้น $\sum\left(\hat{Y}_i - \bar{Y}\right)\left(Y_i - \hat{Y}_i\right) = 0$

จะได้ $$\sum\left(Y_i - \bar{Y}\right)^2 = \sum\left(\hat{Y}_i - \bar{Y}\right)^2 + \sum\left(Y_i - \hat{Y}_i\right)^2 \qquad (2.34)$$

หรือ เขียนสมการนี้ในเทอมของผลรวมกำลังสองใหม่ได้เป็น SST = SSR + SSE

โดย SST (Sum Square of Total) คือ ค่าความผันแปรทั้งหมดของ Y โดยเป็นผลรวมกำลังสองของผลต่างของค่าสังเกต และค่าเฉลี่ยที่ใช้วัดความผันแปรของแต่ละค่าสังเกตจากค่าเฉลี่ย เรียก SST ว่า ผลรวมกำลังสองรวม หรือผลรวมกำลังสองของความผันแปร โดยที่

$$SST = \sum\left(Y_i - \bar{Y}\right)^2 = \sum y^2 \qquad (2.35)$$

SSR (Sum Square of Regression) คือ ค่าความผันแปรที่อธิบายได้ หรือค่าความผันแปรของ Y เนื่องจากอิทธิพลของ $X_1, X_2, \ldots, X_k$ โดยเป็นผลรวมกำลังสองของผลต่างของค่าประมาณและค่าเฉลี่ยที่ใช้วัดความผันแปรของแต่ละค่าประมาณจากค่าเฉลี่ย เรียก SSR ว่า ผลรวมกำลังสองของความผันแปรเนื่องจากการถดถอย โดยที่

$$SSR = \sum\left(\hat{Y}_i - \bar{Y}\right)^2 \qquad (2.36)$$

SSE (Sum Square of Error) คือ ค่าความผันแปรที่อธิบายไม่ได้ หรือค่าความผันแปรของ Y เนื่องจากอิทธิพลอื่น ๆ หรือเรียกค่าความผันแปรอย่างสุ่ม โดยเป็นผลรวมกำลังสองของความคลาดเคลื่อน หรือของผลต่างของค่าสังเกตและค่าประมาณ เรียก SSE ว่า ผลรวมกำลังสองของความผันแปรที่ไม่ได้มาจากการถดถอย โดยที่

$$SSE = \sum\left(Y_i - \hat{Y}_i\right)^2 \qquad (2.37)$$

อย่างไรก็ตาม การเปรียบเทียบ SSR กับ SSE โดยตรงนั้น เป็นการเปรียบเทียบที่เอนเอียง (Biased) เนื่องจากค่าทั้งสองมีระดับความเป็นอิสระที่ต่างกัน ดังนั้น ค่าทางสถิติที่ใช้ในการทดสอบสมการการถดถอยเชิงเส้นพหุคูณ ในการวิเคราะห์ค่าความแปรปรวนนี้จึงใช้ค่าความผันแปรที่ปรับด้วยระดับความเป็นอิสระแล้ว เรียกว่า ค่าความผันแปรเฉลี่ย (Mean Square) โดยที่

ค่าความผันแปรที่อธิบายได้เฉลี่ย (Mean Square of Regression) = MSR = SSR / (k-1)

ค่าความผันแปรที่อธิบายไม่ได้เฉลี่ย (Mean Square of Error) = MSE = SSE / (n-k-1)

และค่าสถิติที่ใช้ในการทดสอบตัวแบบการถดถอยเชิงเส้นพหุคูณ จึงเป็น

$$F = \frac{MSR}{MSE} \quad \text{โดยที่} \quad F \sim F_{(k,n-k-1)} \qquad (2.38)$$

ตารางที่ 2.1 การวิเคราะห์ความแปรปรวนของการวิเคราะห์การถดถอยพหุคูณ

แหล่งความแปรปรวน (SV)	องศาความเป็นอิสระ (df)	ผลบวกกำลังสอง (SS)	ผลบวกกำลังสองเฉลี่ย (MS)	F
ความถดถอย (Regression)	k	SSR	MSR	MSR / MSE
ความคลาดเคลื่อน (Error)	n-k-1	SSE	MSE	
ผลรวม (Total)	n-1	SST		

จากตารางวิเคราะห์ความแปรปรวนจะตั้งสมมติฐานได้ดังนี้

$H_0 : \beta_1 = \beta_2 = \ldots = \beta_k = 0$

H_1 : มี β_i สำหรับ $i = 1, 2, \ldots, k$ อย่างน้อยหนึ่งค่าไม่เป็น 0

สถิติทดสอบ คือ $F = \frac{MSR}{MSE}$

เขตปฏิเสธสมมติฐาน H_0 จะปฏิเสธสมมติฐาน H_0 ถ้า $F > F_{\alpha,(k,n-k-1)}$

ก. ถ้ายอมรับสมมติฐาน $H_0 : \beta_1 = \beta_2 = \ldots = \beta_k = 0$ จะสรุปได้ว่า ตัวแปรอิสระ X_i ทุกตัวไม่มีส่วนในการอธิบายความผันแปรของตัวแปรตาม Y

ข. ถ้าปฏิเสธสมมติฐาน H_0 จะสรุปได้ว่า มีตัวแปรอิสระอย่างน้อยหนึ่งตัวมีส่วนในการอธิบายความผันแปรของตัวแปรตาม Y จึงต้องทำการทดสอบต่อไปว่า X_i ตัวใดมีส่วนในการอธิบายความผันแปรของตัวแปรตาม Y จะใช้ตัวสถิติทดสอบ t โดยตั้งสมมติฐานดังนี้

$$H_0 : \beta_i = 0$$

$$H_1 : \beta_i \neq 0 \qquad , i = 1, 2, \ldots, k$$

สถิติทดสอบ คือ $t = \dfrac{b_i - 0}{S_{b_i}}$

เขตปฏิเสธสมมติฐาน H_0 จะปฏิเสธสมมติฐาน H_0 ถ้า $t < -t_{\frac{\alpha}{2}, n-k-1}$ หรือ $t > t_{\frac{\alpha}{2}, n-k-1}$ หรือกล่าวว่าจะปฏิเสธสมมติฐาน ถ้า $|t| > t_{\frac{\alpha}{2}, n-k-1}$

ก. ถ้ายอมรับสมมติฐาน $H_0 : \beta_i = 0$ จะสรุปได้ว่า ตัวแปรอิสระ X_i ไม่มีส่วนในการอธิบายความผันแปรของตัวแปรตาม Y เมื่อมีตัวแปรอิสระตัวอื่น ๆ อยู่ในตัวแบบการถดถอยแล้ว

ข. ถ้าปฏิเสธสมมติฐาน H_0 จะสรุปได้ว่า ตัวแปรอิสระ X_i มีส่วนในการอธิบายความผันแปรของตัวแปรตาม Y เมื่อมีตัวแปรอิสระตัวอื่น ๆ อยู่ในตัวแบบการถดถอยแล้ว

2.1.2.6 การทดสอบแบบ F บางส่วน

เป็นการทดสอบว่าจากตัวแบบ $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \varepsilon$ มีตัวแปรอิสระบางตัวไม่มีส่วนในการอธิบายความผันแปรของตัวแปรตาม Y เมื่อมีตัวแปรอิสระตัวอื่น ๆ อยู่ในตัวแบบ เช่น การทดสอบว่าตัวแปรอิสระ X_i ไม่มีส่วนในการอธิบายความผันแปรของตัวแปรตาม Y จะกำหนด $H_0 : \beta_i = 0$ กับ $H_1 : \beta_i \neq 0$ ในกรณีที่มีตัวแปรอิสระ 3 ตัวที่มีตัวแบบ $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$ จะตั้งสมมติฐานดังนี้

$$H_0 : \beta_3 = 0$$

$$H_1 : \beta_3 \neq 0$$

สามารถเขียนตัวสถิติทดสอบ F ได้ดังนี้ (ทรงศิริ แต้สมบัติ. 2548 : 106)

$$F = \frac{SSR(X_3/X_1,X_2)/1}{MSE(X_1,X_2,X_3)} = \frac{MSR(X_3/X_1,X_2)}{MSE(X_1,X_2,X_3)} \qquad (2.39)$$

ที่ระดับนัยสำคัญ α จะปฏิเสธสมมติฐาน H_0 ถ้า $F \geq F_{\alpha, (1, n-k-1)}$

ก. ถ้ายอมรับสมมติฐาน $H_0 : \beta_3 = 0$ จะสรุปได้ว่า ตัวแปรอิสระ X_3 ไม่มีส่วนในการอธิบายความผันแปรของตัวแปรตาม Y เมื่อมีตัวแปรอิสระตัวอื่น ๆ อยู่ในตัวแบบการถดถอยแล้ว

ข. ถ้าปฏิเสธสมมติฐาน H_0 จะสรุปได้ว่า ตัวแปรอิสระ X_3 มีส่วนในการอธิบายความผันแปรของตัวแปรตาม Y เมื่อมีตัวแปรอิสระตัวอื่น ๆ อยู่ในตัวแบบการถดถอยแล้ว

2.1.2.7 สัมประสิทธิ์การตัดสินใจเชิงพหุ (Multiple Coefficient of Determination : R^2)

เป็นค่าที่วัดความเหมาะสมของรูปแบบที่แสดงสัดส่วนหรือเปอร์เซ็นต์ที่ตัวแปรอิสระ X ทุกตัวรวมกันมีส่วนในการอธิบายความผันแปรของตัวแปรตาม Y ค่า R^2 จะหาได้จาก

$$R^2 = \frac{SSR}{SST} \quad \text{หรือ} \quad R^2 = 1 - \frac{SSE}{SST} \tag{2.40}$$

โดย R^2 มีค่าอยู่ระหว่าง 0 ถึง 1

ถ้า R^2 เข้าใกล้ 1 นั่นคือ ค่า SSR มีค่าใกล้ค่า SST แสดงว่าตัวแปรอิสระ X ทุกตัวรวมกันมีส่วนในการอธิบายความผันแปรของตัวแปรตาม Y สูง

ถ้า R^2 เข้าใกล้ 0 นั่นคือ ค่า SSR มีค่าห่างจากค่า SST มาก แสดงว่าตัวแปรอิสระ X ทุกตัวรวมกันมีส่วนในการอธิบายความผันแปรของตัวแปรตาม Y น้อย

2.1.2.8 สัมประสิทธิ์สหสัมพันธ์ (Coefficient of Correlation: r)

เป็นค่าที่แสดงความสัมพันธ์เชิงเส้นตรงระหว่างตัวแปรสุ่ม Y และตัวแปรสุ่ม X โดยที่ตัวแปรสุ่ม X และ Y มีการแจกแจงแบบปกติ ซึ่ง r มีค่าอยู่ระหว่าง -1 ถึง 1 ค่าสัมบูรณ์ของ r จะบอกขนาดของความสัมพันธ์ เมื่อค่าสัมบูรณ์ของ r เข้าใกล้ 1 แสดงว่าตัวแปรสุ่ม Y และตัวแปรสุ่ม X มีความสัมพันธ์ในเชิงเส้นตรงสูง ส่วนเครื่องหมายของ r จะบอกทิศทางความสัมพันธ์ว่าเป็นทางเดียวกัน หรือทางตรงกันข้าม โดย r หาได้จาก

$$r = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2 \sum (Y - \bar{Y})^2}} = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}} \tag{2.41}$$

เมื่อ $x = (X - \bar{X})$ และ $y = (Y - \bar{Y})$ โดยที่

1. ค่าของ r บอกทิศทางของความสัมพันธ์ระหว่างตัวแปร X และ Y นั่นคือ

ถ้า r มีค่าเป็นบวก แสดงว่า X และ Y มีความสัมพันธ์เชิงเส้นตรงในทางเดียวกัน กล่าวคือ เมื่อ X มีค่าเพิ่มขึ้น ค่าของ Y จะเพิ่มขึ้นด้วย

ถ้า r มีค่าเป็นลบ แสดงว่า X และ Y มีความสัมพันธ์เชิงเส้นตรงในทางตรงกันข้าม กล่าวคือ เมื่อ X มีค่าเพิ่มขึ้น แต่ค่าของ Y จะลดลง

2. ขนาดของ r จะบอกว่าตัวแปร X และ Y มีความสัมพันธ์ในเชิงเส้นตรงมากหรือน้อยเพียงใด

ถ้า $|r|$ มีค่าเข้าใกล้ 1 แสดงว่าตัวแปร X และ Y มีความสัมพันธ์ในเชิงเส้นตรงสูงมาก

ถ้า $|r|$ มีค่าเข้าใกล้ 0 แสดงว่าตัวแปร X และ Y มีความสัมพันธ์ในเชิงเส้นตรงน้อยมาก

2.1.2.9 สัมประสิทธิ์สหสัมพันธ์เชิงส่วน (Coefficients of Partial Correlation)

$r_{Y1.23...k}$ เป็นค่าที่แสดงขนาดและทิศทางของความสัมพันธ์เชิงเส้นตรงระหว่างตัวแปรสุ่ม Y และ X_1 เมื่อค่าของตัวแปรสุ่ม $X_2, \ldots, X_k$ มีค่าคงที่ การอธิบายขนาดและทิศทางของความสัมพันธ์เป็นทำนองเดียวกันกับการอธิบายขนาด และทิศทางของความสัมพันธ์ค่าสัมประสิทธิ์สหสัมพันธ์ ในกรณีนี้ใช้ตัวแปรอิสระ 3 ตัวแปร นั่นคือ X_1, X_2 และ X_3 จะหาค่าสัมประสิทธิ์สหสัมพันธ์บางส่วนระหว่าง X_1 และ X_2 เมื่อ X_3 คงที่ ($r_{12.3}$) คำนวณได้จาก (ทรงศิริ แต้สมบัติ. 2548 : 110)

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}} \qquad (2.42)$$

โดยที่

$$r_{12} = \frac{\sum_{i=1}^{n}(X_{1i} - \bar{X}_1)(X_{2i} - \bar{X}_2)}{\sqrt{\sum_{i=1}^{n}(X_{1i} - \bar{X}_1)^2 \sum_{i=1}^{n}(X_{2i} - \bar{X}_2)^2}}$$

$$r_{13} = \frac{\sum_{i=1}^{n}(X_{1i} - \bar{X}_1)(X_{3i} - \bar{X}_3)}{\sqrt{\sum_{i=1}^{n}(X_{1i} - \bar{X}_1)^2 \sum_{i=1}^{n}(X_{3i} - \bar{X}_3)^2}}$$

$$r_{23} = \frac{\sum_{i=1}^{n}(X_{2i} - \bar{X}_2)(X_{3i} - \bar{X}_3)}{\sqrt{\sum_{i=1}^{n}(X_{2i} - \bar{X}_2)^2 \sum_{i=1}^{n}(X_{3i} - \bar{X}_3)^2}}$$

และ $r_{13}^2 = (r_{13})^2$, $r_{23}^2 = (r_{23})^2$

2.1.2.10 การเลือกสมการถดถอยที่ดีที่สุด

วิธีเพิ่มตัวแปรอิสระ (Forward Selection Procedure)

วิธีเพิ่มตัวแปรอิสระเป็นวิธีการเลือกตัวแปรอิสระเข้าในตัวแบบการถดถอย โดยเลือกตัวแปรอิสระเข้าครั้งละหนึ่งตัว และเลือกไปจนกว่าไม่มีตัวแปรอิสระใดที่มีส่วนในการอธิบายความผันแปรของตัวแปร Y ได้อีกแล้ว ตัวแปรอิสระใดที่เลือกเข้าสู่ตัวแบบการถดถอยแล้วจะไม่มีการตัดออกภายหลัง

ขั้นตอนของวิธีการเพิ่มตัวแปรอิสระ ตัวอย่างกรณีมีตัวแปรอิสระ 3 ตัว ได้แก่ X_1, X_2 และ X_3 มีดังนี้

1) เลือกตัวแปรอิสระตัวแรกเข้าสู่สมการถดถอย โดยเลือกตัวแปรอิสระที่ให้ค่า r_{Yi} สำหรับ i = 1,2,3 สูงที่สุด สมมติว่าตัวแปรอิสระ X_2 เข้าไปในตัวแบบเป็นตัวแรก เพราะ X_2 เป็นตัวแปรที่ให้ค่า r_{Y2} สูงที่สุด

2) ทดสอบว่าตัวแปรอิสระที่เลือกในขั้นตอนที่ 1 มีส่วนในการอธิบายความผันแปรของตัวแปรตาม Y จากตัวแบบ $Y = \beta_0 + \beta_2 X_2 + \varepsilon$ ทดสอบ $H_0 : \beta_2 = 0$ กับ $H_1 : \beta_2 \neq 0$ ถ้ายอมรับ H_0 จะหยุดการเลือกพิจารณาเลือกตัวแปรอิสระเข้าในตัวแบบการถดถอย และสรุปว่า ตัวแปรอิสระทั้งหมดไม่มีส่วนในการอธิบายความผันแปรของตัวแปรตาม Y แต่ถ้าปฏิเสธ H_0 จะทำขั้นตอนที่ 3 ต่อไป นั่นคือ นำตัวแปรอิสระ X_2 เข้าในตัวแบบการถดถอยเป็นตัวที่ 1 และหาตัวแปรอิสระใหม่เข้าในตัวแบบการถดถอย

3) เลือกตัวแปรอิสระตัวที่ 2 เข้าในตัวแบบการถดถอย โดยเลือกตัวแปรอิสระจาก X_1 และ X_3 ที่ให้ค่า $r_{Yi.2}$ สูงสุด สำหรับ i = 1,3 สมมติว่าเลือกตัวแปรอิสระ X_1 เข้าไปในตัวแบบการถดถอยเป็นตัวต่อไป เพราะ X_1 เป็นตัวแปรที่ให้ค่า $r_{Y1.2}$ สูงที่สุด

4) ทดสอบว่าตัวแปรอิสระ X_1 ที่เลือกเข้ามาในขั้นตอนที่ 3 มีส่วนในการอธิบายความผันแปรของตัวแปรตาม Y จากตัวแบบ $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$ ทดสอบ $H_0 : \beta_1 = 0$ กับ

$H_1 : \beta_1 \neq 0$ ถ้ายอมรับ H_0 จะหยุดการพิจารณาต่อไป และสรุปว่าตัวแบบการถดถอยที่ดีที่สุด คือ $Y = \beta_0 + \beta_2 X_2 + \varepsilon$ แต่ถ้าปฏิเสธ H_0 จะทำขั้นตอนที่ 5 ต่อไป

5) เลือกตัวแปรอิสระตัวที่ 3 เข้าในตัวแบบการถดถอย ในที่นี้เหลือเพียงตัวแปรอิสระจาก X_3 จึงนำตัวแปรอิสระ X_3 ไปทดสอบในขั้นตอนที่ 6

6) ทดสอบตัวแปรอิสระตัวสุดท้าย X_3 จากตัวแบบ $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$ ทดสอบ $H_0 : \beta_3 = 0$ กับ $H_1 : \beta_3 \neq 0$ ถ้ายอมรับ H_0 จะหยุดการพิจารณาต่อไป และสรุปว่าตัวแบบการถดถอยที่ดีที่สุด คือ $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$ แต่ถ้าปฏิเสธ H_0 จะสรุปว่าตัวแบบเต็มเป็นตัวแบบการถดถอยที่ดีที่สุด

การพิจารณาตัวแปรอิสระในตัวแบบการถดถอยกรณีมีตัวแปรมากกว่า 3 ตัวแปรอิสระ ทำได้ในทำนองเดียวกัน

วิธีลดตัวแปรอิสระ (Backward Elimination Procedure)

วิธีลดตัวแปรอิสระ เป็นวิธีการเลือกตัวแบบการถดถอยที่ดีที่สุด โดยเริ่มพิจารณาจากตัวแบบที่ประกอบด้วยตัวแปรอิสระทั้งหมด จากนั้นจึงพิจารณาตัดตัวแปรอิสระที่มีส่วนในการอธิบายความผันแปรของตัวแปร Y น้อยที่สุด ออกไปครั้งละ 1 ตัว

ขั้นตอนของวิธีการลดตัวแปรอิสระ ตัวอย่างกรณีมีตัวแปรอิสระ 3 ตัว ได้แก่ X_1, X_2 และ X_3 มีดังนี้

1) จากตัวแบบ $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$ พิจารณาว่าตัวแปรอิสระใดมีส่วนในการอธิบายความผันแปรของตัวแปรตาม Y น้อย ตัวแปรอิสระที่ตัดออกเป็นตัวแปรอิสระที่มีค่าของตัวสถิติทดสอบ F บางส่วนต่ำที่สุด สมมติว่า X_3 เป็นตัวแปรอิสระที่ $F_3 = \min(F_1, F_2, F_3)$ และยอมรับ $H_0 : \beta_3 = 0$ จะตัดตัวแปร X_3 ออกจากตัวแบบ และทำขั้นตอนที่ 2 ต่อไป แต่ถ้าปฏิเสธ H_0 จะตัดตัวแปรอิสระ X_3 ออกจากตัวแบบไม่ได้ จะหยุดการพิจารณา และสรุปว่าตัวแบบที่ประกอบด้วย X_1, X_2, X_3 เป็นตัวแบบที่เหมาะสม

2) จากตัวแบบ $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$ ตัวแปรอิสระที่จะตัดออกเป็นตัวต่อไปเป็นตัวแปรอิสระที่มีค่าของตัวสถิติทดสอบ F บางส่วนต่ำที่สุด สมมติว่า X_2 เป็นตัวแปรอิสระที่ $F_2 = \min(F_1, F_2)$ และยอมรับ $H_0 : \beta_2 = 0$ จะตัดตัวแปร X_2 ออกจากตัวแบบ และทำขั้นตอนทำนองเดียวกันต่อไป แต่ถ้าปฏิเสธ H_0 จะตัดตัวแปรอิสระ X_2 ออกจากตัวแบบไม่ได้ จะหยุดการพิจารณา และสรุปว่าตัวแบบที่ประกอบด้วย X_1, X_2 เป็นตัวแบบที่เหมาะสม

ขั้นตอนของวิธีการลดตัวแปรอิสระจะเสร็จสิ้น เมื่อตัดตัวแปรอิสระออกจากตัวแบบการถดถอยไม่ได้อีกแล้ว ตัวแบบการถดถอยสุดท้ายจะเป็นตัวแบบที่ใช้ในการสร้างสมการถดถอยที่ดีที่สุด

วิธีการเพิ่มตัวแปรอิสระทีละขั้นตอน (Stepwise Regression Procedure)

วิธีการเพิ่มตัวแปรอิสระทีละขั้นตอน เป็นวิธีการเลือกตัวแปรอิสระเข้าในตัวแบบการถดถอยครั้งละ 1 ตัว ตัวแปรอิสระใดที่เข้าอยู่ในตัวแบบการถดถอยแล้ว อาจถูกตัดทิ้งออกไปได้ภายหลัง นั่นคือ ต้องทดสอบว่าตัวแปรอิสระนี้มีส่วนในการอธิบายความผันแปรของตัวแปรตาม Y หรือไม่ ขณะที่ตัวแปรอิสระตัวอื่นอยู่ในตัวแบบการถดถอย วิธีการเพิ่มตัวแปรอิสระทีละขั้นตอนเป็นวิธีที่รวมขั้นตอนของวิธีการเพิ่มตัวแปรอิสระ และวิธีการลดตัวแปรอิสระ

ขั้นตอนของวิธีการเพิ่มตัวแปรอิสระทีละขั้นตอน ตัวอย่างกรณีมีตัวแปรอิสระ 3 ตัว ได้แก่ X_1, X_2 และ X_3 มีดังนี้

1) เลือกตัวแปรอิสระตัวแรกเข้าในตัวแบบการถดถอย โดยเลือกตัวแปรอิสระที่ให้ค่า r_{Yi} สำหรับ i = 1,2,3 สูงที่สุด นั่นคือ เลือกตัวแปรอิสระ X ที่มีความสัมพันธ์กับตัวแปรตาม Y สูงที่สุด สมมติว่าเลือก X_3 นั่นคือ r_{Y3} มีค่าสูงที่สุด

2) จากตัวแบบ $Y = \beta_0 + \beta_3 X_3 + \varepsilon$ ทดสอบ $H_0 : \beta_3 = 0$ กับ $H_1 : \beta_3 \neq 0$ ด้วยการทดสอบแบบ F บางส่วน

2.1 ถ้ายอมรับสมมติฐาน H_0 แสดงว่าตัวแปรอิสระ X_3 ไม่มีส่วนในการอธิบายความผันแปรของตัวแปรตาม Y และกระบวนการเลือกตัวแปรด้วยวิธีนี้จะสิ้นสุดลง นั่นคือ สรุปว่าไม่มีตัวแปรอิสระตัวใดที่เหมาะสมในการอธิบายความผันแปรของตัวแปรตาม Y

2.2 ถ้าปฏิเสธสมมติฐาน H_0 แสดงว่าตัวแปรอิสระ X_3 มีส่วนในการอธิบายความผันแปรของตัวแปรตาม Y ดังนั้น X_3 จึงอยู่ในตัวแบบการถดถอยได้ และได้ตัวแบบการถดถอย คือ $Y = \beta_0 + \beta_3 X_3 + \varepsilon$

3) เลือกตัวแปรอิสระตัวที่ 2 เข้าในตัวแบบการถดถอย โดยเลือกตัวแปรอิสระที่ให้ค่า $r_{Yi.3}$ สำหรับ i = 1,2 สูงที่สุด เมื่อมีตัวแปรอิสระ X_3 อยู่ในสมการแล้ว สมมติว่าเลือก X_2 นั่นคือ $r_{Y2.3}$ มีค่าสูงที่สุด

4) ดังนั้นตัวแบบการถดถอย คือ $Y = \beta_0 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$ จะต้องทำการทดสอบ 2 การทดสอบ เริ่มจาก $H_0 : \beta_2 = 0$ กับ $H_1 : \beta_2 \neq 0$ เป็นการพิจารณาตัวแปรอิสระ X_2 ที่เข้ามาใหม่ด้วยการทดสอบแบบ F บางส่วน

4.1 ถ้ายอมรับสมมติฐาน H_0 แสดงว่าตัวแปรอิสระ X_2 ไม่มีส่วนในการอธิบายความผันแปรของตัวแปรตาม Y เมื่อมีตัวแปรอิสระ X_3 อยู่แล้ว ดังนั้น ตัวแบบการถดถอย คือ $Y = \beta_0 + \beta_3 X_3 + \varepsilon$ ทำให้กระบวนการเลือกตัวแปรด้วยวิธีนี้จะสิ้นสุดลง และตัวแบบการถดถอยนี้มีเพียง X_3 เท่านั้นที่มีอิทธิพลเพียงพอที่จะอธิบายความผันแปรของตัวแปรตาม Y ได้ เพราะ $r_{Y2.3}$ มีค่าสูงที่สุด แต่ยังไม่มีส่วนในการอธิบายความผันแปรของตัวแปรตาม Y เพราะฉะนั้น ตัวแปรอิสระที่เหลือก็จะไม่มีส่วนในการอธิบายความผันแปรของตัวแปรตาม Y ด้วย

4.2 ถ้าปฏิเสธสมมติฐาน H_0 แสดงว่าตัวแปรอิสระ X_2 มีส่วนในการอธิบายความผันแปรของตัวแปรตาม Y เมื่อมีตัวแปรอิสระ X_3 อยู่แล้ว ทำให้ได้ตัวแบบการถดถอย คือ $Y = \beta_0 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$ จึงต้องทำการทดสอบ $H_0 : \beta_3 = 0$ กับ $H_1 : \beta_3 \neq 0$ เพื่อพิจารณาว่าเมื่อมีตัวแปรอิสระ X_2 เข้ามาอยู่ในตัวแบบการถดถอยแล้ว ตัวแปรอิสระ X_3 จะยังมีส่วนในการอธิบายความผันแปรของตัวแปรตาม Y หรือไม่

4.2.1 ถ้ายอมรับสมมติฐาน H_0 สรุปว่า จะต้องตัดตัวแปรอิสระ X_3 ออกจากตัวแบบการถดถอย แสดงว่า ตัวแบบการถดถอยคือ $Y = \beta_0 + \beta_2 X_2 + \varepsilon$ กระบวนการเลือกตัวแปรด้วยวิธีนี้จะสิ้นสุดลง

4.2.2 ถ้าปฏิเสธสมมติฐาน H_0 แสดงว่าตัวแปรอิสระ X_3 ยังคงอยู่ในตัวแบบการถดถอยต่อไปได้ จะได้ตัวแบบการถดถอยคือ $Y = \beta_0 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$

5) ตัวแปรอิสระตัวสุดท้ายที่เข้ามาในตัวแบบการถดถอย คือ ตัวแปรอิสระ X_1 จากตัวแบบเต็ม $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$ จะต้องทำการทดสอบ 3 การทดสอบ เริ่มจาก $H_0 : \beta_1 = 0$ กับ $H_1 : \beta_1 \neq 0$ เป็นการพิจารณาตัวแปรอิสระ X_1 ที่เข้ามาใหม่ด้วยการทดสอบแบบ F บางส่วน

5.1 ถ้ายอมรับสมมติฐาน H_0 แสดงว่าตัวแปรอิสระ X_1 ไม่มีส่วนในการอธิบายความผันแปรของตัวแปรตาม Y เมื่อมีตัวแปรอิสระ X_2, X_3 อยู่แล้ว ดังนั้น ตัวแบบการถดถอย คือ $Y = \beta_0 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$ ทำให้กระบวนการเลือกตัวแปรด้วยวิธีนี้จะสิ้นสุดลง

5.2 ถ้าปฏิเสธสมมติฐาน H_0 แสดงว่าตัวแปรอิสระ X_1 มีส่วนในการอธิบายความผันแปรของตัวแปรตาม Y เมื่อมีตัวแปรอิสระ X_2, X_3 อยู่แล้ว ทำให้ได้ตัวแบบการถดถอย คือ $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$ ต่อไปทำการทดสอบ $H_0 : \beta_2 = 0$ กับ $H_1 : \beta_2 \neq 0$ เพื่อพิจารณาว่าเมื่อมีตัวแปรอิสระ X_1, X_3 เข้ามาอยู่ในตัวแบบการถดถอยแล้ว ตัวแปรอิสระ X_2 จะยังมีส่วนในการอธิบายความผันแปรของตัวแปรตาม Y หรือไม่

5.2.1 ถ้ายอมรับสมมติฐาน H_0 สรุปว่า จะต้องตัดตัวแปรอิสระ X_2 ออกจากตัวแบบการถดถอย เมื่อมีตัวแปรอิสระ X_1, X_3 เข้ามาอยู่ในตัวแบบการถดถอยแล้ว แสดงว่า ตัวแบบการถดถอยคือ $Y = \beta_0 + \beta_1 X_1 + \beta_3 X_3 + \varepsilon$ กระบวนการเลือกตัวแปรด้วยวิธีนี้จะสิ้นสุดลง

5.2.2 ถ้าปฏิเสธสมมติฐาน H_0 แสดงว่าตัวแปรอิสระ X_2 ยังคงอยู่ในตัวแบบการถดถอยต่อไปได้ เมื่อมีตัวแปรอิสระ X_1, X_3 เข้ามาอยู่ในตัวแบบการถดถอยแล้ว จะได้ตัวแบบการถดถอยคือ $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$ ต่อไปทำการทดสอบ $H_0 : \beta_3 = 0$ กับ $H_1 : \beta_3 \neq 0$ เพื่อพิจารณาว่าเมื่อมีตัวแปรอิสระ X_1, X_2 เข้ามาอยู่ในตัวแบบการถดถอยแล้ว ตัวแปรอิสระ X_3 จะยังมีส่วนในการอธิบายความผันแปรของตัวแปรตาม Y หรือไม่

5.2.2.1 ถ้ายอมรับสมมติฐาน H_0 สรุปว่า จะต้องตัดตัวแปรอิสระ X_3 ออกจากตัวแบบการถดถอย เมื่อมีตัวแปรอิสระ X_1, X_2 เข้ามาอยู่ในตัวแบบการถดถอยแล้ว แสดงว่า ตัวแบบการถดถอยคือ $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$ กระบวนการเลือกตัวแปรด้วยวิธีนี้จะสิ้นสุดลง

5.2.2.2 ถ้าปฏิเสธสมมติฐาน H_0 แสดงว่าตัวแปรอิสระ X_3 ยังคงอยู่ในตัวแบบการถดถอยต่อไปได้ เมื่อมีตัวแปรอิสระ X_1, X_2 เข้ามาอยู่ในตัวแบบการถดถอยแล้ว จะได้ตัวแบบการถดถอยคือ $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$

วิธีการเพิ่มตัวแปรอิสระแบบทีละขั้นตอน จะเสร็จสิ้นเมื่อหาตัวแปรอิสระเข้ามาในตัวแบบการถดถอยไม่ได้อีกแล้ว กรณีที่มีตัวแปรอิสระมากกว่า 3 ตัวแปร การคัดเลือกตัวแปรอิสระจะทำได้ในทำนองเดียวกัน โดยเลือกตัวแปรอิสระครั้งละหนึ่งตัวแปร ตัวแปรอิสระที่เข้ามาอยู่ในตัวแบบการถดถอยนี้ อาจจะถูกตัดออกจากตัวแบบได้

2.2 งานวิจัยที่เกี่ยวข้อง

ชลิต ทับทิมทอง, ปกรณ์ ขันทอง และสุดแดน จุลศิริ (2543) ศึกษาความสัมพันธ์ระหว่างปริมาณยานพาหนะกับปริมาณมลพิษ โดยศึกษาข้อมูลบริเวณหน้าจุดตรวจวัดคุณภาพอากาศริมถนนแบบกึ่งถาวร ของกรมควบคุมมลพิษ จำนวน 3 จุด คือ หน้าสถานีการไฟฟ้าย่อยธนบุรี ถนนอินทรพิทักษ์, สถานีที่พักตำรวจจราจรโชคชัย 4 ถนนลาดพร้าว และสถานีเคหะชุมชนดินแดง ถนนดินแดง จุดละ 1 วัน ตั้งแต่เวลา 07.00 น. – 19.00 น. นอกจากนี้ได้ทำการศึกษาจำนวนยานพาหนะที่ใช้น้ำมันเบนซิน และยานพาหนะที่ใช้น้ำมันดีเซลที่ผ่าน ณ บริเวณจุดตรวจวัด จากการศึกษาพบว่าบริเวณหน้าสถานีการไฟฟ้าย่อยธนบุรี ปริมาณยานพาหนะที่ใช้น้ำมันเบนซินคับคั่งอยู่ในช่วงเวลา 08.00 น. – 09.00 น. และปริมาณยานพาหนะที่ใช้น้ำมันดีเซลคับคั่งในช่วงเวลา 18.00 น. – 19.00 น. ส่วนปริมาณมลพิษก๊าซ CO, NO_2 และ SO_2 มีปริมาณมากที่สุดเท่ากับ 3.2 ppm. 62 ppb. และ 10

ppb. ที่ช่วงเวลา 08.00 น. – 09.00 น. 16.00 น. – 17.00 น. และ 18.00 น. – 19.00 น. ตามลำดับ บริเวณหน้าที่พักตำรวจจราจรโชคชัย 4 ปริมาณยานพาหนะที่ใช้น้ำมันเบนซินคับคั่งอยู่ในช่วงเวลา 08.00 น. – 09.00 น. และปริมาณยานพาหนะที่ใช้น้ำมันดีเซลคับคั่งในช่วงเวลา 16.00 น. – 17.00 น. ส่วนปริมาณมลพิษก๊าซ CO, NO_2 และ SO_2 มีปริมาณมากที่สุดเท่ากับ 1.8 ppm. 55 ppb. และ 13 ppb. ที่ช่วงเวลา 08.00 น. – 09.00 น. 16.00 น. – 17.00 น. และ 18.00 น. – 19.00 น. ตามลำดับ และบริเวณหน้าเคหะชุมชนดินแดง ปริมาณยานพาหนะที่ใช้น้ำมันเบนซินคับคั่งอยู่ในช่วงเวลา 18.00 น. – 19.00 น. และปริมาณยานพาหนะที่ใช้น้ำมันดีเซลคับคั่งในช่วงเวลา 17.00 น. – 18.00 น. ส่วนปริมาณมลพิษก๊าซ CO, NO_2 และ SO_2 มีปริมาณมากที่สุดเท่ากับ 5.9 ppm. 113 ppb. และ 21 ppb. ที่ช่วงเวลา 08.00 น. – 09.00 น. 17.00 น. – 18.00 น. และ 07.00 น. – 09.00 น. ตามลำดับ

พิริยา แก้วประเสริฐศิลป์ (2545) ศึกษาวิธีการนำแบบจำลองการกระจายมลพิษทางอากาศในเขตอุตสาหกรรม มาทำนายความเข้มข้นของมลพิษในเขตเมืองขนาดกลาง โดยพัฒนาการใช้แบบจำลอง Atmospheric Dispersion and Modelling System (ADMS 3) ทำนายความเข้มข้นของ CO, NO_2 , SO_2 และ PM_{10} ในเขตอำเภอเมือง จังหวัดนครราชสีมา การศึกษาแบ่งออกเป็น 2 ส่วน ส่วนแรกเป็นการพัฒนาการใช้แบบจำลอง ADMS 3 ในเขตเมือง ส่วนที่สองเป็นการศึกษาความเป็นไปได้ในการใช้วิธีการที่พัฒนาขึ้น เพื่อพยากรณ์ความเข้มข้นของมลพิษในอนาคต โดยใช้ข้อมูลทางด้านภูมิอากาศแบบคร่าว ๆ จากการเปรียบเทียบทางสถิติพบว่า วิธีการที่พัฒนาขึ้นสามารถทำนายความเข้มข้นของ CO และ PM_{10} ในพื้นที่ศึกษาได้มีความถูกต้องสูง และสามารถทำนายความเข้มข้นของ NO_2 และ SO_2 ได้ผลอยู่ในช่วงที่ยอมรับได้ จากนั้นก็ศึกษาความเป็นไปได้ในการใช้วิธีการที่พัฒนาขึ้นในการพยากรณ์ความเข้มข้นของมลพิษในพื้นที่ศึกษา โดยใช้แบบจำลองพยากรณ์ความเข้มข้นของมลพิษด้วยข้อมูลภูมิอากาศในปี พ.ศ. 2542 อย่างคร่าว ๆ เมื่อเปรียบเทียบความเข้มข้นจากการตรวจวัดพบว่า แบบจำลองสามารถพยากรณ์ความเข้มข้นของ CO และ PM_{10} ได้ดี และทำนายความเข้มข้นของ NO_2 และ SO_2 ได้ผลอยู่ในช่วงที่ยอมรับได้

วีรญา แพ่งแสง (2547) ศึกษาเพื่อกำหนดพื้นที่เสี่ยงต่อการเกิดภาวะมลพิษทางอากาศ ในเขตกรุงเทพมหานคร โดยการนำระบบสารสนเทศภูมิศาสตร์ การวิเคราะห์อนุกรมเวลา และการวิเคราะห์การถดถอยพหุคูณมาบูรณาการกัน โดยทำการศึกษาก๊าซมลพิษทั้ง 3 ประเภท ได้แก่ คาร์บอนมอนอกไซด์ ไนโตรเจนไดออกไซด์ และซัลเฟอร์ไดออกไซด์ ซึ่งเป็นก๊าซที่เกิดจากการเผาไหม้เชื้อเพลิงของเครื่องยนต์เป็นส่วนใหญ่ จากการประมาณค่าเชิงพื้นที่จะได้ผลลัพธ์แบ่งเป็นช่วงความเข้มของก๊าซมลพิษแต่ละชนิด ผลที่ได้จากการวิจัยครั้งนี้ คือ พื้นที่เสี่ยงต่อการเกิดภาวะมลพิษก๊าซคาร์บอนมอนอกไซด์ ได้แก่ พื้นที่ที่ครอบคลุมบริเวณเขตจตุจักร บางรัก บึงกุ่ม ปทุมวัน ราชเทวี ลาดพร้าว สาทร และพื้นที่เสี่ยงต่อการเกิดภาวะมลพิษก๊าซไนโตรเจนไดออกไซด์ ได้แก่ พื้นที่ที่ครอบคลุมบริเวณเขตบางรัก บึงกุ่ม ปทุมวัน ราชเทวี ลาดพร้าว วัฒนา สาทร ซึ่งพื้นที่ที่เสี่ยงต่อ

การเกิดภาวะมลพิษทางอากาศของก๊าซมลพิษทั้ง 2 ชนิดนี้ครอบคลุมพื้นที่ในย่านใกล้เคียงกัน ผลการศึกษาพบว่า บริเวณดังกล่าวมีช่วงความเข้มข้นของก๊าซคาร์บอนมอนอกไซด์ และไนโตรเจนไดออกไซด์ในระดับที่สูง คิดเป็น 28.53% และ 23.21% ของพื้นที่ศึกษา ตามลำดับ สำหรับพื้นที่เสี่ยงต่อการเกิดภาวะมลพิษก๊าซซัลเฟอร์ไดออกไซด์ ได้แก่ เขตบางกะปิและสวนหลวง คิดเป็น 29.06% ของพื้นที่ศึกษา

ศรัณย์ ปานศรีพงษ์ (2550) ศึกษาแบบจำลองโครงข่ายประสาทเทียม และแบบจำลองการถดถอยเชิงเส้นพหุคูณ เพื่อทำนายความเข้มข้นของ PM_{10} ล่วงหน้าในพื้นที่กรุงเทพมหานคร และเทศบาลนครราชสีมา และเพื่อเปรียบเทียบประสิทธิภาพของแบบจำลองทั้งสองแบบ โดยใช้ข้อมูลจากกรมควบคุมมลพิษ และกรมอุตุนิยมวิทยา ในช่วงปี พ.ศ. 2543 – 2547 มาทำการสร้างแบบจำลอง ผลการศึกษาพบว่า PM_{10} ในวันรุ่งขึ้น มีความสัมพันธ์สูงที่สุดในทิศทางเดียวกันกับ PM_{10} และ NO_2 โดยมีความสัมพันธ์กับข้อมูลในวันปัจจุบันมากที่สุด รองลงมาคือข้อมูลในวันย้อนหลังที่ลดหลั่นกันไป แบบจำลองโครงข่ายประสาทเทียมที่ได้ต้องทำการแปลงค่าตัวแปรให้อยู่ในรูปลอการิทึมธรรมชาติ และมีจำนวนชั้นซ่อนมากกว่า 1 ชั้น โดยมีค่า MAPE อยู่ในช่วงร้อยละ 17.4 – 12.2 ส่วนแบบจำลองการถดถอยเชิงเส้นพหุคูณที่ผ่านการทดสอบค่าคลาดเคลื่อนแล้ว เป็นแบบจำลองที่มีการคัดเลือกตัวแปรโดยใช้วิธี Backward และต้องทำการแปลงค่าตัวแปรให้อยู่ในรูปลอการึทึมธรรมชาติด้วยเช่นกัน โดยมีค่า Adjusted R^2 อยู่ในช่วง 0.699 – 0.770

บทที่ 3

วิธีการดำเนินการวิจัย

ขั้นตอนการดำเนินงานวิจัย มีขั้นตอนต่าง ๆ ดังนี้

1. การเก็บรวบรวมข้อมูล
2. การวิเคราะห์ข้อมูลเพื่อสร้างตัวแบบ

รายละเอียดแต่ละขั้นตอน มีดังนี้

3.1 การเก็บรวบรวมข้อมูล

ข้อมูลที่ใช้สำหรับการศึกษาวิจัยในครั้งนี้ เป็นข้อมูลประเภททุติยภูมิที่เก็บรวบรวมโดย กรมควบคุมมลพิษ กระทรวงทรัพยากรธรรมชาติและสิ่งแวดล้อม และกรมอุตุนิยมวิทยา กระทรวงเทคโนโลยีสารสนเทศและการสื่อสาร

3.1.1 เก็บรวบรวมค่าที่ใช้วัดมลพิษทางอากาศจำนวน 6 ตัวแปร

1. ก๊าซโอโซน (O_3) เฉลี่ย 1 ชั่วโมง (X_1)
2. ก๊าซไนโตรเจนไดออกไซด์ (NO_2) เฉลี่ย 1 ชั่วโมง (X_2)
3. ก๊าซคาร์บอนมอนอกไซด์ (CO) เฉลี่ย 1 ชั่วโมง (X_3)
4. ก๊าซคาร์บอนมอนอกไซด์ (CO) เฉลี่ย 8 ชั่วโมง (X_4)
5. ก๊าซซัลเฟอร์ไดออกไซด์ (SO_2) เฉลี่ย 24 ชั่วโมง (X_5)
6. ฝุ่นละอองขนาดเล็กกว่า 10 ไมครอน (PM_{10}) เฉลี่ย 24 ชั่วโมง (X_6)

ที่ทำการตรวจวัดในบริเวณต่าง ๆ ทั่วกรุงเทพมหานครทั้งหมด 7 สถานี นั่นคือ สถานีการเคหะชุมชนแห่งชาติ เขตบางกะปิ, สถานีห้วยขวาง เขตห้วยขวาง, สถานียานนาวา เขตยานนาวา, สถานีบางขุนเทียน เขตบางขุนเทียน, สถานีอินทรพิทักษ์ เขตธนบุรี, สถานีลาดพร้าว เขตลาดพร้าว และสถานีดินแดง เขตดินแดง ตั้งแต่วันที่ 1 มกราคม 2549 ถึง 31 ธันวาคม 2549

3.1.2 เก็บรวบรวมค่าที่ใช้วัดสภาพอากาศจำนวน 13 ตัวแปร

1. การระเหยของน้ำในแต่ละวัน (X_7)
2. ข้อมูลเมฆ (X_8)
3. อุณหภูมิตุ้มเปียก (X_9)

4. ทัศนวิสัยโดยเฉลี่ย (X_{10})
5. อุณหภูมิสูงสุดในแต่ละวัน (X_{11})
6. อุณหภูมิต่ำสุดในแต่ละวัน (X_{12})
7. อุณหภูมิตุ้มแห้ง (X_{13})
8. ความยาวนานของแสงแดดในแต่ละวัน (X_{14})
9. ความชื้นสัมพัทธ์เฉลี่ยในแต่ละวัน (X_{15})
10. ปริมาณน้ำฝนในแต่ละวัน (X_{16})
11. ความกดอากาศเฉลี่ยในแต่ละวัน (X_{17})
12. ทิศทางลมในแต่ละวัน (X_{18})
13. ความเร็วลมสูงสุดในแต่ละวัน (X_{19})

เป็นค่าที่ได้จากการตรวจวัดในบริเวณกรมอุตุนิยมวิทยาบางนา ตั้งแต่วันที่ 1 มกราคม 2549 ถึง 31 ธันวาคม 2549 โดยในแต่ละสถานีจะมีจำนวนข้อมูลไม่เท่ากัน เนื่องจากมีการตัดข้อมูลที่ไม่สมบูรณ์ออกไป

3.2 การวิเคราะห์ข้อมูลเพื่อสร้างตัวแบบ

การวิเคราะห์ข้อมูลเพื่อสร้างตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศ ใช้โปรแกรมสำเร็จรูปทางสถิติ SPSS/PC for windows (Statistical Package for the Social Science/Personal Computer for Windows) Version 11.5 ขั้นตอนในการวิเคราะห์ข้อมูลเพื่อสร้างตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศ มีขั้นตอนดังนี้

1. หาค่า KMO ของแต่ละสถานี โดยนำค่าตัวแปรอิสระ (X_1 - X_{19}) ของแต่ละสถานีมาวิเคราะห์เพื่อดูว่าวิธีวิเคราะห์ปัจจัยเหมาะสมกับข้อมูลที่มีอยู่หรือไม่

2. ใช้วิธีการวิเคราะห์ปัจจัย (Factor Analysis) เพื่อจัดตัวแปรอิสระที่มีความสัมพันธ์กันไว้ในกลุ่มเดียวกัน มีขั้นตอนดังนี้

2.1 ทำการแปลงค่าสังเกต X_{ji} ให้เป็นค่ามาตรฐาน (standardized) Z_{ji} เนื่องจากตัวแปรแต่ละตัวมีหน่วยไม่เหมือนกัน

2.2 สร้างเมตริกซ์สหสัมพันธ์ (Correlation Matrix, **R**) หาค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปรจากค่ามาตรฐานที่หาได้จากข้อ 1 จะได้เมตริกซ์สหสัมพันธ์สำหรับสร้างตัวแบบดัชนีคุณภาพอากาศ ดังนี้

$$R = \begin{bmatrix} 1 & r_{z_1z_2} & r_{z_1z_3} & \cdots & r_{z_1z_{19}} \\ r_{z_2z_1} & 1 & r_{z_2z_3} & \cdots & r_{z_2z_{19}} \\ r_{z_3z_1} & r_{z_3z_2} & 1 & \cdots & r_{z_3z_{19}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{z_{19}z_1} & r_{z_{19}z_2} & r_{z_{19}z_3} & \cdots & 1 \end{bmatrix}_{19\times19} \quad (3.1)$$

2.3 สกัดปัจจัย ทำการประมาณค่าน้ำหนักปัจจัย (factor loading) ด้วยวิธี Principal Component และพิจารณาว่าจะมีจำนวนปัจจัยร่วมกี่ปัจจัย แต่ละปัจจัยร่วมประกอบด้วยตัวแปรอะไรบ้าง

2.4 หมุนแกนปัจจัย ถ้าพิจารณาจำนวนปัจจัยร่วม หรือจัดตัวแปรให้เข้าไปอยู่ในปัจจัยร่วมต่าง ๆ ไม่ได้ ตามข้อ 2 จะทำการหมุนแกนปัจจัย โดยวิธี Varimax จะทำให้ได้ค่าปัจจัยใหม่ที่เกิดจากการหมุนแกนปัจจัย

2.5 การคำนวณค่าคะแนนปัจจัย (factor score) เมื่อจัดตัวแปรให้เข้าไปอยู่ปัจจัยร่วมแล้ว ขั้นต่อไปจะทำการคำนวณหาคะแนนปัจจัยด้วยวิธี Regression Method ค่าคะแนนปัจจัยจะถือว่าเป็นตัวแทนของแต่ละปัจจัย และถือว่าเป็นตัวแปรใหม่ที่จะนำไปวิเคราะห์หาสมการสำหรับพยากรณ์ดัชนีคุณภาพอากาศด้วยวิธีการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ

3. สร้างตัวแบบโดยใช้วิธีการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ (Multiple Linear Regression Analysis) เมื่อได้ค่าคะแนนปัจจัย (factor score) ซึ่งเป็นตัวแปรใหม่ที่ได้จากการวิเคราะห์ปัจจัยทั้งหมด m ตัวแปร ได้แก่ $F_1, F_2, F_3, \ldots, F_m$ นำตัวแปรใหม่ที่ได้มาวิเคราะห์หาตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศด้วยวิธีการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ โดยถือว่าตัวแปรอิสระแต่ละตัวมีความสัมพันธ์เชิงเส้นตรงกับตัวแปรตาม Y มีขั้นตอนดังนี้

3.1 ตรวจสอบลักษณะความสัมพันธ์ของตัวแปรตาม Y กับตัวแปรอิสระ F ทุกตัวทีละคู่ โดยการสร้างแผนภาพการกระจายว่ามีความสัมพันธ์ในแบบเชิงเส้นตรงหรือไม่ หากพบว่าความสัมพันธ์ระหว่างตัวแปรตาม Y กับตัวแปรอิสระ F ไม่มีความสัมพันธ์ในแบบเชิงเส้นตรง จะกำหนดความสัมพันธ์ใหม่ โดยการแปลงตัวแปรอิสระ F ไปเป็นฟังก์ชันของตัวแปรอิสระ F เพื่อให้ได้ความสัมพันธ์ในแบบเชิงเส้นตรงกับตัวแปรตาม Y

3.2 ประมาณค่าพารามิเตอร์ของการถดถอยโดยการประมาณค่าของ $\beta_0, \beta_1, \beta_2, \ldots, \beta_m$ ด้วยวิธีกำลังสองน้อยสุด (Least Square Method)

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

โดย

$$\mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_m \end{bmatrix} \qquad \mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \qquad X = \begin{bmatrix} 1 & F_{11} & F_{12} & \cdots & F_{1n} \\ 1 & F_{21} & F_{22} & \cdots & F_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & F_{m1} & F_{m2} & \cdots & F_{mn} \end{bmatrix}$$

จะได้สมการสำหรับพยากรณ์ดัชนีคุณภาพอากาศ ซึ่งอาจมีรูปแบบดังนี้

$$\hat{Y}_i = b_0 + b_1F_{1i} + b_2F_{2i} + ... + b_mF_{mi} \qquad (3.2)$$

เมื่อ $i = 1, 2, ..., n$

n = จำนวนวันในการพยากรณ์ดัชนีคุณภาพอากาศ

3.3 นำตัวแบบที่ได้ในข้อ 3.2 มาพยากรณ์ดัชนีคุณภาพอากาศ

3.4 หาค่าเศษเหลือ (e_i) ของการพยากรณ์ นำค่าพยากรณ์ดัชนีคุณภาพอากาศที่ได้จากข้อ 3.3 มาหาค่าเศษเหลือ จากสมการ

$$e_i = Y_i - \hat{Y}_i \qquad (3.3)$$

เมื่อ Y คือ ค่าดัชนีคุณภาพอากาศที่วัดได้

$\hat{Y}$ คือ ค่าพยากรณ์ดัชนีคุณภาพอากาศ

3.5 ตรวจสอบข้อสมมติของค่าความคลาดเคลื่อน (Assumption) และแก้ปัญหาของการวิเคราะห์การถดถอย ดังนี้

3.4.1 ตรวจสอบการแจกแจงแบบปกติ (Test of normality) โดยการพล๊อตกราฟ Normal Probability Plot และทดสอบด้วยวิธีของ Kolmogorov-Smirnov ที่ปรับปรุงด้วยการทดสอบของ Lilliefors

3.4.2 ตรวจสอบความเป็นอิสระของค่าความคลาดเคลื่อน โดยใช้การทดสอบของ Durbin-Watson

3.4.3 ตรวจสอบความแปรปรวนของค่าความคลาดเคลื่อน (Heteroscedasticity) โดยการพล็อตกราฟระหว่างค่ามาตรฐานของค่าเศษเหลือ (e_i) กับค่าประมาณ $\hat{Y}_i$ ถ้าความแปรปรวนของค่าความคลาดเคลื่อนคงที่ กราฟที่ได้จะต้องแตกต่างจาก 0 ไม่มากนัก และมีความเคลื่อนไหวอยู่ในแนวขนานกับแกน X

3.4.4 ตรวจสอบความเป็นอิสระของตัวแปรอิสระ (multi-collinearity) โดยใช้ค่า VIF ในการทดสอบ

บทที่ 4

ผลการวิเคราะห์ข้อมูล

ผลการวิจัยในบทนี้ จะประกอบไปด้วยผลการวิเคราะห์ข้อมูลในการสร้างตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีต่าง ๆ ทั้งหมด 7 สถานี ประกอบด้วย สถานีการเคหะชุมชนแห่งชาติ สถานีห้วยขวาง สถานียานนาวา สถานีบางขุนเทียน สถานีอินทรพิทักษ์ สถานีลาดพร้าว และสถานีดินแดง โดยใช้วิธีการวิเคราะห์ปัจจัย (Factor Analysis) และวิธีการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ (Multiple Linear Regression Analysis) ได้ผลวิเคราะห์ข้อมูลดังนี้

4.1 ผลการวิเคราะห์ข้อมูลในการสร้างตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีเคหะชุมชนแห่งชาติ

4.1.1 ผลการวิเคราะห์ข้อมูลโดยใช้วิธีการวิเคราะห์ปัจจัย

ผลการพิจารณาค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปรอิสระ พบว่าค่า $|r|$ มีค่ามากกว่า 0.5 จึงสรุปได้ว่าตัวแปรอิสระมีความสัมพันธ์กันสูง

ผลการพิจารณาค่า KMO (Kaiser-Meyer-Olkin) ซึ่งเป็นค่าที่ใช้วัดความเหมาะสมของข้อมูลที่จะนำมาวิเคราะห์ปัจจัย พบว่าค่า KMO = 0.720 ซึ่งมีค่ามากกว่า 0.5 จึงสรุปได้ว่าข้อมูลที่มีอยู่เหมาะสมที่จะใช้วิธีการวิเคราะห์ปัจจัย

สำหรับผลการวิเคราะห์ปัจจัยของตัวแปรอิสระ 19 ตัวแปร โดยวิธีวิเคราะห์องค์ประกอบหลัก แสดงดังตารางที่ 4.1

ตารางที่ 4.1 แสดงผลการวิเคราะห์ปัจจัยของตัวแปรอิสระสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีการเคหะชุมชนแห่งชาติ

Component	Initial Eigenvalues			Extraction Sums of Squared			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5.234	27.547	27.547	5.234	27.547	27.547	3.445	18.132	18.132
2	3.101	16.323	43.869	3.101	16.323	43.869	3.157	16.615	34.747
3	1.751	9.214	53.083	1.751	9.214	53.083	3.084	16.231	50.978
4	1.168	6.147	59.229	1.168	6.147	59.229	1.535	8.079	59.057
5	1.068	5.623	64.853	1.068	5.623	64.853	1.101	5.796	64.853
6	.985	5.185	70.038						
7	.844	4.443	74.481						
8	.825	4.343	78.823						
9	.768	4.044	82.867						
10	.690	3.630	86.497						
11	.498	2.623	89.120						
12	.460	2.423	91.543						
13	.432	2.274	93.818						
14	.349	1.836	95.653						
15	.261	1.371	97.025						
16	.226	1.191	98.216						
17	.196	1.030	99.246						
18	.142	.747	99.993						
19	.001	.007	100.000						

Extraction Method: Principal Component

จากตารางที่ 4.1 พบว่าตัวแปรอิสระต่าง ๆ มีการจับกลุ่มรวมกันเป็นปัจจัยทั้งหมด 5 ปัจจัย และสามารถอธิบายความแปรปรวนรวมทั้งหมดของตัวแปรได้ร้อยละ 64.853

ทำการหมุนแกนปัจจัยด้วยวิธีวาริแมกซ์ แสดงดังตารางที่ 4.2

ตารางที่ 4.2 แสดงผลการหมุนแกนปัจจัยด้วยวิธีวาริแมกซ์ของสถานีการเคหะชุมชนแห่งชาติ

	Component				
	1	2	3	4	5
Zscore(X14)	.859				
Zscore(X8)	-.801			.243	
Zscore(X15)	-.775	.317			.208
Zscore(X7)	.647	.323	-.332	.222	
Zscore(X1)	.533	-.240		-.352	.254
Zscore(X13)		.881	-.231		
Zscore(X9)	-.375	.826	-.276		
Zscore(X12)		.743			
Zscore(X17)	.399	-.529	.257	-.415	.213
Zscore(X5)		.435			-.212
Zscore(X2)			.819		
Zscore(X4)			.784		
Zscore(X6)	.366	-.235	.751	-.226	
Zscore(X10)		.249	-.651	.229	-.401
Zscore(X3)			.560		-.272
Zscore(X19)			-.242	.660	
Zscore(X18)				.642	
Zscore(X16)	-.358			.311	.592
Zscore(X11)	.319	.439			.492

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.

จากตารางที่ 4.2 การจัดตัวแปรอิสระให้อยู่ในปัจจัยใด จะพิจารณาจากขนาดของค่า Factor Loading ที่มีค่ามากที่สุด ซึ่งแต่ละปัจจัยประกอบด้วยตัวแปรต่าง ๆ ดังนี้

ปัจจัยที่ 1 F_1 จะสามารถอธิบายความแปรปรวนรวมทั้งหมดของตัวแปรได้ร้อยละ 18.132 ประกอบด้วยสมาชิกทั้งหมด 5 ตัวแปร ดังนี้

1. ก๊าซโอโซน (O_3) เฉลี่ย 1 ชั่วโมง (X_1)
2. การระเหยของน้ำในแต่ละวัน (X_7)
3. ข้อมูลเมฆ (X_8)
4. ความยาวนานของแสงแดดในแต่ละวัน (X_{14})
5. ความชื้นสัมพัทธ์เฉลี่ยในแต่ละวัน (X_{15})

ปัจจัยที่ 2 F_2 จะสามารถอธิบายความแปรปรวนรวมทั้งหมดของตัวแปรได้ร้อยละ 16.615 ประกอบด้วยสมาชิกทั้งหมด 5 ตัวแปร ดังนี้

1. ก๊าซซัลเฟอร์ไดออกไซด์ (SO_2) เฉลี่ย 24 ชั่วโมง (X_5)
2. อุณหภูมิตุ้มเปียก (X_9)
3. อุณหภูมิต่ำสุดในแต่ละวัน (X_{12})
4. อุณหภูมิตุ้มแห้ง (X_{13})
5. ความกดอากาศเฉลี่ยในแต่ละวัน (X_{17})

ปัจจัยที่ 3 F_3 จะสามารถอธิบายความแปรปรวนรวมทั้งหมดของตัวแปรได้ร้อยละ 16.231 ประกอบด้วยสมาชิกทั้งหมด 5 ตัวแปร ดังนี้

1. ก๊าซไนโตรเจนไดออกไซด์ (NO_2) เฉลี่ย 1 ชั่วโมง (X_2)
2. ก๊าซคาร์บอนมอนอกไซด์ (CO) เฉลี่ย 1 ชั่วโมง (X_3)
3. ก๊าซคาร์บอนมอนอกไซด์ (CO) เฉลี่ย 8 ชั่วโมง (X_4)
4. ฝุ่นละอองขนาดเล็กกว่า 10 ไมครอน (PM_{10}) เฉลี่ย 24 ชั่วโมง (X_6)
5. ทัศนวิสัยโดยเฉลี่ย (X_{10})

ปัจจัยที่ 4 F_4 จะสามารถอธิบายความแปรปรวนรวมทั้งหมดของตัวแปรได้ร้อยละ 8.079 ประกอบด้วยสมาชิกทั้งหมด 2 ตัวแปร ดังนี้

1. ทิศทางลมในแต่ละวัน (X_{18})
2. ความเร็วลมสูงสุดในแต่ละวัน (X_{19})

ปัจจัยที่ 5 F_5 จะสามารถอธิบายความแปรปรวนรวมทั้งหมดของตัวแปรได้ร้อยละ 5.796 ประกอบด้วยสมาชิกทั้งหมด 2 ตัวแปร ดังนี้

1. อุณหภูมิสูงสุดในแต่ละวัน (X_{11})
2. ปริมาณน้ำฝนในแต่ละวัน (X_{16})

ผลการหาสมการสำหรับคำนวณค่าคะแนนปัจจัย (factor score) ด้วยวิธีวิเคราะห์การถดถอย (Regression Method) ได้ผลดังนี้

กำหนดให้ F_1, F_2, F_3, F_4 และ F_5 คือ ค่าคะแนนปัจจัยของปัจจัยที่ 1, 2, 3, 4 และ 5 ตามลำดับ จะได้ว่า

$$F_1 = 0.533Z_1 + 0.647Z_7 - 0.801Z_8 + 0.859Z_{14} - 0.775Z_{15} \quad (4.1)$$

$$F_2 = 0.435Z_5 + 0.826Z_9 + 0.743Z_{12} + 0.881Z_{13} - 0.529Z_{17} \quad (4.2)$$

$$F_3 = 0.819Z_2 + 0.560Z_3 + 0.784Z_4 + 0.751Z_6 - 0.651Z_{10} \quad (4.3)$$

$$F_4 = 0.642Z_{18} + 0.660Z_{19} \quad (4.4)$$

$$F_5 = 0.492Z_{11} + 0.592Z_{16} \quad (4.5)$$

4.1.2 ผลการวิเคราะห์ข้อมูลโดยใช้วิธีการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ

ตรวจสอบลักษณะความสัมพันธ์ของตัวแปรตาม Y กับตัวแปรอิสระ F ทีละคู่ พบว่า ตัวแปรตาม Y ที่มีความสัมพันธ์กับตัวแปรอิสระ F เป็นเส้นตรง คือ Y กับ F_1, Y กับ F_3 และ Y กับ F_4 ส่วนตัวแปรตาม Y ที่มีความสัมพันธ์กับตัวแปรอิสระ F ไม่เป็นเส้นตรง จะต้องทำการแปลงข้อมูล นั่นคือ Y กับ F_2 แปลง F_2 เป็นแบบ Cubic และ Y กับ F_5 แปลง F_5 เป็นแบบ Cubic

นำค่าคะแนนปัจจัยที่ได้จากการวิเคราะห์ปัจจัยในข้อ 4.1.1 มาวิเคราะห์หาตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศ ประมาณค่าพารามิเตอร์ของการถดถอยด้วยวิธีกำลังสองน้อยที่สุด (Least Square Method) และเลือกสมการที่ดีที่สุดจากวิธีการคัดเลือกตัวแปรทั้ง 3 วิธี ดังนี้

1) วิธีการเพิ่มตัวแปรอิสระ (Forward Selection Procedure)

ผลการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยด้วยวิธีการเพิ่มตัวแปรอิสระ แสดงดังตารางที่ 4.3

ตารางที่ 4.3 แสดงผลการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยโดยวิธีการเพิ่มตัวแปรอิสระของตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีการเคหะชุมชนแห่งชาติ

Model		Unstandardized Coefficients		t	Sig.	Collinearity Statistics
		B	Std. Error			VIF
5	(Constant)	55.548	.765	72.600	.000	
	FAC3_1	7.246	.859	8.434	.000	1.022
	FAC1_1	7.657	.768	9.974	.000	1.015
	FAC5_1	5.698	.946	6.022	.000	1.020
	FAC4_1	-4.853	.817	-5.944	.000	1.007
	FAC2_1	-2.356	.775	-3.041	.003	1.013

a. Dependent Variable: Y

ผลจากตารางที่ 4.3 พบว่าตัวแปรที่มีผลต่อดัชนีคุณภาพอากาศ คือ ตัวแปรในปัจจัยร่วมที่ 1 (FAC1_1), ตัวแปรในปัจจัยร่วมที่ 2 (FAC2_1), ตัวแปรในปัจจัยร่วมที่ 3 (FAC3_1), ตัวแปรในปัจจัยร่วมที่ 4 (FAC4_1) และตัวแปรในปัจจัยร่วมที่ 5 (FAC5_1) ได้สมการการถดถอย ดังนี้

$$\hat{Y} = 55.548 + 7.657F_1 - 2.356F_2 + 7.246F_3 - 4.853F_4 + 5.698F_5$$

และมีค่า $R^2 = 0.480$

2) วิธีการลดตัวแปรอิสระ (Backward Elimination Procedure)

ผลการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยด้วยวิธีการลดตัวแปรอิสระแสดงดังตารางที่ 4.4

ตารางที่ 4.4 แสดงผลการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยโดยวิธีการลดตัวแปรอิสระของตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีการเคหะชุมชนแห่งชาติ

Model		Unstandardized Coefficients		t	Sig.	Collinearity Statistics
		B	Std. Error			VIF
1	(Constant)	54.829	.942	58.184	.000	
	FAC1_1	7.267	.823	8.826	.000	1.170
	FAC2_1	-1.521	1.004	-1.515	.003	1.706
	FAC2_2	.725	.556	1.304	.009	2.018
	FAC3_1	7.136	.862	8.276	.000	1.032
	FAC4_1	-4.987	.822	-6.067	.000	1.023
	FAC5_1	6.051	.983	6.156	.000	1.103

a. Dependent Variable: Y

ผลจากตารางที่ 4.4 พบว่าตัวแปรที่มีผลต่อดัชนีคุณภาพอากาศ คือ ตัวแปรในปัจจัยร่วมที่ 1 (FAC1_1), ตัวแปรในปัจจัยร่วมที่ 2 (FAC2_1), ตัวแปรในปัจจัยร่วมที่ 2 ยกกำลังสอง (FAC2_2), ตัวแปรในปัจจัยร่วมที่ 3 (FAC3_1), ตัวแปรในปัจจัยร่วมที่ 4 (FAC4_1) และตัวแปรในปัจจัยร่วมที่ 5 (FAC5_1) ได้สมการการถดถอยดังนี้

$$\hat{Y} = 54.829 + 7.267F_1 - 1.521F_2 + 0.725F_2^2 + 7.136F_3 - 4.987F_4 + 6.051F_5$$

และมีค่า $R^2 = 0.483$

3) วิธีการเพิ่มตัวแปรอิสระทีละขั้นตอน (Stepwise Regression Procedure)

ผลการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยด้วยวิธีการเพิ่มตัวแปรอิสระทีละขั้นตอน ได้ผลเหมือนกับวิธีการเพิ่มตัวแปรอิสระ

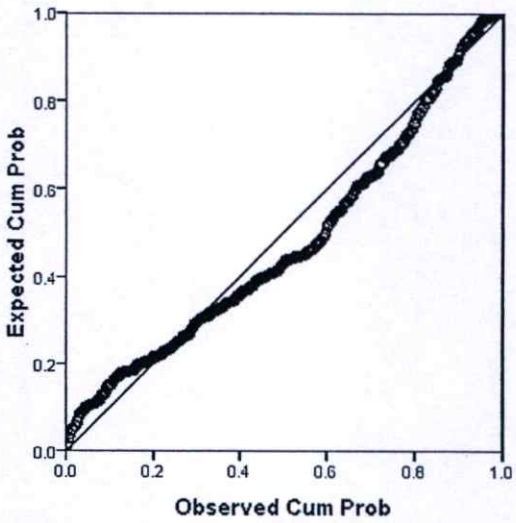
จากวิธีการคัดเลือกตัวแปรทั้ง 3 วิธี จะเลือกสมการที่ดีที่สุดด้วยวิธีลดตัวแปรอิสระ เพราะให้ค่า R^2 สูงที่สุด ดังนั้นสมการถดถอยที่ดีที่สุดสำหรับการพยากรณ์ คือ

$$\hat{Y} = 54.829 + 7.267F_1 - 1.521F_2 + 0.725F_2^2 + 7.136F_3 - 4.987F_4 + 6.051F_5$$

4.1.3 ผลการตรวจสอบข้อสมมติของค่าความคลาดเคลื่อน

1) การตรวจสอบค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติ

1.1) การพล๊อตกราฟ Normal Probability Plot



รูปที่ 4.1 กราฟ Normal Probability Plot ของค่าเศษเหลือของสมการถดถอยสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีการเคหะชุมชนแห่งชาติ

จากรูปที่ 4.1 พบว่ากราฟที่ได้ไม่เป็นเส้นตรง จึงสรุปว่าค่าความคลาดเคลื่อนไม่มีการแจกแจงแบบปกติ

1.2) ผลการทดสอบของ Kolmogorov-Smirnov ที่ปรับปรุงด้วยการทดสอบของ Lilliefors ได้ผลดังนี้

สมมติฐานของการทดสอบ คือ

H_0 : ค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติ

H_1 : ค่าความคลาดเคลื่อนไม่มีการแจกแจงแบบปกติ

ตารางที่ 4.5 แสดงการทดสอบค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติของสถานีการเคหะชุมชนแห่งชาติ

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Unstandardized Residual	.110	293	.000	.943	293	.000

a. Lilliefors Significance Correction

สถิติทดสอบ Kolmogorov-Smirnov

$$D = \max\left|F(x)-S(x)\right| = 0.110$$

ค่า p-value = 0.000 มีค่าน้อยกว่า $\alpha = 0.01$ ดังนั้นจึงปฏิเสธ H_0 สรุปได้ว่าค่าความคลาดเคลื่อนไม่มีการแจกแจงแบบปกติ

2) การตรวจสอบความเป็นอิสระกันของค่าความคลาดเคลื่อน

สมมติฐานของการทดสอบ คือ

H_0 : ค่าความคลาดเคลื่อนเป็นอิสระกัน

H_1 : ค่าความคลาดเคลื่อนไม่เป็นอิสระกัน

ตารางที่ 4.6 แสดงค่า Durbin-Watson ที่ได้จากการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยด้วยวิธีการลดตัวแปรอิสระของสถานีการเคหะชุมชนแห่งชาติ

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.695[a]	.483	.472	13.03305	1.481

a. Predictors: (Constant), FAC5_1, FAC2_1, FAC4_1, FAC1_1, FAC3_1, FAC2_2

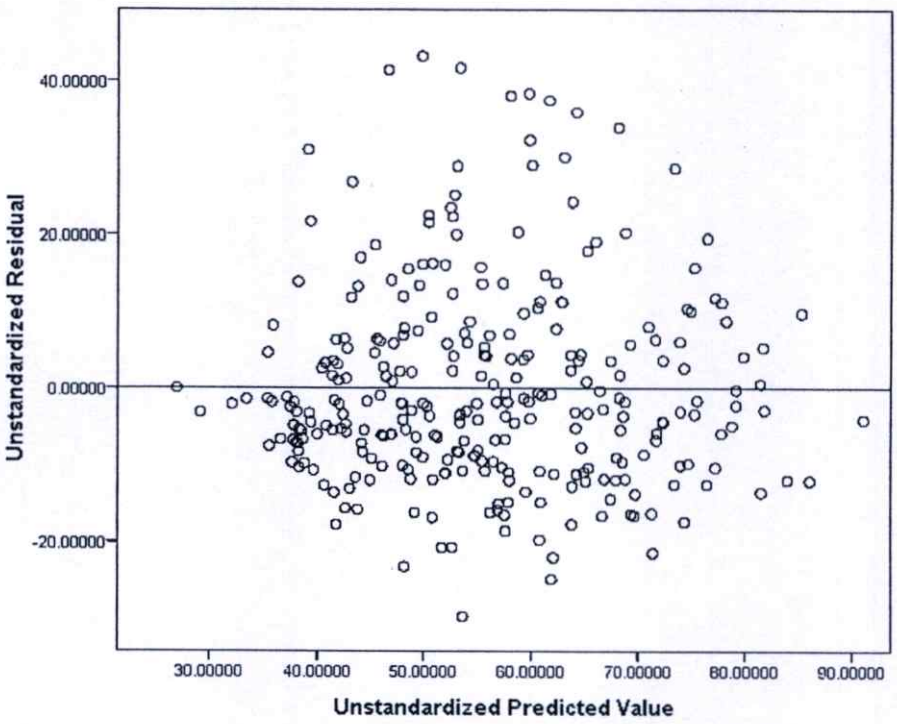
b. Dependent Variable: Y

สถิติทดสอบ

$$d = \frac{\sum_{i=2}^{n}(e_i - e_{i-1})}{\sum_{i=1}^{n} e_i^2} = 1.481$$

ค่า Durbin-Watson = 1.481 มีค่าน้อยกว่า 1.5 ดังนั้นจึงปฏิเสธ H_0 จะสรุปว่า e_i กับ e_j ไม่เป็นอิสระกัน

3) การตรวจสอบค่าความแปรปรวนของค่าความคลาดเคลื่อน โดยการพล็อตกราฟระหว่าง $E(e)$ กับ $E(\hat{Y})$



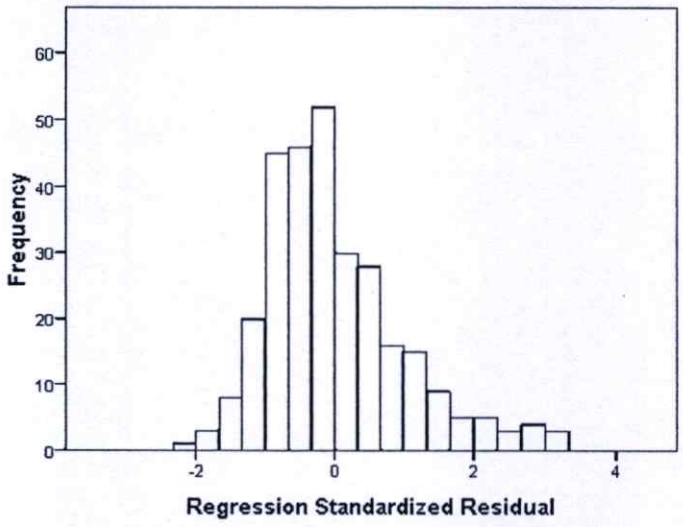
รูปที่ 4.2 ความสัมพันธ์ระหว่าง $E(e)$ กับ $E(\hat{Y})$ ของสถานีการเคหะชุมชนแห่งชาติ

จากรูปที่ 4.2 พบว่าค่า e_i มีการกระจายอยู่รอบค่าศูนย์ที่ด้านล่างมากกว่าด้านบน และไม่ขนานไปกับแกน X สรุปได้ว่าค่าความแปรปรวนของค่าความคลาดเคลื่อนมีค่าไม่คงที่

4) การตรวจสอบความเป็นอิสระกันของตัวแปรอิสระ

จากตารางที่ 4.4 พบว่า ค่า VIF ของตัวแปรอิสระทุกตัวมีค่าไม่เกิน 10 จึงสรุปได้ว่าตัวแปรอิสระไม่มีความสัมพันธ์กัน

เนื่องจากตัวแบบที่ได้ไม่ผ่านข้อสมมติเบื้องต้นของการทดสอบ ดังนั้นจะใช้การแปลงข้อมูลของตัวแปรตามด้วยวิธีของ Box-Cox Transformation จะได้ $Y' = \log Y$ เนื่องจากการแจกแจงของความคลาดเคลื่อนเบ้ขวา ตามรูปที่ 4.3 ดังนั้น หลังจากแปลงข้อมูลตัวแปรตามแล้วจึงนำไปวิเคราะห์การถดถอยเชิงเส้นพหุคูณอีกครั้ง จะได้ว่า



รูปที่ 4.3 แสดง Histogram ค่าเศษเหลือของสถานีการเคหะชุมชนแห่งชาติ

4.1.4 ผลการวิเคราะห์ข้อมูลโดยใช้วิธีการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ หลังจากการแปลงข้อมูลตัวแปรตาม $Y' = \log Y$

ตรวจสอบลักษณะความสัมพันธ์ของตัวแปรตาม Y' กับตัวแปรอิสระ F ทีละคู่ พบว่า ตัวแปรตาม Y' ที่มีความสัมพันธ์กับตัวแปรอิสระ F เป็นเส้นตรง คือ Y' กับ F_1, Y' กับ F_3 และ Y' กับ F_4 ส่วนตัวแปรตาม Y' ที่มีความสัมพันธ์กับตัวแปรอิสระ F ไม่เป็นเส้นตรง จะต้องทำการแปลงข้อมูล นั่นคือ Y' กับ F_2 แปลง F_2 เป็นแบบ Cubic และ Y' กับ F_5 แปลง F_5 เป็นแบบ Cubic

1) วิธีการเพิ่มตัวแปรอิสระ (Forward Selection Procedure)

ผลการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยด้วยวิธีการเพิ่มตัวแปรอิสระ แสดงดังตารางที่ 4.7

ตารางที่ 4.7 แสดงผลการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยโดยวิธีการเพิ่มตัวแปรอิสระของตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีการเคหะชุมชนแห่งชาติ หลังจากการแปลงข้อมูลตัวแปรตาม

Model		Unstandardized Coefficients		t	Sig.	Collinearity Statistics
		B	Std. Error			VIF
1	(Constant)	1.715	.007	234.209	.000	
	FAC1_1	.064	.006	10.060	.000	1.170
	FAC2_1	-.010	.008	-1.272	.002	1.706
	FAC3_1	.054	.007	8.031	.000	1.032
	FAC4_1	-.041	.006	-6.360	.000	1.023
	FAC5_1	.051	.008	6.646	.000	1.103
	FAC2_2	.006	.004	1.453	.001	2.018

a. Dependent Variable:

ผลจากตารางที่ 4.7 พบว่าตัวแปรที่มีผลต่อดัชนีคุณภาพอากาศ คือ ตัวแปรในปัจจัยร่วมที่ 1 (FAC1_1), ตัวแปรในปัจจัยร่วมที่ 2 (FAC2_1), ตัวแปรในปัจจัยร่วมที่ 2 ยกกำลังสอง (FAC2_2), ตัวแปรในปัจจัยร่วมที่ 3 (FAC3_1), ตัวแปรในปัจจัยร่วมที่ 4 (FAC4_1) และตัวแปรในปัจจัยร่วมที่ 5 (FAC5_1) ได้สมการการถดถอย ดังนี้

$$\text{LOG } \hat{Y} = 1.715 + 0.064F_1 - 0.010F_2 + 0.006F_2^2 + 0.054F_3 - 0.041F_4 + 0.051F_5$$

และมีค่า $R^2 = 0.511$

2) วิธีการลดตัวแปรอิสระ (Backward Elimination Procedure)

ผลการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยด้วยวิธีการลดตัวแปรอิสระ ได้ผลเหมือนกับวิธีการเพิ่มตัวแปรอิสระ

3) วิธีการเพิ่มตัวแปรอิสระทีละขั้นตอน (Stepwise Regression Procedure)

ผลการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยด้วยวิธีการเพิ่มตัวแปรอิสระทีละขั้นตอน ได้ผลเหมือนกับวิธีการเพิ่มตัวแปรอิสระ

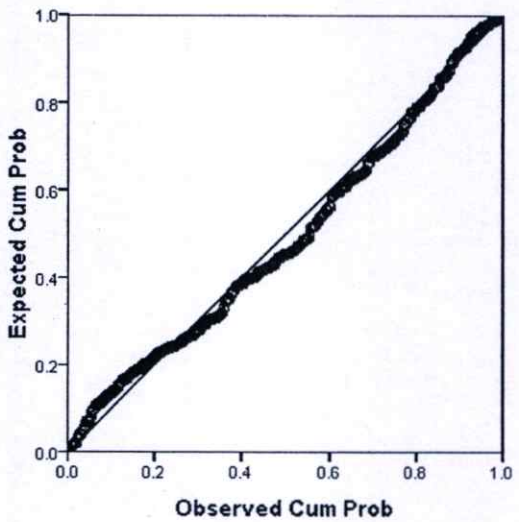
จากวิธีการคัดเลือกตัวแปรทั้ง 3 วิธี จะเลือกสมการที่ดีที่สุดด้วยวิธีการเพิ่มตัวแปรอิสระทีละขั้นตอน เพราะให้ค่า R^2 สูงที่สุด ดังนั้นสมการถดถอยที่ดีที่สุดสำหรับการพยากรณ์ คือ

$$LOG\ \hat{Y} = 1.715 + 0.064F_1 - 0.010F_2 + 0.006F_2^2 + 0.054F_3 - 0.041F_4 + 0.051F_5$$

4.1.5 ผลการตรวจสอบข้อสมมติของค่าความคลาดเคลื่อน หลังจากการแปลงข้อมูลตัวแปรตาม $Y' = \log Y$

1) การตรวจสอบค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติ

1.1) การพล๊อตกราฟ Normal Probability Plot



รูปที่ 4.4 กราฟ Normal Probability Plot ของค่าเศษเหลือของสมการถดถอยสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีการเคหะชุมชนแห่งชาติ หลังจากการแปลงข้อมูลตัวแปรตาม

จากรูปที่ 4.4 พบว่ากราฟที่ได้เป็นเส้นตรง จึงสรุปว่าค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติ

1.2) ผลการทดสอบของ Kolmogorov-Smirnov ที่ปรับปรุงด้วยการทดสอบของ Lilliefors ได้ผลดังนี้

สมมติฐานของการทดสอบ คือ

H_0 : ค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติ

H_1 : ค่าความคลาดเคลื่อนไม่มีการแจกแจงแบบปกติ

ตารางที่ 4.8 แสดงการทดสอบค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติของสถานีการเคหะชุมชนแห่งชาติ หลังจากการแปลงข้อมูลตัวแปรตาม

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Unstandardized Residual	.064	293	.036	.986	293	.006

a. Lilliefors Significance Correction

สถิติทดสอบ Kolmogorov-Smirnov

$$D = \max\left|F(x)-S(x)\right| = 0.064$$

ค่า p-value = 0.036 มีค่ามากกว่า $\alpha = 0.01$ ดังนั้นจึงยอมรับ H_0 สรุปได้ว่าค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติ

2) การตรวจสอบความเป็นอิสระกันของค่าความคลาดเคลื่อน

สมมติฐานของการทดสอบ คือ

H_0 : ค่าความคลาดเคลื่อนเป็นอิสระกัน

H_1 : ค่าความคลาดเคลื่อนไม่เป็นอิสระกัน

ตารางที่ 4.9 แสดงค่า Durbin-Watson ที่ได้จากการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยด้วยวิธีการเพิ่มตัวแปรอิสระทีละขั้นตอนของสถานีการเคหะชุมชนแห่งชาติ หลังจากการแปลงข้อมูลตัวแปรตาม

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.715[a]	.511	.501	.10128364	1.508

a. Predictors: (Constant), FAC2_2, FAC4_1, FAC3_1, FAC5_1, FAC1_1, FAC2_1

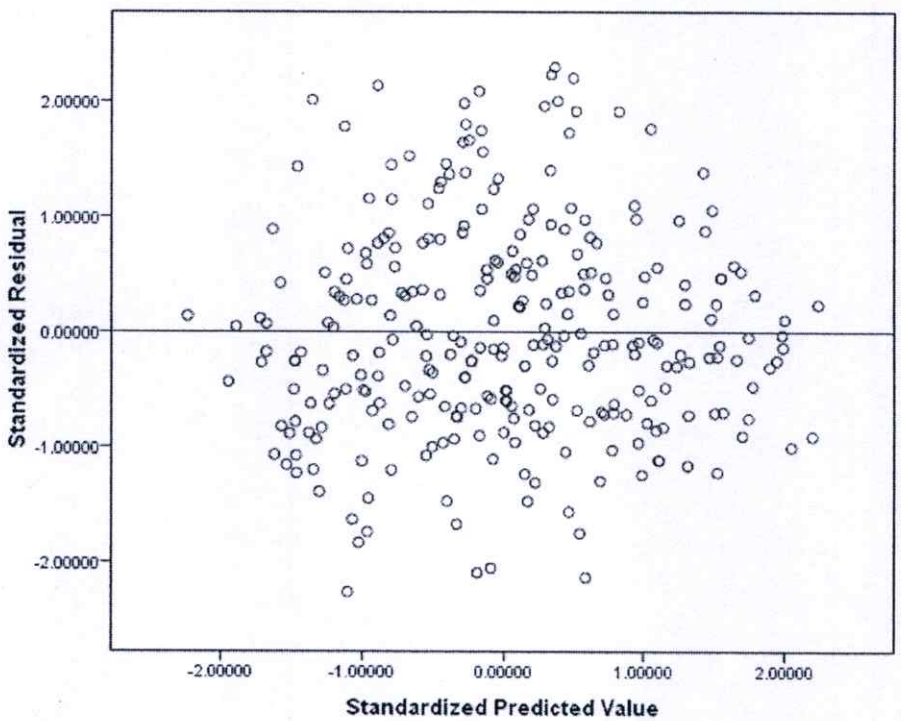
b. Dependent Variable: LOGY

สถิติทดสอบ

$$d = \frac{\sum_{i=2}^{n}(e_i - e_{i-1})}{\sum_{i=1}^{n} e_i^2} = 1.508$$

ค่า Durbin-Watson = 1.508 (นั่นคือ มีค่าในช่วง 1.5 - 2.5) ดังนั้นจึงยอมรับ H_0 จะสรุปว่า e_i กับ e_j เป็นอิสระกัน

3) การตรวจสอบค่าความแปรปรวนของค่าความคลาดเคลื่อน โดยการพล๊อตกราฟระหว่าง $E(e)$ กับ $E(\hat{Y})$



รูปที่ 4.5 ความสัมพันธ์ระหว่าง $E(e)$ กับ $E(\hat{Y})$ ของสถานีการเคหะชุมชนแห่งชาติ หลังจากการแปลงข้อมูลตัวแปรตาม

จากรูปที่ 4.5 พบว่าค่า e_i มีการกระจายอยู่รอบค่าศูนย์ และขนานไปกับแกน X สรุปได้ว่าค่าความแปรปรวนของค่าความคลาดเคลื่อนมีค่าคงที่

4) การตรวจสอบความเป็นอิสระกันของตัวแปรอิสระ

จากตารางที่ 4.7 พบว่า ค่า VIF ของตัวแปรอิสระทุกตัวมีค่าไม่เกิน 10 จึงสรุปได้ว่าตัวแปรอิสระไม่มีความสัมพันธ์กัน

เนื่องจากตัวแบบที่ได้ผ่านข้อสมมติเบื้องต้นของตัวแบบทุกข้อ ดังนั้นจะได้สมการถดถอยที่ดีที่สุดสำหรับการพยากรณ์ คือ

$$\text{LOG } \hat{Y} = 1.715 + 0.064F_1 - 0.010F_2 + 0.006F_2^2 + 0.054F_3 - 0.041F_4 + 0.051F_5 \quad (4.6)$$

และมีค่า $R^2 = 0.511$

จากสมการ 4.6 พบว่าตัวแปรอิสระที่มีผลต่อดัชนีคุณภาพอากาศ คือ ตัวแปรในปัจจัยร่วมที่ 1 (F_1), ตัวแปรในปัจจัยร่วมที่ 2 (F_2), ตัวแปรในปัจจัยร่วมที่ 2 ยกกำลังสอง (F_2^2), ตัวแปรในปัจจัยร่วมที่ 3 (F_3), ตัวแปรในปัจจัยร่วมที่ 4 (F_4) และตัวแปรในปัจจัยร่วมที่ 5 (F_5) ซึ่งสามารถอธิบายตัวแปรตามได้ 51.1%

4.2 ผลการวิเคราะห์ข้อมูลในการสร้างตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีห้วยขวาง

4.2.1 ผลการวิเคราะห์ข้อมูลโดยใช้วิธีการวิเคราะห์ปัจจัย

ผลการพิจารณาค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปรอิสระ พบว่าค่า $|r|$ มีค่ามากกว่า 0.5 จึงสรุปได้ว่าตัวแปรอิสระมีความสัมพันธ์กันสูง

ผลการพิจารณาค่า KMO (Kaiser-Meyer-Olkin) ซึ่งเป็นค่าที่ใช้วัดความเหมาะสมของข้อมูลที่จะนำมาวิเคราะห์ปัจจัย พบว่าค่า KMO = 0.677 ซึ่งมีค่ามากกว่า 0.5 จึงสรุปได้ว่าข้อมูลที่มีอยู่เหมาะสมที่จะใช้วิธีการวิเคราะห์ปัจจัย

สำหรับผลการวิเคราะห์ปัจจัยของตัวแปรอิสระ 19 ตัวแปร โดยวิธีวิเคราะห์องค์ประกอบหลัก แสดงดังตารางที่ 4.10

ตารางที่ 4.10 แสดงผลการวิเคราะห์ปัจจัยของตัวแปรอิสระสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีห้วยขวาง

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.972	26.168	26.168	4.972	26.168	26.168	2.935	15.448	15.448
2	2.967	15.617	41.785	2.967	15.617	41.785	2.839	14.941	30.390
3	1.966	10.349	52.134	1.966	10.349	52.134	2.508	13.197	43.587
4	1.199	6.309	58.444	1.199	6.309	58.444	2.475	13.028	56.615
5	1.083	5.702	64.145	1.083	5.702	64.145	1.431	7.530	64.145
6	.967	5.088	69.233						
7	.943	4.964	74.197						
8	.832	4.377	78.574						
9	.819	4.308	82.882						
10	.681	3.585	86.467						
11	.636	3.346	89.813						
12	.473	2.491	92.304						
13	.419	2.204	94.508						
14	.362	1.907	96.415						
15	.310	1.632	98.046						
16	.208	1.095	99.141						
17	.151	.796	99.938						
18	.011	.056	99.994						
19	.001	.006	100.000						

Extraction Method: Principal Component Analysis.

จากตารางที่ 4.10 พบว่าตัวแปรอิสระต่าง ๆ มีการจับกลุ่มรวมกันเป็นปัจจัยทั้งหมด 5 ปัจจัยและสามารถอธิบายความแปรปรวนรวมทั้งหมดของตัวแปรได้ร้อยละ 64.145

ทำการหมุนแกนปัจจัยด้วยวิธีวาริแมกซ์ แสดงดังตารางที่ 4.11

ตารางที่ 4.11 แสดงผลการหมุนแกนปัจจัยด้วยวิธีวาริแมกซ์ของสถานีห้วยขวาง

	Component				
	1	2	3	4	5
Zscore(X14)	.851				
Zscore(X8)	-.773		-.215	.245	
Zscore(X15)	-.734	.347		.282	.237
Zscore(X7)	.663	.354	-.280		
Zscore(X13)		.878			
Zscore(X9)	-.348	.838		.326	
Zscore(X12)		.767			
Zscore(X11)	.272	.537		-.245	.224
Zscore(X4)			.837	-.320	
Zscore(X3)			.833	-.323	
Zscore(X5)	.220		.644	.213	
Zscore(X6)	.287	-.314	.340	-.721	
Zscore(X10)	.229	.213	-.314	.703	
Zscore(X1)				-.605	
Zscore(X2)	.268		.466	-.515	
Zscore(X17)					-.598
Zscore(X18)				.304	.559
Zscore(X16)	-.363				.519
Zscore(X19)				.290	.425

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization

จากตารางที่ 4.11 การจัดตัวแปรอิสระให้อยู่ในปัจจัยใด จะพิจารณาจากขนาดของค่า Factor Loading ที่มีค่ามากที่สุด ซึ่งแต่ละปัจจัยประกอบด้วยตัวแปรต่าง ๆ ดังนี้

ปัจจัยที่ 1 F_1 จะสามารถอธิบายความแปรปรวนรวมทั้งหมดของตัวแปรได้ร้อยละ 15.488 ประกอบด้วยสมาชิกทั้งหมด 4 ตัวแปร ดังนี้

1. การระเหยของน้ำในแต่ละวัน (X_7)
2. ข้อมูลเมฆ (X_8)
3. ความยาวนานของแสงแดดในแต่ละวัน (X_{14})
4. ความชื้นสัมพัทธ์เฉลี่ยในแต่ละวัน (X_{15})

ปัจจัยที่ 2 F_2 จะสามารถอธิบายความแปรปรวนรวมทั้งหมดของตัวแปรได้ร้อยละ 14.941 ประกอบด้วยสมาชิกทั้งหมด 4 ตัวแปร ดังนี้

1. อุณหภูมิตุ้มเปียก (X_9)
2. อุณหภูมิสูงสุดในแต่ละวัน (X_{11})
3. อุณหภูมิต่ำสุดในแต่ละวัน (X_{12})
4. อุณหภูมิตุ้มแห้ง (X_{13})

ปัจจัยที่ 3 F_3 จะสามารถอธิบายความแปรปรวนรวมทั้งหมดของตัวแปรได้ร้อยละ 13.197 ประกอบด้วยสมาชิกทั้งหมด 3 ตัวแปร ดังนี้

1. ก๊าซคาร์บอนมอนอกไซด์ (CO) เฉลี่ย 1 ชั่วโมง (X_3)
2. ก๊าซคาร์บอนมอนอกไซด์ (CO) เฉลี่ย 8 ชั่วโมง (X_4)
3. ก๊าซซัลเฟอร์ไดออกไซด์ (SO_2) เฉลี่ย 24 ชั่วโมง (X_5)

ปัจจัยที่ 4 F_4 สามารถอธิบายความแปรปรวนรวมทั้งหมดของตัวแปรได้ร้อยละ 13.028 ประกอบด้วยสมาชิกทั้งหมด 4 ตัวแปร ดังนี้

1. ก๊าซโอโซน (O_3) เฉลี่ย 1 ชั่วโมง (X_1)
2. ก๊าซไนโตรเจนไดออกไซด์ (NO_2) เฉลี่ย 1 ชั่วโมง (X_2)
3. ฝุ่นละอองขนาดเล็กกว่า 10 ไมครอน (PM_{10}) เฉลี่ย 24 ชั่วโมง (X_6)
4. ทัศนวิสัยโดยเฉลี่ย (X_{10})

ปัจจัยที่ 5 F_5 จะสามารถอธิบายความแปรปรวนรวมทั้งหมดของตัวแปรได้ร้อยละ 7.530 ประกอบด้วยสมาชิกทั้งหมด 4 ตัวแปร ดังนี้

1. ปริมาณน้ำฝนในแต่ละวัน (X_{16})
2. ความกดอากาศเฉลี่ยในแต่ละวัน (X_{17})
3. ทิศทางลมในแต่ละวัน (X_{18})
4. ความเร็วลมสูงสุดในแต่ละวัน (X_{19})

ผลการหาสมการสำหรับคำนวณค่าคะแนนปัจจัย (factor score) ด้วยวิธีวิเคราะห์การถดถอย (Regression Method) ได้ผลดังนี้

กำหนดให้ F_1, F_2, F_3, F_4 และ F_5 คือ ค่าคะแนนปัจจัยของปัจจัยที่ 1, 2, 3, 4 และ 5 ตามลำดับ จะได้ว่า

$$F_1 = 0.663Z_7 - 0.773Z_8 + 0.851Z_{14} - 0.734Z_{15} \quad (4.7)$$

$$F_2 = 0.838Z_9 + 0.537Z_{11} + 0.767Z_{12} + 0.878Z_{13} \quad (4.8)$$

$$F_3 = 0.833Z_3 + 0.837Z_4 + 0.644Z_5 \quad (4.9)$$

$$F_4 = -0.605Z_1 - 0.515Z_2 - 0.721Z_6 + 0.703Z_{10} \quad (4.10)$$

$$F_5 = 0.519Z_{16} - 0.598Z_{17} + 0.559Z_{18} + 0.425Z_{19} \quad (4.11)$$

4.2.2 ผลการวิเคราะห์ข้อมูลโดยใช้วิธีการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ

ตรวจสอบลักษณะความสัมพันธ์ของตัวแปรตาม Y กับตัวแปรอิสระ F ทีละคู่ พบว่า ตัวแปรตาม Y ที่มีความสัมพันธ์กับตัวแปรอิสระ F เป็นเส้นตรง คือ Y กับ F_4 ส่วนตัวแปรตาม Y ที่มีความสัมพันธ์กับตัวแปรอิสระ F ไม่เป็นเส้นตรง จะต้องทำการแปลงข้อมูล นั่นคือ Y กับ F_1 แปลง F_1 เป็นแบบ Exponential, Y กับ F_2 แปลง F_2 เป็นแบบ Cubic, Y กับ F_3 แปลง F_3 เป็นแบบ Exponential และ Y กับ F_5 แปลง F_5 เป็นแบบ Cubic

นำค่าคะแนนปัจจัยที่ได้จากการวิเคราะห์ปัจจัยในข้อ 4.2.1 มาวิเคราะห์หาตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศ ประมาณค่าพารามิเตอร์ของการถดถอยด้วยวิธีกำลังสองน้อยที่สุด (Least Square Method) และเลือกสมการที่ดีที่สุดจากวิธีการคัดเลือกตัวแปรทั้ง 3 วิธี ดังนี้

1) วิธีการเพิ่มตัวแปรอิสระ (Forward Selection Procedure)

ผลการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยด้วยวิธีการเพิ่มตัวแปรอิสระ แสดงดังตารางที่ 4.12

ตารางที่ 4.12 แสดงผลการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยโดยวิธีการเพิ่มตัวแปรอิสระของตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีห้วยขวาง

Model		Unstandardized Coefficients		t	Sig.	Collinearity Statistics
		B	Std. Error			VIF
4	(Constant)	47.175	.824	57.230	.000	
	FAC4_1	-10.708	.524	-20.415	.000	1.020
	FAC1_EXP	2.200	.398	5.526	.000	1.016
	FAC3_EXP	.850	.143	5.966	.000	1.015
	FAC2_1	-2.415	.499	-4.837	.000	1.020

a. Dependent Variable: Y

ผลจากตารางที่ 4.12 พบว่าตัวแปรอิสระที่มีผลต่อดัชนีคุณภาพอากาศ คือ ตัวแปรในปัจจัยร่วมที่ 1 เป็นแบบเอกซ์โปเนนเชียล (FAC1_EXP), ตัวแปรในปัจจัยร่วมที่ 2 (FAC2_1), ตัวแปรในปัจจัยร่วมที่ 3 เป็นแบบเอกซ์โปเนนเชียล (FAC3_EXP) และตัวแปรในปัจจัยร่วมที่ 4 (FAC4_1) จะได้สมการการถดถอย ดังนี้

$$\hat{Y} = 47.175 + 2.200EXP(F_1) - 2.415F_2 + 0.850EXP(F_3) - 10.708F_4$$

และมีค่า $R^2 = 0.677$

2) วิธีการลดตัวแปรอิสระ (Backward Elimination Procedure)

ผลการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยด้วยวิธีการลดตัวแปรอิสระแสดงดังตารางที่ 4.13

ตารางที่ 4.13 แสดงผลการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยโดยวิธีการลดตัวแปรอิสระของตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีห้วยขวาง

Model		Unstandardized Coefficients		t	Sig.	Collinearity Statistics
		B	Std. Error			VIF
1	(Constant)	47.737	.935	51.064	.000	.
	FAC2_1	-2.280	.492	-4.637	.000	1.036
	FAC4_1	-10.980	.534	-20.557	.000	1.108
	FAC5_1	3.930	1.168	3.365	.001	3.217
	FAC1_EXP	2.025	.395	5.122	.000	1.049
	FAC3_EXP	.778	.141	5.517	.000	1.040
	FAC5_2	.588	.777	.757	.004	1.722
	FAC5_3	-2.201	.605	-3.636	.000	4.180

a. Dependent Variable: Y

ผลจากตารางที่ 4.13 พบว่าตัวแปรที่มีผลต่อดัชนีคุณภาพอากาศ คือ ตัวแปรในปัจจัยร่วมที่ 1 เป็นแบบเอกซ์โปเนนเชียล (FAC1_EXP), ตัวแปรในปัจจัยร่วมที่ 2 (FAC2_1), ตัวแปรในปัจจัยร่วมที่ 3 เป็นแบบเอกซ์โปเนนเชียล (FAC3_EXP), ตัวแปรในปัจจัยร่วมที่ 4 (FAC4_1), ตัวแปรในปัจจัยร่วมที่ 5 (FAC5_1), ตัวแปรในปัจจัยร่วมที่ 5 ยกกำลังสอง (FAC5_2) และตัวแปรในปัจจัยร่วมที่ 5 ยกกำลังสาม (FAC5_3) ได้สมการการถดถอย ดังนี้

$$\hat{Y} = 47.737+2.025EXP(F_1)-2.280F_2+0.778EXP(F_3)-10.980F_4+3.930F_5+0.588F_5^2-2.201F_5^3$$

และมีค่า $R^2 = 0.695$

3) วิธีการเพิ่มตัวแปรอิสระทีละขั้นตอน (Stepwise Regression Procedure)

ผลการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยด้วยวิธีการเพิ่มตัวแปรอิสระทีละขั้นตอน ได้ผลเหมือนกับวิธีเพิ่มตัวแปรอิสระ

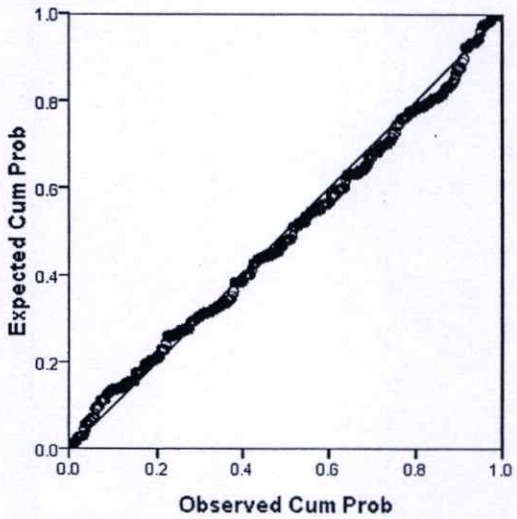
จากวิธีการคัดเลือกตัวแปรทั้ง 3 วิธี จะเลือกสมการที่ดีที่สุดด้วยวิธีการลดตัวแปรอิสระ เพราะให้ค่า R^2 สูงที่สุด ดังนั้นสมการถดถอยที่ดีที่สุดสำหรับการพยากรณ์ คือ

$$\hat{Y} = 47.737+2.025EXP(F_1)-2.280F_2+0.778EXP(F_3)-10.980F_4+3.930F_5+0.588F_5^2-2.201F_5^3$$

4.2.3 ผลการตรวจสอบข้อสมมติของค่าความคลาดเคลื่อน

1) การตรวจสอบค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติ

1.1) การพล๊อตกราฟ Normal Probability Plot



รูปที่ 4.6 กราฟ Normal Probability Plot ของค่าเศษเหลือของสมการถดถอยสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีห้วยขวาง

จากรูปที่ 4.6 พบว่ากราฟที่ได้เป็นเส้นตรง จึงสรุปว่าค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติ

1.2) ผลการทดสอบของ Kolmogorov-Smirnov ที่ปรับปรุงด้วยการทดสอบของ Lilliefors ได้ผลดังนี้

สมมติฐานของการทดสอบ คือ

H_0 : ค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติ

H_1 : ค่าความคลาดเคลื่อนไม่มีการแจกแจงแบบปกติ

ตารางที่ 4.14 แสดงการทดสอบค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติของสถานีห้วยขวาง

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Unstandardized Residual	.046	276	.200*	.984	276	.004

a. Lilliefors Significance Correction

*. This is a lower bound of the true significance.

สถิติทดสอบ Kolmogorov-Smirnov

$$D = \max\left|F(x)-S(x)\right| = 0.046$$

ค่า p-value = 0.200 มีค่ามากกว่า $\alpha = 0.01$ ดังนั้นจึงยอมรับ H_0 สรุปได้ว่าค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติ

2) การตรวจสอบความเป็นอิสระกันของค่าความคลาดเคลื่อน

สมมติฐานของการทดสอบ คือ

H_0 : ค่าความคลาดเคลื่อนเป็นอิสระกัน

H_1 : ค่าความคลาดเคลื่อนไม่เป็นอิสระกัน

ตารางที่ 4.15 แสดงค่า Durbin-Watson ที่ได้จากการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยด้วยวิธีการลดตัวแปรอิสระของสถานีห้วยขวาง

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.833[a]	.695	.687	7.86623	1.583

a. Predictors: (Constant), FAC5_3, FAC2_1, FAC3_EX, FAC1_EX, FAC4_1, FAC5_2, FAC5_1

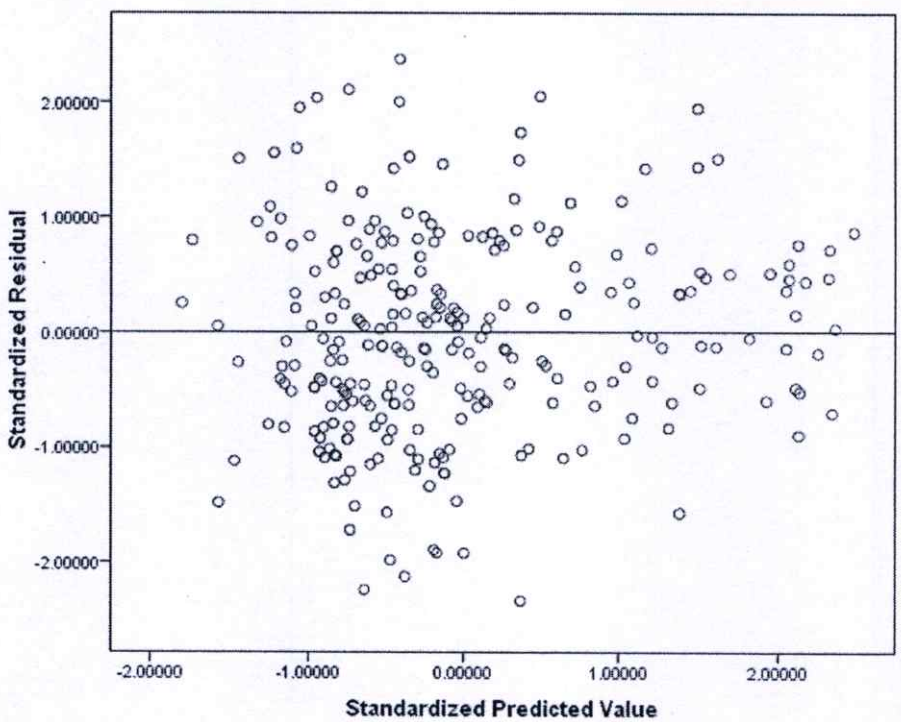
b. Dependent Variable: Y

สถิติทดสอบ

$$d = \frac{\sum_{i=2}^{n}(e_i - e_{i-1})}{\sum_{i=1}^{n} e_i^2} = 1.583$$

ค่า Durbin-Watson = 1.583 (นั่นคือ มีค่าในช่วง 1.5 - 2.5) ดังนั้นจึงยอมรับ H_0 จะสรุปว่า e_i กับ e_j เป็นอิสระกัน

3) การตรวจสอบค่าความแปรปรวนของค่าความคลาดเคลื่อน โดยการพล๊อตกราฟระหว่าง $E(e)$ กับ $E(\hat{Y})$



รูปที่ 4.7 ความสัมพันธ์ระหว่าง $E(e)$ กับ $E(\hat{Y})$ ของสถานีห้วยขวาง

จากรูปที่ 4.7 พบว่ามีการกระจายอยู่รอบค่าศูนย์ และขนานไปกับแกน X สรุปได้ว่าค่าความแปรปรวนของค่าความคลาดเคลื่อนมีค่าคงที่

4) การตรวจสอบความเป็นอิสระกันของตัวแปรอิสระ

จากตารางที่ 4.13 พบว่า ค่า VIF ของตัวแปรอิสระทุกตัวมีค่าไม่เกิน 10 จึงสรุปได้ว่าตัวแปรอิสระไม่มีความสัมพันธ์กัน

เนื่องจากตัวแบบที่ได้ผ่านข้อสมมติเบื้องต้นของตัวแบบทุกข้อ ดังนั้นจะได้สมการถดถอยที่ดีที่สุดสำหรับการพยากรณ์ คือ

$$\hat{Y} = 47.737+2.025EXP(F_1)-2.280F_2+0.778EXP(F_3)-10.980F_4+3.930F_5+0.588F_5^2-2.201F_5^3 \quad (4.12)$$

และมีค่า $R^2 = 0.695$

จากสมการ 4.12 พบว่าตัวแปรอิสระที่มีผลต่อดัชนีคุณภาพอากาศ คือ ตัวแปรในปัจจัยร่วมที่ 1 เป็นแบบเอกซ์โปเนนเชียล ($EXP(F_1)$), ตัวแปรในปัจจัยร่วมที่ 2 (F_2), ตัวแปรในปัจจัยร่วมที่ 3 เป็นแบบเอกซ์โปเนนเชียล ($EXP(F_3)$), ตัวแปรในปัจจัยร่วมที่ 4 (F_4), ตัวแปรในปัจจัยร่วมที่ 5 (F_5), ตัวแปรในปัจจัยร่วมที่ 5 ยกกำลังสอง (F_5^2) และตัวแปรในปัจจัยร่วมที่ 5 ยกกำลังสาม (F_5^3) ซึ่งสามารถอธิบายตัวแปรตามได้ 69.5%

4.3 ผลการวิเคราะห์ข้อมูลในการสร้างตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานียานนาวา

4.3.1 ผลการวิเคราะห์ข้อมูลโดยใช้วิธีการวิเคราะห์ปัจจัย

ผลการพิจารณาค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปรอิสระ พบว่าค่า $|r|$ มีค่ามากกว่า 0.5 จึงสรุปได้ว่าตัวแปรอิสระมีความสัมพันธ์กันสูง

ผลการพิจารณาค่า KMO (Kaiser-Meyer-Olkin) ซึ่งเป็นค่าที่ใช้วัดความเหมาะสมของข้อมูลที่จะนำมาวิเคราะห์ปัจจัย พบว่าค่า KMO = 0.732 ซึ่งมีค่ามากกว่า 0.5 จึงสรุปได้ว่าข้อมูลที่มีอยู่เหมาะสมที่จะใช้วิธีการวิเคราะห์ปัจจัย

สำหรับผลการวิเคราะห์ปัจจัยของตัวแปรอิสระ 19 ตัวแปร โดยวิธีวิเคราะห์องค์ประกอบหลัก แสดงดังตารางที่ 4.16

ตารางที่ 4.16 แสดงผลการวิเคราะห์ปัจจัยของตัวแปรอิสระสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานียานนาวา

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5.709	30.047	30.047	5.709	30.047	30.047	4.278	22.514	22.514
2	2.916	15.350	45.397	2.916	15.350	45.397	2.969	15.624	38.138
3	1.690	8.896	54.293	1.690	8.896	54.293	2.664	14.019	52.157
4	1.146	6.034	60.328	1.146	6.034	60.328	1.417	7.456	59.613
5	1.058	5.570	65.897	1.058	5.570	65.897	1.194	6.284	65.897
6	.955	5.025	70.922						
7	.924	4.864	75.786						
8	.839	4.418	80.204						
9	.799	4.204	84.409						
10	.691	3.637	88.046						
11	.514	2.703	90.749						
12	.396	2.082	92.831						
13	.371	1.953	94.784						
14	.321	1.690	96.474						
15	.299	1.574	98.048						
16	.198	1.042	99.089						
17	.159	.837	99.926						
18	.013	.067	99.993						
19	.001	.007	100.000						

Extraction Method: Principal Component Analysis.

จากตารางที่ 4.16 พบว่าตัวแปรอิสระต่าง ๆ มีการจับกลุ่มรวมกันเป็นปัจจัยทั้งหมด 5 ปัจจัยและสามารถอธิบายความแปรปรวนรวมทั้งหมดของตัวแปรได้ร้อยละ 65.897

ทำการหมุนแกนปัจจัยด้วยวิธีวาริแมกซ์ แสดงดังตารางที่ 4.17

ตารางที่ 4.17 แสดงผลการหมุนแกนปัจจัยด้วยวิธีวาริแมกซ์ของสถานียานนาวา

	Component				
	1	2	3	4	5
Zscore(X6)	.813			-.211	
Zscore(X4)	.809				.225
Zscore(X3)	.804				.218
Zscore(X2)	.752	.270			
Zscore(X10)	-.710	.342			.261
Zscore(X1)	.578	.275	-.207	-.264	
Zscore(X14)	.210	.816			
Zscore(X15)	-.230	-.751	.358	.273	
Zscore(X7)	-.233	.733	.318		
Zscore(X8)	-.393	-.720			
Zscore(X13)	-.335		.821		
Zscore(X9)	-.398	-.351	.790		
Zscore(X12)	-.267		.718		
Zscore(X11)		.272	.570		-.242
Zscore(X17)				-.557	
Zscore(X19)	-.303			.546	
Zscore(X18)				.545	
Zscore(X16)		-.306		.399	-.649
Zscore(X5)	.363		.230		.605

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.

จากตารางที่ 4.17 การจัดตัวแปรอิสระให้อยู่ในปัจจัยใด จะพิจารณาจากขนาดของค่า Factor Loading ที่มีค่าที่สุด และหากมีค่า Factor Loading เพียงค่าเดียวในปัจจัยใด ให้ถือว่าตัวแปรอิสระควรอยู่ในปัจจัยนั้นซึ่งแต่ละปัจจัยประกอบด้วยตัวแปรต่าง ๆ ดังนี้

ปัจจัยที่ 1 F_1 สามารถอธิบายความแปรปรวนรวมทั้งหมดของตัวแปรได้ร้อยละ 22.514 ประกอบด้วยสมาชิกทั้งหมด 6 ตัวแปร ดังนี้

1. ก๊าซโอโซน (O_3) เฉลี่ย 1 ชั่วโมง (X_1)
2. ก๊าซไนโตรเจนไดออกไซด์ (NO_2) เฉลี่ย 1 ชั่วโมง (X_2)
3. ก๊าซคาร์บอนมอนอกไซด์ (CO) เฉลี่ย 1 ชั่วโมง (X_3)
4. ก๊าซคาร์บอนมอนอกไซด์ (CO) เฉลี่ย 8 ชั่วโมง (X_4)
5. ฝุ่นละอองขนาดเล็กกว่า 10 ไมครอน (PM_{10}) เฉลี่ย 24 ชั่วโมง (X_6)
6. ทัศนวิสัยโดยเฉลี่ย (X_{10})

ปัจจัยที่ 2 F_2 สามารถอธิบายความแปรปรวนรวมทั้งหมดของตัวแปรได้ร้อยละ 15.624 ประกอบด้วยสมาชิกทั้งหมด 4 ตัวแปร ดังนี้

1. การระเหยของน้ำในแต่ละวัน (X_7)
2. ข้อมูลเมฆ (X_8)
3. ความยาวนานของแสงแดดในแต่ละวัน (X_{14})
4. ความชื้นสัมพัทธ์เฉลี่ยในแต่ละวัน (X_{15})

ปัจจัยที่ 3 F_3 สามารถอธิบายความแปรปรวนรวมทั้งหมดของตัวแปรได้ร้อยละ 14.019 ประกอบด้วยสมาชิกทั้งหมด 4 ตัวแปร ดังนี้

1. อุณหภูมิตุ้มเปียก (X_9)
2. อุณหภูมิสูงสุดในแต่ละวัน (X_{11})
3. อุณหภูมิต่ำสุดในแต่ละวัน (X_{12})
4. อุณหภูมิตุ้มแห้ง (X_{13})

ปัจจัยที่ 4 F_4 สามารถอธิบายความแปรปรวนรวมทั้งหมดของตัวแปรได้ร้อยละ 7.456 ประกอบด้วยสมาชิกทั้งหมด 3 ตัวแปร ดังนี้

1. ความกดอากาศเฉลี่ยในแต่ละวัน (X_{17})
2. ทิศทางลมในแต่ละวัน (X_{18})
3. ความเร็วลมสูงสุดในแต่ละวัน (X_{19})

ปัจจัยที่ 5 F_5 สามารถอธิบายความแปรปรวนรวมทั้งหมดของตัวแปรได้ร้อยละ 6.284 ประกอบด้วยสมาชิกทั้งหมด 2 ตัวแปร ดังนี้

1. ก๊าซซัลเฟอร์ไดออกไซด์ (SO_2) เฉลี่ย 24 ชั่วโมง (X_5)
2. ปริมาณน้ำฝนในแต่ละวัน (X_{16})

ผลการหาสมการสำหรับคำนวณค่าคะแนนปัจจัย (factor score) ด้วยวิธีวิเคราะห์การถดถอย (Regression Method) ได้ผลดังนี้

กำหนดให้ F_1, F_2, F_3, F_4 และ F_5 คือ ค่าคะแนนปัจจัยของปัจจัยที่ 1, 2, 3, 4 และ 5 ตามลำดับ จะได้ว่า

$$F_1 = 0.578Z_1 + 0.752Z_2 + 0.804Z_3 + 0.809Z_4 + 0.813Z_6 - 0.710Z_{10} \quad (4.13)$$

$$F_2 = 0.733Z_7 - 0.720Z_8 + 0.816Z_{14} - 0.751Z_{15} \quad (4.14)$$

$$F_3 = 0.790Z_9 + 0.570Z_{11} + 0.718Z_{12} + 0.821Z_{13} \quad (4.15)$$

$$F_4 = -0.557Z_{17} + 0.545Z_{18} + 0.546Z_{19} \quad (4.16)$$

$$F_5 = 0.605Z_5 - 0.649Z_{16} \quad (4.17)$$

4.3.2 ผลการวิเคราะห์ข้อมูลโดยใช้วิธีการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ

ตรวจสอบลักษณะความสัมพันธ์ของตัวแปรตาม Y กับตัวแปรอิสระ F ทีละคู่ พบว่า ตัวแปรตาม Y ที่มีความสัมพันธ์กับตัวแปรอิสระ F เป็นเส้นตรง คือ Y กับ F_1 ส่วนตัวแปรตาม Y ที่มีความสัมพันธ์กับตัวแปรอิสระ F ไม่เป็นเส้นตรง จะต้องทำการแปลงข้อมูล นั่นคือ Y กับ F_2 แปลง F_2 เป็นแบบ Cubic, Y กับ F_3 แปลง F_3 เป็นแบบ Cubic, Y กับ F_4 แปลง F_4 เป็นแบบ Cubic และ Y กับ F_5 แปลง F_5 เป็นแบบ Cubic

นำค่าคะแนนปัจจัยที่ได้จากการวิเคราะห์ปัจจัยในข้อ 4.3.1 มาวิเคราะห์หาตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศ ประมาณค่าพารามิเตอร์ของการถดถอยด้วยวิธีกำลังสองน้อยที่สุด (Least Square Method) และเลือกสมการที่ดีที่สุดจากวิธีการคัดเลือกตัวแปรทั้ง 3 วิธี ดังนี้

1) วิธีการเพิ่มตัวแปรอิสระ (Forward Selection Procedure)

ผลการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยด้วยวิธีการเพิ่มตัวแปรอิสระ แสดงดังตารางที่ 4.18

ตารางที่ 4.18 แสดงผลการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยโดยวิธีการเพิ่มตัวแปรอิสระของตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานียานนาวา

Model		Unstandardized Coefficients		t	Sig.	Collinearity Statistics
		B	Std. Error			VIF
1	(Constant)	57.963	.793	73.105	.000	
	FAC1_1	13.187	.514	25.673	.000	1.030
	FAC2_1	.897	.979	.917	.003	3.809
	FAC3_1	-2.278	.505	-4.507	.000	1.017
	FAC4_1	-2.772	.608	-4.560	.000	1.027
	FAC2_2	.322	.561	.574	.005	1.893
	FAC2_3	.583	.381	1.533	.001	5.318
	FAC4_2	1.218	.438	2.780	.006	1.056

a. Dependent Variable: Y

ผลจากตารางที่ 4.18 พบว่าตัวแปรที่มีผลต่อดัชนีคุณภาพอากาศ คือ ตัวแปรในปัจจัยร่วมที่ 1 (FAC1_1), ตัวแปรในปัจจัยร่วมที่ 2 (FAC2_1), ตัวแปรในปัจจัยร่วมที่ 2 ยกกำลังสอง (FAC2_2), ตัวแปรในปัจจัยร่วมที่ 2 ยกกำลังสาม (FAC2_3), ตัวแปรในปัจจัยร่วมที่ 3 (FAC3_1), ตัวแปรในปัจจัยร่วมที่ 4 (FAC4_1) และตัวแปรในปัจจัยร่วมที่ 4 ยกกำลังสอง (FAC4_2) จะได้สมการการถดถอย ดังนี้

$$\hat{Y} = 57.963 + 13.187F_1 + 0.897F_2 + 0.322F_2^2 + 0.583F_2^3 - 2.278F_3 - 2.772F_4 + 1.218F_4^2$$

และมีค่า $R^2 = 0.716$

2) วิธีการลดตัวแปรอิสระ (Backward Elimination Procedure)

ผลการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยด้วยวิธีการลดตัวแปรอิสระ ได้ผลเหมือนกับวิธีการเพิ่มตัวแปรอิสระ

3) วิธีการเพิ่มตัวแปรอิสระทีละขั้นตอน (Stepwise Regression Procedure)

ผลการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยด้วยวิธีการเพิ่มตัวแปรอิสระทีละขั้นตอน ได้ผลเหมือนกับวิธีการเพิ่มตัวแปรอิสระ

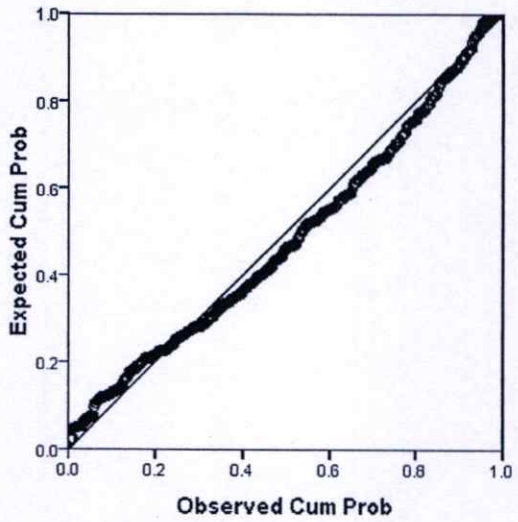
จากวิธีการคัดเลือกตัวแปรทั้ง 3 วิธี จะเลือกสมการที่ดีที่สุดด้วยวิธีการเพิ่มตัวแปรอิสระทีละขั้นตอน เพราะให้ค่า R^2 สูงที่สุด ดังนั้นสมการถดถอยที่ดีที่สุดสำหรับการพยากรณ์ คือ

$$\hat{Y} = 57.963 + 13.187F_1 + 0.897F_2 + 0.322F_2^2 + 0.583F_2^3 - 2.278F_3 - 2.772F_4 + 1.218F_4^2$$

4.3.3 ผลการตรวจสอบข้อสมมติของค่าความคลาดเคลื่อน

1) การตรวจสอบค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติ

1.1) การพล๊อตกราฟ Normal Probability Plot



รูปที่ 4.8 กราฟ Normal Probability Plot ของค่าเศษเหลือของสมการถดถอยสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานียานนาวา

จากรูปที่ 4.8 พบว่ากราฟที่ได้เป็นเส้นตรง จึงสรุปว่าค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติ

1.2) ผลการทดสอบของ Kolmogorov-Smirnov ที่ปรับปรุงด้วยการทดสอบของ Lilliefors ได้ผลดังนี้

สมมติฐานของการทดสอบ คือ

H_0 : ค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติ

H_1 : ค่าความคลาดเคลื่อนไม่มีการแจกแจงแบบปกติ

ตารางที่ 4.19 แสดงการทดสอบค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติของสถานียานนาวา

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Unstandardized Residual	.044	305	.200*	.986	305	.004

a. Lilliefors Significance Correction

*. This is a lower bound of the true significance.

สถิติทดสอบ Kolmogorov-Smirnov

$$D = \max\left|F(x)\text{-}S(x)\right| = 0.044$$

ค่า p-value = 0.200 มีค่ามากกว่า $\alpha = 0.01$ ดังนั้นจึงยอมรับ H_0 สรุปได้ว่าค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติ

2) การตรวจสอบความเป็นอิสระกันของค่าความคลาดเคลื่อน

สมมติฐานของการทดสอบ คือ

H_0 : ค่าความคลาดเคลื่อนเป็นอิสระกัน

H_1 : ค่าความคลาดเคลื่อนไม่เป็นอิสระกัน

ตารางที่ 4.20 แสดงค่า Durbin-Watson ที่ได้จากการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยด้วยวิธีเพิ่มตัวแปรอิสระทีละขั้นตอนของสถานียานนาวา

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.846[a]	.716	.710	8.85896	1.595

a. Predictors: (Constant), FAC4_2, FAC2_1, FAC3_1, FAC1_1, FAC4_1, FAC2_2,

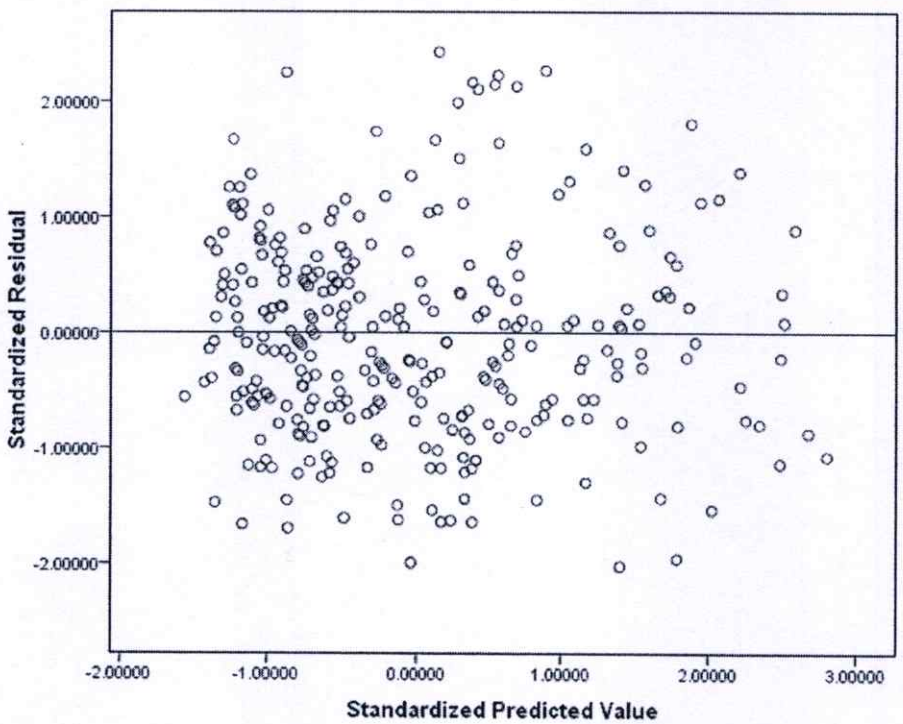
b. Dependent Variable: Y

สถิติทดสอบ

$$d = \frac{\sum_{i=2}^{n}(e_i - e_{i-1})}{\sum_{i=1}^{n} e_i^2} = 1.595$$

ค่า Durbin-Watson = 1.595 (นั่นคือ มีค่าในช่วง 1.5 - 2.5) ดังนั้นจึงยอมรับ H_0 จะสรุปว่า e_i กับ e_j เป็นอิสระกัน

3) การตรวจสอบค่าความแปรปรวนของค่าความคลาดเคลื่อน โดยการพล๊อตกราฟระหว่าง $E(e)$ กับ $E(\hat{Y})$



รูปที่ 4.9 ความสัมพันธ์ระหว่าง $E(e)$ กับ $E(\hat{Y})$ ของสถานียานนาวา

จากรูปที่ 4.9 พบว่าค่า e_i มีการกระจายอยู่รอบค่าศูนย์ และขนานไปกับแกน X สรุปได้ว่าค่าความแปรปรวนของค่าความคลาดเคลื่อนมีค่าคงที่

4) การตรวจสอบความเป็นอิสระกันของตัวแปรอิสระ

จากตารางที่ 4.18 พบว่า ค่า VIF ของตัวแปรอิสระทุกตัวมีค่าไม่เกิน 10 จึงสรุปได้ว่าตัวแปรอิสระไม่มีความสัมพันธ์กัน

เนื่องจากตัวแบบที่ได้ผ่านข้อสมมติเบื้องต้นของตัวแบบทุกข้อ ดังนั้นจะได้สมการถดถอยที่ดีที่สุดสำหรับการพยากรณ์ คือ

$$\hat{Y} = 57.963 + 13.187F_1 + 0.897F_2 + 0.322F_2^2 + 0.583F_2^3 - 2.278F_3 - 2.772F_4 + 1.218F_4^2 \quad (4.18)$$

และมีค่า $R^2 = 0.716$

จากสมการ 4.18 พบว่าตัวแปรอิสระที่มีผลต่อดัชนีคุณภาพอากาศ คือ ตัวแปรในปัจจัยร่วมที่ 1 (F_1), ตัวแปรในปัจจัยร่วมที่ 2 (F_2), ตัวแปรในปัจจัยร่วมที่ 2 ยกกำลังสอง (F_2^2), ตัวแปรในปัจจัยร่วมที่ 2 ยกกำลังสาม (F_2^3), ตัวแปรในปัจจัยร่วมที่ 3 (F_3), ตัวแปรในปัจจัยร่วมที่ 4 (F_4) และตัวแปรในปัจจัยร่วมที่ 4 ยกกำลังสอง (F_4^2) ซึ่งสามารถอธิบายตัวแปรตามได้ 71.6%

4.4 ผลการวิเคราะห์ข้อมูลในการสร้างตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีบางขุนเทียน

4.4.1 ผลการวิเคราะห์ข้อมูลโดยใช้วิธีการวิเคราะห์ปัจจัย

ผลการพิจารณาค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปรอิสระ พบว่าค่า $|r|$ มีค่ามากกว่า 0.5 จึงสรุปได้ว่าตัวแปรอิสระมีความสัมพันธ์กันสูง

ผลการพิจารณาค่า KMO (Kaiser-Meyer-Olkin) ซึ่งเป็นค่าที่ใช้วัดความเหมาะสมของข้อมูลที่จะนำมาวิเคราะห์ปัจจัย พบว่าค่า KMO = 0.716 ซึ่งมีค่ามากกว่า 0.5 จึงสรุปได้ว่าข้อมูลที่มีอยู่เหมาะสมที่จะใช้วิธีการวิเคราะห์ปัจจัย

สำหรับผลการวิเคราะห์ปัจจัยของตัวแปรอิสระ 19 ตัวแปร โดยวิธีวิเคราะห์องค์ประกอบหลัก แสดงดังตารางที่ 4.21

ตารางที่ 4.21 แสดงผลการวิเคราะห์ปัจจัยของตัวแปรอิสระสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีบางขุนเทียน

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5.481	28.850	28.850	5.481	28.850	28.850	3.542	18.644	18.644
2	3.086	16.243	45.093	3.086	16.243	45.093	2.963	15.593	34.237
3	1.518	7.988	53.081	1.518	7.988	53.081	2.887	15.194	49.431
4	1.160	6.106	59.188	1.160	6.106	59.188	1.854	9.757	59.188
5	.985	5.186	64.373						
6	.965	5.080	69.453						
7	.917	4.829	74.282						
8	.840	4.422	78.703						
9	.804	4.233	82.936						
10	.675	3.552	86.488						
11	.581	3.057	89.545						
12	.484	2.549	92.094						
13	.408	2.147	94.241						
14	.359	1.890	96.131						
15	.281	1.481	97.612						
16	.188	.991	98.603						
17	.136	.714	99.318						
18	.128	.676	99.994						
19	.001	.006	100.000						

Extraction Method: Principal Component

จากตารางที่ 4.21 พบว่าตัวแปรอิสระต่าง ๆ มีการจับกลุ่มรวมกันเป็นปัจจัยทั้งหมด 4 ปัจจัย และสามารถอธิบายความแปรปรวนรวมทั้งหมดของตัวแปรได้ร้อยละ 59.188

ทำการหมุนแกนปัจจัยด้วยวิธีวาริแมกซ์ แสดงดังตารางที่ 4.22

ตารางที่ 4.22 แสดงผลการหมุนแกนปัจจัยด้วยวิธีวาริแมกซ์ของสถานีบางขุนเทียน

	Component			
	1	2	3	4
Zscore(X6)	.824	-.242	.204	-.268
Zscore(X3)	.790	-.320		
Zscore(X2)	.746	-.377	.252	
Zscore(X10)	-.720		.350	
Zscore(X4)	.549		.241	
Zscore(X5)	.319			
Zscore(X13)	-.278	.860		
Zscore(X9)	-.377	.794	-.344	
Zscore(X12)	-.208	.744		
Zscore(X11)		.648	.239	
Zscore(X14)			.807	-.210
Zscore(X7)	-.222	.362	.725	
Zscore(X15)	-.266	.301	-.722	.300
Zscore(X8)	-.271		-.707	.367
Zscore(X16)			-.245	.637
Zscore(X19)	-.254			.560
Zscore(X1)	.440		.247	-.510
Zscore(X18)				.500
Zscore(X17)				-.390

Extraction Method: Principal Component Analysis.

จากตารางที่ 4.22 การจัดตัวแปรอิสระให้อยู่ในปัจจัยใด จะพิจารณาจากขนาดของค่า Factor Loading ที่มีค่ามากที่สุด และหากมีค่า Factor Loading เพียงค่าเดียวในปัจจัยใด ให้ถือว่าตัวแปรอิสระควรอยู่ในปัจจัยนั้น ซึ่งแต่ละปัจจัยประกอบด้วยตัวแปรต่าง ๆ ดังนี้

ปัจจัยที่ 1 F_1 สามารถอธิบายความแปรปรวนรวมทั้งหมดของตัวแปรได้ร้อยละ 18.644 ประกอบด้วยสมาชิกทั้งหมด 6 ตัวแปร ดังนี้

1. ก๊าซไนโตรเจนไดออกไซด์ (NO_2) เฉลี่ย 1 ชั่วโมง (X_2)
2. ก๊าซคาร์บอนมอนอกไซด์ (CO) เฉลี่ย 1 ชั่วโมง (X_3)
3. ก๊าซคาร์บอนมอนอกไซด์ (CO) เฉลี่ย 8 ชั่วโมง (X_4)
4. ก๊าซซัลเฟอร์ไดออกไซด์ (SO_2) เฉลี่ย 24 ชั่วโมง (X_5)
5. ฝุ่นละอองขนาดเล็กกว่า 10 ไมครอน (PM_{10}) เฉลี่ย 24 ชั่วโมง (X_6)
6. ทัศนวิสัยโดยเฉลี่ย (X_{10})

ปัจจัยที่ 2 F_2 สามารถอธิบายความแปรปรวนรวมทั้งหมดของตัวแปรได้ร้อยละ 15.593 ประกอบด้วยสมาชิกทั้งหมด 4 ตัวแปร ดังนี้

1. อุณหภูมิตุ้มเปียก (X_9)
2. อุณหภูมิสูงสุดในแต่ละวัน (X_{11})
3. อุณหภูมิต่ำสุดในแต่ละวัน (X_{12})
4. อุณหภูมิตุ้มแห้ง (X_{13})

ปัจจัยที่ 3 F_3 สามารถอธิบายความแปรปรวนรวมทั้งหมดของตัวแปรได้ร้อยละ 15.194 ประกอบด้วยสมาชิกทั้งหมด 4 ตัวแปร ดังนี้

1. การระเหยของน้ำในแต่ละวัน (X_7)
2. ข้อมูลเมฆ (X_8)
3. ความยาวนานของแสงแดดในแต่ละวัน (X_{14})
4. ความชื้นสัมพัทธ์เฉลี่ยในแต่ละวัน (X_{15})

ปัจจัยที่ 4 F_4 สามารถอธิบายความแปรปรวนรวมทั้งหมดของตัวแปรได้ร้อยละ 9.757 ประกอบด้วยสมาชิกทั้งหมด 5 ตัวแปร ดังนี้

1. ก๊าซโอโซน (O_3) เฉลี่ย 1 ชั่วโมง (X_1)
2. ปริมาณน้ำฝนในแต่ละวัน (X_{16})
3. ความกดอากาศเฉลี่ยในแต่ละวัน (X_{17})
4. ทิศทางลมในแต่ละวัน (X_{18})
5. ความเร็วลมสูงสุดในแต่ละวัน (X_{19})

ผลการหาสมการสำหรับคำนวณค่าคะแนนปัจจัย (factor score) ด้วยวิธีวิเคราะห์การถดถอย (Regression Method) ได้ผลดังนี้

กำหนดให้ F_1, F_2, F_3 และ F_4 คือ ค่าคะแนนปัจจัยของปัจจัยที่ 1, 2, 3 และ 4 ตามลำดับ จะได้ว่า

$$F_1 = 0.746Z_2 + 0.790Z_3 + 0.549Z_4 + 0.319Z_5 + 0.824Z_6 - 0.720Z_{10} \quad (4.19)$$

$$F_2 = 0.794Z_9 + 0.648Z_{11} + 0.744Z_{12} + 0.860Z_{13} \quad (4.20)$$

$$F_3 = 0.725Z_7 - 0.707Z_8 + 0.807Z_{14} - 0.722Z_{15} \quad (4.21)$$

$$F_4 = -0.510Z_1 + 0.637Z_{16} - 0.390Z_{17} + 0.500Z_{18} + 0.560Z_{19} \quad (4.22)$$

4.4.2 ผลการวิเคราะห์ข้อมูลโดยใช้วิธีการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ

ตรวจสอบลักษณะความสัมพันธ์ของตัวแปรตาม Y กับตัวแปรอิสระ F ทีละคู่ พบว่า ตัวแปรตาม Y ที่มีความสัมพันธ์กับตัวแปรอิสระ F เป็นเส้นตรง คือ Y กับ F_1 ส่วนตัวแปรตาม Y ที่มีความสัมพันธ์กับตัวแปรอิสระ F ไม่เป็นเส้นตรง จะต้องทำการแปลงข้อมูล นั่นคือ Y กับ F_2 แปลง F_2 เป็นแบบ Cubic, Y กับ F_3 แปลง F_3 เป็นแบบ Cubic และ Y กับ F_4 แปลง F_4 เป็นแบบ Cubic

นำค่าคะแนนปัจจัยที่ได้จากการวิเคราะห์ปัจจัยในข้อ 4.4.1 มาวิเคราะห์หาตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศ ประมาณค่าพารามิเตอร์ของการถดถอยด้วยวิธีกำลังสองน้อยที่สุด (Least Square Method) และเลือกสมการที่ดีที่สุดจากวิธีการคัดเลือกตัวแปรทั้ง 3 วิธี ดังนี้

1) วิธีการเพิ่มตัวแปรอิสระ (Forward Selection Procedure)

ผลการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยด้วยวิธีการเพิ่มตัวแปรอิสระ แสดงดังตารางที่ 4.23

ตารางที่ 4.23 แสดงผลการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยโดยวิธีการเพิ่มตัวแปรอิสระของตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีบางขุนเทียน

Model		Unstandardized Coefficients		t	Sig.	Collinearity Statistics
		B	Std. Error			VIF
7	(Constant)	55.560	.699	79.433	.000	
	FAC1_1	15.101	.464	32.514	.000	1.107
	FAC4_1	-3.908	.747	-5.234	.000	2.319
	FAC2_1	-3.556	.442	-8.043	.000	1.029
	FAC3_1	3.173	.468	6.782	.000	1.155
	FAC3_2	-1.440	.365	-3.940	.000	1.211
	FAC4_3	-1.202	.246	-4.878	.000	6.436
	FAC4_2	2.355	.600	3.928	.000	4.218

a. Dependent Variable: Y

ผลจากตารางที่ 4.23 พบว่าตัวแปรที่มีผลต่อดัชนีคุณภาพอากาศ คือ ตัวแปรในปัจจัยร่วมที่ 1 (FAC1_1), ตัวแปรในปัจจัยร่วมที่ 2 (FAC2_1), ตัวแปรในปัจจัยร่วมที่ 3 (FAC3_1), ตัวแปรในปัจจัยร่วมที่ 3 ยกกำลังสอง (FAC3_2), ตัวแปรในปัจจัยร่วมที่ 4 (FAC4_1), ตัวแปรในปัจจัยร่วมที่ 4 ยกกำลังสอง (FAC4_2) และตัวแปรในปัจจัยร่วมที่ 4 ยกกำลังสาม (FAC4_3) จะได้สมการการถดถอย ดังนี้

$$\hat{Y} = 55.560 + 15.101F_1 - 3.556F_2 + 3.173F_3 - 1.440F_3^2 - 3.908F_4 + 2.355F_4^2 - 1.202F_4^3$$

และมีค่า $R^2 = 0.837$

2) วิธีการลดตัวแปรอิสระ (Backward Elimination Procedure)

ผลการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยด้วยวิธีการลดตัวแปรอิสระ ได้ผลเหมือนกับวิธีเพิ่มตัวแปรอิสระ

3) วิธีการเพิ่มตัวแปรอิสระทีละขั้นตอน (Stepwise Regression Procedure)

ผลการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยด้วยวิธีการเพิ่มตัวแปรอิสระทีละขั้นตอน ได้ผลเหมือนกับวิธีการเพิ่มตัวแปรอิสระ

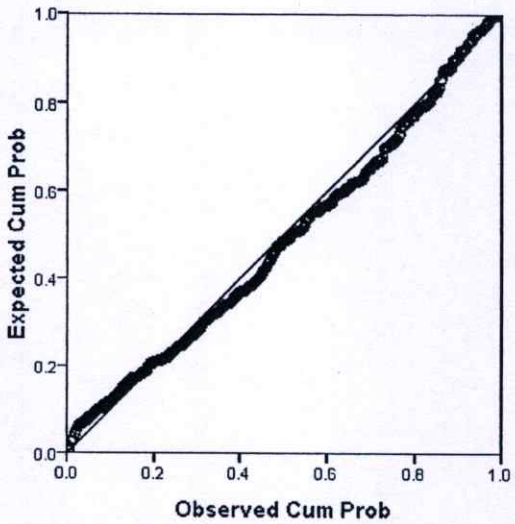
จากวิธีการคัดเลือกตัวแปรทั้ง 3 วิธี จะเลือกสมการที่ดีที่สุดด้วยวิธีการเพิ่มตัวแปรอิสระทีละขั้นตอน เพราะให้ค่า R^2 สูงที่สุด ดังนั้นสมการถดถอยที่ดีที่สุดสำหรับการพยากรณ์ คือ

$$\hat{Y} = 55.560 + 15.101F_1 - 3.556F_2 + 3.173F_3 - 1.440F_3^2 - 3.908F_4 + 2.355F_4^2 - 1.202F_4^3$$

4.4.3 ผลการตรวจสอบข้อสมมติของค่าความคลาดเคลื่อน

1) การตรวจสอบค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติ

1.1) การพล๊อตกราฟ Normal Probability Plot



รูปที่ 4.10 กราฟ Normal Probability Plot ของค่าเศษเหลือของสมการถดถอยสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีบางขุนเทียน

จากรูปที่ 4.10 พบว่ากราฟที่ได้เป็นเส้นตรง จึงสรุปว่าค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติ

1.2) ผลการทดสอบของ Kolmogorov-Smirnov ที่ปรับปรุงด้วยการทดสอบของ Lilliefors ได้ผลดังนี้

สมมติฐานของการทดสอบ คือ

H_0 : ค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติ

H_1 : ค่าความคลาดเคลื่อนไม่มีการแจกแจงแบบปกติ

ตารางที่ 4.24 แสดงการทดสอบค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติสถานีบางขุนเทียน

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Unstandardized Residual	.054	302	.035	.980	302	.000

a. Lilliefors Significance Correction

สถิติทดสอบ Kolmogorov-Smirnov

$$D = \max\left|F(x)-S(x)\right| = 0.054$$

ค่า p-value = 0.035 มีค่ามากกว่า $\alpha = 0.01$ ดังนั้นจึงยอมรับ H_0 สรุปได้ว่าค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติ

2) การตรวจสอบความเป็นอิสระกันของค่าความคลาดเคลื่อน

สมมติฐานของการทดสอบ คือ

H_0 : ค่าความคลาดเคลื่อนเป็นอิสระกัน

H_1 : ค่าความคลาดเคลื่อนไม่เป็นอิสระกัน

ตารางที่ 4.25 แสดงค่า Durbin-Watson ที่ได้จากการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยด้วยวิธีการเพิ่มตัวแปรอิสระทีละขั้นตอนของสถานีบางขุนเทียน

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
4	.915[d]	.837	.833	7.59991	1.854

d. Predictors: (Constant), FAC4_3, FAC1_1, FAC3_2, FAC2_1, FAC4_1, FAC3_1, FAC4_2

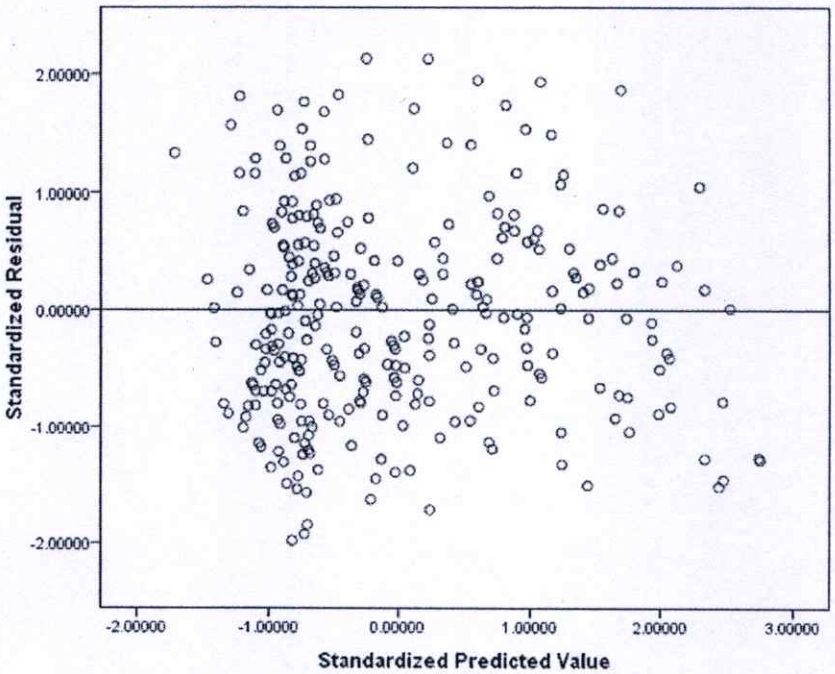
e. Dependent Variable: Y

สถิติทดสอบ

$$d = \frac{\sum_{i=2}^{n}(e_i - e_{i-1})}{\sum_{i=1}^{n} e_i^2} = 1.854$$

ค่า Durbin-Watson = 1.854 (นั่นคือ มีค่าในช่วง 1.5 - 2.5) ดังนั้นจึงยอมรับ H_0 จะสรุปว่า e_i กับ e_j เป็นอิสระกัน

3) การตรวจสอบค่าความแปรปรวนของค่าความคลาดเคลื่อน โดยการพล๊อตกราฟระหว่าง $E(e)$ กับ $E(\hat{Y})$



รูปที่ 4.11 ความสัมพันธ์ระหว่าง $E(e)$ กับ $E(\hat{Y})$ ของสถานีบางขุนเทียน

จากรูปที่ 4.11 พบว่าค่า e_i มีการกระจายอยู่รอบค่าศูนย์ และขนานไปกับแกน X สรุปได้ว่าค่าความแปรปรวนของค่าความคลาดเคลื่อนมีค่าคงที่

4) การตรวจสอบความเป็นอิสระกันของตัวแปรอิสระ

จากตารางที่ 4.23 พบว่า ค่า VIF ของตัวแปรอิสระทุกตัวมีค่าไม่เกิน 10 จึงสรุปได้ว่าตัวแปรอิสระไม่มีความสัมพันธ์กัน

เนื่องจากตัวแบบที่ได้ผ่านข้อสมมติเบื้องต้นของตัวแบบทุกข้อ ดังนั้นจะได้สมการถดถอยที่ดีที่สุดสำหรับการพยากรณ์ คือ

$$\hat{Y} = 55.560 + 15.101F_1 - 3.556F_2 + 3.173F_3 - 1.440F_3^2 - 3.908F_4 + 2.355F_4^2 - 1.202F_4^3 \quad (4.23)$$

และมีค่า $R^2 = 0.837$

จากสมการ 4.23 พบว่าตัวแปรอิสระที่มีผลต่อดัชนีคุณภาพอากาศ คือ ตัวแปรในปัจจัยร่วมที่ 1 (F_1), ตัวแปรในปัจจัยร่วมที่ 2 (F_2), ตัวแปรในปัจจัยร่วมที่ 3 (F_3), ตัวแปรในปัจจัยร่วมที่ 3 ยกกำลังสอง (F_3^2), ตัวแปรในปัจจัยร่วมที่ 4 (F_4), ตัวแปรในปัจจัยร่วมที่ 4 ยกกำลังสอง (F_4^2), และตัวแปรในปัจจัยร่วมที่ 4 ยกกำลังสาม (F_4^3), ซึ่งสามารถอธิบายตัวแปรตามได้ 83.7%

4.5 ผลการวิเคราะห์ข้อมูลในการสร้างตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีอินทรพิทักษ์

4.5.1 ผลการวิเคราะห์ข้อมูลโดยใช้วิธีการวิเคราะห์ปัจจัย

ผลการพิจารณาค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปรอิสระ พบว่าค่า |r| มีค่ามากกว่า 0.5 จึงสรุปได้ว่าตัวแปรอิสระมีความสัมพันธ์กันสูง

ผลการพิจารณาค่า KMO (Kaiser-Meyer-Olkin) ซึ่งเป็นค่าที่ใช้วัดความเหมาะสมของข้อมูลที่จะนำมาวิเคราะห์ปัจจัย พบว่าค่า KMO = 0.717 ซึ่งมีค่ามากกว่า 0.5 จึงสรุปได้ว่าข้อมูลที่มีอยู่เหมาะสมที่จะใช้วิธีการวิเคราะห์ปัจจัย

สำหรับผลการวิเคราะห์ปัจจัยของตัวแปรอิสระ 19 ตัวแปร โดยวิธีวิเคราะห์องค์ประกอบหลัก แสดงดังตารางที่ 4.26

ตารางที่ 4.26 แสดงผลการวิเคราะห์ปัจจัยของตัวแปรอิสระสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีอินทรพิทักษ์

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5.283	27.803	27.803	5.283	27.803	27.803	3.652	19.219	19.219
2	3.471	18.267	46.070	3.471	18.267	46.070	3.480	18.318	37.537
3	1.948	10.255	56.325	1.948	10.255	56.325	2.981	15.689	53.226
4	1.126	5.925	62.249	1.126	5.925	62.249	1.546	8.135	61.361
5	1.034	5.440	67.690	1.034	5.440	67.690	1.202	6.329	67.690
6	.942	4.956	72.646						
7	.919	4.834	77.480						
8	.844	4.442	81.922						
9	.685	3.605	85.528						
10	.638	3.358	88.886						
11	.470	2.476	91.362						
12	.441	2.319	93.681						
13	.364	1.918	95.599						
14	.304	1.602	97.201						
15	.215	1.130	98.330						
16	.180	.947	99.277						
17	.127	.669	99.946						
18	.009	.047	99.993						
19	.001	.007	100.000						

Extraction Method: Principal Component Analysis.

จากตารางที่ 4.26 พบว่าตัวแปรอิสระต่าง ๆ มีการจับกลุ่มรวมกันเป็นปัจจัยทั้งหมด 5 ปัจจัย และสามารถอธิบายความแปรปรวนรวมทั้งหมดของตัวแปรได้ร้อยละ 67.690

ทำการหมุนแกนปัจจัยด้วยวิธีวาริแมกซ์ แสดงดังตารางที่ 4.27

ตารางที่ 4.27 แสดงผลการหมุนแกนปัจจัยด้วยวิธีวาริแมกซ์ของสถานีอินทรพิทักษ์

	Component				
	1	2	3	4	5
Zscore(X4)	.918				
Zscore(X3)	.912				
Zscore(X6)	.710	.362	-.254	-.299	
Zscore(X2)	.710	.401	-.303	-.320	
Zscore(X10)	-.584		.238	.429	-.209
Zscore(X5)	.582			.417	-.244
Zscore(X15)		-.825	.247		.220
Zscore(X14)		.821	.226		
Zscore(X8)		-.816		.234	
Zscore(X7)	-.301	.600	.430	.218	
Zscore(X1)		.584		-.447	
Zscore(X13)	-.244		.890		
Zscore(X9)	-.209	-.498	.787		
Zscore(X12)			.749		
Zscore(X11)		.255	.635		
Zscore(X19)	-.202			.633	
Zscore(X18)				.495	.216
Zscore(X17)					-.727
Zscore(X16)		-.285			.651

Extraction Method: Principal Component Analysis.

จากตารางที่ 4.27 การจัดตัวแปรอิสระให้อยู่ในปัจจัยใด จะพิจารณาจากขนาดของค่า Factor Loading ที่มีค่ามากที่สุด ซึ่งแต่ละปัจจัยประกอบด้วยตัวแปรต่าง ๆ ดังนี้

ปัจจัยที่ 1 F_1 จะสามารถอธิบายความแปรปรวนรวมทั้งหมดของตัวแปรได้ร้อยละ 19.219 ประกอบด้วยสมาชิกทั้งหมด 6 ตัวแปร ดังนี้

1. ก๊าซไนโตรเจนไดออกไซด์ (NO_2) เฉลี่ย 1 ชั่วโมง (X_2)
2. ก๊าซคาร์บอนมอนอกไซด์ (CO) เฉลี่ย 1 ชั่วโมง (X_3)
3. ก๊าซคาร์บอนมอนอกไซด์ (CO) เฉลี่ย 8 ชั่วโมง (X_4)
4. ก๊าซซัลเฟอร์ไดออกไซด์ (SO_2) เฉลี่ย 24 ชั่วโมง (X_5)
5. ฝุ่นละอองขนาดเล็กกว่า 10 ไมครอน (PM_{10}) เฉลี่ย 24 ชั่วโมง (X_6)
6. ทัศนวิสัยโดยเฉลี่ย (X_{10})

ปัจจัยที่ 2 F_2 จะสามารถอธิบายความแปรปรวนรวมทั้งหมดของตัวแปรได้ร้อยละ 18.318 ประกอบด้วยสมาชิกทั้งหมด 5 ตัวแปร ดังนี้

1. ก๊าซโอโซน (O_3) เฉลี่ย 1 ชั่วโมง (X_1)
2. การระเหยของน้ำในแต่ละวัน (X_7)
3. ข้อมูลเมฆ (X_8)
4. ความยาวนานของแสงแดดในแต่ละวัน (X_{14})
5. ความชื้นสัมพัทธ์เฉลี่ยในแต่ละวัน (X_{15})

ปัจจัยที่ 3 F_3 สามารถอธิบายความแปรปรวนรวมทั้งหมดของตัวแปรได้ร้อยละ 15.689 ประกอบด้วยสมาชิกทั้งหมด 4 ตัวแปร ดังนี้

1. อุณหภูมิตุ้มเปียก (X_9)
2. อุณหภูมิสูงสุดในแต่ละวัน (X_{11})
3. อุณหภูมิต่ำสุดในแต่ละวัน (X_{12})
4. อุณหภูมิตุ้มแห้ง (X_{13})

ปัจจัยที่ 4 F_4 สามารถอธิบายความแปรปรวนรวมทั้งหมดของตัวแปรได้ร้อยละ 8.135 ประกอบด้วยสมาชิกทั้งหมด 2 ตัวแปร ดังนี้

1. ทิศทางลมในแต่ละวัน (X_{18})
2. ความเร็วลมสูงสุดในแต่ละวัน (X_{19})

ปัจจัยที่ 5 F_5 สามารถอธิบายความแปรปรวนรวมทั้งหมดของตัวแปรได้ร้อยละ 6.329 ประกอบด้วยสมาชิกทั้งหมด 2 ตัวแปร ดังนี้

1. ปริมาณน้ำฝนในแต่ละวัน (X_{16})
2. ความกดอากาศเฉลี่ยในแต่ละวัน (X_{17})

ผลการหาสมการสำหรับคำนวณค่าคะแนนปัจจัย (factor score) ด้วยวิธีวิเคราะห์การถดถอย (Regression Method) ได้ผลดังนี้

กำหนดให้ F_1, F_2, F_3, F_4 และ F_5 คือ ค่าคะแนนปัจจัยของปัจจัยที่ 1, 2, 3, 4 และ 5 ตามลำดับ จะได้ว่า

$$F_1 = 0.710Z_2 + 0.912Z_3 + 0.918Z_4 + 0.582Z_5 + 0.710Z_6 - 0.584Z_{10} \quad (4.24)$$

$$F_2 = 0.584Z_1 + 0.600Z_7 - 0.816Z_8 + 0.821Z_{14} - 0.825Z_{15} \quad (4.25)$$

$$F_3 = 0.787Z_9 + 0.635Z_{11} + 0.749Z_{12} + 0.890Z_{13} \quad (4.26)$$

$$F_4 = 0.495Z_{18} + 0.633Z_{19} \quad (4.27)$$

$$F_5 = 0.651Z_{16} - 0.727Z_{17} \quad (4.28)$$

4.5.2 ผลการวิเคราะห์ข้อมูลโดยใช้วิธีการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ

ตรวจสอบลักษณะความสัมพันธ์ของตัวแปรตาม Y กับตัวแปรอิสระ F ทีละคู่ พบว่า ตัวแปรตาม Y ที่มีความสัมพันธ์กับตัวแปรอิสระ F เป็นเส้นตรง คือ Y กับ F_1 และ Y กับ F_4 ส่วนตัวแปรตาม Y ที่มีความสัมพันธ์กับตัวแปรอิสระ F ไม่เป็นเส้นตรง จะต้องทำการแปลงข้อมูล นั่นคือ Y กับ F_2 แปลง F_2 เป็นแบบ Cubic, Y กับ F_3 แปลง F_3 เป็นแบบ Cubic และ Y กับ F_5 แปลง F_5 เป็นแบบ Cubic

นำค่าคะแนนปัจจัยที่ได้จากการวิเคราะห์ปัจจัยในข้อ 4.5.1 มาวิเคราะห์หาตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศ ประมาณค่าพารามิเตอร์ของการถดถอยด้วยวิธีกำลังสองน้อยที่สุด (Least Square Method) และเลือกสมการที่ดีที่สุดจากวิธีการคัดเลือกตัวแปรทั้ง 3 วิธี ดังนี้

1) วิธีการเพิ่มตัวแปรอิสระ (Forward Selection Procedure)

ผลการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยด้วยวิธีการเพิ่มตัวแปรอิสระ แสดงดังตารางที่ 4.28

ตารางที่ 4.28 แสดงผลการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยโดยวิธีการเพิ่มตัวแปรอิสระของตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีอินทรพิทักษ์

Model		Unstandardized Coefficients		t	Sig.	Collinearity Statistics	
		B	Std. Error			Toleranc	VIF
4	(Constant)	48.931	.446	109.752	.000		
	FAC1_1	11.436	.448	25.523	.000	.999	1.001
	FAC2_1	6.765	.442	15.308	.000	1.000	1.000
	FAC4_1	-5.951	.447	-13.301	.000	.999	1.001
	FAC3_1	-4.403	.442	-9.969	.000	1.000	1.000

a. Dependent Variable: Y

ผลจากตารางที่ 4.28 พบว่าตัวแปรที่มีผลต่อดัชนีคุณภาพอากาศ คือ ตัวแปรในปัจจัยร่วมที่ 1 (FAC1_1), ตัวแปรในปัจจัยร่วมที่ 2 (FAC2_1), ตัวแปรในปัจจัยร่วมที่ 3 (FAC3_1), และตัวแปรในปัจจัยร่วมที่ 4 (FAC4_1) จะได้สมการการถดถอย ดังนี้

$$\hat{Y} = 48.931 + 11.436F_1 + 6.765F_2 - 4.403F_3 - 5.951F_4$$

และมีค่า $R^2 = 0.790$

2) วิธีการลดตัวแปรอิสระ (Backward Elimination Procedure)

ผลการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยด้วยวิธีการลดตัวแปรอิสระ ได้ผลเหมือนกับวิธีการเพิ่มตัวแปรอิสระ

3) วิธีการเพิ่มตัวแปรอิสระทีละขั้นตอน (Stepwise Regression Procedure)

ผลการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยด้วยวิธีการเพิ่มตัวแปรอิสระทีละขั้นตอน ได้ผลเหมือนกับวิธีการเพิ่มตัวแปรอิสระ

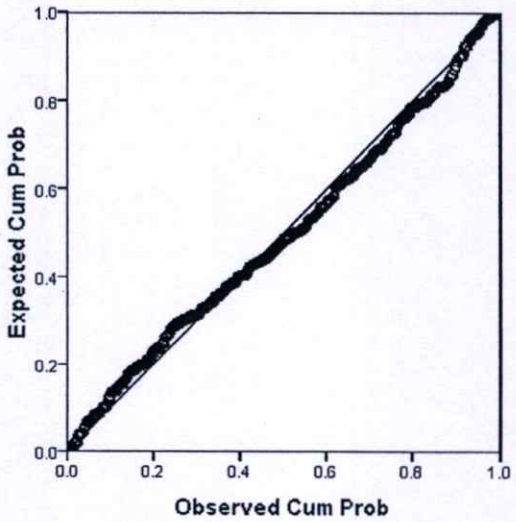
จากวิธีการคัดเลือกตัวแปรทั้ง 3 วิธี จะเลือกสมการที่ดีที่สุดด้วยวิธีการเพิ่มตัวแปรอิสระทีละขั้นตอน เพราะให้ค่า R^2 สูงที่สุด ดังนั้นสมการถดถอยที่ดีที่สุดสำหรับการพยากรณ์ คือ

$$\hat{Y} = 48.931 + 11.436F_1 + 6.765F_2 - 4.403F_3 - 5.951F_4$$

4.5.3 ผลการตรวจสอบข้อสมมติของค่าความคลาดเคลื่อน

1) การตรวจสอบค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติ

1.1) การพล๊อตกราฟ Normal Probability Plot



รูปที่ 4.12 กราฟ Normal Probability Plot ของค่าเศษเหลือของสมการถดถอยสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีอินทรพิทักษ์

จากรูปที่ 4.12 พบว่ากราฟที่ได้เป็นเส้นตรง จึงสรุปว่าค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติ

1.2) ผลการทดสอบของ Kolmogorov-Smirnov ที่ปรับปรุงด้วยการทดสอบของ Lilliefors ได้ผลดังนี้

สมมติฐานของการทดสอบ คือ

H_0 : ค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติ

H_1 : ค่าความคลาดเคลื่อนไม่มีการแจกแจงแบบปกติ

ตารางที่ 4.29 แสดงการทดสอบค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติของสถานีอินทรพิทักษ์

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Unstandardized Residual	.040	308	.200*	.987	308	.008

a. Lilliefors Significance Correction

*. This is a lower bound of the true significance.

สถิติทดสอบ Kolmogorov-Smirnov

$$D = \max\left|F(x)-S(x)\right| = 0.040$$

ค่า p-value = 0.200 มีค่ามากกว่า $\alpha = 0.01$ ดังนั้นจึงยอมรับ H_0 สรุปได้ว่าค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติ

2) การตรวจสอบความเป็นอิสระกันของค่าความคลาดเคลื่อน

สมมติฐานของการทดสอบ คือ

H_0 : ค่าความคลาดเคลื่อนเป็นอิสระกัน

H_1 : ค่าความคลาดเคลื่อนไม่เป็นอิสระกัน

ตารางที่ 4.30 แสดงค่า Durbin-Watson ที่ได้จากการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยด้วยวิธีการเพิ่มตัวแปรอิสระทีละขั้นตอนของสถานีอินทรพิทักษ์

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
4	.889[d]	.790	.787	7.82214	1.647

d. Predictors: (Constant), FAC1_1, FAC2_1, FAC4_1, FAC3_1

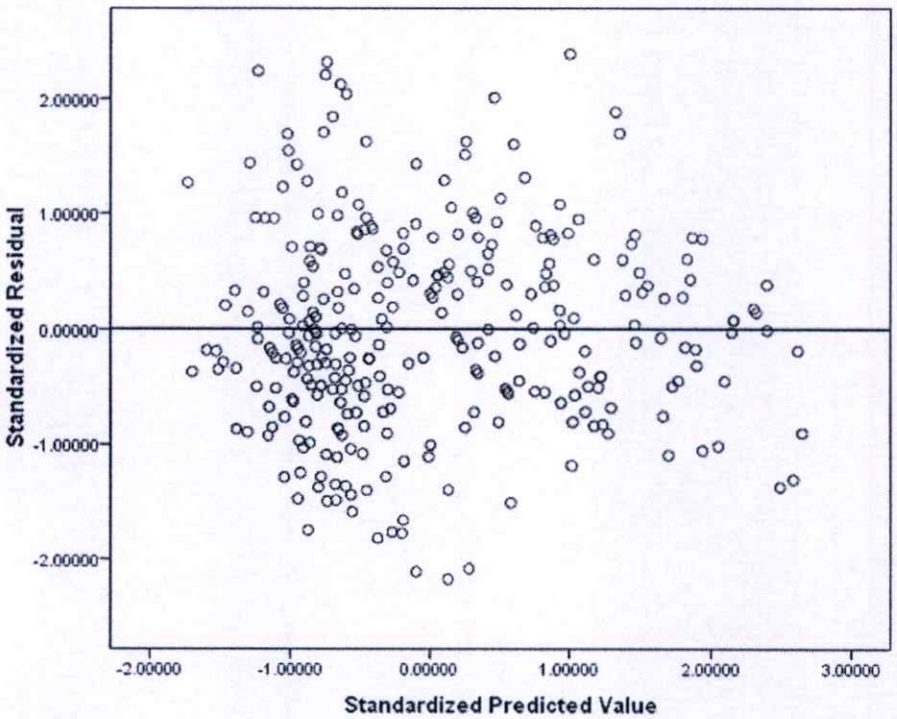
e. Dependent Variable: Y

สถิติทดสอบ

$$d = \frac{\sum_{i=2}^{n}(e_i - e_{i-1})}{\sum_{i=1}^{n} e_i^2} = 1.647$$

ค่า Durbin-Watson = 1.647 (นั่นคือ มีค่าในช่วง 1.5 - 2.5) ดังนั้นจึงยอมรับ H_0 จะสรุปว่า e_i กับ e_j เป็นอิสระกัน

3) การตรวจสอบค่าความแปรปรวนของค่าความคลาดเคลื่อน โดยการพล๊อตกราฟระหว่าง $E(e)$ กับ $E(\hat{Y})$



รูปที่ 4.13 ความสัมพันธ์ระหว่าง $E(e)$ กับ $E(\hat{Y})$ ของสถานีอินทรพิทักษ์

จากรูปที่ 4.13 พบว่าค่า e_i มีการกระจายอยู่รอบค่าศูนย์ และขนานไปกับแกน X สรุปได้ว่าค่าความแปรปรวนของค่าความคลาดเคลื่อนมีค่าคงที่

4) การตรวจสอบความเป็นอิสระกันของตัวแปรอิสระ

จากตารางที่ 4.28 พบว่า ค่า VIF ของตัวแปรอิสระทุกตัวมีค่าไม่เกิน 10 จึงสรุปได้ว่าตัวแปรอิสระไม่มีความสัมพันธ์กัน

เนื่องจากตัวแบบที่ได้ผ่านข้อสมมติเบื้องต้นของตัวแบบทุกข้อ ดังนั้นจะได้สมการถดถอยที่ดีที่สุดสำหรับการพยากรณ์ คือ

$$\hat{Y} = 48.931 + 11.436F_1 + 6.765F_2 - 4.403F_3 - 5.951F_4 \qquad (4.29)$$

และมีค่า $R^2 = 0.790$

จากสมการ 4.29 พบว่าตัวแปรอิสระที่มีผลต่อดัชนีคุณภาพอากาศ คือ ตัวแปรในปัจจัยร่วมที่ 1 (F_1), ตัวแปรในปัจจัยร่วมที่ 2 (F_2), ตัวแปรในปัจจัยร่วมที่ 3 (F_3) และตัวแปรในปัจจัยร่วมที่ 4 (F_4) ซึ่งสามารถอธิบายตัวแปรตามได้ 79%

4.6 ผลการวิเคราะห์ข้อมูลในการสร้างตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีลาดพร้าว

4.6.1 ผลการวิเคราะห์ข้อมูลโดยใช้วิธีการวิเคราะห์ปัจจัย

ผลการพิจารณาค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปรอิสระ พบว่าค่า $|r|$ มีค่ามากกว่า 0.5 จึงสรุปได้ว่าตัวแปรอิสระมีความสัมพันธ์กันสูง

ผลการพิจารณาค่า KMO (Kaiser-Meyer-Olkin) ซึ่งเป็นค่าที่ใช้วัดความเหมาะสมของข้อมูลที่จะนำมาวิเคราะห์ปัจจัย พบว่าค่า KMO = 0.708 ซึ่งมีค่ามากกว่า 0.5 จึงสรุปได้ว่าข้อมูลที่มีอยู่เหมาะสมที่จะใช้วิธีการวิเคราะห์ปัจจัย

สำหรับผลการวิเคราะห์ปัจจัยของตัวแปรอิสระ 19 ตัวแปร โดยวิธีวิเคราะห์องค์ประกอบหลัก แสดงดังตารางที่ 4.31

ตารางที่ 4.31 แสดงผลการวิเคราะห์ปัจจัยของตัวแปรอิสระสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีลาดพร้าว

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.941	26.006	26.006	4.941	26.006	26.006	3.240	17.054	17.054
2	3.342	17.590	43.595	3.342	17.590	43.595	3.119	16.415	33.469
3	2.150	11.314	54.909	2.150	11.314	54.909	3.088	16.251	49.720
4	1.155	6.080	60.989	1.155	6.080	60.989	1.690	8.893	58.613
5	1.017	5.352	66.340	1.017	5.352	66.340	1.468	7.728	66.340
6	.997	5.247	71.588						
7	.941	4.950	76.538						
8	.798	4.201	80.739						
9	.767	4.035	84.774						
10	.702	3.697	88.471						
11	.468	2.463	90.933						
12	.404	2.128	93.061						
13	.389	2.049	95.110						
14	.291	1.533	96.643						
15	.267	1.404	98.047						
16	.196	1.032	99.079						
17	.145	.762	99.841						
18	.029	.152	99.993						
19	.001	.007	100.000						

Extraction Method: Principal Component Analysis.

จากตารางที่ 4.31 พบว่าตัวแปรอิสระต่าง ๆ มีการจับกลุ่มรวมกันเป็นปัจจัยทั้งหมด 5 ปัจจัย และสามารถอธิบายความแปรปรวนรวมทั้งหมดของตัวแปรได้ร้อยละ 66.340

ทำการหมุนแกนปัจจัยด้วยวิธีวาริแมกซ์ แสดงดังตารางที่ 4.32

ตารางที่ 4.32 แสดงผลการหมุนแกนปัจจัยด้วยวิธีวาริแมกซ์ของสถานีลาดพร้าว

	Component				
	1	2	3	4	5
Zscore(X14)	.864				
Zscore(X8)	-.802				.213
Zscore(X15)	-.775		.380		.222
Zscore(X7)	.662	-.209	.368		
Zscore(X1)	.528			-.452	-.267
Zscore(X4)		.914			
Zscore(X3)		.912			
Zscore(X2)		.677	-.230	-.475	
Zscore(X6)	.294	.574	-.326	-.543	
Zscore(X5)		.546	.286		
Zscore(X13)			.881		
Zscore(X9)	-.386		.863		
Zscore(X12)			.767		
Zscore(X11)	.282		.488	-.372	.234
Zscore(X10)		-.402	.267	.709	
Zscore(X16)	-.336				.601
Zscore(X19)		-.210			.556
Zscore(X17)					-.515
Zscore(X18)				.392	.490

Extraction Method: Principal Component Analysis.

จากตารางที่ 4.32 การจัดตัวแปรอิสระให้อยู่ในปัจจัยใด จะพิจารณาจากขนาดของค่า Factor Loading ที่มีค่ามากที่สุด ซึ่งแต่ละปัจจัยประกอบด้วยตัวแปรต่าง ๆ ดังนี้

ปัจจัยที่ 1 F_1 จะสามารถอธิบายความแปรปรวนรวมทั้งหมดของตัวแปรได้ร้อยละ 17.054 ประกอบด้วยสมาชิกทั้งหมด 5 ตัวแปร ดังนี้

1. ก๊าซโอโซน (O_3) เฉลี่ย 1 ชั่วโมง (X_1)
2. การระเหยของน้ำในแต่ละวัน (X_7)
3. ข้อมูลเมฆ (X_8)
4. ความยาวนานของแสงแดดในแต่ละวัน (X_{14})
5. ความชื้นสัมพัทธ์เฉลี่ยในแต่ละวัน (X_{15})

ปัจจัยที่ 2 F_2 จะสามารถอธิบายความแปรปรวนรวมทั้งหมดของตัวแปรได้ร้อยละ 16.415 ประกอบด้วยสมาชิกทั้งหมด 5 ตัวแปร ดังนี้

1. ก๊าซไนโตรเจนไดออกไซด์ (NO_2) เฉลี่ย 1 ชั่วโมง (X_2)
2. ก๊าซคาร์บอนมอนอกไซด์ (CO) เฉลี่ย 1 ชั่วโมง (X_3)
3. ก๊าซคาร์บอนมอนอกไซด์ (CO) เฉลี่ย 8 ชั่วโมง (X_4)
4. ก๊าซซัลเฟอร์ไดออกไซด์ (SO_2) เฉลี่ย 24 ชั่วโมง (X_5)
5. ฝุ่นละอองขนาดเล็กกว่า 10 ไมครอน (PM_{10}) เฉลี่ย 24 ชั่วโมง (X_6)

ปัจจัยที่ 3 F_3 จะสามารถอธิบายความแปรปรวนรวมทั้งหมดของตัวแปรได้ร้อยละ 16.251 ประกอบด้วยสมาชิกทั้งหมด 4 ตัวแปร ดังนี้

1. อุณหภูมิตุ้มเปียก (X_9)
2. อุณหภูมิสูงสุดในแต่ละวัน (X_{11})
3. อุณหภูมิต่ำสุดในแต่ละวัน (X_{12})
4. อุณหภูมิตุ้มแห้ง (X_{13})

ปัจจัยที่ 4 F_4 สามารถอธิบายความแปรปรวนรวมทั้งหมดของตัวแปรได้ร้อยละ 8.893 ประกอบด้วยสมาชิกทั้งหมด 1 ตัวแปร ดังนี้

1. ทัศนวิสัยโดยเฉลี่ย (X_{10})

ปัจจัยที่ 5 F_5 จะสามารถอธิบายความแปรปรวนรวมทั้งหมดของตัวแปรได้ร้อยละ 7.728 ประกอบด้วยสมาชิกทั้งหมด 4 ตัวแปร ดังนี้

1. ปริมาณน้ำฝนในแต่ละวัน (X_{16})
2. ความกดอากาศเฉลี่ยในแต่ละวัน (X_{17})
3. ทิศทางลมในแต่ละวัน (X_{18})
4. ความเร็วลมสูงสุดในแต่ละวัน (X_{19})

ผลการหาสมการสำหรับคำนวณค่าคะแนนปัจจัย (factor score) ด้วยวิธีวิเคราะห์การถดถอย (Regression Method) ได้ผลดังนี้

กำหนดให้ F_1, F_2, F_3, F_4 และ F_5 คือ ค่าคะแนนปัจจัยของปัจจัยที่ 1, 2, 3, 4 และ 5 ตามลำดับ จะได้ว่า

$$F_1 = 0.528Z_1 + 0.662Z_7 - 0.802Z_8 + 0.864Z_{14} - 0.775Z_{15} \quad (4.30)$$

$$F_2 = 0.677Z_2 + 0.912Z_3 + 0.914Z_4 + 0.546Z_5 + 0.574Z_6 \quad (4.31)$$

$$F_3 = 0.863Z_9 + 0.488Z_{11} + 0.767Z_{12} + 0.881Z_{13} \quad (4.32)$$

$$F_4 = 0.709Z_{10} \quad (4.33)$$

$$F_5 = 0.601Z_{16} - 0.515Z_{17} + 0.490Z_{18} + 0.556Z_{19} \quad (4.34)$$

4.6.2 ผลการวิเคราะห์ข้อมูลโดยใช้วิธีการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ

ตรวจสอบลักษณะความสัมพันธ์ของตัวแปรตาม Y กับตัวแปรอิสระ F ทีละคู่ พบว่า ตัวแปรตาม Y ที่มีความสัมพันธ์กับตัวแปรอิสระ F เป็นเส้นตรง คือ Y กับ F_2 และ Y กับ F_4 ส่วนตัวแปรตาม Y ที่มีความสัมพันธ์กับตัวแปรอิสระ F ไม่เป็นเส้นตรง จะต้องทำการแปลงข้อมูล นั่นคือ Y กับ F_1 แปลง F_1 เป็นแบบ Quadratic, Y กับ F_3 แปลง F_3 เป็นแบบ Cubic และ Y กับ F_5 แปลง F_5 เป็นแบบ Cubic

นำค่าคะแนนปัจจัยที่ได้จากการวิเคราะห์ปัจจัยในข้อ 4.6.1 มาวิเคราะห์หาตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศ ประมาณค่าพารามิเตอร์ของการถดถอยด้วยวิธีกำลังสองน้อยที่สุด (Least Square Method) และเลือกสมการที่ดีที่สุดจากวิธีการคัดเลือกตัวแปรทั้ง 3 วิธี ดังนี้

1) วิธีการเพิ่มตัวแปรอิสระ (Forward Selection Procedure)

ผลการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยด้วยวิธีการเพิ่มตัวแปรอิสระ แสดงดังตารางที่ 4.33

ตารางที่ 4.33 แสดงผลการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยโดยวิธีการเพิ่มตัวแปรอิสระของตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีลาดพร้าว

Model		Unstandardized Coefficients		t	Sig.	Collinearity Statistics
		B	Std. Error			VIF
1	(Constant)	57.181	.302	189.383	.000	
	FAC1_1	2.888	.240	12.024	.000	1.027
	FAC2_1	5.366	.249	21.570	.000	1.045
	FAC3_1	-2.657	.238	-11.154	.000	1.010
	FAC4_1	-5.916	.268	-22.037	.000	1.097
	FAC5_1	-.163	.493	-.330	.007	2.678
	FAC5_2	-.064	.410	-.155	.008	3.877
	FAC5_3	-.174	.208	-.838	.004	6.218

a. Dependent Variable: Y

ผลจากตารางที่ 4.33 พบว่าตัวแปรที่มีผลต่อดัชนีคุณภาพอากาศ คือ ตัวแปรในปัจจัยร่วมที่ 1 (FAC1_1), ตัวแปรในปัจจัยร่วมที่ 2 (FAC2_1), ตัวแปรในปัจจัยร่วมที่ 3 (FAC3_1), ตัวแปรในปัจจัยร่วมที่ 4 (FAC4_1), ตัวแปรในปัจจัยร่วมที่ 5 (FAC5_1) ตัวแปรในปัจจัยร่วมที่ 5 ยกกำลังสอง (FAC2_2) และตัวแปรในปัจจัยร่วมที่ 5 ยกกำลังสาม (FAC5_3) จะได้สมการการถดถอยดังนี้

$$\hat{Y} = 57.181 + 2.888F_1 + 5.366F_2 - 2.657F_3 - 5.916F_4 - 0.163F_5 - 0.064F_5^2 - 0.174F_5^3$$

และมีค่า $R^2 = 0.817$

2) วิธีการลดตัวแปรอิสระ (Backward Elimination Procedure)

ผลการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยด้วยวิธีการลดตัวแปรอิสระได้ผลเหมือนกับวิธีเพิ่มตัวแปรอิสระ

3) วิธีการเพิ่มตัวแปรอิสระทีละขั้นตอน (Stepwise Regression Procedure)

ผลการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยด้วยวิธีการเพิ่มตัวแปรอิสระทีละขั้นตอน ได้ผลเหมือนกับวิธีการเพิ่มตัวแปรอิสระ

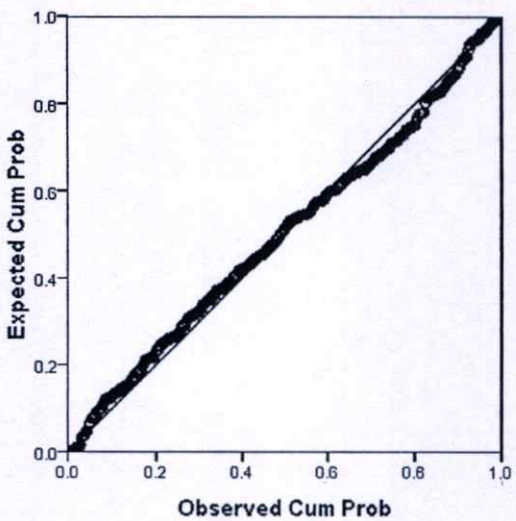
จากวิธีการคัดเลือกตัวแปรทั้ง 3 วิธี จะเลือกสมการที่ดีที่สุดด้วยวิธีการเพิ่มตัวแปรอิสระทีละขั้นตอน เพราะให้ค่า R^2 สูงที่สุด ดังนั้นสมการถดถอยที่ดีที่สุดสำหรับการพยากรณ์ คือ

$$\hat{Y} = 57.181 + 2.888F_1 + 5.366F_2 - 2.657F_3 - 5.916F_4 - 0.163F_5 - 0.064F_5^2 - 0.174F_5^3$$

4.6.3 ผลการตรวจสอบข้อสมมติของค่าความคลาดเคลื่อน

1) การตรวจสอบค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติ

1.1) การพล๊อตกราฟ Normal Probability Plot



รูปที่ 4.14 กราฟ Normal Probability Plot ของค่าเศษเหลือของสมการถดถอยสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีลาดพร้าว

จากรูปที่ 4.14 พบว่ากราฟที่ได้เป็นเส้นตรง จึงสรุปว่าค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติ

1.2) ผลการทดสอบของ Kolmogorov-Smirnov ที่ปรับปรุงด้วยการทดสอบของ Lilliefors ได้ผลดังนี้

สมมติฐานของการทดสอบ คือ

H_0 : ค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติ

H_1 : ค่าความคลาดเคลื่อนไม่มีการแจกแจงแบบปกติ

ตารางที่ 4.34 แสดงการทดสอบค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติของสถานีลาดพร้าว

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Unstandardized Residual	.051	311	.054	.986	311	.003

a. Lilliefors Significance Correction

สถิติทดสอบ Kolmogorov-Smirnov

$$D = \max\left|F(x)-S(x)\right| = 0.051$$

ค่า p-value = 0.054 มีค่ามากกว่า $\alpha = 0.01$ ดังนั้นจึงยอมรับ H_0 สรุปได้ว่าค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติ

2) การตรวจสอบความเป็นอิสระกันของค่าความคลาดเคลื่อน

สมมติฐานของการทดสอบ คือ

H_0 : ค่าความคลาดเคลื่อนเป็นอิสระกัน

H_1 : ค่าความคลาดเคลื่อนไม่เป็นอิสระกัน

ตารางที่ 4.35 แสดงค่า Durbin-Watson ที่ได้จากการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยด้วยวิธีการเพิ่มตัวแปรอิสระทีละขั้นตอนของสถานีลาดพร้าว

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.904[a]	.817	.813	4.18267	1.532

a. Predictors: (Constant), FAC5_3, FAC3_1, FAC1_1, FAC2_1, FAC4_1, FAC5_1,

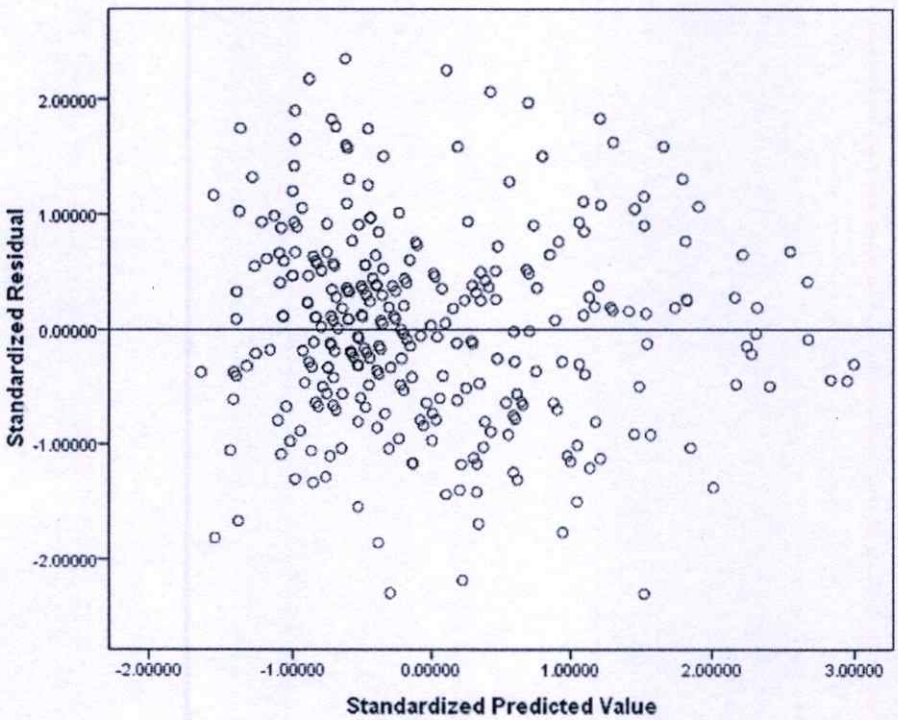
b. Dependent Variable: Y

สถิติทดสอบ

$$d = \frac{\sum_{i=2}^{n}(e_i - e_{i-1})}{\sum_{i=1}^{n} e_i^2} = 1.532$$

ค่า Durbin-Watson = 1.532 (นั่นคือ มีค่าในช่วง 1.5 - 2.5) ดังนั้นจึงยอมรับ H_0 จะสรุปว่า e_i กับ e_j เป็นอิสระกัน

3) การตรวจสอบค่าความแปรปรวนของค่าความคลาดเคลื่อน โดยการพล๊อตกราฟระหว่าง $E(e)$ กับ $E(\hat{Y})$



รูปที่ 4.15 ความสัมพันธ์ระหว่าง $E(e)$ กับ $E(\hat{Y})$ ของสถานีลาดพร้าว

จากรูปที่ 4.15 พบว่าค่า e_i มีการกระจายอยู่รอบค่าศูนย์ และขนานไปกับแกน X สรุปได้ว่าค่าความแปรปรวนของค่าความคลาดเคลื่อนมีค่าคงที่

4) การตรวจสอบความเป็นอิสระกันของตัวแปรอิสระ

จากตารางที่ 4.33 พบว่า ค่า VIF ของตัวแปรอิสระทุกตัวมีค่าไม่เกิน 10 จึงสรุปได้ว่าตัวแปรอิสระไม่มีความสัมพันธ์กัน

เนื่องจากตัวแบบที่ได้ผ่านข้อสมมติเบื้องต้นของตัวแบบทุกข้อ ดังนั้นจะได้สมการถดถอยที่ดีที่สุดสำหรับการพยากรณ์ คือ

$$\hat{Y} = 57.181 + 2.888F_1 + 5.366F_2 - 2.657F_3 - 5.916F_4 - 0.163F_5 - 0.064F_5^2 - 0.174F_5^3 \quad (4.35)$$

และมีค่า $R^2 = 0.817$

จากสมการ 4.35 พบว่าตัวแปรอิสระที่มีผลต่อดัชนีคุณภาพอากาศ คือ ตัวแปรในปัจจัยร่วมที่ 1 (F_1), ตัวแปรในปัจจัยร่วมที่ 2 (F_2), ตัวแปรในปัจจัยร่วมที่ 3 (F_3), ตัวแปรในปัจจัยร่วมที่ 4 (F_4) ตัวแปรในปัจจัยร่วมที่ 5 (F_5), ตัวแปรในปัจจัยร่วมที่ 5 ยกกำลังสอง (F_5^2) และตัวแปรในปัจจัยร่วมที่ 5 ยกกำลังสาม (F_5^3) ซึ่งสามารถอธิบายตัวแปรตามได้ 81.7%

4.7 ผลการวิเคราะห์ข้อมูลในการสร้างตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีดินแดง

4.7.1 ผลการวิเคราะห์ข้อมูลโดยใช้วิธีการวิเคราะห์ปัจจัย

ผลการพิจารณาค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปรอิสระ พบว่าค่า $|r|$ มีค่ามากกว่า 0.5 จึงสรุปได้ว่าตัวแปรอิสระมีความสัมพันธ์กันสูง

ผลการพิจารณาค่า KMO (Kaiser-Meyer-Olkin) ซึ่งเป็นค่าที่ใช้วัดความเหมาะสมของข้อมูลที่จะนำมาวิเคราะห์ปัจจัย พบว่าค่า KMO = 0.714 ซึ่งมีค่ามากกว่า 0.5 จึงสรุปได้ว่าข้อมูลที่มีอยู่เหมาะสมที่จะใช้วิธีการวิเคราะห์ปัจจัย

สำหรับผลการวิเคราะห์ปัจจัยของตัวแปรอิสระ 19 ตัวแปร โดยวิธีวิเคราะห์องค์ประกอบหลัก แสดงดังตารางที่ 4.36

ตารางที่ 4.36 แสดงผลการวิเคราะห์ปัจจัยของตัวแปรอิสระสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีดินแดง

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of	Cumulative	Total	% of	Cumulative	Total	% of	Cumulative
1	5.367	28.246	28.246	5.367	28.246	28.246	4.302	22.643	22.643
2	3.010	15.842	44.088	3.010	15.842	44.088	3.013	15.857	38.500
3	1.840	9.686	53.774	1.840	9.686	53.774	2.647	13.930	52.430
4	1.205	6.345	60.118	1.205	6.345	60.118	1.325	6.974	59.404
5	1.020	5.366	65.485	1.020	5.366	65.485	1.155	6.081	65.485
6	.971	5.113	70.598						
7	.921	4.848	75.446						
8	.905	4.765	80.211						
9	.758	3.988	84.199						
10	.688	3.621	87.820						
11	.573	3.017	90.837						
12	.441	2.320	93.157						
13	.384	2.023	95.179						
14	.325	1.708	96.888						
15	.209	1.099	97.986						
16	.198	1.040	99.027						
17	.167	.881	99.908						
18	.016	.086	99.994						
19	.001	.006	100.000						

Extraction Method: Principal Component Analysis.

จากตารางที่ 4.36 พบว่าตัวแปรอิสระต่าง ๆ มีการจับกลุ่มรวมกันเป็นปัจจัยทั้งหมด 5 ปัจจัย และสามารถอธิบายความแปรปรวนรวมทั้งหมดของตัวแปรได้ร้อยละ 65.485

ทำการหมุนแกนปัจจัยด้วยวิธีวาริแมกซ์ แสดงดังตารางที่ 4.37

ตารางที่ 4.37 แสดงผลการหมุนแกนปัจจัยด้วยวิธีวาริแมกซ์ของสถานีดินแดง

	Component				
	1	2	3	4	5
Zscore(X3)	.922				
Zscore(X4)	.914				
Zscore(X2)	.841	-.279			
Zscore(X6)	.816	-.269			
Zscore(X10)	-.692	-.237			
Zscore(X14)		-.836			
Zscore(X8)	-.201	.815			
Zscore(X15)		.791	.313	-.209	
Zscore(X7)	-.406	-.610	.365		
Zscore(X13)	-.297		.854		
Zscore(X9)	-.328	.449	.784		
Zscore(X12)			.752		
Zscore(X11)		-.221	.565		
Zscore(X17)				.797	.322
Zscore(X16)		.368		-.488	.256
Zscore(X1)				.418	
Zscore(X18)				-.361	.247
Zscore(X5)	.252				.730
Zscore(X19)	-.425				.478

Extraction Method: Principal Component Analysis.

จากตารางที่ 4.37 การจัดตัวแปรอิสระให้อยู่ในปัจจัยใด จะพิจารณาจากขนาดของค่า Factor Loading ที่มีค่ามากที่สุด และหากมีค่า Factor Loading เพียงค่าเดียวในปัจจัยใด ให้ถือว่าตัวแปรอิสระควรอยู่ในปัจจัยนั้น ซึ่งแต่ละปัจจัยประกอบด้วยตัวแปรต่าง ๆ ดังนี้

ปัจจัยที่ 1 F_1 สามารถอธิบายความแปรปรวนรวมทั้งหมดของตัวแปรได้ร้อยละ 22.643 ประกอบด้วยสมาชิกทั้งหมด 5 ตัวแปร ดังนี้

1. ก๊าซไนโตรเจนไดออกไซด์ (NO_2) เฉลี่ย 1 ชั่วโมง (X_2)
2. ก๊าซคาร์บอนมอนอกไซด์ (CO) เฉลี่ย 1 ชั่วโมง (X_3)
3. ก๊าซคาร์บอนมอนอกไซด์ (CO) เฉลี่ย 8 ชั่วโมง (X_4)
4. ฝุ่นละอองขนาดเล็กกว่า 10 ไมครอน (PM_{10}) เฉลี่ย 24 ชั่วโมง (X_6)
5. ทัศนวิสัยโดยเฉลี่ย (X_{10})

ปัจจัยที่ 2 F_2 สามารถอธิบายความแปรปรวนรวมทั้งหมดของตัวแปรได้ร้อยละ 15.857 ประกอบด้วยสมาชิกทั้งหมด 4 ตัวแปร ดังนี้

1. การระเหยของน้ำในแต่ละวัน (X_7)
2. ข้อมูลเมฆ (X_8)
3. ความยาวนานของแสงแดดในแต่ละวัน (X_{14})
4. ความชื้นสัมพัทธ์เฉลี่ยในแต่ละวัน (X_{15})

ปัจจัยที่ 3 F_3 สามารถอธิบายความแปรปรวนรวมทั้งหมดของตัวแปรได้ร้อยละ 13.930 ประกอบด้วยสมาชิกทั้งหมด 4 ตัวแปร ดังนี้

1. อุณหภูมิตุ้มเปียก (X_9)
2. อุณหภูมิสูงสุดในแต่ละวัน (X_{11})
3. อุณหภูมิต่ำสุดในแต่ละวัน (X_{12})
4. อุณหภูมิตุ้มแห้ง (X_{13})

ปัจจัยที่ 4 F_4 สามารถอธิบายความแปรปรวนรวมทั้งหมดของตัวแปรได้ร้อยละ 6.974 ประกอบด้วยสมาชิกทั้งหมด 4 ตัวแปร ดังนี้

1. ก๊าซโอโซน (O_3) เฉลี่ย 1 ชั่วโมง (X_1)
2. ปริมาณน้ำฝนในแต่ละวัน (X_{16})
3. ความกดอากาศเฉลี่ยในแต่ละวัน (X_{17})
4. ทิศทางลมในแต่ละวัน (X_{18})

ปัจจัยที่ 5 F_5 สามารถอธิบายความแปรปรวนรวมทั้งหมดของตัวแปรได้ร้อยละ 6.081 ประกอบด้วยสมาชิกทั้งหมด 2 ตัวแปร ดังนี้

1. ก๊าซซัลเฟอร์ไดออกไซด์ (SO_2) เฉลี่ย 24 ชั่วโมง (X_5)
2. ความเร็วลมสูงสุดในแต่ละวัน (X_{19})

ผลการหาสมการสำหรับคำนวณค่าคะแนนปัจจัย (factor score) ด้วยวิธีวิเคราะห์การถดถอย (Regression Method) ได้ผลดังนี้

กำหนดให้ F_1, F_2, F_3, F_4 และ F_5 คือ ค่าคะแนนปัจจัยของปัจจัยที่ 1, 2, 3, 4 และ 5 ตามลำดับ จะได้ว่า

$$F_1 = 0.841Z_2 + 0.922Z_3 + 0.914Z_4 + 0.816Z_6 - 0.692Z_{10} \quad (4.36)$$

$$F_2 = -0.610Z_7 + 0.815Z_8 - 0.836Z_{14} + 0.791Z_{15} \quad (4.37)$$

$$F_3 = 0.784Z_9 + 0.565Z_{11} + 0.752Z_{12} + 0.854Z_{13} \quad (4.38)$$

$$F_4 = 0.418Z_1 - 0.488Z_{16} + 0.797Z_{17} - 0.361Z_{18} \quad (4.39)$$

$$F_5 = 0.730Z_5 + 0.478Z_{19} \quad (4.40)$$

4.7.2 ผลการวิเคราะห์ข้อมูลโดยใช้วิธีการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ

ตรวจสอบลักษณะความสัมพันธ์ของตัวแปรตาม Y กับตัวแปรอิสระ F ทีละคู่ พบว่า ตัวแปรตาม Y ที่มีความสัมพันธ์กับตัวแปรอิสระ F เป็นเส้นตรง คือ Y กับ F_1 ส่วนตัวแปรตาม Y ที่มีความสัมพันธ์กับตัวแปรอิสระ F ไม่เป็นเส้นตรง จะต้องทำการแปลงข้อมูล นั่นคือ Y กับ F_2 แปลง F_2 เป็นแบบ Cubic, Y กับ F_3 แปลง F_3 เป็นแบบ Cubic, Y กับ F_4 แปลง F_4 เป็นแบบ Cubic และ Y กับ F_5 แปลง F_5 เป็นแบบ Cubic

นำค่าคะแนนปัจจัยที่ได้จากการวิเคราะห์ปัจจัยในข้อ 4.7.1 มาวิเคราะห์หาตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศ ประมาณค่าพารามิเตอร์ของการถดถอยด้วยวิธีกำลังสองน้อยที่สุด (Least Square Method) และเลือกสมการที่ดีที่สุดจากวิธีการคัดเลือกตัวแปรทั้ง 3 วิธี ดังนี้

1) วิธีการเพิ่มตัวแปรอิสระ (Forward Selection Procedure)

ผลการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยด้วยวิธีการเพิ่มตัวแปรอิสระ แสดงดังตารางที่ 4.38

ตารางที่ 4.38 แสดงผลการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยโดยวิธีการเพิ่มตัวแปรอิสระของตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีดินแดง

Model		Unstandardized Coefficients		t	Sig.	Collinearity Statistics
		B	Std. Error			VIF
6	(Constant)	80.770	.664	121.558	.000	
	FAC1_1	17.163	.556	30.887	.000	1.001
	FAC2_1	-5.515	.576	-9.566	.000	1.078
	FAC3_1	-3.582	.555	-6.458	.000	1.002
	FAC5_1	4.419	.746	5.926	.000	1.213
	FAC4_1	6.742	1.211	5.569	.000	1.391
	FAC4_2	3.074	1.172	2.622	.009	1.248

a. Dependent Variable: Y

ผลจากตารางที่ 4.38 พบว่าตัวแปรที่มีผลต่อดัชนีคุณภาพอากาศ คือ ตัวแปรในปัจจัยร่วมที่ 1 (FAC1_1), ตัวแปรในปัจจัยร่วมที่ 2 (FAC2_1), ตัวแปรในปัจจัยร่วมที่ 3 (FAC3_1), ตัวแปรในปัจจัยร่วมที่ 4 (FAC4_1), ตัวแปรในปัจจัยร่วมที่ 4 ยกกำลังสอง (FAC4_2) และตัวแปรในปัจจัยร่วมที่ 5 (FAC5_1) จะได้สมการการถดถอย ดังนี้

$$\hat{Y} = 80.770 + 17.163F_1 - 5.515F_2 - 3.582F_3 + 6.742F_4 + 3.074F_4^2 + 4.419F_5$$

และมีค่า $R^2 = 0.818$

2) วิธีการลดตัวแปรอิสระ (Backward Elimination Procedure)

ผลการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยด้วยวิธีการลดตัวแปรอิสระ ได้ผลเหมือนกับวิธีการเพิ่มตัวแปรอิสระ

3) วิธีการเพิ่มตัวแปรอิสระทีละขั้นตอน (Stepwise Regression Procedure)

ผลการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยด้วยวิธีการเพิ่มตัวแปรอิสระทีละขั้นตอน ได้ผลเหมือนกับวิธีการเพิ่มตัวแปรอิสระ

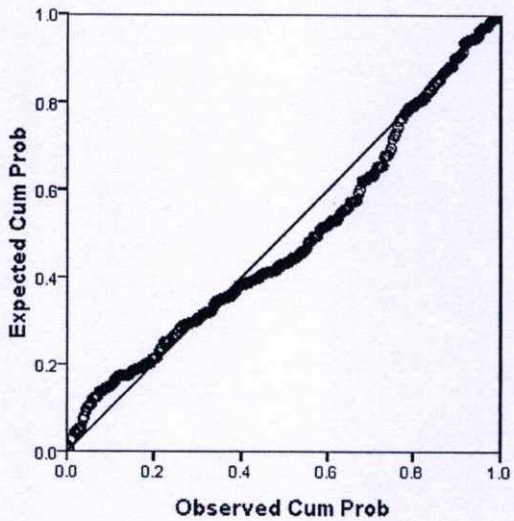
จากวิธีการคัดเลือกตัวแปรทั้ง 3 วิธี จะเลือกสมการที่ดีที่สุดด้วยวิธีการเพิ่มตัวแปรอิสระทีละขั้นตอน เพราะให้ค่า R^2 สูงที่สุด ดังนั้นสมการถดถอยที่ดีที่สุดสำหรับการพยากรณ์ คือ

$$\hat{Y} = 80.770 + 17.163F_1 - 5.515F_2 - 3.582F_3 + 6.742F_4 + 3.074F_4^2 + 4.419F_5$$

4.1.3 ผลการตรวจสอบข้อสมมติของค่าความคลาดเคลื่อน

1) การตรวจสอบค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติ

1.1) การพล๊อตกราฟ Normal Probability Plot



รูปที่ 4.16 กราฟ Normal Probability Plot ของค่าเศษเหลือของสมการถดถอยสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีดินแดง

จากรูปที่ 4.16 พบว่ากราฟที่ได้ไม่เป็นเส้นตรง จึงสรุปว่าค่าความคลาดเคลื่อนไม่มีการแจกแจงแบบปกติ

1.2) ผลการทดสอบของ Kolmogorov-Smirnov ที่ปรับปรุงด้วยการทดสอบของ Lilliefors ได้ผลดังนี้

สมมติฐานของการทดสอบ คือ

H_0 : ค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติ

H_1 : ค่าความคลาดเคลื่อนไม่มีการแจกแจงแบบปกติ

ตารางที่ 4.39 แสดงการทดสอบค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติของสถานีดินแดง

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Unstandardized Residual	.095	262	.000	.951	262	.000

a. Lilliefors Significance Correction

สถิติทดสอบ Kolmogorov-Smirnov

$$D = \max\left|F(x)-S(x)\right| = 0.064$$

ค่า p-value = 0.000 มีค่าน้อยกว่า $\alpha = 0.01$ ดังนั้นจึงปฏิเสธ H_0 สรุปได้ว่าค่าความคลาดเคลื่อนไม่มีการแจกแจงแบบปกติ

2) การตรวจสอบความเป็นอิสระกันของค่าความคลาดเคลื่อน

สมมติฐานของการทดสอบ คือ

H_0 : ค่าความคลาดเคลื่อนเป็นอิสระกัน

H_1 : ค่าความคลาดเคลื่อนไม่เป็นอิสระกัน

ตารางที่ 4.40 แสดงค่า Durbin-Watson ที่ได้จากการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยด้วยวิธีการเพิ่มตัวแปรอิสระทีละขั้นตอนของสถานีดินแดง

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
6	.904[f]	.818	.814	8.98343	1.366

f. Predictors: (Constant), FAC1_1, FAC2_1, FAC3_1, FAC5_1, FAC4_1, FAC4_2

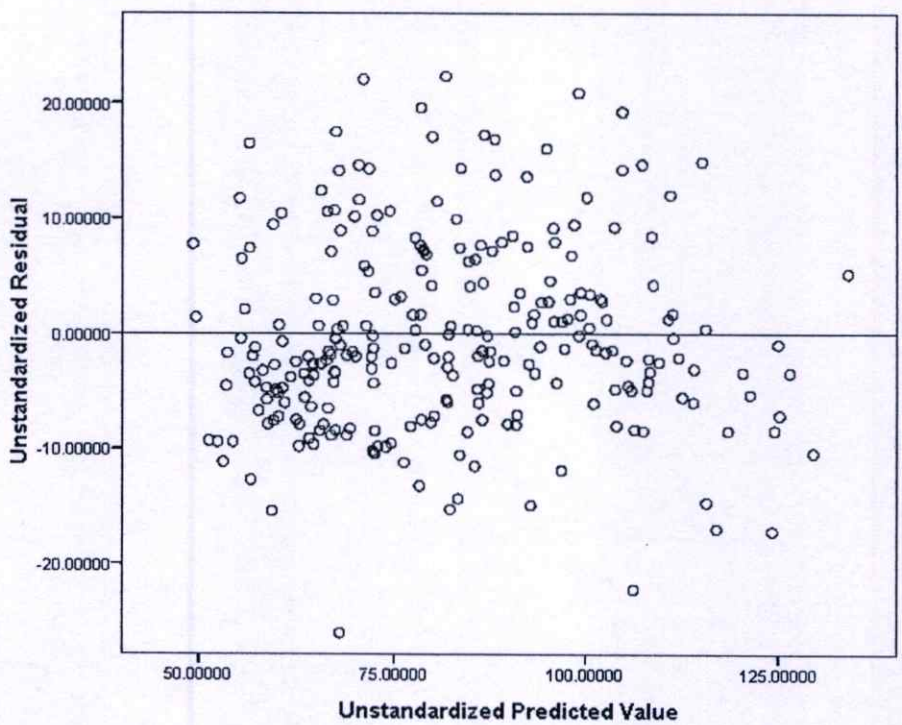
g. Dependent Variable: Y

สถิติทดสอบ

$$d = \frac{\sum_{i=2}^{n}(e_i - e_{i-1})}{\sum_{i=1}^{n} e_i^2} = 1.366$$

ค่า Durbin-Watson = 1.366 มีค่าน้อยกว่า 1.5 ดังนั้นจึงปฏิเสธ H_0 จะสรุปว่า e_i กับ e_j ไม่เป็นอิสระกัน

3) การตรวจสอบค่าความแปรปรวนของค่าความคลาดเคลื่อน โดยการพล็อตกราฟระหว่าง $E(e)$ กับ $E(\hat{Y})$



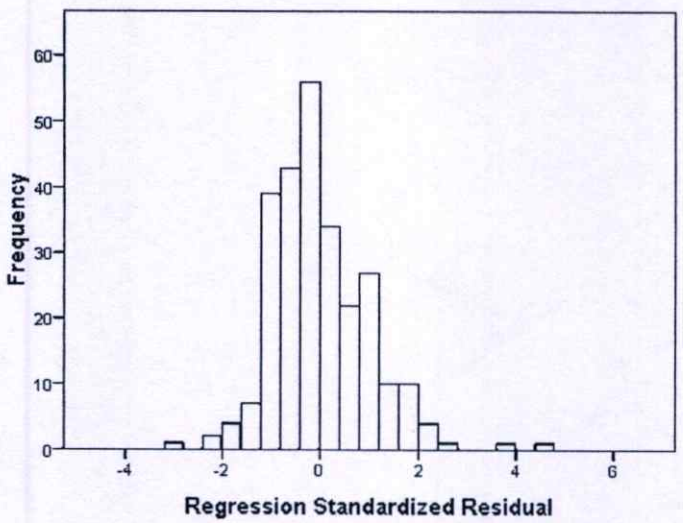
รูปที่ 4.17 ความสัมพันธ์ระหว่าง $E(e)$ กับ $E(\hat{Y})$ ของสถานีดินแดง

จากรูปที่ 4.17 พบว่าค่า e_i มีการกระจายอยู่รอบค่าศูนย์ และขนานไปกับแกน X สรุปได้ว่าค่าความแปรปรวนของค่าความคลาดเคลื่อนมีค่าคงที่

4) การตรวจสอบความเป็นอิสระกันของตัวแปรอิสระ

จากตารางที่ 4.38 พบว่า ค่า VIF ของตัวแปรอิสระทุกตัวมีค่าไม่เกิน 10 จึงสรุปได้ว่าตัวแปรอิสระไม่มีความสัมพันธ์กัน

เนื่องจากตัวแบบที่ได้ไม่ผ่านข้อสมมติเบื้องต้นของการทดสอบ ดังนั้นจะใช้การแปลงข้อมูลของตัวแปรตามด้วยวิธีของ Box-Cox Transformation จะได้ $Y' = \log Y$ เนื่องจากการแจกแจงของค่าความคลาดเคลื่อนเบ้ขวา ตามรูปที่ 4.18 ดังนั้น หลังจากแปลงข้อมูลตัวแปรตามแล้ว จึงนำไปวิเคราะห์การถดถอยเชิงเส้นพหุคูณอีกครั้ง จะได้ว่า



รูปที่ 4.18 แสดง Histogram ค่าเศษเหลือของสถานีดินแดง

4.7.4 ผลการวิเคราะห์ข้อมูลโดยใช้วิธีการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ หลังจากการแปลงข้อมูลตัวแปรตาม $Y' = \log Y$

ตรวจสอบลักษณะความสัมพันธ์ของตัวแปรตาม Y' กับตัวแปรอิสระ F ทีละคู่ พบว่า ตัวแปรตาม Y' ที่มีความสัมพันธ์กับตัวแปรอิสระ F เป็นเส้นตรง คือ Y' กับ F_1 ส่วนตัวแปรตาม Y' ที่มีความสัมพันธ์กับตัวแปรอิสระ F ไม่เป็นเส้นตรง จะต้องทำการแปลงข้อมูล นั่นคือ Y' กับ F_2 แปลง F_2 เป็นแบบ Cubic, Y' กับ F_3 แปลง F_3 เป็นแบบ Cubic, Y' กับ F_4 แปลง F_4 เป็นแบบ Cubic และ Y' กับ F_5 แปลง F_5 เป็นแบบ Cubic

1) วิธีการเพิ่มตัวแปรอิสระ (Forward Selection Procedure)

ผลการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยด้วยวิธีการเพิ่มตัวแปรอิสระ แสดงดังตารางที่ 4.41

ตารางที่ 4.41 แสดงผลการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยโดยวิธีการเพิ่มตัวแปรอิสระของตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีดินแดง หลังจากการแปลงข้อมูลตัวแปรตาม

Model		Unstandardized Coefficients		t	Sig.	Collinearity Statistics
		B	Std. Error			VIF
5	(Constant)	1.898	.003	573.762	.000	
	FAC1_1	.093	.003	28.314	.000	1.000
	FAC2_1	-.028	.003	-8.370	.000	1.006
	FAC3_1	-.016	.003	-4.930	.000	1.002
	FAC5_1	.022	.004	4.877	.000	1.204
	FAC4_1	.028	.007	4.165	.000	1.208

a. Dependent Variable:

ผลจากตารางที่ 4.41 พบว่าตัวแปรที่มีผลต่อดัชนีคุณภาพอากาศ คือ ตัวแปรในปัจจัยร่วมที่ 1 (FAC1_1), ตัวแปรในปัจจัยร่วมที่ 2 (FAC2_1), ตัวแปรในปัจจัยร่วมที่ 3 (FAC3_1), ตัวแปรในปัจจัยร่วมที่ 4 (FAC4_1) และตัวแปรในปัจจัยร่วมที่ 5 (FAC5_1) จะได้สมการการถดถอย ดังนี้

$$\text{LOG } \hat{Y} = 1.898 + 0.093F_1 - 0.028F_2 - 0.016F_3 + 0.028F_4 + 0.022F_5$$

และมีค่า $R^2 = 0.785$

2) วิธีการลดตัวแปรอิสระ (Backward Elimination Procedure)

ผลการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยด้วยวิธีการลดตัวแปรอิสระแสดงดังตารางที่ 4.42

ตารางที่ 4.42 แสดงผลการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยโดยวิธีการลดตัวแปรอิสระของตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีดินแดง หลังจากการแปลงข้อมูลตัวแปรตาม

Model		Unstandardized Coefficients		t	Sig.	Collinearity Statistics
		B	Std. Error			VIF
8	(Constant)	1.893	.004	483.795	.000	
	FAC1_1	.093	.003	28.478	.000	1.001
	FAC2_1	-.030	.003	-8.767	.000	1.078
	FAC3_1	-.016	.003	-5.008	.000	1.002
	FAC4_1	.034	.007	4.770	.000	1.391
	FAC5_1	.022	.004	5.102	.000	1.213
	FAC4_2	.016	.007	2.358	.009	1.248

a. Dependent Variable:

ผลจากตารางที่ 4.42 พบว่าตัวแปรที่มีผลต่อดัชนีคุณภาพอากาศ คือ ตัวแปรในปัจจัยร่วมที่ 1 (FAC1_1), ตัวแปรในปัจจัยร่วมที่ 2 (FAC2_1), ตัวแปรในปัจจัยร่วมที่ 3 (FAC3_1), ตัวแปรในปัจจัยร่วมที่ 4 (FAC4_1), ตัวแปรในปัจจัยร่วมที่ 4 ยกกำลังสอง (FAC4_2) และตัวแปรในปัจจัยร่วมที่ 5 (FAC5_1) ได้สมการการถดถอย ดังนี้

$$\text{LOG } \hat{Y} = 1.893 + 0.093F_1 - 0.030F_2 - 0.016F_3 + 0.034F_4 + 0.016F_4^2 + 0.022F_5$$

และมีค่า $R^2 = 0.789$

3) วิธีการเพิ่มตัวแปรอิสระทีละขั้นตอน (Stepwise Regression Procedure)

ผลการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยด้วยวิธีการเพิ่มตัวแปรอิสระทีละขั้นตอน ได้ผลเหมือนกับวิธีการเพิ่มตัวแปรอิสระ

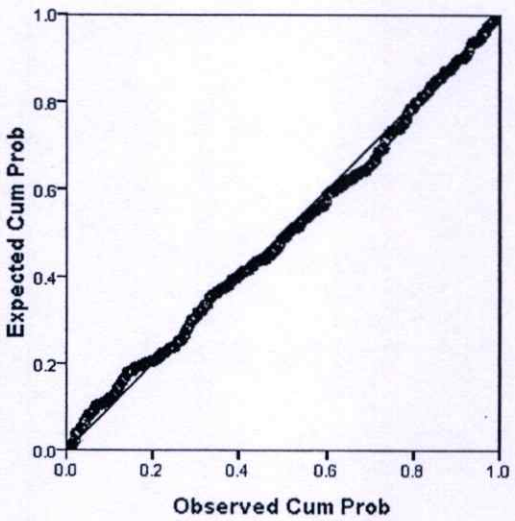
จากวิธีการคัดเลือกตัวแปรทั้ง 3 วิธี จะเลือกสมการที่ดีที่สุดด้วยวิธีลดตัวแปรอิสระ เพราะให้ค่า R^2 สูงที่สุด ดังนั้นสมการถดถอยที่ดีที่สุดสำหรับการพยากรณ์ คือ

$$\text{LOG } \hat{Y} = 1.893 + 0.093F_1 - 0.030F_2 - 0.016F_3 + 0.034F_4 + 0.016F_4^2 + 0.022F_5 \qquad (4.41)$$

4.7.5 ผลการตรวจสอบข้อสมมติของค่าความคลาดเคลื่อน หลังจากการแปลงข้อมูลตัวแปรตาม $Y' = \log Y$

1) การตรวจสอบค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติ

1.1) การพล๊อตกราฟ Normal Probability Plot



รูปที่ 4.19 กราฟ Normal Probability Plot ของค่าเศษเหลือของสมการถดถอยสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีดินแดง หลังจากการแปลงข้อมูลตัวแปรตาม

จากรูปที่ 4.19 พบว่ากราฟที่ได้เป็นเส้นตรง จึงสรุปว่าค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติ

1.2) ผลการทดสอบของ Kolmogorov-Smirnov ที่ปรับปรุงด้วยการทดสอบของ Lilliefors ได้ผลดังนี้

สมมติฐานของการทดสอบ คือ

H_0 : ค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติ

H_1 : ค่าความคลาดเคลื่อนไม่มีการแจกแจงแบบปกติ

ตารางที่ 4.43 แสดงการทดสอบค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติของสถานีดินแดง หลังจากการแปลงข้อมูลตัวแปรตาม

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Unstandardized Residual	.046	262	.200*	.985	262	.006

a. Lilliefors Significance Correction

*. This is a lower bound of the true significance.

สถิติทดสอบ Kolmogorov-Smirnov

$$D = \max\left|F(x)\text{-}S(x)\right| = 0.046$$

ค่า p-value = 0.200 มีค่ามากกว่า $\alpha = 0.01$ ดังนั้นจึงยอมรับ H_0 สรุปได้ว่าค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติ

2) การตรวจสอบความเป็นอิสระกันของค่าความคลาดเคลื่อน

สมมติฐานของการทดสอบ คือ

H_0 : ค่าความคลาดเคลื่อนเป็นอิสระกัน

H_1 : ค่าความคลาดเคลื่อนไม่เป็นอิสระกัน

ตารางที่ 4.44 แสดงค่า Durbin-Watson ที่ได้จากการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยด้วยวิธีลดตัวแปรอิสระของสถานีดินแดง

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
8	.888[h]	.789	.784	.05289826	1.546

h. Predictors: (Constant), FAC3_1, FAC1_1, FAC4_1, FAC4_2, FAC5_1, FAC2_1

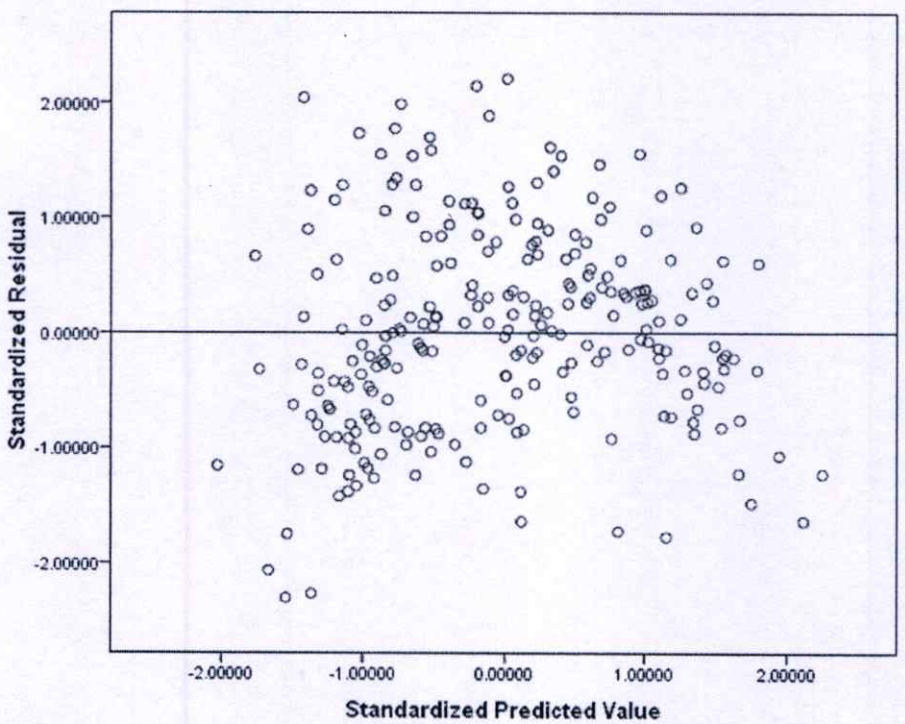
i. Dependent Variable: LOGY

สถิติทดสอบ

$$d = \frac{\sum_{i=2}^{n}(e_i - e_{i-1})}{\sum_{i=1}^{n} e_i^2} = 1.546$$

ค่า Durbin-Watson = 1.546 (นั่นคือ มีค่าในช่วง 1.5 - 2.5) ดังนั้นจึงยอมรับ H_0 จะสรุปว่า e_i กับ e_j เป็นอิสระกัน

3) การตรวจสอบค่าความแปรปรวนของค่าความคลาดเคลื่อน โดยการพล๊อตกราฟระหว่าง $E(e)$ กับ $E(\hat{Y})$



รูปที่ 4.20 ความสัมพันธ์ระหว่าง $E(e)$ กับ $E(\hat{Y})$ ของสถานีดินแดง หลังจากการแปลงข้อมูลตัวแปรตาม

จากรูปที่ 4.20 พบว่าค่า e_i มีการกระจายอยู่รอบค่าศูนย์ และขนานไปกับแกน X สรุปได้ว่าค่าความแปรปรวนของค่าความคลาดเคลื่อนมีค่าคงที่

4) การตรวจสอบความเป็นอิสระกันของตัวแปรอิสระ

จากตารางที่ 4.42 พบว่า ค่า VIF ของตัวแปรอิสระทุกตัวมีค่าไม่เกิน 10 จึงสรุปได้ว่าตัวแปรอิสระไม่มีความสัมพันธ์กัน

เนื่องจากตัวแบบที่ได้ผ่านข้อสมมติเบื้องต้นของตัวแบบทุกข้อ ดังนั้นจะได้สมการถดถอยที่ดีที่สุดสำหรับการพยากรณ์ คือ

$$\text{LOG } \hat{Y} = 1.893 + 0.093F_1 - 0.030F_2 - 0.016F_3 + 0.034F_4 + 0.016F_4^2 + 0.022F_5 \qquad (4.41)$$

และมีค่า $R^2 = 0.789$

จากสมการ 4.41 พบว่าตัวแปรอิสระที่มีผลต่อดัชนีคุณภาพอากาศ คือ ตัวแปรในปัจจัยร่วมที่ 1 (F_1), ตัวแปรในปัจจัยร่วมที่ 2 (F_2), ตัวแปรในปัจจัยร่วมที่ 3 (F_3), ตัวแปรในปัจจัยร่วมที่ 4 (F_4), ตัวแปรในปัจจัยร่วมที่ 4 ยกกำลังสอง (F_4^2), และตัวแปรในปัจจัยร่วมที่ 5 (F_5) ซึ่งสามารถอธิบายตัวแปรตามได้ 78.9%

บทที่ 5

สรุปผลการวิเคราะห์และข้อเสนอแนะ

5.1 สรุปผลการวิจัย

จากผลการวิเคราะห์ข้อมูลเพื่อนำมาสร้างตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศในเขตกรุงเทพมหานคร โดยใช้วิธีการวิเคราะห์ปัจจัยเพื่อจัดกลุ่มตัวแปรอิสระ และสร้างตัวแปรใหม่ต่อมาใช้วิธีการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ โดยใช้วิธีการคัดเลือกตัวแปรทั้ง 3 วิธี ได้แก่ วิธีการเพิ่มตัวแปรอิสระ (Forward Selection Procedure) วิธีการลดตัวแปรอิสระ (Backward Elimination Procedure) และวิธีการเพิ่มตัวแปรอิสระทีละขั้นตอน (Stepwise Regression Procedure) ในการสร้างตัวแบบ จะพิจารณาเลือกสมการที่ดีที่สุดจาก R^2 ที่มีค่าสูงที่สุด สามารถสรุปผลการวิเคราะห์ข้อมูลได้ดังนี้

5.1.1 สรุปผลการวิเคราะห์ข้อมูลเพื่อสร้างตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีการเคหะชุมชนแห่งชาติ

จากการวิเคราะห์ปัจจัยสามารถจัดกลุ่มตัวแปรจากตัวแปรอิสระ 19 ตัวแปร ได้ปัจจัยร่วมที่สำคัญ 5 ปัจจัย หรือได้ตัวแปรใหม่ 5 ตัวแปร ทำการแปลง $Y' = \log Y$ และเมื่อนำตัวแปรใหม่ไปวิเคราะห์ข้อมูลโดยใช้วิธีการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ และคัดเลือกตัวแปรด้วยวิธีการเพิ่มตัวแปรอิสระทีละขั้นตอน จะได้ตัวแบบที่เหมาะสมในการพยากรณ์ดัชนีคุณภาพอากาศของสถานีการเคหะชุมชนแห่งชาติ ดังนี้

$$\text{LOG } \hat{Y} = 1.715 + 0.064F_1 - 0.010F_2 + 0.006F_2^2 + 0.054F_3 - 0.041F_4 + 0.051F_5$$

จากผลการศึกษาดัชนีคุณภาพอากาศด้วยวิธีการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ จะได้ค่า $R^2 = 0.511$ นั่นคือ ตัวแปรอิสระสามารถอธิบายความผันแปรของตัวแปรตามได้ 51.1 % ผลการศึกษาพบว่า ตัวแปรที่มีผลต่อดัชนีคุณภาพอากาศ มีดังนี้คือ ตัวแปรในปัจจัยร่วมที่ 1, ตัวแปรในปัจจัยร่วมที่ 2, ตัวแปรในปัจจัยร่วมที่ 2 ยกกำลังสอง, ตัวแปรในปัจจัยร่วมที่ 3, ตัวแปรในปัจจัยร่วมที่ 4 และตัวแปรในปัจจัยร่วมที่ 5 ในกรณีที่ ปัจจัยร่วมที่ 1, ปัจจัยร่วมที่ 2 ยกกำลังสอง, ปัจจัยร่วมที่ 3 และปัจจัยร่วมที่ 5 มีค่าเพิ่มขึ้น ค่าดัชนีคุณภาพอากาศจะมีค่ามากขึ้น ส่งผลให้

คุณภาพอากาศแย่ลง และในกรณีที่ปัจจัยร่วมที่ 2 และปัจจัยร่วมที่ 4 มีค่าเพิ่มขึ้น ค่าดัชนีคุณภาพอากาศจะมีค่าน้อยลง ส่งผลให้คุณภาพอากาศดีขึ้น

5.1.2 สรุปผลการวิเคราะห์ข้อมูลเพื่อสร้างตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีห้วยขวาง

จากการวิเคราะห์ปัจจัยสามารถจัดกลุ่มตัวแปรจากตัวแปรอิสระ 19 ตัวแปร ได้ปัจจัยร่วมที่สำคัญ 5 ปัจจัย หรือได้ตัวแปรใหม่ 5 ตัวแปร และเมื่อนำตัวแปรใหม่ไปวิเคราะห์ข้อมูลโดยใช้วิธีการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ และคัดเลือกตัวแปรด้วยวิธีการลดตัวแปรอิสระ จะได้ตัวแบบที่เหมาะสมในการพยากรณ์ดัชนีคุณภาพอากาศของสถานีห้วยขวาง ดังนี้

$$\hat{Y} = 47.737+2.025EXP(F_1)-2.280F_2+0.778EXP(F_3)-10.980F_4+3.930F_5+0.588F_5^2-2.201F_5^3$$

โดยที่ $EXP(F_1)$ คือ ปัจจัยร่วมที่ 1 ที่ใช้การแปลงแบบเอกซ์โปเนนเชียล

$EXP(F_3)$ คือ ปัจจัยร่วมที่ 3 ที่ใช้การแปลงแบบเอกซ์โปเนนเชียล

จากผลการศึกษาดัชนีคุณภาพอากาศด้วยวิธีการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ จะได้ค่า $R^2 = 0.695$ นั่นคือ ตัวแปรอิสระสามารถอธิบายความผันแปรของตัวแปรตามได้ 69.5 % ผลการศึกษาพบว่า ตัวแปรที่มีผลต่อดัชนีคุณภาพอากาศ มีดังนี้คือ ตัวแปรในปัจจัยร่วมที่ 1 ที่ใช้การแปลงแบบเอกซ์โปเนนเชียล, ตัวแปรในปัจจัยร่วมที่ 2, ตัวแปรในปัจจัยร่วมที่ 3 ที่ใช้การแปลงแบบเอกซ์โปเนนเชียล, ตัวแปรในปัจจัยร่วมที่ 4, ตัวแปรในปัจจัยร่วมที่ 5, ตัวแปรในปัจจัยร่วมที่ 5 ยกกำลังสอง และตัวแปรในปัจจัยร่วมที่ 5 ยกกำลังสาม ในกรณีที่ ปัจจัยร่วมที่ 1, ปัจจัยร่วมที่ 3, ปัจจัยร่วมที่ 5 และปัจจัยร่วมที่ 5 ยกกำลังสอง มีค่าเพิ่มขึ้น ค่าดัชนีคุณภาพอากาศจะมีค่ามากขึ้น ส่งผลให้คุณภาพอากาศแย่ลง และในกรณีที่ปัจจัยร่วมที่ 2, ปัจจัยร่วมที่ 4 และปัจจัยร่วมที่ 5 ยกกำลังสาม มีค่าเพิ่มขึ้น ค่าดัชนีคุณภาพอากาศจะมีค่าน้อยลง ส่งผลให้คุณภาพอากาศดีขึ้น

5.1.3 สรุปผลการวิเคราะห์ข้อมูลเพื่อสร้างตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานียานนาวา

จากการวิเคราะห์ปัจจัยสามารถจัดกลุ่มตัวแปรจากตัวแปรอิสระ 19 ตัวแปร ได้ปัจจัยร่วมที่สำคัญ 5 ปัจจัย หรือได้ตัวแปรใหม่ 5 ตัวแปร และเมื่อนำตัวแปรใหม่ไปวิเคราะห์ข้อมูลโดยใช้วิธีการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ และคัดเลือกตัวแปรด้วยวิธีการเพิ่มตัวแปรอิสระทีละขั้นตอน จะได้ตัวแบบที่เหมาะสมในการพยากรณ์ดัชนีคุณภาพอากาศของสถานียานนาวา ดังนี้

$$\hat{Y} = 57.963 + 13.187F_1 + 0.897F_2 + 0.322F_2^2 + 0.583F_2^3 - 2.278F_3 - 2.772F_4 + 1.218F_4^2$$

จากผลการศึกษาดัชนีคุณภาพอากาศด้วยวิธีการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ จะได้ค่า $R^2 = 0.716$ นั่นคือ ตัวแปรอิสระสามารถอธิบายความผันแปรของตัวแปรตามได้ 71.6 % ผลการศึกษาพบว่า ตัวแปรที่มีผลต่อดัชนีคุณภาพอากาศ มีดังนี้คือ ตัวแปรในปัจจัยร่วมที่ 1, ตัวแปรในปัจจัยร่วมที่ 2, ตัวแปรในปัจจัยร่วมที่ 2 ยกกำลังสอง, ตัวแปรในปัจจัยร่วมที่ 2 ยกกำลังสาม, ตัวแปรในปัจจัยร่วมที่ 3, ตัวแปรในปัจจัยร่วมที่ 4 และตัวแปรในปัจจัยร่วมที่ 4 ยกกำลังสอง ในกรณีที่ ปัจจัยร่วมที่ 1, ปัจจัยร่วมที่ 2, ปัจจัยร่วมที่ 2 ยกกำลังสอง, ปัจจัยร่วมที่ 2 ยกกำลังสาม และปัจจัยร่วมที่ 4 ยกกำลังสอง มีค่าเพิ่มขึ้น ส่งผลให้คุณภาพอากาศแย่ลง ค่าดัชนีคุณภาพอากาศจะมีค่ามากขึ้น และในกรณีที่ปัจจัยร่วมที่ 3 และปัจจัยร่วมที่ 4 มีค่าเพิ่มขึ้น ค่าดัชนีคุณภาพอากาศจะมีค่าน้อยลง ส่งผลให้คุณภาพอากาศดีขึ้น

5.1.4 สรุปผลการวิเคราะห์ข้อมูลเพื่อสร้างตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีบางขุนเทียน

จากการวิเคราะห์ปัจจัยสามารถจัดกลุ่มตัวแปรจากตัวแปรอิสระ 19 ตัวแปร ได้ปัจจัยร่วมที่สำคัญ 4 ปัจจัย หรือได้ตัวแปรใหม่ 4 ตัวแปร และเมื่อนำตัวแปรใหม่ไปวิเคราะห์ข้อมูลโดยใช้วิธีการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ และคัดเลือกตัวแปรด้วยวิธีการเพิ่มตัวแปรอิสระทีละขั้นตอน จะได้ตัวแบบที่เหมาะสมในการพยากรณ์ดัชนีคุณภาพอากาศของสถานีบางขุนเทียน ดังนี้

$$\hat{Y} = 55.560 + 15.101F_1 - 3.556F_2 + 3.173F_3 - 1.440F_3^2 - 3.908F_4 + 2.355F_4^2 - 1.202F_4^3$$

จากผลการศึกษาดัชนีคุณภาพอากาศด้วยวิธีการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ จะได้ค่า $R^2 = 0.837$ นั่นคือ ตัวแปรอิสระสามารถอธิบายความผันแปรของตัวแปรตามได้ 83.7 % ผลการศึกษาพบว่า ตัวแปรที่มีผลต่อดัชนีคุณภาพอากาศ มีดังนี้คือ ตัวแปรในปัจจัยร่วมที่ 1, ตัวแปรในปัจจัยร่วมที่ 2, ตัวแปรในปัจจัยร่วมที่ 3, ตัวแปรในปัจจัยร่วมที่ 3 ยกกำลังสอง, ตัวแปรในปัจจัยร่วมที่ 4, ตัวแปรในปัจจัยร่วมที่ 4 ยกกำลังสอง และตัวแปรในปัจจัยร่วมที่ 4 ยกกำลังสาม ในกรณีที่ ปัจจัยร่วมที่ 1, ปัจจัยร่วมที่ 3 และปัจจัยร่วมที่ 4 ยกกำลังสอง มีค่าเพิ่มขึ้น ค่าดัชนีคุณภาพอากาศจะมีค่ามากขึ้น ส่งผลให้คุณภาพอากาศแย่ลง และในกรณีที่ปัจจัยร่วมที่ 2, ปัจจัยร่วมที่ 3 ยกกำลังสอง, ปัจจัยร่วมที่ 4 และปัจจัยร่วมที่ 4 ยกกำลังสาม มีค่าเพิ่มขึ้น ค่าดัชนีคุณภาพอากาศจะมีค่าน้อยลง ส่งผลให้คุณภาพอากาศดีขึ้น

5.1.5 สรุปผลการวิเคราะห์ข้อมูลเพื่อสร้างตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีอินทรพิทักษ์

จากการวิเคราะห์ปัจจัยสามารถจัดกลุ่มตัวแปรจากตัวแปรอิสระ 19 ตัวแปร ได้ปัจจัยร่วมที่สำคัญ 5 ปัจจัย หรือได้ตัวแปรใหม่ 5 ตัวแปร และเมื่อนำตัวแปรใหม่ไปวิเคราะห์ข้อมูลโดยใช้วิธีการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ และคัดเลือกตัวแปรด้วยวิธีการเพิ่มตัวแปรอิสระทีละขั้นตอน จะได้ตัวแบบที่เหมาะสมในการพยากรณ์ดัชนีคุณภาพอากาศของสถานีอินทรพิทักษ์ ดังนี้

$$\hat{Y} = 48.931 + 11.436F_1 + 6.765F_2 - 4.403F_3 - 5.951F_4$$

จากผลการศึกษาดัชนีคุณภาพอากาศด้วยวิธีการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ จะได้ค่า $R^2 = 0.790$ นั่นคือ ตัวแปรอิสระสามารถอธิบายความผันแปรของตัวแปรตามได้ 79 % ผลการศึกษาพบว่า ตัวแปรที่มีผลต่อดัชนีคุณภาพอากาศ มีดังนี้คือ ตัวแปรในปัจจัยร่วมที่ 1, ตัวแปรในปัจจัยร่วมที่ 2, ตัวแปรในปัจจัยร่วมที่ 3 และตัวแปรในปัจจัยร่วมที่ 4 ในกรณีที่ ปัจจัยร่วมที่ 1 และปัจจัยร่วมที่ 2 มีค่าเพิ่มขึ้น ค่าดัชนีคุณภาพอากาศจะมีค่ามากขึ้น ส่งผลให้คุณภาพอากาศแย่ลง และในกรณีที่ปัจจัยร่วมที่ 3 และปัจจัยร่วมที่ 4 มีค่าเพิ่มขึ้น ค่าดัชนีคุณภาพอากาศจะมีค่าน้อยลง ส่งผลให้คุณภาพอากาศดีขึ้น

5.1.6 สรุปผลการวิเคราะห์ข้อมูลเพื่อสร้างตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีลาดพร้าว

จากการวิเคราะห์ปัจจัยสามารถจัดกลุ่มตัวแปรจากตัวแปรอิสระ 19 ตัวแปร ได้ปัจจัยร่วมที่สำคัญ 5 ปัจจัย หรือได้ตัวแปรใหม่ 5 ตัวแปร และเมื่อนำตัวแปรใหม่ไปวิเคราะห์ข้อมูลโดยใช้วิธีการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ และคัดเลือกตัวแปรด้วยวิธีการเพิ่มตัวแปรอิสระทีละขั้นตอน จะได้ตัวแบบที่เหมาะสมในการพยากรณ์ดัชนีคุณภาพอากาศของสถานีลาดพร้าว ดังนี้

$$\hat{Y} = 57.181 + 2.888F_1 + 5.366F_2 - 2.657F_3 - 5.916F_4 - 0.163F_5 - 0.064F_5^2 - 0.174F_5^3$$

จากผลการศึกษาดัชนีคุณภาพอากาศด้วยวิธีการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ จะได้ค่า $R^2 = 0.817$ นั่นคือ ตัวแปรอิสระสามารถอธิบายความผันแปรของตัวแปรตามได้ 81.7 % ผลการศึกษาพบว่า ตัวแปรที่มีผลต่อดัชนีคุณภาพอากาศ มีดังนี้คือ ตัวแปรในปัจจัยร่วม

ที่ 1, ตัวแปรในปัจจัยร่วมที่ 2, ตัวแปรในปัจจัยร่วมที่ 3, ตัวแปรในปัจจัยร่วมที่ 4, ตัวแปรในปัจจัยร่วมที่ 5, ตัวแปรในปัจจัยร่วมที่ 5 ยกกำลังสอง และตัวแปรในปัจจัยร่วมที่ 5 ยกกำลังสาม ในกรณีที่ปัจจัยร่วมที่ 1, ปัจจัยร่วมที่ 2 และปัจจัยร่วมที่ 5 ยกกำลังสอง มีค่าเพิ่มขึ้น ค่าดัชนีคุณภาพอากาศจะมีค่ามากขึ้น ส่งผลให้คุณภาพอากาศแย่ลง และในกรณีที่ปัจจัยร่วมที่ 3, ปัจจัยร่วมที่ 4, ปัจจัยร่วมที่ 5 และปัจจัยร่วมที่ 5 ยกกำลังสาม มีค่าเพิ่มขึ้น ค่าดัชนีคุณภาพอากาศจะมีค่าน้อยลง ส่งผลให้คุณภาพอากาศดีขึ้น

5.1.7 สรุปผลการวิเคราะห์ข้อมูลเพื่อสร้างตัวแบบสำหรับพยากรณ์ดัชนีคุณภาพอากาศของสถานีดินแดง

จากการวิเคราะห์ปัจจัยสามารถจัดกลุ่มตัวแปรจากตัวแปรอิสระ 19 ตัวแปร ได้ปัจจัยร่วมที่สำคัญ 5 ปัจจัย หรือได้ตัวแปรใหม่ 5 ตัวแปร และเมื่อนำตัวแปรใหม่ไปวิเคราะห์ข้อมูลโดยใช้วิธีการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ และคัดเลือกตัวแปรด้วยวิธีลดตัวแปรอิสระ จะได้ตัวแบบที่เหมาะสมในการพยากรณ์ดัชนีคุณภาพอากาศของสถานีดินแดง ดังนี้

$$\text{LOG } \hat{Y} = 1.893 + 0.093F_1 - 0.030F_2 - 0.016F_3 + 0.034F_4 + 0.016F_4^2 + 0.022F_5$$

จากผลการศึกษาดัชนีคุณภาพอากาศด้วยวิธีการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ จะได้ค่า $R^2 = 0.789$ นั่นคือ ตัวแปรอิสระสามารถอธิบายความผันแปรของตัวแปรตามได้ 78.9 % ผลการศึกษาพบว่า ตัวแปรที่มีผลต่อดัชนีคุณภาพอากาศ มีดังนี้คือ ตัวแปรในปัจจัยร่วมที่ 1, ตัวแปรในปัจจัยร่วมที่ 2, ตัวแปรในปัจจัยร่วมที่ 3, ตัวแปรในปัจจัยร่วมที่ 4, ตัวแปรในปัจจัยร่วมที่ 4 ยกกำลังสอง และตัวแปรในปัจจัยร่วมที่ 5 ในกรณีที่ ปัจจัยร่วมที่ 1, ปัจจัยร่วมที่ 4, ปัจจัยร่วมที่ 4 ยกกำลังสอง, และปัจจัยร่วมที่ 5 มีค่าเพิ่มขึ้น ค่าดัชนีคุณภาพอากาศจะมีค่ามากขึ้น ส่งผลให้คุณภาพอากาศแย่ลง และในกรณีที่ปัจจัยร่วมที่ 2 และปัจจัยร่วมที่ 3 มีค่าเพิ่มขึ้น ค่าดัชนีคุณภาพอากาศจะมีค่าน้อยลง ส่งผลให้คุณภาพอากาศดีขึ้น

5.2 ข้อเสนอแนะ

หากต้องการให้การพยากรณ์ดัชนีคุณภาพอากาศในเขตกรุงเทพมหานครมีประสิทธิภาพมากขึ้น ควรใช้เทคนิคอื่น ๆ เพิ่มเติม เพื่อให้ได้ตัวแบบที่เหมาะสมในการพยากรณ์ดัชนีคุณภาพอากาศในแต่ละสถานี และเมื่อมีการใช้ตัวแบบไปได้สักระยะหนึ่ง ควรมีการตรวจสอบ ถ้าพบว่าค่าความคลาดเคลื่อนสูงขึ้นเรื่อย ๆ ก็ควรมีการปรับตัวแบบให้เหมาะสม นอกจากนี้อาจพิจารณาตัว

แปรอื่น ๆ ที่อาจเกี่ยวข้องกับดัชนีคุณภาพอากาศเข้ามาเพิ่มในตัวแบบ เพื่อเพิ่มประสิทธิภาพในการพยากรณ์ให้ดีมากยิ่งขึ้น

บรรณานุกรม

กัลยา วานิชย์บัญชา. 2546. **การวิเคราะห์สถิติขั้นสูงด้วย SPSS.** พิมพ์ครั้งที่ 3. กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

กัลยา วานิชย์บัญชา. 2548. **การใช้ SPSS for Windows ในการวิเคราะห์ข้อมูล.** พิมพ์ครั้งที่ 7. กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

ควบคุมมลพิษ, กรม. 2544. **การสำรวจและวิเคราะห์องค์ประกอบของโอโซนเพื่อแก้ไขปัญหามลพิษทางอากาศในกรุงเทพมหานคร.** กรุงเทพฯ: กรมควบคุมมลพิษ.

จันทนี เกียรติโพธา. 2542. "ความตระหนักในมลพิษทางอากาศของตำรวจจราจรกรุงเทพมหานคร." วิทยานิพนธ์การศึกษามหาบัณฑิต สาขาวิชาสุขศึกษา บัณฑิตวิทยาลัย, มหาวิทยาลัยศรีนครินทรวิโรฒ.

ชลิต ทับทิมทอง, ปกรณ์ ขันทอง และสุดแดน จุลศิริ. 2543. "ความสัมพันธ์ระหว่างปริมาณยานพาหนะกับปริมาณมลพิษทางอากาศในเขตกรุงเทพมหานคร" ภาคนิพนธ์วิศวกรรมศาสตรบัณฑิต สาขาวิศวกรรมโยธา คณะวิศวกรรมศาสตร์, สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ

ณัฐวสา เจริญ. 2541. "ความรู้และความตระหนักเกี่ยวกับมลพิษทางอากาศของนักเรียนชั้น ป.6 ในโรงเรียนสังกัดกรุงเทพมหานคร." วิทยานิพนธ์การศึกษามหาบัณฑิต สาขาวิชาสุขศึกษา บัณฑิตวิทยาลัย, มหาวิทยาลัยศรีนครินทรวิโรฒ.

ทรงศิริ แต้สมบัติ. 2548. **การวิเคราะห์การถดถอย.** กรุงเทพฯ: สำนักพิมพ์ฟิสิกส์เซ็นเตอร์.

บงกช ชัยเชาวรินทร์. 2548. "การแบ่งกลุ่มจังหวัดในประเทศไทยโดยใช้ตัวชี้วัดด้านการพัฒนา." วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต สาขาวิชาสถิติประยุกต์ บัณฑิตวิทยาลัย, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง.

พิริยา แก้วประเสริฐศิลป. 2545. "การพัฒนาเทคนิคในการประยุกต์ใช้แบบจำลองการกระจายตัวของมลพิษสำหรับเขตอุตสาหกรรมในการทำนายความเข้มข้นของมลพิษทางอากาศในเขตเมือง." วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสิ่งแวดล้อม บัณฑิตวิทยาลัย, มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี.

วิรัชช พานิชวงศ์. 2547. **การวิเคราะห์การถดถอย.** กรุงเทพฯ: สำนักพิมพ์สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ.

วีรญา แพ่งแสง. 2547. "ระบบสารสนเทศภูมิศาสตร์เพื่อกำหนดพื้นที่เสี่ยงต่อภาวะมลพิษทางอากาศในเขตกรุงเทพมหานคร." วิทยานิพนธ์อักษรศาสตรมหาบัณฑิต สาขาวิชาภูมิศาสตร์ บัณฑิตวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย.

ศรัณย์ ปานศรีพงษ์. 2550. "การพัฒนาแบบจำลองโครงข่ายใยประสาทประดิษฐ์และแบบจำลองความถดถอยเชิงพหุ เพื่อทำนายความเข้มข้นของ PM_{10} ในพื้นที่กรุงเทพมหานคร และเทศบาลนครราชสีมา" วิทยานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมสิ่งแวดล้อม บัณฑิตวิทยาลัย, มหาวิทยาลัยเทคโนโลยีสุรนารี.

สุชาติ ประสิทธิ์รัฐสินธุ์. 2540. **เทคนิคการวิเคราะห์ตัวแปรหลายตัวสำหรับการวิจัยทางสังคมศาสตร์และพฤติกรรมศาสตร์.** พิมพ์ครั้งที่ 4 กรุงเทพฯ: โรงพิมพ์เลี่ยงเชียง.

สุชาติ ประสิทธิ์รัฐสินธุ์ และกรรณิการ์ สุขเกษม. 2533. **เทคนิคทางสถิติขั้นสูงสำหรับการวิเคราะห์ข้อมูลด้วยไมโครคอมพิวเตอร์และโปรแกรมสำเร็จรูป SPSS PC+ เล่ม 1.** กรุงเทพฯ โรงพิมพ์ภาพพิมพ์.

โสภณา เพ็งอุบล. 2544. "ปัจจัยที่มีอิทธิพลต่อพฤติกรรมการป้องกันตนเองจากมลพิษทางอากาศและเสียงของตำรวจจราจรในเขตกรุงเทพมหานคร." วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต สาขาวิชาสุขศึกษา บัณฑิตวิทยาลัย, มหาวิทยาลัยเกษตรศาสตร์.

อภิลักษณ์ เดนไชยวงศ์. 2549. "ตัวแบบสำหรับพยากรณ์อุณหภูมิต่ำสุดของจังหวัดเชียงใหม่และจังหวัดเลย." วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต สาขาวิชาสถิติประยุกต์ บัณฑิตวิทยาลัย, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง.

อุมาพร จันทศร. 2542. **สถิติที่ไม่ใช้พารามิเตอร์.** กรุงเทพฯ: สำนักพิมพ์ฟิสิกส์เซ็นเตอร์.

Alvin C. Rencher. 2002. **Methods of Multivariate Analysis.** 2nd ed. A Wiley Interscience Publication.

Douglas C. Montgomery, et.al. 2001. **Introduction to Linear Regression Analysis.** 3rd ed. New York : John Wiley & sons, Inc.

Johnson J.D. 1992. **Applied Multivariate Data Analysis. Volume 2.** New York : Springer-Verlag.

Neter , J. et. al. 1990. **Applied Linear Statistical Model.** 3rd ed. Tokyo : Toppan Company, Ltd.

Richard A. Johnson. 2002. **Applied Multivariate Statistical Analysis.** 5th ed. New Jersey: Prentice-Hall.

Available : http://www.pcd.go.th

Available : http://www.tmd.go.th

ภาคผนวก ก.

ข้อมูลที่ใช้ในการพยากรณ์ดัชนีคุณภาพอากาศในเขตกรุงเทพมหานคร

ตารางที่ ก-1 ข้อมูลที่ใช้ในการวิเคราะห์ดัชนีคุณภาพอากาศของสถานีการเคหะชุมชนแห่งชาติ *

ตารางที่ ก-2 ข้อมูลที่ใช้ในการวิเคราะห์ดัชนีคุณภาพอากาศของสถานีห้วยขวาง *

ตารางที่ ก-3 ข้อมูลที่ใช้ในการวิเคราะห์ดัชนีคุณภาพอากาศของสถานียานนาวา *

ตารางที่ ก-4 ข้อมูลที่ใช้ในการวิเคราะห์ดัชนีคุณภาพอากาศของสถานีบางขุนเทียน *

ตารางที่ ก-5 ข้อมูลที่ใช้ในการวิเคราะห์ดัชนีคุณภาพอากาศของสถานีอินทรพิทักษ์ *

ตารางที่ ก-6 ข้อมูลที่ใช้ในการวิเคราะห์ดัชนีคุณภาพอากาศของสถานีลาดพร้าว *

ตารางที่ ก-7 ข้อมูลที่ใช้ในการวิเคราะห์ดัชนีคุณภาพอากาศของสถานีดินแดง *

* หมายเหตุทั้ง 7 ตาราง ได้เก็บไว้ในแผ่น CD ที่แนบมาพร้อมกับเอกสาร

ประวัติผู้เขียน

ชื่อ-นามสกุล	นางสาวเกศกาญจน์ บริบูรณ์
วัน เดือน ปีเกิด	19 กุมภาพันธ์ 2525 ที่จังหวัดจันทบุรี
ที่อยู่	53/3 หมู่ 7 ถนนกมลวาปี ตำบลหนองบัว อำเภอเมือง จังหวัดจันทบุรี 22000

ประวัติการศึกษา

พ.ศ.2547	สำเร็จการศึกษาวิทยาศาสตรบัณฑิต สาขาคณิตศาสตร์ จากมหาวิทยาลัยศรีนครินทรวิโรฒ
พ.ศ.2548	ศึกษาต่อระดับปริญญามหาบัณฑิต สาขาสถิติประยุกต์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง