



ใบรับรองวิทยานิพนธ์

บัณฑิตวิทยาลัย มหาวิทยาลัยเกษตรศาสตร์

วิทยาศาสตร์มหาบัณฑิต (สถิติ)

ปริญญา

สถิติ

สถิติ

สาขา

ภาควิชา

เรื่อง การวิเคราะห์ปัจจัยสำหรับตัวแบบคาดการณ์โอกาสการเกิดฝน ภาคตะวันออกเฉียงเหนือ
ของประเทศไทย

Factor Analysis for Predictive Model of Rainfall Occurrence in the Northeastern Part of
Thailand

นามผู้วิจัย นางสาวดลพรพร พันธ์พานิชย์

ได้พิจารณาเห็นชอบโดย

อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

(ผู้ช่วยศาสตราจารย์ลีลี อิงศรีสว่าง, Ph.D.)

อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม

(รองศาสตราจารย์เปรมใจ ตริสรานูวัฒนา, พศ.บ.)

อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม

(รองศาสตราจารย์ปรียานุช อภิบุญโยภาส, Ph.D.)

หัวหน้าภาควิชา

(อาจารย์อำไพ ทองธีรภาพ, Ph.D.)

บัณฑิตวิทยาลัย มหาวิทยาลัยเกษตรศาสตร์รับรองแล้ว

(รองศาสตราจารย์กัญญา ชีระกุล, D.Agr.)

คณบดีบัณฑิตวิทยาลัย

วันที่ เดือน พ.ศ.

วิทยานิพนธ์

เรื่อง

การวิเคราะห์ปัจจัยสำหรับตัวแบบคาดการณ์โอกาสการเกิดฝน ภาคตะวันออกเฉียงเหนือ
ของประเทศไทย

Factor Analysis for Predictive Model of Rainfall Occurrence in the Northeastern Part of Thailand

โดย

นางสาวดลพรพร พันธ์พานิชย์

เสนอ

บัณฑิตวิทยาลัย มหาวิทยาลัยเกษตรศาสตร์
เพื่อความสมบูรณ์แห่งปริญญาวิทยาศาสตรมหาบัณฑิต (สถิติ)

พ.ศ. 2552

ดลพรพร พันธ์พานิชย์ 2552: การวิเคราะห์ปัจจัยสำหรับตัวแบบคาดการณ์โอกาสการเกิดฝน
ภาคตะวันออกเฉียงเหนือ ของประเทศไทย ปริญญาวิทยาศาสตรมหาบัณฑิต (สถิติ) สาขาวิชาสถิติ
ภาควิชาสถิติ อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก: ผู้ช่วยศาสตราจารย์ลีลี อิงศรีสว่าง, Ph.D. 146 หน้า

งานวิจัยนี้มีวัตถุประสงค์เพื่อวิเคราะห์ปัจจัยร่วมที่เกี่ยวกับข้อมูลผลการตรวจอากาศชั้นบน สำหรับ
สร้างตัวแบบคาดการณ์โอกาสการเกิดฝนในภาคตะวันออกเฉียงเหนือ ของประเทศไทย โดยใช้ข้อมูลผลการ
ตรวจอากาศชั้นบนเป็นรายวันของสำนักฝนหลวงและการบินเกษตร ร่วมกับข้อมูลปริมาณน้ำฝน 24 ชั่วโมง
ของกรมอุตุนิยมวิทยา จำนวน 277 วัน ที่เป็นวันที่ติดต่อกันบ้างและไม่ติดต่อกันบ้าง เริ่มต้นตั้งแต่วันที่ 12 มี.ค.
2547 ถึง 14 ก.ย. 2549 มีตัวแปรทั้งหมด 49 ตัวแปร ในกลุ่มอุณหภูมิ ความชื้น ความสูงและความกดอากาศ
กระแสลม การทรงตัวของบรรยากาศและค่าพยากรณ์อากาศ และค่าพยากรณ์โอกาสความสำเร็จในการทำฝน
จากเมฆอุ่น โดยที่ตัวแปรปริมาณน้ำฝน 24 ชั่วโมง ถูกใช้เป็นตัวแปรเป้าหมายสำหรับคาดการณ์การเกิดฝน
รายวันของภาคตะวันออกเฉียงเหนือ วิธีการศึกษาใช้เทคนิคการวิเคราะห์ปัจจัย ร่วมกับการสร้างตัวแบบการ
ถดถอยโลจิสติกที่ตัวแปรตามมีค่าเป็น 2 ลักษณะ คือ ฝนตก หรือ ฝนไม่ตก ในการพัฒนาตัวแบบคาดการณ์การ
เกิดฝน ใช้ผลลัพธ์การวิเคราะห์ปัจจัยที่พิจารณาความแตกต่างใน 3 ประเด็น คือ 1) ใช้ตัวแปรเชิงปริมาณทั้งหมด
45 ตัว หรือใช้เฉพาะตัวแปรเชิงปริมาณ 28 ตัว ที่ตรวจพบว่ามีความ $VIF \geq 10$ 2) วิธีการสกัดปัจจัยแบบ Principal
Component หรือ แบบ Maximum Likelihood และ 3) จำนวนค่าคะแนนปัจจัยด้วยวิธี Regression หรือวิธี
Factor Based Score ทำให้เกิดเป็นกรณีศึกษาการวิเคราะห์ปัจจัย 8 กรณีแตกต่างกัน นำมารวมกับตัวแปรกลุ่มอีก
3 ตัว คือ โอกาสที่อุณหภูมิผิวพื้นจะถึงค่าในการเกิดเมฆ โอกาสความสำเร็จในการทำฝนจากเมฆอุ่น และตัว
แปรการเกิดฝนของวันก่อนหน้า สร้างตัวแบบการถดถอยโลจิสติกเป็น 8 ตัวแบบ

ผลการศึกษาพบว่าทั้ง 8 ตัวแบบ ด้วยค่าร้อยละความถูกต้องของการพยากรณ์การเกิดฝนรายวัน อยู่
ในช่วงร้อยละ 88.8 ถึง 91.0 ส่วนตัวแบบที่พิจารณาว่ามีความเหมาะสมที่สุด คือตัวแบบที่ใช้ผลการวิเคราะห์
ปัจจัยจากตัวแปรเชิงปริมาณ 28 ตัวแปร ที่มีค่า $VIF \geq 10$ สกัดปัจจัยด้วยวิธี Maximum Likelihood และจำนวน
ค่าคะแนนปัจจัยด้วยวิธี Factor Based Score ได้ผลลัพธ์ 3 ปัจจัยร่วม รวมกับตัวแปรกลุ่มอีก 3 ตัวข้างต้น ทำการ
คัดเลือกตัวแปรด้วยวิธีขั้นตอนก้าวหน้าแบบอัตราส่วนภาวะน่าจะเป็นที่กำหนดระดับนัยสำคัญ 0.15 สำหรับ
เลือกตัวแปรเข้า และ 0.20 สำหรับเลือกตัวแปรออก พบว่าปัจจัยที่ส่งผลต่อการคาดการณ์โอกาสการเกิดฝน ที่
ระดับนัยสำคัญทางสถิติ 0.05 ได้แก่ ปัจจัยที่เกิดจากตัวแปรในกลุ่มความชื้น กับตัวแปรการเกิดฝนของวันก่อน
หน้า โดยที่ ทุก ๆ 1 คะแนนมาตรฐานที่เพิ่มขึ้นของปัจจัยที่เกี่ยวกับความชื้น จะมีโอกาสที่ฝนจะตกในวันนี้
เพิ่มขึ้นเป็น 1.38 เท่า และถ้าวันก่อนหน้าเป็นวันที่มีฝนตกแล้วในวันนี้ฝนจะตกมีโอกาสมากกว่า 3.80 เท่า เมื่อ
เทียบกับวันก่อนหน้าที่ฝนไม่ตก เมื่อตัวแปรอื่น ๆ ในตัวแบบมีค่าคงที่ และตัวแบบนี้ให้ค่าความถูกต้องของการ
พยากรณ์สูงถึงร้อยละ 90.3

Dollapas Punpanich 2009: Factor Analysis for Predictive Model of Rainfall Occurrence in the Northeastern Part of Thailand. Master of Science (Statistics), Major Field: Statistics, Department of Statistics. Thesis Advisor: Assistant Professor Lily Ingsrisawang, Ph.D. 146 pages.

The purpose of this research was the analysis of the common factors of daily upper air data for the development of a model of rainfall occurrences in Northeastern Thailand. The daily upper air observations were provided by The Bureau of the Royal Rain Making and Agricultural Aviation and the records of rainfall in 24 hour period were obtained from the Department of Meteorology. Both data sets contained 277 inconsistent records for the period of 12 March 2004 – 14 September 2006. Each record consisted of 49 variables such as temperature, humidity, pressure, wind, stability and forecasting indices, and warm cloud seeding potential. The data on rainfall volume were categorized into rain or no-rain events. Two statistical methods of factor analysis and logistic regression were applied for developing the prediction model of rain event in Northeastern Thailand. The factor analysis was first conducted by considering three interesting issues as following: 1) the selection of 45 quantitative variables or 28 quantitative variables with $VIF \geq 10$, 2) the extraction of common factors using the Principal Component or the Maximum Likelihood method, and 3) the estimation of factor score using the method of Regression or Factor Based Score. As a result, the different eight alternatives of factor analysis were studied. Next the logistic regression was performed using the number of factors obtained from the above result of each alternative and three additional qualitative variables: convective temperature reachable, warm cloud seeding potential, and yesterday's rain event as the independent variables in the development of eight difference predictive models of rainfall occurrences.

The predictive power of the eight rainfall models were measured by the percentages of prediction accuracy or hit rate statistics. It appeared that these models could achieve the hit rates ranged from 88.8% to 91.0%. The most suitable model was the model that extracted three common factors from the combination of using 28 quantitative variables with $VIF \geq 10$, factor extraction by maximum likelihood, and estimation of factor score by the Factor Based Score method. These three factors and the three stated qualitative variables were modeled under the procedure of forward stepwise variable selection with a significance level of 0.15 for entry and 0.20 for removal. The results showed that the statistically significant factors at 0.05 level consisted of one common factor of five humidity variables and the other qualitative variable of yesterday's rain event with their odds ratios (OR) of 1.38 and 3.80, respectively. This model gave the accuracy of prediction about 90.3%

Student's signature

Thesis Advisor's signature

กิตติกรรมประกาศ

กราบขอบพระคุณ ผศ.ดร.ลีลี อิงศรีสว่าง ประธานกรรมการที่ปรึกษาวิทยานิพนธ์ สำหรับความรู้และประสบการณ์ที่มีประโยชน์อย่างมากในชีวิตการทำงานในปัจจุบัน รศ.เปรมใจ ตรีสรานุวัฒนา รศ.ดร.ปริยานุช อภิบุญโยภาส อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม และผศ.ดร.นิภา โรจน์รุ่งวศินกุล ผู้ทรงคุณวุฒิภายนอก สำหรับคำแนะนำที่ดีต่อการแก้ปัญหาตลอดจนการตรวจแก้ไขวิทยานิพนธ์กระทั่งเสร็จสมบูรณ์

ขอขอบคุณเจ้าหน้าที่กรมอุตุวิทยุ และเจ้าหน้าที่ของศูนย์ปฏิบัติการฝนหลวงเฉลิมพระเกียรติ อำเภอหัวหิน จังหวัดประจวบคีรีขันธ์ สำหรับข้อมูลการศึกษา

กราบขอบพระคุณ คุณพ่อ คุณแม่ และพี่ชาย สำหรับความอบอุ่นในครอบครัวพันธุ์พานิชย์ที่ทำให้ข้าพเจ้าเต็มทีกับชีวิตได้ทุกวัน ขอขอบคุณคุณโชคชัย สุเวชวัฒนกุล สำหรับพลังชีวิต และสุดท้ายขอขอบคุณเพื่อนที่แสนดี นายดิษฐ์พล มั่นธรรม สำหรับความเข้าใจ และกำลังใจที่ดีเสมอมา

ดลพรพร พันธุ์พานิชย์

ตุลาคม 2552

สารบัญ

	หน้า
สารบัญ	(1)
สารบัญตาราง	(2)
สารบัญภาพ	(6)
คำนำ	1
วัตถุประสงค์	4
การตรวจเอกสาร	10
อุปกรณ์และวิธีการ	61
อุปกรณ์	61
วิธีการ	61
ผลและวิจารณ์	80
ผล	80
วิจารณ์	132
สรุปและข้อเสนอแนะ	134
สรุป	134
ข้อเสนอแนะ	136
เอกสารและสิ่งอ้างอิง	137
ภาคผนวก	141
ภาคผนวก ก การพล็อตค่า Residual คู่กับ Predicted probability (\hat{Y}) ของแต่ละตัวแบบ	142
ประวัติการศึกษา และการทำงาน	146

สารบัญตาราง

ตารางที่		หน้า
1	รายละเอียดของตัวแปรที่ใช้ในการศึกษา	5
2	ชนิดของเมฆ จำแนกตามลักษณะการเกิดและความสูง	10
3	ตัวแปรที่เกี่ยวข้องกับการพยากรณ์อากาศ และโอกาสความสำเร็จในการเกิดฝน	16
4	แสดงขนาดตัวอย่างกับค่าน้ำหนักปัจจัย ในการพิจารณาจำนวนปัจจัยที่เหมาะสม	44
5	ระยะเวลาข้อมูล จำนวนตัวแปร และจำนวนรายการข้อมูล (วัน) ของข้อมูลต่าง ๆ	63
6	จำนวนรายการ (วัน) ที่มีข้อมูลสูญหาย ในแต่ละตัวแปร	64
7	ค่าสถิติพื้นฐานของตัวแปรเชิงปริมาณ จากชุดข้อมูล 277 รายการ	80
8	ค่าสถิติพื้นฐานของตัวแปรเชิงคุณภาพ จากชุดข้อมูล 277 รายการ	83
9	ผลการตรวจสอบความเหมาะสมของข้อมูล 45 ตัวแปร (กรณี 1 – 4) ใช้เทคนิคการวิเคราะห์ปัจจัยด้วยค่าสถิติ KMO และค่าสถิติ Bartlett's Test of Sphericity	84
10	ผลการวิเคราะห์น้ำหนักปัจจัยก่อนการหมุนแกน และร้อยละความผันแปรของแต่ละปัจจัย ที่ได้จากการสกัดปัจจัยแบบ PCM จากตัวแปรเชิงปริมาณ 45 ตัวแปร	85
11	ผลการวิเคราะห์น้ำหนักปัจจัยหลังการหมุนแกนแบบ Varimax และร้อยละความผันแปรของแต่ละปัจจัย ที่ได้จากการสกัดปัจจัยแบบ PCM จากตัวแปรเชิงปริมาณ 45 ตัวแปร	88
12	เมตริกซ์สัมประสิทธิ์คะแนนปัจจัยร่วม (Component Score Coefficient Matrix) ที่ได้จากการสกัดปัจจัยแบบ PCM จากตัวแปรเชิงปริมาณ 45 ตัวแปร	91
13	แสดงตัวอย่างการคำนวณค่าคะแนนปัจจัยโดยใช้วิธี Regression (กรณีที่ 1) และ Factor Based Score (กรณีที่ 2) เมื่อใช้วิธีสกัดปัจจัยด้วยเทคนิค PCM	94
14	ผลการวิเคราะห์น้ำหนักปัจจัยก่อนการหมุนแกนแบบ Varimax และร้อยละความผันแปรของแต่ละปัจจัย ที่ได้จากการสกัดปัจจัยแบบ MLM จากตัวแปรเชิงปริมาณ 45 ตัวแปร	95
15	ผลการวิเคราะห์น้ำหนักปัจจัยหลังการหมุนแกนแบบ Varimax และร้อยละความผันแปรของแต่ละปัจจัย ที่ได้จากการสกัดปัจจัยแบบ MLM จากตัวแปรเชิงปริมาณ 45 ตัวแปร	98

สารบัญตาราง (ต่อ)

ตารางที่		หน้า
16	เมตริกซ์สัมประสิทธิ์คะแนนปัจจัยร่วม (Component score coefficient matrix) ที่ได้จากการสกัดปัจจัยแบบ MLM จากตัวแปรเชิงปริมาณ 45 ตัวแปร	101
17	แสดงตัวอย่างการคำนวณค่าคะแนนปัจจัยโดยใช้วิธี Regression (กรณีที่ 3) และ Factor Based Score (กรณีที่ 4) เมื่อใช้วิธีสกัดปัจจัยด้วยเทคนิค MLM	104
18	ผลการตรวจสอบความเหมาะสมของตัวแปรเชิงปริมาณ 28 ตัวแปร (กรณี 5 – 8) ใช้เทคนิคการวิเคราะห์ปัจจัยด้วยค่าสถิติ KMO และค่าสถิติ Bartlett's Test of Sphericity	105
19	ผลการวิเคราะห์น้ำหนักปัจจัยก่อนการหมุนแกนแบบ Varimax และร้อยละความผันแปรของแต่ละปัจจัย ที่ได้จากการสกัดปัจจัยแบบ PCM จากตัวแปรเชิงปริมาณ 28 ตัวแปร	105
20	ผลการวิเคราะห์น้ำหนักปัจจัยหลังการหมุนแกนแบบ Varimax และร้อยละความผันแปรของแต่ละปัจจัย ที่ได้จากการสกัดปัจจัยแบบ PCM จากตัวแปรเชิงปริมาณ 28 ตัวแปร	108
21	เมตริกซ์สัมประสิทธิ์คะแนนปัจจัยร่วม (Component score coefficient matrix) ที่ได้จากการสกัดปัจจัยแบบ PCM จากตัวแปรเชิงปริมาณ 28 ตัวแปร	110
22	แสดงตัวอย่างการคำนวณค่าคะแนนปัจจัยโดยใช้วิธี Regression (กรณีที่ 5) และ Factor Based Score (กรณีที่ 6) เมื่อใช้วิธีสกัดปัจจัยด้วยเทคนิค PCM	112
23	ผลการวิเคราะห์น้ำหนักปัจจัยก่อนการหมุนแกนแบบ Varimax และร้อยละความผันแปรของแต่ละปัจจัย ที่ได้จากการสกัดปัจจัยแบบ MLM จากตัวแปรเชิงปริมาณ 28 ตัวแปร	113
24	ผลการวิเคราะห์น้ำหนักปัจจัยหลังการหมุนแกนแบบ Varimax และร้อยละความผันแปรของแต่ละปัจจัย ที่ได้จากการสกัดปัจจัยแบบ MLM จากตัวแปรเชิงปริมาณ 28 ตัวแปร	115
25	เมตริกซ์สัมประสิทธิ์คะแนนปัจจัย (Component score coefficient matrix) ที่ได้จากการสกัดปัจจัยแบบ MLM จากตัวแปรเชิงปริมาณ 28 ตัวแปร	118

สารบัญตาราง (ต่อ)

ตารางที่		หน้า
26	แสดงตัวอย่างการคำนวณค่าคะแนนปัจจัยโดยใช้วิธี Regression (กรณีที่ 7) และ Factor Based Score (กรณีที่ 8) เมื่อใช้วิธีสกัดปัจจัยด้วยเทคนิค MLM	119
27	การตรวจสอบความเหมาะสมด้วยค่าสถิติ ในการวิเคราะห์ปัจจัยทั้ง 8 กรณี	120
28	ผลการวิเคราะห์ถดถอยโลจิสติก กรณีที่ไม่พิจารณาค่า VIF จึงทำให้ใช้ตัวแปรที่เป็นข้อมูลเชิงปริมาณทั้ง 45 ตัวแปร โดยสกัดปัจจัยด้วย PCM แล้วคำนวณค่าคะแนนปัจจัยด้วยวิธี Regression (กรณีที่ 1)	122
29	ผลการวิเคราะห์ถดถอยโลจิสติก กรณีที่ไม่พิจารณาค่า VIF จึงทำให้ใช้ตัวแปรที่เป็นข้อมูลเชิงปริมาณทั้ง 45 ตัวแปร โดยสกัดปัจจัยด้วย PCM แล้วคำนวณค่าคะแนนปัจจัยด้วยวิธี Factor Based Score (กรณีที่ 2)	123
30	ผลการวิเคราะห์ถดถอยโลจิสติก กรณีที่ไม่พิจารณาค่า VIF จึงทำให้ใช้ตัวแปรที่เป็นข้อมูลเชิงปริมาณทั้ง 45 ตัวแปร โดยสกัดปัจจัยด้วย MLM แล้วคำนวณค่าคะแนนปัจจัยด้วยวิธี Regression (กรณีที่ 3)	124
31	ผลการวิเคราะห์ถดถอยโลจิสติก กรณีที่ไม่พิจารณาค่า VIF จึงทำให้ใช้ตัวแปรที่เป็นข้อมูลเชิงปริมาณทั้ง 45 ตัวแปร โดยสกัดปัจจัยด้วย MLM แล้วคำนวณค่าคะแนนปัจจัยด้วยวิธี Factor Based Score (กรณีที่ 4)	125
32	ผลการวิเคราะห์ถดถอยโลจิสติก กรณีที่พิจารณาตัวแปรที่มีค่า VIF มากกว่าหรือเท่ากับ 10 จึงทำให้ใช้ตัวแปรที่เป็นข้อมูลเชิงปริมาณ 28 ตัวแปร โดยสกัดปัจจัยด้วย PCM แล้วคำนวณค่าคะแนนปัจจัยด้วยวิธี Regression (กรณีที่ 5)	126
33	ผลการวิเคราะห์ถดถอยโลจิสติก กรณีที่พิจารณาตัวแปรที่มีค่า VIF มากกว่าหรือเท่ากับ 10 จึงทำให้ใช้ตัวแปรที่เป็นข้อมูลเชิงปริมาณ 28 ตัวแปร โดยสกัดปัจจัยด้วย PCM แล้วคำนวณค่าคะแนนปัจจัยด้วยวิธี Factor Based Score (กรณีที่ 6)	127
34	ผลการวิเคราะห์ถดถอยโลจิสติก กรณีที่พิจารณาตัวแปรที่มีค่า VIF มากกว่าหรือเท่ากับ 10 จึงทำให้ใช้ตัวแปรที่เป็นข้อมูลเชิงปริมาณ 28 ตัวแปร โดยสกัดปัจจัยด้วย MLM แล้วคำนวณค่าคะแนนปัจจัยด้วยวิธี Regression (กรณีที่ 7)	128

สารบัญตาราง (ต่อ)

ตารางที่		หน้า
35	ผลการวิเคราะห์ถดถอยโลจิสติก กรณีที่พิจารณาตัวแปรที่มีค่า VIF มากกว่าหรือเท่ากับ 10 จึงทำให้ใช้ตัวแปรที่เป็นข้อมูลเชิงปริมาณ 28 ตัวแปร โดยสกัดปัจจัยด้วย MLM แล้วคำนวณค่าคะแนนปัจจัยด้วยวิธี Factor base (กรณีที่มี 8)	129
36	ผลการวิเคราะห์ถดถอยโลจิสติกแบบทวิ 8 ตัวแบบ	130
37	สรุปผลการวิเคราะห์ปัจจัย 8 กรณี และการวิเคราะห์ถดถอยโลจิสติกแบบทวิ 8 ตัวแบบ	131

สารบัญภาพ

ภาพที่		หน้า
1	แผนภาพการวิเคราะห์ปัจจัยสำหรับตัวแบบคาดการณ์โอกาสการเกิดฝน	9
2	การทดสอบ Scree test	38
3	การหมุนแกนแบบตั้งฉากในกรณีที่มี 2 ปัจจัย	42
4	การหมุนแกนแบบตั้งฉากในกรณีที่มี 2 ปัจจัย ด้วยวิธี Varimax	43
5	กราฟ Logistic เมื่อ $\beta_1 > 0$ และกราฟ Logistic เมื่อ $\beta_1 < 0$	48
6	แสดงการเปรียบเทียบฟังก์ชันการถดถอย ระหว่างรูปแบบเส้นโค้งการถดถอย โลจิสติก กับรูปแบบเส้นตรงของตัวแบบการถดถอยอย่างง่าย	50
7	แสดงรูปแบบการเคลื่อนไหว (แนวโน้ม) ของข้อมูลปริมาณน้ำฝน 24 ชั่วโมง ระหว่างวันที่ 12 มีนาคม พ.ศ. 2547 ถึงวันที่ 14 กันยายน พ.ศ. 2549	65
8	แผนภาพการกระจาย ระหว่างตัวแปร RAOB_SURFACE_TEMPERATURE กับตัวแปรปริมาณน้ำฝน 24 ชม.	66
9	แผนภาพการกระจาย ระหว่างตัวแปร PRECIPITABLE_WATER_SFC850 กับตัวแปรปริมาณน้ำฝน 24 ชม.	66
10	แผนภาพการกระจาย ระหว่างตัวแปร LEVEL_OF_FREE_CONVECTION_LFC_wrtTma กับตัวแปรปริมาณน้ำฝน 24 ชม.	67
11	แผนภาพการกระจาย ระหว่างตัวแปร MEAN_WIND_1000_5000_FT_DEG กับตัวแปรปริมาณน้ำฝน 24 ชม.	67
12	แผนภาพการกระจาย ระหว่างตัวแปร K_INDEX กับตัวแปรปริมาณน้ำฝน 24 ชม.	68
13	แผนภาพการกระจาย ระหว่างตัวแปร ISOTHERM_HEIGHTS_minus5c กับตัวแปรปริมาณน้ำฝน 24 ชม.	68
14	แผนภาพการกระจาย ระหว่างตัวแปร TOP_OF_LATENT_INSTABILITY_LAYER_wrt กับตัวแปรปริมาณน้ำฝน 24 ชม.	69
15	แผนภาพการกระจาย ระหว่างตัวแปร POTENTIAL_BUOYANCY_INDEX กับตัวแปรปริมาณน้ำฝน 24 ชม.	69

สารบัญภาพ (ต่อ)

ภาพที่		หน้า
16	แผนผังแสดงจำนวนตัวแปรอิสระเชิงปริมาณ ที่ใช้สำหรับการวิเคราะห์ปัจจัย 8 กรณี	73
17	แผนผังแสดงลำดับการวิเคราะห์ปัจจัยเพื่อสร้างตัวแบบโลจิสติก 8 ตัวแบบ	78
18	กราฟ Scree Plot จากการวิเคราะห์ตัวแปรเชิงปริมาณ 45 ตัวแปร ด้วยวิธี PCM	87
19	กราฟ Scree Plot จากการวิเคราะห์ตัวแปรเชิงปริมาณ 45 ตัวแปร ด้วยวิธี MLM	97
20	กราฟ Scree Plot จากการวิเคราะห์ตัวแปรเชิงปริมาณ 28 ตัวแปร ด้วยวิธี PCM	107
21	กราฟ Scree Plot จากการวิเคราะห์ตัวแปรเชิงปริมาณ 28 ตัวแปร ด้วยวิธี MLM	114
22	แผนผังแสดงการลำดับขั้นตอนของการวิเคราะห์เหตุการณ์ทั้ง 8 ตัวแบบ	121

การวิเคราะห์ปัจจัยสำหรับตัวแบบคาดการณ์โอกาสการเกิดฝน

ภาคตะวันออกเฉียงเหนือ ของประเทศไทย

Factor Analysis for Predictive Model of Rainfall Occurrence

in the Northeastern Part of Thailand

คำนำ

ภาคตะวันออกเฉียงเหนือประกอบด้วย 19 จังหวัด คือ หนองคาย นครพนม มุกดาหาร ชัยภูมิ สกลนคร เลย ขอนแก่น มหาสารคาม กาฬสินธุ์ ร้อยเอ็ด ยโสธร นครราชสีมา บุรีรัมย์ สุรินทร์ ศรีสะเกษ อุบลราชธานี อุดรธานี หนองบัวลำภู และอำนาจเจริญ เป็นภูมิภาคที่มีพื้นที่และประชากรมากที่สุดในประเทศไทย ลักษณะภูมิประเทศเป็นที่ราบสูงที่เอียงลาดจากด้านตะวันตกไปทางด้านตะวันออก มีเทือกเขาสูงที่สำคัญรวม 5 เทือกเขา คือ เทือกเขาเพชรบูรณ์ เทือกเขาแดงพญาเย็น เทือกเขาสันกำแพง เทือกเขาพนมดงรัก และเทือกเขาภูพาน โดยที่เทือกเขาเพชรบูรณ์ และเทือกเขาแดงพญาเย็น มีความสูงประมาณ 800 ถึง 1,300 เมตร และเทือกเขาพนมดงรักมีความสูงประมาณ 400 เมตร เป็นแนวกั้นกระแสลมตะวันตกเฉียงใต้ เป็นสาเหตุที่ทำให้บริเวณด้านหลังเขา ซึ่งเป็นพื้นที่ทางด้านตะวันออกของภาคตะวันออกเฉียงเหนือมีปริมาณฝนน้อยกว่าทางภาคตะวันออก และมีแม่น้ำสำคัญ 2 สาย คือ แม่น้ำชี และน้ำมูล ซึ่งมีต้นกำเนิดอยู่ทางตะวันตกของภูมิภาค ไหลลงทางทิศตะวันออก ไปบรรจบกันที่บริเวณจังหวัดอุบลราชธานี แล้วไหลลงสู่แม่น้ำโขง บริเวณอำเภอโขงเจียม จังหวัดอุบลราชธานี (รังสรรค์, 2547)

ในปี 2551 กรมพัฒนาที่ดินได้สำรวจพื้นที่แห้งแล้งซ้ำซาก พบว่าพื้นที่แห้งแล้งส่วนใหญ่จะอยู่บริเวณภาคภาคตะวันออกเฉียงเหนือตอนล่าง ประกอบไปด้วยพื้นที่บางส่วนของจังหวัด นครราชสีมา ชัยภูมิ บุรีรัมย์ สุรินทร์ และศรีสะเกษ ส่วนในภาคตะวันออกเฉียงเหนือตอนบน เป็นพื้นที่จังหวัดขอนแก่น ชัยภูมิ (ตอนบน) และบางส่วนของทุ่งกุลาร้องไห้ มีพื้นที่ประมาณ 2 ล้านไร่ ครอบคลุม 5 จังหวัด คือ ร้อยเอ็ด สุรินทร์ มหาสารคาม ยโสธร และศรีสะเกษ โดยปัจจุบันรัฐบาลได้พยายามปรับปรุงพื้นที่ให้ดีขึ้น ด้วยการใช้นวัตกรรมชลประทานสมัยใหม่ ทำให้สามารถทำการเพาะปลูกได้จนกลายเป็นแหล่งเพาะปลูกข้าวหอมมะลิที่ดีที่สุดแห่งหนึ่งของประเทศไทย แต่ถึง

อย่างไรก็ยังปลูกได้เฉพาะฤดูฝนเท่านั้น ในช่วงฤดูแล้งสามารถทำการเพาะปลูกได้เฉพาะบางส่วน ซึ่งยังไม่ครอบคลุมบริเวณพื้นที่ที่ต้องการเพาะปลูกทั้งหมด

ฝ่ายทรัพยากรธรรมชาติและสิ่งแวดล้อม สถาบันวิจัยเพื่อการพัฒนาประเทศไทย ได้รายงานผลการสำรวจข้อมูลความต้องการใช้น้ำในการอุปโภคบริโภค และกิจกรรมด้านต่าง ๆ ปี พ.ศ. 2536 พ.ศ.2539 และ พ.ศ.2549 จำแนกตามลุ่มน้ำของภาคต่าง ๆ ทั่วประเทศ โดยเฉพาะลุ่มน้ำทางภาคตะวันออกเฉียงเหนือ พบว่ามีความต้องการใช้น้ำส่วนใหญ่ในกิจกรรมภาคการเกษตร คือความต้องการใช้น้ำเพิ่มขึ้นจาก 6,389 ล้านลูกบาศก์เมตร ในปี พ.ศ. 2536 เป็น 8,409 ล้านลูกบาศก์เมตรในปี พ.ศ. 2539 และ 11,814 ล้านลูกบาศก์เมตรในปี พ.ศ. 2549 และมีรายงานการขาดแคลนน้ำเพิ่มขึ้นจาก 961 ล้านลูกบาศก์เมตรในปี พ.ศ. 2536 เป็น 1,003 ล้านลูกบาศก์เมตรในปี พ.ศ. 2539 และ 2,637 ล้านลูกบาศก์เมตรในปี พ.ศ. 2549 ตามลำดับ ทำให้ส่งผลกระทบต่อภาคการเกษตรของภูมิภาคนี้ เนื่องจากประชากรส่วนใหญ่ประกอบอาชีพเกษตรกรเป็นหลัก เมื่อพิจารณาพื้นที่การเกษตรของทั้งประเทศที่มี 122.4 ล้านไร่ จะเป็นพื้นที่ที่มีระบบชลประทานเพื่อการเกษตรเพียง 31.2 ล้านไร่ หรือร้อยละ 25.5 ของพื้นที่การเกษตร แต่ภาคตะวันออกเฉียงเหนือมีพื้นที่ทางการเกษตรที่มากที่สุดของประเทศ คือมีสัดส่วนพื้นที่การเกษตร 57.7 ล้านไร่ หรือร้อยละ 44.4 ของพื้นที่การเกษตรทั้งประเทศ แต่มีพื้นที่ชลประทานเพียง 3.7 ล้านไร่ หรือร้อยละ 6.4 ของพื้นที่เพาะปลูก อีกทั้งลักษณะสภาพดินภูมิภาคนี้ยังเป็นดินเค็ม ดินเป็นกรด และดินค่อนข้างเป็นทรายอีกด้วย ทำให้ดินขาดอินทรีย์วัตถุประมาณร้อยละ 76 ส่งผลกระทบต่อการใช้ประโยชน์ทางด้านเกษตรกรรมได้ไม่เต็มประสิทธิภาพ (สถาบันวิจัยเพื่อการพัฒนาประเทศไทย, ม.ป.ป.)

“น้ำ” จึงเป็นปัจจัยที่สำคัญต่อการพัฒนาพื้นที่ทางการเกษตรของภูมิภาคนี้ ถ้ามีปริมาณน้ำที่พอเพียงต่อภาคการเกษตร ประชากรในภูมิภาคก็จะสามารถประกอบอาชีพได้เต็มศักยภาพ ทำให้ปัญหาสังคมของประเทศลดลงได้ เช่น ลดปัญหาความยากจนของประชากร ลดอัตราการย้ายถิ่น และเพิ่มคุณภาพชีวิตของประชากรในภูมิภาคนี้ได้ยั่งยืน โดยหน่วยงานที่มีความสำคัญและให้ความช่วยเหลือด้านชลประทานของชาวเกษตรกรมาโดยตลอดก็คือ สำนักฝนหลวงและการบินเกษตร ซึ่งปัจจุบันมีศูนย์ปฏิบัติการฝนหลวงทั่วประเทศจำนวน 8 ศูนย์หลัก อยู่ในภาคตะวันออกเฉียงเหนือ 2 ศูนย์ คือศูนย์ปฏิบัติการฝนหลวงดอนบนที่จังหวัดขอนแก่น และศูนย์ปฏิบัติการฝนหลวงดอนล่างที่จังหวัดนครราชสีมา โดยศูนย์ปฏิบัติการฝนหลวงแต่ละแห่งมีหน้าที่ปฏิบัติการฝนหลวงช่วยเหลือประชาชนและมีการรวบรวมข้อมูลที่เป็นประโยชน์ต่อระบบสารสนเทศฝนหลวง สำหรับใช้วางแผนการตัดสินใจเกี่ยวกับการปฏิบัติการฝนหลวง (ทรง, 2551) และในการปฏิบัติการฝน

หลวงสำหรับแต่ละพื้นที่ จำเป็นต้องมีการนำข้อมูลที่ได้จากการคาดการณ์ลักษณะสภาพภูมิอากาศ และผลการพยากรณ์อากาศจากข้อมูลของส่วนต่าง ๆ มาประกอบการตัดสินใจการขึ้นปฏิบัติการประจำวัน (กระทรวงเกษตรและสหกรณ์, ม.ป.ป.)

โดยข้อมูลผลการตรวจอากาศชั้นบนนั้นจะมาจากการประมวลผลด้วยโปรแกรม GPCM (Great Plains Cloud Model) ซึ่งเป็นแบบจำลองมิติเดียวชนิด One-dimensional หรือ Steady-state model (วราวุธ, 2539) ซึ่งต้องอาศัยข้อมูลทั้งหมดจาก 6 กลุ่มข้อมูล คือ 1) ข้อมูลอุณหภูมิ 2) ความชื้น 3) ความสูงและความกดอากาศ 4) กระแสลม 5) การทรงตัวของบรรยากาศและค่าพยากรณ์อากาศ และ 6) ค่าพยากรณ์โอกาสความสำเร็จในการทำฝนจากเมฆอุ่น จึงเป็นหน้าที่ของผู้ที่มีส่วนเกี่ยวข้องกับการปฏิบัติการฝนหลวงที่ต้องให้ความสำคัญกับตัวแปรต่าง ๆ ที่มีผลต่อโอกาสการเกิดฝนตกประจำวันก่อนตัดสินใจขึ้นปฏิบัติการทำฝน ทั้งนี้เพื่อเป็นการเพิ่มประสิทธิภาพของการปฏิบัติการฝนหลวงให้ประสบความสำเร็จสูงสุด

ดังนั้นวิธีทางสถิติสามารถนำมาประยุกต์เพื่อช่วยตอบคำถามเกี่ยวกับตัวแปรที่มีความสำคัญต่อการคาดการณ์โอกาสการเกิดฝน และเนื่องจากตัวแปรข้อมูลผลการตรวจอากาศชั้นบนมีจำนวนมาก และมักจะมีความสัมพันธ์กัน ผู้วิจัยจึงได้พิจารณาวิธีการวิเคราะห์ปัจจัย (Factor analysis) ซึ่งเป็นวิธีทางสถิติที่จะช่วยจัดกลุ่มตัวแปรที่มีความสัมพันธ์กันในระดับเดียวกันเป็นตัวแปรใหม่เรียกแทนว่าปัจจัย (Factor) แล้วจึงนำปัจจัยที่ได้เหล่านี้ ไปสร้างตัวแบบคาดการณ์โอกาสการเกิดฝนด้วยวิธีการวิเคราะห์การถดถอยโลจิสติกแบบทวิ (Binary Logistic Regression Analysis) ที่ตัวแปรตามมีค่าเป็น 2 ลักษณะ คือ ฝนตก หรือฝนไม่ตก เพื่อใช้ในการพิจารณาประกอบการตัดสินใจการขึ้นปฏิบัติการฝนหลวงประจำวัน

วัตถุประสงค์

1. เพื่อหาปัจจัยร่วมที่เกี่ยวข้องกับข้อมูลผลการตรวจอากาศชั้นบน ของภาคตะวันออกเฉียงเหนือ
2. เพื่อสร้างตัวแบบคาดการณ์โอกาสการเกิดฝนในภาคตะวันออกเฉียงเหนือ

ขอบเขตการวิจัย

การศึกษานี้ได้กำหนดขอบเขตของการศึกษา ดังนี้

1. พื้นที่ศึกษา

เป็นพื้นที่รับผิดชอบของศูนย์ปฏิบัติการฝนหลวงภาคตะวันออกเฉียงเหนือ 2 ศูนย์ คือ

1.1 ศูนย์ปฏิบัติการฝนหลวงตอนบน

ตั้งอยู่ที่จังหวัดขอนแก่น รับผิดชอบครอบคลุมพื้นที่จังหวัดขอนแก่น มหาสารคาม กาฬสินธุ์ ร้อยเอ็ด ชัยภูมิ (ตอนบน) สกลนคร หนองคาย มุกดาหาร นครพนม เลย หนองบัวลำภู และอุดรธานี

1.2 ศูนย์ปฏิบัติการฝนหลวงตอนล่าง

ตั้งอยู่ที่จังหวัดนครราชสีมา รับผิดชอบครอบคลุมพื้นที่จังหวัดนครราชสีมา ชัยภูมิ (ตอนล่าง) บุรีรัมย์ สุรินทร์ อุบลราชธานี ยโสธร อำนาจเจริญ และศรีสะเกษ

(กระทรวงเกษตรและสหกรณ์, 2552)

2. ข้อมูลและแหล่งข้อมูล

2.1 สำนักฝนหลวงและการบินเกษตร

ข้อมูลจากฐานข้อมูลศูนย์ปฏิบัติการฝนหลวงเฉลิมพระเกียรติ อำเภอหัวหิน จังหวัดประจวบคีรีขันธ์ เป็นชุดข้อมูลการตรวจอากาศชั้นบนประจำวัน โดยต่อไปจะเรียกชุดข้อมูลนี้ว่าข้อมูล GPCM เริ่มระหว่างวันที่ 1 เมษายน พ.ศ.2547 ถึงวันที่ 14 กันยายน พ.ศ.2549

2.2 กรมอุตุนิยมวิทยา

ข้อมูลปริมาณน้ำฝน 24 ชั่วโมง (รายวัน) จาก 18 สถานีฐาน เริ่มระหว่างวันที่ 1 มกราคม พ.ศ.2545 ถึงวันที่ 30 กันยายน พ.ศ.2549

3. ชุดข้อมูลที่ใช้วิเคราะห์

ประกอบด้วยข้อมูลจากข้อ 2.1 (ข้อมูล GPCM) รวมกับปริมาณน้ำฝนรายวัน จากกรมอุตุนิยมวิทยา โดยคิดปริมาณฝนรายวันด้วยค่าเฉลี่ยของ 18 สถานีฐานในภาคตะวันออกเฉียงเหนือ

4. ตัวแปรที่ใช้ในการวิจัย

4.1 ตัวแปรตาม (Y) คือเหตุการณ์การมีฝน ที่มีค่าเป็น 2 ลักษณะ โดยกำหนดให้

$Y = 1$ ถ้าค่าเฉลี่ยปริมาณฝนรายวันของ 18 สถานีฐาน มีค่ามากกว่า 0 มิลลิเมตร กำหนดให้เป็นฝนตก

$Y = 0$ ถ้าค่าเฉลี่ยปริมาณฝนรายวันของ 18 สถานีฐาน มีค่าเท่ากับ 0 มิลลิเมตร กำหนดให้เป็นฝนไม่ตก

4.2 ตัวแปรอิสระ (X) ประกอบด้วย 3 กลุ่มตัวแปร แบ่งตามลักษณะข้อมูล คือตัวแปรอิสระกลุ่มที่ 1 ได้จากข้อมูล GPCM ที่เป็นข้อมูลเชิงปริมาณ 45 ตัวแปร ตัวแปรอิสระกลุ่มที่ 2 ได้จากข้อมูล GPCM ที่เป็นข้อมูลเชิงคุณภาพ 2 ตัวแปร และกลุ่มตัวแปรอิสระกลุ่มที่ 3 ได้จากการแปลงค่าข้อมูลปริมาณน้ำฝนรายวัน พิจารณาให้เป็นตัวแปรเหตุการณ์การเกิดฝนของวันก่อนหน้า (ฝนตก/ฝนไม่ตก) เป็นข้อมูลเชิงคุณภาพเพิ่มขึ้นอีก 1 ตัวแปร ดังนั้นในการศึกษาครั้งนี้จึงมีจำนวนตัวแปรอิสระทั้งสิ้น 48 ตัวแปร รายละเอียดดังตารางที่ 1

ตารางที่ 1 รายละเอียดของตัวแปรที่ใช้ในการศึกษา

ตัวแปร	คำอธิบาย (หน่วย)	รหัสตัวแปร
1. ตัวแปรตาม		
Rain	เหตุการณ์การเกิดฝนรายวัน มี 2 เหตุการณ์	Y
	- ถ้าวันนี้ฝนไม่ตก	Y = 0
	- ถ้าวันนี้ฝนตก	Y = 1
2. ตัวแปรอิสระ		
2.1 ข้อมูลเชิงปริมาณ 45 ตัวแปร		
2.1.1 ตัวแปรในกลุ่มอุณหภูมิ		
1) RAOB_SURFACE_TEMPERATURE	RAOB_SURFACE_TEMPERATURE (celsius)	X111
2) CONVECTIVE_TEMPERATURE	อุณหภูมิในการเกิดเมฆคิวมูลัส (celsius)	X121
3) CLOUD_BASE_HEIGHT_LCL	ความสูงของฐานเมฆ LCL (feet)	X131

ตารางที่ 1 (ต่อ)

ตัวแปร	คำอธิบาย (หน่วย)	รหัส ตัวแปร
4) CLOUD_BASE_TEMPERATURE_LCL	อุณหภูมิที่ระดับฐานเมฆ LCL (celsius)	X132
5) SFC_TEMP_RISE_REQUIRED_FOR_CCL	SFC_TEMP_RISE_REQUIRED_FOR _CCL (celsius)	X141
6) CLOUD_BASE_PRESSURE_CCL	ความกดอากาศที่ระดับฐานเมฆ CCL (mbar)	X142
7) CLOUD_BASE_HEIGHT_CCL	ความสูงของฐานเมฆ CCL (feet)	X143
8) CLOUD_BASE_TEMPERATURE_CCL	อุณหภูมิที่ระดับฐานเมฆ CCL (celsius)	X144
9) SUB_CLOUD_MIXING_RATIO_CCL	SUB_CLOUD_MIXING_RATIO_CCL(g/k g)	X145
<u>2.1.2 ตัวแปรในกลุ่มความชื้น</u>		
10) PRECIPITABLE_WATER_SFC850	ปริมาณน้ำในบรรยากาศตั้งแต่ผิวพื้นถึง 850 มิลลิบาร์ (cm.)	X211
11) PRECIPITABLE_WATER_SFC700	ปริมาณน้ำในบรรยากาศตั้งแต่ผิวพื้นถึง 700 มิลลิบาร์ (cm.)	X212
12) PRECIPITABLE_WATER_SFC500	ปริมาณน้ำในบรรยากาศตั้งแต่ผิวพื้นถึง 500 มิลลิบาร์ (cm.)	X213
13) PRECIPITABLE_WATER_TOTAL	ปริมาณน้ำในบรรยากาศรวม (cm.)	X214
14) MEAN_MIXING_RATIO_LOWEST _100_MB	อัตราส่วนไอน้ำในอากาศเฉลี่ย จากพื้นผิว ขึ้นไป 100 มิลลิบาร์ (g/kg)	X221
15) MEAN_MIXING_RATIO_LOWEST _50_MB	อัตราส่วนไอน้ำในอากาศเฉลี่ย จากพื้นผิว ขึ้นไป 50 มิลลิบาร์ (g/kg)	X222
16) MEAN_DEW_POINT_TEMP_LOWEST _50_MB	อุณหภูมิจุดน้ำค้างเฉลี่ยจากพื้นผิวขึ้นไป 50 มิลลิบาร์ (celsius)	X231
17) RH_AT_THE_CCL	ความชื้นสัมพัทธ์ที่ระดับฐานเมฆ CCL (%)	X241
18) AVG_RH_0_10000_FT	ความชื้นสัมพัทธ์เฉลี่ยจาก 0-10,000 (%)	X242
19) AVG_RH_10_18000_FT	ความชื้นสัมพัทธ์เฉลี่ยจาก 10-18,000 (%)	X243
20) MEAN_RH_1000_5000_FT	ความชื้นสัมพัทธ์เฉลี่ยจาก 1,001-5,000 ฟุต	X244
21) MEAN_RH_5000_10000_FT	ความชื้นสัมพัทธ์เฉลี่ยจาก 5,001-10,000 ฟุต	X245
22) MEAN_RH_10000_15000_FT	ความชื้นสัมพัทธ์เฉลี่ยจาก 10,001-15,000 ฟุต	X246
23) MEAN_RH_20000_25000_FT	ความชื้นสัมพัทธ์เฉลี่ยจาก 20,001-25,000 ฟุต	X247

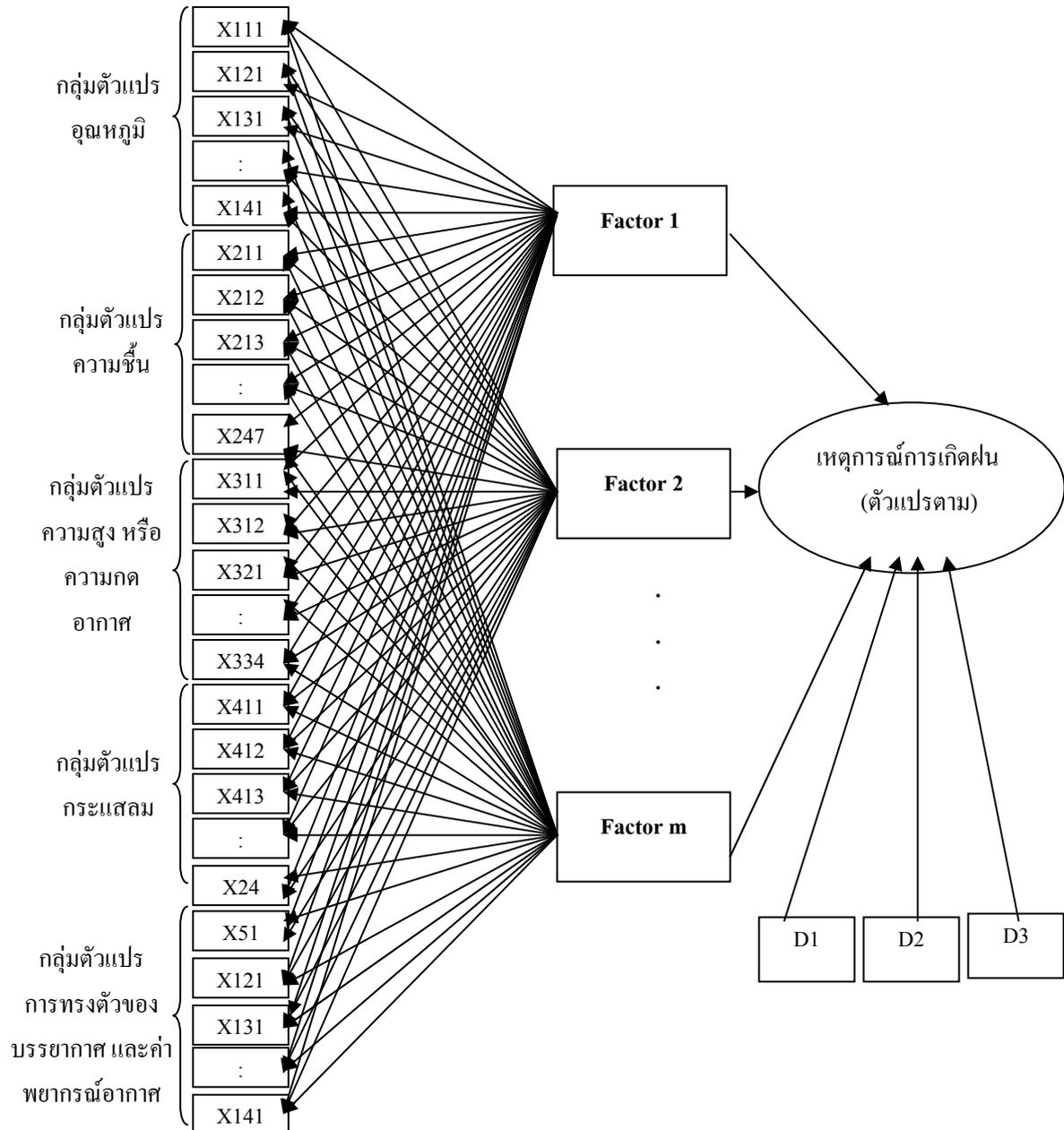
ตารางที่ 1 (ต่อ)

ตัวแปร	คำอธิบาย (หน่วย)	รหัส ตัวแปร
<u>2.1.3 ตัวแปรในกลุ่มความสูงหรือความกดอากาศ</u>		
24) LEVEL_OF_FREE_CONVECTION _LFC_wrtTma	ความสูงของระดับยกตัวอิสระ (mbar)	X311
25) TOP_OF_LATENT_INSTABILITY _LAYER_wrtTma	TOP_OF_LATENT_INSTABILITY _LAYER_wrtTma (mbar)	X312
26) LEVEL_OF_NEUTRAL_BUOYANCY _LNB	ความสูงของระดับสมดุลของแรงลอยตัว (mbar)	X321
27) ISOTHERM_HEIGHTS_0c	ความสูงของระดับอุณหภูมิเท่า 0 c. (feet)	X331
28) ISOTHERM_HEIGHTS_minus5c	ความสูงของระดับอุณหภูมิเท่า -5 c.(feet)	X332
29) ISOTHERM_HEIGHTS_minus10c	ความสูงของระดับอุณหภูมิเท่า -10 c. (feet)	X333
30) ISOTHERM_HEIGHTS_minus15c	ความสูงของระดับอุณหภูมิเท่า -15 c. (feet)	X334
<u>2.1.4 ตัวแปรในกลุ่มกระแสลม</u>		
31) MEAN_WIND_1000_5000_FT_DEG	ทิศทางลมเฉลี่ยจาก 1,000-5,000 ฟุต (deg)	X411
32) MEAN_WIND_5000_10000_FT_DEG	ทิศทางลมเฉลี่ยจาก 5,001-10,000 ฟุต (deg)	X412
33) MEAN_WIND_10000_15000_FT_DEG	ทิศทางลมเฉลี่ยจาก 10,001-15,000 ฟุต (deg)	X413
34) MEAN_WIND_20000_25000_FT_DEG	ทิศทางลมเฉลี่ยจาก 20,001-25,000 ฟุต (deg)	X414
35) MEAN_WIND_1000_5000_FT_KTS	ทิศทางลมเฉลี่ยจาก 1,001-5,000 ฟุต (knots)	X421
36) MEAN_WIND_5000_10000_FT_KTS	ทิศทางลมเฉลี่ยจาก 5,001-10,000 ฟุต (knots)	X422
37) MEAN_WIND_10000_15000_FT_KTS	ทิศทางลมเฉลี่ยจาก 10,001-15,000 ฟุต(knots)	X423
38) MEAN_WIND_20000_25000_FT_KTS	ทิศทางลมเฉลี่ยจาก 20,001-25,000 ฟุต(knots)	X424
<u>2.1.5 ข้อมูลในกลุ่มการทรงตัวของบรรยากาศและค่าพยากรณ์อากาศ</u>		
39) LIFTED_INDEX_100_MBAR_LAYER _ADIABATIC	ดัชนีการยกตัวของมวลอากาศ ที่ลากจาก ค่าเฉลี่ยของอากาศผิวพื้นหนา 100 มิลลิบาร์ (ไม่มีหน่วย)	X51
40) LIFTED_INDEX_50_MBAR_LAYER _MEAN_VALUES	ดัชนีการยกตัวของมวลอากาศ ที่ลากจาก ค่าเฉลี่ยของอากาศผิวพื้นหนา 50 มิลลิบาร์ (ไม่มีหน่วย)	X52
41) SHOWALTER_INDEX	Showalter Index (ไม่มีหน่วย)	X53
42) TOTAL_TOTALS_INDEX	Total totals index (ไม่มีหน่วย)	X54
43) K_INDEX	K Index (ไม่มีหน่วย)	X55

ตารางที่ 1 (ต่อ)

ตัวแปร	คำอธิบาย (หน่วย)	รหัส ตัวแปร
44) SWEAT_INDEX	Sweat index (ไม่มีหน่วย)	X56
45) POTENTIAL_BUOYANCY_INDEX	Potential buoyancy index (ไม่มีหน่วย)	X57
2.2 ข้อมูลเชิงคุณภาพ 3 ตัวแปร		
2.2.1 ตัวแปรในกลุ่มอุณหภูมิ		
46) CONVECTIVE_TEMP_REACHABLE	โอกาสที่อุณหภูมิผิวพื้นจะถึงค่าอุณหภูมิ ในการเกิดเมฆคิวมูลัส - ไม่มีโอกาส ($D_1 = 0$) - มีโอกาส ($D_1 = 1$)	D_1
2.2.2 ตัวแปรพยากรณ์โอกาสความสำเร็จในการทำฝน		
47) WARM_CLOUD_SEEDING_POTENTIAL	โอกาสความสำเร็จในการทำฝนจากเมฆอุ่น - ไม่สำเร็จ ($D_2 = 0$) - สำเร็จ ($D_2 = 1$)	D_2
2.2.3 ตัวแปรการเกิดฝนของวันก่อนหน้า		
48) YESTERDAY_RAIN	เหตุการณ์การเกิดฝนของวันก่อนหน้า - ถ้าฝนไม่ตกวันก่อนหน้า ($D_3 = 0$) - ถ้าฝนตกวันก่อนหน้า ($D_3 = 1$)	D_3

5. กรอบแนวคิดการวิจัย



ภาพที่ 1 แผนภาพการวิเคราะห์ปัจจัยสำหรับตัวแบบคาดการณ์โอกาสการเกิดฝน

การตรวจเอกสาร

การตรวจเอกสารแยกออกเป็น 2 ส่วน คือ

ส่วนที่ 1 ความรู้ทางอุตุนิยมวิทยา และงานวิจัยที่เกี่ยวข้อง

ส่วนที่ 2 ทฤษฎีทางสถิติ

2.1 การตรวจสอบความสัมพันธ์กันเองระหว่างตัวแปรอิสระ(Multicollinearity)

2.2 การวิเคราะห์ปัจจัย (Factor analysis)

2.3 การถดถอยโลจิสติกแบบทวิ (Binary logistic regression)

ส่วนที่ 1 ความรู้ทางอุตุนิยมวิทยา และงานวิจัยที่เกี่ยวข้อง

1.1 เมฆ (Cloud) คือมวลอากาศที่ลอยตัวขึ้นจากผิวดินจนถึงระดับความสูงหนึ่ง จึงเกิดการควบแน่น เนื่องจากสภาวะอากาศในขณะนั้น มีความชื้นสัมพัทธ์สูงกว่า 100 เปอร์เซ็นต์ เป็นผลให้เกิดหยดน้ำ หรือผลึกน้ำแข็ง แขนงลอยอยู่ในมวลอากาศ เมฆจึงแบ่งเป็น 4 ประเภทตามระดับชั้นความสูง คือ 1) เมฆชั้นสูง 2) เมฆชั้นกลาง 3) เมฆชั้นต่ำ และ 4) เมฆที่ก่อตัวตามแนวตั้ง โดยเมฆแต่ละประเภทยังแบ่งออกเป็นชนิดต่าง ๆ มีชื่อเรียกตามลักษณะทางกายภาพที่แตกต่างกัน สำหรับเมฆที่เกี่ยวข้องในการทำฝนหลวงส่วนใหญ่ คือเมฆคิวมูลัส สรุปได้ดังตารางที่ 2 ดังนี้

ตารางที่ 2 ชนิดของเมฆ จำแนกตามลักษณะการเกิดและความสูง

ชนิดเมฆ	ชื่อเมฆ	ความสูง (กิโลเมตร)	ลักษณะ
ชั้นสูง	เซอร์รัส (Cirrus, Ci)	6 - 18	ริ้วขาวบาง
	เซอร์โรสเตรตัส (Cirrostratus, Cs)		แผ่นขาวบาง
	เซอร์โรคิวมูลัส (Cirrocumulus, Cc)		ก้อนเล็กคล้ายระลอกทราย
ชั้นกลาง	อัลโตสเตรตัส (Altostratus, As)	2 - 8	แผ่นทึบและต่ำกว่าเซอร์โรสเตรตัส
	อัลโตคิวมูลัส (Alto cumulus, Ac)		คล้ายฝูงแกะ
ชั้นต่ำ	สเตรโตคิวมูลัส (Stratocumulus, Sc)	ผิวพื้น - 2	อยู่ต่ำ ก้อนกลมเป็นคลื่น
	นิมโบสเตรตัส (Nimbostratus, Ns)		แผ่นหนาฟ้ามืดครึ้ม
ก่อตัวแนวตั้ง	คิวมูลัส (Cumulus, Cu)	ผิวพื้น - 18	เมฆก้อนหนา
	คิวมูโลนิมบัส (Cumulonimbus)		ก้อนหนาทึบยอดแผ่สูง

ที่มา: ความรู้ในการปฏิบัติการฝนหลวง กระทรวงเกษตรและสหกรณ์ (2552)

1.2 ฝน (Rain) คือ หยาดน้ำฟ้าที่อยู่ในรูปของเหลว โดยมีแหล่งกำเนิดจากเมฆอัลโตสเตรตัส นิมโบสเตรตัส สเตรโตคิวมูลัส คิวมูลัส และคิวมูโลนิมบัส (เกษม, 2526) กล่าวว่า การเกิดหยาดน้ำฟ้า (Precipitation) ประกอบด้วยปัจจัยที่จำเป็น 3 ประการ คือ 1) ไอ้ในอากาศ 2) กลวิธีในการควบแน่น (Mechanism of condensation) และ 3) ฝุ่นละอองในบรรยากาศ (Condensation nuclei) ในการเพิ่มขนาดของน้ำในเมฆ นอกจากกระบวนการควบแน่นที่ทำให้หยดน้ำขนาดเล็กเพิ่มขนาดและปริมาณขึ้น จะต้องประกอบด้วยกระบวนการอื่นๆ ร่วมด้วย คือ กระบวนการชนและรวมตัวกัน (Collision-coalescence process) กับกระบวนการเพิ่มขนาดของผลึกน้ำแข็ง (Ice crystal growth process)

1.2.1 กระบวนการชนและรวมตัวกัน

ตันสกูล (2549) กระบวนการชนและรวมตัวกันของหยดน้ำในเมฆเกิดขึ้นในเมฆที่มีการรวมตัวกันของหยดน้ำที่มีอุณหภูมิสูงกว่าจุดเยือกแข็ง เรียกว่า เมฆอุ่น (Warm clouds) สำหรับเมฆที่ความชื้นพอเพียง การชนและรวมตัวจะขึ้นอยู่กับขนาดของหยดน้ำที่เหมาะสม ประจุไฟฟ้าในหยดน้ำรวมทั้งสนามไฟฟ้าในเมฆ ความหนาแน่นของเมฆ และกระแสอากาศไหลขึ้น (Updraft) ในเมฆขนาดของหยดน้ำในเมฆต่างกันทำให้ความเร็วปลายในการตกลงสู่เบื้องล่างด้วยแรงดึงดูดของโลกน้ำแตกต่างกัน หยดน้ำขนาดใหญ่มีความเร็วปลายมากกว่าหยดน้ำขนาดเล็ก หยดน้ำขนาดเส้นผ่านศูนย์กลาง 0.1 มิลลิเมตร มีความเร็วปลายในการตกลงมา 0.27 เมตรต่อวินาที ขณะที่หยดน้ำขนาดเส้นผ่านศูนย์กลาง 5.8 มิลลิเมตร มีความเร็วปลายในการตกลงมา 9.17 เมตรต่อวินาที ดังนั้นหยดน้ำขนาดใหญ่ที่ตกลงมาเร็วกว่าจะเกิดการชนและรวมตัวกับหยดน้ำขนาดเล็กกว่า จนเกิดการเพิ่มขนาดขึ้นตามลำดับ แต่ในบางกรณีหยดน้ำอาจไม่ชนกันก็ได้ แม้ว่าจะตกลงมาในแนวเดียวกัน เนื่องจากหยดน้ำอาจลอยผ่านกันไปเพราะถูกกระแสลมพัดให้ห่างกัน หรือหยดน้ำอาจรวมตัวกันชั่วขณะแล้วกลับแยกออกเป็นขนาดเท่าเดิม หรือหยดน้ำอาจรวมตัวกันชั่วขณะแล้วแตกเป็นหยดน้ำเล็กๆ จำนวนมาก

1.2.2 กระบวนการเพิ่มขนาดของผลึกน้ำแข็ง

กระบวนการเพิ่มขนาดผลึกน้ำแข็ง กระบวนการเพิ่มขนาดผลึกน้ำแข็งเป็นกระบวนการที่สำคัญที่ทำให้เกิดหยาดน้ำฟ้า ในบริเวณเขตอบอุ่นหรือเขตหนาว กระบวนการนี้เกิดขึ้นในเมฆส่วนที่อยู่ในระดับความสูงกว่าระดับเยือกแข็ง จึงเรียกเมฆชนิดนี้ว่า เมฆเย็น (Cold cloud) ในเมฆเย็นมีหยดน้ำที่อุณหภูมิต่ำกว่าศูนย์องศาเซียส เรียกว่าหยดน้ำเย็นยิ่งยวด (Super cooled droplets) อยู่ล้อมรอบผลึกน้ำแข็งจากความแตกต่างของความดันไอ ทำให้เกิดการ

แพร่กระจาย (Diffusion) ของไอน้ำจากหยดน้ำเย็นยิ่งยวดไปเกาะบนผลึกน้ำแข็งจากนั้นเปลี่ยนสถานะเป็นน้ำแข็งพอกอยู่ภายนอกผลึกน้ำแข็ง จนกระทั่งน้ำหนักของผลึกน้ำแข็งมีมากกว่าแรงพยุงเนื่องจากกระแสอากาศพัดขึ้น ผลึกน้ำแข็งจะเริ่มตกลงมาเกิดการชนกับหยดน้ำเย็นยิ่งยวดข้างล่างและเคลือบเพิ่มขนาดขึ้นเรื่อยๆ เรียกกระบวนการนี้ว่า กระบวนการเพิ่มขนาดหรือเพิ่มขอบ (Accretion or riming process) และเรียกหยดน้ำฟ้านี้ว่า ผลึกน้ำแข็งที่เพิ่มขนาด (Graupel) จากนั้นผลึกน้ำแข็งอาจเกิดการชนกันและรวมตัวกันระหว่างผลึกน้ำแข็ง (Aggregation) จนเป็นเกล็ดหิมะหรือตกลงมาที่อุณหภูมิสูงกว่า 0 องศาเซลเซียส และละลายเป็นหยดน้ำฝนตกลงสู่พื้นดินในที่สุด กระบวนการทั้งหมดนี้เรียกว่า กระบวนการเมฆเย็น (Cold cloud process) ซึ่งขนาดนามตามชื่อ นักวิทยาศาสตร์ที่เริ่มตามลำดับ เรียกว่า กระบวนการเวเนเนอร์-เบอร์เกอร์-เฟรดเดอริก (Wehener-Bergeron-Frederick Process)

1.3 เครื่องมือการตรวจวัดฝน

ในปัจจุบันการตรวจวัดฝนนิยมใช้กันในประเทศไทยมี 2 แบบคือ

1.3.1 การตรวจวัดฝนด้วยเครื่องวัดน้ำฝนภาคพื้นดิน

การตรวจวัดฝนด้วยเครื่องวัดน้ำฝนภาคพื้นดินสามารถแบ่งเครื่องมือวัดปริมาณฝนได้เป็น 2 ชนิด คือ เครื่องวัดน้ำฝนแบบไม่บันทึกข้อมูลต่อเนื่อง (Non-recording rain gauge) และเครื่องวัดน้ำฝนแบบบันทึกข้อมูลต่อเนื่อง (Recording rain gauges)

1.3.2 การตรวจวัดฝนด้วยเรดาร์ตรวจอากาศ

มีเป้าหมายที่จะทำการตรวจวัดตามวัตถุประสงค์ของสำนักฝนหลวงและการบินเกษตร คือ กลุ่มฝนหรือน้ำที่อยู่ในอากาศที่เกิดจากการทำฝนเทียม แต่เนื่องจากการตรวจวัดด้วยเรดาร์ไม่ได้เป็นการตรวจวัดปริมาณฝนที่จะตกโดยตรง แต่เป็นการวัดค่าการสะท้อนกลับ (Reflectivity factor, Z) จึงนำค่าดังกล่าวมาสร้างสมการความสัมพันธ์ระหว่างความเข้มฝน (Rain intensity, R) ในรูปแบบของสมการยกกำลังดังนี้

$$Z=aR^b$$

เมื่อ R คือ ความเข้มฝน มีหน่วยเป็น mm/hr

Z คือ แฟกเตอร์การสะท้อนกลับ มีหน่วยเป็น mm^6/m^3

a และ b คือ ค่าคงที่ ขึ้นอยู่กับการกระจายของขนาดเม็ดน้ำ (Drop size distribution) และชนิดของฝน (ต้นสกุล, 2549)

1.4 ตัวแปรที่เกี่ยวข้องกับการพยากรณ์อากาศและโอกาสความสำเร็จในการเกิดฝน

วรารุช (2539) ได้ทำการศึกษาวิเคราะห์ผลพยากรณ์อากาศโดย GPCM ซึ่งได้ใช้วิธีการวัดค่าข้อมูลของตัวแปรต่าง ๆ ในแต่ละกลุ่มที่เกี่ยวข้องกับการศึกษาในครั้งนี้ โดยแบ่งข้อมูลออกเป็น 6 กลุ่ม คือ 1) ตัวแปรกลุ่มอุณหภูมิ 2) ตัวแปรกลุ่มความชื้น 3) ตัวแปรกลุ่มความสูงและความกดอากาศ 4) ตัวแปรกลุ่มกระแสลม 5) ตัวแปรกลุ่มการทรงตัวของบรรยากาศและค่าพยากรณ์อากาศ และ 6) ตัวแปรพยากรณ์โอกาสความสำเร็จในการทำฝน รายละเอียดมีดังต่อไปนี้

1.4.1 ตัวแปรในกลุ่มอุณหภูมิ มี 4 ตัวแปรหลัก คือ

- 1) อุณหภูมิผิวพื้น (Surface temperature, T_{SFC}) หมายถึง ระดับความร้อนบนของอากาศที่ตรวจวัดที่ระดับผิวพื้น
- 2) อุณหภูมิของการเกิดเมฆคิวมูลัส (Convective temperature, T_C) เป็นค่าวิเคราะห์ของอุณหภูมิต่ำสุดที่กล่อ่งอากาศ (Parcel) ที่ผิวพื้นพึงจะมี ก่อนจะมีพลังงานเพียงพอที่จะยกตัวลอยสูงขึ้น ไปจนถึงระดับอิมตัวด้วยไอน้ำ (Convective condensation level, CCL)
- 3) อุณหภูมิที่ระดับฐานเมฆ LCL (LCL temperature, T_{LCL}) เป็นค่าวิเคราะห์ของอุณหภูมิมบรรยากาศที่ระดับ LCL หรือ Lifting condensation level ซึ่งเป็นระดับที่กล่อ่งอากาศ (parcel) จะอิมตัวด้วยไอน้ำ หลังจากถูกยกตัวให้ลอยสูงขึ้นมาจากผิวพื้น โดยปราศจากพลังงานที่เกิดจากความร้อน หรือบางที่เรียกว่า อุณหภูมิของระดับฐานเมฆสเตเรตัส
- 4) อุณหภูมิที่ระดับฐานเมฆ CCL (CCL temperature, T_{CCL}) เป็นค่าวิเคราะห์ของอุณหภูมิมบรรยากาศที่ระดับ CCL หรือระดับ convective condensation level ซึ่งเป็นระดับที่กล่อ่งอากาศ (parcel) จะอิมตัวด้วยไอน้ำหลังจากการยกตัวให้ลอยสูงขึ้นมาจากผิวพื้น เนื่องมาจากพลังงานที่เกิดจากความร้อน หรือบางที่เรียกว่า อุณหภูมิของระดับฐานเมฆคิวมูลัส

1.4.2 ตัวแปรในกลุ่มความชื้น มี 4 ตัวแปรหลัก คือ

- 1) ปริมาณน้ำในบรรยากาศ (Precipitable water, P_w) หรือเรียกว่าปริมาณไอน้ำในบรรยากาศ (Precipitable water vapor) หมายถึง ปริมาณไอน้ำที่บรรจุอยู่ในหนึ่งหน่วยปริมาตรของบรรยากาศรูปทรงกระบอกที่มีฐานล่าง และฐานบนอยู่ระหว่างชั้นบรรยากาศใด ๆ ที่ต้องการทราบค่าปริมาณน้ำในบรรยากาศ
- 2) อัตราส่วนไอน้ำในอากาศ (Mixing ratio, MR) หมายถึง อัตราส่วนโดยน้ำหนักของไอน้ำในอากาศกับอากาศแห้ง
- 3) อุณหภูมิจุดน้ำค้าง (Dew point temperature, T_d) หมายถึง อุณหภูมิที่ซึ่งอากาศจะถูกทำให้เย็นลง (ลดอุณหภูมิลง) ภายใต้สภาวะความกดอากาศ และปริมาณไอน้ำคงที่ เพื่อทำให้ไอน้ำในอากาศเกิดการกลั่นตัว

4) ความชื้นสัมพัทธ์ (Relative humidity, RH) หมายถึง อัตราส่วนของค่าความดันไอน้ำจริง (Actual vapor pressure) ของอากาศ กับค่าความดันไอน้ำอิ่มตัวจริง (Saturation vapor pressure)

1.4.3 ตัวแปรในกลุ่มความสูงหรือความกดอากาศ มี 3 ตัวแปรหลัก คือ

1) ความสูงของระดับยกตัวอิสระ (Level of Free Convective, LFC) เป็นความสูงที่ระดับ LFC ที่วิเคราะห์บนแผน Thermodynamic diagram แสดงถึงระดับที่ก้อนอากาศ (parcel) เมื่อลอยขึ้นมาพ้นระดับแล้วจะมีอุณหภูมิสูงกว่าอุณหภูมิของบรรยากาศในระดับเดียวกัน ทำให้สามารถลอยตัวขึ้นต่อไปได้โดยอิสระ โดยไม่ต้องอาศัยพลังงานจากภายนอก

2) ความสูงระดับสมดุลของแรงลอยตัว (Level of Neutral Buoyancy, LNB) เป็นความสูงที่ระดับ LNB ที่วิเคราะห์บนแผน Thermodynamic diagram แสดงถึงระดับที่ก้อนอากาศ (Parcel) เมื่อลอยขึ้นมาถึงระดับแล้วจะมีอุณหภูมิเท่ากับอุณหภูมิของบรรยากาศในระดับเดียวกัน ทำให้ไม่สามารถลอยตัวขึ้นโดยอิสระต่อไปได้อีก ระดับนี้บางครั้งเรียกว่า ระดับสมดุล (Equilibrium Level, EL) ซึ่งใช้ทำนายความสูงของยอดเมฆในวันนั้น ๆ

3) ความสูงของอุณหภูมิเท่า (Isotherm heights) โดยใช้ค่าเฉลี่ยที่ระดับต่าง ๆ 4 ระดับ คือ 0, -5, -10 และ -15 องศาเซลเซียส ตามลำดับ

1.4.4 ตัวแปรในกลุ่มกระแสลม มี 2 ตัวแปรหลัก คือ

1) ทิศทางลม (Wind direction, WD) เป็นทิศทางของการเคลื่อนที่ของกระแสอากาศโดยวัดทิศทางจากจุดที่ลมพัดเข้าหาจุดสังเกต

2) ความเร็วลม (Wind speed, WS) เป็นความเร็วของการเคลื่อนที่ของกระแสอากาศเทียบกับตำแหน่งบนผิวพื้น โดยใช้ค่าเฉลี่ยที่ความสูงต่าง ๆ 4 ระดับ คือ ระดับ 1,000 – 5,000 ฟุต ระดับ 5,000 – 10,000 ฟุต ระดับ 10,000 – 15,000 ฟุต และระดับ 20,000 – 25,000 ฟุต ตามลำดับ

1.4.5 ข้อมูลในกลุ่มดัชนีการทรงตัวของบรรยากาศและค่าพยากรณ์อากาศ มี 5 ตัวแปรหลัก คือ

1) Lifted Index (LI) เป็นค่าดัชนีที่หาจากผลต่างของอุณหภูมิของก้อนอากาศที่ถูกยกตัวขึ้นมาจากบริเวณใกล้ผิวพื้นภายใต้ขบวนการ Thermodynamic กับอุณหภูมิของบรรยากาศที่ระดับ 500 มิลลิบาร์ ถ้าค่าที่ได้เป็นลบ LI จะแสดงถึงความไม่เสถียรภาพ (Unstable) ของบรรยากาศ

2) Showalter Index (SI) เป็นค่าดัชนีที่หาจากผลต่างของอุณหภูมิของก้อนอากาศที่ถูกยกตัวขึ้นมาจากระดับ 850 มิลลิบาร์ ภายใต้ขบวนการ Thermodynamic กับอุณหภูมิของ

บรรยากาศที่ระดับ 500 มิลลิบาร์ ถ้าค่าที่ได้เป็นลบ SI จะแสดงถึงความไม่เสถียรภาพ (Unstable) ของบรรยากาศ

3) Total Total Index (TTI) เป็นค่าดัชนีที่ใช้ในการประเมินโอกาสการเกิดพายุฝน เช่น ค่าระหว่าง 45-49 จะพยากรณ์โอกาสเกิดพายุฝน และค่าในช่วง 50 ขึ้นไปจะพยากรณ์พายุฝนฟ้าคะนองที่มีความรุนแรง

4) K Index (KI) เป็นค่าดัชนีที่ใช้ในการประเมินโอกาสเกิดพายุฝน เช่นเดียวกัน โดยใช้ค่าอัตราลดอุณหภูมิตามความสูง ปริมาณความชื้นในบรรยากาศชั้นล่าง ๆ และระดับการยกตัวของความชื้นขึ้นเบื้องบน

5) Shweta Index (SWI) เป็นค่าดัชนีเพื่อใช้ในการพยากรณ์พายุฝนฟ้าคะนอง มาจากคำว่า Severe weather threat ค่า SWI คำนวณจากความเสถียรภาพของบรรยากาศ ความแตกต่างของลมแต่ละระดับทั้งด้านความเร็วและทิศทาง และความชื้นที่ระดับมาตรฐานการตรวจอากาศชั้นบน

1.4.6 ตัวแปรในกลุ่มตัวแปรพยากรณ์โอกาสความสำเร็จในการทำฝน

มี 1 ตัวแปรหลัก คือ โอกาสความสำเร็จในการทำฝนจากเมฆอุ่น (Warm Cloud Seeding Potential, WCSP) โดยอาศัยพิสัย (Range) ของข้อมูลอุณหภูมิตามระดับความเอื้ออำนวยต่อการเกิดฝนในธรรมชาติ โดยอาศัยการให้คะแนนแบบถ่วงน้ำหนัก ภายใต้สมมติฐานที่ว่าถ้าโอกาสการเกิดฝนในธรรมชาติสูง โอกาสความสำเร็จในการทำฝนจากเมฆอุ่นจะสูงด้วย โดยแบ่งเป็น 2 กลุ่ม คือ โอกาสทำฝนจากเมฆอุ่นไม่สำเร็จ ($D_2=0$) และ โอกาสทำฝนจากเมฆอุ่นสำเร็จ ($D_2=1$)

การศึกษาครั้งนี้ทำการศึกษาปัจจัยในการเกิดฝนเพิ่ม 1 ตัวแปร คือ

1.4.7 ตัวแปรในกลุ่มผลตรวจปริมาณฝนด้วยเครื่องมือวัดฝน

คือตัวแปรปริมาณน้ำฝน 24 ชั่วโมง หรือปริมาณน้ำฝนรายวัน หมายถึง ปริมาณรวมของความลึกฝนที่วัดด้วยเครื่องมือวัดฝนที่ผิวดินภายใน 24 ชั่วโมง ที่เวลาเริ่มต้นการวัด 7:00 น. โดยใช้ข้อมูลปริมาณน้ำฝน 24 ชั่วโมง สร้างเป็นตัวแปรใหม่ 2 ตัวแปร คือ

1) ตัวแปรการเกิดฝนวันนี้ โดยแบ่งเป็น 2 กลุ่ม คือ วันนี้ฝนตก ($Y=1$) และ วันนี้ฝนไม่ตก ($Y=0$)

2) ตัวแปรการเกิดฝนวันก่อนหน้า โดยแบ่งเป็น 2 กลุ่ม คือ ถ้าวันก่อนหน้า ฝนตก ($D_3=1$) หรือ ถ้าวันก่อนหน้า ฝนไม่ตก ($D_3=0$)

สรุปรายละเอียดตัวแปรในแต่ละกลุ่ม ดังตารางที่ 3 (วารุช,2539)

ตารางที่ 3 ตัวแปรที่เกี่ยวข้องกับการพยากรณ์อากาศ และ โอกาสความสำเร็จในการเกิดฝน

ตัวแปร	หน่วยวัด	วิธีการได้มาซึ่งข้อมูล
1. ตัวแปรกลุ่มอุณหภูมิ		
1.1 อุณหภูมิผิวพื้น (T_{SFC})	องศาเซลเซียส	อ่านจากเครื่องวัดเทอร์มอมิเตอร์
1.2 อุณหภูมิของการเกิดเมฆคิวมูลัส (T_C)	องศาเซลเซียส	การวิเคราะห์ผลตรวจอากาศชั้นบน
1.3 อุณหภูมิที่รับฐานเมฆ LCL (T_{LCL})	องศาเซลเซียส	การวิเคราะห์ผลตรวจอากาศชั้นบน
1.4 อุณหภูมิที่รับฐานเมฆ CCL (T_{CCL})	องศาเซลเซียส	การวิเคราะห์ผลตรวจอากาศชั้นบน
2. ตัวแปรในกลุ่มความชื้น		
2.1 ปริมาณน้ำในบรรยากาศ (P_w)	เซนติเมตร (นิ้ว)	การอินทีเกรตค่าของ Mixing Ratio ตลอดชั้นบรรยากาศที่สนใจ
2.2 อัตราส่วนไอน้ำในอากาศ (MR)	กรัม/กิโลกรัม	จากการวิเคราะห์Thermodynamic diagram
2.3 อุณหภูมิจุดน้ำค้าง (T_d)	องศาเซลเซียส	วัดด้วยเครื่อง Dewpoint hygrometer
2.4 ความชื้นสัมพัทธ์ (RH)	เปอร์เซ็นต์	คำนวณจากอุณหภูมิจุดน้ำค้าง และ อุณหภูมิ
3. ตัวแปรในกลุ่มความสูงหรือความกดอากาศ		
3.1 ความสูงของระดับฐานเมฆ (LCL)	มิลลิบาร์	อ่านค่าบนแผ่นThermodynamic diagram
3.2 ความสูงของระดับฐานเมฆ (CCL)	มิลลิบาร์	อ่านค่าบนแผ่นThermodynamic diagram
3.3 ความสูงของระดับยกตัวอิสระ (LFC)	มิลลิบาร์	อ่านค่าบนแผ่นThermodynamic diagram
3.4 ความสูงของระดับสมดุลลอยตัว (LNB)	มิลลิบาร์	อ่านค่าบนแผ่นThermodynamic diagram
3.5 ความสูงของระดับเยือกแข็ง (FL)	มิลลิบาร์	อ่านค่าบนแผ่น Thermodynamic diagram
4. ตัวแปรในกลุ่มกระแสลม		
4.1 ทิศทางลม (WD)	องศา	คำนวณตำแหน่งการเคลื่อนที่ของบอลูนตรวจอากาศชั้นบน
4.2 ความเร็วลม (WS)	เมตร/นาที่ หรือ knots	การคำนวณความเร็วของบอลูนตรวจอากาศชั้นบน
5. ข้อมูลในกลุ่มการทรงตัวของบรรยากาศ และค่าพยากรณ์อากาศ		
5.1 Lified Index (LI)	ไม่มีหน่วย	อ่านค่าบนแผ่นThermodynamic diagram
5.2 Showalter Index (SI)	ไม่มีหน่วย	อ่านค่าบนแผ่นThermodynamic diagram
5.3 Total Total Index (TTI)	ไม่มีหน่วย	คำนวณจากสูตร (วราวุธ, 2539)
5.4 K Index (KI)	ไม่มีหน่วย	คำนวณจากสูตร (วราวุธ, 2539)
5.5 Sweat Index (SWI)	ไม่มีหน่วย	คำนวณจากสูตร (วราวุธ, 2539)
6. ตัวแปรพยากรณ์โอกาสความสำเร็จในการทำฝน		
6.1 Warm Cloud Seeding Potential (WCSP)	ไม่มีหน่วย	ผลการตรวจอากาศชั้นบน
7. ตัวแปรปริมาณน้ำฝนรายวัน	มิลลิเมตร	ด้วยเครื่องมือวัดฝน

งานวิจัยที่เกี่ยวข้อง

วรารุช (2539) ศึกษาและพัฒนาเรื่องความผันแปรและความสัมพันธ์ของข้อมูลตัวแปรทางอุตุนิยมวิทยา ที่เกี่ยวข้องกับการเกิดเมฆและฝน และการพยากรณ์โอกาสความสำเร็จในการทำฝน โดยเปรียบเทียบผลของตัวแบบ GPCM กับปรากฏการณ์ที่เกิดขึ้นจริงตามธรรมชาติ เพื่อพัฒนาปรับแก้ผลพยากรณ์ของตัวแบบ GPCM ให้มีความแม่นยำมากขึ้น ก่อนที่จะนำไปใช้ประกอบการวางแผนตัดสินใจในการทำฝนหลวงของแต่ละวัน ใช้ในการทำนายปรากฏการณ์ที่เกี่ยวข้องกับการเกิดเมฆและฝน ผลจากการศึกษาครั้งนี้ได้แสดงให้เห็นถึงศักยภาพของความสัมพันธ์ของตัวแปรทางอุตุนิยมวิทยาและโครงสร้างทางไดนามิกส์ของบรรยากาศ แต่ยังพบปัญหาของค่าดัชนีพยากรณ์อากาศ ซึ่งส่วนใหญ่ยังไม่มีความน่าเชื่อถือ โดยได้แสดงความคิดเห็นในตอนท้ายว่าควรพัฒนาหลักการพยากรณ์โอกาสการเกิดฝน ในระดับกว้าง (Synoptic Patterns) และให้ครอบคลุมมากยิ่งขึ้น สำหรับการศึกษารั้งต่อไป

คุณฉวี (2543) องค์กรประกอบทางอุตุนิยมวิทยา อยู่ในรูปแบบสมการอนุพันธ์หลายตัวแปรแบบไม่เป็นเชิงเส้น (Non - linear partial differential equation) คือ ไม่สามารถหาคำตอบของสมการที่แท้จริงได้ ต้องใช้การหาคำตอบแบบการประมาณเพื่อหาค่าความถูกต้องให้ใกล้เคียงมากที่สุด การพยากรณ์อากาศส่วนใหญ่มักใช้แบบจำลองทางสถิติ แต่ในช่วงหลังได้มีการนำโครงข่ายประสาทเทียมมาใช้แทนแบบจำลองทางสถิติสำหรับการสร้างรูปแบบความสัมพันธ์ที่ไม่เป็นเชิงเส้น ปรากฏว่าวิธีโครงข่ายประสาทเทียมให้ผลความผิดพลาดที่ร้อยละ 30 ขณะที่แบบจำลองทางสถิติให้ผลความผิดพลาดที่ร้อยละ 48 หลังจากช่วงต้นทศวรรษที่ 90 (พ.ศ.2533) ได้มีการนำโครงข่ายประสาทเทียมมาประยุกต์ใช้ในการพยากรณ์อากาศอีกหลายลักษณะ โดยได้มีผู้ศึกษาวิธีโครงข่ายประสาทเทียมวิเคราะห์อัตราการตกของฝน (มม.) ให้ค่าความถูกต้องที่ร้อยละ 56.2 ในขณะที่แบบจำลองทางสถิติให้ความถูกต้องร้อยละ 46.2 และเมื่อมีการปล่อยบอลูน/เครื่องบินตรวจอากาศ เก็บค่าตัวแปรอุตุนิยมวิทยา เช่น อุณหภูมิในแต่ละระดับความสูง ผลการศึกษาพบว่าการใช้โครงข่ายประสาทเทียมในการพยากรณ์ให้ความแม่นยำมากกว่าแบบจำลองทางสถิติ ร้อยละ 20

ปกรณ (2545) เสนอระบบการพยากรณ์ฝนในเขตกรุงเทพฯ ครอบคลุมระยะเวลา 24 ชั่วโมง ทำการพยากรณ์การเกิดฝน (ฝนตก/ไม่ตก) เฉพาะเขตพื้นที่กรุงเทพฯ เนื่องจากมีลักษณะภูมิอากาศเฉพาะตัวจึงเลือกใช้ความเสถียรภาพของบรรยากาศและข้อมูลฤดูกาลเป็นข้อมูลหลักในการพยากรณ์ โดยใช้โครงข่ายประสาทเทียมชนิดเพอเซปตรอนหลายชั้น (Multilayer Perceptron) ควบคู่กับการเรียนรู้แบบกระจายกลับ พบว่าการเรียนรู้โดยใช้ข้อมูลของกรมอุตุนิยมวิทยาช่วงปี พ.ศ.2537, พ.ศ.2539 และพ.ศ.2542 ระบบโครงข่ายให้ค่าความถูกต้องของการพยากรณ์สูงถึงร้อยละ 83.74 เมื่อเทียบกับสภาพอากาศปี พ.ศ.2543 โดยเขาได้เสนอว่าหากมีข้อมูลสำคัญอื่นร่วมด้วย เช่น ความสูงของฐานเมฆอาจจะให้ผลลัพธ์ที่แม่นยำกว่านี้ ผลของการศึกษาค้นคว้าครั้งนี้ใช้ในการวางแผนการระบายน้ำและการจราจรในเขตกรุงเทพฯ

สมยศ (2547) ประยุกต์การใช้โครงข่ายประสาทเทียมที่ใช้ข้อมูลระดับน้ำแม่น้ำปิง อ.เมือง จ.เชียงใหม่ เป็นรายชั่วโมง ทดสอบค่าประสิทธิภาพการทำนายเทียบกับการใช้ข้อมูลระดับน้ำราย 3 ชั่วโมง จากการศึกษาพบจำนวนโหนดในชั้นรับข้อมูล จำนวนโหนดในชั้นแฝง จำนวนโหนดในชั้นแสดงผล พบว่าโครงสร้างที่ให้ค่าประสิทธิภาพดีที่สุด ได้แก่ 10-6-1 ซึ่งให้ค่าประสิทธิภาพการทำนายถูกต้องร้อยละ 88.60 และเมื่อเปรียบเทียบทฤษฎีที่ใช้ในขั้นตอนคำนวณย้อนกลับพบว่าทฤษฎี Levenberg Marquardt ให้ค่าประสิทธิภาพการทำนายสูงกว่าและใช้เวลาในการคำนวณน้อยกว่าการใช้ทฤษฎี Gradient descent

ลีลี (2550) ศึกษาความสัมพันธ์ระหว่างข้อมูลตรวจอากาศที่เกี่ยวข้องกับการเกิดฝนในภาคตะวันออกเฉียงเหนือ เพื่อจัดทำตัวแบบจำลองอย่างง่ายในการพยากรณ์ และโอกาสความสำเร็จในการปฏิบัติการฝนหลวงในภาคตะวันออกเฉียงเหนือ โดยใช้ข้อมูลจาก 3 แหล่ง คือ 1) ข้อมูลผลการตรวจวัดอากาศชั้นบน กับข้อมูลการขึ้นปฏิบัติการฝนหลวง (GPCM) 2) ข้อมูลปริมาณน้ำฝนรายวันของกรมอุตุนิยมวิทยา และ 3) ข้อมูลเรดาร์ (RADAR) ตรวจวัดและติดตามกลุ่มฝน ของการศึกษานี้ใช้วิธีทางสถิติควบคู่กับวิธีการการเรียนรู้ทางคอมพิวเตอร์ พัฒนาแบบจำลองการพยากรณ์เป็น 3 กรณี คือ 1) พยากรณ์ผลลัพธ์การเกิดฝน (ฝนตก / ไม่ตก) 2) พยากรณ์ปริมาณน้ำฝน (มม.) และ 3) พยากรณ์ปริมาณน้ำฝน 3 ระดับ คือ ไม่มีฝนตก (0 ถึง 0.1 มม.) ฝนน้อย (มากกว่า 0.1 ถึง 10 มม.) และ ฝนปานกลาง (10 มม. ขึ้นไป) โดยมีการคัดเลือกตัวแปรที่มีความสัมพันธ์ต่อการเกิดฝน ผลการศึกษาพบว่า แบบจำลองที่ 1 ใช้วิธีการการเรียนรู้ต้นไม้ตัดสินใจ (C4.5) พยากรณ์การเกิดฝน (ฝนตก/ไม่ตก) ให้ค่าความถูกต้องของการจำแนกการเกิดฝนได้ ร้อยละ 87.06 (จาก 13 ตัวแปร) สำหรับข้อมูล GPCM และร้อยละ 94.41 (จาก 13 ตัวแปร) จากข้อมูล GPCM + RADAR

แบบจำลองที่ 2 ใช้วิธีโครงข่ายประสาทเทียม พยากรณ์ปริมาณน้ำฝน (มม.) ที่ตกภายในวันเดียวกัน ให้ค่าความผิดพลาดของการพยากรณ์แสดงด้วยค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) เท่ากับ 0.155 (จาก 7 ตัวแปร) และ 0.171 (จาก 5 ตัวแปร) สำหรับข้อมูล GPCM และข้อมูล GPCM + RADAR ตามลำดับ ซึ่งเป็นค่าที่ต่ำกว่าค่า RMSE ที่ได้จากการพยากรณ์ปริมาณน้ำฝนที่ตกวันถัดไปและสองวันถัดไป แบบจำลองที่ 3 ใช้วิธีโครงข่ายประสาทเทียม พยากรณ์ปริมาณน้ำฝน 3 ระดับ (ไม่มีฝน/ฝนตกน้อย/และฝนตกปานกลาง) ได้ค่าความถูกต้องของการจำแนกปริมาณน้ำฝน 3 ระดับ ที่ตกภายในวันเดียวกันมีค่าร้อยละ 85.44 (จาก 13 ตัวแปร) และร้อยละ 68.16 (จาก 9 ตัวแปร) สำหรับข้อมูล GPCM และข้อมูล GPCM + RADAR ตามลำดับ ซึ่งเป็นค่าที่สูงกว่าค่าความถูกต้องที่ได้จากการพยากรณ์ระดับปริมาณน้ำฝนที่ตกในวันถัดไปและสองวันถัดไป

จากการทบทวนวรรณกรรม จะเห็นว่าได้มีการใช้เทคนิคโครงข่ายประสาทเทียมในการพัฒนาระบบการพยากรณ์อากาศและโอกาสความสำเร็จในการปฏิบัติการทำฝนหลวงไปบ้างแล้วในบางพื้นที่ จึงทำให้ผู้วิจัยมีแนวคิดที่จะใช้วิธีการและตัวแบบทางสถิติ กับข้อมูลที่ได้จาก ลีลี (2550) เพื่อสร้างตัวแบบคาดการณ์โอกาสการเกิดฝน ของภาคตะวันออกเฉียงเหนือ โดยใช้เฉพาะข้อมูลตรวจอากาศชั้นบน กับข้อมูลปริมาณน้ำฝน 24 ชั่วโมง โดยไม่มีข้อมูลเรดาร์ และเพิ่มตัวแปรเหตุการณ์เกิดฝนของวันก่อนหน้า เป็นตัวแปรอิสระอีกหนึ่งตัว

ดังนั้นผู้วิจัยจึงคาดการณ์ว่าตัวแปรที่ส่งผลต่อการพยากรณ์การเกิดฝน (ฝนตก/ฝนไม่ตก) น่าจะมีความสอดคล้องกับผลการศึกษาของลีลี (2550) ที่พบว่าตัวแปรที่ผ่านการคัดเลือกจากชุดข้อมูล GPCM มี 13 ตัวแปร ดังนี้

- | | |
|---|---------------------------------|
| 1. Precipitable_water_SPC500_mb | 8. RH_AT_THE_CCL |
| 2. Precipitable_water_Total | 9. AVG_RH_10_18000_FT |
| 3. Mean_mixing_ratio_lowest_50_mb | 10. Mean_wind_1000_5000_FT_DEG |
| 4. Roab_surface_temperature | 11. Mean_wind_5000_10000_FT_DEG |
| 5. Cloud_base_height_CCL | 12. Mean_RH_1000_15000_FT |
| 6. Lifted_index_50_mb_layer_mean_values | 13. Mean_RH_2000_25000_FT |
| 7. K_index | |

การศึกษานี้จึงใช้ตัวแปรที่ผ่านการคัดเลือกดังกล่าวเป็นแนวทาง การสร้างตัวแบบทางสถิติ ตลอดจนวิเคราะห์เปรียบเทียบกับตัวแปรที่ได้จากการคัดเลือกด้วยวิธีการทางสถิติของการศึกษานี้

ส่วนที่ 2 ทฤษฎีทางสถิติ

2.1 การตรวจสอบปัญหาตัวแปรอิสระมีความสัมพันธ์กันเอง (Multicollinearity)

เทคนิคการวิเคราะห์ทางสถิติที่เกี่ยวข้องกับตัวแปรอิสระหลายตัว โดยเฉพาะ Multiple linear regression, Non-linear regression หรือ Logistic regression ที่มีความสัมพันธ์ระหว่าง Y กับ X_1, X_2, \dots, X_p จะมีข้อกำหนดว่า ตัวแปรอิสระจะต้องเป็นอิสระต่อกัน ดังนั้น ก่อนการวิเคราะห์สิ่งแรกจึงควรทำการตรวจสอบความสัมพันธ์กันเองระหว่างตัวแปรอิสระ ด้วยวิธีการง่าย ๆ 2 วิธี คือ

วิธีที่ 1) ตรวจสอบจากค่าสัมประสิทธิ์สหสัมพันธ์ (Pearson correlation coefficient, r)

สามารถทำได้โดยพิจารณาจากเมตริกซ์ของค่า r ควบคู่กับแผนภาพการกระจายของ ตัวแปร 2 ตัว ที่ละคู่

$$r_{X_i, Y} = \frac{n \sum X_i Y - \sum X_i \sum Y}{\sqrt{[n \sum X_i^2 - (\sum X_i)^2]} \times \sqrt{[n \sum Y^2 - (\sum Y)^2]}} ; i = 1, 2, \dots, p$$

เป็นการศึกษาความสัมพันธ์เชิงเส้นระหว่างตัวแปรเชิงปริมาณ 2 ตัวขึ้นไป ด้วยค่า r เป็นตัวบอกราคของความสัมพันธ์เชิงเส้นระหว่างตัวแปรที่ละคู่ โดยที่ $-1 \leq r \leq 1$ ถ้า r มีค่าเข้าใกล้ -1 หรือ 1 แสดงว่าตัวแปรคู่่นั้นมีความสัมพันธ์กันแบบเชิงเส้นในระดับสูง แต่ถ้ามีค่าเข้าใกล้ 0 แสดงถึงการมีความสัมพันธ์กันแบบเชิงเส้นในระดับน้อย หรือไม่มีเลย และเครื่องหมาย $+$ หรือ $-$ หน้า r บอกราคทิศทางของความสัมพันธ์ คือ

ถ้า r มีเครื่องหมาย $+$ แสดงว่าตัวแปรหนึ่งมีความสัมพันธ์แบบเชิงเส้นกับอีกตัวแปรหนึ่งไปในทิศทางบวก คือเมื่อตัวแปรหนึ่งมีค่าเพิ่มขึ้นอีกตัวแปรหนึ่งก็มีค่าเพิ่มขึ้นด้วย

ถ้า r มีเครื่องหมาย $-$ แสดงว่าตัวแปรหนึ่งมีความสัมพันธ์แบบเชิงเส้นกับอีกตัวแปรหนึ่งไปในทิศทางลบ คือเมื่อตัวแปรหนึ่งมีค่าเพิ่มขึ้นอีกตัวแปรหนึ่งจะมีค่าลดลง

(Tabachnick and Fidell, 2007)

โดยทดสอบสมมติฐานเกี่ยวกับค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรเชิงปริมาณ ได้ดังนี้

สมมติฐานของการทดสอบ

$H_0 : \rho = 0$ (ตัวแปร X และ Y ไม่มีความสัมพันธ์แบบเชิงเส้น)

$H_1 : \rho \neq 0$ (ตัวแปร X และ Y มีความสัมพันธ์กันแบบเชิงเส้น)

สถิติทดสอบ ด้วยการทดสอบแบบสองทาง

$$t = \frac{r_{XY} \sqrt{n-2}}{\sqrt{1-r_{XY}^2}} \quad ; df = n-2$$

ขอบเขตการปฏิเสธสมมติฐาน H_0

จะปฏิเสธ H_0 เมื่อค่า t ที่คำนวณได้มีค่ามากกว่าหรือเท่ากับค่า $t_{\alpha, n-2}$ ที่เปิดจากตาราง หรือ t ที่คำนวณได้มีค่าน้อยกว่าหรือเท่ากับค่า $-t_{\alpha, n-2}$

(Christensen and Stoup, 1991)

วิธีที่ 2) ตรวจสอบโดยพิจารณาค่า Variance Inflation Factor (VIF)

เป็นการตรวจสอบปัญหา Multicollinearity คือ ตัวแปรอิสระ X_1, X_2, \dots, X_p ที่เป็นข้อมูลเชิงปริมาณมีความสัมพันธ์กันเองหรือไม่ โดยพิจารณาจากตัวแปรอิสระที่นำมาศึกษา มีผลทำให้ความแปรปรวนของค่าประมาณสัมประสิทธิ์การถดถอย (β 's) มีค่าเพิ่มมากขึ้นหรือไม่เมื่อเทียบกับตัวแปรอิสระเหล่านี้ไม่มีความสัมพันธ์กันในเชิงเส้น โดยคำนวณ VIF จากสูตร

$$(VIF)_k = \frac{1}{1 - R_k^2} \quad ; k = 1, 2, \dots, p-1$$

$(VIF)_k$ เป็น Variance Inflation Factor สำหรับ $\hat{\beta}_k$

$\hat{\beta}_k$ คือ สัมประสิทธิ์ถดถอยของตัวแปรอิสระ X_k

โดย R_k^2 เป็นค่าสัมประสิทธิ์แห่งการกำหนด (Multiple coefficient of determination) เมื่อทำการวิเคราะห์การถดถอยระหว่างตัวแปรอิสระ X_k กับตัวแปรอิสระอื่นที่เหลือ $p-1$ ตัว มีค่าอยู่ในช่วง 0 ถึง 1 (Chatterjee, 1991)

เช่น R_1^2 คือ ค่าสัมประสิทธิ์การตัดสินใจเชิงพหุระหว่าง X_1 กับ X_2, X_3, \dots, X_p ดังนั้นค่า VIF จึงมีค่าที่เป็นไปได้ในช่วง 1 ถึง ∞ โดยค่าที่เข้าใกล้ ∞ หมายถึง ตัวแปรอิสระนั้นจะมีความสัมพันธ์กับตัวแปรอิสระตัวอื่น ๆ ในสมการถดถอยเชิงเส้นแบบพหุในระดับสูง และถ้าค่า $VIF \geq 10$ แสดงว่าตัวแปรอิสระที่ทดสอบนั้นเกิดปัญหา Multicollinearity กับตัวแปรอิสระตัวอื่นที่เหลือ (Chatterjee, 1991; Kutner *et al.*, 2005) และจากการที่ตัวแปรอิสระตัวหนึ่งมีค่า VIF สูง ส่งผลให้ค่า Standard Error (SE) ของสัมประสิทธิ์การถดถอย b ของตัวแปรอิสระตัวนั้นมีค่าสูงด้วย ดังนั้นเมื่อนำ $SE(b)$ มาคำนวณ 95% CI ของ β และ 95% CI ของ OR จะได้ช่วงของการประมาณค่ามีค่ากว้างมากกว่าปกติ ตัวแปรนั้นจึงไม่เหมาะที่จะนำไปใช้ในการสร้างตัวแบบ ส่งผลให้ค่าความถูกต้องในการพยากรณ์ลดลง

ดังนั้นวิธีการจัดการปัญหา Multicollinearity อาจทำได้โดยการตัดตัวแปรใดตัวแปรหนึ่งที่มีความสัมพันธ์กันเองออก เพื่อให้ตัวแปรอิสระที่เหลือเป็นตัวแปรที่จะนำไปวิเคราะห์สร้างตัวแบบต่อได้ หรืออาจแก้ปัญหาด้วยการใช้เทคนิคการวิเคราะห์ปัจจัย ซึ่งเป็นวิธีการจัดการข้อมูลที่สามารถอธิบายความซ้ำซ้อนของข้อมูลตัวแปรอิสระ X_1, X_2, \dots, X_p ที่มีความสัมพันธ์กันเอง โดยไม่ต้องตัดตัวแปรอิสระที่ทำให้เกิดปัญหา Multicollinearity จึงเป็นวิธีการรักษาตัวแปรอิสระไว้ โดยจัดเป็นกลุ่มของตัวแปรอิสระขึ้นมาใหม่แล้วเรียกแทนกลุ่มตัวแปรนี้ว่าปัจจัย (Factor)

2.2 การวิเคราะห์ปัจจัย (Factor Analysis : FA)

การวิเคราะห์ปัจจัยเป็นเทคนิคหนึ่งของการวิเคราะห์หลายตัวแปร (Multivariate Analysis) ซึ่งมีเป้าหมายที่จะอธิบายความซ้ำซ้อนของข้อมูลตัวแปรอิสระ X_1, X_2, \dots, X_p ที่มีความสัมพันธ์กันเอง (Multicollinearity) ด้วยการจัดกลุ่มตัวแปรที่มีความสัมพันธ์ในระดับเดียวกัน หรือใกล้เคียงกันให้อยู่ในกลุ่มเดียวกัน จึงทำให้จำนวนตัวแปรลดลงแล้วเรียกตัวแปรใหม่นี้ว่าปัจจัยร่วม F_1, F_2, \dots, F_m (Common factor) การจัดกลุ่มตัวแปรลักษณะเช่นนี้ ทำให้มีจำนวนปัจจัยร่วม (m) น้อยกว่าจำนวนตัวแปรเดิม (p) (Rencher, 1995)

การวิเคราะห์ปัจจัย เป็นการแสดงโครงสร้างของตัวแปรเดิม ($X_i ; i = 1, 2, \dots, p$) โดยแสดงในรูปแบบสมการผลรวมเชิงเส้น (Linear combination) ของปัจจัยร่วมโดยนำค่าความสัมพันธ์ (Correlation) หรือความแปรปรวนร่วม (Covariance) ระหว่างตัวแปรมาพิจารณา การวิเคราะห์ปัจจัยมี 2 แบบ คือการวิเคราะห์ปัจจัยเชิงสำรวจ (Exploratory Factor Analysis, EFA) และการวิเคราะห์ปัจจัยเชิงยืนยัน (Confirmatory Factor Analysis, CFA) โดยที่ EFA พยายามอธิบายและสรุปข้อมูลด้วยการจัดกลุ่มตัวแปรที่มีความสัมพันธ์กันและมักจะทำในขั้นตอนเริ่มแรกของการทำวิจัย เพื่อใช้เป็นเครื่องมือในการสร้างตัวแปรแฝง (Latent variable) ที่ไม่ถูกสังเกต (Unobserved variable) และการทำ EFA ยังเหมาะสำหรับการตรวจสอบความตรงเชิงโครงสร้าง (Construct validity) ของข้อมูล ว่าเป็นไปตามนิยามทางทฤษฎีหรือไม่ และสอดคล้องกับสภาพที่เป็นจริงอย่างไร ส่วน CFA เป็นเทคนิคการวิเคราะห์ที่ใช้ทดสอบยืนยันโครงสร้างของปัจจัยร่วมที่แฝงอยู่ โดย CFA มักจะถูกวิเคราะห์ด้วยการสร้างตัวแบบสมการโครงสร้าง (Structural Equation Modeling, SEM) (Garson, 2009)

ในการศึกษานี้จะทำการวิเคราะห์ปัจจัยด้วยวิธี EFA เพื่อหาจำนวนปัจจัยร่วมที่เหมาะสม โดยแบ่งออกเป็น 5 ขั้นตอน คือ 1) นิยามข้อตกลงเบื้องต้นของการวิเคราะห์ปัจจัย 2) การประมาณค่าน้ำหนักปัจจัยและค่าความร่วมกัน 3) การหาจำนวนปัจจัยที่เหมาะสม 4) การหมุนแกน และ 5) การคำนวณค่าคะแนนปัจจัย

1. นิยามและข้อตกลงของตัวแบบการวิเคราะห์ปัจจัย (Factor Analysis Model Definition and Assumption)

ในการวิเคราะห์ปัจจัยจะสมมติให้ตัวแปร X เป็นเวกเตอร์ของตัวแปรอิสระที่วัดค่าได้ (X_1, X_2, \dots, X_p) มีทั้งหมด p ตัวแปร โดยสุ่มจากประชากรที่มีความคล้ายคลึงกัน (Homogenous

population) คือ $\underline{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}$ มีค่าเฉลี่ย $\underline{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix}$ และมี Σ เป็นเมตริกซ์ความแปรปรวนร่วมที่

แสดงความสัมพันธ์ระหว่างตัวแปร X_1, X_2, \dots, X_p ซึ่งสามารถเขียนอธิบายตัวแปร X_1, X_2, \dots, X_p ด้วยตัวแบบปัจจัย (Factor model) ในรูปแบบสมการผลรวมเชิงเส้น (Linear combination) ของปัจจัยร่วม (F_1, F_2, \dots, F_m) กับค่าความคลาดเคลื่อนที่เป็นค่าเฉพาะ ($\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p$) ของตัวแปร X_1, X_2, \dots, X_p (Johnson and Wichern, 2002)

1.1 ตัวแบบปัจจัยเชิงตั้งฉาก (Orthogonal Factor Model)

ตัวแบบปัจจัยเชิงตั้งฉาก เป็นตัวแบบที่ปัจจัยทุกตัว ไม่มีความสัมพันธ์กัน หรือเป็นอิสระกัน ตัวแบบ FA เขียนได้ดังนี้

$$\begin{aligned} X_1 - \mu_1 &= 1_{11}F_1 + 1_{12}F_2 + \dots + 1_{1m}F_m + \varepsilon_1 \\ X_2 - \mu_2 &= 1_{21}F_1 + 1_{22}F_2 + \dots + 1_{2m}F_m + \varepsilon_2 \\ &\vdots \\ X_p - \mu_p &= 1_{p1}F_1 + 1_{p2}F_2 + \dots + 1_{pm}F_m + \varepsilon_p \end{aligned} \quad \dots\dots\dots(1)$$

โดยที่ X_i เป็นตัวแปรอิสระที่สังเกตและวัดค่าได้ (Observed variable) หรือเรียกว่าตัวแปรเดิม

$$; i = 1, 2, \dots, p$$

μ_i เป็นค่าเฉลี่ยของตัวแปร X_i ; $i = 1, 2, \dots, p$

1_{ij} เป็นค่าน้ำหนัก (Loading) ที่แสดงความสัมพันธ์ระหว่างตัวแปรเดิม (X_i) กับปัจจัยร่วม (F_j) ซึ่ง 1_{ij} จะเป็นตัวที่บอกว่าตัวแปรเดิม (X_i) แต่ละตัวมีความสัมพันธ์กับปัจจัยร่วม (F_j) มากน้อยเพียงใด โดยที่ $-1 \leq 1_{ij} \leq 1$; $i = 1, 2, \dots, p$ และ $j = 1, 2, \dots, m$; $m < p$

F_j คือ ปัจจัยร่วม (Common factors) ที่ j หรือเรียกว่าตัวแปรใหม่ ; $j=1,2,\dots,m$
 ε_i เป็นปัจจัยเฉพาะ (Specific factor) ที่สัมพันธ์กับตัวแปร X_i เท่านั้น ; $i=1,2,\dots,p$

ดังนั้นตัวแปร $\mu_i, 1_{ij}, F_j$ และ ε_i เป็นค่าที่ไม่ได้มีการสังเกตหรือวัดค่าไว้ (Unobserved variables) หรือเขียนตัวแปรปัจจัยในรูปแบบเมตริก คือ

$$\underset{(p \times 1)}{\mathbf{X}} - \underset{(p \times 1)}{\boldsymbol{\mu}} = \underset{(p \times m)}{\mathbf{L}} \underset{(m \times 1)}{\mathbf{F}} + \underset{(p \times 1)}{\boldsymbol{\varepsilon}} \quad \dots\dots\dots(2)$$

(Johnson and Wichern, 2002)

1.2 ข้อตกลงของเวกเตอร์ปัจจัยร่วมและเวกเตอร์ความคลาดเคลื่อน มี 3 ข้อ คือ

1) ปัจจัยร่วมไม่มีความสัมพันธ์กัน

ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างปัจจัยร่วม F_i และ F_j เป็นศูนย์ ตัวแปรปัจจัยร่วมจะสามารถอธิบายตัวแปร X_i ด้วยสมการผลรวมเชิงเส้นระหว่างปัจจัยร่วม (F_1, F_2, \dots, F_m) กับค่าความคลาดเคลื่อนที่เป็นค่าเฉพาะ ($\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p$) โดยที่ส่วนของปัจจัยร่วมไม่มีความสัมพันธ์กันและถ้าข้อมูลถูกจัดให้อยู่ในรูปคะแนนมาตรฐาน จะได้ว่า

$$E(\mathbf{F}) = \underset{(m \times 1)}{\mathbf{0}}, \quad \text{Cov}(\mathbf{F}) = E(\mathbf{F}\mathbf{F}') = \underset{(m \times m)}{\mathbf{I}}$$

หรือ $E(F_j) = 0$, $\text{Var}(F_j) = 1$ และ $\text{Cov}(F_i, F_j) = 0$; $i=1,2,\dots,p$, $j=1,2,\dots,m$ และ $i \neq j$
 ดังนั้น $\text{Corr}(F_i, F_j) = 0$; $i \neq j$ (3)

(Rencher, 1995)

2) ค่าเฉพาะของ ε_i และ ε_j ไม่มีความสัมพันธ์กัน

ส่วนของปัจจัยเฉพาะ (Specific factor) หรือเทอมความคลาดเคลื่อน ε มีค่าเฉลี่ยและความแปรปรวน ดังนี้

$$E(\underline{\varepsilon}) = \underline{0}_{(p \times 1)}, \quad \text{Cov}(\underline{\varepsilon}) = E[\underline{\varepsilon}\underline{\varepsilon}'] = \underline{\theta}_{(p \times p)} = \begin{bmatrix} \theta_1 & 0 & L & 0 \\ 0 & \theta_2 & L & 0 \\ M & M & O & M \\ 0 & 0 & L & \theta_p \end{bmatrix}$$

ซึ่งเป็นเมตริกซ์ตามเส้นทแยงมุม (Diagonal matrix) เรียก θ_i ว่าค่าความแปรปรวนเฉพาะ (Specific variance) ของ ε_i

หรือ $E(\varepsilon_i) = 0$, $\text{Var}(\varepsilon_i) = \theta_i$ และ $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0$; $i = 1, 2, \dots, p$, $j = 1, 2, \dots, m$ และ $i \neq j$
 ดังนั้น $\text{Corr}(\varepsilon_i, \varepsilon_j) = 0$; $i \neq j$ (4)

3) ปัจจัยร่วม F_j และปัจจัยเฉพาะ ε_i ไม่มีความสัมพันธ์ร่วมกัน

โดยที่ \underline{F} และ $\underline{\varepsilon}$ เป็นอิสระต่อกันจะได้

$$\text{Cov}(\underline{\varepsilon}, \underline{F}) = E(\underline{\varepsilon}\underline{F}') = \underline{0}_{(p \times m)}$$

$$= \begin{bmatrix} \text{Cov}(F_1, \varepsilon_1) & \text{Cov}(F_1, \varepsilon_2) & L & \text{Cov}(F_1, \varepsilon_p) \\ \text{Cov}(F_2, \varepsilon_1) & \text{Cov}(F_2, \varepsilon_2) & L & \text{Cov}(F_2, \varepsilon_p) \\ M & M & O & M \\ \text{Cov}(F_m, \varepsilon_1) & \text{Cov}(F_m, \varepsilon_2) & L & \text{Cov}(F_m, \varepsilon_p) \end{bmatrix} \dots\dots\dots(5)$$

หรือ $\text{Cov}(\varepsilon_i, F_j) = 0$; $i = 1, 2, \dots, p$, $j = 1, 2, \dots, m$ และ $i \neq j$

ดังนั้น $\text{Corr}(F_j, \varepsilon_i) = 0$; $i \neq j$ (6)

(Rencher, 1995)

ข้อตกลงทั้ง 3 ข้อนี้มีวัตถุประสงค์เพื่อสร้างรูปแบบสมการวิเคราะห์ปัจจัยที่เป็นแบบอิสระต่อกัน ดังนั้นหลักของการวิเคราะห์ปัจจัยจะเป็นการแสดงค่าความแปรปรวนและค่าความแปรปรวนร่วมของตัวแปรเดิม X_1, X_2, \dots, X_p ในรูปของสมการผลรวมเชิงเส้นระหว่างค่านี้กับปัจจัย (1_{ij}) จำนวน $p \times m$ ตัว กับค่าเฉพาะ $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p$ โดยที่เมตริกซ์ความแปรปรวนร่วมของ $\underline{X} - \underline{\mu}$ คือ $\text{Cov}(\underline{X} - \underline{\mu})$ ซึ่งคำนวณได้ ดังนี้

$$\begin{aligned}
 \Sigma &= \text{Cov}(\underline{X} - \underline{\mu}) \\
 &= E(\underline{X} - \underline{\mu})(\underline{X} - \underline{\mu})' \\
 &= E(\underline{LF} + \underline{\varepsilon})(\underline{LF} + \underline{\varepsilon})' \\
 &= E(\underline{LF} + \underline{\varepsilon})(\underline{LF})' + E(\underline{\varepsilon}\underline{\varepsilon}') \\
 &= E\left[\underline{LF}(\underline{LF})' + \underline{\varepsilon}(\underline{LF})' + \underline{LF}\underline{\varepsilon}' + \underline{\varepsilon}\underline{\varepsilon}'\right] \\
 &= LE(\underline{FF}')L' + E(\underline{\varepsilon}\underline{F}')L' + LE(\underline{F}\underline{\varepsilon}') + E(\underline{\varepsilon}\underline{\varepsilon}') \quad \dots\dots\dots(7)
 \end{aligned}$$

จากข้อตกลงข้อ (1) และ (2)

จะได้ $E(\underline{FF}') = I, \quad E(\underline{\varepsilon}\underline{F}') = 0, \quad E(\underline{F}\underline{\varepsilon}') = 0 \quad \text{และ} \quad E(\underline{\varepsilon}\underline{\varepsilon}') = \theta$

ดังนั้นสมการที่ (7) คือ $\Sigma = \text{Cov}(\underline{X} - \underline{\mu}) = LL' + \theta \quad \dots\dots\dots(8)$

หรือ $\Sigma_{(p \times p)} = L_{(p \times m)} L'_{(m \times p)} + \theta_{(p \times p)} \quad \dots\dots\dots(9)$

$$\begin{aligned}
 \begin{bmatrix} \sigma_{11} & \sigma_{12} & L & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & L & \sigma_{2p} \\ M & M & O & M \\ \sigma_{p1} & \sigma_{p2} & L & \sigma_{pp} \end{bmatrix}_{(p \times p)} &= \begin{bmatrix} l_{11} & l_{12} & L & l_{1m} \\ l_{21} & l_{22} & L & l_{2m} \\ M & M & O & M \\ l_{p1} & l_{p2} & L & l_{pm} \end{bmatrix}_{(p \times m)} \begin{bmatrix} l_{11} & l_{21} & L & l_{p1} \\ l_{12} & l_{22} & L & l_{p2} \\ M & M & O & M \\ l_{1m} & l_{2m} & L & l_{pm} \end{bmatrix}_{(m \times p)} \\
 &+ \begin{bmatrix} \theta_1 & 0 & L & 0 \\ 0 & \theta_2 & L & 0 \\ M & M & O & M \\ 0 & 0 & L & \theta_p \end{bmatrix}_{(p \times p)} \quad \dots\dots\dots(10)
 \end{aligned}$$

(Rencher, 1995)

ในการหาค่าความแปรปรวนร่วมระหว่างตัวแปรเดิม (X) กับปัจจัยร่วม (F) ในรูปของค่าน้ำหนักปัจจัย (Factor loading) คำนวณได้จาก

$$\begin{aligned}
 \text{Cov}(\underline{X}, \underline{F}) &= E(\underline{X} - \underline{\mu})\underline{F}' \\
 &= E(\underline{LF} + \underline{\varepsilon})\underline{F}' \\
 &= LE(\underline{FF}') + E(\underline{\varepsilon}\underline{F}') \\
 &= L \quad \dots\dots\dots(11)
 \end{aligned}$$

หรือ $\text{Cov}(X_i, F_j) = l_{ij} \quad ; i = 1, 2, \dots, p, \quad j = 1, 2, \dots, m \quad \dots\dots\dots(12)$

ความแปรปรวนของ X [$\text{Var}(X_i)$] หาได้จาก

$$\begin{aligned}\text{Var}(X_i) &= (1_{i1}^2 + 1_{i2}^2 + \dots + 1_{im}^2) + \theta_i \\ &= h_i^2 + \theta_i \quad \dots\dots\dots(13) \\ &= \text{Communality} + \text{Specific/Unique variance}\end{aligned}$$

โดยที่ $h_i^2 = 1_{i1}^2 + 1_{i2}^2 + \dots + 1_{im}^2$ หรือเรียกว่าค่าความร่วมกัน (Communality : h_i^2) หรือความแปรปรวนร่วม (Common variance) ซึ่งเกิดจากผลรวมกำลังสองของค่าน้ำหนักของตัวแปร X_i ที่อธิบายได้จากแต่ละปัจจัยร่วม F_1, F_2, \dots, F_m และ θ_i คือ ค่าแปรปรวนเฉพาะของ ε_i เรียกเป็น Specificity variance หรือ Unique variance ; $i = 1, 2, \dots, p$ (Rencher, 1995)

2. การประมาณค่าน้ำหนักปัจจัย (Factor Loading : 1_{ij}) และค่าความร่วมกัน (Communalities : h_i^2)

วิธีการประมาณค่าน้ำหนักปัจจัย (1_{ij}) เป็นวิธีการหาจำนวนปัจจัยร่วมที่เหมาะสม โดยเริ่มคำนวณจากการกำหนดให้มีจำนวนปัจจัยร่วมเป็น 1 ปัจจัย (F_j ; $j = 1, 2, \dots, m$) ที่สามารถอธิบายค่าความผันแปรต่าง ๆ ของตัวแปรเดิมด้วยปัจจัยร่วม ทำให้ปัจจัยร่วมเป็นตัวแทนของข้อมูลตัวแปรเดิมได้

การพิจารณาค่าความร่วมกัน (h_i^2) สำหรับการวิเคราะห์ปัจจัยจะตรวจสอบตัวแบบปัจจัยร่วม (F_1, F_2, \dots, F_j) ที่สร้างขึ้นจากการประมาณค่า 1_{ij} และค่าความแปรปรวนของค่าเฉพาะ (θ_i) โดยสามารถหาค่า h_i^2 จากการกำหนดให้ X_1, X_2, \dots, X_p เป็นตัวแปรอิสระที่มีความสัมพันธ์กัน p ตัวแปร จากสมการ

$$\sigma_{ik} = \text{Cov}(X_i, X_k) = (1_{i1}^1 1_{k1}^1 + 1_{i2}^1 1_{k2}^1 + \dots + 1_{ij}^1 1_{kj}^1) + \theta_i$$

จากข้อตกลงข้อที่ว่า $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0$ จึงทำให้

$$\sigma_{ik} = \text{Cov}(X_i, X_k) = (1_{i1}^1 1_{k1}^1 + 1_{i2}^1 1_{k2}^1 + \dots + 1_{ij}^1 1_{kj}^1) \quad \dots\dots\dots(14)$$

ดังนั้น $\Sigma = \text{Cov}(\underline{X} - \underline{\mu}) = LL'$

หรือเขียนในรูปเมตริกซ์ คือ

$$\Sigma = \underset{(p \times p)}{L} \underset{(p \times m) (m \times p)}{L'}$$

$$\begin{bmatrix} \sigma_{11} & \sigma_{12} & L & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & L & \sigma_{2p} \\ M & M & O & M \\ \sigma_{p1} & \sigma_{p2} & L & \sigma_{pp} \end{bmatrix}_{(p \times p)} = \begin{bmatrix} l_{11} & l_{12} & L & l_{1m} \\ l_{21} & l_{22} & L & l_{2m} \\ M & M & O & M \\ l_{p1} & l_{p2} & L & l_{pm} \end{bmatrix}_{(p \times m)} \begin{bmatrix} l_{11} & l_{21} & L & l_{p1} \\ l_{12} & l_{22} & L & l_{p2} \\ M & M & O & M \\ l_{1m} & l_{2m} & L & l_{pm} \end{bmatrix}_{(m \times p)} \quad \dots(15)$$

เช่น $\sigma_{11} = l_{11}l_{11} + l_{12}l_{12} + l_{13}l_{13} + \dots + l_{1m}l_{1m}$

$\sigma_{12} = l_{11}l_{21} + l_{12}l_{22} + l_{13}l_{23} + \dots + l_{1m}l_{2m}$ (Rencher, 1995)

เมื่อได้ค่านำหนักปัจจัย (l_{ij}) แล้วจะสามารถคำนวณค่าความร่วมกัน (h_i^2) และค่าแปรปรวนของค่าเฉพาะ (θ_i) ได้ตามลำดับ ; $i = 1, 2, \dots, p$, $k = 1, 2, \dots, p$ และ $j = 1, 2, \dots, m$

การประมาณค่า l_{ij} ในปัจจุบันมีหลายวิธี เช่น วิธีองค์ประกอบหลักปัจจัย (Principal Component, PC) วิธีแกนหลัก (Principal Axis, PA) วิธีกำลังสองน้อยที่สุดไม่ปรับน้ำหนัก (Unweighted Least Square, ULS) วิธีกำลังสองน้อยที่สุดทั่วไป (Generalized Least Square, GLS) วิธีภาวะน่าจะเป็นสูงสุด (Maximum Likelihood, ML) วิธีอัลฟา (Alpha) และวิธีเงา (Image) ดังนั้น ในการเลือกวิธีในการประมาณค่า l_{ij} ผู้ศึกษาจึงต้องตรวจสอบความเหมาะสมของแต่ละวิธี เพื่อใช้ในการวิเคราะห์ข้อมูลได้อย่างเหมาะสม เพื่อสามารถรวมกลุ่มตัวแปรเดิมที่มีความสัมพันธ์กันเองได้ครอบคลุมที่สุด โดยมีจำนวนปัจจัยร่วมที่สร้างขึ้นน้อยที่สุดด้วย (Garson, 2009)

การศึกษาครั้งนี้จึงเลือกการประมาณค่า l_{ij} จากวิธีที่ได้รับความนิยมสูงสุด (Garson, 2009; Rencher, 1995; Tabachnick and Fidell, 2007) คือ วิธีองค์ประกอบหลัก (PCM) ซึ่งเป็นวิธีการคำนวณค่าแบบไม่วนซ้ำ และวิธีภาวะน่าจะเป็นสูงสุด (MLM) ที่มีการคำนวณค่าความร่วมกันแบบวนซ้ำ เพื่อใช้ในการเปรียบเทียบทั้ง 2 วิธี ว่าวิธีใดที่เมื่อสร้างปัจจัยร่วมแล้วสามารถอธิบายผลจากการสร้างปัจจัยร่วมได้ดีที่สุด โดยวัดค่าสัดส่วนสะสมของความแปรปรวนที่อธิบายได้ (Σh_i^2) โดยถ้าผลรวมมีค่าสูง แสดงว่าวิธีดังกล่าวสร้างปัจจัยร่วมโดยการรวมกลุ่มตัวแปรเดิมที่มีความสัมพันธ์กันเองให้อยู่ในกลุ่มเดียวกันได้ดี

2.1 วิธีองค์ประกอบหลัก (Principal Component Method, PCM)

Rencher (1995) ได้อธิบายวิธีองค์ประกอบหลัก จากการประมาณค่าน้ำหนัก (Loadings, l_{ij}) ที่ใช้หลักการแทนที่ความผันแปรทั้งหมดของตัวแปรเดิม X_1, X_2, \dots, X_p ด้วยส่วนของปัจจัยร่วม (Common factors) จึงทำให้จำนวนตัวแปรที่สังเกตได้ (p) เท่ากับจำนวนปัจจัย (m)

เขียนตัวแบบของวิธีองค์ประกอบหลัก (PCM) ได้ดังนี้

$$\begin{aligned} PC_1 &= l_{11}(X_1 - \mu_1) + l_{12}(X_2 - \mu_2) + \dots + l_{1p}(X_p - \mu_p) \\ PC_2 &= l_{21}(X_1 - \mu_1) + l_{22}(X_2 - \mu_2) + \dots + l_{2p}(X_p - \mu_p) \\ &\vdots \\ PC_m &= l_{m1}(X_1 - \mu_1) + l_{m2}(X_2 - \mu_2) + \dots + l_{mp}(X_p - \mu_p) \end{aligned} \quad \dots\dots\dots(16)$$

โดยที่ PC_j คือ ตัวแปรใหม่หรือปัจจัยร่วมเป็นค่าที่ไม่ได้ถูกสังเกต (Unobservable) ; $j = 1, 2, \dots, m$

l_{ij} เป็นค่าน้ำหนัก (Loading) ที่แสดงความสัมพันธ์ระหว่างตัวแปรเดิมตัวที่ i (X_i) กับปัจจัยร่วม ตัวที่ j (PC_j) ซึ่ง l_{ij} จะเป็นตัวที่บอกว่าตัวแปรเดิม (X_i) แต่ละตัวมีความสัมพันธ์กับปัจจัยร่วม (PC) มากน้อยเพียงใดโดยที่ $-1 \leq l_{ij} \leq 1$; $i = 1, 2, \dots, p$ และ $j = 1, 2, \dots, m$; $m \leq p$

X_i เป็นตัวแปรอิสระที่สังเกตและวัดค่าได้ (Observed variables) หรือเรียกว่าตัวแปรเดิม ; $i = 1, 2, \dots, p$

μ_i เป็นค่าเฉลี่ยของตัวแปร X_i ; $i = 1, 2, \dots, p$

ดังนั้นค่าของตัวแปร μ_i , PC_i และ l_{ij} เป็นค่าที่ไม่ได้มีการสังเกตหรือวัดค่าไว้

(Unobserved variable) เขียนให้อยู่ในรูปแบบเมตริก ได้คือ

$$\underset{(m \times 1)}{PC} = \underset{(p \times p)}{L} \underset{(p \times 1)}{(X - \mu)} \quad \dots\dots\dots(17)$$

ตัวอย่าง สมมุติสุ่มตัวอย่างตัวแปรเดิม X_1, X_2, \dots, X_p ที่มีเมตริกซ์ความแปรปรวนร่วมของข้อมูลตัวอย่าง (S) และต้องการหาตัวประมาณค่า \hat{L} ดังนั้นจึงต้องพิจารณาสมการของ $\Sigma = LL' + \theta$ จะได้ว่า

$$S \cong \hat{L}\hat{L}' + \hat{\theta} \quad \dots\dots\dots(18)$$

ตามหลักการ PCM จะพิจารณาค่าความแปรปรวนของค่าเฉพาะ $\hat{\theta}$ ให้มีค่าเป็นศูนย์ จึงทำให้ $S = \hat{L}\hat{L}'$ ในการหาค่าของเมตริกซ์ S ซึ่งจะต้องอาศัยวิธีการแยกส่วนเมตริกซ์ S (Spectral decomposition) ได้ดังนี้

$$S = CDC' \quad \text{.....(19)}$$

โดยที่ C เป็นเมตริกซ์เชิงตั้งฉาก (Orthogonal matrix) ที่สร้างด้วยเวกเตอร์ไอเกนที่ปรับปกติ (Normalized eigenvectors) ทำให้ค่า $c_i'c_i = 1$ เมื่อพิจารณาตามคอลัมน์ของเมตริกซ์ S และเมตริกซ์ D คือ

$$D = \begin{bmatrix} \lambda_1 & 0 & L & 0 \\ 0 & \lambda_2 & L & 0 \\ M & M & O & M \\ 0 & 0 & L & \lambda_p \end{bmatrix}_{(p \times p)} \quad \text{.....(20)}$$

เมื่อ $\lambda_1, \lambda_2, \dots, \lambda_p$ เป็นค่าไอเกน (Eigenvalues) ของเมตริกซ์ S ซึ่งจะเปลี่ยนรูปเมตริกซ์ $S = CDC'$ ให้อยู่ในรูปของ $\hat{L}\hat{L}'$ ดังนั้นเมตริกซ์ S จึงเป็นเมตริกซ์จำกัดเขตที่เป็นบวก (Positive definite matrix) ส่งผลให้ค่าไอเกน λ_i เป็นบวกทั้งหมด และสามารถแยกเมตริกซ์ D ได้ดังนี้

$$D = D^{1/2}D^{1/2} \quad \text{.....(21)}$$

และ

$$D^{1/2} = \begin{bmatrix} \sqrt{\lambda_1} & 0 & L & 0 \\ 0 & \sqrt{\lambda_2} & L & 0 \\ M & M & O & M \\ 0 & 0 & L & \sqrt{\lambda_p} \end{bmatrix}_{(p \times p)} \quad \text{.....(22)}$$

(Rencher, 1995)

แทนค่า D ในสมการ (19)

$$\begin{aligned} S &= CDC' \\ &= CD^{1/2}D^{1/2}C' \\ &= (CD^{1/2})(CD^{1/2})' \end{aligned} \quad \text{.....(23)}$$

จึงสามารถเขียนเมตริกซ์ S ให้อยู่ในรูปแบบของเมตริกซ์ $\hat{L}\hat{L}'$ แต่เราไม่สามารถหาเมตริกซ์ \hat{L} ได้โดยตรงในเมตริกซ์ $CD^{1/2}$ เนื่องจากเมตริกซ์ $CD^{1/2}$ มีขนาด $p \times p$ จึงจะต้องประมาณค่าเมตริกซ์ \hat{L} ให้มีขนาด $p \times m$ เพื่อให้จำนวน $m < p$ ต้องสร้างเมตริกซ์ C_1 ให้มีขนาด $p \times m$ และเมตริกซ์ D_1 ขนาด $m \times m$ จึงจะหาค่าเมตริกซ์ \hat{L} ได้จาก

$$\hat{L} = C_1 D_1^{1/2} \quad \dots\dots\dots(24)$$

เมื่อ \hat{L} มีขนาด $p \times m$ โดยที่ $m < p$

C_1 มีขนาด $p \times m$ ประกอบด้วยเวกเตอร์ไอเกน c_1, c_2, \dots, c_m

$D_1^{1/2}$ มีขนาด $m \times m$ ที่ประกอบด้วยค่าไอเกนที่มากที่สุด m ค่า โดยที่ $\lambda_1 > \lambda_2 > \dots > \lambda_m > 0$

ตัวอย่าง ถ้ามีตัวแปรอิสระ 5 ตัว ($p=5$) และมีปัจจัยร่วม 2 ปัจจัย ($m=2$) จะได้

$$\begin{bmatrix} \hat{l}_{11} & \hat{l}_{12} \\ \hat{l}_{21} & \hat{l}_{22} \\ \hat{l}_{31} & \hat{l}_{32} \\ \hat{l}_{41} & \hat{l}_{42} \\ \hat{l}_{51} & \hat{l}_{52} \end{bmatrix}_{5 \times 2} = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \\ c_{31} & c_{32} \\ c_{41} & c_{42} \\ c_{51} & c_{52} \end{bmatrix}_{5 \times 2} \begin{bmatrix} \sqrt{\lambda_1} & 0 \\ 0 & \sqrt{\lambda_2} \end{bmatrix}_{2 \times 2} = \begin{bmatrix} \sqrt{\lambda_1} c_{11} & \sqrt{\lambda_2} c_{12} \\ \sqrt{\lambda_1} c_{21} & \sqrt{\lambda_2} c_{22} \\ \sqrt{\lambda_1} c_{31} & \sqrt{\lambda_2} c_{32} \\ \sqrt{\lambda_1} c_{41} & \sqrt{\lambda_2} c_{42} \\ \sqrt{\lambda_1} c_{51} & \sqrt{\lambda_2} c_{52} \end{bmatrix}_{5 \times 2} \quad \dots\dots\dots(25)$$

(Rencher, 1995)

จากสมการที่ (25) ได้คำตอบในเทอมของการวิเคราะห์ตัวประกอบหลัก เพื่อใช้อธิบายในแต่ละคอลัมน์ของเมตริกซ์ \hat{L} ซึ่งเป็นสัดส่วนของเวกเตอร์ไอเกน (c_1, c_2, \dots, c_m) ของเมตริกซ์ความแปรปรวนร่วม (S) ดังนั้นค่า Loading ในแต่ละปัจจัย คือสัดส่วนของค่าสัมประสิทธิ์ขององค์ประกอบหลักที่ j (PC_j) กับเวกเตอร์ไอเกนที่ i (λ_i) และในการแปลผลของ PCM หลังจากที่มีการหมุนแกน (Rotation) จะทำให้ค่าน้ำหนักของแต่ละปัจจัยจะเปลี่ยนไป ส่งผลให้การจัดกลุ่มตัวแปร X_1, X_2, \dots, X_p เข้าในปัจจัยร่วมจะแตกต่างจากก่อนการหมุนแกน ทำให้สมาชิกในแนวทแยงมุมของเมตริกซ์ $\hat{L}\hat{L}'$ ก็คือค่าผลรวมกำลังสองของแถวที่ i ของ \hat{L} หรือ $\hat{l}'_i \hat{l}_i = \sum_{j=1}^m l_{ij}^2$ ในการประมาณค่าของเมตริกซ์ S ก็จะเสร็จสมบูรณ์

$$\text{คือ} \quad \hat{\theta}_i = s_{ii} - \sum_{j=1}^m l_{ij}^2 \quad \dots\dots\dots(26)$$

$$S \cong \hat{L}\hat{L}' + \hat{\theta} \quad \dots\dots\dots(27)$$

เมื่อ $\hat{\theta} = \text{diag}(\hat{\theta}_1, \hat{\theta}_1, \dots, \hat{\theta}_p)$ คือความแปรปรวน(Variance) ของสมาชิกในแนวทแยงมุมของเมตริกซ์ S จะถูกสร้างตัวแบบที่ดี แต่ความแปรปรวนร่วมที่อยู่นอกเส้นทแยงมุมจะไม่ได้นำมาสร้างตัวแบบ (Rencher, 1995)

ในวิธีการประมาณค่า PCM ซึ่งก็คือผลรวมกำลังสองของแถวและคอลัมน์ของเมตริกซ์ \hat{L} จะได้เท่ากับค่าความร่วมกัน (h_i^2) และค่าไอเกน (λ_j) ตามลำดับ การคำนวณค่า h_i^2 ของตัวแปร X_i ในแต่ละปัจจัยร่วม (F_1, F_2, \dots, F_j) ประมาณค่าได้จาก

$$h_i^2 = \sum_{j=1}^m l_{ij}^2 \quad \dots\dots\dots(28)$$

ซึ่งเป็นผลรวมกำลังสองของค่าน้ำหนักปัจจัย ในแถวที่ i ของเมตริกซ์ \hat{L} และผลรวมกำลังสองของค่าน้ำหนักปัจจัย ในคอลัมน์ที่ j คือ ค่าไอเกนที่ j ของเมตริกซ์ S

$$\begin{aligned} \sum_{i=1}^p l_{ij}^2 &= \sum_{i=1}^p (\sqrt{\lambda_i} c_{ij})^2 \\ &= \lambda_j \sum_{i=1}^p c_{ij}^2 \\ &= \lambda_j \quad ; i = 1, 2, \dots, p, j = 1, 2, \dots, m \quad \dots\dots\dots(29) \end{aligned}$$

เนื่องจากเวกเตอร์ไอเกนที่ปรับปกติแล้วจะมีค่าเป็น 1 หรือ $c_i'c_i = \sum_{i=1}^p c_{ij}^2 = 1$

ถ้าตัวแปรเดิม X_1, X_2, \dots, X_p มีค่าหน่วยวัดไม่เป็นหน่วยเดียวกัน โดยทั่วไปจะปรับข้อมูลให้อยู่ในรูปคะแนนมาตรฐาน (Standardize score) ก่อน เพื่อที่จะใช้เมตริกซ์ค่าแปรปรวนร่วม (covariance matrices, S) ก็จะสามารถใช้เป็นเมตริกซ์สัมประสิทธิ์สหสัมพันธ์ (correlation matrices, R) ได้ ดังนั้น ค่าไอเกน λ_j และเวกเตอร์ไอเกน c_1, c_2, \dots, c_m ของเมตริกซ์ R จะถูกนำไปใช้เพื่อหาค่าน้ำหนัก Loading ที่ได้จากโปรแกรมสำเร็จรูปทางสถิติทั่วไป และในการวิเคราะห์ด้วย PCM ก็จะใช้เมตริกซ์ R ด้วย (Rencher, 1995)

โดยพิจารณาจากสมการ (26) คือ $\hat{\theta}_i = s_{ii} - \sum_{j=1}^m \hat{1}_{ij}^2$

ความแปรปรวนร่วมของตัวอย่างที่ i (s_{ii}) แยกออกเป็น 2 ส่วน คือ ส่วนของปัจจัยร่วม (Common factors) และส่วนของค่าเฉพาะ (Specific) คือ

$$\begin{aligned} s_{ii} &= \sum_{j=1}^m \hat{1}_{ij}^2 + \hat{\theta}_i \\ &= \hat{h}_i^2 + \hat{\theta}_i \\ &= (\hat{1}_{i1}^2 + \hat{1}_{i2}^2 + \dots + \hat{1}_{im}^2) + \hat{\theta}_i \end{aligned} \quad \dots\dots\dots(30)$$

จะสังเกตได้ว่าปัจจัยร่วมที่ j มีส่วนร่วมใน $\hat{1}_{ij}^2$ ถึง s_{ii} ซึ่งการมีส่วนร่วมของปัจจัยที่ j คือ ความแปรปรวนทั้งหมดของตัวอย่าง [$\text{tr}(S) = s_{11} + s_{22} + \dots + s_{pp}$] คำนวณได้จาก $\sum_{i=1}^p \hat{1}_{ij}^2$ จะเท่ากับ ผลรวมกำลังสองของค่านำหน้าในคอลัมน์ที่ j ของเมตริกซ์ \hat{L} คือ

$$\hat{1}_{1j}^2 + \hat{1}_{2j}^2 + \dots + \hat{1}_{pj}^2 \quad \dots\dots\dots(31)$$

ซึ่งเท่ากับค่าไอเกนที่ j (λ_j) ดังนั้นสัดส่วนความแปรปรวนร่วมของตัวอย่างทั้งหมดที่ได้มาจาก ปัจจัยที่ j คือ

$$\frac{\lambda_j}{\text{tr}(S)} = \frac{\hat{1}_{1j}^2 + \hat{1}_{2j}^2 + \dots + \hat{1}_{pj}^2}{\text{tr}(S)} \quad \dots\dots\dots(32)$$

และถ้าใช้การแยกปัจจัยในเมตริกซ์ R ก็จะสอดคล้องกับค่าสัดส่วน ของ

$$\frac{\hat{1}_{1j}^2 + \hat{1}_{2j}^2 + \dots + \hat{1}_{pj}^2}{\text{tr}(R)} = \frac{\lambda_j}{p} \quad \dots\dots\dots(33)$$

โดยที่ p คือ จำนวนตัวแปร (Rencher, 1995)

จากสมการที่ (33) จะเห็นได้ว่าถ้าค่าความร่วมกัน h_i^2 มีค่าน้อย จะส่งผลให้สัดส่วนความแปรปรวนของ ค่าไอเกนที่ j (λ_j) ของเมตริกซ์ S หรือค่าไอเกนที่ j (λ_j) ของเมตริกซ์ R ก็จะมีค่าน้อยด้วย ซึ่งในทางปฏิบัติจะนิยมใช้เมตริกซ์ R มากกว่าใช้เมตริกซ์ S ในการวิเคราะห์หาค่านำหนักปัจจัย (Rencher, 1995)

2.2 วิธีภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Method, MLM)

Johnson and Wichern (2002) ในการประมาณค่า μ_j และ θ ด้วยวิธี MLM มีข้อสมมุติว่า ปัจจัยร่วม (F_j) และค่าเฉพาะ (ε_i) ที่มีการแจกแจงแบบปกติหลายตัวแปร (Multivariate normal distribution) มีเวกเตอร์ค่าเฉลี่ยเป็นศูนย์ เมตริกซ์ค่าแปรปรวนร่วมเป็น Σ (Common variance-covariance matrix) หรือเขียนในรูปสัญลักษณ์ คือ $X: Np(0, \Sigma)$ ดังนั้นจึงทำให้สมการ $\underline{X}_i - \underline{\mu}_i = LF_j + \underline{\varepsilon}_i$ มีการแจกแจงแบบปกติหลายตัวแปรด้วย จากสมการ Likelihood function

$$\begin{aligned}
 l(\underline{\mu}, \Sigma) &= (2\pi)^{-\frac{np}{2}} |\Sigma|^{-\frac{n}{2}} e^{-\left(\frac{1}{2}\right) \text{tr} \left[\Sigma^{-1} \left(\sum_{j=1}^n (x_j - \bar{x})(x_j - \bar{x})' + n(\bar{x} - \mu)(\bar{x} - \mu)' \right) \right]} \\
 &= (2\pi)^{-\frac{(n-1)p}{2}} |\Sigma|^{-\frac{(n-1)}{2}} e^{-\left(\frac{1}{2}\right) \text{tr} \left[\Sigma^{-1} \left(\sum_{j=1}^n (x_j - \bar{x})(x_j - \bar{x})' \right) \right]} \times (2\pi)^{-\frac{p}{2}} |\Sigma|^{-\frac{1}{2}} e^{-\left(\frac{n}{2}\right) (\bar{x} - \mu)' \Sigma^{-1} (\bar{x} - \mu)} \\
 &\dots\dots\dots(34)
 \end{aligned}$$

ค่าที่ได้จะขึ้นกับค่า \hat{L} และ $\hat{\theta}$ ที่ผ่านสมการ $\Sigma = LL' + \theta$ โดยปกติเวกเตอร์ค่าเฉลี่ย ตัวอย่าง (Sample mean vector, \bar{X}) เป็นค่า MLE ของ μ ในการหา MLE ของ \hat{L} และ $\hat{\theta}$ จะใช้วิธีการคำนวณแบบวนซ้ำเชิงตัวเลข โดยใช้เมตริกซ์ S (Sample variance-covariance matrix) เป็นค่าประมาณเริ่มต้นของ เมตริกซ์ Σ เนื่องจากการแจกแจงแบบปกติหลายตัวแปร (Multivariate normal distribution) ดังนั้น $(n-1)S$ จะมีการแจกแจงแบบ Wishart distribution ที่มี $df = n - 1$ และค่าคาดหวังของ S เท่ากับ Σ หรือ $E(S) = \Sigma$ โดยฟังก์ชันภาวะน่าจะเป็นของ S คือ

$$f(S) = c \times |\Sigma|^{-\frac{(n-1)}{2}} \times |S|^{\frac{(n-1)}{2} - \frac{(p+1)}{2}} \times e^{-\frac{(n-1)}{2} \text{tr}(\Sigma^{-1}S)}$$

โดยที่ c เป็นค่าคงที่ ดังนั้น Log-likelihood ของ \hat{L} และ $\hat{\theta}$ เมื่อ $\Sigma = LL' + \theta$ จะเท่ากับ

$$\ln c - \left(\frac{n-1}{2} \right) \left\{ \text{tr} \left[(LL' + \theta)^{-1} S \right] - \ln \left| (LL' + \theta)^{-1} S \right| \right\}$$

MLE ของ \hat{L} และ $\hat{\theta}$ ได้จากการทำให้สมการดังกล่าวมีค่ามากที่สุดเมื่อเทียบกับเงื่อนไขเฉพาะที่ว่า $L'\theta^{-1}L = \Delta$ a diagonal matrix จะต้องอยู่บนเส้นทแยงมุมของเมตริกซ์ Σ (Orthogonal transformation) (Johnson and Wichern, 2002; Ravindra and Dayanand, 2000)

ดังนั้น ค่าประมาณภาวะน่าจะเป็นสูงสุดของค่าความร่วมกัน คือ

$$\hat{h}_i^2 = \hat{i}_{i1}^2 + \hat{i}_{i2}^2 + \dots + \hat{i}_{im}^2 ; i = 1, 2, \dots, p \quad \dots\dots\dots(35)$$

และสัดส่วนความแปรปรวนร่วมของตัวอย่างปัจจัยที่ j มีค่าเป็น $\frac{\hat{i}_{1j}^2 + \hat{i}_{2j}^2 + \dots + \hat{i}_{pj}^2}{S_1^2 + S_2^2 + \dots + S_p^2} \quad \dots\dots\dots(36)$

วิธี MLM จะมีประสิทธิภาพกว่าวิธี PCM ในกรณีที่ค่าประมาณน้ำหนักปัจจัยที่ได้ไม่ขึ้นกับหน่วย หรือมีหน่วยเหมือนกันทุกตัวแปรที่นำมาวิเคราะห์ ซึ่งแตกต่างจาก PCM หน่วยจะมีผลต่อการประมาณค่าน้ำหนักปัจจัย จึงจะต้องปรับข้อมูลให้อยู่ในเมตริกซ์ R ก่อนนำไปวิเคราะห์ปัจจัย ซึ่งถ้าเป็น MLM จะสามารถใช้เมตริกซ์ Σ หรือ S ได้ทันที แต่ถึงอย่างไร MLM ก็สามารถที่จะเริ่มต้นด้วยการใช้เมตริกซ์สหสัมพันธ์ R ได้และยังสามารถปรับเมตริกซ์ R ใหม่ ทำให้ได้ค่าสัดส่วนความแปรปรวนร่วมของตัวอย่างที่ใช้เมตริกซ์ R เป็นค่าคะแนนมาตรฐานของข้อมูล ปัจจัยที่ j คือ

$$\text{สัดส่วนความแปรปรวนร่วมของเมตริกซ์ R ปัจจัยที่ } j = \frac{\hat{i}_{1j}^2 + \hat{i}_{2j}^2 + \dots + \hat{i}_{pj}^2}{p} \quad \dots\dots\dots(37)$$

ปัญหาที่เกิดขึ้นของวิธีภาวะน่าจะเป็นสูงสุด คือค่าประมาณที่คำนวณได้มีค่าไม่แน่นอน ซึ่งหมายความว่ามิได้หลายค่าเนื่องจากค่าต่าง ๆ ในเมตริกซ์ L และ θ ไม่เป็นอิสระกัน จึงต้องใช้วิธีการวนซ้ำ (Iteration) เพื่อประมาณค่าความร่วมกัน ซึ่งในแต่ละรอบการทำงานค่าความร่วมกันจะเปลี่ยนแปลงไป และจะหยุดการประมาณค่าความร่วมกันก็ต่อเมื่อค่าประมาณที่ได้เข้าสู่ค่าคงที่ (Rencher, 1995)

จากทั้ง 2 วิธีที่ได้กล่าวมา คือวิธีองค์ประกอบหลัก (PCM) และวิธีภาวะความเป็นไปได้ (MLM) ทั้ง 2 วิธีสามารถใช้โปรแกรมสำเร็จรูปทางสถิติ เช่น SAS และ SPSS ในการประมาณค่าความร่วมกันได้ โดยเฉพาะวิธี MLM ใช้การคำนวณแบบวนซ้ำ ซึ่งต้องใช้เวลาในการคำนวณมาก

เมื่อหาค่าความร่วมกัน (h^2) จากค่านำหนักปัจจัย (l_{ij}) ได้แล้ว ดังนั้นจะได้เมตริกซ์ R ขนาด $p \times p$ ที่แสดงรายละเอียดของค่า l_{ij} จำนวน $p \times m$ ตัว โดยจำนวนปัจจัยร่วมดังกล่าวยังไม่เป็นจำนวนปัจจัยร่วมที่เหมาะสม เนื่องจากมีจำนวนปัจจัยร่วม (m) เท่ากับจำนวนตัวแปรเดิม (p), $m = p$ (Rencher, 1995)

3. การหาจำนวนปัจจัยที่เหมาะสม (Choosing the Number of Factors, m)

จำนวนปัจจัยร่วมที่เหมาะสมจะต้องสามารถอธิบายความแปรปรวนร่วมของตัวแปรเดิมทุกตัวให้ได้มากที่สุด และควรมีจำนวนปัจจัยร่วมน้อยที่สุด จึงจะถือว่าเป็นตัวแทนที่ดีได้ โดยมีเกณฑ์การพิจารณาจำนวนปัจจัยร่วมดังนี้

3.1 กำหนดค่าไอเกน (Eigen value Specification)

โดยทั่วไปนิยมกำหนดค่าไอเกนที่เกินหนึ่ง ($\lambda_i \geq 1$) เป็นเกณฑ์ในการเลือกจำนวนปัจจัยร่วมไปใช้ โดยปรับค่าหน่วยวัดของตัวแปรเดิมให้เป็นหน่วยเดียวกันก่อนหรือทำให้อยู่ในรูปของเมตริกซ์ R ก่อนการพิจารณา วิธีการนี้เหมาะสำหรับข้อมูลที่มีตัวแปรจำนวน 20 – 50 ตัวแปร (Hair and Black, 1998)

3.2 พิจารณาค่าเฉลี่ยของค่าความร่วมกัน (Average Communalities)

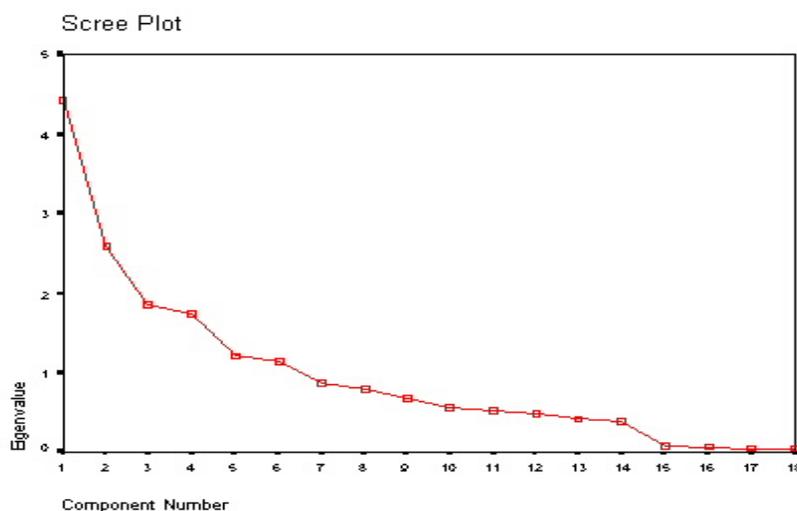
เลือกจำนวนปัจจัย (m) ที่มีค่า $\lambda_i \geq 1$ ถ้าใช้เมตริกซ์ R โดยเลือกเฉพาะปัจจัยร่วมมีค่า λ_i มากกว่าค่าเฉลี่ยเป็นเกณฑ์ หรือถ้าเป็นเมตริกซ์ S จะพิจารณาค่า $\sum_{i=1}^p \frac{\theta_i}{p}$ โดยเลือกเฉพาะปัจจัยร่วมมีค่า λ_i มากกว่าค่าเฉลี่ยของ $\sum_{i=1}^p \frac{\theta_i}{p}$ ซึ่งวิธีการพิจารณาค่าเฉลี่ยของค่าความร่วมกันจะทำให้มีจำนวนปัจจัยร่วมน้อยเกินไป

3.3 พิจารณาค่าสัดส่วนสะสมของความแปรปรวนที่อธิบายได้ (Cumulative Proportion)

Rencher (1995) ควรเลือกจำนวนปัจจัย (m) ที่เหมาะสมสามารถอธิบายตัวแปรเดิมเท่าที่จำเป็น โดยพิจารณาจากค่าสัดส่วนสะสมของความแปรปรวนที่อธิบายได้ 80 เปอร์เซ็นต์ขึ้นไป

3.4 การทดสอบสกรี้ (Scree Test)

โดยกำหนดให้ค่าไอเกนอยู่ในแกน Y และจำนวนปัจจัยรวมอยู่ในแกน X ซึ่งการพิจารณาจำนวนปัจจัยรวม ว่าควรมีกี่ปัจจัยจะพิจารณาจากการวาดเส้นกราฟซึ่งจุดตัดจะเป็นจุดเปลี่ยนของเส้นกราฟ หรือจุดตัดที่เส้นกราฟเริ่มเปลี่ยนจากเส้นโค้งเป็นเส้นตรง วิธีนี้เป็นวิธีที่ดีเมื่อนักวิจัยสนใจศึกษาปัจจัยรวมที่สำคัญ (Garson, 2009) ดังภาพที่ 2



ภาพที่ 2 การทดสอบ Scree test

ที่มา: <http://faculty.chass.ncsu.edu/garson/PA765/factor.htm>

เห็นได้ว่าจุดตัดที่เส้นกราฟเริ่มเปลี่ยนจากเส้นโค้งเป็นเส้นตรง และมีค่าไอเกนมากกว่า 1 คือจุดตัดที่เลข 4 ซึ่งหมายความว่า ค่าไอเกนมากกว่า 1 มีจำนวน 4 ปัจจัย

3.5 การทดสอบสมมติฐานทางสถิติ

การตรวจสอบความเหมาะสมในการใช้เทคนิคการวิเคราะห์ปัจจัย จะใช้เมตริกซ์ R ซึ่งค่าเมตริกซ์ R ที่เหมาะสมควรแตกต่างจากศูนย์ เนื่องจากถ้าตัวแปรไม่มีความสัมพันธ์กัน แสดงว่าข้อมูลตัวแปรเหล่านั้นก็ไม่มีค่าความร่วมกัน (Communality) และไม่มีประโยชน์ที่จะนำเมตริกซ์ R ไปวิเคราะห์ปัจจัยต่อ ดังนั้นจึงมีการทดสอบข้อสมมติฐานที่ว่าเมตริกซ์สหสัมพันธ์นั้นเป็นเมตริกซ์เอกลักษณ์ (Identity matrix) หรือไม่ โดยใช้การทดสอบทางสถิติ 2 วิธี ดังนี้

1) การทดสอบแบบ Bartlett's Sphericity

จากการแจกแจงโดยการประมาณค่าแบบไคกำลังสอง (χ^2) เป็นฟังก์ชันของค่าดีเทอร์มิแนนต์ของเมตริกซ์ R มีข้อสมมติฐานว่าตัวแปรจะต้องมีการแจกแจงแบบปกติหลายตัวแปร หรือ $x : N_p(\mu, \Sigma)$ (Rencher, 1995)

สมมติฐานการทดสอบ คือ

$$H_0 : \Sigma_{(p \times p)} = L_{(p \times m)} L'_{(m \times p)} + \theta_{(p \times p)}$$

และ $H_1 : \Sigma_{(p \times p)} \neq L_{(p \times m)} L'_{(m \times p)} + \theta_{(p \times p)}$

หรือ $H_0 : \text{เมตริกซ์สหสัมพันธ์เป็นเมตริกซ์เอกลักษณ์ } (R = I)$

$H_1 : \text{เมตริกซ์สหสัมพันธ์ไม่เป็นเมตริกซ์เอกลักษณ์ } (R \neq I)$

การทดสอบแบบ Bartlett's Sphericity คำนวณจาก

$$\chi^2 = - \left[(n-1) - \frac{(2p+5)}{6} \right] \ln |R| \quad \dots\dots\dots(38)$$

โดยที่ χ^2 มีองศาอิสระเท่ากับ $\frac{1}{2}[p(p-1)]$ หากค่าดีเทอร์มิแนนต์ของเมตริกซ์ R จากการคำนวณค่าไอเกน

จากสูตร $|R| = \prod_{i=1}^p \lambda_i$ (39)

โดยที่ λ_i คือค่าไอเกนของตัวแปรที่ i ; $i=1,2,\dots,p$

p คือจำนวนตัวแปร

n คือจำนวนข้อมูล

สถิติทดสอบ χ^2 อธิบายได้ว่า

ถ้าตัวแปรไม่มีความสัมพันธ์กัน เมตริกซ์ $R =$ เมตริกซ์ I ทำให้ค่า λ_i มีค่าใกล้ 1 หรือ $|R|$ มีค่าใกล้ 1 ทำให้ $\ln|R|$ เข้าสู่ศูนย์ ส่งผลให้ค่า χ^2 ต่ำ จึงไม่สามารถปฏิเสธ H_0 ได้ คือตัวแปรอิสระนำมาวิเคราะห์ไม่เหมาะสมในการวิเคราะห์ปัจจัย

ถ้าตัวแปรมีความสัมพันธ์กัน เมตริกซ์ $R \neq$ เมตริกซ์ I ทำให้ค่า λ_i มีค่ามากกว่า 1 หรือ $|R|$ มีค่ามากกว่า 1 ทำให้ ขณะที่ λ_i บางตัวมีค่าใกล้ 0 ทำให้ $\ln|R|$ มีค่าติดลบมาก ซึ่งมีผลให้ค่า χ^2 สูง จึงปฏิเสธ H_0 ได้ คือตัวแปรอิสระที่นำมาวิเคราะห์มีความเหมาะสมในการวิเคราะห์ปัจจัย

2) การคำนวณค่าดัชนีไคเซอร์-ไมเยอร์-ออลกิน (Kaiser-Meyer-Olkin)

เป็นดัชนีบอกความแตกต่างระหว่างเมตริกซ์สหสัมพันธ์ของตัวแปรตั้งเกิดได้ กับ เมตริกซ์สหสัมพันธ์แอนติอิมเมจหรือปฏิภาพ (Anti-image correlation matrix) ซึ่งเป็นเมตริกซ์ของสหสัมพันธ์พาร์เซิลระหว่างตัวแปรแต่ละคู่ เมื่อจัดความแปรปรวนของตัวแปรอื่น ๆ ออกไป (Hair and Black, 1998)

การทดสอบแบบ Kaiser-Meyer-Olkin (KMO) คำนวณจาก

$$KMO = \frac{\sum_{i \neq j} r_{ij}^2}{\sum_{i \neq j} r_{ij}^2 + \sum_{i \neq j} a_{ij}^2} ; 0 < KMO < 1 \quad \text{.....(40)}$$

โดยที่ r_{ij} คือ สัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปร X_i และ X_j ; $i \neq j$

a_{ij}^2 คือ สัมประสิทธิ์สหสัมพันธ์บางส่วนระหว่างตัวแปร X_i และ X_j เมื่อควบคุมตัวแปรอื่น ๆ ให้มีค่าคงที่

การแปลความหมายค่าดัชนีไคเซอร์-ไมเยอร์-ออลกิน (Hair and Black, 1998)

- ถ้า $KMO \geq 0.90$ หมายถึง ข้อมูลสามารถใช้ในการวิเคราะห์ปัจจัยระดับดีมาก
 0.80 – 0.89 หมายถึง ข้อมูลสามารถใช้ในการวิเคราะห์ปัจจัยระดับดี
 0.70 – 0.69 หมายถึง ข้อมูลสามารถใช้ในการวิเคราะห์ปัจจัยระดับปานกลาง
 ≤ 0.60 หมายถึง ข้อมูลไม่เหมาะสมในการใช้เทคนิควิเคราะห์ปัจจัย

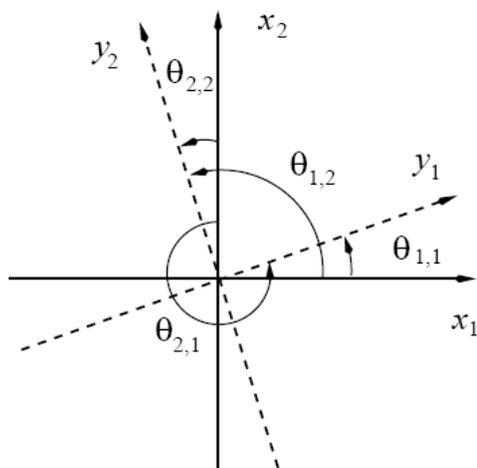
4. การหมุนแกน (Method of Rotation)

ใช้เพื่อให้ง่ายต่อการแปลผล และอธิบายโครงสร้างทางทฤษฎีได้ชัดเจนขึ้น เช่น บางตัวแปรที่มีค่าน้ำหนักปัจจัย (Factor loading) ใกล้เคียงกัน ทำให้ลำบากต่อการจัดตัวแปรนั้น ๆ ว่าควร จะอยู่ในปัจจัยร่วมตัวใด ดังนั้นจึงต้องมีการหมุนแกนปัจจัย (Factor rotation) โดยการให้ค่าน้ำหนัก ของตัวแปรนั้น ๆ มีการเปลี่ยนแปลงสามารถจัดให้อยู่ในปัจจัยที่เหมาะสม และมีความชัดเจน ยิ่งขึ้น (Johnson and Wichern, 2002) การหมุนแกนปัจจัยมี 2 ประเภท ได้แก่

4.1 การหมุนแกนแบบมุมฉาก (Orthogonal Rotation)

เป็นการหมุนแกนปัจจัยในลักษณะที่ แกนของปัจจัยยังคงตั้งฉากกัน ($\theta=90^\circ$) ผล ของการหมุนจะทำให้ค่าได้ค่าน้ำหนักที่หมุนแกนแล้ว (Rotation factor loading) เพิ่มขึ้นหรือลดลง ได้ แต่พิสัยยังคงที่ตำแหน่งเดิม การหมุนแกนปัจจัยแบบมุมฉากจะทำให้ตัวแปรปัจจัยแต่ละตัว ไม่มี ความสัมพันธ์ระหว่างกัน ซึ่งมีวิธีย่อยด้วยกัน 3 วิธี คือ 1) วิธี Varimax 2) วิธี Quartmax และ 3) วิธี Equamax

วิธีการหมุนแกนปัจจัย (Factor rotation) ที่สามารถอธิบายให้เห็นภาพได้ง่ายที่สุด คือการพิจารณาด้วยกราฟ แต่มีข้อจำกัดที่สามารถอธิบายได้เฉพาะกรณีที่มีจำนวนปัจจัยร่วม 2 ปัจจัยเท่านั้น ($m=2$) (Herve, 2003)



ภาพที่ 3 การหมุนแกนแบบตั้งฉากในกรณีที่มี 2 ปัจจัย

ที่มา: <http://www.utdallas.edu/~herve/Abdi-rotations-pretty.pdf>

จากภาพที่ 3 การหมุนแกน x_1 และ x_2 ยังคงทำให้ได้แกน y_1 ที่ตั้งฉากกับ y_2 อยู่เช่นเดิม ซึ่งสามารถคำนวณค่านำหนักที่หมุนแกนแล้ว (Rotation factor loading) จากสูตรดังนี้

กรณีตามเข็มนาฬิกา

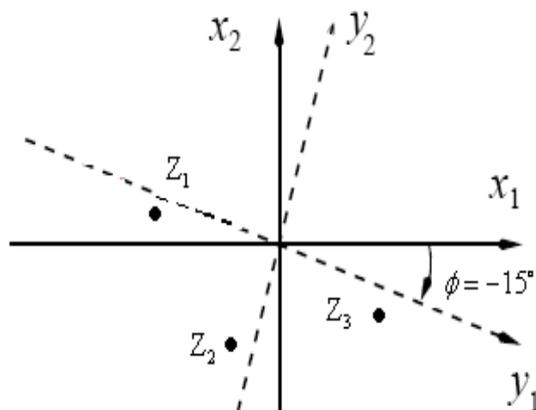
$$R = \begin{bmatrix} \cos \theta_{1,1} & \cos \theta_{1,2} \\ \cos \theta_{2,1} & \cos \theta_{2,2} \end{bmatrix} = \begin{bmatrix} \cos \theta_{1,1} & \sin \theta_{1,1} \\ -\sin \theta_{1,1} & \cos \theta_{1,1} \end{bmatrix} \quad \dots\dots\dots(41)$$

และ กรณีทวนเข็มนาฬิกา

$$R = \begin{bmatrix} \cos \theta_{1,1} & \cos \theta_{2,1} \\ \cos \theta_{1,2} & \cos \theta_{2,2} \end{bmatrix} = \begin{bmatrix} \cos \theta_{1,1} & -\sin \theta_{1,1} \\ \sin \theta_{1,1} & \cos \theta_{1,1} \end{bmatrix} \quad \dots\dots\dots(42)$$

(Johnson and Wichern, 2002)

ตัวอย่าง เมตริกซ์ค่านำหนักปัจจัย โดยมีแกนอ้างอิงเป็น x_1 และ x_2 จะได้จุดพิกัดของตัวแปร Z_1 (ตัวแปร X_1 ที่อยู่ในรูปคะแนนมาตรฐาน) วัดระยะห่างตามแกน x_1 เป็นระยะทาง 0.9 หน่วย และวัดไปตามแกน x_2 เป็นระยะทาง 0.2 หน่วย จุดพิกัด Z_1 เป็น $(-0.44, 0.06)$ จะได้จุดพิกัดของ Z_2 เป็น $(-0.05, -0.73)$ และ Z_3 เป็น $(0.42, -0.31)$ ตามลำดับ



ภาพที่ 4 การหมุนแกนแบบตั้งฉากในกรณีที่มี 2 ปัจจัย ด้วยวิธี Varimax

ที่มา: <http://www.utdallas.edu/~herve/Abdi-rotations-pretty.pdf>

จากตัวอย่างจะต้องพล็อตกราฟได้ดังภาพที่ 4 แล้วหมุนแกน x_1 และ x_2 ตามเข็มนาฬิกาให้ทำมุม $\theta = (-15^\circ)$ กับแกนเดิม สมมติให้แกนใหม่เป็น y_1 และ y_2 โดยที่แกนทั้งสองยังคงตั้งฉากกัน จะได้จุดพิกัดของตัวแปรดังนี้

$$\text{พิกัดที่ 1} \quad y_1 = (a_1)\cos(-15^\circ) + (a_2)\sin(-15^\circ) \text{ และ}$$

$$\text{พิกัดที่ 2} \quad y_2 = (a_2)\cos(15^\circ) - (a_1)\sin(15^\circ)$$

$$\text{พิกัดของ } Z_1 : y_1 = (-0.44)(0.97) + (0.06)(-0.26) = (-0.44)$$

$$y_2 = (0.06)(0.97) - (-0.44)(-0.26) = (-0.05)$$

ได้พิกัดของ Z_1 คือ (-0.44, -0.05)

$$\text{พิกัดของ } Z_2 : y_1 = (-0.26)(0.97) + (-0.58)(-0.26) = (-0.11)$$

$$y_2 = (-0.58)(0.97) - (-0.26)(-0.26) = (-0.63)$$

ได้พิกัดของ Z_2 คือ (-0.11, -0.63)

$$\text{พิกัดของ } Z_3 : y_1 = (0.42)(0.97) + (-0.31)(-0.26) = 0.48$$

$$y_2 = (-0.31)(0.97) - (0.42)(-0.26) = (-0.20)$$

ได้พิกัดของ Z_3 คือ (0.48, -0.20)

4.2 การหมุนแกนแบบเฉียง (Oblique Rotation)

เป็นการหมุนแกนที่คล้ายกับวิธีหมุนแกนแบบตั้งฉาก แต่วิธีนี้จะต้องกำหนดมุมของแกนปัจจัยด้วยตนเองว่าจะกำหนดให้มุมเป็นเท่าใด โดยจะต้องไม่เป็นการหมุนแบบตั้งฉาก ($\theta \neq 90^\circ$) ซึ่งทำให้การหมุนแกนปัจจัยแบบเฉียงตัวแปรปัจจัยอาจมีความสัมพันธ์ระหว่างกันได้ ซึ่งมีวิธีย่อย 2 วิธี คือ 1) วิธี Direct oblimin และ 2) วิธี Promax

ในกรณีที่มีจำนวนปัจจัยรวมมากกว่า 2 ปัจจัย ($m > 2$) จะไม่สามารถใช้วิธี Graphical approach ได้ จึงต้องเลือกใช้วิธีการหมุนแกนแบบมุมฉาก (Orthogonal rotation) ด้วยวิธีแวร์ริแมกซ์ (Varimax) เพื่อลดจำนวนตัวแปรให้เหลือน้อยที่สุด และเป็นวิธีที่นิยมใช้กันมากที่สุด โดยใช้เกณฑ์การพิจารณาคัดเลือกตัวแปรของปัจจัยรวม มี 2 หลักการ คือ

หลักการที่ 1 เป็นกฎทั่วไป (Rule of Thumb) โดยพิจารณาค่าน้ำหนักตัวแปรที่มีค่ามากกว่า 0.3 ($|a_{ij}| > 0.3$) (Hair and Black, 1998; Rencher, 1995)

หลักการที่ 2 พิจารณาจากขนาดตัวอย่าง ซึ่งค่าน้ำหนักปัจจัยจะแปรผกผันกับขนาดตัวอย่าง ดังตารางที่ 4 (Hair and Black, 1998)

ตารางที่ 4 แสดงขนาดตัวอย่างกับค่าน้ำหนักปัจจัย ในการพิจารณาจำนวนปัจจัยที่เหมาะสม

ค่าน้ำหนักปัจจัย	0.30	0.35	0.40	0.45	0.50	0.55	0.60	0.65	0.70	0.75
ขนาดตัวอย่าง	350	250	200	150	120	100	85	70	60	50

5. การคำนวณค่าคะแนนปัจจัย (Factor Score)

เมื่อหมุนแกนแล้วจะสามารถจัดตัวแปรหรือให้ความหมายแก่ปัจจัยร่วมแต่ละปัจจัยได้ โดยปัจจัยร่วมที่สร้างขึ้นใหม่ จะมีค่าคะแนนของแต่ละปัจจัยร่วม จากการคำนวณด้วยวิธีการที่ได้รับความนิยมสูงสุด (ยูทช, 2551; Johnson and Wichern, 2002; Rencher, 1995) 2 วิธี คือวิธีการถดถอย (Regression) และการใช้ค่าน้ำหนักปัจจัยเป็นฐาน (Factor Based Scores) ที่มีน้ำหนักปัจจัยน้อยกว่า 0.70 เพื่อลดจำนวนตัวแปร (Hatcher, 1998) ดังนี้

5.1 วิธีการถดถอย (Regression)

วิธีการถดถอยเหมาะสำหรับใช้ในการประมาณค่าการทำนาย เพราะค่าที่ได้จะมีความใกล้เคียงกับค่าจริงของตัวแปรเดิมมาก เนื่องจากเป็นผลรวมของค่าน้ำหนักปัจจัยร่วมกับค่าตัวแปรอิสระเดิมทุกตัวให้เป็นค่าคะแนนปัจจัยร่วมหนึ่งค่าที่สร้างขึ้นใหม่ เพื่อนำไปเป็นค่าเริ่มต้นของการวิเคราะห์ทางสถิติลำดับต่อไป หรือถ้านำปัจจัยร่วมที่สร้างขึ้นใหม่ไปวิเคราะห์บางปัจจัย ผลที่ได้ยังคงทำให้ตัวแบบการทำนายที่สร้างขึ้นไม่แตกต่างจากความจริงมาก จึงทำให้ปัจจัยร่วมที่สร้างขึ้นใหม่ไม่จำเป็นที่จะต้องใช้ทุกปัจจัยก็ได้ (ศิริชัย, 2543)

ดังนั้นการสร้างตัวแปรปัจจัยด้วยวิธี Regression จึงเป็นการสร้างค่าคะแนนจากผลบวกเชิงเส้นของเมตริกซ์สัมประสิทธิ์คะแนนปัจจัยร่วม (Component score coefficient matrix) กับตัวแปรเดิมที่ตัวที่ j ให้ข้อมูลอยู่ในรูปคะแนนมาตรฐานก่อนการคำนวณค่าคะแนนปัจจัย ดังสมการดังนี้

$$\begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_m \end{bmatrix} = \begin{bmatrix} \beta_{11}(X_1 - \bar{X}_1) + \beta_{12}(X_2 - \bar{X}_2) + \dots + \beta_{1p}(X_p - \bar{X}_p) \\ \beta_{21}(X_1 - \bar{X}_1) + \beta_{22}(X_2 - \bar{X}_2) + \dots + \beta_{2p}(X_p - \bar{X}_p) \\ \vdots \\ \beta_{m1}(X_1 - \bar{X}_1) + \beta_{m2}(X_2 - \bar{X}_2) + \dots + \beta_{mp}(X_p - \bar{X}_p) \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{bmatrix} \quad \dots\dots\dots(43)$$

หรือหาค่าน้ำหนักปัจจัยร่วมของแต่ละรายการ (Case) ได้จาก

$$F_i = \beta'(X_i - \bar{X}_i) + \varepsilon_i \quad ; i=1,2,\dots,n \quad \dots\dots\dots(44)$$

กำหนดให้

F_i คือคะแนนปัจจัยร่วมของ Case ที่ i ; $i=1,2,\dots,n$

X_i คือตัวแปรอิสระของ Case ที่ i ที่ปรับข้อมูลให้อยู่ในรูปคะแนนมาตรฐานแล้วหรือ

เท่ากับตัวแปร Z_i ; $i=1,2,\dots,n$

n = จำนวนรายการข้อมูล (Case)

β' = ฟังก์ชันของน้ำหนักปัจจัยร่วม โดยผลที่ได้จากวิเคราะห์ด้วยโปรแกรม SPSS จะให้ค่าข้อมูลในตารางที่ชื่อว่า Component score coefficient matrix ซึ่งเป็นค่าเมตริกซ์สัมประสิทธิ์คะแนนปัจจัยร่วมของแต่ละตัวแปรเดิม

5.2 การคำนวณคะแนนปัจจัยโดยใช้ค่าน้ำหนักปัจจัยเป็นฐาน (Factor Based Scores)

พิจารณาจากค่าเมตริกซ์น้ำหนักปัจจัยหลังการหมุนแกน (Rotated component matrix) หรือค่า Factor loading ที่จัดกลุ่มตัวแปรใหม่ให้อยู่ในแต่ละปัจจัยได้แล้ว จึงนำค่าข้อมูลตัวแปรทั้งหมดที่อยู่ในแต่ละปัจจัยรวมบวกรวมกัน ซึ่งเป็นการลดจำนวนตัวแปรให้น้อยลงได้ เนื่องจากกำหนดค่าน้ำหนักปัจจัยในแต่ละตัวแปรให้มีค่ามากกว่า 0.70 พิจารณาเฉพาะตัวแปรที่มีค่าความสัมพันธ์กันสูงเท่านั้น (Hatcher, 1998) ดังนั้นจำนวนตัวแปรเดิมที่อยู่ภายในแต่ละปัจจัยรวมจะมีจำนวนตัวแปรเดิมน้อยกว่าวิธี Regression จากการคำนวณด้วยสมการ

$$F_{ik} = \sum_{j=1}^p Z_{ij} \quad ;i=1,2,\dots,n ;k=1,2,\dots,m ;j=1,2,\dots,p$$

กำหนดให้

F_{ik} คือ คะแนนปัจจัยร่วมที่ k ของ case ที่ i ; $i=1,2,\dots,n ;k=1,2,\dots,m$

Z_{ij} คือ ตัวแปรอิสระ X_{ij} ที่ข้อมูลอยู่ในรูปคะแนนมาตรฐานแล้วมีค่าเฉลี่ยเป็นศูนย์ และค่าความแปรปรวนเป็น 1 โดยที่ j จะต้องเป็นตัวแปรที่ถูกจัดกลุ่มใหม่ให้อยู่ในปัจจัยร่วมเดียวกันที่ k ; $i=1,2,\dots,n ;j=1,2,\dots,p$

ตัวอย่าง เช่นรายการข้อมูล (Case) ที่ 1 ของปัจจัยร่วมที่ 1 ประกอบด้วยตัวแปรอิสระ 5 ตัวแปร คือ X_1, X_2, X_7, X_8, X_9 ปัจจัยร่วมที่ 2 ประกอบด้วยตัวแปรอิสระ 4 ตัวแปร คือ X_3, X_4, X_5, X_6 ดังนั้น จะคำนวณคะแนนปัจจัยโดยใช้ค่า Factor Based Scores ได้จาก

$$F_{11} = Z_{11} + Z_{12} + Z_{17} + Z_{18} + Z_{19}$$

$$F_{12} = Z_{13} + Z_{14} + Z_{15} + Z_{16}$$

เมื่อจัดกลุ่มตัวแปรและคำนวณค่าปัจจัยร่วมได้แล้วก็จะสามารถตั้งชื่อให้กับตัวแปรใหม่หรือตั้งชื่อให้กับปัจจัยที่สร้างขึ้น โดยให้ชื่อมีความหมายสอดคล้องกับกลุ่มตัวแปรเดิม เพื่อให้ง่ายต่อการนำไปใช้ในการอธิบายผลการศึกษา และเป็นข้อมูลตัวแปรอิสระในการวิเคราะห์ความถดถอยโลจิสติก

2.3 การวิเคราะห์การถดถอยโลจิสติกแบบทวิ (Binary Logistic Regression)

การวิเคราะห์การถดถอยโลจิสติก เป็นเทคนิคการวิเคราะห์การถดถอยที่ตัวแปรตาม (Y) เป็นตัวแปรผลลัพธ์ที่มีระดับการวัดของข้อมูลเป็นแบบนามบัญญัติ (Nominal) หรือเป็นแบบอันดับมาตรา (Ordinal) ในที่นี้จะกล่าวถึงการวิเคราะห์การถดถอยที่ตัวแปรตามมีค่าเป็นไปได้ 2 ค่า เท่านั้น คือ Y มีค่าเท่ากับ 1 เมื่อเกิดเหตุการณ์ที่สนใจ และ Y มีค่าเท่ากับ 0 เมื่อเกิดเหตุการณ์ที่ไม่สนใจ จึงเรียกการวิเคราะห์การถดถอยทวิ (Binary regression) ส่วนตัวแปรอิสระ (X) เป็นตัวแปรที่มีระดับการวัดของข้อมูลเป็นแบบใดก็ได้ คือเป็นได้ทั้งนามบัญญัติ (Nominal), อันดับมาตรา (Ordinal), อนุภาค (Interval) และอัตราส่วน (Ratio) โดยแบ่งออกเป็น 7 ขั้นตอน คือ

1. รูปแบบการวิเคราะห์การถดถอยโลจิสติกแบบทวิ (Binary Logistic Regression Analysis)

สามารถอธิบายหลักการพิจารณารูปแบบวิเคราะห์การถดถอยโลจิสติก ด้วยการวิเคราะห์การถดถอยเชิงเส้นอย่างง่าย (Simple linear regression analysis) จากการแสดงความสัมพันธ์ระหว่างตัวแปรตาม (Y) กับตัวแปรอิสระ (X) หนึ่งตัว คือ

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad ; i = 1, 2, \dots, n \quad \dots\dots\dots(45)$$

โดยที่ Y_i = ค่าผลลัพธ์ของตัวแปรตาม Y ในรายการที่ i มีได้ 2 ค่า คือ 1 และ 0

X_i = ค่าที่สังเกตได้ของตัวแปรอิสระ X ในรายการที่ i

β_0, β_1 = ค่าสัมประสิทธิ์การถดถอย (Regression coefficients)

ε_i = ค่าความคลาดเคลื่อนจากการพยากรณ์ค่าตัวแปรตาม Y ในรายการที่ i

n = จำนวนรายการ

ข้อสมมติของการวิเคราะห์การถดถอย คือ ค่าความคลาดเคลื่อน (ε_i) จะต้องมีการแจกแจงแบบปกติมีค่าเฉลี่ย (μ) เป็น 0 และความแปรปรวนเท่ากับ σ^2 หรือ $\varepsilon_i : \text{Normal}(0, \sigma^2) ; i = 1, 2, \dots, n$ นอกจากนี้ ε_i และ ε_j จะต้องมีการแจกแจงที่เป็นอิสระต่อกันด้วย คือ $\text{COV}(\varepsilon_i, \varepsilon_j) = 0 ; i \neq j$ และค่าคาดหวังของตัวแปรตามสามารถคำนวณได้จากสมการ

$$\begin{aligned} E(Y_i) &= E(\beta_0 + \beta_1 X_i + \varepsilon_i) \\ &= \beta_0 + \beta_1 X_i + E(\varepsilon_i) \\ &= \beta_0 + \beta_1 X_i \end{aligned} \quad \dots\dots\dots(46)$$

สำหรับกรณีที่ตัวแปรตาม (Y) มีค่าที่เป็นไปได้ 2 ค่า คือ 1 และ 0 ดังนั้นตัวแปรตาม (Y) จึงมีการแจกแจงแบบเบอร์นูลลี (Bernoulli distribution) ด้วยคือ

$$P(Y = Y_i) = P^{Y_i}(1 - P)^{1 - Y_i} \quad ; Y_i = 1, 0 \quad \dots\dots\dots(47)$$

ถ้า $Y_i = 1$ จะได้ $P(Y_i = 1) = P^1(1 - P)^{1 - 1} = P_i \quad \dots\dots\dots(48)$

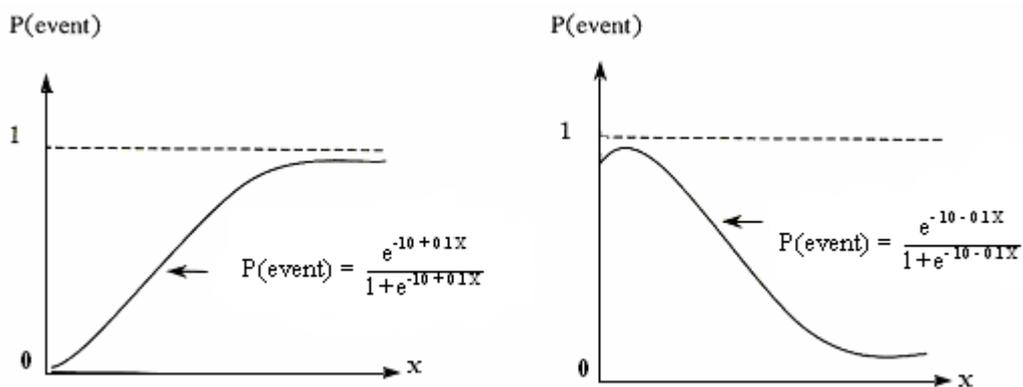
และ $Y_i = 0$ จะได้ $P(Y_i = 0) = P_i^0(1 - P_i)^{1 - 0} = 1 - P_i \quad \dots\dots\dots(49)$

เมื่อ P_i คือ ความน่าจะเป็นของการเกิดเหตุการณ์ที่สนใจ
 $1 - P_i$ คือ ความน่าจะเป็นของการเกิดเหตุการณ์ที่ไม่สนใจ

ดังนั้น จากสมการที่ (46) และสมการที่ (48) ได้ค่าคาดหวังของตัวแปรตามที่มีค่าได้ 2 ลักษณะ คือ

$$E(Y_i) = \beta_0 + \beta_1 X_i = P_i \quad \text{(Kutner et al., 2005)}$$

วัตถุประสงค์ของการวิเคราะห์การถดถอยโลจิสติกมีเพื่อศึกษาความสัมพันธ์หรืออิทธิพลของตัวแปรอิสระ (X_i) ที่มีผลต่อความน่าจะเป็นของการเกิดเหตุการณ์ที่สนใจ $P(Y_i = 1)$ โดยมีความเป็นไปได้ในช่วง 0 ถึง 1 เท่านั้น ดังนั้นค่าคาดหวังของตัวแปรตาม $E(Y_i)$ ก็คือความน่าจะเป็นของการเกิดเหตุการณ์ที่สนใจ ณ ที่ระดับของตัวแปรอิสระ (X_i) ทำให้ความสัมพันธ์ระหว่าง Y กับ X ไม่ได้อยู่ในรูปของสมการเชิงเส้น เมื่อนำไป Plot scatter diagram จะพบว่าฟังก์ชันแสดงความสัมพันธ์เป็นได้ทั้งบวกหรือค่าลบ ตามค่าสัมประสิทธิ์การถดถอย (β)



ภาพที่ 5 กราฟ Logistic เมื่อ $\beta_1 > 0$ และกราฟ Logistic เมื่อ $\beta_1 < 0$

จากภาพแสดงค่าความน่าจะเป็นของเหตุการณ์ที่สนใจเมื่อตัวแปรตามมีค่าเป็นได้ 2 ลักษณะ ซึ่งใช้การแปลงแบบโลจิท (Logit transformation) คำนวณหาความสัมพันธ์เชิงเส้นกับตัวแปรอิสระด้วยสมการ

$$\text{logit}(P_i) = \log_e \left(\frac{P_i}{1-P_i} \right)$$

ทำให้ได้ความสัมพันธ์ในรูปแบบเชิงเส้นกับตัวแปรอิสระ คือ

$$\log_e \left(\frac{P_i}{1-P_i} \right) = \beta_0 + \beta_1 X_{i1} + \dots + \beta_m X_{mi} \quad \dots\dots\dots(50)$$

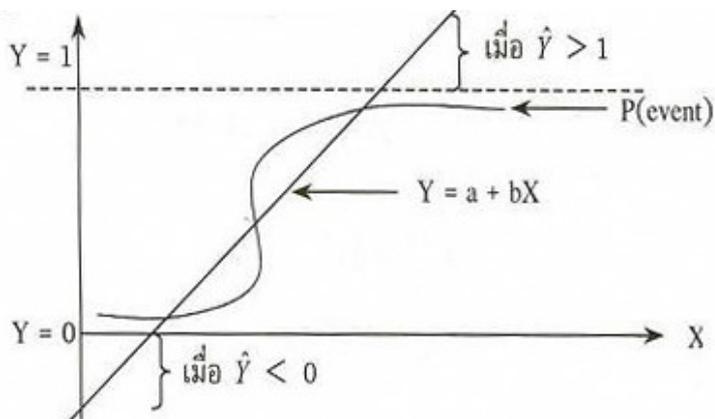
และเรียกอัตราส่วนของ $\frac{P}{1-P}$ ว่า “Odds” ซึ่ง หมายถึง ความเป็นต่อของการเกิดเหตุการณ์ที่สนใจเทียบกับการเกิดเหตุการณ์ที่ไม่สนใจ ดังนั้นสมการที่ (50) ก็คือ Logarithm ของ Odds ที่อยู่ในรูปแบบความสัมพันธ์เชิงเส้นกับตัวแปรอิสระ หรือเรียกว่าสมการตัวแบบการถดถอยโลจิสติกเชิงเส้น (Linear logistic model) แยกเป็น 2 เหตุการณ์ คือ

$$P_i = E(Y_i = 1) = \frac{e^{\beta_0 + \beta_1 X_{i1} + \dots + \beta_m X_{mi}}}{1 + e^{\beta_0 + \beta_1 X_{i1} + \dots + \beta_m X_{mi}}} = P \text{ (เกิดเหตุการณ์ที่สนใจ)} \quad \dots\dots(51)$$

$$\text{และ } 1-P_i = E(Y_i = 0) = 1 - \frac{e^{\beta_0 + \beta_1 X_{i1} + \dots + \beta_m X_{mi}}}{1 + e^{\beta_0 + \beta_1 X_{i1} + \dots + \beta_m X_{mi}}} = P \text{ (เกิดเหตุการณ์ที่ไม่ได้สนใจ)} \quad ..(52)$$

(พิมลรัตน์, 2547; Kutner *et al.*, 2005)

ค่าความสัมพันธ์ระหว่าง $E(Y)$ และ X จึงมีลักษณะเป็นเส้นโค้งแบบตัวอักษร S (S- Shaped or sigmoidal curve) มีขีดต่ำสุดและขีดสูงสุดอยู่ที่ 0 และ 1 ตามลำดับ โดยเส้นโค้งของฟังก์ชันการถดถอยโลจิสติกจะมีลักษณะสมมาตรที่ 0.5 ($P = 0.5$) โดยที่ส่วนปลายของทั้ง 2 ข้าง มีค่าคาดหวังของตัวแปรตามที่มีค่าจำกัดอยู่ในช่วง 0 และ 1 เท่านั้น ไม่ว่าตัวแปรอิสระจะเป็นค่าใด ๆ ก็ตาม ซึ่งสามารถอธิบายความสัมพันธ์ของเหตุการณ์ที่เกิดขึ้นได้ตามภาพที่ 6



ภาพที่ 6 แสดงการเปรียบเทียบฟังก์ชันการถดถอย ระหว่างรูปแบบเส้นโค้งการถดถอยโลจิสติก กับรูปแบบเส้นตรงของตัวแบบการถดถอยอย่างง่าย

ที่มา: <http://www.thebuggyprofessor.org/archives/00000254.php>

ดังนั้นจึงแสดงให้เห็นว่ารูปแบบการถดถอยโลจิสติกเป็นรูปแบบการถดถอยที่ไม่ใช่เชิงเส้นตรง (Nonlinear regression model) แต่สามารถปรับให้อยู่ในรูปแบบการถดถอยเชิงเส้นได้จากการแปลงค่าด้วยการประมาณค่าพารามิเตอร์ในรูปแบบโลจิท (Logit transformation) เพื่อคำนวณค่าสัมประสิทธิ์การถดถอย (β) ของตัวแบบโลจิสติก แทนได้

2. การประมาณค่าสัมประสิทธิ์การถดถอยของตัวแบบโลจิสติก

โดยทั่วไปการวิเคราะห์การถดถอยเชิงเส้นแบบพหุ ที่ตัวแปรตาม (X) มีมากกว่า 1 ตัวแปร และข้อมูลมีสเกลการวัดเป็นแบบต่อเนื่อง (Continuous data) นิยมใช้การประมาณค่าสัมประสิทธิ์ความถดถอยด้วยวิธีกำลังสองน้อยที่สุด (Ordinary Least Squares, OLS) (Hosmer and Lemeshow, 2000) เพื่อต้องการให้ค่าความแตกต่างระหว่างตัวแปรค่าสังเกต Y กับค่าพยากรณ์ที่คำนวณได้ \hat{Y} มีค่าผลรวมกำลังสองของความคลาดเคลื่อนแตกต่างกันน้อยที่สุด ดังนั้นวิธี OLS จึงไม่เหมาะที่จะนำมาใช้ในการประมาณค่าสัมประสิทธิ์ความถดถอยโลจิสติกที่มีค่าเป็น 2 ลักษณะ เนื่องจากสมการถดถอยโลจิสติกแบบทวิมีการแจกแจงแบบเบอร์นูลลี ดังนั้นค่าความคลาดเคลื่อน (ε) จึงมีการแจกแจงแบบเบอร์นูลลีด้วย (Neter *et al.*, 1989)

$$Y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_m X_{im} + \varepsilon_i \quad ; i = 1, 2, \dots, n$$

โดยที่ Y_i = ค่าผลลัพธ์ของตัวแปรตาม มีได้ 2 ค่า คือ 1 และ 0; $i = 0, 1$
 ε_i = ค่าคลาดเคลื่อนจากการพยากรณ์ตัวแปรตาม Y_i ; $i = 1, 2, \dots, n$
 $\beta_0, \beta_1, \dots, \beta_m$ = ค่าสัมประสิทธิ์การถดถอย (Regression coefficients)
 X_i = ค่าที่สังเกตได้ของตัวแปรอิสระที่ i ; $i = 1, 2, \dots, m$
 n = จำนวนรายการ
 m = จำนวนค่าสังเกตของตัวแปรอิสระ (X)

$$\text{ถ้าให้ } Y_i = 1 \text{ จะทำให้ } \varepsilon_i = 1 - \beta_0 - \beta_1 X_i \quad \dots\dots\dots(53)$$

$$\text{และ } Y_i = 0 \text{ จะทำให้ } \varepsilon_i = -\beta_0 - \beta_1 X_i \quad \dots\dots\dots(54)$$

จากสมการที่ (53) และ (54) ทำให้สรุปได้ว่าค่าคลาดเคลื่อนมีการแจกแจงไม่เป็นแบบปกติ (Non-normal error terms) จึงไม่เป็นไปตามข้อตกลงของการใช้เทคนิควิธีการวิเคราะห์การถดถอยเชิงเส้นแบบพหุ ที่ว่า $\varepsilon_i : \text{Normal}(0, \sigma^2)$

ดังนั้น ในการประมาณค่าสัมประสิทธิ์ความถดถอยโลจิสติก ที่มีค่าเป็น 2 ลักษณะ จึงเลือกวิธีการประมาณค่าพารามิเตอร์ 2 วิธี คือวิธีภาวะน่าจะเป็นสูงสุด (ML) และวิธีกำลังสองน้อยที่สุดแบบถ่วงน้ำหนัก (WLS) โดยการศึกษาครั้งนี้เลือกใช้วิธี ML เนื่องจากเป็นการประมาณค่าภาวะน่าจะเป็นสูงสุดแบบคำนวณซ้ำ (Iterative maximum likelihood estimator) ทำได้ด้วยโปรแกรมวิเคราะห์สำเร็จรูป SPSS และ SAS ทำให้ค่าที่ได้มีความน่าจะเป็นสูงสุดสำหรับค่าสังเกต (X) ในการประมาณค่าสัมประสิทธิ์ความถดถอยโลจิสติก และเหมาะสำหรับการวิเคราะห์ข้อมูลที่มีหลายตัวแปร

วิธีภาวะน่าจะเป็นสูงสุด (Maximum likelihood, ML) เป็นการประมาณค่า β_0 และ β_j ; $j = 0, 1, 2, \dots, k$ โดยพิจารณาค่าสังเกตตัวแปรตาม (Y) ของแต่ละรายการข้อมูล ที่มีค่าได้อย่างใดอย่างหนึ่ง คือ เป็น 0 (เหตุการณ์ที่ไม่สนใจ) หรือเป็น 1 (เหตุการณ์ที่สนใจ) เท่านั้น ในการเริ่มต้นเพื่อสร้างฟังก์ชันภาวะน่าจะเป็น (Likelihood function, $L(\beta)$) จากสมการภาวะน่าจะเป็น (Probability density function) คือ

$$f_i(Y_i) = P_i^{Y_i} (1 - P_i)^{1 - Y_i} \quad \dots\dots\dots(55)$$

โดย $Y_i = 0, 1$; $i = 1, 2, \dots, n$

และ P_i = ความน่าจะเป็นของการเกิดเหตุการณ์ที่สนใจในรายการข้อมูลที่ i

ถ้าให้ Y_i เป็นอิสระต่อกัน จะได้ฟังก์ชันความน่าจะเป็นร่วม (Joint probability density function) คือ

$$g(Y_1, Y_2, \dots, Y_n) = \prod_{i=1}^n f_i(Y_i) \quad \dots\dots\dots(56)$$

$$= \prod_{i=1}^n P_i^{Y_i} (1 - P_i)^{1 - Y_i} \quad \dots\dots\dots(57)$$

และเรียกสมการที่ (57) ว่าฟังก์ชันภาวะน่าจะเป็น (Likelihood function, $L(\beta)$) พิมลรัตน์ (2547)

หลักการหาตัวประมาณค่าพารามิเตอร์ด้วยวิธี ML ใช้วิธีการหาอนุพันธ์ของ $L(\beta)$ เทียบกับ

พารามิเตอร์ที่ต้องการประมาณ โดยใช้ Log-likelihood function หรือ $\log_e L(\beta)$ แทน $L(\beta)$

$$\text{ฟังก์ชันการถดถอยโลจิสติก คือ } P_i = E(Y_i = 1) = \frac{e^{(\beta_0 + \beta_1 X_i)}}{1 + e^{(\beta_0 + \beta_1 X_i)}}$$

เมื่อ $\beta'X_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik}$ แทนในสมการที่ (58) จะได้ Log-likelihood function ของตัวแบบโลจิสติก คือ

$$\log_e L(\beta) = - \sum_{i=1}^n Y_i (\beta'X_i) - \sum_{i=1}^n \log_e (1 + e^{\beta'X_i}) \quad \dots\dots\dots(58)$$

(พิมลรัตน์, 2547; Kutner *et al.*, 2005)

พิมลรัตน์ (2547) หลักของวิธี ML คือการประมาณค่า β_0 และ β_1 เพื่อให้สมการ Log-likelihood function มีค่ามากที่สุด โดยการหาอนุพันธ์ลำดับที่ 1 ของสมการ Log-likelihood function เทียบกับ

พารามิเตอร์ β_0 และ β_j ; $j = 0, 1, 2, \dots, k$ ได้สมการภาวะน่าจะเป็น (Likelihood equations)

ทั้งหมด $k+1$ สมการ ดังนี้

$$\frac{\partial \{\log_e L(\beta)\}}{\partial \beta_0} = \sum_{i=1}^n Y_i - \sum_{i=1}^n \left(\frac{e^{\beta'X_i}}{1 + e^{\beta'X_i}} \right) \quad \dots\dots\dots(59)$$

$$\frac{\partial \{\log_e L(\beta)\}}{\partial \beta_j} = \sum_{i=1}^n Y_i X_{ji} - \sum_{i=1}^n \left(\frac{X_{ji} e^{\beta'X_i}}{1 + e^{\beta'X_i}} \right) \quad \dots\dots\dots(60)$$

และนำสมการที่ (59) และ (60) เทียบกับ β_0 และ β_j แล้วให้เท่ากับ 0 เพื่อแก้สมการหาตัวประมาณ ภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Estimators, MLE) คือ $\hat{\beta}_0$ และ $\hat{\beta}_j$ จึงต้องทำการหาอนุพันธ์ ถ้าตัวอย่างมีขนาดใหญ่เข้าใกล้อนันต์ ตัวประมาณภาวะน่าจะเป็นสูงสุด ก็จะมีการแจกแจงแบบปกติตามหลักการของ Asymptotic distribution

ในบางกรณีที่ไม่สามารถประมาณค่า β_0 และ β_j ด้วยวิธี MLE ได้โดยตรง เนื่องจากสมการปกติไม่ได้อยู่ในรูปเชิงเส้น จึงใช้วิธีการของนิวตัน-ราฟสัน (Newton-Raphson method) พบว่าเป็นวิธีที่มีประสิทธิภาพ เป็นวิธีการแก้สมการไม่เชิงเส้น (Nonlinear equations) โดยเทคนิคการทำซ้ำ (Iteration techniques) ที่ใช้โปรแกรมสำรूपทางสถิติ ในการหาค่าประมาณ $\hat{\beta}_0$ และ $\hat{\beta}_j$ ได้ตามลำดับ

3. วิธีคัดเลือกตัวแบบอัตโนมัติ (Automatic Model Selection Methods)

การคัดเลือกตัวแปรสำหรับตัวแบบการถดถอยโลจิสติก เป็นกระบวนการเลือกตัวแปรอิสระที่มีความสามารถในการพยากรณ์ค่าความน่าจะเป็นที่เกิดเหตุการณ์ที่สนใจได้ โดยมีเทคนิคการคัดเลือกดังนี้

3.1 Enter

เป็นเทคนิคการเลือกตัวแปรอิสระทั้งหมด (p ตัว) เข้าในตัวแบบการถดถอยโลจิสติกขั้นตอนเดียว โดยผู้วิจัยจะเป็นผู้ตัดสินใจเองว่าตัวแปรอิสระตัวใดที่มีความสัมพันธ์กับตัวแปรตาม หรือเหมาะสมที่จะอยู่ในตัวแบบการถดถอยโลจิสติก โดยพิจารณาจากค่าสถิติทดสอบ หรือ ค่า Significance ของสถิติทดสอบ

3.2 Forward stepwise : Wald หรือ Likelihood ratio

เป็นเทคนิคการคัดเลือกตัวแปรอิสระเข้าตัวแบบการถดถอยโลจิสติก โดยการคัดเลือกตัวแปรอิสระเข้าตัวแบบนั้นอาศัยวิธีการถดถอยซ้ำ ๆ กัน ซึ่งสมการที่สร้างขึ้นจะมีการเพิ่มตัวแปรอิสระเข้าในตัวแบบทีละ 1 ตัว ในแต่ละขั้น พร้อมทั้งพิจารณาว่าควรตัดตัวแปรอิสระตัวใดออกจากตัวแบบหรือไม่ ด้วยสถิติ Wald หรือ Likelihood ratio statistic

3.3 Backward stepwise : Wald หรือ Likelihood ratio

เป็นเทคนิคการคัดเลือกตัวแปรอิสระเข้าตัวแบบการถดถอยโลจิสติก ที่ตรงข้ามกับเทคนิค Forward stepwise โดยเทคนิค Backward stepwise จะมีตัวแปรอิสระทั้งหมด (p ตัว) อยู่ในตัวแบบการถดถอยโลจิสติกตั้งแต่เริ่มแรก จากนั้นจึงคัดเลือกตัวแปรอิสระออกจากตัวแบบทีละ 1 ตัว ในแต่ละขั้น พร้อมทั้งพิจารณาว่าควรตัดตัวแปรอิสระตัวใดออกจากตัวแบบหรือไม่ ด้วยสถิติ Wald หรือ Likelihood ratio statistic

(Garson, 2009)

4. การแปลความหมายจากค่าสัมประสิทธิ์การถดถอยของตัวแบบโลจิสติก

จากการประมาณค่าสัมประสิทธิ์การถดถอยโลจิสติกด้วยวิธีภาวะน่าจะเป็นสูงสุด (MLE) เพื่อสร้างตัวแบบโลจิสติก ที่ตัวแปรอิสระอยู่ในรูปผลรวมเชิงเส้น สามารถเขียนในรูปแบบได้ คือ

$$\text{logit}(P_i) = \beta_0 + \beta_1 X_{i1} + \dots + \beta_m X_{mi}$$

ซึ่ง $\text{logit}(P_i)$ ก็คือการแปลงตัวแบบโลจิท ให้อยู่ในรูป Logarithm ของ Odds ได้ดังนี้

$$\text{logit}(P_i) = \log_e \left(\frac{P_i}{1 - P_i} \right)$$

ดังนั้นการอธิบายความหมายของค่าสัมประสิทธิ์การถดถอยของตัวแบบโลจิสติก ($\hat{\beta}$) สามารถแบ่งตามลักษณะของข้อมูลตัวแปรอิสระได้ 2 กรณี คือ กรณีที่ตัวแปรอิสระเป็นตัวแปรแบ่งกลุ่ม และกรณีที่ตัวแปรอิสระเป็นตัวแปรต่อเนื่อง

4.1 กรณีที่ตัวแปรอิสระเป็นตัวแปรแบ่งกลุ่ม (Categorical variable)

ตัวอย่าง การศึกษาตัวแปรผลการเกิดฝนของวันก่อนหน้า จะมีผลต่อการเกิดฝนของวันนี้ (ประจำวัน) หรือไม่ โดยแบ่งผลการเกิดฝนเป็น 2 กลุ่ม คือ วันก่อนหน้าฝนตก และวันก่อนหน้าฝนไม่ตก

จากตัวอย่าง ตัวแปรตาม (Y) มีค่าเป็น 2 ลักษณะ คือ วันนี้ฝนตก (Y=1) และวันนี้ฝนไม่ตก (Y=0)
ตัวแปรอิสระ (X) มีค่าเป็น 2 ลักษณะ คือ วันก่อนหน้าฝนตก และวันก่อนหน้าฝนไม่ตก ดังนั้น

ถ้า วันก่อนหน้าฝนไม่ตก จะทำให้ X มีค่าเป็น 0

และถ้าวันก่อนหน้าฝนตกจะทำให้ X มีค่าเป็น 1

P_1 คือ ความน่าจะเป็นของการเกิดฝนตกภายในวันนี้ ถ้าวันก่อนหน้าฝนไม่ตก จากสมการ

$$\text{logit}(P_1) = \log_e \left(\frac{P_1}{1-P_1} \right) = \beta_0 + \beta_1 X_1 = \beta_0 + \beta_1(0) = \beta_0 \quad \dots\dots\dots(61)$$

ดังนั้น $\left(\frac{P_1}{1-P_1} \right) = e^{\beta_0}$ คือค่า Odds ที่วันก่อนหน้าฝนไม่ตก มีค่าเท่ากับ e^{β_0}

ให้ P_2 คือความน่าจะเป็นของการเกิดฝนตกภายในวันนี้ ถ้าวันก่อนหน้าฝนตก จากสมการ

$$\text{logit}(P_2) = \log_e \left(\frac{P_2}{1-P_2} \right) = \beta_0 + \beta_1 X_1 = \beta_0 + \beta_1(1) = \beta_0 + \beta_1 \quad \dots\dots\dots(62)$$

ดังนั้น $\left(\frac{P_2}{1-P_2} \right) = e^{\beta_0 + \beta_1}$ คือค่า Odds ที่วันก่อนหน้าฝนตก มีค่าเท่ากับ $e^{\beta_0 + \beta_1}$

นำสมการที่ (62) เทียบกับสมการที่ (61) จะได้

$$\frac{P_2 / (1-P_2)}{P_1 / (1-P_1)} = \frac{e^{\beta_0 + \beta_1}}{e^{\beta_0}} = e^{\beta_1} \quad \dots\dots\dots(63)$$

จะได้ค่า Odds ของวันก่อนหน้าฝนตก ($e^{\beta_0 + \beta_1}$) เทียบกับค่า Odds ของวันก่อนหน้าฝนไม่ตก (e^{β_0})
จะได้ Odds ratio (OR) เป็น e^{β_1} หรือ $\beta_1 = \log_e(\text{OR})$ สามารถแปลความหมายจากค่าสัมประสิทธิ์
การถดถอยของตัวแบบโลจิสติก ในกรณีนี้ X เป็นตัวแปรแบบทวิ ได้ว่า “ $\hat{\beta}_1$ เป็นค่าประมาณของ
log ของ OR หรือ ค่า OR ที่ได้จะมีค่าเท่ากับ $e^{\hat{\beta}_1}$ ” (พิมลรัตน์, 2547)

ตัวอย่าง ถ้า $\hat{\beta}_1 = 1.596$ แล้วจะทำให้ค่า $\text{OR} = e^{1.596} = 4.933$ แปลได้ว่าถ้าวันก่อนหน้าฝนตก
จะมีโอกาสที่วันนี้ฝนจะตกด้วยเป็น 4.933 เท่า เมื่อเทียบกับวันก่อนหน้าที่ฝนไม่ตก

4.2 กรณีที่ตัวแปรอิสระเป็นตัวแปรต่อเนื่อง (Continuous variable)

สำหรับกรณีที่ตัวแปรอิสระเป็นตัวแปรต่อเนื่อง การเปรียบเทียบค่าสัมประสิทธิ์การถดถอย (β) ที่มีต่างกันเพียง 1 หน่วย จะไม่ให้ความสำคัญมากนัก คือ จะไม่เทียบโอกาสที่ฝนจะตกเมื่อค่าความชื้นเพิ่มขึ้นเพียง 1 หน่วย เช่น จะไม่สนใจถ้าค่าความชื้นเพิ่มขึ้นจาก 250 เป็น 251 แต่จะสนใจเมื่อตัวแปรดังกล่าวมีค่าเพิ่มขึ้นมากกว่า 1 หน่วย เช่น ค่าความชื้นเพิ่มจาก 250 เป็น 300 จะเขียนแทนให้อยู่ในรูปทั่วไปได้ คือ $X + m$ หน่วย

$$OR = \frac{e^{\beta_0 + \beta_1(X+m)}}{e^{\beta_0 + \beta_1 X}} = e^{m\beta_1} \quad \dots\dots\dots(64)$$

แปลความหมายของค่าสัมประสิทธิ์การถดถอยได้ว่า “ถ้าตัวแปรอิสระเปลี่ยนค่าไปทุก ๆ m หน่วยแล้วค่าโอกาสที่จะเกิดเหตุการณ์ที่สนใจ เทียบกับโอกาสที่จะไม่เกิดเหตุการณ์ที่สนใจ $P/(1-P)$ มีค่าเพิ่มขึ้นเป็น $e^{m\beta_1}$ เท่า”

(พิมลรัตน์, 2547)

5. การทดสอบสมมติฐานเกี่ยวกับค่าสัมประสิทธิ์การถดถอยโลจิสติก (β_m)

เมื่อมีตัวแปรอิสระ m ตัว (X_1, X_2, \dots, X_m) จะทำการอนุมานทางสถิติเกี่ยวกับ ค่าสัมประสิทธิ์การถดถอยโลจิสติก (β 's) ด้วยการทดสอบสมมติฐานทางสถิติ โดยจัดการทดสอบออกเป็น 2 กรณี คือ ทดสอบตัวแปรอิสระทุกตัวพร้อมกัน และทดสอบตัวแปรอิสระทีละตัว

5.1 การทดสอบนัยสำคัญของตัวแปรอิสระทุกตัวพร้อมกัน

สมมติฐานการทดสอบ คือ

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_{m-1} = 0 \text{ (Intercept Model)}$$

$$H_1 : \beta_i \neq 0 \text{ อย่างน้อย 1 ตัว ; } i = 1, 2, \dots, m - 1$$

ค่าสถิติทดสอบนัยสำคัญของตัวแปรอิสระทุกตัวพร้อมกัน คือ

$$\begin{aligned} G^2 &= -2 \log_e \left(\frac{L_{\text{reduced}}}{L_{\text{Full}}} \right) \\ &= -2 [\log_e(L_{\text{reduced}}) - \log_e(L_{\text{Full}})] \end{aligned}$$

โดยที่ G^2 มีการแจกแจงแบบ χ^2 ($df = m - 1$) ถ้า $G^2 \leq \chi^2(1-\alpha; m - 1)$ จะยอมรับสมมติฐานหลัก
 ถ้า $G^2 > \chi^2(1-\alpha; m - 1)$ จะปฏิเสธสมมติฐานหลัก

กำหนดให้

L_{reduced} คือ ค่าฟังก์ชัน likelihood ในตัวแบบที่ไม่มีตัวแปรอิสระที่ต้องการทดสอบ
 หรือมีเฉพาะค่าคงที่

L_{Full} คือ ค่าฟังก์ชัน likelihood ในตัวแบบที่มีตัวแปรอิสระที่ต้องการทดสอบ
 ที่ i (X_1, X_2, \dots, X_p)

$$\text{โดยที่} \quad \log_e L_{\text{reduced}} = n_0 \left[\log_e \left(\frac{n_0}{n} \right) \right] + n_1 \left[\log_e \left(\frac{n_1}{n} \right) \right]$$

เมื่อ n คือ จำนวนตัวอย่างทั้งหมด

n_0 คือ จำนวนตัวอย่างที่ไม่เกิดเหตุการณ์ที่สนใจ

n_1 คือ จำนวนตัวอย่างที่เกิดเหตุการณ์ที่สนใจ

ค่าที่ได้จากการคำนวณสถิติ G^2 (Likelihood ratio test) มีการแจกแจงโดยการประมาณแบบไคลสแควร์ ท้องาคความเป็นอิสระเท่ากับจำนวนของสัมประสิทธิ์ที่ทำการทดสอบในสมมติฐานหลัก จะคล้ายกับการทดสอบโดยใช้สถิติ F-test ในการวิเคราะห์การถดถอยเชิงเส้นทั่วไป

5.2 การทดสอบนัยสำคัญของตัวแปรอิสระทีละตัว

สมมติฐานการทดสอบ คือ

$$H_0 : \beta_i = 0$$

$$H_1 : \beta_i \neq 0 \quad ; i=1,2,\dots,m$$

สถิติที่ใช้ทดสอบสัมประสิทธิ์การถดถอยโลจิสติก คือ สถิติทดสอบวอลด์ (Wald statistic)

$$\text{Wald} = \left[\frac{\hat{\beta}_i}{\text{SE}(\hat{\beta}_i)} \right]^2 \quad ; i=1,2,\dots,m$$

(พิมลรัตน์, 2547; Kutner *et al.*, 2005)

โดยที่ Wald มีการแจกแจงแบบ χ^2 ($df = 1$) ถ้า $W \leq \chi^2(1-\alpha/2)$ จะยอมรับสมมติฐานหลัก
 ถ้า $W > \chi^2(1-\alpha/2)$ จะปฏิเสธสมมติฐานหลัก

6. การทดสอบความเหมาะสมของตัวแบบการวิเคราะห์ถดถอยโลจิสติก (Goodness of fit test)

6.1 ทดสอบ Hosmer and Lemeshow

เป็นสถิติทดสอบความเหมาะสมของตัวแบบ Logistic regression โดยการจัดรายการข้อมูล (case) ออกเป็น k กลุ่มเท่า ๆ กัน ตามค่าของ Predict probability ของ $Y = 1$ ที่ได้จากตัวแบบ Logistic regression โดย case ที่คาดไว้ (E_i) จะต้องมากกว่า 5 และไม่มีกลุ่มใดที่มีค่าน้อยกว่า 1 ($E_i < 1$) แล้ววิเคราะห์ความเหมาะสมของตัวแบบ ด้วยสถิติทดสอบไคสแควร์

สมมติฐานการทดสอบ คือ

H_0 : ตัวแบบการวิเคราะห์การถดถอยมีความเหมาะสม

H_1 : ตัวแบบการวิเคราะห์การถดถอยไม่เหมาะสม

สถิติทดสอบไคสแควร์ (χ^2)

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} ; df = k-2, 5 \leq k \leq 10$$

(Kutner, 2005)

โดยที่ O_i = ความถี่ที่ได้จากข้อมูล

E_i = ความถี่ที่คาดว่าจะเกิดขึ้นเมื่อสมมติฐาน H_0 เป็นจริง

6.2 การวัดอำนาจของการพยากรณ์ (Measuring Predictive Power)

จากตัวแปรอิสระที่ผ่านการคัดเลือกสามารถบอกอำนาจของการพยากรณ์ หรือบอกเปอร์เซ็นต์ความสามารถในการอธิบายตัวแบบ Logistic regression ด้วยค่า Cox-Snell R^2 และ Nagelkerke's R^2 ได้ดังนี้

ก. Cox- Snell R^2

$$\text{Cox- Snell } R^2 = 1 - \exp\left[\frac{-(G^2)}{n}\right]$$

หรือ

$$\text{Cox- Snell } R^2 = 1 - \left[\frac{L_{\text{Intercept}}}{L_{\text{Full}}}\right]^{2/n}$$

กำหนดให้

n คือ จำนวนตัวอย่างทั้งหมด

$L_{\text{Intercept}}$ คือ ค่าฟังก์ชัน likelihood ในตัวแบบที่มีเฉพาะค่าคงที่ (Intercept Model)

L_{full} คือ ค่าฟังก์ชัน likelihood ในตัวแบบที่มีตัวแปรอิสระที่ต้องการทดสอบ (Full Model)

ข. Nagelkerke's R^2 (R_N^2)

$$R_N^2 = \frac{\text{Cox-Snell } R^2}{1 - (L_{\text{Intercept}})^{2/n}}$$

ค่าที่ R_N^2 จะมากกว่า $R_{\text{cox-snell}}^2$ เสมอ และจะอยู่ระหว่าง 0 ถึง 1 แต่มักจะมีค่าต่ำกว่า R^2 ของเทคนิคการวิเคราะห์การถดถอยเชิงเส้นทั่วไป โดยถ้าตัวแบบที่พัฒนาขึ้นมีค่า R_N^2 เข้าใกล้ 1 จะเป็นตัวแบบที่ดีมาก

(Williams, n.d.)

7. ค่าความถูกต้องของการจำแนก (Accuracy of Classification)

เป็นการตรวจสอบความแม่นยำในการพยากรณ์ของตัวแบบในการพยากรณ์การเกิดฝนรายวัน เทียบกับเหตุการณ์จริงของรายงานการเกิดฝนรายวัน คิดเป็นค่าร้อยละ เรียกว่า Percentage of correct classification หรือ Hit rate ซึ่งคำนวณได้จากสูตร

$$\text{Hit rate} = \frac{\sum_{i=1}^n y_i}{n} \times 100$$

(Garson, 2009)

โดย $\sum_{i=1}^n y_i$ คือ จำนวนวันที่ตัวแบบพยากรณ์ได้ถูกต้องตรงกับเหตุการณ์จริงของการรายงานผล
การเกิดฝนประจำวัน
n คือ จำนวนวันทั้งหมดที่ใช้ในการวิเคราะห์ข้อมูล

อุปกรณ์และวิธีการ

อุปกรณ์

การวิเคราะห์และประมวลผลข้อมูลใช้โปรแกรม Microsoft Excel โปรแกรม SAS และโปรแกรม SPSS โดยเครื่องคอมพิวเตอร์ของภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตบางเขน

วิธีการ

การวิเคราะห์ปัจจัยสำหรับตัวแบบคาดการณ์โอกาสการเกิดฝน ในการปฏิบัติการฝนหลวงของภาคตะวันออกเฉียงเหนือ แบ่งเป็น 2 ส่วน คือ

ส่วนที่ 1 ศึกษาลักษณะทั่วไปของข้อมูล GPCM ซึ่งประกอบด้วย

- 1.1 แหล่งข้อมูลและการรวบรวมข้อมูล
- 1.2 ตรวจสอบความถูกต้องและความสมบูรณ์ของข้อมูล
 - 1.2.1 ศึกษาแนวโน้มของข้อมูล
 - 1.2.2 รูปแบบความสัมพันธ์ของข้อมูล

ส่วนที่ 2 วิธีการวิเคราะห์ทางสถิติ

- 2.1 การวิเคราะห์ปัจจัย
- 2.2 การสร้างตัวแบบการถดถอยโลจิสติกแบบทวิ

ส่วนที่ 1 ศึกษาลักษณะทั่วไปของข้อมูล GPCM

1.1 แหล่งข้อมูลและการรวบรวมข้อมูล

การศึกษานี้ได้รับความอนุเคราะห์ข้อมูลจากศูนย์ปฏิบัติการฝนหลวงเฉลิมพระเกียรติ อำเภอหัวหิน จังหวัดประจวบคีรีขันธ์ และกรมอุตุนิยมวิทยาสำนักงานเขตบางนา ซึ่งข้อมูลที่ได้มีระยะเวลาแตกต่างกัน คือข้อมูล GPCM ได้จากการปฏิบัติการฝนหลวง ๑ ระหว่างวันที่ 1 เมษายน พ.ศ.2547 ถึงวันที่ 14 กันยายน พ.ศ.2549 ส่วนข้อมูลปริมาณน้ำฝน 24 ชั่วโมง (รายวัน) ของกรมอุตุนิยมวิทยาสำนักงานเขตบางนา ได้ข้อมูลระหว่างวันที่ 1 มกราคม พ.ศ.2545 ถึงวันที่ 30 กันยายน พ.ศ.2549

จากข้อมูลทั้ง 2 หน่วยงาน นำข้อมูลทั้งสองส่วนมาเชื่อมต่อกันโดยใช้วันที่ (วัน/เดือน/ปี) รายงานเป็นตัวเชื่อม ทำให้ได้ข้อมูลทั้งหมด 918 วัน (ราชการ) ระหว่างวันที่ 11 มีนาคม พ.ศ. 2547 ถึงวันที่ 14 กันยายน พ.ศ.2549 มีจำนวนตัวแปรทั้งสิ้น 49 ตัวแปร ประกอบด้วยข้อมูล GPCM มี 47 ตัวแปร เป็นตัวแปรเชิงปริมาณ 45 ตัวแปร และตัวแปรเชิงกลุ่ม 2 ตัวแปร และข้อมูลปริมาณน้ำฝน 24 ชั่วโมง ที่เป็นตัวแปรเชิงปริมาณ แต่ในการศึกษาครั้งนี้จะมีได้ใช้ค่าข้อมูลปริมาณน้ำฝน 24 ชั่วโมงสร้างตัวแบบโดยตรง เนื่องจากวัตถุประสงค์ของการศึกษา คือต้องการทราบโอกาสการเกิดฝนที่มีค่าเป็น 2 ลักษณะ คือ มีฝนตก หรือไม่มีฝนตก ในวันนั้น ๆ ดังนั้นผู้ศึกษาจึงใช้ข้อมูลปริมาณน้ำฝน 24 ชั่วโมง สร้างตัวแปรใหม่ 2 ตัวแปร คือตัวแปรเหตุการณ์การเกิดฝนของวันก่อนหน้าหรือเมื่อวานนี้ (ฝนตก/ฝนไม่ตก) กับตัวแปรเหตุการณ์การเกิดฝนของวันนี้ (ฝนตก/ฝนไม่ตก) เพื่อใช้ประกอบการศึกษาครั้งนี้

โดยกำหนดให้ตัวแปรตาม (Dependent variable, Y) คือตัวแปรเหตุการณ์การเกิดฝนตกในวันนี้ (ประจำวัน) ที่มีค่าเป็น 2 ลักษณะ คือฝนตกในวันนี้ ($Y_i=1$) หรือฝนไม่ตกในวันนี้ ($Y_i=0$) ส่วนตัวแปรอิสระที่สามารถสังเกตและวัดค่าได้ (Observed variables) มีทั้งหมด 48 ตัวแปร ประกอบด้วยตัวแปรเชิงปริมาณ 45 ตัวแปร และตัวแปรเชิงคุณภาพ 3 ตัวแปร ดังแสดงในตารางที่ 1

1.2 ตรวจสอบความถูกต้องและความสมบูรณ์ของข้อมูล

เมื่อตรวจสอบความครบถ้วนของข้อมูลทั้ง 918 วัน พบว่าบางวันมีข้อมูลบางตัวแปรสูญหายไป ฉะนั้นจึงเหลือจำนวนวันที่มีตัวแปรครบทั้ง 49 ตัวแปร เพียง 308 วัน และเนื่องจากการศึกษาครั้งนี้สนใจศึกษาตัวแปรเหตุการณ์การเกิดฝนของวันก่อนหน้า (D_t) ร่วมด้วย เช่น ถ้าข้อมูลการเกิดฝนของวันนี้ (วันที่ 12 มีนาคม พ.ศ.2547) จะนำไปวิเคราะห์ได้ก็ต่อเมื่อจะต้องมีข้อมูลการเกิดฝนของวันก่อนหน้า (วันที่ 11 มีนาคม พ.ศ.2547) ดังแสดงไว้ในตารางที่ 5

ตารางที่ 5 ระยะเวลาข้อมูล จำนวนตัวแปร และจำนวนรายการข้อมูล (วัน) ของข้อมูลต่าง ๆ

ข้อมูล	เริ่ม - สิ้นสุด	ตัวแปร	รายการข้อมูลทั้งหมด	รายการข้อมูลที่วิเคราะห์ได้
1 GPCM	1 เม.ย.43 - 14 ก.ย.49	47	2,359	875
2 กรมอุตุนิยมวิทยา	1 ม.ค.45 - 30 ก.ย.49	2	1,734	1,734
3 GPCM + กรมอุตุนิยมวิทยา	11 มี.ค.47 - 14 ก.ย.49	49	918	308
4 GPCM + กรมอุตุนิยมวิทยา ที่มีข้อมูลของ 2 วันติดต่อกัน	12 มี.ค.47 - 14 ก.ย.49	49	917	277*

หมายเหตุ (*) ชุดข้อมูลที่จะใช้ในการวิเคราะห์ต่อไป คือข้อมูลที่ 4

สรุปได้ว่าจากข้อมูล GPCM และข้อมูลกรมอุตุนิยมวิทยา ที่เชื่อมต่อกันด้วยวันที่รายงานข้อมูลพร้อมทั้งมีข้อมูลของวันก่อนหน้าซึ่งเป็นวันที่ติดกันทำให้จัดข้อมูลที่จะใช้วิเคราะห์ได้เพียง 277 วัน (รายการ) ที่มีตัวแปรครบทั้ง 49 ตัวแปร เริ่มต้นตั้งแต่วันที่ 12 มีนาคม พ.ศ. 2547 ถึง 14 กันยายน พ.ศ. 2549

จากการตรวจสอบข้อมูลตัวแปรอิสระที่เป็นข้อมูลเชิงปริมาณมีทั้งหมด 45 ตัวแปร พบว่ามี 18 ตัวแปร ที่มีข้อมูลขาดความสมบูรณ์ เนื่องจากมีข้อมูลสูญหาย ในบางรายการ (วัน) ดังนี้

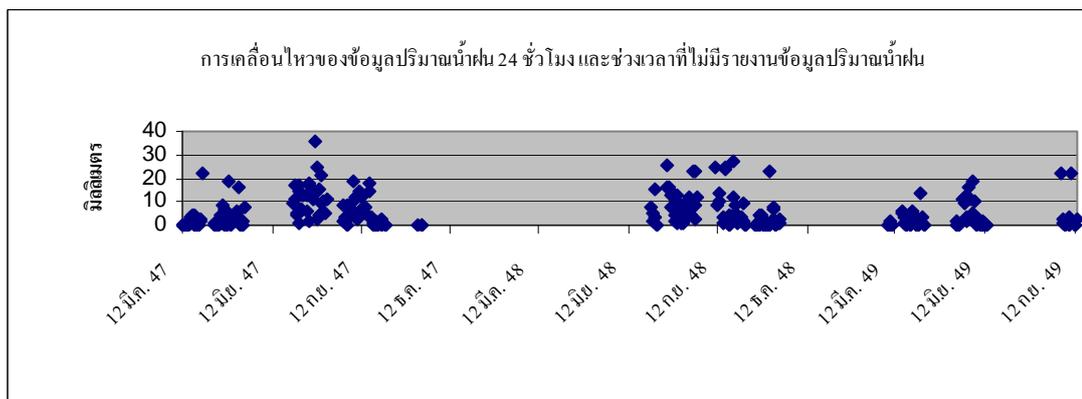
ตารางที่ 6 จำนวนรายการ (วัน) ที่มีข้อมูลสูญหาย ในแต่ละตัวแปร

ชื่อตัวแปร	จำนวนข้อมูลที่สูญหาย (จาก 277รายการ)
1. SFC_TEMP_RISE_REQUIRED_FOR_CCL	2
2. MEAN_WIND_5000_10000_FT_DEG	3
3. MEAN_WIND_5000_10000_FT_KTS	3
4. MEAN_WIND_1000_5000_FT_DEG	4
5. MEAN_WIND_1000_5000_FT_KTS	4
6. CLOUD_BASE_HEIGHT_LCL	5
7. SWEAT_INDEX	7
8. MEAN_WIND_10000_15000_FT_DEG	7
9. MEAN_WIND_10000_15000_FT_KTS	7
10. MEAN_RH_1000_5000_FT	8
11. MEAN_RH_5000_10000_FT	8
12. MEAN_RH_10000_15000_FT	8
13. MEAN_RH_20000_25000_FT	8
14. MEAN_WIND_20000_25000_FT_DEG	12
15. MEAN_WIND_20000_25000_FT_KTS	12
16. LEVEL_OF_FREE_CONVECTION_LFC_wrtTma	107
17. LEVEL_OF_NEUTRAL_BUOYANCY_LNB	111
18. TOP_OF_LATENT_INSTABILITY_LAYER_wrtTma	157

ดังนั้น จึงใช้การแทนค่าข้อมูลสูญหายด้วยค่าเฉลี่ยรายเดือนของแต่ละตัวแปร เพื่อให้มีข้อมูลครบถ้วน และยังคงมีจำนวนตัวอย่างที่ศึกษามากพอ

1.2.1 ศึกษาแนวโน้มของข้อมูล

จากภาพที่ 7 เห็นได้ว่าข้อมูลปริมาณน้ำฝน 24 ชั่วโมง ไม่อาจอธิบายได้ชัดเจนถึงอิทธิของแนวโน้ม (Trend) วัฏจักร (Cyclical) ฤดูกาล (Seasonal) และความผันแปรที่ไม่แน่นอน (Irregular) เนื่องจากใน 277 รายการ มีค่าข้อมูลปริมาณน้ำฝน 24 ชั่วโมงครบถ้วนที่ของข้อมูลไม่ได้มีวันที่ติดกัน คือมีวันเว้นช่วง เป็นวันที่ติดต่อกันบ้างไม่ติดต่อกันบ้าง จึงไม่สามารถศึกษาลักษณะของข้อมูลปริมาณน้ำฝน 24 ชั่วโมง ที่เปลี่ยนแปลงไปตามลำดับของเวลาได้ ดังนั้นจึงใช้การวิเคราะห์แบบอนุกรมเวลาไม่ได้



ภาพที่ 7 แสดงรูปแบบการเคลื่อนไหว (แนวโน้ม) ของข้อมูลปริมาณน้ำฝน 24 ชั่วโมง ระหว่างวันที่ 12 มีนาคม พ.ศ. 2547 ถึงวันที่ 14 กันยายน พ.ศ. 2549

1.2.2 รูปแบบความสัมพันธ์ของข้อมูล

การตรวจสอบความสัมพันธ์เชิงเส้นระหว่างตัวแปรเชิงปริมาณสองตัวใด ๆ โดยเฉพาะความสัมพันธ์เชิงเส้น ระหว่างตัวแปรผลการตรวจอากาศชั้นบนที่เป็นตัวแปรปริมาณ กับตัวแปรปริมาณน้ำฝน 24 ชั่วโมง สามารถพิจารณาจากแผนภาพการกระจาย (Scatter Plot) ซึ่งจะแสดงขนาดและทิศทางความสัมพันธ์เชิงเส้น ด้วยค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สัน (Pearson correlation coefficient, r) ที่มีค่าระหว่าง -1 ถึง 1 การทดสอบสมมติฐานเกี่ยวกับค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรเชิงปริมาณ 2 ตัว สามารถทำได้ดังนี้

กำหนดสมมติฐานของการทดสอบ

$$H_0 : \rho = 0 \text{ (ตัวแปร X และ Y ไม่มีความสัมพันธ์กัน)}$$

$$H_1 : \rho \neq 0 \text{ (ตัวแปร X และ Y มีความสัมพันธ์กัน)}$$

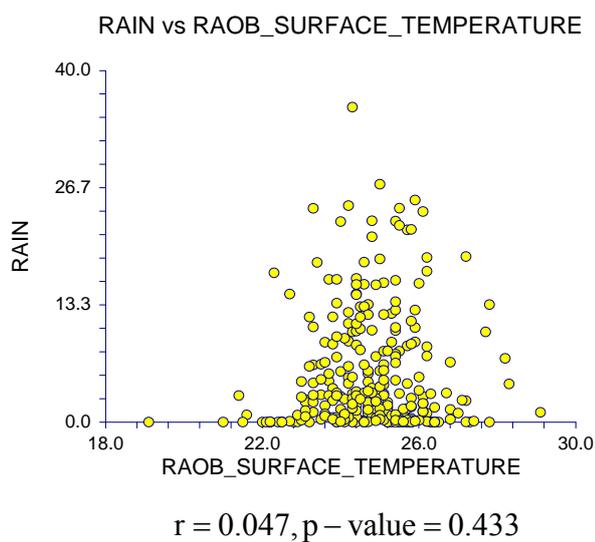
สถิติทดสอบ เป็นการทดสอบแบบสองทาง

$$t = \frac{r_{XY}\sqrt{n-2}}{\sqrt{1-r_{XY}^2}}, \text{ df} = n-2$$

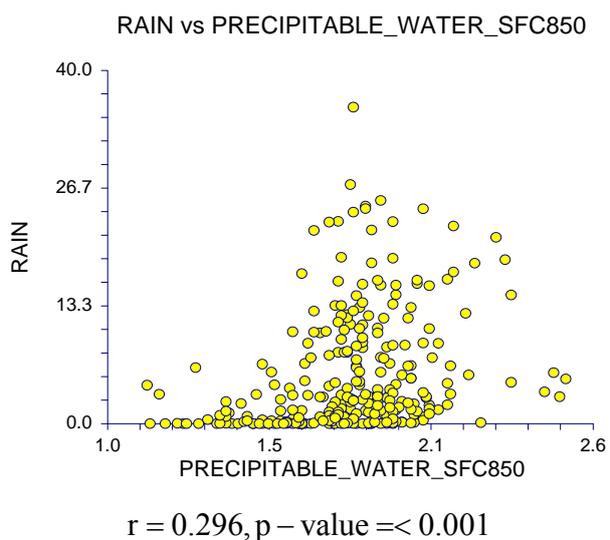
เมื่อ ρ คือ พารามิเตอร์ของสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปร X และ Y ของข้อมูลประชากร

r_{XY} คือ ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปร X และ Y ของข้อมูลตัวอย่าง

ขอบเขตการปฏิเสธสมมติฐาน H_0 : จะปฏิเสธ H_0 เมื่อค่า $|t|$ ที่คำนวณได้มีค่ามากกว่า หรือเท่ากับค่า $t_{\alpha/2, n-2}$ หรือ $|t| \geq t_{\alpha/2, n-2}$ ที่เปิดจากตาราง (Tabachnick and Fidell, 2007)

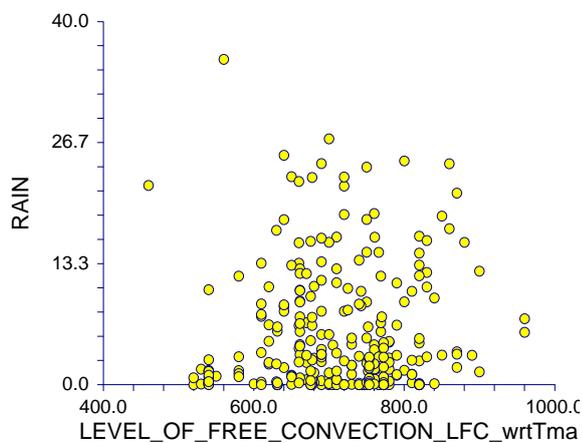


ภาพที่ 8 แผนภาพการกระจาย ระหว่างตัวแปร RAOB_SURFACE_TEMPERATURE กับตัวแปรปริมาณน้ำฝน 24 ชม.



ภาพที่ 9 แผนภาพการกระจาย ระหว่างตัวแปร PRECIPITABLE_WATER_SFC850 กับตัวแปรปริมาณน้ำฝน 24 ชม.

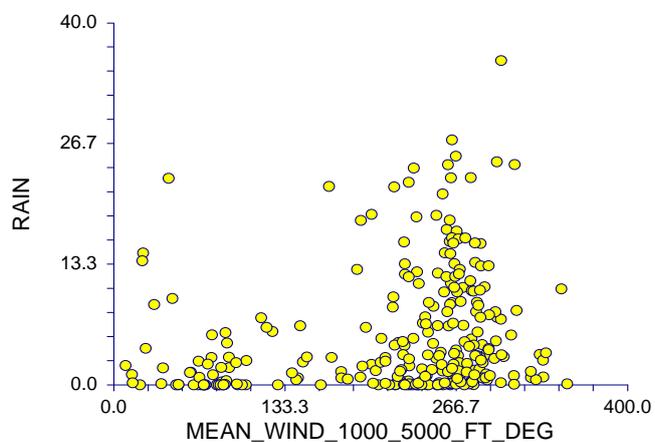
RAIN vs LEVEL_OF_FREE_CONVECTION_LFC_w



$$r = 0.045, p\text{-value} = 0.455$$

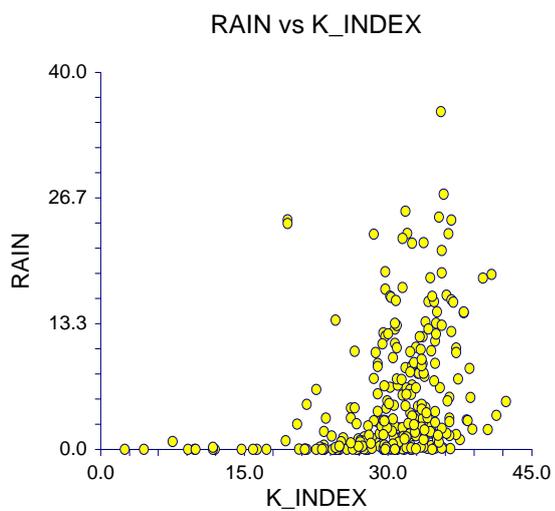
ภาพที่ 10 แผนภาพการกระจาย ระหว่างตัวแปร LEVEL_OF_FREE_CONVECTION_LFC_wrtTma กับตัวแปรปริมาณน้ำฝน 24 ชม.

RAIN vs MEAN_WIND_1000_5000_FT_DEG



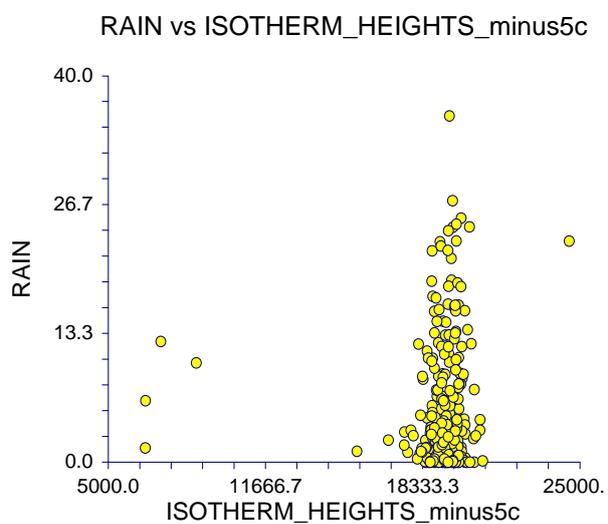
$$r = 0.223, p\text{-value} = < 0.001$$

ภาพที่ 11 แผนภาพการกระจาย ระหว่างตัวแปร MEAN_WIND_1000_5000_FT_DEG กับตัวแปรปริมาณน้ำฝน 24 ชม.



$$r = 0.313, p\text{-value} \leq 0.001$$

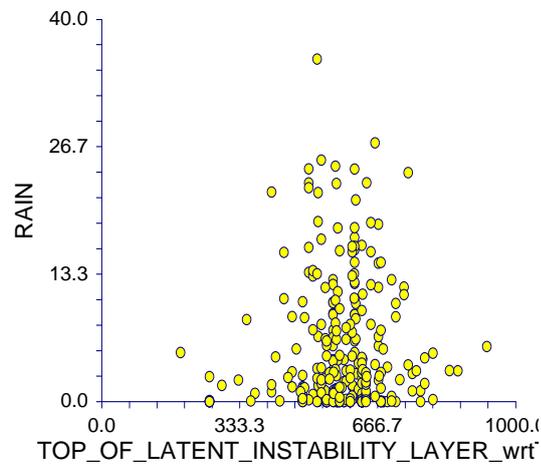
ภาพที่ 12 แผนภาพการกระจาย ระหว่างตัวแปร K_INDEX
กับตัวแปรปริมาณน้ำฝน 24 ชม.



$$r = 0.044, p\text{-value} = 0.461$$

ภาพที่ 13 แผนภาพการกระจาย ระหว่างตัวแปร ISOTHERM_HEIGHTS_minus5c
กับตัวแปรปริมาณน้ำฝน 24 ชม.

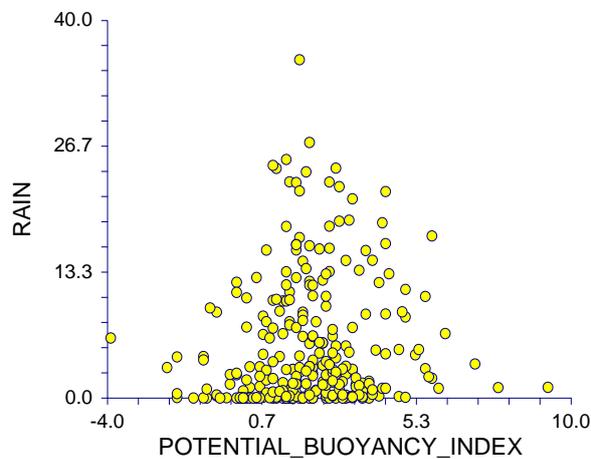
N vs TOP_OF_LATENT_INSTABILITY_LAYER



$$r = 0.033, p - \text{value} = 0.587$$

ภาพที่ 14 แผนภาพการกระจาย ระหว่างตัวแปร TOP_OF_LATENT_INSTABILITY_LAYER_wrt กับตัวแปรปริมาณน้ำฝน 24 ชม.

RAIN vs POTENTIAL_BUOYANCY_INDEX



$$r = 0.088, p - \text{value} = 0.145$$

ภาพที่ 15 แผนภาพการกระจาย ระหว่างตัวแปร POTENTIAL_BUOYANCY_INDEX กับตัวแปรปริมาณน้ำฝน 24 ชม.

จากภาพที่ 8 - ภาพที่ 15 พบว่าค่า r ของตัวแปรตรวจอากาศชั้นบนบางตัวมีความสัมพันธ์เชิงเส้น อย่างมีนัยสำคัญทางสถิติที่ 0.05 กับข้อมูลปริมาณน้ำฝน 24 ชั่วโมง ($P\text{-value} \leq 0.05$) โดยที่ระดับความสัมพันธ์เชิงเส้นอยู่ในระดับต่ำ

ส่วนที่ 2 วิธีการวิเคราะห์ทางสถิติ

จากข้อมูลตัวแปรอิสระ 48 ตัวแปร ที่เป็นตัวแปรเชิงปริมาณ 45 ตัวแปร (X_1, X_2, \dots, X_{45}) จึงมีโอกาที่ตัวแปรเหล่านี้จะเกิดปัญหา Multicollinearity ขึ้น ดังนั้นวิธีการสถิติที่จะใช้วิเคราะห์ข้อมูลที่มีความสัมพันธ์กันเองได้ ก็การใช้เทคนิคการวิเคราะห์ปัจจัย (Factor analysis) ทำการจับกลุ่มหรือรวมตัวแปรอิสระที่มีความสัมพันธ์กันไว้ด้วยกัน แล้วจึงสร้างเป็นตัวแปรใหม่หรือปัจจัยร่วม (F_1, F_2, \dots, F_m) โดยปัจจัยร่วมที่สร้างขึ้นจะไม่มีความสัมพันธ์ระหว่างกัน แล้วจึงนำปัจจัยร่วม F_1, F_2, \dots, F_m ไปใช้เป็นตัวแปรอิสระ เพื่อสร้างตัวแบบการถดถอยโลจิสติก (Logistic regression analysis) ในลำดับต่อไป

2.1 วิธีการวิเคราะห์ปัจจัย

แบ่งการวิเคราะห์ปัจจัยออกเป็น 5 ขั้นตอน ดังนี้

2.1.1 การตรวจสอบความเหมาะสมของข้อมูลตัวแปรอิสระที่จะใช้วิธีการวิเคราะห์ปัจจัย

1) ค่า Kaiser Meger Olkin (KMO)

การคำนวณค่า KMO เพื่อวัดความเหมาะสมของข้อมูลตัวแปรอิสระที่จะนำมาวิเคราะห์ปัจจัย (factor analysis) โดยที่ $0 \leq KMO \leq 1$ ซึ่งแปลผลได้ว่า

- ถ้าค่า KMO มีค่าน้อยเข้าใกล้ 0 แสดงว่าข้อมูลที่มีอยู่ไม่เหมาะสมกับการวิเคราะห์ปัจจัย
- ถ้าค่า KMO มีค่ามากเข้าใกล้ 1 แสดงว่าข้อมูลที่มีอยู่เหมาะสมกับการวิเคราะห์ปัจจัย

เกณฑ์การวิเคราะห์ปัจจัยทั่วไป ถ้ามีค่า KMO มากกว่า 0.50 จึงจะถือว่ามีความเหมาะสมในการวิเคราะห์ปัจจัยได้ (Hari and Black, 1998)

2) สถิติทดสอบ Barlett's Test of sphericity

สมมติฐานการทดสอบ คือ

$$H_0 : \text{Correlation matrix} = \text{Identity matrix}$$

$$H_1 : \text{Correlation matrix} \neq \text{Identity matrix}$$

การทดสอบแบบ Bartlett's Sphericity มีการแจกแจงโดยวิธีการประมาณค่าแบบไคกำลังสอง (χ^2) ซึ่งเป็นฟังก์ชันของค่า ดีเทอร์มิแนนต์ ของเมตริกซ์ R

$$\chi^2 = -\left[(n-1) - \frac{(2p+5)}{6}\right] \ln|R|$$

$$\text{องศาความเป็นอิสระ (df)} = \frac{1}{2}[p(p-1)]$$

- ถ้ายอมรับ H_0 แสดงว่าตัวแปรอิสระ $X_i; i=1,2,\dots,p$ ไม่มีความสัมพันธ์กัน จึงไม่ควรใช้การวิเคราะห์ปัจจัยสำหรับวิเคราะห์ข้อมูลตัวแปรดังกล่าว

- ถ้าปฏิเสธ H_0 แสดงว่าตัวแปรอิสระ $X_i; i=1,2,\dots,p$ มีความสัมพันธ์กัน จึงเหมาะสมในการใช้เทคนิคการวิเคราะห์ปัจจัย (Rencher, 1995)

2.1.2 การสกัดปัจจัยหรือหาปัจจัยเริ่มต้น (Factor extraction หรือ Initial factors)

การสกัดปัจจัย โดยการหาค่าความสัมพันธ์จากการวิเคราะห์ความแปรปรวนร่วมของตัวแปรต่าง ๆ ไปไว้ในปัจจัยร่วม เพื่อให้ปัจจัยร่วมเป็นตัวแทนของตัวแปรเดิม โดยขั้นตอนนี้เลือกวิธีการสกัดปัจจัย 2 วิธี คือ การเทคนิคตัวประกอบหลักปัจจัย (Principal Component Method : PCM) และเทคนิคภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Method : MLM)

2.1.3 การกำหนดปัจจัยร่วมที่เหมาะสม

จากขั้นตอนในการสกัดปัจจัย ทำให้จำนวนตัวแปรเดิมลดลง และให้ตัวแปรปัจจัยร่วมเป็นตัวแทนของตัวแปรเดิม โดยที่จำนวนปัจจัยร่วมที่เหมาะสมจะต้องมีจำนวนน้อยกว่าจำนวนตัวแปรเดิมที่นำมาวิเคราะห์ และสามารถอธิบายการรวมกลุ่มของตัวแปรเดิมได้อย่างดี ซึ่งในการศึกษาครั้งนี้พิจารณาจากค่าเฉพาะปัจจัยร่วมที่มีค่าไอเกนมากกว่า 1 ร่วมกับการใช้พล็อตกราฟ Scree plot

2.1.4 การหมุนแกนปัจจัย (Factor Rotation)

ในกรณีที่ค่าน้ำหนักปัจจัย (Factor loading) มีค่ากลาง ๆ ไม่ชัดเจน ทำให้ไม่สามารถตัดสินใจได้ว่าควรจัดตัวแปรให้อยู่ในปัจจัยใด จึงต้องใช้เทคนิคการหมุนแกนด้วยวิธีแวนริแมกซ์ (Varimax) ในการทำให้ค่าน้ำหนักปัจจัยของแต่ละตัวแปร มีค่ามากขึ้นหรือลดลง แล้วจึงสามารถจัดตัวแปรนั้น ๆ ได้ว่าควรจะอยู่ในปัจจัยใด โดยวิธี Varimax เป็นการหมุนแกนแบบตั้งฉาก (Orthogonal rotation) จึงทำให้ปัจจัยร่วมยังคงเป็นอิสระต่อกันอยู่เช่นเดิม ซึ่งในขั้นตอนนี้ผู้ศึกษาเลือกใช้โปรแกรมสำเร็จรูปทางสถิติ SPSS ในการหมุนแกนให้ได้ค่าน้ำหนักที่เหมาะสมในการจัดกลุ่มปัจจัยให้ชัดเจนขึ้น

2.1.5 การคำนวณค่าปัจจัยร่วม หรือหาค่าคะแนนปัจจัยร่วม (Factor score) มี 2 วิธี คือ

วิธีที่ 1 ใช้การประมาณค่าคะแนนปัจจัย ด้วยวิธี Regression จาก

$$F_i = \beta'(X_i - \bar{X}_i) + \varepsilon_i$$

วิธีที่ 2 ใช้ค่าคะแนนพื้นฐานของตัวแปรที่อยู่ภายในปัจจัยเดียวกันเท่านั้น (Factor Based Score) คือ

$$F_{ik} = Z_{1k} + Z_{2k} + Z_{3k} + \dots + Z_{pk}$$

เมื่อจัดกลุ่มตัวแปรและคำนวณค่าคะแนนปัจจัยใหม่ได้แล้ว ต่อไปจะเป็นการตั้งชื่อให้กับปัจจัยที่สร้างขึ้น โดยให้ชื่อมีความหมายสอดคล้องกับกลุ่มตัวแปรเดิม

ในการศึกษาครั้งนี้จะเป็นการเปรียบเทียบผลการศึกษา แยกเป็น 8 กรณี และแสดงในแผนภาพที่ 16 คือ

กรณีที่ 1 วิเคราะห์ปัจจัยจากตัวแปรปริมาณทั้งหมด 45 ตัวแปร ด้วยวิธี PCM และคำนวณค่าคะแนนปัจจัยด้วยวิธี Regression

กรณีที่ 2 วิเคราะห์ปัจจัยจากตัวแปรปริมาณทั้งหมด 45 ตัวแปร ด้วยวิธี PCM และคำนวณค่าคะแนนปัจจัยด้วยวิธี Factor Based Score

กรณีที่ 3 วิเคราะห์ปัจจัยจากตัวแปรปริมาณทั้งหมด 45 ตัวแปร ด้วยวิธี MLM และคำนวณค่าคะแนนปัจจัยด้วยวิธี Regression

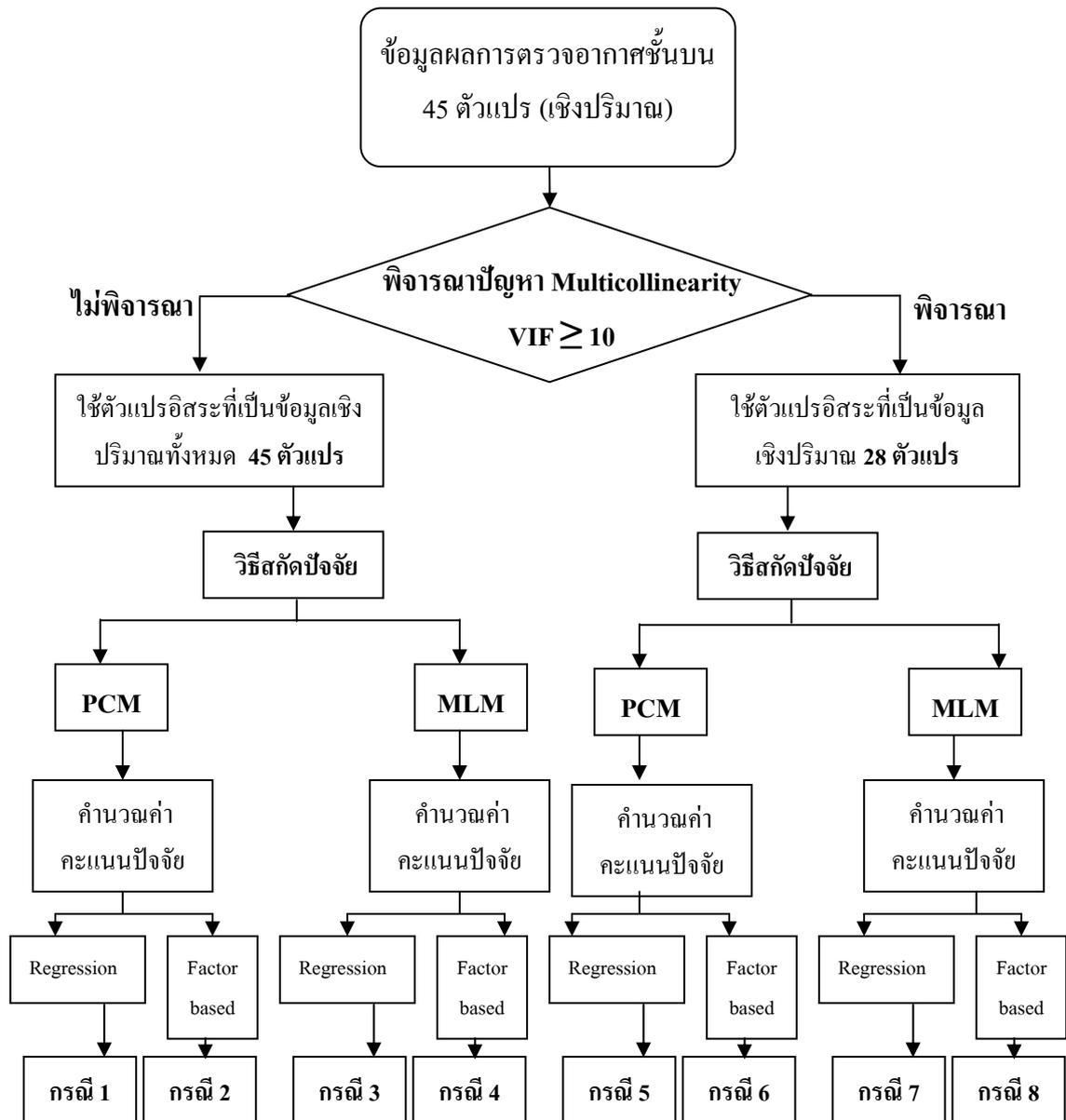
กรณีที่ 4 วิเคราะห์ปัจจัยจากตัวแปรปริมาณทั้งหมด 45 ตัวแปร ด้วยวิธี MLM และคำนวณค่าคะแนนปัจจัยด้วยวิธี Factor Based Score

กรณีที่ 5 วิเคราะห์ปัจจัยจากตัวแปรปริมาณ 28 ตัวแปร ที่มีปัญหา Multicollinearity ($VIF \geq 10$) ด้วยวิธี PCM และคำนวณค่าคะแนนปัจจัยด้วยวิธี Regression

กรณีที่ 6 วิเคราะห์ปัจจัยจากตัวแปรปริมาณ 28 ตัวแปร ที่มีปัญหา Multicollinearity ($VIF \geq 10$) ด้วยวิธี PCM และคำนวณค่าคะแนนปัจจัยด้วยวิธี Factor Based Score

กรณีที่ 7 วิเคราะห์ปัจจัยจากตัวแปรปริมาณ 28 ตัวแปร ที่มีปัญหา Multicollinearity ($VIF \geq 10$) ด้วยวิธี MLM และคำนวณค่าคะแนนปัจจัยด้วยวิธี Regression

กรณีที่ 8 วิเคราะห์ปัจจัยจากตัวแปรปริมาณ 28 ตัวแปร ที่มีปัญหา Multicollinearity ($VIF \geq 10$) ด้วยวิธี MLM และคำนวณค่าคะแนนปัจจัยด้วยวิธี Factor Based Score



ภาพที่ 16 แผนผังแสดงจำนวนตัวแปรอิสระเชิงปริมาณ ที่ใช้สำหรับการวิเคราะห์ปัจจัย 8 กรณี

เพื่อให้ง่ายต่อการนำไปใช้ในการอธิบายผลการศึกษา และนำผลคะแนนปัจจัยที่ได้จาก
ทั้ง 8 กรณี ไปเป็นตัวแปรอิสระสำหรับการสร้างตัวแบบวิเคราะห์การถดถอยโลจิสติกแบบทวิต่อไป

2.2 การสร้างตัวแบบการถดถอยโลจิสติกแบบทวิ

ในการศึกษาความสัมพันธ์ของเหตุการณ์การเกิดฝน (Y_i ; $i = 0, 1$) ที่มีค่าได้ 2 ค่า คือ ฝนไม่ตก (0) และฝนตก (1) ตัวแปรอิสระของการวิเคราะห์การถดถอยโลจิสติกของการศึกษานี้จะเป็นตัวแปรอิสระที่เกิดจากการวิเคราะห์ปัจจัยร่วม (F_1, F_2, \dots, F_m) ทำให้มีจำนวนปัจจัยร่วมน้อยกว่าจำนวนตัวแปรอิสระเดิม (X_1, X_2, \dots, X_{45}) โดยปัจจัยร่วมจะไม่มีความสัมพันธ์ระหว่างกันสามารถนำไปพิจารณาพร้อมกับตัวแปรเชิงคุณภาพที่เหลืออีก 3 ตัวแปร คือ CONVECTIVE_TEMP_REACHABLE (D_1), โอกาสความสำเร็จในการทำฝนจากเมฆอุ่น (Warm Cloud Seeding Potential (D_2)) และตัวแปรการเกิดฝนวันก่อนหน้า YESTERDAY_RAIN (D_3) เพื่อสร้างตัวแบบคาดการณ์โอกาสการเกิดฝน ตามขั้นตอนการวิเคราะห์การถดถอยโลจิสติกแบบทวิได้

ขั้นตอนในการวิเคราะห์โลจิสติกมีดังต่อไปนี้

2.2.1 การคัดเลือกตัวแปรอิสระ

นำผลการวิเคราะห์ที่ได้จากการสร้างตัวแปรปัจจัย ร่วมกับตัวแปรเชิงคุณภาพอีก 3 ตัวแปร เพื่อพยากรณ์โอกาสในแต่ละรายการ (วัน) จะอยู่ในกลุ่มฝนตก หรือกลุ่มฝนไม่ตก ด้วยวิธีการเพิ่มตัวแปรอิสระแบบขั้นตอน (Stepwise) ซึ่งเป็นวิธีการเลือกตัวแปรอิสระเข้าในตัวแบบการถดถอย โดยการเลือกตัวแปรอิสระเข้าตัวแบบครั้งละหนึ่งตัวแปร ตัวแปรใดที่เข้าอยู่ในตัวแบบการถดถอยแล้วอาจถูกตัดออกไปได้ภายหลัง หมายความว่า จะต้องทำการทดสอบตัวแปรที่อยู่ในตัวแบบการถดถอยว่าตัวแปรที่เข้ามาอยู่ในตัวแบบมีส่วนช่วยในการอธิบายความผันแปรของตัวแปรตาม ในขณะที่ตัวแปรอิสระตัวอื่น ๆ อยู่ในตัวแบบการถดถอยหรือไม่ ในขั้นตอนนี้ผู้ศึกษาคำนวณด้วยโปรแกรมสำเร็จรูป SPSS โดยเลือกวิธีการ (Forward stepwise : Likelihood ratio) กำหนดค่า Probability of stepwise entry (P_E) = 0.15 และ Probability of stepwise removal (P_R) = 0.20 ซึ่งเป็นค่าที่ Hosmer and Lemshow (2000) แนะนำให้ใช้ในการคัดเลือกตัวแปรอิสระเข้า และออก ตามลำดับ

2.2.2 การประมาณค่าพารามิเตอร์ $\beta_0, \beta_1, \dots, \beta_m$

การประมาณค่าสัมประสิทธิ์ความถดถอยโลจิสติก ($\beta_0, \beta_1, \dots, \beta_m$) โดยพิจารณา ค่าตัวแปรตาม (Y) ของแต่ละรายการข้อมูลที่มีค่าเป็น 1 ถ้าฝนตก และมีค่าเป็น 0 ถ้าฝนไม่ตก ด้วยวิธีประมาณค่าแบบภาวะน่าจะเป็นสูงสุด (Maximum Likelihood, ML) เป็นการกำหนดค่าเริ่มต้นของ $\beta_0', \beta_1', \dots, \beta_m'$ เป็นค่าใด ๆ และจะให้ค่า $\beta_0', \beta_1', \dots, \beta_m'$ เปลี่ยนแปลงไปทั้งค่าและทิศทางที่จะทำให้ฟังก์ชันภาวะน่าจะเป็นของตัวแบบที่มีตัวแปรอิสระ m ตัว (Log likelihood function, -2LL) ใช้สำหรับเปรียบเทียบความคลาดเคลื่อนกรณีที่ไม่มี และมีตัวแปรอิสระอยู่ในตัวแบบ สังเกตได้ว่าค่าความแตกต่างจะมีค่าลดลงเรื่อย ๆ เมื่อมีตัวแปรอิสระถูกใส่เพิ่มเข้าไปทีละตัว ค่า Chi-square (χ^2) ที่ถูกคำนวณโดยวิธีการเปรียบเทียบค่าคลาดเคลื่อนของตัวแบบจะมีค่า $df = m - n$ โดยที่ m คือจำนวนตัวแปรอิสระทั้งหมด และ n คือ จำนวนตัวประมาณค่าพารามิเตอร์ในสมการ ค่าสถิติทดสอบ -2LL หรือ Deviance วัดความเหมาะสมของสมการโลจิสติก ถ้า -2LL มีค่าลดลงมากที่สุด ตัวแบบโลจิสติกก็จะเหมาะสมมากที่สุด โดยทำซ้ำ ๆ เช่นนี้ไปหลาย ๆ รอบจนกระทั่งค่า -2LL ไม่เพิ่มขึ้นหรือลดลงอีก

2.2.3 การตรวจสอบความเหมาะสม (Goodness of fit test) ของสมการถดถอยโลจิสติก

1) ตรวจสอบตัวแปรอิสระทุกตัวพร้อมกัน

สมมติฐานการทดสอบ คือ

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_m = 0$$

$$H_1 : \beta_i \neq 0 \text{ อย่างน้อย 1 ค่า ; } i=1, 2, \dots, m$$

โดยใช้สถิติ Model Chi-Square ซึ่งเป็นผลต่างของ -2LL₀ กรณีที่มีเฉพาะค่าคงที่อยู่ในตัวแบบกรณี -2LL₁ มีตัวแปรอิสระที่อยู่ในตัวแบบ ผลต่างจะมีการแจกแจงแบบไคกำลังสอง และเป็นค่าไคกำลังสองของค่าคลาดเคลื่อน ดังนั้นถ้าตัวแปรอิสระที่คาดไว้มีผลต่อ ความสัมพันธ์กับโอกาสที่เหตุการณ์จะเกิดแล้วค่า -2LL จะต้องต่ำกว่า -2LL กรณีที่ไม่มีตัวแปรอิสระดังกล่าว

$$df = [(N-1) - (N-(p+1))] = p \text{ หรือ } df = p \text{ จำนวนตัวแปรอิสระในตัวแบบ}$$

2) ตรวจสอบตัวแปรอิสระทีละตัว

สมมติฐานการทดสอบ คือ

$$H_0 : \beta_i = 0$$

$$H_1 : \beta_i \neq 0 ; i=1,2,\dots,p$$

ใช้การทดสอบสมมติฐานเกี่ยวกับสัมประสิทธิ์ความถดถอยโลจิสติก โดยสถิติทดสอบวอลด์ (Wald- test) มีการแจกแจงแบบ Chi-square (χ^2) ที่มี $df = 1$ และใช้สถิติทดสอบอัตราส่วนความควรจะเป็น (Block chi-square) ในกรณีที่ค่าสัมประสิทธิ์มีค่ามากหรือตัวแปรอิสระเป็นตัวแปรเชิงคุณภาพ

2.2.4 การวัดระดับความสัมพันธ์

การวัดระดับความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระหรือปัจจัยใหม่จากการใช้ค่าการวัดอำนาจของการพยากรณ์ด้วยค่า Nagelkerke's R^2 (R_N^2)

2.2.5 สร้างตัวแบบฟังก์ชันตอบสนองโลจิท (Logit response function)

ตัวแบบ Logistic regression ที่แสดงความสัมพันธ์ในรูปแบบ Multiple linear regression จะอยู่ในรูปของ $\log_e(\text{odds})$ หรือ Logit response function จาก

$$\text{Odds} = E\left(\frac{Y=1}{Y=0}\right) = \frac{e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_m X_m}}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_m X_m}}$$

$$\log_e(\text{odds}) = \log_e\left(\frac{\hat{P}}{1-\hat{P}}\right) = \beta'_0 + \beta'_1 X_1 + \beta'_2 X_2 + \dots + \beta'_m X_m$$

โดยที่ \hat{P} คือค่าพยากรณ์ที่ได้จากการประมาณค่าของตัวแบบจากข้อมูลที่ศึกษา

$\beta_0, \beta_1, \dots, \beta_m$ หมายถึง ค่าพารามิเตอร์ เมื่อทำการประมาณค่าจะได้

$\beta'_0, \beta'_1, \dots, \beta'_m$ หมายถึง ค่าสัมประสิทธิ์ที่ประมาณได้จากข้อมูล

X_1, X_2, \dots, X_m หมายถึง ตัวแปรอิสระซึ่งมีทั้งหมด m ตัว

2.2.6 อธิบายผลการวิเคราะห์จากตัวแบบที่สร้างขึ้น

ค่า e^{β_j} จะแสดงการเปลี่ยนแปลงของ Odds คือ ทุกครั้งที่ X_j เพิ่มขึ้น 1 หน่วย จะทำให้ Odds เพิ่มขึ้น e^{β_j} เท่า จากอัตราส่วนเดิม แต่ในการตีความต้องคำนึงถึงเครื่องหมายด้วย ในกรณีที่มีค่ามากกว่า 0 หรือ มีเครื่องหมายเป็นบวก ก็จะทำให้ Odds เพิ่มขึ้น แต่ถ้ามีค่าน้อยกว่า 0 หรือมีเครื่องหมายเป็นลบ ก็จะทำให้ Odds ลดลง และถ้า β_j มีค่าเท่ากับ 0 ก็จะทำให้ ค่า e^{β_j} มีค่าเท่ากับ 1 หมายความว่าไม่ทำให้ Odds เปลี่ยนแปลง

2.2.7 พล็อตกราฟแสดงการกระจายของข้อมูล

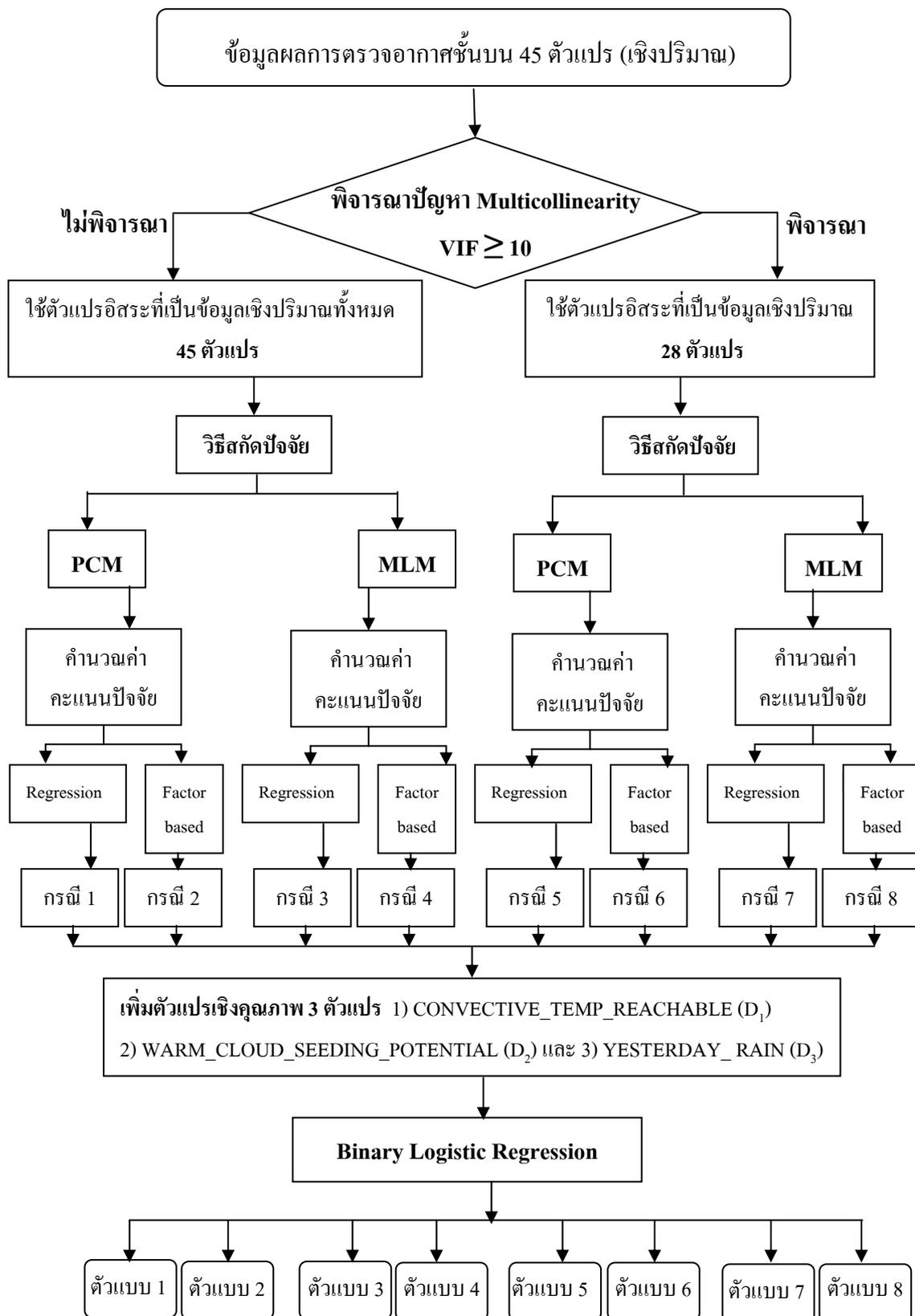
ด้วยการนำข้อมูลที่ได้จากการวิเคราะห์ตัวแบบหาค่าความคลาดเคลื่อน (Residual) พล็อตคู่กับค่าพยากรณ์ (Predicted, \hat{Y}) ของตัวแบบแต่ละกรณี เพื่อตรวจสอบการกระจายของข้อมูล ให้เห็นรูปแบบแนวโน้มค่าพยากรณ์ในแต่ละตัวแบบ โดยที่ค่า Probabilities เป็นค่าพยากรณ์โอกาสที่จะเกิดเหตุการณ์ในแต่ละรายการ (case) และค่า Residuals คำนวณจาก $Y - \hat{Y}$ เป็นการเปรียบเทียบค่าพยากรณ์กับค่าจริง ซึ่งกำหนดค่า Cut value = 0.50 นั่นคือ

ถ้า $\hat{P}(\text{Rain}) \leq 0.50$ จะให้เป็นเหตุการณ์ฝนไม่ตก

และถ้า $\hat{P}(\text{Rain}) > 0.50$ จะให้เป็นเหตุการณ์ฝนตก

สำหรับเกณฑ์ที่ใช้ในการเปรียบเทียบการสร้างตัวแบบจากการวิเคราะห์โลจิสติกแบบทวิ จะพิจารณาจากค่าสถิติทดสอบ G^2 ค่า R^2_N ค่าร้อยละความถูกต้องของการพยากรณ์จากตัวแบบ และทดสอบความเหมาะสมของตัวแบบจากค่าสถิติ Chi-Square ร่วมกับการพิจารณาการใช้จำนวนตัวแปรอิสระในการสร้างตัวแบบ นั่นคือ ถ้าตัวแบบใดให้ค่า Chi-Square ต่ำ และค่าสถิติทดสอบ G^2 ค่า R^2_N และค่าร้อยละความถูกต้องของการพยากรณ์ มีค่าสูง ร่วมกับการใช้จำนวนตัวแปรอิสระจำนวนน้อยที่สุด แสดงว่าตัวแบบในกรณีนั้น ๆ ให้ความเหมาะสมมากที่สุด

สรุปได้ดังแผนภาพที่ 17



ภาพที่ 17 แผนผังแสดงลำดับการวิเคราะห์ปัจจัยเพื่อสร้างตัวแบบโลจิสติก 8 ตัวแบบ

สถานที่และระยะเวลาทำการวิจัย

ดำเนินการวิจัยที่ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ โดยใช้
ระยะเวลาดำเนินการวิจัยระหว่างเดือนตุลาคม พ.ศ. 2549 ถึงเดือนตุลาคม พ.ศ. 2552

ประโยชน์ที่ได้รับจากการวิจัย

1. ได้ปัจจัยร่วมที่ใช้แทนจำนวนตัวแปรเดิมเฉพาะตัวแปรเชิงปริมาณที่เกี่ยวข้องกับผลการตรวจอากาศชั้นบน (GPCM) สำหรับนำไปสร้างตัวแบบการคาดการณ์โอกาสการเกิดฝนของภาคตะวันออกเฉียงเหนือ เพื่อประกอบการตัดสินใจในการขึ้นปฏิบัติการฝนหลวง
2. ได้เรียนรู้วิธีการทางสถิติ การวิเคราะห์ปัจจัย เพื่อลดจำนวนตัวแปร และการแก้ปัญหาข้อมูลสูญหาย ตลอดจนการสร้างตัวแบบพยากรณ์โอกาสการเกิดฝนของภาคตะวันออกเฉียงเหนือ
3. ผู้ปฏิบัติการที่เกี่ยวข้องสามารถนำวิธีการทางสถิติไปวิเคราะห์สร้างตัวแบบการเกิดฝนของพื้นที่อื่น ๆ ในประเทศไทย

ผลและวิจารณ์

ผล

จากการวิเคราะห์ปัจจัยสำหรับตัวแบบคาดการณ์โอกาสการเกิดฝนรายวัน ของภาคตะวันออกเฉียงเหนือ เพื่อตอบวัตถุประสงค์ 2 ข้อ คือ 1) เพื่อหาปัจจัยร่วมที่เกี่ยวข้องกับข้อมูลผลการตรวจอากาศชั้นบน ของภาคตะวันออกเฉียงเหนือ และ 2) เพื่อหาตัวแบบการคาดการณ์โอกาสการเกิดฝนในภาคตะวันออกเฉียงเหนือ สรุปผลการศึกษาแยกเป็น 3 ส่วน ดังนี้

ส่วนที่ 1 ผลการวิเคราะห์ข้อมูลเบื้องต้น

1.1 ข้อมูลเชิงปริมาณ

จากข้อมูลทั้งหมด 277 รายการ ที่ประกอบด้วยตัวแปรอิสระที่เป็นข้อมูลเชิงปริมาณ 45 ตัวแปร วิเคราะห์ค่าสถิติพื้นฐานของตัวแปรเหล่านี้ ด้วยค่าเฉลี่ย ค่าเบี่ยงเบนมาตรฐาน ค่าน้อยสุด ค่ามากที่สุด และค่าสถิติ VIF ที่ตรวจสอบปัญหา Multicollinearity ของตัวแปรอิสระ กับตัวแปรตามอีก 1 ตัวแปรที่เป็นข้อมูลปริมาณน้ำฝน 24 ชั่วโมง ดังแสดงในตารางที่ 7

ตารางที่ 7 ค่าสถิติพื้นฐานของตัวแปรเชิงปริมาณ จากชุดข้อมูล 277 รายการ

รายการ	ค่าเฉลี่ย	ค่าเบี่ยงเบนมาตรฐาน	ค่าน้อยสุด	ค่ามากที่สุด	VIF ^a
ตัวแปรตาม					
- RAIN24hr.	5.43	6.66	0.00	35.85	-
ตัวแปรอิสระ					
กลุ่มที่ 1 ตัวแปรในกลุ่มอุณหภูมิ					
1 RAOB_SURFACE_TEMPERATURE	24.78	1.30	19.10	29.10	71.83*
2 CONVECTIVE_TEMPERATURE	33.81	3.36	24.00	45.80	758.01*
3 CLOUD_BASE_HEIGHT_LCL	2,766.16	853.20	1,035.00	6,020.00	30.53*

ตารางที่ 7 (ต่อ)

	รายการ	ค่าเฉลี่ย	ค่า เบี่ยงเบน มาตรฐาน	ค่า น้อยสุด	ค่า มากที่สุด	VIF ^a
4	CLOUD_BASE_TEMPERATURE_LCL	19.04	2.52	8.10	25.50	71.95*
5	SFC_TEMP_RISE_REQUIRED_FOR_CCL	9.08	3.40	0.90	23.40	381.73*
6	CLOUD_BASE_PRESSURE_CCL	795.34	56.57	600.00	980.00	686.00*
7	CLOUD_BASE_HEIGHT_CCL	6,832.86	1,994.24	797.00	14,552.00	201.04*
8	CLOUD_BASE_TEMPERATURE_CCL	15.82	3.10	3.10	24.20	540.74*
9	SUB_CLOUD_MIXING_RATIO_CCL	14.49	1.82	8.28	19.63	2,615.6*
กลุ่มที่ 2 ตัวแปรในกลุ่มความชื้น						
10	PRECIPITABLE_WATER_SFC850	1.80	0.23	1.13	2.51	149.62*
11	PRECIPITABLE_WATER_SFC700	3.19	0.43	1.87	4.47	344.23*
12	PRECIPITABLE_WATER_SFC500	4.19	0.67	2.05	6.04	608.67*
13	PRECIPITABLE_WATER_TOTAL	4.53	0.81	2.10	6.59	122.57*
14	MEAN_MIXING_RATIO_LOWEST_100_MB	13.58	1.75	8.16	18.96	372.12*
15	MEAN_MIXING_RATIO_LOWEST_50_MB	14.24	1.81	8.17	19.61	4,033.2*
16	MEAN_DEW_POINT_TEMP_LOWEST_50_MB	18.42	2.08	10.12	23.62	518.33*
17	RH_AT_THE_CCL	66.44	16.45	10.20	100.00	9.87
18	AVG_RH_0_10000_FT	68.04	10.46	34.40	97.70	63.64*
19	AVG_RH_10_18000_FT	62.12	19.82	9.70	100.00	95.03*
20	MEAN_RH_1000_5000_FT	70.62	9.44	37.90	99.00	14.68*
21	MEAN_RH_5000_10000_FT	67.81	13.63	17.80	100.00	9.34
22	MEAN_RH_10000_15000_FT	63.59	17.67	9.90	100.00	12.97*
23	MEAN_RH_20000_25000_FT	48.80	24.11	5.80	100.00	6.57
กลุ่มที่ 3 ตัวแปรในกลุ่มความสูง หรือความกดอากาศ						
24	LEVEL_OF_FREE_CONVECTION_LFC_wrtTma	711.81	86.51	460.00	960.00	2.62
25	TOP_OF_LATENT_INSTABILITY_LAYER_wrtTma	579.79	101.93	190.00	930.00	1.71
26	LEVEL_OF_NEUTRAL_BUOYANCY_LNB	539.05	92.00	190.00	940.00	2.44

ตารางที่ 7 (ต่อ)

	รายการ	ค่าเฉลี่ย	ค่า เบี่ยงเบน มาตรฐาน	ค่า น้อยสุด	ค่า มากที่สุด	VIF ^a
27	ISOTHERM_HEIGHTS_0c	16,103.7	1,308.39	6,233.00	17,944.00	10.61*
28	ISOTHERM_HEIGHTS_minus5c	19,135.0	1,606.08	6,568.00	24,546.00	16.84*
29	ISOTHERM_HEIGHTS_minus10c	21,978.8	1,965.65	6,904.00	32,805.00	10.45*
30	ISOTHERM_HEIGHTS_minus15c	24,849.7	1,338.59	7,239.00	32,805.00	3.02
กลุ่มที่ 4 ตัวแปรในกลุ่มกระแสลม						
31	MEAN_WIND_1000_5000_FT_DEG	217.62	82.91	9.20	352.90	2.41
32	MEAN_WIND_5000_10000_FT_DEG	199.67	89.90	8.20	357.10	3.58
33	MEAN_WIND_10000_15000_FT_DEG	199.24	92.45	1.90	359.60	3.00
34	MEAN_WIND_20000_25000_FT_DEG	187.62	93.44	6.10	358.70	1.52
35	MEAN_WIND_1000_5000_FT_KTS	13.26	7.29	0.20	31.50	3.84
36	MEAN_WIND_5000_10000_FT_KTS	11.78	8.00	0.50	42.50	7.01
37	MEAN_WIND_10000_15000_FT_KTS	10.86	6.73	0.50	40.00	3.68
38	MEAN_WIND_20000_25000_FT_KTS	10.46	6.17	0.70	30.00	1.64
กลุ่มที่ 5 ข้อมูลในกลุ่มการทรงตัวของบรรยากาศ และค่าพยากรณ์อากาศ						
39	LIFTED_INDEX_100_MBAR_LAYER _ADIABATIC	0.09	2.27	-9.00	8.40	114.13*
40	LIFTED_INDEX_50_MBAR_LAYER_MEAN _VALUES	1.02	2.49	-8.00	11.50	64.24*
41	SHOWALTER_INDEX	2.18	2.17	-5.60	9.10	153.89*
42	TOTAL_TOTALS_INDEX	42.06	3.09	32.60	57.40	193.72*
43	K_INDEX	30.24	6.10	2.50	42.30	9.00
44	SWEAT_INDEX	196.12	36.58	69.40	355.00	7.89
45	POTENTIAL_BUOYANCY_INDEX	2.10	1.70	-3.90	9.30	17.96*

หมายเหตุ (a) VIF เป็นสถิติที่ใช้ตรวจสอบตัวแปรที่มีปัญหา Multicollinearity

(*) ตัวแปรที่มีปัญหา Multicollinearity ซึ่งมีค่า VIF ≥ 10

จากตารางที่ 7 สรุปผลการวิเคราะห์เบื้องต้นของข้อมูลเชิงปริมาณ เช่น ตัวแปร RAOB_SURFACE_TEMPERATURE มีค่าเฉลี่ยเท่ากับ 24.78 เซลเซียส ค่าเบี่ยงเบนมาตรฐานเท่ากับ 1.30 เซลเซียส ค่าน้อยสุดเท่ากับ 19.10 เซลเซียส ค่ามากที่สุดเท่ากับ 29.10 เซลเซียส และมีค่า VIF เท่ากับ 71.83 ซึ่งมีค่ามากกว่าหรือเท่ากับ 10 แสดงว่าตัวแปรมีปัญหา Multicollinearity คือมีความสัมพันธ์กับตัวแปรอิสระตัวอื่นด้วย นอกจากนี้ยังมีตัวแปรอิสระที่มีค่า VIF ≥ 10 นับได้ทั้งหมด 28 ตัวแปร

1.2 ข้อมูลเชิงคุณภาพ

มีตัวแปรอิสระที่เป็นข้อมูลเชิงคุณภาพ 3 ตัวแปร และตัวแปรตาม คือเหตุการณ์การเกิดฝนประจำวัน 1 ตัวแปร วิเคราะห์ข้อมูลเบื้องต้นด้วยจำนวน และร้อยละ ได้ผลดังนี้

ตารางที่ 8 ค่าสถิติพื้นฐานของตัวแปรเชิงคุณภาพ จากชุดข้อมูล 277 รายการ

	รายการ	จำนวน	ร้อยละ
กลุ่มที่ 1 ตัวแปรในกลุ่มอุณหภูมิ			
1	CONVECTIVE_TEMP_REACHABLE		
	- ไม่มีโอกาสที่อุณหภูมิผิวพื้นจะถึงค่าอุณหภูมิในการเกิดเมฆ	162	58.5
	- มีโอกาสที่อุณหภูมิผิวพื้นจะถึงค่าอุณหภูมิในการเกิดเมฆ	115	41.5
กลุ่มที่ 2 ตัวแปรพยากรณ์โอกาสความสำเร็จในการทำฝน			
2	WARM_CLOUD_SEEDING_POTENTIAL		
	- โอกาสทำฝนจากเมฆอุ่นไม่สำเร็จ	268	96.8
	- โอกาสทำฝนจากเมฆอุ่นสำเร็จ	9	3.2
กลุ่มที่ 3 ตัวแปรปริมาณน้ำฝน 24 ชั่วโมง			
3	YESTERDAY_RAIN		
	- เหตุการณ์ที่ไม่มีฝนตกของวันก่อนหน้า	41	14.8
	- เหตุการณ์ที่มีฝนตกของวันก่อนหน้า	236	85.2
4	เหตุการณ์เกิดฝนประจำวัน		
	- ฝนไม่ตกวันนี้	38	13.7
	- ฝนตกวันนี้	239	86.3

จากการวิเคราะห์ปัญหา Multicollinearity ระหว่างตัวแปรอิสระที่เป็นตัวแปรเชิงปริมาณ ทำให้การวิเคราะห์ในส่วนถัดไปพิจารณาแยกเป็น 2 กรณี คือ 1) ใช้จำนวนตัวแปรอิสระทั้งหมดที่เป็นข้อมูลเชิงปริมาณ 45 ตัวแปร ไม่มีการตัดตัวแปรที่มีค่า VIF ≥ 10 ออก และ 2) ใช้จำนวนตัวแปรอิสระที่มีปัญหา Multicollinearity มีทั้งหมด 28 ตัวแปร

ส่วนที่ 2 ผลการวิเคราะห์ปัจจัย

2.1 ผลการวิเคราะห์ปัจจัย (กรณีที่ 1 – 4)

เป็นกรณีที่ใช้ตัวแปรอิสระที่เป็นตัวแปรเชิงปริมาณทั้งหมด 45 ตัวแปร ที่ผ่านการแปลงข้อมูลให้อยู่ในรูปคะแนนมาตรฐาน (Standardized score) แล้ว ก่อนทำการวิเคราะห์ต่อไปโดยใช้วิธีสกัดปัจจัย 2 วิธี คือ PCM และ MLM ดังนี้

2.1.1 ผลการตรวจสอบความเหมาะสมของการใช้เทคนิคการวิเคราะห์ปัจจัยกับชุดข้อมูล 45 ตัวแปร

จากการตรวจสอบความเหมาะสมของข้อมูลที่ใช้เทคนิคการวิเคราะห์ปัจจัยด้วยสถิติ KMO และ Bartlett's Test of Sphericity พบว่าได้ค่า KMO เท่ากับ 0.862 อยู่ในระดับดี และ ค่า Bartlett's Test of Sphericity มีค่า P-value น้อยกว่า 0.001 ซึ่งทำให้ไม่มีหลักฐานเพียงพอที่จะปฏิเสธสมมติฐานหลักที่ว่าข้อมูลชุดดังกล่าวสามารถวิเคราะห์ปัจจัยได้ และผลจากทดสอบสถิติทั้งสองมีความสอดคล้องกัน จึงสรุปได้ว่าข้อมูลตัวแปรอิสระทั้ง 45 ตัวแปร มีความเหมาะสมที่จะวิเคราะห์ด้วยเทคนิคการวิเคราะห์ปัจจัย ดังตารางที่ 9

ตารางที่ 9 ผลการตรวจสอบความเหมาะสมของข้อมูล 45 ตัวแปร (กรณีที่ 1 – 4) ใช้เทคนิคการวิเคราะห์ปัจจัยด้วยค่าสถิติ KMO และค่าสถิติ Bartlett's Test of Sphericity

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.862
Bartlett's Test of Sphericity	Approx. Chi-Square	27,129.710
	Df.	990
	Sig.	0.001**

หมายเหตุ ระดับนัยสำคัญของการทดสอบที่ 0.05 ($\alpha = 0.05$)

2.1.2 ผลการสกัดปัจจัยแบบ PCM (กรณีที่ 1 - 2)

1) จำนวนปัจจัยร่วมที่เหมาะสม

พิจารณาจากปัจจัยร่วมที่มีค่าไอเกนมากกว่า 1 ในตารางที่ 10 ร่วมกับการพล็อตกราฟ Scree Plot พบว่าปัจจัยร่วมที่มีค่าไอเกนมากกว่า 1 ก่อนการหมุนแกน มีทั้งหมด 8 ปัจจัย (factors) มีค่าไอเกนเท่ากับ 18.91, 5.04, 3.69, 3.37, 2.15, 1.65, 1.49 และ 1.18 ตามลำดับ ซึ่งสามารถอธิบายความผันแปรของตัวแปรเดิมได้ร้อยละ 83.3

ตารางที่ 10 ผลการวิเคราะห์น้ำหนักปัจจัยก่อนการหมุนแกน และร้อยละความผันแปรของแต่ละปัจจัย ที่ได้จากการสกัดปัจจัยแบบ PCM จากตัวแปรเชิงปริมาณ 45 ตัวแปร

ชื่อตัวแปร	Factor code	Factor							
		1	2	3	4	5	6	7	8
1 RAOB_SURFACE_TEMPERATURE	X111	0.108	-0.221	0.523	0.057	-0.071	0.105	0.668	-0.127
2 CONVECTIVE_TEMPERATURE	X121	-0.804	0.001	0.417	0.164	-0.058	0.012	-0.172	0.178
3 CLOUD_BASE_HEIGHT_LCL	X131	-0.724	0.078	0.203	-0.106	-0.296	0.256	0.388	-0.046
4 CLOUD_BASE_TEMPERATURE_LCL	X132	0.805	-0.236	0.054	0.146	0.281	-0.197	-0.064	-0.026
5 SFC_TEMP_RISE_REQUIRED_FOR_CCL	X141	-0.822	0.073	0.186	0.019	-0.014	-0.043	-0.419	0.206
6 CLOUD_BASE_PRESSURE_CCL	X142	0.908	-0.106	-0.263	-0.122	0.130	-0.068	0.097	-0.151
7 CLOUD_BASE_HEIGHT_CCL	X143	-0.913	0.073	0.227	0.120	-0.124	0.079	-0.119	0.180
8 CLOUD_BASE_TEMPERATURE_CCL	X144	0.937	-0.179	-0.031	-0.034	0.195	-0.108	0.049	-0.139
9 SUB_CLOUD_MIXING_RATIO_CCL	X145	0.921	-0.195	0.100	0.126	0.212	-0.113	-0.002	-0.071
10 PRECIPITABLE_WATER_SFC850	X211	0.909	-0.222	-0.006	0.149	0.024	0.022	-0.072	-0.005
11 PRECIPITABLE_WATER_SFC700	X212	0.917	0.025	0.019	0.243	-0.211	0.073	0.025	0.065
12 PRECIPITABLE_WATER_SFC500	X213	0.892	0.324	0.047	0.110	-0.228	-0.040	-0.029	0.055
13 PRECIPITABLE_WATER_TOTAL	X214	0.871	0.399	0.042	0.065	-0.177	-0.065	-0.056	0.057
14 MEAN_MIXING_RATIO_LOWEST_100_MB	X221	0.933	-0.168	0.062	0.137	0.118	-0.033	-0.034	-0.059
15 MEAN_MIXING_RATIO_LOWEST_50_MB	X222	0.929	-0.178	0.096	0.121	0.190	-0.094	-0.003	-0.081
16 MEAN_DEW_POINT_TEMP_LOWEST_50_MB	X231	0.927	-0.182	0.097	0.118	0.188	-0.094	0.011	-0.116
17 RH_AT_THE_CCL	X241	0.749	0.181	-0.283	-0.062	-0.313	0.122	0.232	-0.039
18 AVG_RH_0_10000_FT	X242	0.856	0.225	-0.286	0.067	-0.215	0.062	-0.050	0.056

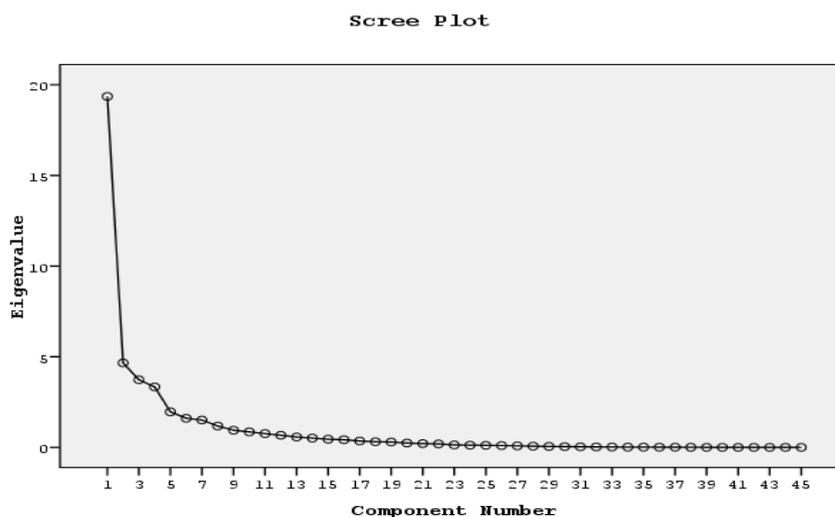
ตารางที่ 10 (ต่อ)

ชื่อตัวแปร	Factor code	Factor							
		1	2	3	4	5	6	7	8
19AVG_RH_10_18000_FT	X243	0.631	0.613	0.041	-0.139	-0.223	-0.176	-0.131	0.035
20MEAN_RH_1000_5000_FT	X244	0.827	0.004	-0.286	0.132	0.079	-0.017	-0.281	0.023
21MEAN_RH_5000_10000_FT	X245	0.680	0.286	-0.260	0.008	-0.406	0.085	0.134	0.097
22MEAN_RH_10000_15000_FT	X246	0.625	0.563	-0.007	-0.098	-0.278	-0.170	-0.069	0.068
23MEAN_RH_20000_25000_FT	X247	0.581	0.523	-0.008	-0.165	0.032	-0.048	-0.153	-0.012
24LEVEL_OF_FREE_CONVECTION_LFC_wrtTma	X311	0.327	-0.124	-0.495	0.102	0.267	0.078	-0.007	0.439
25TOP_OF_LATENT_INSTABILITY_LAYER_wrtT	X312	0.151	0.023	-0.329	0.046	0.369	-0.011	0.337	0.613
26LEVEL_OF_NEUTRAL_BUOYANCY_LNB	X321	-0.125	-0.013	-0.657	0.132	0.333	0.061	0.257	0.179
27ISOTHERM_HEIGHTS_0c	X331	-0.138	0.069	0.115	0.894	-0.019	0.056	0.116	0.007
28ISOTHERM_HEIGHTS_minus5c	X332	-0.122	0.171	0.124	0.904	-0.035	0.042	0.097	-0.049
29ISOTHERM_HEIGHTS_minus10c	X333	-0.108	0.129	0.121	0.913	-0.079	0.028	-0.023	-0.030
30ISOTHERM_HEIGHTS_minus15c	X334	0.033	0.226	-0.020	0.660	-0.039	-0.124	-0.077	0.034
31MEAN_WIND_1000_5000_FT_DEG	X411	0.211	0.187	0.633	-0.088	0.122	-0.296	0.293	0.144
32MEAN_WIND_5000_10000_FT_DEG	X412	0.201	0.497	0.534	-0.137	0.261	-0.264	0.115	0.120
33MEAN_WIND_10000_15000_FT_DEG	X413	0.151	0.513	0.503	-0.134	0.232	-0.229	0.074	0.232
34MEAN_WIND_20000_25000_FT_DEG	X414	-0.010	0.125	0.387	-0.088	0.274	0.147	-0.047	0.045
35MEAN_WIND_1000_5000_FT_KTS	X421	-0.023	0.749	0.110	-0.032	0.306	0.171	0.003	0.031
36MEAN_WIND_5000_10000_FT_KTS	X422	0.150	0.764	0.068	-0.049	0.299	0.326	0.002	-0.212
37MEAN_WIND_10000_15000_FT_KTS	X423	0.127	0.650	0.067	-0.036	0.345	0.425	-0.073	-0.249
38MEAN_WIND_20000_25000_FT_KTS	X424	-0.080	-0.063	-0.080	0.046	0.278	0.509	-0.085	0.008
39LIFTED_INDEX_100_MBAR_LAYER_ADIABATIC	X51	-0.786	0.458	-0.318	0.055	-0.057	-0.106	0.047	0.001
40LIFTED_INDEX_50_MBAR_LAYER_MEAN_VALUES	X52	-0.801	0.434	-0.298	0.049	-0.121	-0.073	-0.050	0.039
41SHOWALTER_INDEX	X53	-0.712	0.346	-0.260	0.089	0.233	-0.332	0.028	-0.235
42TOTAL_TOTALS_INDEX	X54	0.573	-0.451	0.365	-0.175	-0.215	0.351	-0.061	0.228
43K_INDEX	X55	0.753	0.224	0.125	-0.100	-0.400	0.033	0.043	0.144
44SWEAT_INDEX	X56	0.666	0.310	0.090	-0.022	0.048	0.529	-0.022	0.098
45POTENTIAL_BUOYANCY_INDEX	X57	0.425	-0.431	0.578	0.100	0.056	0.122	-0.340	0.112

ตารางที่ 10 (ต่อ)

ชื่อตัวแปร	Factor	Factor	Factor	Factor	Factor	Factor	Factor	Factor
	1	2	3	4	5	6	7	8
Eigen values	18.905	5.044	3.691	3.374	2.150	1.655	1.496	1.182
Proportion of total variance	0.420	0.112	0.082	0.075	0.048	0.037	0.033	0.026
Cumulative proportion	0.420	0.532	0.614	0.689	0.737	0.774	0.807	0.833

เมื่อพิจารณาจากกราฟ Scree Plot ดังภาพที่ 18 ค่าไอเกนลดลงอย่างรวดเร็ว หรือเกือบจะขนานกับแกนอน เริ่มจากค่าไอเกนของปัจจัยที่ 5 ดังนั้น จึงพิจารณาใช้เพียง 4 ปัจจัยรวม (Factors) ที่สามารถอธิบายเกี่ยวกับการเกิดฝนได้



ภาพที่ 18 กราฟ Scree Plot จากการวิเคราะห์ตัวแปรเชิงปริมาณ 45 ตัวแปร ด้วยวิธี PCM

ดังนั้นวิธีการสกัดปัจจัยแบบ PCM ที่ใช้เกณฑ์การเลือกปัจจัยที่ให้ค่าไอเกน มากกว่า 1 และจากกราฟได้จำนวนปัจจัยที่เหมาะสม 4 ปัจจัยที่ให้ค่าไอเกนเท่ากับ 18.91, 5.04, 3.69 และ 3.37 ตามลำดับ ซึ่งมีค่าสัดส่วนสะสมของความแปรปรวนที่อธิบายตัวแปรเดิมได้ร้อยละ 68.9

2) ปัจจัยที่สกัดได้จากวิธีการหมุนแกน

ปัจจัยร่วมที่ได้ในตารางที่ 10 ยังไม่สามารถจัดตัวแปรให้อยู่ในปัจจัยใดปัจจัยหนึ่งได้ชัดเจนซึ่งในที่นี้มองปัจจัยเสมือนเป็นแกนในทางพีชคณิต โดยเลือกการหมุนแกนแบบมุมฉาก (Varimax rotation) เพื่อให้ปัจจัยที่สกัดได้มีความเป็นอิสระต่อกัน แต่จะทำให้ค่าน้ำหนักปัจจัย (Factor loading) เปลี่ยนแปลงไปจากเดิมก่อนที่มีการหมุนแกน เพื่อความสะดวกและง่ายต่อการจัดตัวแปรเดิมให้อยู่ในปัจจัยใหม่ และตั้งชื่อปัจจัยให้สอดคล้องกับกลุ่มตัวแปรเดิมให้มากที่สุด ดังแสดงไว้ในตารางที่ 11

ตารางที่ 11 ผลการวิเคราะห์น้ำหนักปัจจัยหลังการหมุนแกนแบบ Varimax และร้อยละความผันแปรของแต่ละปัจจัย ที่ได้จากการสกัดปัจจัยแบบ PCM จากตัวแปรเชิงปริมาณ 45 ตัวแปร

ที่	ตัวแปร	code	ค่าน้ำหนักปัจจัย (Loading, l_{im})				ค่าความ	ค่าปัจจัย
			Factor 1	Factor 2	Factor 3	Factor 4	ร่วมกัน	เฉพาะ
			Factor Factor Factor Factor				h_i^2	e_i
1	MEAN_MIXING_RATIO_LOWEST_100_MB	X221	0.950	0.057	0.110	0.058	0.921	0.079
2	PRECIPITABLE_WATER_SFC850	X211	0.945	-0.008	0.046	0.056	0.898	0.102
3	CLOUD_BASE_TEMPERATURE_CCL	X144	0.944	0.066	0.036	-0.122	0.912	0.088
4	MEAN_MIXING_RATIO_LOWEST_50_MB	X222	0.944	0.051	0.147	0.046	0.918	0.082
5	MEAN_DEW_POINT_TEMP_LOWEST_50_MB	X231	0.943	0.048	0.148	0.042	0.916	0.084
6	SUB_CLOUD_MIXING_RATIO_CCL	X145	0.942	0.033	0.151	0.048	0.913	0.087
7	CLOUD_BASE_PRESSURE_CCL	X142	0.904	0.123	-0.194	-0.220	0.919	0.081
8	PRECIPITABLE_WATER_SFC700	X212	0.903	0.219	0.037	0.188	0.901	0.099
9	CLOUD_BASE_HEIGHT_CCL	X143	-0.900	-0.158	0.163	0.209	0.905	0.095
10	LIFTED_INDEX_50_MBAR_LAYER_MEAN_VALUES	X52	-0.851	0.185	-0.382	0.130	0.922	0.078
11	CLOUD_BASE_TEMPERATURE_LCL	X132	0.844	-0.041	0.103	0.064	0.728	0.272
12	LIFTED_INDEX_100_MBAR_LAYER_ADIABATIC	X51	-0.840	0.210	-0.404	0.137	0.932	0.068
13	MEAN_RH_1000_5000_FT	X244	0.829	0.169	-0.256	0.048	0.783	0.217
14	SFC_TEMP_RISE_REQUIRED_FOR_CCL	X141	-0.819	-0.124	0.135	0.100	0.715	0.285

ตารางที่ 11 (ต่อ)

ที่	ตัวแปร	code	ค่าน้ำหนักปัจจัย (Loading, l_{im})				ค่าความ	ค่าปัจจัย
			Factor 1	Factor 2	Factor 3	Factor 4	ร่วมกัน	เฉพาะ
			Factor Factor Factor Factor				h_i^2	e_i
15	AVG_RH_0_10000_FT	X242	0.799	0.398	-0.271	0.016	0.870	0.130
16	PRECIPITABLE_WATER_SFC500	X213	0.795	0.520	0.045	0.109	0.916	0.084
17	CONVECTIVE_TEMPERATURE	X121	-0.785	-0.191	0.359	0.256	0.847	0.153
18	PRECIPITABLE_WATER_TOTAL	X214	0.752	0.592	0.035	0.076	0.923	0.077
19	SHOWALTER_INDEX	X53	-0.742	0.121	-0.335	0.155	0.702	0.298
20	CLOUD_BASE_HEIGHT_LCL	X131	-0.740	-0.075	0.168	-0.026	0.582	0.418
21	RH_AT_THE_CCL	X241	0.692	0.347	-0.258	-0.110	0.678	0.322
22	K_INDEX	X55	0.657	0.425	0.145	-0.095	0.642	0.358
23	TOTAL_TOTALS_INDEX	X54	0.617	-0.236	0.451	-0.235	0.696	0.304
24	MEAN_RH_5000_10000_FT	X245	0.607	0.422	-0.256	-0.018	0.613	0.387
25	SWEAT_INDEX	X56	0.563	0.471	0.090	-0.005	0.548	0.452
26	MEAN_WIND_5000_10000_FT_KTS	X422	-0.041	0.779	-0.001	0.068	0.613	0.387
27	AVG_RH_10_18000_FT	X243	0.450	0.766	0.020	-0.076	0.795	0.205
28	MEAN_WIND_1000_5000_FT_KTS	X421	-0.206	0.722	0.033	0.097	0.575	0.425
29	MEAN_RH_10000_15000_FT	X246	0.463	0.707	-0.026	-0.049	0.717	0.283
30	MEAN_RH_20000_25000_FT	X247	0.423	0.667	-0.019	-0.117	0.638	0.362
31	MEAN_WIND_10000_15000_FT_KTS	X423	-0.035	0.662	0.009	0.064	0.444	0.556
32	MEAN_WIND_5000_10000_FT_DEG	X412	0.032	0.585	0.499	-0.011	0.592	0.408
33	MEAN_WIND_10000_15000_FT_DEG	X413	-0.018	0.584	0.464	-0.006	0.557	0.443
34	MEAN_WIND_20000_25000_FT_KTS	X424	-0.053	-0.093	-0.080	0.032	0.019	0.981
35	LEVEL_OF_NEUTRAL_BUOYANCY_LNB	X321	-0.064	-0.112	-0.666	0.062	0.464	0.536
36	POTENTIAL_BUOYANCY_INDEX	X57	0.486	-0.276	0.627	0.069	0.711	0.289
37	MEAN_WIND_1000_5000_FT_DEG	X411	0.113	0.291	0.625	0.000	0.488	0.512
38	RAOB_SURFACE_TEMPERATURE	X111	0.130	-0.153	0.540	0.073	0.337	0.663

ตารางที่ 11 (ต่อ)

ที่	ตัวแปร	code	ค่าน้ำหนักปัจจัย (Loading, l_{im})				ค่าความ ร่วมกัน h_i^2	ค่าปัจจัย เฉพาะ e_i
			Factor 1	Factor 2	Factor 3	Factor 4		
39	LEVEL_OF_FREE_CONVECTION_LFC_wrtTma	X311	0.385	-0.088	-0.470	0.006	0.378	0.622
40	MEAN_WIND_20000_25000_FT_DEG	X414	-0.071	0.158	0.378	-0.024	0.173	0.827
41	TOP_OF_LATENT_INSTABILITY_LAYER_wrtTma	X312	0.165	0.029	-0.325	0.003	0.134	0.866
42	ISOTHERM_HEIGHTS_minus10c	X333	-0.047	-0.024	0.017	0.934	0.876	0.124
43	ISOTHERM_HEIGHTS_minus5c	X332	-0.072	0.014	0.016	0.933	0.876	0.124
44	ISOTHERM_HEIGHTS_0c	X331	-0.064	-0.087	0.018	0.908	0.837	0.163
45	ISOTHERM_HEIGHTS_minus15c	X334	0.049	0.128	-0.103	0.677	0.488	0.512
Eigen values			17.938	5.863	3.737	3.475		
Proportion of total variance			0.399	0.130	0.083	0.077		
Cumulative proportion			0.399	0.529	0.612	0.689		

หมายเหตุ 1. คอลัมน์ที่แรเงา คือตัวแปรอิสระที่เป็นองค์ประกอบของแต่ละปัจจัย (Factor)

2. ค่าน้ำหนักปัจจัยที่เป็นตัวเข้ม คือค่า Loading ที่มีค่ามากกว่าหรือเท่ากับ 0.7

3) ความสามารถของความแปรปรวนที่อธิบายได้ (Total Variance Explained)

ปัจจัยที่สกัดได้ทั้ง 4 มีความสามารถในการอธิบายการผันแปรของตัวแปรเดิม
ทั้ง 45 ตัวแปร คิดเป็นร้อยละ 68.9 โดยที่

ปัจจัยที่ 1 อธิบายการผันแปรได้มากที่สุด มีจำนวนตัวแปรเดิม 25 ตัวแปร คิดเป็นร้อยละ 39.9

ปัจจัยที่ 2 อธิบายการผันแปรของตัวแปรเดิม 9 ตัวแปร คิดเป็นร้อยละ 13.0

ปัจจัยที่ 3 อธิบายการผันแปรของตัวแปรเดิม 7 ตัวแปร คิดเป็นร้อยละ 8.3

ปัจจัยที่ 4 อธิบายการผันแปรของตัวแปรเดิม 4 ตัวแปร คิดเป็นร้อยละ 7.7

4) ประมวลค่าคะแนนปัจจัย (Estimated Factor Score)

ก. กรณีที่ 1 วิเคราะห์ปัจจัยจากตัวแปรปริมาณทั้งหมด 45 ตัวแปร

ด้วยวิธี PCM และคำนวณค่าคะแนนปัจจัยด้วยวิธี Regression

เป็นการสร้างคะแนนปัจจัยจากผลรวมของค่าสัมประสิทธิ์คะแนนปัจจัย

(Component score coefficient matrix) ที่ได้จากรายที่ 12 คูณด้วยค่าของตัวแปรแต่ละตัวแปร

(X_i) ที่อยู่ในรูปคะแนนมาตรฐาน โดยคำนวณค่าคะแนนปัจจัยเป็นรายวัน สามารถคำนวณได้จาก

$$F_i = \beta'(X_i - \bar{X}_i) + \varepsilon_i$$

กำหนดให้ $i = 1, 2, \dots, 45$ $j = 1, 2, 3, 4$ และ $k = 1, 2, 3, \dots, 277$

นำผลของเมตริกซ์สัมประสิทธิ์คะแนนปัจจัย ที่ได้จากการประมวลผลโปรแกรม SPSS ดังที่ปรากฏในตารางที่ 12 แล้วคำนวณค่าคะแนนปัจจัยด้วยวิธี Regression ได้ดังนี้

ตัวอย่าง การคำนวณค่าคะแนนปัจจัยพื้นฐาน ปัจจัยที่ 1 ถึงปัจจัยที่ 3 ของข้อมูลวันที่ 12 มีนาคม 2547 (วันแรก) ได้ดังนี้

$$F_{11} = (0.009)(0.631) + (-0.043)(0.919) + (-0.047)(1.146) + \dots + (0.035)(0.291) = -0.518$$

$$F_{21} = (-0.032)(0.631) + (-0.009)(0.919) + (0.014)(1.146) + \dots + (-0.068)(0.291) = -1.408$$

$$F_{31} = (0.143)(0.631) + (0.105)(0.919) + (0.054)(1.146) + \dots + (0.162)(0.291) = 0.362$$

$$F_{41} = (0.025)(0.631) + (0.063)(0.919) + (-0.020)(1.146) + \dots + (0.032)(0.291) = 0.485$$

สรุป เมื่อสกัดปัจจัยด้วยวิธี PCM แล้วคำนวณค่าคะแนนปัจจัย ด้วย วิธี Regression จะต้องใช้ตัวแปรเดิมทั้งหมด 45 ตัวแปร ในการสร้างตัวแปรใหม่ (ปัจจัยร่วม) 4 ปัจจัย

ตารางที่ 12 เมตริกซ์สัมประสิทธิ์คะแนนปัจจัยร่วม (Component Score Coefficient Matrix)

ที่ได้จากการสกัดปัจจัยแบบ PCM จากตัวแปรเชิงปริมาณ 45 ตัวแปร

ที่	ชื่อตัวแปร	code	Factor	Factor	Factor	Factor
			1	2	3	4
1	RAOB_SURFACE_TEMPERATURE	X111	0.009	-0.032	0.143	0.025
2	CONVECTIVE_TEMPERATURE	X121	-0.043	-0.009	0.105	0.063
3	CLOUD_BASE_HEIGHT_LCL	X131	-0.047	0.014	0.054	-0.020
4	CLOUD_BASE_TEMPERATURE_LCL	X132	0.056	-0.039	0.017	0.034
5	SFC_TEMP_RISE_REQUIRED_FOR_CCL	X141	-0.048	0.006	0.046	0.016
6	CLOUD_BASE_PRESSURE_CCL	X142	0.052	-0.008	-0.063	-0.050
7	CLOUD_BASE_HEIGHT_CCL	X143	-0.050	0.001	0.054	0.047

ตารางที่ 12 (ต่อ)

ที่	ชื่อตัวแปร	code	Factor 1	Factor 2	Factor 3	Factor 4
8	CLOUD_BASE_TEMPERATURE_CCL	X144	0.056	-0.021	-0.001	-0.019
9	SUB_CLOUD_MIXING_RATIO_CCL	X145	0.058	-0.028	0.030	0.031
10	PRECIPITABLE_WATER_SFC850	X211	0.061	-0.037	0.001	0.033
11	PRECIPITABLE_WATER_SFC700	X212	0.053	0.007	0.000	0.069
12	PRECIPITABLE_WATER_SFC500	X213	0.033	0.069	0.005	0.040
13	PRECIPITABLE_WATER_TOTAL	X214	0.027	0.085	0.003	0.029
14	MEAN_MIXING_RATIO_LOWEST_100_MB	X221	0.059	-0.024	0.019	0.034
15	MEAN_MIXING_RATIO_LOWEST_50_MB	X222	0.058	-0.025	0.028	0.030
16	MEAN_DEW_POINT_TEMP_LOWEST_50_MB	X231	0.058	-0.025	0.029	0.029
17	RH_AT_THE_CCL	X241	0.033	0.041	-0.076	-0.024
18	AVG_RH_0_10000_FT	X242	0.040	0.045	-0.081	0.015
19	AVG_RH_10_18000_FT	X243	-0.001	0.131	0.004	-0.023
20	MEAN_RH_1000_5000_FT	X244	0.051	0.000	-0.078	0.027
21	MEAN_RH_5000_10000_FT	X245	0.026	0.058	-0.074	0.001
22	MEAN_RH_10000_15000_FT	X246	0.003	0.119	-0.009	-0.014
23	MEAN_RH_20000_25000_FT	X247	0.001	0.113	-0.007	-0.034
24	LEVEL_OF_FREE_CONVECTION_LFC_wrtTma	X311	0.034	-0.033	-0.132	0.010
25	TOP_OF_LATENT_INSTABILITY_LAYER_wrtTma	X312	0.014	-0.002	-0.090	0.004
26	LEVEL_OF_NEUTRAL_BUOYANCY_LNB	X321	0.009	-0.023	-0.180	0.019
27	ISOTHERM_HEIGHTS_0c	X331	0.015	-0.024	0.005	0.266
28	ISOTHERM_HEIGHTS_minus5c	X332	0.012	-0.005	0.005	0.272
29	ISOTHERM_HEIGHTS_minus10c	X333	0.014	-0.013	0.004	0.273
30	ISOTHERM_HEIGHTS_minus15c	X334	0.012	0.015	-0.028	0.198
31	MEAN_WIND_1000_5000_FT_DEG	X411	-0.011	0.055	0.169	-0.001
32	MEAN_WIND_5000_10000_FT_DEG	X412	-0.026	0.113	0.137	-0.009
33	MEAN_WIND_10000_15000_FT_DEG	X413	-0.028	0.115	0.128	-0.009
34	MEAN_WIND_20000_25000_FT_DEG	X414	-0.015	0.035	0.104	-0.010
35	MEAN_WIND_1000_5000_FT_KTS	X421	-0.038	0.145	0.015	0.017

ตารางที่ 12 (ต่อ)

ที่	ชื่อตัวแปร	code	Factor 1	Factor 2	Factor 3	Factor 4
36	MEAN_WIND_5000_10000_FT_KTS	X422	-0.030	0.150	0.004	0.011
37	MEAN_WIND_10000_15000_FT_KTS	X423	-0.026	0.128	0.006	0.011
38	MEAN_WIND_20000_25000_FT_KTS	X424	0.002	-0.017	-0.022	0.009
39	LIFTED_INDEX_100_MBAR_LAYER_ADIABATIC	X51	-0.054	0.068	-0.098	0.023
40	LIFTED_INDEX_50_MBAR_LAYER_MEAN_VALUES	X52	-0.054	0.063	-0.092	0.021
41	SHOWALTER_INDEX	X53	-0.045	0.047	-0.081	0.031
42	TOTAL_TOTALS_INDEX	X54	0.039	-0.063	0.113	-0.055
43	K_INDEX	X55	0.023	0.059	0.034	-0.021
44	SWEAT_INDEX	X56	0.018	0.070	0.020	0.004
45	POTENTIAL_BUOYANCY_INDEX	X57	0.035	-0.068	0.162	0.032

ข. กรณีที่ 2 วิเคราะห์ปัจจัยจากตัวแปรปริมาณทั้งหมด 45 ตัวแปร

ด้วยวิธี PCM และคำนวณค่าคะแนนปัจจัยด้วยวิธี Factor Based Score

สร้างสมการคำนวณค่าคะแนนพื้นฐานของปัจจัย ด้วยการรวมคะแนน

เฉพาะตัวแปรที่อยู่ในปัจจัยเดียวกัน ที่มีค่าน้ำหนักปัจจัย (i_{ij}) ในตารางที่ 11 มากกว่า 0.70 เท่านั้น และตัวแปรเหล่านี้ผ่านการแปลงค่าข้อมูลให้อยู่ในรูปคะแนนมาตรฐานแล้ว

ตัวอย่าง การคำนวณค่าคะแนนปัจจัยพื้นฐาน ปัจจัยที่ 1 ถึงปัจจัยที่ 4 ของข้อมูลวันที่ 12 มีนาคม 2547 (วันแรก) ได้ดังนี้

$$\begin{aligned}
 F_{11} &= X221 + X211 + X144 + X222 + X231 + X145 + X142 \\
 &\quad + X212 + X143 + X52 + X132 + X51 + X244 + X141 \\
 &\quad + X242 + X213 + X121 + X214 + X53 + X131 \\
 &= -3.880
 \end{aligned}$$

$$F_{21} = X422 + X243 + X421 + X246 = -5.710$$

$$F_{31} = \text{ตัวแปร } X_i \text{ ที่จัดอยู่ในปัจจัยนี้ทุกตัวมีค่า loading น้อยกว่า 0.70 (ไม่นำมาวิเคราะห์)}$$

$$F_{41} = X333 + X332 + X331 = 1.400$$

สรุป เมื่อสกัดปัจจัยด้วยวิธี PCM แล้วคำนวณค่าคะแนนปัจจัย ด้วยวิธีสร้างค่าคะแนนพื้นฐานของปัจจัยจะใช้ตัวแปรเดิม 27 ตัวแปรในการสร้างตัวแปรใหม่ (ปัจจัยร่วม) 3 ปัจจัย

ตารางที่ 13 แสดงตัวอย่างการคำนวณค่าคะแนนปัจจัยโดยใช้วิธี Regression (กรณีที่ 1) และ Factor Based Score (กรณีที่ 2) เมื่อใช้วิธีสกัดปัจจัยด้วยเทคนิค PCM

ที่	วัน	Regression (กรณีที่ 1)				Factor Based Score (กรณีที่ 2)			
		Fac1	Fac2	Fac3	Fac4	Fac1	Fac2	Fac3	Fac4
1	12 มี.ค.47	-0.518	-1.408	0.362	0.485	-3.882	-5.709	-	1.400
2	13 มี.ค.47	0.103	-1.49	-0.202	0.456	-4.077	-5.1	-	1.373
3	14 มี.ค.47	0.116	-1.833	0.039	0.338	-3.305	-5.221	-	1.142
4	15 มี.ค.47	-0.106	-1.863	-0.351	0.38	-4.824	-7.161	-	1.030
5	16 มี.ค.47	0.855	-2.294	-0.506	-0.024	-1.121	-6.239	-	0.010

2.1.3 ผลการสกัดปัจจัยแบบ MLM (กรณีที่ 3 - 4)

1) จำนวนปัจจัยร่วมที่เหมาะสม

พิจารณาจากปัจจัยร่วมที่มีค่าไอเก้น มากกว่า 1 ในตารางที่ 14 ร่วมกับการพล็อตกราฟ Scree Plot พบว่าปัจจัยร่วมที่มีค่าไอเก้นมากกว่า 1 ก่อนการการหมุนแกน มีทั้งหมด 8 ปัจจัย (factors) คือ มีค่าไอเก้นเท่ากับ 16.66, 4.04, 3.27, 2.14, 2.21, 2.15, 1.58 และ 1.27 ตามลำดับ ซึ่งสามารถอธิบายความผันแปรของตัวแปรเดิมได้ร้อยละ 74.0

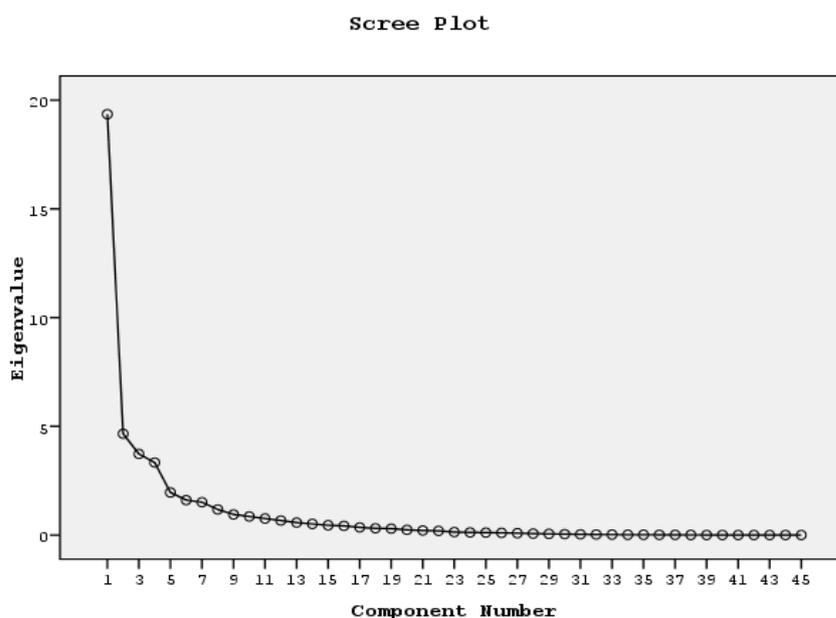
ตารางที่ 14 ผลการวิเคราะห์น้ำหนักปัจจัยก่อนการหมุนแกนแบบ Varimax และร้อยละ
ความผันแปรของแต่ละปัจจัย ที่ได้จากการสกัดปัจจัยแบบ MLM จากตัวแปร
เชิงปริมาณ 45 ตัวแปร

ชื่อตัวแปร	Factor code	Factor							
		1	2	3	4	5	6	7	8
1 RAOB_SURFACE_TEMPERATURE	X111	0.194	-0.804	0.510	-0.234	0.015	0.006	0.003	-0.006
2 CONVECTIVE_TEMPERATURE	X121	-0.683	-0.463	-0.020	0.563	0.003	-0.009	0.001	-0.010
3 CLOUD_BASE_HEIGHT_LCL	X131	-0.728	-0.371	0.284	-0.102	-0.170	0.229	-0.354	0.031
4 CLOUD_BASE_TEMPERATURE_LCL	X132	0.865	-0.046	-0.060	-0.020	0.171	-0.242	0.384	-0.035
5 SFC_TEMP_RISE_REQUIRED_FOR_CCL	X141	-0.743	-0.135	-0.237	0.587	0.036	-0.050	-0.057	-0.102
6 CLOUD_BASE_PRESSURE_CCL	X142	0.856	0.309	-0.019	-0.396	0.020	-0.037	0.004	-0.074
7 CLOUD_BASE_HEIGHT_CCL	X143	-0.860	-0.296	-0.015	0.395	-0.049	0.036	0.007	0.044
8 CLOUD_BASE_TEMPERATURE_CCL	X144	0.960	0.093	-0.019	-0.201	0.058	-0.082	0.020	-0.070
9 SUB_CLOUD_MIXING_RATIO_CCL	X145	0.996	-0.024	-0.009	-0.007	0.034	-0.039	0.059	-0.003
10 PRECIPITABLE_WATER_SFC850	X211	0.924	0.110	0.062	0.030	-0.246	0.014	0.026	-0.113
11 PRECIPITABLE_WATER_SFC700	X212	0.842	0.233	0.372	0.082	-0.272	0.054	0.101	-0.042
12 PRECIPITABLE_WATER_SFC500	X213	0.757	0.380	0.511	0.141	0.022	-0.012	-0.005	0.002
13 PRECIPITABLE_WATER_TOTAL	X214	0.724	0.414	0.501	0.139	0.105	-0.031	-0.029	0.016
14 MEAN_MIXING_RATIO_LOWEST_100_MB	X221	0.977	0.058	0.044	0.023	-0.112	0.031	-0.021	-0.036
15 MEAN_MIXING_RATIO_LOWEST_50_MB	X222	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
16 MEAN_DEW_POINT_TEMP_LOWEST_50_MB	X231	0.995	-0.007	0.007	-0.026	0.002	-0.010	0.009	0.008
17 RH_AT_THE_CCL	X241	0.549	0.459	0.414	-0.332	-0.152	0.013	0.029	-0.054
18 AVG_RH_0_10000_FT	X242	0.696	0.568	0.305	-0.043	-0.173	0.111	0.008	-0.106
19 AVG_RH_10_18000_FT	X243	0.432	0.528	0.535	0.144	0.385	-0.158	-0.159	0.091
20 MEAN_RH_1000_5000_FT	X244	0.766	0.456	-0.047	0.029	-0.087	0.020	0.120	-0.062
21 MEAN_RH_5000_10000_FT	X245	0.473	0.488	0.491	-0.117	-0.172	0.053	0.045	-0.121
22 MEAN_RH_10000_15000_FT	X246	0.426	0.496	0.528	0.106	0.318	-0.120	-0.090	0.053
23 MEAN_RH_20000_25000_FT	X247	0.410	0.462	0.313	0.018	0.294	-0.133	-0.135	0.120
24 LEVEL_OF_FREE_CONVECTION_LFC_wrtTma	X311	0.299	0.203	-0.172	-0.151	-0.008	-0.007	0.284	-0.089

ตารางที่ 14 (ต่อ)

ชื่อตัวแปร	Factor code	Factor	Factor	Factor	Factor	Factor	Factor	Factor	Factor
		1	2	3	4	5	6	7	8
25 TOP_OF_LATENT_INSTABILITY_LAYER_wrtTma	X312	0.139	0.091	-0.073	-0.136	0.029	0.130	0.142	-0.118
26 LEVEL_OF_NEUTRAL_BUOYANCY_LNB	X321	-0.121	0.174	-0.248	-0.281	0.010	0.295	0.254	0.476
27 ISOTHERM_HEIGHTS_0c	X331	-0.039	-0.177	0.123	0.301	-0.200	0.454	0.474	0.522
28 ISOTHERM_HEIGHTS_minus5c	X332	-0.039	-0.117	0.174	0.311	-0.135	0.497	0.476	0.552
29 ISOTHERM_HEIGHTS_minus10c	X333	-0.039	-0.059	0.121	0.341	-0.186	0.385	0.453	0.240
30 ISOTHERM_HEIGHTS_minus15c	X334	0.065	0.058	0.129	0.307	0.024	0.297	0.319	0.155
31 MEAN_WIND_1000_5000_FT_DEG	X411	0.235	-0.268	0.264	0.111	0.259	-0.141	-0.160	0.192
32 MEAN_WIND_5000_10000_FT_DEG	X412	0.179	-0.006	0.228	0.174	0.368	-0.081	-0.220	0.123
33 MEAN_WIND_10000_15000_FT_DEG	X413	0.126	0.012	0.250	0.235	0.344	-0.039	-0.225	0.099
34 MEAN_WIND_20000_25000_FT_DEG	X414	0.026	-0.129	0.012	0.114	0.077	-0.083	-0.153	0.079
35 MEAN_WIND_1000_5000_FT_KTS	X421	-0.084	0.179	0.210	0.213	0.383	0.164	-0.162	0.174
36 MEAN_WIND_5000_10000_FT_KTS	X422	0.048	0.242	0.273	0.053	0.359	0.131	-0.167	0.220
37 MEAN_WIND_10000_15000_FT_KTS	X423	0.046	0.210	0.172	0.038	0.290	0.094	-0.155	0.118
38 MEAN_WIND_20000_25000_FT_KTS	X424	-0.066	-0.023	-0.175	-0.083	-0.100	-0.003	0.078	-0.097
39 LIFTED_INDEX_100_MBAR_LAYER_ADIABATIC	X51	-0.832	0.179	0.031	0.064	0.293	0.374	0.125	-0.134
40 LIFTED_INDEX_50_MBAR_LAYER_MEAN_VALUES	X52	-0.851	0.221	0.005	0.157	0.178	0.367	0.062	0.065
41 SHOWALTER_INDEX	X53	-0.625	0.025	-0.230	0.034	0.518	0.425	0.061	0.022
42 TOTAL_TOTALS_INDEX	X54	0.504	-0.138	0.145	-0.055	-0.533	-0.584	-0.125	-0.023
43 K_INDEX	X55	0.569	0.289	0.548	0.021	-0.040	-0.247	-0.061	-0.006
44 SWEAT_INDEX	X56	0.525	0.249	0.339	0.027	-0.157	-0.057	-0.090	0.236
45 POTENTIAL_BUOYANCY_INDEX	X57	0.537	-0.350	-0.107	0.369	-0.264	-0.450	-0.124	
Eigen values		16.660	4.038	3.266	2.143	2.208	2.146	1.584	1.269
Proportion of total variance		0.370	0.090	0.073	0.048	0.049	0.048	0.035	0.028
Cumulative proportion		0.370	0.460	0.533	0.580	0.629	0.677	0.712	0.740

เมื่อพิจารณาจากกราฟ Scree Plot ดังภาพที่ 18 จะมีลักษณะเดียวกันกับภาพที่ 19 คือพบว่าค่าไอเก็นลดลงอย่างรวดเร็ว หรือเกือบที่จะขนานกับแกนนอน เริ่มจากค่าไอเก็นของปัจจัยที่ 5 ดังนั้นจึงพิจารณาใช้เพียง 4 ปัจจัยร่วม (factors) ที่สามารถอธิบายเกี่ยวกับการเกิดฝนได้



ภาพที่ 19 กราฟ Scree Plot จากการวิเคราะห์ตัวแปรเชิงปริมาณ 45 ตัวแปร ด้วยวิธี MLM

ดังนั้นวิธีการสกัดปัจจัยแบบ MLM ที่ใช้เกณฑ์การเลือกปัจจัยที่ให้ค่าไอเก็นมากกว่า 1 และจากกราฟได้จำนวนปัจจัยที่เหมาะสม 4 ปัจจัยที่ให้ค่าไอเก็นเท่ากับ 16.66, 4.04, 3.27 และ 2.14 ตามลำดับ ซึ่งมีค่าสัดส่วนสะสมของความแปรปรวนที่อธิบายตัวแปรเดิมได้ร้อยละ 58.0

2) ปัจจัยที่สกัดได้จากวิธีการหมุนแกน

ปัจจัยร่วมที่ได้ในตารางที่ 14 ยังไม่สามารถจัดตัวแปรให้อยู่ในปัจจัยใดปัจจัยหนึ่งได้ชัดเจนซึ่งในที่นี้มองปัจจัยเสมือนเป็นแกนในทางพีชคณิต โดยเลือกการหมุนแกนแบบมูมจาก (Varimax rotation) เพื่อทำให้ปัจจัยที่สกัดได้มีความเป็นอิสระต่อกัน แต่จะทำให้ค่าน้ำหนักปัจจัย (Factor loading) เปลี่ยนแปลงไปจากเดิมก่อนที่มีการหมุนแกน เพื่อความสะดวกและง่ายต่อการจัดตัวแปรเดิมให้อยู่ในปัจจัยใหม่ และตั้งชื่อปัจจัยให้สอดคล้องกับกลุ่มตัวแปรเดิมให้มากที่สุด ดังแสดงไว้ในตารางที่ 15

ตารางที่ 15 ผลการวิเคราะห์น้ำหนักปัจจัยหลังการหมุนแกนแบบ Varimax และร้อยละ
ความผันแปรของแต่ละปัจจัย ที่ได้จากการสกัดปัจจัยแบบ MLM จากตัวแปร
เชิงปริมาณ 45 ตัวแปร

ที่	ตัวแปร	code	ค่าน้ำหนักปัจจัย (Loading, l_{im})				ค่าความ ร่วมกัน h_i^2	ค่าปัจจัย เฉพาะ e_i
			Factor 1	Factor 2	Factor 3	Factor 4		
1	MEAN_MIXING_RATIO_LOWEST_50_MB	X222	0.973	0.160	-0.155	0.020	0.997	0.003
2	MEAN_DEW_POINT_TEMP_LOWEST_50_MB	X231	0.972	0.154	-0.149	0.005	0.991	0.009
3	SUB_CLOUD_MIXING_RATIO_CCL	X145	0.968	0.154	-0.181	0.005	0.995	0.005
4	MEAN_MIXING_RATIO_LOWEST_100_MB	X221	0.965	0.150	-0.058	0.100	0.967	0.033
5	CLOUD_BASE_TEMPERATURE_CCL	X144	0.963	0.152	-0.034	-0.188	0.987	0.013
6	PRECIPITABLE_WATER_SFC850	X211	0.942	0.085	0.063	0.172	0.928	0.072
7	CLOUD_BASE_PRESSURE_CCL	X142	0.899	0.145	0.223	-0.342	0.996	0.004
8	CLOUD_BASE_HEIGHT_CCL	X143	-0.894	-0.180	-0.208	0.342	0.993	0.007
9	LIFTED_INDEX_50_MBAR_LAYER_MEAN_VALUES	X52	-0.856	0.032	0.193	0.010	0.771	0.229
10	PRECIPITABLE_WATER_SFC700	X212	0.853	0.285	0.265	0.345	0.999	0.001
11	CLOUD_BASE_TEMPERATURE_LCL	X132	0.846	0.125	-0.232	-0.066	0.789	0.211
12	LIFTED_INDEX_100_MBAR_LAYER_ADIABATIC	X51	-0.842	0.059	0.165	-0.077	0.747	0.253
13	SFC_TEMP_RISE_REQUIRED_FOR_CCL	X141	-0.803	-0.107	-0.320	0.261	0.827	0.173
14	MEAN_RH_1000_5000_FT	X244	0.778	0.226	0.187	-0.017	0.692	0.308
15	CONVECTIVE_TEMPERATURE	X121	-0.755	-0.146	-0.431	0.449	0.979	0.021
16	AVG_RH_0_10000_FT	X242	0.721	0.390	0.482	0.100	0.914	0.086
17	CLOUD_BASE_HEIGHT_LCL	X131	-0.714	-0.195	0.086	0.179	0.587	0.413
18	PRECIPITABLE_WATER_SFC500	X213	0.706	0.631	0.230	0.222	0.999	0.001
19	SHOWALTER_INDEX	X53	-0.674	0.010	-0.164	-0.285	0.562	0.438
20	RH_AT_THE_CCL	X241	0.605	0.318	0.581	-0.049	0.808	0.192
21	TOTAL_TOTALS_INDEX	X54	0.552	-0.100	0.062	0.231	0.372	0.628

ตารางที่ 15 (ต่อ)

ที่	ตัวแปร	ค่าน้ำหนักปัจจัย (Loading, l_{im})				ค่าความ ร่วมกัน	ค่าปัจจัย เฉพาะ	
		code	1	2	3			4
22	SWEAT_INDEX	X56	0.516	0.338	0.226	0.171	0.461	0.539
23	LEVEL_OF_FREE_CONVECTION_LFC_wrtTma	X311	0.342	-0.075	0.142	-0.151	0.165	0.835
24	TOP_OF_LATENT_INSTABILITY_LAYER_wrtTma	X312	0.165	-0.053	0.093	-0.102	0.049	0.951
25	AVG_RH_10_18000_FT	X243	0.321	0.916	0.163	-0.041	0.970	0.030
26	MEAN_RH_10000_15000_FT	X246	0.337	0.824	0.211	-0.006	0.837	0.163
27	PRECIPITABLE_WATER_TOTAL	X214	0.659	0.691	0.205	0.157	0.979	0.021
28	MEAN_RH_20000_25000_FT	X247	0.337	0.672	0.136	-0.154	0.608	0.392
29	K_INDEX	X55	0.539	0.547	0.270	0.167	0.692	0.308
30	MEAN_WIND_5000_10000_FT_KTS	X422	-0.036	0.519	0.020	-0.104	0.282	0.718
31	MEAN_WIND_10000_15000_FT_DEG	X413	0.007	0.491	-0.232	0.041	0.297	0.703
32	MEAN_WIND_5000_10000_FT_DEG	X412	0.063	0.483	-0.255	-0.023	0.303	0.697
33	MEAN_WIND_1000_5000_FT_KTS	X421	-0.186	0.483	-0.082	-0.021	0.275	0.725
34	MEAN_WIND_10000_15000_FT_KTS	X423	-0.025	0.413	-0.007	-0.115	0.184	0.816
35	LEVEL_OF_NEUTRAL_BUOYANCY_LNB	X321	-0.045	-0.260	0.231	-0.244	0.182	0.818
36	MEAN_WIND_20000_25000_FT_KTS	X424	-0.034	-0.206	0.011	-0.044	0.046	0.954
37	MEAN_RH_5000_10000_FT	X245	0.508	0.397	0.579	0.142	0.771	0.229
38	POTENTIAL_BUOYANCY_INDEX	X57	0.481	-0.050	-0.481	0.338	0.578	0.422
39	MEAN_WIND_1000_5000_FT_DEG	X411	0.139	0.314	-0.334	0.071	0.235	0.765
40	RAOB_SURFACE_TEMPERATURE	X111	0.179	-0.087	-0.248	0.233	0.155	0.845
41	MEAN_WIND_20000_25000_FT_DEG	X414	-0.023	0.085	-0.209	0.037	0.053	0.947
42	ISOTHERM_HEIGHTS_0c	X331	-0.040	-0.131	-0.047	0.534	0.306	0.694
43	ISOTHERM_HEIGHTS_minus5c	X332	-0.052	-0.042	-0.024	0.529	0.285	0.715
44	ISOTHERM_HEIGHTS_minus10c	X333	-0.048	-0.046	-0.013	0.513	0.267	0.733

ตารางที่ 15 (ต่อ)

ที่	ตัวแปร	ค่าน้ำหนักปัจจัย (Loading, l_{im})				ค่าความ ร่วมกัน h_i^2	ค่าปัจจัย เฉพาะ e_i	
		Factor code	1	2	3			4
45	ISOTHERM_HEIGHTS_minus15c	X334	0.026	0.142	-0.034	0.342	0.139	0.861
Eigen values			16.600	5.506	2.606	2.307		
Proportion of total variance			0.369	0.122	0.058	0.051		
Cumulative proportion			0.369	0.491	0.549	0.600		

หมายเหตุ 1. คอลัมน์ที่แรเงา คือตัวแปรอิสระที่เป็นองค์ประกอบของแต่ละปัจจัย (Factor)

2. ค่าน้ำหนักปัจจัยที่เป็นตัวเข้ม คือค่า Loading ที่มีค่ามากกว่าหรือเท่ากับ 0.7

3) ความสามารถของความแปรปรวนที่อธิบายได้ (Total Variance Explained)

ปัจจัยที่สกัดได้ทั้ง 4 มีความสามารถในการอธิบายการผันแปรของตัวแปรเดิมทั้ง

45 ตัวแปร คิดเป็นร้อยละ 60.0 โดยที่

ปัจจัยที่ 1 อธิบายการผันแปรได้มากที่สุด มีจำนวนตัวแปรเดิม 24 ตัวแปร คิดเป็นร้อยละ 36.9

ปัจจัยที่ 2 อธิบายการผันแปรของตัวแปรเดิม 12 ตัวแปร คิดเป็นร้อยละ 12.2

ปัจจัยที่ 3 อธิบายการผันแปรของตัวแปรเดิม 5 ตัวแปร คิดเป็นร้อยละ 5.8

ปัจจัยที่ 4 อธิบายการผันแปรของตัวแปรเดิม 4 ตัวแปร คิดเป็นร้อยละ 5.1

4) ประมาณค่าคะแนนปัจจัย (Estimated Factor Score)

ก. กรณีที่ 3 วิเคราะห์ปัจจัยจากตัวแปรปริมาณทั้งหมด 45 ตัวแปร

ด้วยวิธี MLM และคำนวณค่าคะแนนปัจจัยด้วยวิธี Regression

เป็นการสร้างคะแนนปัจจัยจากผลรวมของค่าสัมประสิทธิ์คะแนนปัจจัย

(Component score coefficient matrix) ที่ได้จากรายการที่ 16 คูณด้วยค่าของตัวแปรแต่ละตัวแปร

(X_i) ที่อยู่ในรูปคะแนนมาตรฐาน โดยคำนวณค่าคะแนนปัจจัยเป็นรายวัน สามารถคำนวณได้จาก

$$F_i = \beta'(X_i - \bar{X}_i) + \varepsilon_i$$

กำหนดให้ $i = 1, 2, \dots, 45$ $j = 1, 2, 3, 4$ และ $k = 1, 2, 3, \dots, 277$

นำผลของเมตริกซ์สัมประสิทธิ์คะแนนปัจจัย ที่ได้จากการประมวลผลโปรแกรม SPSS ดังที่ปรากฏในตารางที่ 16 แล้วคำนวณค่าคะแนนปัจจัยด้วยวิธี Regression ได้ดังนี้

ตัวอย่าง การคำนวณค่าคะแนนปัจจัยพื้นฐาน ปัจจัยที่ 1 ถึงปัจจัยที่ 3 ของข้อมูลวันที่ 12 มีนาคม 2547 (วันแรก) ได้ดังนี้

$$F_{11} = (0.001)(0.631) + (-0.018)(0.919) + (-0.001)(1.146) + \dots + (0.002)(0.291) = -0.273$$

$$F_{21} = (-0.001)(0.631) + (0.049)(0.919) + (-0.002)(1.146) + \dots + (-0.001)(0.291) = -1.602$$

$$F_{31} = (-0.004)(0.631) + (-0.249)(0.919) + (0.005)(1.146) + \dots + (-0.016)(0.291) = -0.392$$

$$F_{41} = (0.003)(0.631) + (0.184)(0.919) + (0.004)(1.146) + \dots + (0.008)(0.291) = 0.687$$

สรุป เมื่อสกัดปัจจัยด้วยวิธี MLM แล้วคำนวณค่าคะแนนปัจจัย ด้วย วิธี Regression จะต้องใช้ตัวแปรเดิมทั้งหมด 45 ตัวแปร ในการสร้างตัวแปรใหม่ (ปัจจัยร่วม) 4 ปัจจัย

ตารางที่ 16 เมตริกซ์สัมประสิทธิ์คะแนนปัจจัยร่วม (Component score coefficient matrix)

ที่ได้จากการสกัดปัจจัยแบบ MLM จากตัวแปรเชิงปริมาณ 45 ตัวแปร

ที่	ชื่อตัวแปร	code	Factor 1	Factor 2	Factor 3	Factor 4
1	RAOB_SURFACE_TEMPERATURE	X111	0.001	-0.001	-0.004	0.003
2	CONVECTIVE_TEMPERATURE	X121	-0.018	0.049	-0.249	0.184
3	CLOUD_BASE_HEIGHT_LCL	X131	-0.001	-0.002	0.005	0.004
4	CLOUD_BASE_TEMPERATURE_LCL	X132	0.004	0.003	-0.018	-0.001
5	SFC_TEMP_RISE_REQUIRED_FOR_CCL	X141	-0.003	0.008	-0.022	0.014
6	CLOUD_BASE_PRESSURE_CCL	X142	0.152	-0.157	0.575	-0.680
7	CLOUD_BASE_HEIGHT_CCL	X143	-0.080	0.045	-0.284	0.404
8	CLOUD_BASE_TEMPERATURE_CCL	X144	0.063	0.002	-0.091	-0.112
9	SUB_CLOUD_MIXING_RATIO_CCL	X145	0.164	0.038	-0.607	0.016
10	PRECIPITABLE_WATER_SFC850	X211	0.013	-0.024	-0.001	0.015
11	PRECIPITABLE_WATER_SFC700	X212	0.462	-1.411	1.438	1.527
12	PRECIPITABLE_WATER_SFC500	X213	-0.322	1.799	-0.308	-0.152
13	PRECIPITABLE_WATER_TOTAL	X214	-0.023	0.120	-0.037	-0.034

ตารางที่ 16 (ต่อ)

ที่	ชื่อตัวแปร	code	Factor 1	Factor 2	Factor 3	Factor 4
14	MEAN_MIXING_RATIO_LOWEST_100_MB	X221	0.027	-0.018	-0.056	0.018
15	MEAN_MIXING_RATIO_LOWEST_50_MB	X222	0.277	0.027	-0.945	0.046
16	MEAN_DEW_POINT_TEMP_LOWEST_50_MB	X231	0.100	0.006	-0.327	0.006
17	RH_AT_THE_CCL	X241	0.000	-0.006	0.031	-0.009
18	AVG_RH_0_10000_FT	X242	0.000	-0.009	0.048	-0.009
19	AVG_RH_10_18000_FT	X243	-0.043	0.181	-0.064	-0.088
20	MEAN_RH_1000_5000_FT	X244	0.001	-0.002	0.004	-0.002
21	MEAN_RH_5000_10000_FT	X245	-0.001	-0.004	0.024	-0.003
22	MEAN_RH_10000_15000_FT	X246	-0.007	0.027	-0.006	-0.014
23	MEAN_RH_20000_25000_FT	X247	-0.002	0.010	-0.003	-0.007
24	LEVEL_OF_FREE_CONVECTION_LFC_wrtTma	X311	0.001	-0.002	0.003	-0.001
25	TOP_OF_LATENT_INSTABILITY_LAYER_wrtTma	X312	0.000	-0.001	0.002	-0.001
26	LEVEL_OF_NEUTRAL_BUOYANCY_LNB	X321	0.000	-0.003	0.005	-0.002
27	ISOTHERM_HEIGHTS_0c	X331	0.000	-0.003	-0.001	0.006
28	ISOTHERM_HEIGHTS_minus5c	X332	0.000	-0.002	-0.001	0.005
29	ISOTHERM_HEIGHTS_minus10c	X333	0.000	-0.002	-0.001	0.005
30	ISOTHERM_HEIGHTS_minus15c	X334	0.000	0.000	-0.002	0.002
31	MEAN_WIND_1000_5000_FT_DEG	X411	0.000	0.004	-0.008	0.000
32	MEAN_WIND_5000_10000_FT_DEG	X412	-0.001	0.007	-0.008	-0.001
33	MEAN_WIND_10000_15000_FT_DEG	X413	-0.001	0.006	-0.007	-0.001
34	MEAN_WIND_20000_25000_FT_DEG	X414	0.000	0.002	-0.003	0.000
35	MEAN_WIND_1000_5000_FT_KTS	X421	-0.001	0.006	-0.004	-0.002
36	MEAN_WIND_5000_10000_FT_KTS	X422	-0.001	0.006	-0.003	-0.003
37	MEAN_WIND_10000_15000_FT_KTS	X423	-0.001	0.004	-0.002	-0.002
38	MEAN_WIND_20000_25000_FT_KTS	X424	0.000	-0.001	0.001	0.000
39	LIFTED_INDEX_100_MBAR_LAYER_ADIABATIC	X51	-0.004	0.005	0.009	-0.004
40	LIFTED_INDEX_50_MBAR_LAYER_MEAN_VALUES	X52	-0.005	0.004	0.011	-0.002
41	SHOWALTER_INDEX	X53	-0.002	0.006	-0.004	-0.004

ตารางที่ 16 (ต่อ)

ที่	ชื่อตัวแปร	code	Factor 1	Factor 2	Factor 3	Factor 4
42	TOTAL_TOTALS_INDEX	X54	0.001	-0.004	0.001	0.003
43	K_INDEX	X55	-0.001	0.005	0.002	-0.001
44	SWEAT_INDEX	X56	0.000	0.000	0.002	0.000
45	POTENTIAL_BUOYANCY_INDEX	X57	0.002	0.000	-0.016	0.008

ข. กรณีที่ 4 วิเคราะห์ปัจจัยจากตัวแปรปริมาณทั้งหมด 45 ตัวแปร

ด้วยวิธี MLM และคำนวณค่าคะแนนปัจจัยด้วยวิธี Factor Based Score

สร้างสมการคำนวณค่าคะแนนพื้นฐานของปัจจัย ด้วยการรวมคะแนน

เฉพาะตัวแปรที่อยู่ในปัจจัยเดียวกัน ที่มีค่าน้ำหนักปัจจัย (l_{ij}) ในตารางที่ 15 มากกว่า 0.70 เท่านั้น และตัวแปรเหล่านี้ผ่านการแปลงค่าข้อมูลให้อยู่ในรูปคะแนนมาตรฐานแล้ว

ตัวอย่าง การคำนวณค่าคะแนนปัจจัยพื้นฐาน ปัจจัยที่ 1 ถึงปัจจัยที่ 4 ของข้อมูลวันที่ 12 มีนาคม 2547 (วันแรก) ได้ดังนี้

$$F_{11} = X_{222} + X_{231} + X_{145} + X_{221} + X_{144} + X_{211} + X_{142} + X_{143} + X_{52} \\ + X_{212} + X_{132} + X_{51} + X_{141} + X_{244} + X_{121} + X_{242} + X_{131} + X_{213} \\ = -3.143$$

$$F_{21} = X_{243} + X_{246} = -3.597$$

$$F_{31} = \text{ตัวแปร } X_i \text{ ที่จัดอยู่ในปัจจัยนี้ทุกตัวมีค่า loading น้อยกว่า 0.7 (ไม่นำมาวิเคราะห์)}$$

$$F_{41} = \text{ตัวแปร } X_i \text{ ที่จัดอยู่ในปัจจัยนี้ทุกตัวมีค่า loading น้อยกว่า 0.7 (ไม่นำมาวิเคราะห์)}$$

สรุป เมื่อสกัดปัจจัยด้วยวิธี MLM แล้วคำนวณค่าคะแนนปัจจัย ด้วยวิธีสร้างค่าคะแนนพื้นฐานของปัจจัยจะใช้ตัวแปรเดิม 20 ตัวแปรในการสร้างตัวแปรใหม่ (ปัจจัยรวม) 2 ปัจจัย

ตารางที่ 17 แสดงตัวอย่างการคำนวณค่าคะแนนปัจจัยโดยใช้วิธี Regression (กรณีที่ 3) และ Factor Based Score (กรณีที่ 4) เมื่อใช้วิธีสกัดปัจจัยด้วยเทคนิค MLM

ที่	วัน	Regression (กรณี 3)				Factor Based Score (กรณี 4)			
		Fac1	Fac2	Fac3	Fac4	Fac1	Fac2	Fac3	Fac4
1	12 มี.ค.47	-0.273	-1.602	-0.392	0.687	-3.143	-3.597	-	-
2	13 มี.ค.47	0.217	-2.362	1.204	1.600	-2.006	-3.458	-	-
3	14 มี.ค.47	0.343	-2.040	-0.619	0.405	-1.913	-2.808	-	-
4	15 มี.ค.47	0.238	-2.894	0.926	0.695	-2.415	-5.603	-	-
5	16 มี.ค.47	1.318	-2.916	-0.298	-0.349	0.363	-4.114	-	-

2.2 ผลการวิเคราะห์ปัจจัย (กรณีที่ 5 – 8)

ในส่วนนี้ใช้ตัวแปรอิสระเชิงปริมาณ 28 ตัวแปร ที่มีค่า VIF มากกว่าหรือเท่ากับ 10 ในการวิเคราะห์ปัจจัย แต่เนื่องจากข้อมูลที่เป็นตัวแปรเชิงปริมาณ มีหน่วยที่เหมือนและไม่เหมือนกัน ดังนั้นจึงต้องทำการแปลงข้อมูลให้อยู่ในรูปคะแนนมาตรฐาน ที่จะทำการวิเคราะห์ปัจจัยต่อไป โดยใช้วิธีการสกัดปัจจัย 2 วิธี คือ PCM และ MLM ดังนี้

2.2.1 ผลการตรวจสอบความเหมาะสมของการใช้เทคนิคการวิเคราะห์ปัจจัยกับชุดข้อมูล 28 ตัวแปร

จากการตรวจสอบความเหมาะสมของข้อมูลที่ใช้เทคนิคการวิเคราะห์ปัจจัยด้วยสถิติ KMO และ Bartlett's Test of Sphericity พบว่าได้ค่า KMO เท่ากับ 0.848 อยู่ในระดับดี และ ค่า Bartlett's Test of Sphericity มีค่า P-value น้อยกว่า 0.001 ซึ่งทำให้ไม่มีหลักฐานเพียงพอที่จะปฏิเสธสมมติฐานหลักที่ว่าข้อมูลชุดดังกล่าวสามารถวิเคราะห์ปัจจัยได้ และผลจากตัวทดสอบสถิติทั้งสองมีความสอดคล้องกัน จึงสรุปได้ว่าข้อมูลตัวแปรอิสระทั้ง 28 ตัวแปร มีความเหมาะสมที่จะวิเคราะห์ด้วยเทคนิคการวิเคราะห์ปัจจัย ดังตารางที่ 18

ตารางที่ 18 ผลการตรวจสอบความเหมาะสมของตัวแปรเชิงปริมาณ 28 ตัวแปร (กรณี 5 – 8)
ใช้เทคนิคการวิเคราะห์ปัจจัยด้วยค่าสถิติ KMO และค่าสถิติ Bartlett's Test of Sphericity

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.848
Bartlett's Test of Sphericity	Approx. Chi-Square	22,485.815
	Df.	378
	Sig.	0.001**

หมายเหตุ ระดับนัยสำคัญของการทดสอบที่ 0.05 ($\alpha = 0.05$)

1) จำนวนปัจจัยร่วมที่เหมาะสม

พิจารณาจากปัจจัยร่วมที่มีค่าไอเก้น มากกว่า 1 ในตารางที่ 19 ร่วมกับการพล็อตกราฟ Scree Plot พบว่าปัจจัยร่วมที่มีค่าไอเก้นมากกว่า 1 ก่อนการหมุนแกน มีทั้งหมด 5 ปัจจัย (factors) คือ มีค่าไอเก้นเท่ากับ 16.435, 3.157, 3.016, 1.632 และ 1.282 ตามลำดับ ซึ่งสามารถอธิบายความผันแปรของตัวแปรเดิมได้ร้อยละ 91.1

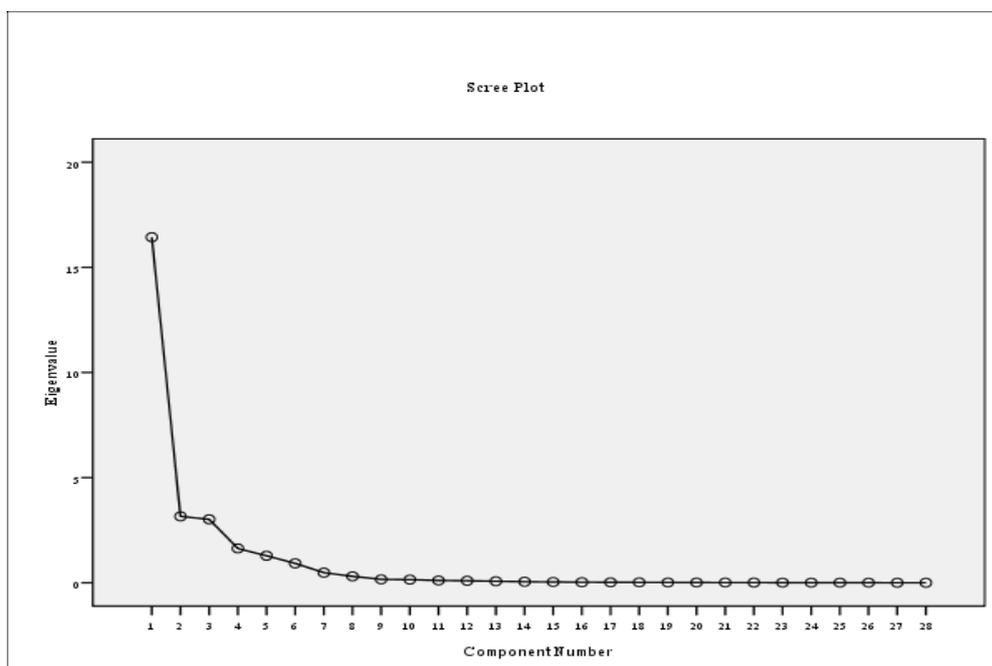
ตารางที่ 19 ผลการวิเคราะห์น้ำหนักปัจจัยก่อนการหมุนแกนแบบ Varimax และร้อยละความผันแปรของแต่ละปัจจัย ที่ได้จากการสกัดปัจจัยแบบ PCM จากตัวแปรเชิงปริมาณ 28 ตัวแปร

ชื่อตัวแปร	code	Factor	Factor	Factor	Factor	Factor	
		1	2	3	4	5	
1	RAOB_SURFACE_TEMPERATURE	X111	0.121	-0.509	0.176	0.299	-0.697
2	CONVECTIVE_TEMPERATURE	X121	-0.796	-0.273	0.256	0.265	0.234
3	CLOUD_BASE_HEIGHT_LCL	X131	-0.746	-0.149	-0.009	0.357	-0.421
4	CLOUD_BASE_TEMPERATURE_LCL	X132	0.840	-0.139	0.101	-0.235	0.078
5	SFC_TEMP_RISE_REQUIRED_FOR_CCL	X141	-0.818	-0.072	0.064	0.133	0.482
6	CLOUD_BASE_PRESSURE_CCL	X142	0.916	0.123	-0.198	-0.251	-0.144
7	CLOUD_BASE_HEIGHT_CCL	X143	-0.919	-0.132	0.184	0.227	0.169
8	CLOUD_BASE_TEMPERATURE_CCL	X144	0.960	-0.051	-0.074	-0.193	-0.079
9	SUB_CLOUD_MIXING_RATIO_CCL	X145	0.950	-0.130	0.104	-0.140	-0.003

ตารางที่ 19 (ต่อ)

	ชื่อตัวแปร	code	Factor	Factor	Factor	Factor	Factor
			1	2	3	4	5
10	PRECIPITABLE_WATER_SFC850	X211	0.933	-0.092	0.111	-0.053	0.064
11	PRECIPITABLE_WATER_SFC700	X212	0.903	0.104	0.251	0.182	-0.024
12	PRECIPITABLE_WATER_SFC500	X213	0.851	0.349	0.158	0.338	-0.005
13	PRECIPITABLE_WATER_TOTAL	X214	0.823	0.399	0.115	0.337	0.014
14	MEAN_MIXING_RATIO_LOWEST_100_MB	X221	0.956	-0.087	0.115	-0.077	0.021
15	MEAN_MIXING_RATIO_LOWEST_50_MB	X222	0.955	-0.114	0.102	-0.122	-0.006
16	MEAN_DEW_POINT_TEMP_LOWEST_50_MB	X231	0.955	-0.118	0.098	-0.129	-0.027
17	AVG_RH_0_10000_FT	X242	0.820	0.420	0.023	0.107	0.024
18	AVG_RH_10_18000_FT	X243	0.565	0.592	-0.061	0.467	0.053
19	MEAN_RH_1000_5000_FT	X244	0.832	0.245	0.046	-0.176	0.258
20	MEAN_RH_10000_15000_FT	X246	0.559	0.586	-0.032	0.451	0.009
21	ISOTHERM_HEIGHTS_0c	X331	-0.134	0.007	0.930	-0.107	-0.068
22	ISOTHERM_HEIGHTS_minus5c	X332	-0.125	0.117	0.949	-0.078	-0.087
23	ISOTHERM_HEIGHTS_minus10c	X333	-0.106	0.084	0.921	-0.057	0.040
24	LIFTED_INDEX_100_MBAR_LAYER_ADIABATIC	X51	-0.827	0.512	0.029	-0.068	-0.064
25	LIFTED_INDEX_50_MBAR_LAYER_MEAN_VALUES	X52	-0.842	0.489	0.027	-0.019	0.040
26	SHOWALTER_INDEX	X53	-0.719	0.417	0.041	-0.347	-0.081
27	TOTAL_TOTALS_INDEX	X54	0.591	-0.584	-0.109	0.371	0.120
28	POTENTIAL_BUOYANCY_INDEX	X57	0.472	-0.651	0.181	0.226	0.392
Eigen values			16.435	3.157	3.016	1.632	1.282
Proportion of total variance			0.587	0.113	0.108	0.058	0.046
Cumulative proportion			0.587	0.700	0.807	0.866	<u>0.911</u>

เมื่อพิจารณาจากกราฟ Scree Plot ดังภาพที่ 20 ค่าไอเกนลดลงอย่างรวดเร็ว หรือเกือบที่จะขนานกับแกนนอน เริ่มจากค่าไอเกนของปัจจัยที่ 4 ดังนั้นจึงพิจารณาใช้เพียง 3 ปัจจัยรวม (factors) ที่สามารถอธิบายเกี่ยวกับการเกิดฝนได้



ภาพที่ 20 กราฟ Scree Plot จากการวิเคราะห์ตัวแปรเชิงปริมาณ 28 ตัวแปร ด้วยวิธี PCM

ดังนั้นวิธีการสกัดปัจจัยแบบ PCM ที่ใช้เกณฑ์การเลือกปัจจัยที่ให้ค่าไอเก้นมากกว่า 1 และจากกราฟได้จำนวนปัจจัยที่เหมาะสม 4 ปัจจัยที่ให้ค่าไอเก้นเท่ากับ 16.435, 3.157 และ 3.016 ตามลำดับ ซึ่งมีค่าสัดส่วนสะสมของความแปรปรวนที่อธิบายตัวแปรเดิมได้ร้อยละ 80.7

2) ปัจจัยที่สกัดได้จากวิธีการหมุนแกน

ปัจจัยร่วมที่ได้ในตารางที่ 20 ยังไม่สามารถจัดตัวแปรให้อยู่ในปัจจัยใดปัจจัยหนึ่งได้ชัดเจนซึ่งในที่นี้มองปัจจัยเสมือนเป็นแกนในทางพีชคณิต โดยเลือกการหมุนแกนแบบมุมฉาก (Varimax rotation) เพื่อทำให้ปัจจัยที่สกัดได้มีความเป็นอิสระต่อกัน แต่จะทำให้ค่าน้ำหนักปัจจัย (Factor loading) เปลี่ยนแปลงไปจากเดิมก่อนที่มีการหมุนแกน เพื่อความสะดวกและง่ายต่อการจัดตัวแปรเดิมให้อยู่ในปัจจัยใหม่ และตั้งชื่อปัจจัยให้สอดคล้องกับกลุ่มตัวแปรเดิมให้มากที่สุด ดังแสดงไว้ในตารางที่ 20

ตารางที่ 20 ผลการวิเคราะห์น้ำหนักปัจจัยหลังการหมุนแกนแบบ Varimax และร้อยละ
ความผันแปรของแต่ละปัจจัย ที่ได้จากการสกัดปัจจัยแบบ PCM จากตัวแปร
เชิงปริมาณ 28 ตัวแปร

ที่	ตัวแปร	ค่าน้ำหนักปัจจัย (Loading, l_{im})			ค่าความ ร่วมกัน h_i^2	ค่า ปัจจัย เฉพาะ e_i	
		Factor code	Factor 1	Factor 2			Factor 3
1	AVG_RH_0_10000_FT	X242	0.914	0.120	-0.017	0.850	0.150
2	PRECIPITABLE_WATER_TOTAL	X214	0.907	0.148	0.073	0.850	0.150
3	PRECIPITABLE_WATER_SFC500	X213	0.902	0.210	0.109	0.870	0.130
4	MEAN_RH_1000_5000_FT	X244	0.825	0.273	-0.007	0.755	0.245
5	CLOUD_BASE_HEIGHT_CCL	X143	-0.825	-0.391	0.251	0.896	0.104
6	CLOUD_BASE_PRESSURE_CCL	X142	0.817	0.395	-0.265	0.894	0.106
7	PRECIPITABLE_WATER_SFC700	X212	0.808	0.450	0.181	0.889	0.111
8	CONVECTIVE_TEMPERATURE	X121	-0.802	-0.198	0.302	0.773	0.227
9	AVG_RH_10_18000_FT	X243	0.799	-0.173	-0.068	0.673	0.327
10	MEAN_RH_10000_15000_FT	X246	0.792	-0.168	-0.039	0.657	0.343
11	CLOUD_BASE_TEMPERATURE_CCL	X144	0.758	0.575	-0.157	0.930	0.070
12	MEAN_MIXING_RATIO_LOWEST_100_MB	X221	0.740	0.622	0.028	0.935	0.065
13	MEAN_MIXING_RATIO_LOWEST_50_MB	X222	0.723	0.642	0.014	0.936	0.064
14	MEAN_DEW_POINT_TEMP_LOWEST_50_MB	X231	0.720	0.645	0.010	0.935	0.065
15	PRECIPITABLE_WATER_SFC850	X211	0.718	0.612	0.026	0.891	0.109
16	SFC_TEMP_RISE_REQUIRED_FOR_CCL	X141	-0.711	-0.395	0.127	0.678	0.322
17	SUB_CLOUD_MIXING_RATIO_CCL	X145	0.710	0.653	0.015	0.930	0.070
18	CLOUD_BASE_HEIGHT_LCL	X131	-0.698	-0.299	0.043	0.578	0.422
19	CLOUD_BASE_TEMPERATURE_LCL	X132	0.614	0.598	0.021	0.735	0.265
20	LIFTED_INDEX_100_MBAR_LAYER_ADIABATIC	X51	-0.388	-0.883	0.133	0.948	0.052
21	LIFTED_INDEX_50_MBAR_LAYER_MEAN_VALUES	X52	-0.413	-0.872	0.130	0.948	0.052

ตารางที่ 20 (ต่อ)

ที่	ตัวแปร	ค่าน้ำหนักปัจจัย (Loading, l_{im})			ค่าความ ร่วมกัน h_i^2	ค่า ปัจจัย เฉพาะ e_i	
		Factor code	Factor 1	Factor 2			Factor 3
22	POTENTIAL_BUOYANCY_INDEX	X57	0.023	0.818	0.096	0.679	0.321
23	TOTAL_TOTALS_INDEX	X54	0.150	0.800	-0.197	0.702	0.298
24	SHOWALTER_INDEX	X53	-0.353	-0.743	0.129	0.693	0.307
25	RAOB_SURFACE_TEMPERATURE	X111	-0.185	0.503	0.131	0.305	0.695
26	ISOTHERM_HEIGHTS_minus5c	X332	-0.008	-0.069	0.961	0.929	0.071
27	ISOTHERM_HEIGHTS_0c	X331	-0.078	0.015	0.936	0.882	0.118
28	ISOTHERM_HEIGHTS_minus10c	X333	-0.011	-0.034	0.931	0.868	0.132
Eigen values			12.124	7.374	3.110		
Proportion of total variance			0.433	0.263	0.111		
Cumulative proportion			0.433	0.696	0.807		

หมายเหตุ 1. คอลัมน์ที่แรเงาคือตัวแปรอิสระที่เป็นองค์ประกอบของแต่ละปัจจัย (Factor)

2. ค่าน้ำหนักปัจจัยที่เป็นตัวเข้ม คือค่า Loading ที่มีค่ามากกว่าหรือเท่ากับ 0.7

3) ความสามารถของความแปรปรวนที่อธิบายได้ (Total Variance Explained)

ปัจจัยที่สกัดได้ทั้ง 3 มีความสามารถในการอธิบายการผันแปรของตัวแปรเดิมทั้ง 28 ตัวแปร คิดเป็นร้อยละ 80.74 โดยที่

ปัจจัยที่ 1 อธิบายการผันแปรได้มากที่สุด มีจำนวนตัวแปรเดิม 19 ตัวแปร คิดเป็นร้อยละ 43.3

ปัจจัยที่ 2 อธิบายการผันแปรของตัวแปรเดิม 6 ตัวแปร คิดเป็นร้อยละ 26.3

ปัจจัยที่ 3 อธิบายการผันแปรของตัวแปรเดิม 3 ตัวแปร คิดเป็นร้อยละ 11.1

4) ประมาณค่าคะแนนปัจจัย (Estimated Factor Score)

ก. กรณีที่ 5 วิเคราะห์ปัจจัยจากตัวแปรปริมาณทั้งหมด 28 ตัวแปร ที่มีปัญหา

Multicollinearity ด้วยวิธี PCM และคำนวณค่าคะแนนปัจจัยด้วยวิธี

Regression

เป็นการสร้างคะแนนปัจจัยจากผลรวมของค่าสัมประสิทธิ์คะแนนปัจจัย

(Component score coefficient matrix) ที่ได้จากรายที่ 21 คูณด้วยค่าของตัวแปรแต่ละตัวแปร

(X_i) ที่อยู่ในรูปคะแนนมาตรฐาน โดยคำนวณค่าคะแนนปัจจัยเป็นรายวัน สามารถคำนวณได้จาก

$$F_i = \beta'(X_i - \bar{X}_i) + \varepsilon_i$$

กำหนดให้ $i = 1, 2, \dots, 28$ $j = 1, 2, 3$ และ $k = 1, 2, 3, \dots, 277$

นำผลของเมตริกซ์สัมประสิทธิ์คะแนนปัจจัย ที่ได้จากการประมวลผลโปรแกรม SPSS ดังที่ปรากฏในตารางที่ 21 แล้วคำนวณค่าคะแนนปัจจัยด้วยวิธี Regression ได้ดังนี้

ตัวอย่าง การคำนวณค่าคะแนนปัจจัยพื้นฐาน ปัจจัยที่ 1 ถึงปัจจัยที่ 3 ของข้อมูลวันที่ 12 มีนาคม 2547 (วันแรก) ได้ดังนี้

$$F_{11} = (-0.084)(0.631) + (-0.086)(0.919) + (-0.064)(1.146) + \dots + (-0.092)(0.291) = -1.317$$

$$F_{21} = (0.142)(0.631) + (0.052)(0.919) + (0.013)(1.146) + \dots + (0.191)(0.291) = 0.630$$

$$F_{31} = (0.047)(0.631) + (0.083)(0.919) + (-0.002)(1.146) + \dots + (0.043)(0.291) = 0.490$$

สรุป เมื่อสกัดปัจจัยด้วยวิธี PCM แล้วคำนวณค่าคะแนนปัจจัย ด้วย วิธี Regression จะต้องใช้ตัวแปรเดิมทั้งหมด 28 ตัวแปร ในการสร้างตัวแปรใหม่ (ปัจจัยรวม) 3 ปัจจัย

ตารางที่ 21 เมตริกซ์สัมประสิทธิ์คะแนนปัจจัยรวม (Component score coefficient matrix)

ที่ได้จากการสกัดปัจจัยแบบ PCM จากตัวแปรเชิงปริมาณ 28 ตัวแปร

ที่	ชื่อตัวแปร	code	Factor 1	Factor 2	Factor 3
1	RAOB_SURFACE_TEMPERATURE	X111	-0.084	0.142	0.047
2	CONVECTIVE_TEMPERATURE	X121	-0.086	0.052	0.083
3	CLOUD_BASE_HEIGHT_LCL	X131	-0.064	0.013	-0.002
4	CLOUD_BASE_TEMPERATURE_LCL	X132	0.018	0.068	0.026
5	SFC_TEMP_RISE_REQUIRED_FOR_CCL	X141	-0.053	-0.007	0.024
6	CLOUD_BASE_PRESSURE_CCL	X142	0.066	-0.007	-0.067
7	CLOUD_BASE_HEIGHT_CCL	X143	-0.068	0.009	0.063
8	CLOUD_BASE_TEMPERATURE_CCL	X144	0.038	0.044	-0.030

ตารางที่ 21 (ต่อ)

ที่	ชื่อตัวแปร	code	Factor 1	Factor 2	Factor3
9	SUB_CLOUD_MIXING_RATIO_CCL	X145	0.025	0.070	0.027
10	PRECIPITABLE_WATER_SFC850	X211	0.031	0.060	0.030
11	PRECIPITABLE_WATER_SFC700	X212	0.066	0.013	0.080
12	PRECIPITABLE_WATER_SFC500	X213	0.107	-0.056	0.055
13	PRECIPITABLE_WATER_TOTAL	X214	0.114	-0.072	0.042
14	MEAN_MIXING_RATIO_LOWEST_100_MB	X221	0.033	0.059	0.031
15	MEAN_MIXING_RATIO_LOWEST_50_MB	X222	0.028	0.066	0.026
16	MEAN_DEW_POINT_TEMP_LOWEST_50_MB	X231	0.027	0.067	0.025
17	AVG_RH_0_10000_FT	X242	0.117	-0.080	0.012
18	AVG_RH_10_18000_FT	X243	0.134	-0.136	-0.010
19	MEAN_RH_1000_5000_FT	X244	0.086	-0.034	0.016
20	MEAN_RH_10000_15000_FT	X246	0.133	-0.134	-0.001
21	ISOTHERM_HEIGHTS_0c	X331	0.004	0.025	0.307
22	ISOTHERM_HEIGHTS_minus5c	X332	0.024	-0.002	0.316
23	ISOTHERM_HEIGHTS_minus10c	X333	0.019	0.006	0.306
24	LIFTED_INDEX_100_MBAR_LAYER_ADIABATIC	X51	0.051	-0.160	0.025
25	LIFTED_INDEX_50_MBAR_LAYER_MEAN_VALUES	X52	0.046	-0.155	0.023
26	SHOWALTER_INDEX	X53	0.040	-0.131	0.026
27	TOTAL_TOTALS_INDEX	X54	-0.077	0.168	-0.051
28	POTENTIAL_BUOYANCY_INDEX	X57	-0.092	0.191	0.043

ข. กรณีที่ 6 วิเคราะห์ปัจจัยจากตัวแปรปริมาณทั้งหมด 28 ตัวแปร ที่มีปัญหา

Multicollinearity ด้วยวิธี MLM และคำนวณค่าคะแนนปัจจัยด้วยวิธี

Factor Based Score

สร้างสมการคำนวณค่าคะแนนพื้นฐานของปัจจัย ด้วยการรวมคะแนน

เฉพาะตัวแปรที่อยู่ในปัจจัยเดียวกัน ที่มีค่าน้ำหนักปัจจัย (t_{ij}) ในตารางที่ 20 มากกว่า 0.70 เท่านั้น และตัวแปรเหล่านี้ผ่านการแปลงค่าข้อมูลให้อยู่ในรูปคะแนนมาตรฐานแล้ว

ตัวอย่าง การคำนวณค่าคะแนนปัจจัยพื้นฐาน ปัจจัยที่ 1 ถึงปัจจัยที่ 3 ของข้อมูลวันที่ 12 มีนาคม 2547 (วันแรก) ได้ดังนี้

$$F_{11} = X_{242} + X_{214} + X_{213} + X_{244} + X_{143} + X_{142} + X_{212} + X_{121} + X_{243} + X_{246} + X_{144} + X_{221} + X_{222} + X_{231} + X_{211} + X_{141} + X_{145} = -8.537$$

$$F_{21} = X_{51} + X_{52} + X_{57} + X_{54} + X_{53} = 0.787$$

$$F_{31} = X_{332} + X_{331} + X_{333} = 1.400$$

สรุป เมื่อสกัดปัจจัยด้วยวิธี PCM แล้วคำนวณค่าคะแนนปัจจัย ด้วยวิธีสร้างค่าคะแนนพื้นฐานของปัจจัยจะใช้ตัวแปรเดิม 25 ตัวแปรในการสร้างตัวแปรใหม่ (ปัจจัยร่วม) 3 ปัจจัย

ตารางที่ 22 แสดงตัวอย่างการคำนวณค่าคะแนนปัจจัยโดยใช้วิธี Regression (กรณีที่ 5) และ Factor Based Score (กรณีที่ 6) เมื่อใช้วิธีสกัดปัจจัยด้วยเทคนิค PCM

ที่	วัน	Regression (กรณี 5)			Factor Based Score (กรณี 6)		
		Fac1	Fac2	Fac3	Fac1	Fac2	Fac3
1	12 มี.ค.47	-1.317	0.630	0.490	-8.537	0.787	1.400
2	13 มี.ค.47	-1.041	1.114	0.390	-6.102	-1.060	1.373
3	14 มี.ค.47	-1.101	1.269	0.360	-5.598	-0.390	1.142
4	15 มี.ค.47	-1.281	1.159	0.149	-9.365	-0.378	1.030
5	16 มี.ค.47	-0.675	1.793	-0.186	-2.353	-1.066	0.010

2.1.4 ผลการสกัดปัจจัยแบบ MLM (กรณีที่ 7-8)

1) จำนวนปัจจัยร่วมที่เหมาะสม

พิจารณาจากปัจจัยร่วมที่มีค่าไอเก้น มากกว่า 1 ในตารางที่ 23 ร่วมกับการพล็อตกราฟ Scree plot พบว่าปัจจัยร่วมที่มีค่าไอเก้นมากกว่า 1 ก่อนการหมุนแกน มีทั้งหมด 5 ปัจจัย (factors) คือ มีค่าไอเก้นเท่ากับ 12.753, 3.453, 4.411, 1.963 และ 1.102 ตามลำดับ ซึ่งสามารถอธิบายความผันแปรของตัวแปรเดิมได้ร้อยละ 84.6

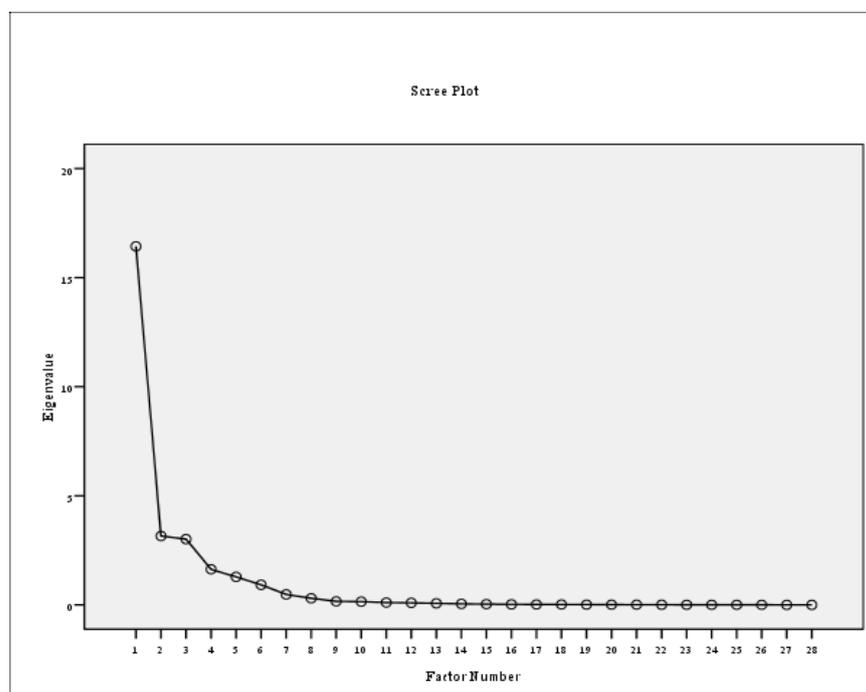
ตารางที่ 23 ผลการวิเคราะห์น้ำหนักปัจจัยก่อนการหมุนแกนแบบ Varimax และร้อยละ
ความผันแปรของแต่ละปัจจัย ที่ได้จากการสกัดปัจจัยแบบ MLM จากตัวแปร
เชิงปริมาณ 28 ตัวแปร

	ชื่อตัวแปร	Code	Factor1	Factor2	Factor3	Factor4	Factor5
1	RAOB_SURFACE_TEMPERATURE	X111	0.211	-0.195	0.032	0.319	-0.036
2	CONVECTIVE_TEMPERATURE	X121	-0.564	-0.283	-0.532	0.538	-0.076
3	CLOUD_BASE_HEIGHT_LCL	X131	-0.502	-0.272	-0.502	-0.011	0.118
4	CLOUD_BASE_TEMPERATURE_LCL	X132	0.628	0.157	0.562	0.188	-0.133
5	SFC_TEMP_RISE_REQUIRED_FOR_CCL	X141	-0.636	-0.221	-0.496	0.333	-0.109
6	CLOUD_BASE_PRESSURE_CCL	X142	0.675	0.255	0.614	-0.312	0.018
7	CLOUD_BASE_HEIGHT_CCL	X143	-0.680	-0.277	-0.602	0.302	0.010
8	CLOUD_BASE_TEMPERATURE_CCL	X144	0.735	0.211	0.631	-0.026	-0.065
9	SUB_CLOUD_MIXING_RATIO_CCL	X145	0.749	0.214	0.582	0.215	-0.050
10	PRECIPITABLE_WATER_SFC850	X211	0.808	0.179	0.429	0.140	0.204
11	PRECIPITABLE_WATER_SFC700	X212	0.878	0.341	0.150	0.090	0.276
12	PRECIPITABLE_WATER_SFC500	X213	0.836	0.548	-0.005	0.000	0.000
13	PRECIPITABLE_WATER_TOTAL	X214	0.796	0.580	-0.015	-0.031	-0.074
14	MEAN_MIXING_RATIO_LOWEST_100_MB	X221	0.789	0.234	0.500	0.189	0.081
15	MEAN_MIXING_RATIO_LOWEST_50_MB	X222	0.760	0.228	0.570	0.208	-0.025
16	MEAN_DEW_POINT_TEMP_LOWEST_50_MB	X231	0.756	0.222	0.577	0.194	-0.028
17	AVG_RH_0_10000_FT	X242	0.738	0.494	0.127	-0.225	0.254
18	AVG_RH_10_18000_FT	X243	0.573	0.643	-0.213	-0.187	-0.393
19	MEAN_RH_1000_5000_FT	X244	0.644	0.354	0.375	-0.071	0.147
20	MEAN_RH_10000_15000_FT	X246	0.556	0.617	-0.185	-0.184	-0.299
21	ISOTHERM_HEIGHTS_0c	X331	-0.116	0.173	-0.128	0.471	0.439
22	ISOTHERM_HEIGHTS_minus5c	X332	-0.123	0.274	-0.156	0.454	0.409
23	ISOTHERM_HEIGHTS_minus10c	X333	-0.076	0.201	-0.177	0.416	0.371
24	LIFTED_INDEX_100_MBAR_LAYER_ADIABATIC	X51	-0.820	0.269	-0.412	-0.164	0.081

ตารางที่ 23 (ต่อ)

ชื่อตัวแปร	Code	Factor1	Factor2	Factor3	Factor4	Factor5
25 LIFTED_INDEX_50_MBAR_LAYER_MEAN_VALUES	X52	-0.792	0.224	-0.475	-0.135	0.137
26 SHOWALTER_INDEX	X53	-0.913	0.352	-0.025	0.005	-0.122
27 TOTAL_TOTALS_INDEX	X54	0.832	-0.555	-0.002	0.000	0.000
28 POTENTIAL_BUOYANCY_INDEX	X57	0.579	-0.388	0.123	0.524	-0.213
Eigen values		12.753	3.453	4.411	1.963	1.102
Proportion of total variance		0.455	0.123	0.158	0.070	0.039
Cumulative proportion		0.455	0.579	0.736	0.806	<u>0.846</u>

เมื่อพิจารณาจากกราฟ Scree Plot ดังภาพที่ 20 มีลักษณะเดียวกันกับภาพที่ 21 คือพบว่า ค่าไอเก็นลดลงอย่างรวดเร็ว หรือเกือบที่จะขนานกับแกนนอน เริ่มจากค่าไอเก็นของปัจจัยที่ 4 ดังนั้น จึงพิจารณาใช้เพียง 3 ปัจจัยรวม (factors)



ภาพที่ 21 กราฟ Scree Plot จากการวิเคราะห์ตัวแปรเชิงปริมาณ 28 ตัวแปร ด้วยวิธี MLM

ดังนั้นวิธีการสกัดปัจจัยแบบ MLM ที่ใช้เกณฑ์การเลือกปัจจัยที่ให้ค่าไอเก้นมากกว่า 1 และจากกราฟได้จำนวนปัจจัยที่เหมาะสม 3 ปัจจัยที่ให้ค่าไอเก้นเท่ากับ 12.753, 3.453 และ 4.411 ตามลำดับ ซึ่งสามารถอธิบายความผันแปรของตัวแปรเดิมได้ร้อยละ 73.6

2) ปัจจัยที่สกัดได้จากวิธีการหมุนแกน

ปัจจัยร่วมที่ได้ในตารางที่ 23 ยังไม่สามารถจัดตัวแปรให้อยู่ในปัจจัยใดปัจจัยหนึ่งได้ จึงเลือกการหมุนแกนแบบมุมฉาก (Varimax rotation) เพื่อให้ปัจจัยที่สกัดได้มีความเป็นอิสระต่อกัน แต่จะทำให้ค่าน้ำหนักปัจจัย (Factor loading) เปลี่ยนแปลงไปจากเดิม ก่อนที่มีการหมุนแกน เพื่อความสะดวกและง่ายต่อการจัดตัวแปรเดิมให้อยู่ในปัจจัยใหม่ และตั้งชื่อปัจจัยให้สอดคล้องกับกลุ่มตัวแปรเดิมให้มากที่สุด ดังแสดงไว้ในตารางที่ 24

ตารางที่ 24 ผลการวิเคราะห์น้ำหนักปัจจัยหลังการหมุนแกนแบบ Varimax และร้อยละความผันแปรของแต่ละปัจจัย ที่ได้จากการสกัดปัจจัยแบบ MLM จากตัวแปรเชิงปริมาณ 28 ตัวแปร

ที่	ตัวแปร	code	ค่าน้ำหนักปัจจัย (Loading, l_{im})			ค่าความร่วมกัน h_i^2	ค่าปัจจัยเฉพาะ e_i
			Factor 1	Factor 2	Factor 3		
1	SUB_CLOUD_MIXING_RATIO_CCL	X145	0.937	0.334	-0.069	0.994	0.006
2	MEAN_MIXING_RATIO_LOWEST_50_MB	X222	0.931	0.356	-0.070	0.997	0.003
3	MEAN_DEW_POINT_TEMP_LOWEST_50_MB	X231	0.928	0.350	-0.086	0.991	0.009
4	MEAN_MIXING_RATIO_LOWEST_100_MB	X221	0.886	0.412	-0.056	0.957	0.043
5	CLOUD_BASE_TEMPERATURE_CCL	X144	0.872	0.364	-0.302	0.985	0.015
6	LIFTED_INDEX_50_MBAR_LAYER_MEAN_VALUES	X52	-0.859	-0.144	0.090	0.767	0.233
7	CLOUD_BASE_TEMPERATURE_LCL	X132	0.845	0.233	-0.084	0.776	0.224
8	LIFTED_INDEX_100_MBAR_LAYER_ADIABATIC	X51	-0.839	-0.151	0.036	0.729	0.271
9	PRECIPITABLE_WATER_SFC850	X211	0.823	0.432	-0.070	0.868	0.132
10	CLOUD_BASE_PRESSURE_CCL	X142	0.713	0.434	-0.547	0.996	0.004

ตารางที่ 24 (ต่อ)

ที่	ตัวแปร	code	ค่าน้ำหนักปัจจัย (Loading, l_{im})			ค่าความ ร่วมกัน h_i^2	ค่า ปัจจัย เฉพาะ e_i
			Factor 1	Factor 2	Factor 3		
11	CLOUD_BASE_HEIGHT_CCL	X143	-0.711	-0.453	0.531	0.992	0.008
12	CLOUD_BASE_HEIGHT_LCL	X131	-0.661	-0.293	0.209	0.566	0.434
13	POTENTIAL_BUOYANCY_INDEX	X57	0.636	-0.060	0.414	0.580	0.420
14	MEAN_RH_1000_5000_FT	X244	0.610	0.509	-0.225	0.682	0.318
15	SFC_TEMP_RISE_REQUIRED_FOR_CCL	X141	-0.592	-0.441	0.518	0.813	0.187
16	SHOWALTER_INDEX	X53	-0.547	-0.349	-0.003	0.422	0.578
17	TOTAL_TOTALS_INDEX	X54	0.488	0.177	0.020	0.270	0.730
18	RAOB_SURFACE_TEMPERATURE	X111	0.268	-0.081	0.262	0.147	0.853
19	AVG_RH_10_18000_FT	X243	0.127	0.882	-0.002	0.794	0.206
20	PRECIPITABLE_WATER_TOTAL	X214	0.443	0.879	0.022	0.969	0.031
21	PRECIPITABLE_WATER_SFC500	X213	0.482	0.875	0.041	1.000	0.000
22	MEAN_RH_10000_15000_FT	X246	0.134	0.847	-0.020	0.735	0.265
23	AVG_RH_0_10000_FT	X242	0.423	0.792	-0.237	0.862	0.138
24	PRECIPITABLE_WATER_SFC700	X212	0.640	0.698	0.024	0.897	0.103
25	CONVECTIVE_TEMPERATURE	X121	-0.500	-0.471	0.710	0.976	0.024
26	ISOTHERM_HEIGHTS_minus5c	X332	-0.032	0.053	0.441	0.198	0.802
27	ISOTHERM_HEIGHTS_0c	X331	-0.004	-0.021	0.438	0.192	0.808
28	ISOTHERM_HEIGHTS_minus10c	X333	-0.033	0.053	0.420	0.180	0.820
Eigen values			11.472	6.421	2.443		
Proportion of total variance			0.410	0.229	0.087		
Cumulative proportion			0.410	0.639	0.726		

หมายเหตุ 1. คอลัมน์ที่แรเงาคือตัวแปรอิสระที่เป็นองค์ประกอบของแต่ละปัจจัย (Factor)

2. ค่าน้ำหนักปัจจัยที่เป็นตัวเข้ม คือค่า Loading ที่มีค่ามากกว่าหรือเท่ากับ 0.7

3) ความสามารถของความแปรปรวนที่อธิบายได้ (Total Variance Explained)

ปัจจัยที่สกัดได้ทั้ง 3 มีความสามารถในการอธิบายการผันแปรของตัวแปรเดิม ทั้ง 28 ตัวแปร คิดเป็นร้อยละ 72.6 โดยที่

ปัจจัยที่ 1 อธิบายการผันแปรได้มากที่สุด มีจำนวนตัวแปรเดิม 18 ตัวแปร คิดเป็นร้อยละ 41.0

ปัจจัยที่ 2 อธิบายการผันแปรของตัวแปรเดิม 6 ตัวแปร คิดเป็นร้อยละ 22.9

ปัจจัยที่ 3 อธิบายการผันแปรของตัวแปรเดิม 4 ตัวแปร คิดเป็นร้อยละ 8.7

4) ประมาณค่าคะแนนปัจจัย (Estimated factor score)

ก. กรณีที่ 7 วิเคราะห์ปัจจัยจากตัวแปรปริมาณ 28 ตัวแปร ที่มีปัญหา

Multicollinearity ด้วยวิธี MLM และคำนวณค่าคะแนนปัจจัยด้วยวิธี

Regression (กรณีที่ 7)

เป็นการสร้างคะแนนปัจจัยจากผลรวมของค่าสัมประสิทธิ์คะแนนปัจจัย

(Component score coefficient matrix) ที่ได้จากตารางที่ 25 คูณด้วยค่าของตัวแปรแต่ละตัวแปร

(X_i) ที่อยู่ในรูปคะแนนมาตรฐาน โดยคำนวณค่าคะแนนปัจจัยเป็นรายวัน สามารถคำนวณได้จาก

$$F_i = \beta'(X_i - \bar{X}_i) + \varepsilon_i$$

กำหนดให้ $i = 1, 2, \dots, 28$ $j = 1, 2, 3$ และ $k = 1, 2, 3, \dots, 277$

นำผลของเมทริกซ์สัมประสิทธิ์คะแนนปัจจัย ที่ได้จากการประมวลผลโปรแกรม SPSS ดังที่ปรากฏในตารางที่ 25 แล้วคำนวณค่าคะแนนปัจจัยด้วยวิธี Regression ได้ดังนี้

ตัวอย่าง การคำนวณค่าคะแนนปัจจัยพื้นฐาน ปัจจัยที่ 1 ถึงปัจจัยที่ 3 ของข้อมูลวันที่ 12 มีนาคม 2547 (วันแรก) ได้ดังนี้

$$F_{11} = (0.002)(0.631) + (0.069)(0.919) + (-0.002)(1.146) + \dots + (0.008)(0.291) = 0.081$$

$$F_{21} = (-0.001)(0.631) + (-0.051)(0.919) + (0.001)(1.146) + \dots + (-0.005)(0.291) = -1.373$$

$$F_{31} = (0.004)(0.631) + (0.269)(0.919) + (0.001)(1.146) + \dots + (0.015)(0.291) = 0.491$$

สรุป เมื่อสกัดปัจจัยด้วยวิธี MLM แล้วคำนวณค่าคะแนนปัจจัย ด้วยวิธี Regression

จะต้องใช้ตัวแปรเดิมทั้งหมด 28 ตัวแปรในการสร้างตัวแปรใหม่ (ปัจจัยร่วม) 3 ปัจจัย

ตารางที่ 25 เมตริกซ์สัมประสิทธิ์คะแนนปัจจัย (Component score coefficient matrix)
 ที่ได้จากการสกัดปัจจัยแบบ MLM จากตัวแปรเชิงปริมาณ 28 ตัวแปร

ที่	ชื่อตัวแปร	Code	Factor1	Factor2	Factor3
1	RAOB_SURFACE_TEMPERATURE	X111	0.002	-0.001	0.004
2	CONVECTIVE_TEMPERATURE	X121	0.069	-0.051	0.269
3	CLOUD_BASE_HEIGHT_LCL	X131	-0.002	0.001	0.001
4	CLOUD_BASE_TEMPERATURE_LCL	X132	0.009	-0.005	0.006
5	SFC_TEMP_RISE_REQUIRED_FOR_CCL	X141	0.004	-0.003	0.023
6	CLOUD_BASE_PRESSURE_CCL	X142	-0.093	0.097	-0.965
7	CLOUD_BASE_HEIGHT_CCL	X143	0.050	-0.052	0.517
8	CLOUD_BASE_TEMPERATURE_CCL	X144	0.069	-0.035	-0.063
9	SUB_CLOUD_MIXING_RATIO_CCL	X145	0.336	-0.197	0.266
10	PRECIPITABLE_WATER_SFC850	X211	0.012	-0.007	0.008
11	PRECIPITABLE_WATER_SFC700	X212	0.008	-0.004	0.007
12	PRECIPITABLE_WATER_SFC500	X213	-0.451	1.365	0.546
13	PRECIPITABLE_WATER_TOTAL	X214	-0.006	0.004	-0.011
14	MEAN_MIXING_RATIO_LOWEST_100_MB	X221	0.042	-0.025	0.033
15	MEAN_MIXING_RATIO_LOWEST_50_MB	X222	0.745	-0.438	0.580
16	MEAN_DEW_POINT_TEMP_LOWEST_50_MB	X231	0.225	-0.132	0.161
17	AVG_RH_0_10000_FT	X242	-0.007	0.005	-0.020
18	AVG_RH_10_18000_FT	X243	-0.006	0.004	-0.008
19	MEAN_RH_1000_5000_FT	X244	0.001	0.000	-0.004
20	MEAN_RH_10000_15000_FT	X246	-0.005	0.003	-0.006
21	ISOTHERM_HEIGHTS_0c	X331	0.002	-0.001	0.006
22	ISOTHERM_HEIGHTS_minus5c	X332	0.002	-0.001	0.005
23	ISOTHERM_HEIGHTS_minus10c	X333	0.001	-0.001	0.005
24	LIFTED_INDEX_100_MBAR_LAYER_ADIABATIC	X51	-0.008	0.005	-0.008
25	LIFTED_INDEX_50_MBAR_LAYER_MEAN_VALUES	X52	-0.009	0.005	-0.007
26	SHOWALTER_INDEX	X53	-0.002	0.001	-0.002
27	TOTAL_TOTALS_INDEX	X54	0.002	-0.001	0.002
28	POTENTIAL_BUOYANCY_INDEX	X57	0.008	-0.005	0.015

ข. กรณีที่ 8 วิเคราะห์ปัจจัยจากตัวแปรปริมาณทั้งหมด 28 ตัวแปร ที่มีปัญหา

Multicollinearity ด้วยวิธี MLM และคำนวณค่าคะแนนปัจจัยด้วยวิธี

Factor Based Score

สร้างสมการคำนวณค่าคะแนนพื้นฐานของปัจจัย ด้วยการรวมคะแนนเฉพาะตัวแปรที่อยู่ในปัจจัยเดียวกัน ที่มีค่าน้ำหนักปัจจัย (β_{ij}) ในตารางที่ 24 มากกว่า 0.70 เท่านั้น และตัวแปรเหล่านี้ผ่านการแปลงค่าข้อมูลให้อยู่ในรูปคะแนนมาตรฐานแล้ว

ตัวอย่าง การคำนวณค่าคะแนนปัจจัยพื้นฐาน ปัจจัยที่ 1 ถึงปัจจัยที่ 3 ของข้อมูลวันที่ 12 มีนาคม 2547 (วันแรก) ได้ดังนี้

$$F_{11} = X145 + X222 + X231 + X221 + X144 + X52 + X132 + X51 + X211 \\ + X142 + X143$$

$$= -2.877$$

$$F_{21} = X243 + X214 + X213 + X246 + X242 = -6.787$$

$$F_{31} = X121 = 0.919$$

สรุป เมื่อสกัดปัจจัยด้วยวิธี MLM แล้วคำนวณค่าคะแนนปัจจัย ด้วยวิธีสร้างค่าคะแนนพื้นฐานของปัจจัยจะใช้ตัวแปรเดิม 17 ตัวแปรในการสร้างตัวแปรใหม่ (ปัจจัยรวม) 3 ปัจจัย

ตารางที่ 26 แสดงตัวอย่างการคำนวณค่าคะแนนปัจจัยโดยใช้วิธี Regression (กรณี 7)

และ Factor Based Score (กรณี 8) เมื่อใช้วิธีสกัดปัจจัยด้วยเทคนิค MLM

ที่	วัน	Regression (กรณี 7)			Factor base score (กรณี 8)		
		Fac1	Fac2	Fac3	Fac1	Fac2	Fac3
1	12 มี.ค.47	0.081	-1.372	0.491	-2.877	-6.787	0.919
2	13 มี.ค.47	-0.033	-0.813	0.139	-1.848	-5.332	0.384
3	14 มี.ค.47	0.781	-1.709	0.175	0.225	-6.344	0.562
4	15 มี.ค.47	0.154	-1.535	-0.548	-1.290	-8.607	0.057
5	16 มี.ค.47	1.653	-2.085	-0.956	3.214	-6.381	-0.687

ตารางที่ 27 การตรวจสอบความเหมาะสมด้วยค่าสถิติ ในการวิเคราะห์ปัจจัยทั้ง 8 กรณี

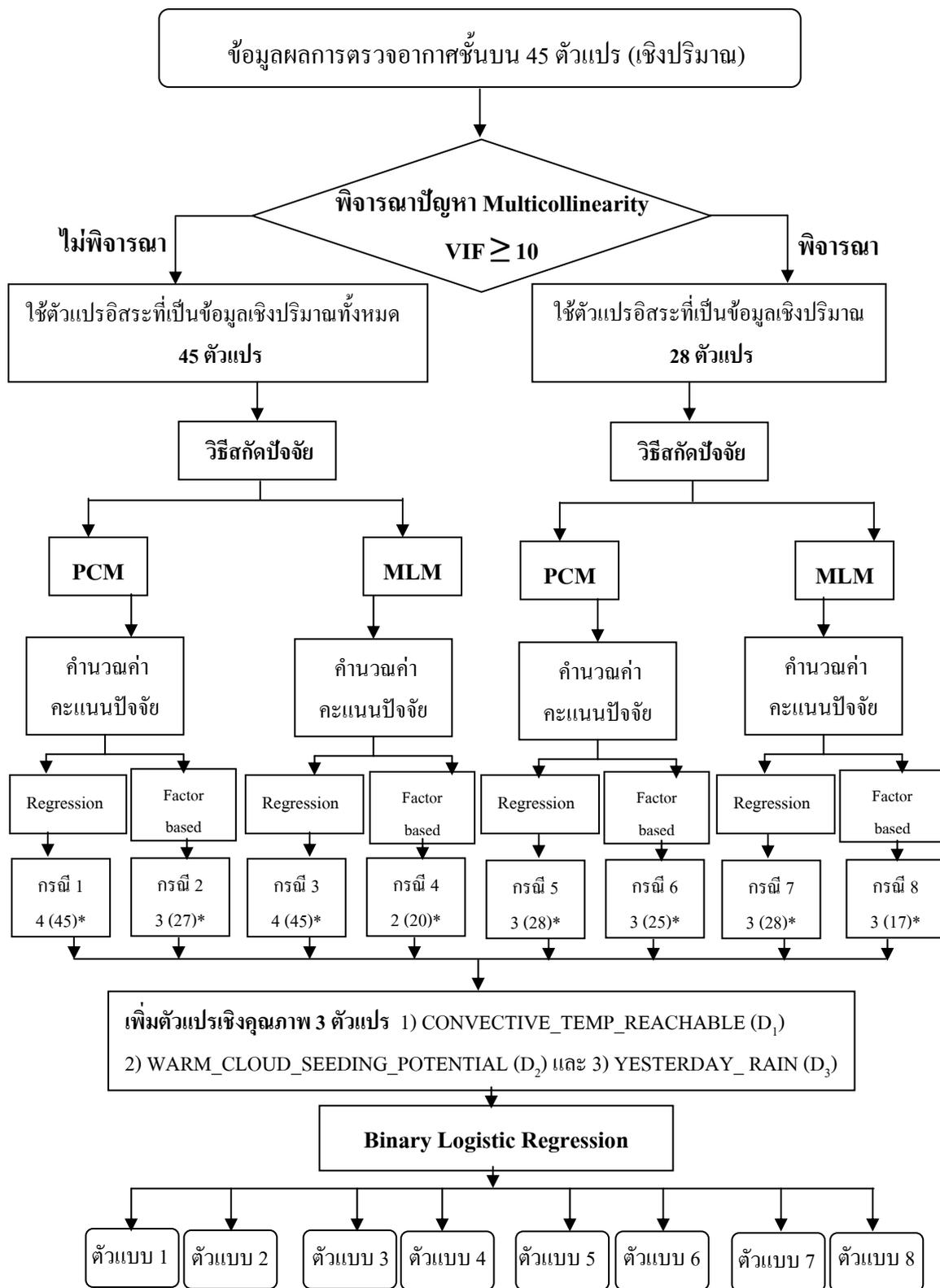
กรณีที่	จำนวน ตัวแปรเริ่มต้น (X's)	วิธีสกัด ปัจจัย	วิธีคำนวณ ค่าน้ำหนักปัจจัย	จำนวน ปัจจัยที่สกัด (F's)	จำนวนตัวแปรเดิม หลังการสกัดปัจจัย (X's)	ค่าสัดส่วน ของความผัน แปรที่อธิบายได้ (%)
1	45	PCM	Regression	4	45	68.9 %
2	45	PCM	Factor Based Score	3	27	68.9 %
3	45	MLM	Regression	4	45	60.0 %
4	45	MLM	Factor Based Score	2	20	60.0 %
5	28	PCM	Regression	3	28	80.7 %
6	28	PCM	Factor Based Score	3	25	80.7 %
7	28	MLM	Regression	3	28	73.6 %
8	28	MLM	Factor Based Score	3	17	73.6 %

ผลสรุปการวิเคราะห์ปัจจัย ด้วยค่าสัดส่วนสะสมของความแปรปรวนที่อธิบายได้ ทั้ง 8 กรณี มีประเด็นที่น่าสนใจ คือ กรณีศึกษาที่ 6 ให้ค่าสัดส่วนของความผันแปรที่อธิบายได้ร้อยละ 80.7 ใช้จำนวนตัวแปรเดิม 25 ตัว ประกอบเป็น 3 ปัจจัย และกรณีศึกษาที่ 8 ให้ค่าสัดส่วนของความผันแปรที่อธิบายได้ร้อยละ 73.6 ใช้จำนวนตัวแปรเดิม 17 ตัว ประกอบเป็น 3 ปัจจัย

ส่วนที่ 3 ผลการสร้างตัวแบบการคาดการณ์โอกาสการเกิดฝนในภาคตะวันออกเฉียงเหนือ ด้วยวิธีวิเคราะห์ถดถอยโลจิสติก (Logistic Regression Analysis)

3.1 ผลการสร้างตัวแบบการคาดการณ์การเกิดฝน ด้วยการวิเคราะห์การถดถอยโลจิสติกแบบทวิ

จากการวิเคราะห์ปัจจัยใน 8 กรณี สรุปในตารางที่ 27 นำปัจจัยรวมที่ได้จากแต่ละกรณี และตัวแปรเชิงคุณภาพอีก 3 ตัว มาเป็นตัวแปรอิสระในการสร้างตัวแบบการคาดการณ์โอกาสการเกิดฝน ภาคตะวันออกเฉียงเหนือ ใช้เทคนิคการคัดเลือกตัวแปรจากโปรแกรมสำเร็จรูป SPSS ด้วยวิธี Forward stepwise likelihood ratio กำหนดค่า $P_E = 0.15$ และ $P_R = 0.20$ ทำให้ได้ผลการประมาณค่าพารามิเตอร์ (β') และค่าสถิติทดสอบต่าง ๆ ที่มีผลต่อการเลือกตัวแปรอิสระเข้าในตัวแบบการพยากรณ์โอกาสการเกิดฝนของแต่ละกรณีได้ 8 ตัวแบบ แสดงรายละเอียดดังภาพที่ 22 และตารางที่ 28 - ตารางที่ 35 ดังต่อไปนี้



หมายเหตุ (*) คือ จำนวนปัจจัย (จำนวนตัวแปรเดิม)

ภาพที่ 22 แผนผังแสดงการลำดับขั้นตอนของการวิเคราะห์เหตุการณ์ทั้ง 8 ตัวแบบ

ตารางที่ 28 ผลการวิเคราะห์ถดถอยโลจิสติก กรณีที่ไม่พิจารณาค่า VIF จึงทำให้ใช้ตัวแปรที่เป็นข้อมูลเชิงปริมาณทั้ง 45 ตัวแปร โดยสกัดปัจจัยด้วย PCM แล้วคำนวณค่าคะแนนปัจจัยด้วยวิธี Regression (กรณีที่ 1)

ตัวแปรอิสระ	สัมประสิทธิ์					95% CI for Exp (β)		
	การถดถอย (β)	S.E.	Wald	df	Sig.	Exp (β)	Lower	Upper
ค่าคงที่	2.282	0.577	15.631	1	<0.01**	9.795		
F ₁	1.288	0.270	22.755	1	<0.01**	3.624	2.135	6.151
F ₂	1.502	0.342	19.241	1	<0.01**	4.490	2.295	8.785
F ₃	0.530	0.240	4.870	1	0.027*	1.698	1.061	2.718
การเกิดฝนของวันก่อนหน้า	1.144	0.488	5.484	1	0.019*	3.138	1.205	8.175

(**) มีนัยสำคัญทางสถิติที่ระดับ 0.01 และ (*) มีนัยสำคัญทางสถิติที่ระดับ 0.05

ผลการสร้างตัวแบบการถดถอยโลจิสติก

$$\text{ตัวแบบที่ 1} \quad \log_e \left(\frac{P}{1-P} \right) = 2.282 + 1.288(F_1) + 1.502(F_2) + 0.53(F_3) + 1.144(D_3)$$

สรุปตัวแบบที่ 1 ปัจจัยที่มีผลต่อโอกาสการเกิดฝน ของภาคตะวันออกเฉียงเหนือ อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ประกอบด้วย ปัจจัยร่วม F₁, F₂, F₃ และตัวแปรการเกิดฝนของวันก่อนหน้า (D₃) ได้ค่า R²_N = 0.529 และค่าความถูกต้องของการพยากรณ์ 91 เปอร์เซ็นต์

การอธิบายผล

ทุก ๆ 1 คะแนนมาตรฐานของ F₁ ที่เพิ่มขึ้น จะมีโอกาสที่ฝนจะตกภายในวันนี้เพิ่มขึ้นเป็น 3.62 เท่า โดยที่ตัวแปรอื่น ๆ ในตัวแบบมีค่าคงที่ [Exp (1.288) = 3.62]

ทุก ๆ 1 คะแนนมาตรฐานของ F₂ ที่เพิ่มขึ้น จะมีโอกาสที่ฝนจะตกภายในวันนี้เพิ่มขึ้นเป็น 4.49 เท่า โดยที่ตัวแปรอื่น ๆ ในตัวแบบมีค่าคงที่ [Exp (1.502) = 4.49]

ทุก ๆ 1 คะแนนมาตรฐานของ F₃ ที่เพิ่มขึ้น จะมีโอกาสที่ฝนจะตกภายในวันนี้เพิ่มขึ้นเป็น 1.70 เท่า โดยที่ตัวแปรอื่น ๆ ในตัวแบบมีค่าคงที่ [Exp (0.530) = 1.70]

ถ้าวันก่อนหน้า (D₃) เป็นวันที่มีฝนตก แล้วโอกาสที่วันนี้ฝนจะตกสูงกว่าเป็น 3.14 เท่าของวันก่อนหน้าที่ฝนไม่ตก โดยที่ตัวแปรอื่น ๆ ในตัวแบบมีค่าคงที่ [Exp (1.144) = 3.14]

ตารางที่ 29 ผลการวิเคราะห์ถดถอยโลจิสติก กรณีที่ไม่พิจารณาค่า VIF จึงทำให้ใช้ตัวแปรที่เป็นข้อมูลเชิงปริมาณทั้ง 45 ตัวแปร โดยสกัดปัจจัยด้วย PCM แล้วคำนวณค่าคะแนนปัจจัยด้วยวิธี Factor Based Score (กรณีที่ 2)

ตัวแปรอิสระ	สัมประสิทธิ์					95% CI for Exp (β)		
	การถดถอย (β)	S.E.	Wald	df	Sig.	Exp (β)	Lower	Upper
ค่าคงที่	2.216	0.566	15.317	1	<0.01**	9.174		
F ₁	0.163	0.046	12.542	1	<0.01**	1.177	1.075	1.287
F ₂	0.287	0.095	9.066	1	<0.01**	1.333	1.105	1.606
F ₄	-0.679	0.236	8.317	1	<0.01**	0.507	0.320	0.804
การเกิดฝนของวันก่อนหน้า	1.116	0.497	5.047	1	0.025*	3.051	1.153	8.075

(**) มีนัยสำคัญทางสถิติที่ระดับ 0.01 และ (*) มีนัยสำคัญทางสถิติที่ระดับ 0.05

ผลการสร้างตัวแบบการถดถอยโลจิสติก

$$\text{ตัวแบบที่ 2} \quad \log_e \left(\frac{P}{1-P} \right) = 2.216 + 0.163(F_1) + 0.287(F_2) - 0.679(F_4) + 1.116(D_3)$$

สรุปตัวแบบที่ 2 ปัจจัยที่มีผลต่อโอกาสการเกิดฝน ของภาคตะวันออกเฉียงเหนือ อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ประกอบด้วย ปัจจัยร่วม F₁, F₂, F₄ และตัวแปรการเกิดฝนของวันก่อนหน้า (D₃) ได้ค่า R²_N = 0.507 และค่าความถูกต้องของการพยากรณ์ 89.89 เปอร์เซ็นต์ การอธิบายผล

ทุก ๆ 1 คะแนนมาตรฐานของ F₁ ที่เพิ่มขึ้น จะมีโอกาสที่ฝนจะตกภายในวันนี้เพิ่มขึ้นเป็น 1.18 เท่า โดยที่ตัวแปรอื่น ๆ ในตัวแบบมีค่าคงที่ [Exp (0.163) = 1.18]

ทุก ๆ 1 คะแนนมาตรฐานของ F₂ ที่เพิ่มขึ้น จะมีโอกาสที่ฝนจะตกภายในวันนี้เพิ่มขึ้นเป็น 1.33 เท่า โดยที่ตัวแปรอื่น ๆ ในตัวแบบมีค่าคงที่ [Exp (0.287) = 1.33]

ทุก ๆ 1 คะแนนมาตรฐานของ F₄ ที่เพิ่มขึ้น จะมีโอกาสที่ฝนจะตกภายในวันนี้ลดลงเป็น 0.51 เท่า โดยที่ตัวแปรอื่น ๆ ในตัวแบบมีค่าคงที่ [Exp (-0.679) = 0.51]

ถ้าวันก่อนหน้า (D₃) เป็นวันที่มีฝนตก แล้วโอกาสที่วันนี้ฝนจะตกสูงกว่าเป็น 3.05 เท่าของวันก่อนหน้าที่ฝนไม่ตก โดยที่ตัวแปรอื่น ๆ ในตัวแบบมีค่าคงที่ [Exp (1.116) = 3.05]

ตารางที่ 30 ผลการวิเคราะห์ถดถอยโลจิสติก กรณีที่ไม่พิจารณาค่า VIF จึงทำให้ใช้ตัวแปรที่เป็นข้อมูลเชิงปริมาณทั้ง 45 ตัวแปร โดยสกัดปัจจัยด้วย MLM แล้วคำนวณค่าคะแนนปัจจัยด้วยวิธี Regression (กรณีที่ 3)

ตัวแปรอิสระ	สัมประสิทธิ์					95% CI for Exp (β)		
	การถดถอย (β)	S.E.	Wald	df	Sig.	Exp (β)	Lower	Upper
ค่าคงที่	2.241	0.557	16.184	1	<0.01**	9.407		
F ₁	1.231	0.262	22.065	1	<0.01**	3.426	2.050	5.727
F ₂	1.383	0.277	24.856	1	<0.01**	3.988	2.315	6.870
การเกิดฝนของวันก่อนหน้า	0.986	0.497	3.938	1	0.047*	2.680	1.012	7.093

(**) มีนัยสำคัญทางสถิติที่ระดับ 0.01 และ (*) มีนัยสำคัญทางสถิติที่ระดับ 0.05

ผลการสร้างตัวแบบการถดถอยโลจิสติก

$$\text{ตัวแบบที่ 3} \quad \log_e \left(\frac{P}{1-P} \right) = 2.241 + 1.231(F_1) + 1.383(F_2) + 0.986(D_3)$$

สรุปตัวแบบที่ 3 ปัจจัยที่มีผลต่อโอกาสการเกิดฝนของภาคตะวันออกเฉียงเหนือ อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ประกอบด้วย ปัจจัยร่วม F₁, F₂ และตัวแปรการเกิดฝนของวันก่อนหน้า (D₃) ได้ค่า R²_N = 0.513 และค่าความถูกต้องของการพยากรณ์ 89.89 เปอร์เซ็นต์การอธิบายผล

ทุก ๆ 1 คะแนนมาตรฐานของ F₁ ที่เพิ่มขึ้น จะมีโอกาสที่ฝนจะตกภายในวันนี้เพิ่มขึ้นเป็น 3.43 เท่า โดยที่ตัวแปรอื่น ๆ ในตัวแบบมีค่าคงที่ [Exp (1.231) = 3.43]

ทุก ๆ 1 คะแนนมาตรฐานของ F₂ ที่เพิ่มขึ้น จะมีโอกาสที่ฝนจะตกภายในวันนี้เพิ่มขึ้นเป็น 3.99 เท่า โดยที่ตัวแปรอื่น ๆ ในตัวแบบมีค่าคงที่ [Exp (1.383) = 3.99]

ถ้าวันก่อนหน้า (D₃) เป็นวันที่มีฝนตก แล้วโอกาสที่วันนี้ฝนจะตกสูงกว่าเป็น 2.68 เท่าของวันก่อนหน้าที่ฝนไม่ตก โดยที่ตัวแปรอื่น ๆ ในตัวแบบมีค่าคงที่ [Exp (0.986) = 2.68]

ตารางที่ 31 ผลการวิเคราะห์ถดถอยโลจิสติก กรณีที่ไม่พิจารณาค่า VIF จึงทำให้ใช้ตัวแปรที่เป็นข้อมูลเชิงปริมาณทั้ง 45 ตัวแปร โดยสกัดปัจจัยด้วย MLM แล้วคำนวณค่าคะแนนปัจจัยด้วยวิธี Factor Based Score (กรณีที่ 4)

ตัวแปรอิสระ	สัมประสิทธิ์					95% CI for Exp (β)		
	การถดถอย (β)	S.E.	Wald	df	Sig.	Exp (β)	Lower	Upper
ค่าคงที่	1.827	0.517	12.464	1	<0.01**	6.215		
F ₁	0.099	0.042	5.393	1	0.020*	1.104	1.015	1.199
F ₂	0.561	0.133	17.832	1	<0.01**	1.753	1.351	2.275
การเกิดฝนของวันก่อนหน้า	1.218	0.487	6.254	1	0.012*	3.381	1.301	8.785

(**) มีนัยสำคัญทางสถิติที่ระดับ 0.01 และ (*) มีนัยสำคัญทางสถิติที่ระดับ 0.05

ผลการสร้างตัวแบบการถดถอยโลจิสติก

$$\text{ตัวแบบที่ 4} \quad \log_e \left(\frac{P}{1-P} \right) = 1.827 + 0.099(F_1) + 0.561(F_2) + 1.218(D_3)$$

สรุปตัวแบบที่ 4 ปัจจัยที่มีผลต่อโอกาสการเกิดฝนของภาคตะวันออกเฉียงเหนือ อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ประกอบด้วย ปัจจัยร่วม F₁, F₂ และตัวแปรการเกิดฝนของวันก่อนหน้า (D₃) ได้ค่า R²_N = 0.478 และค่าความถูกต้องของการพยากรณ์ 88.81 เปอร์เซ็นต์การอธิบายผล

ทุก ๆ 1 คะแนนมาตรฐานของ F₁ ที่เพิ่มขึ้น จะมีโอกาสที่ฝนจะตกภายในวันนี้เพิ่มขึ้นเป็น 1.10 เท่า โดยที่ตัวแปรอื่น ๆ ในตัวแบบมีค่าคงที่ [Exp (0.099) = 1.10]

ทุก ๆ 1 คะแนนมาตรฐานของ F₂ ที่เพิ่มขึ้น จะมีโอกาสที่ฝนจะตกภายในวันนี้เพิ่มขึ้นเป็น 1.75 เท่า โดยที่ตัวแปรอื่น ๆ ในตัวแบบมีค่าคงที่ [Exp (0.561) = 1.75]

ถ้าวันก่อนหน้า (D₃) เป็นวันที่มีฝนตก แล้วโอกาสที่วันนี้ฝนจะตกสูงกว่าเป็น 3.38 เท่าของวันก่อนหน้าที่ฝนไม่ตก โดยที่ตัวแปรอื่น ๆ ในตัวแบบมีค่าคงที่ [Exp (1.218) = 3.38]

ตารางที่ 32 ผลการวิเคราะห์ถดถอยโลจิสติก กรณีที่พิจารณาตัวแปรที่มีค่า VIF มากกว่าหรือเท่ากับ 10 จึงทำให้ใช้ตัวแปรที่เป็นข้อมูลเชิงปริมาณ 28 ตัวแปร โดยสกัดปัจจัยด้วย PCM แล้วคำนวณค่าคะแนนปัจจัยด้วยวิธี Regression (กรณีที่ 5)

ตัวแปรอิสระ	สัมประสิทธิ์					95% CI for Exp (β)		
	การถดถอย (β)	S.E.	Wald	df	Sig.	Exp (β)	Lower	Upper
ค่าคงที่	1.419	0.482	8.653	1	<0.01**	4.134		
F ₁	1.414	0.287	24.221	1	<0.01**	4.111	2.341	7.218
F ₃	-0.929	0.466	3.976	1	0.046*	0.395	0.158	0.984
การเกิดฝนของวันก่อนหน้า	1.641	0.475	11.918	1	<0.01**	5.159	2.032	13.095

(**) มีนัยสำคัญทางสถิติที่ระดับ 0.01 และ (*) มีนัยสำคัญทางสถิติที่ระดับ 0.05

ผลการสร้างตัวแบบการถดถอยโลจิสติก

$$\text{ตัวแบบที่ 5} \quad \log_e \left(\frac{P}{1-P} \right) = 1.419 + 1.414(F_1) - 0.929(F_3) + 1.641(D_3)$$

สรุปตัวแบบที่ 5 ปัจจัยที่มีผลต่อโอกาสการเกิดฝนของภาคตะวันออกเฉียงเหนือ อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ประกอบด้วย ปัจจัยร่วม F₁, F₃ และตัวแปรการเกิดฝนของวันก่อนหน้า (D₃) ได้ค่า R²_N = 0.459 และค่าความถูกต้องของการพยากรณ์ 89.89 เปอร์เซ็นต์การอธิบายผล

ทุก ๆ 1 คะแนนมาตรฐานของ F₁ ที่เพิ่มขึ้น จะมีโอกาสที่ฝนจะตกภายในวันนี้เพิ่มขึ้นเป็น 4.11 เท่า โดยที่ตัวแปรอื่น ๆ ในตัวแบบมีค่าคงที่ [Exp (1.414) = 4.11]

ทุก ๆ 1 คะแนนมาตรฐานของ F₃ ที่เพิ่มขึ้น จะมีโอกาสที่ฝนจะตกภายในวันนี้ลดลงเป็น 0.40 เท่า โดยที่ตัวแปรอื่น ๆ ในตัวแบบมีค่าคงที่ [Exp (-0.929) = 0.40]

ถ้าวันก่อนหน้า (D₃) เป็นวันที่มีฝนตก แล้วโอกาสที่วันนี้ฝนจะตกสูงกว่าเป็น 4.13 เท่าของวันก่อนหน้าที่ฝนไม่ตก โดยที่ตัวแปรอื่น ๆ ในตัวแบบมีค่าคงที่ [Exp (1.641) = 5.16]

ตารางที่ 33 ผลการวิเคราะห์ถดถอยโลจิสติก กรณีที่พิจารณาตัวแปรที่มีค่า VIF มากกว่า หรือเท่ากับ 10 จึงทำให้ใช้ตัวแปรที่เป็นข้อมูลเชิงปริมาณ 28 ตัวแปร โดยสกัด ปัจจัยด้วย PCM แล้วคำนวณค่าคะแนนปัจจัยด้วยวิธี Factor Based Score (กรณีที่ 6)

ตัวแปรอิสระ	สัมประสิทธิ์			Sig.	Exp (β)	95% CI for Exp (β)	
	การถดถอย (β)	S.E.	Wald df			Lower	Upper
ค่าคงที่	1.751	0.515	11.564 1	<0.01**	5.758		
F ₁	0.150	0.030	25.314 1	<0.01**	1.161	1.096	1.231
F ₃	-0.551	0.226	5.928 1	0.015*	0.577	0.370	0.898
การเกิดฝนของวันก่อนหน้า	1.425	0.484	8.670 1	<0.01**	4.157	1.610	10.732

(**) มีนัยสำคัญทางสถิติที่ระดับ 0.01 และ (*) มีนัยสำคัญทางสถิติที่ระดับ 0.05

ผลการสร้างตัวแบบการถดถอยโลจิสติก

$$\text{ตัวแบบที่ 6} \quad \log_e \left(\frac{P}{1-P} \right) = 1.751 + 0.150(F_1) - 0.551(F_3) + 1.425(D_3)$$

สรุปตัวแบบที่ 6 ปัจจัยที่มีผลต่อโอกาสการเกิดฝน ของภาคตะวันออกเฉียงเหนือ อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ประกอบด้วย ปัจจัยร่วม F₁, F₃ และตัวแปรการเกิดฝนของวันก่อนหน้า (D₃) ได้ค่า R²_N = 0.477 และค่าความถูกต้องของการพยากรณ์ 89.17 เปอร์เซ็นต์การอธิบายผล

ทุก ๆ 1 คะแนนมาตรฐานของ F₁ ที่เพิ่มขึ้น จะมีโอกาสที่ฝนจะตกภายในวันนี้เพิ่มขึ้นเป็น 1.16 เท่า โดยที่ตัวแปรอื่น ๆ ในตัวแบบมีค่าคงที่ [Exp (0.150) = 1.16]

ทุก ๆ 1 คะแนนมาตรฐานของ F₃ ที่เพิ่มขึ้น จะมีโอกาสที่ฝนจะตกภายในวันนี้ลดลงเป็น 0.58 เท่า โดยที่ตัวแปรอื่น ๆ ในตัวแบบมีค่าคงที่ [Exp (-0.551) = 0.58]

ถ้าวันก่อนหน้า (D₃) เป็นวันที่มีฝนตก แล้วโอกาสที่วันนี้ฝนจะตกสูงกว่าเป็น 5.16 เท่าของวันก่อนหน้าที่ฝนไม่ตก โดยที่ตัวแปรอื่น ๆ ในตัวแบบมีค่าคงที่ [Exp (1.43) = 4.16]

ตารางที่ 34 ผลการวิเคราะห์ถดถอยโลจิสติก กรณีที่พิจารณาตัวแปรที่มีค่า VIF มากกว่า หรือเท่ากับ 10 จึงทำให้ใช้ตัวแปรที่เป็นข้อมูลเชิงปริมาณ 28 ตัวแปร โดยสกัดปัจจัย ด้วย MLM แล้วคำนวณค่าคะแนนปัจจัยด้วยวิธี Regression (กรณีที่ 7)

ตัวแปรอิสระ	สัมประสิทธิ์					95% CI for Exp (β)		
	การถดถอย (β)	S.E.	Wald	df	Sig.	Exp (β)	Lower	Upper
ค่าคงที่	1.601	0.483	10.969	1	<0.01**	4.957		
F ₁	0.844	0.224	14.219	1	<0.01**	2.326	1.500	3.607
F ₂	1.255	0.258	23.660	1	<0.01**	3.506	2.115	5.813
การเกิดฝนของวันก่อนหน้า	1.420	0.476	8.921	1	<0.01**	4.138	1.629	10.510

(**) มีนัยสำคัญทางสถิติที่ระดับ 0.01

ผลการสร้างตัวแบบการถดถอยโลจิสติก

$$\text{ตัวแบบที่ 7} \quad \log_e \left(\frac{P}{1-P} \right) = 1.601 + 0.844(F_1) + 1.255(F_2) + 1.420(D_3)$$

สรุปตัวแบบที่ 7 ปัจจัยที่มีผลต่อโอกาสการเกิดฝน ของภาคตะวันออกเฉียงเหนือ อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ประกอบด้วย ปัจจัยร่วม F₁, F₂ และตัวแปรการเกิดฝนของวันก่อนหน้า (D₃) ได้ค่า R²_N = 0.469 และค่าความถูกต้องของการพยากรณ์ 90.25 เปอร์เซ็นต์การอธิบายผล

ทุก ๆ 1 คะแนนมาตรฐานของ F₁ ที่เพิ่มขึ้น จะมีโอกาสที่ฝนจะตกภายในวันนี้เพิ่มขึ้นเป็น 2.33 เท่า โดยที่ตัวแปรอื่น ๆ ในตัวแบบมีค่าคงที่ [Exp (0.844) = 2.33]

ทุก ๆ 1 คะแนนมาตรฐานของ F₂ ที่เพิ่มขึ้น จะมีโอกาสที่ฝนจะตกภายในวันนี้เพิ่มขึ้นเป็น 3.51 เท่า โดยที่ตัวแปรอื่น ๆ ในตัวแบบมีค่าคงที่ [Exp (1.255) = 3.51]

ถ้าวันก่อนหน้า (D₃) เป็นวันที่มีฝนตก แล้วโอกาสที่วันนี้ฝนจะตกสูงกว่าเป็น 4.14 เท่าของวันก่อนหน้าที่ฝนไม่ตก โดยที่ตัวแปรอื่น ๆ ในตัวแบบมีค่าคงที่ [Exp (1.420) = 4.14]

ตารางที่ 35 ผลการวิเคราะห์ถดถอยโลจิสติก กรณีที่พิจารณาตัวแปรที่มีค่า VIF มากกว่า หรือเท่ากับ 10 จึงทำให้ใช้ตัวแปรที่เป็นข้อมูลเชิงปริมาณ 28 ตัวแปร โดยสกัดปัจจัย ด้วย MLM แล้วคำนวณค่าคะแนนปัจจัยด้วยวิธี Factor base (กรณีที่ 8)

ตัวแปรอิสระ	สัมประสิทธิ์					95% CI for Exp (β)		
	การถดถอย (β)	S.E.	Wald	df	Sig.	Exp (β)	Lower	Upper
Constant	1.675	0.498	11.296	1	<0.01**	5.339		
F ₂	0.322	0.061	27.737	1	<0.01**	1.379	1.224	1.555
การเกิดฝนของวันก่อนหน้า	1.335	0.482	7.668	1	<0.01**	3.799	1.477	9.773

(**) มีนัยสำคัญทางสถิติที่ระดับ 0.01

ผลการสร้างตัวแบบการถดถอยโลจิสติก

$$\text{ตัวแบบที่ 8} \quad \log_e \left(\frac{P}{1-P} \right) = 1.675 + 0.322(F_2) + 1.335(D_3)$$

สรุปตัวแบบที่ 8 ปัจจัยที่มีผลต่อโอกาสการเกิดฝน ของภาคตะวันออกเฉียงเหนือ อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ประกอบด้วย ปัจจัยร่วม F₂ และตัวแปรการเกิดฝนของวันก่อนหน้า (D₃) ได้ค่า R²_N = 0.472 และค่าความถูกต้องของการพยากรณ์ 90.25 เปอร์เซ็นต์ การอธิบายผล

ทุก ๆ 1 คะแนนมาตรฐานของ F₂ ที่เพิ่มขึ้น จะมีโอกาสที่ฝนจะตกภายในวันนี้เพิ่มขึ้นเป็น 1.38 เท่า โดยที่ตัวแปรอื่น ๆ ในตัวแบบมีค่าคงที่ [Exp (0.322) = 1.38]

ถ้าวันก่อนหน้า (D₃) เป็นวันที่มีฝนตก แล้วโอกาสที่วันนี้ฝนจะตกสูงกว่าเป็น 3.80 เท่าของวันก่อนหน้าที่ฝนไม่ตก โดยที่ตัวแปรอื่น ๆ ในตัวแบบมีค่าคงที่ [Exp (1.34) = 3.80]

ตารางที่ 36 ผลการวิเคราะห์ถดถอยโลจิสติกแบบทวิ 8 ตัวแบบ

ตัวแบบ	ค่าสถิติ G^2 [Sig.]	R^2_N	ความถูกต้องของการพยากรณ์ (%)	ค่าสถิติ Chi-Square [Sig.]	ปัจจัยที่ผ่านการคัดเลือก (F's)	ตัวแปรเดิม (X's) ที่ประกอบเป็นปัจจัยที่ผ่านการคัดเลือก	ตัวแปรเชิงกลุ่มที่ผ่านการคัดเลือก	จำนวนตัวแปรเดิมและตัวแปรเชิงกลุ่มที่ผ่านการคัดเลือก
1	95.248 [<0.01]**	0.529	91.00	3.318 [0.913]	3	41	1	42
2	90.630 [<0.01]**	0.507	89.89	3.913 [0.865]	3	27	1	28
3	91.913 [<0.01]**	0.513	89.89	2.226 [0.973]	2	34	1	35
4	84.638 [<0.01]**	0.478	88.81	2.793 [0.947]	2	20	1	21
5	80.770 [<0.01]**	0.459	89.89	5.730 [0.677]	2	22	1	23
6	84.450 [<0.01]**	0.477	89.17	2.590 [0.957]	2	20	1	21
7	82.677 [<0.01]**	0.469	90.25	2.481 [0.963]	2	24	1	25
8	82.264 [<0.01]**	0.472	90.25	3.834 [0.872]	1	5	1	6

(**) มีนัยสำคัญทางสถิติที่ระดับ 0.01

ผลสรุปการตรวจสอบความเหมาะสมด้วยค่าสถิติ ในการวิเคราะห์ถดถอยโลจิสติกทั้ง 8 ตัวแบบ ในตารางที่ 36 ผลจากการสร้างตัวแบบทั้ง 8 ตัวแบบ พบว่าค่าสถิติ G^2 มีค่า P-value < 0.01 ทุกตัวแบบ ดังนั้นจึงสรุปได้ว่าโอกาสที่จะเกิดเหตุการณ์การเกิดฝนตก ขึ้นอยู่กับตัวแปรอิสระอย่างน้อย 1 ตัวในทุกตัวแบบ โดยค่า R^2_N บอกอำนาจของการพยากรณ์ตัวแบบที่สร้างขึ้นอยู่ในช่วง 0.478 ถึง 0.529 มีค่าร้อยละความถูกต้องของการพยากรณ์อยู่ในช่วงร้อยละ 88.8 ถึง 91.0 และทดสอบความเหมาะสมของตัวแบบจากค่าสถิติ Chi-Square มีค่า P-value > 0.05 ทุกตัวแบบ จึงมีความเหมาะสมในการพยากรณ์ค่าโอกาสการเกิดฝน ของภาคตะวันออกเฉียงเหนือ โดยมีการคัดเลือกตัวแปรอิสระที่แตกต่างกันในแต่ละตัวแบบ

ตารางที่ 37 สรุปผลการวิเคราะห์ปัจจัย 8 กรณี และการวิเคราะห์ถดถอยโลจิสติกแบบทวิ 8 ตัวแบบ

รายชื่อตัวแปร	ไม่พิจารณา VIF				พิจารณา VIF ≥ 10			
	PCM		MLM		PCM		MLM	
	Factor Regression Based Score		Factor Regression Based Score		Factor Regression Based Score		Factor Regression Based Score	
	กรณี 1	กรณี 2	กรณี 3	กรณี 4	กรณี 5	กรณี 6	กรณี 7	กรณี 8
● การวิเคราะห์ปัจจัย								
- จำนวนตัวแปรเริ่มต้น (X's)	45	45	45	45	28	28	28	28
- จำนวนปัจจัยที่สกัดได้ (F's)	4	3	4	2	3	3	3	3
[จำนวนตัวแปรที่ใช้หาค่า คะแนนปัจจัย]	[45]	[27]	[45]	[20]	[28]	[25]	[28]	[17]
- ค่าสัดส่วนของความผันแปร ที่อธิบายได้ (Cumulative Proportion)	68.9 %	68.9 %	60.0 %	60.0 %	80.7 %	80.7 %	73.6 %	73.6 %
● การวิเคราะห์ถดถอย โลจิสติกแบบทวิ	ตัวแบบ 1	ตัวแบบ 2	ตัวแบบ 3	ตัวแบบ 4	ตัวแบบ 5	ตัวแบบ 6	ตัวแบบ 7	ตัวแบบ 8
- ค่าสถิติ G^2 [Sig.]	95.248 [<0.01]**	90.630 [<0.01]**	91.913 [<0.01]**	84.638 [<0.01]**	80.770 [<0.01]**	84.450 [<0.01]**	82.677 [<0.01]**	82.264 [<0.01]**
- R^2_N	0.529	0.507	0.513	0.478	0.459	0.477	0.469	0.472
- Hit rate (%)	91.00 %	89.89 %	89.89 %	88.81 %	89.89 %	89.17 %	90.25 %	90.25 %
- Goodness of fit test ค่าสถิติ Chi-Square [Sig.]	3.318 [0.913]	3.913 [0.865]	2.226 [0.973]	2.793 [0.947]	5.730 [0.677]	2.590 [0.957]	2.481 [0.963]	3.834 [0.872]
- จำนวนปัจจัยที่ผ่าน การคัดเลือก (F's)	3	3	2	2	2	2	2	1
- ตัวแปรเดิม (X's) ที่ประกอบเป็นปัจจัย ที่ผ่านการคัดเลือก	41	27	34	20	22	20	24	5
- ตัวแปรเชิงกลุ่มที่ผ่านการ คัดเลือก	1	1	1	1	1	1	1	1
- จำนวนตัวแปรเดิม และตัวแปร เชิงกลุ่มที่ผ่านการคัดเลือก	42	28	35	21	23	21	25	6

(**) มีนัยสำคัญทางสถิติที่ระดับ 0.01

จากตารางที่ 37 เป็นการสรุปผลการศึกษาการวิเคราะห์ปัจจัย 8 กรณี และการวิเคราะห์ถดถอยโลจิสติกแบบทวิ 8 ตัวแบบ ตัวแบบที่มีความเหมาะสมที่สุด คือ ตัวแบบที่ 8 ที่วิเคราะห์ปัจจัยจากตัวแปรปริมาณ 28 ตัวแปร ที่มีปัญหา Multicollinearity ($VIF \geq 10$) โดยสกัดปัจจัยด้วย MLM และคำนวณค่าคะแนนปัจจัยด้วยวิธี Factor Based Score ใช้ตัวแปรอิสระในการสร้างตัวแบบน้อยสุด 6 ตัวแปร คือ 1) AVG_RH_10_18000_FT (X243) 2) PRECIPITABLE_WATER_TOTAL (X214) 3) PRECIPITABLE_WATER_SFC500 (X213) 4) MEAN_RH_10000_15000_FT (X246) 5) AVG_RH_0_10000_FT (X242) และ 6) การเกิดฝนของวันก่อนหน้า โดยที่ 5 ตัวแปรแรก เป็นข้อมูล GPCM ในกลุ่มความชื้น สร้างเป็นตัวแปรปัจจัยร่วมที่ 2 (F_2) และตัวแปรการเกิดฝนของวันก่อนหน้า (D_3) ที่เป็นตัวแปรเชิงคุณภาพอีก 1 ตัวแปร ที่ผ่านการคัดเลือกจากตัวแบบคาดการณ์โอกาสการเกิดฝน ของภาคตะวันออกเฉียงเหนือ โดยมีค่า $R^2_N = 0.472$ และค่าความถูกต้องของการพยากรณ์สูงถึงร้อยละ 90.25

วิจารณ์

1. ผลที่ได้จากการวิเคราะห์ปัจจัยทั้ง 8 กรณี ทำให้สร้างตัวแบบคาดการณ์โอกาสการเกิดฝน ของภาคตะวันออกเฉียงเหนือ ได้ 8 ตัวแบบ โดยที่ตัวแบบทั้ง 8 มีค่า R^2_N ในช่วง 0.478 ถึง 0.529 และมีค่าร้อยละความถูกต้องของการพยากรณ์ในช่วงร้อยละ 88.8 ถึง ร้อยละ 91 จะเห็นได้ว่าการวิเคราะห์ปัจจัยก่อนการสร้างตัวแบบทำให้ใช้จำนวนตัวแปรลดน้อยลง แต่ยังคงให้ค่าความถูกต้องของการพยากรณ์โอกาสการเกิดฝนอยู่ในระดับสูง

2. ในการวิเคราะห์ปัจจัยหลังจากทำการหมุนแกนแล้วค่าที่ได้ยังไม่สามารถตั้งชื่อกลุ่มปัจจัยให้มีความสอดคล้องสัมพันธ์กับข้อมูล GPCM ที่ทำให้เกิดฝนทั้ง 6 กลุ่ม ที่ประกอบด้วย 1) ตัวแปรกลุ่มอุณหภูมิ 2) ตัวแปรกลุ่มความชื้น 3) ตัวแปรกลุ่มความสูงและความกดอากาศ 4) ตัวแปรกลุ่มกระแสลม 5) ตัวแปรกลุ่มการทรงตัวของบรรยากาศและค่าพยากรณ์อากาศ และ 6) ตัวแปรกลุ่มค่าพยากรณ์โอกาสความสำเร็จในการทำฝนจากเมฆอุ่น อีกทั้งตัวแปรที่เกี่ยวข้องกับการเกิดฝนบางตัวแปรเป็นผลลัพธ์จากการคำนวณมาจากตัวแปรอื่น ๆ ทำให้มีความซ้ำซ้อนของข้อมูลเกิดขึ้นระหว่างตัวแปร จึงทำให้ยากต่อการแยกหรือจัดกลุ่มให้ได้ตามทฤษฎีของกลุ่มตัวแปรการเกิดฝน

3. ตัวแบบที่ 8 ใช้ 28 ตัวแปร มาสกัดปัจจัยได้ 3 ปัจจัยร่วม แต่อีก 17 ตัวแปรปริมาณที่เหลือจาก 45 ตัวแปร ไม่ได้นำมาพิจารณาในตัวแบบถดถอยโลจิสติก อย่างไรก็ตามผลจากตัวแบบที่ 8 ใช้ตัวแปรเดิมทั้งหมด 6 ตัว ให้ค่าความถูกต้องของการพยากรณ์สูงถึงร้อยละ 90.25 ซึ่งไม่สามารถเปรียบเทียบกับผลการศึกษาของลีลี (2550) ได้โดยตรง เนื่องจากใช้เทคนิคการวิเคราะห์ที่แตกต่างกัน โดยที่การศึกษานี้ทำการวิเคราะห์จากข้อมูลทั้งหมด 277 รายการ 49 ตัวแปร และทำการวิเคราะห์เพียงหนึ่งครั้ง ขณะที่การศึกษาของลีลี (2550) วิเคราะห์จาก 309 รายการ 52 ตัวแปร โดยไม่มีตัวแปรการเกิดฝนของวันก่อนหน้า มีการแบ่งข้อมูลเป็นชุดข้อมูลฝึกสอน และชุดข้อมูลทดสอบตามหลัก Cross – validation มีการเรียนรู้ และทดสอบจากชุดข้อมูลทั้งสองหลายครั้ง จนได้ข้อสรุปตัวแปรที่มีความสัมพันธ์ต่อการเกิดฝน 13 ตัวแปร ค่าความถูกต้องของการพยากรณ์ร้อยละ 87.06 แต่อย่างไรก็ตามยังพบว่าทั้งสองการศึกษา มีตัวแปรที่มีความสำคัญต่อโอกาสการเกิดฝนตรงกัน 5 ตัวแปร คือ 1) AVG_RH_10_18000_FT 2) PRECIPITABLE_WATER_TOTA 3) PRECIPITABLE_WATER_SFC500 4) MEAN_RH_10000_15000_FT และ 5) AVG_RH_0_10000_FT

สรุปและข้อเสนอแนะ

สรุป

การวิเคราะห์ปัจจัยสำหรับตัวแบบคาดการณ์โอกาสการเกิดฝน ภาคตะวันออกเฉียงเหนือของประเทศไทย เพื่อตอบวัตถุประสงค์หลัก 2 ข้อ คือ 1) วิเคราะห์ปัจจัยร่วมที่เกี่ยวกับข้อมูลผลการตรวจอากาศชั้นบน ของภาคตะวันออกเฉียงเหนือ และ 2) หาตัวแบบการคาดการณ์โอกาสการเกิดฝนในภาคตะวันออกเฉียงเหนือ ชุดข้อมูลที่นำมาศึกษาครั้งนี้ประกอบด้วยข้อมูล GPCM ซึ่งเป็นข้อมูลรายวันของผลการตรวจวัดอากาศชั้นบนจากสำนักฝนหลวงและการบินเกษตร เชื่อมต่อกับข้อมูลปริมาณน้ำฝน 24 ชั่วโมง ของภาคตะวันออกเฉียงเหนือ จากกรมอุตุนิยมวิทยา มีตัวแปรที่ใช้ในการศึกษาทั้งสิ้น 49 ตัวแปร เริ่มต้นตั้งแต่วันที่ 12 มีนาคม 2547 ถึงวันที่ 14 กันยายน 2549 มีข้อมูลที่สามารถวิเคราะห์ได้ 277 รายการ โดยเป็นวันที่ต่อเนื่องกันบ้างไม่ต่อเนื่องกันบ้าง

วัตถุประสงค์ข้อแรกเกี่ยวกับการวิเคราะห์ปัจจัยร่วมที่เกี่ยวกับการเกิดฝนในภาคตะวันออกเฉียงเหนือ พบว่ามีตัวแปรหรือปัจจัยข้อมูลตรวจอากาศที่เกี่ยวกับการเกิดฝนมีจำนวนมาก คือ 48 ตัวแปร จึงได้มีการศึกษาหาตัวแปรที่มีความสำคัญและสัมพันธ์ต่อการเกิดฝน โดยใช้วิธีการวิเคราะห์ปัจจัยเพื่อจัดกลุ่มตัวแปรให้เป็นปัจจัยผ่านกระบวนการสกัดปัจจัย ทำให้ได้ปัจจัยที่สามารถอธิบายตัวแปรเดิมได้อย่างเหมาะสม แยกการพิจารณาข้อมูลเบื้องต้นด้วยการไม่พิจารณาค่า VIF และพิจารณาค่า VIF ที่มากกว่าหรือเท่ากับ 10 ในขั้นตอนแรก ทำให้เมื่อไม่พิจารณาค่า VIF จะใช้ตัวแปรอิสระที่เป็นข้อมูลเชิงปริมาณ 45 ตัวแปร และเมื่อพิจารณาค่า VIF ที่มากกว่าหรือเท่ากับ 10 ในขั้นตอนแรกจะใช้ตัวแปรอิสระที่เป็นข้อมูลเชิงปริมาณ 28 ตัวแปร ใช้ในการวิเคราะห์ปัจจัยด้วยเทคนิคการสกัดปัจจัย 2 แบบ คือ เทคนิคตัวประกอบหลัก (PCM) และเทคนิคภาวะน่าจะเป็นสูงสุด (MLM) ในขั้นตอนนี้ทำให้ได้ค่าน้ำหนักปัจจัย (1_{ij}) ค่าความร่วมกัน (h_i^2) ค่าเฉพาะ (e_i) ค่าไอเก้น (λ_i) ค่าสัดส่วนความแปรปรวนที่อธิบายได้ (Proportion of Total Variance) และค่าสัดส่วนสะสมของความแปรปรวนที่อธิบายได้ (Cumulative Proportion) โดยค่า 1_{ij} ที่ได้จากการวิเคราะห์ในบางตัวแปรที่ยังคลุมเครือว่าควรอยู่ในปัจจัยใดจึงต้องทำการหมุนแกนปัจจัยด้วยวิธี Varimax เพื่อให้ปัจจัยร่วมยังคงเป็นอิสระต่อกัน และตัวแปรอิสระจะถูกจัดอยู่ในปัจจัยใดปัจจัยหนึ่งอย่างชัดเจน โดยกำหนดจำนวนปัจจัยร่วมที่เหมาะสมจากค่าไอเก้นที่มากกว่า 1 ร่วมกับใช้การพล็อตกราฟ Scree Plot สุดท้ายสำหรับการวิเคราะห์ปัจจัยจะคำนวณค่าคะแนนปัจจัย

จาก 2 วิธี คือวิธี Regression และวิธี Factor Based Score ทำให้มีกรณีการศึกษาทั้งหมด 8 กรณีสรุปได้ว่าเทคนิคตัวประกอบหลัก (PCM) และเทคนิคภาวะเป็นไปได้สูงสุด (MLM) ได้จำนวนปัจจัยร่วมที่เหมาะสมเท่ากัน คือได้จำนวนปัจจัยร่วม 4 ปัจจัย เมื่อไม่พิจารณาค่า VIF และได้จำนวนปัจจัยร่วม 3 ปัจจัย ในการพิจารณาค่า VIF ที่มากกว่าหรือเท่ากับ 10 โดยปัจจัยร่วมที่สร้างขึ้นให้ค่าสัดส่วนสะสมของความแปรปรวนที่อธิบายได้ทั้ง 8 ตัวแบบ เฉลี่ยร้อยละ 71 โดยที่วิธีการหาค่าคะแนนปัจจัยแบบ Factor Based Score จะใช้ข้อมูลตัวแปรเดิมน้อยกว่าวิธี Regression

ผลจากการวิเคราะห์ปัจจัยจะนำไปใช้ในการพัฒนาตัวแบบการคาดการณ์เพื่อตอบสนองวัตถุประสงค์ข้อที่ 2 เพื่อหาตัวแบบการณโอกาสการเกิดฝนในภาคตะวันออกเฉียงเหนือ โดยเลือกการพัฒนาตัวแบบการวิเคราะห์ถดถอยโลจิสติกแบบทวิ ศึกษาความสัมพันธ์ระหว่างปัจจัยที่สกัดได้จากวัตถุประสงค์ข้อที่ 1 ร่วมกับตัวแปรเชิงกลุ่มอีก 3 ตัวแปร คือ 1) CONVECTIVE_TEMP_REACHABLE (D_1) 2) WARM_CLOUD_SEEDING_POTENTIAL (D_2) และ 3) การเกิดฝนของวันก่อนหน้า (D_3) ที่มีผลต่อตัวแปรตาม คือ โอกาสการเกิดฝนประจำวันที่มีค่าเป็น 2 ลักษณะ คือ ฝนไม่ตก (0) และฝนตก (1) ทำการคัดเลือกตัวแปรอิสระด้วยวิธีขั้นตอนก้าวหน้าแบบอัตราส่วนภาวะน่าจะเป็น ที่กำหนดค่า Probability of Stepwise Entry (P_E) = 0.15 และ Probability of Stepwise Removal (P_R) = 0.20 สรุปได้ว่าทั้ง 8 ตัวแบบให้ค่าความถูกต้องเฉลี่ยร้อยละ 90

ผลจากการสร้างตัวแบบทั้ง 8 ตัวแบบ ในการพยากรณ์โอกาสความสำเร็จในเกิดฝนของภาคตะวันออกเฉียงเหนือ ที่มีความเหมาะสมมากที่สุด คือตัวแบบที่ 8 ที่พิจารณาตัวแปรที่มีค่า VIF มากกว่าหรือเท่ากับ 10 ก่อนการสกัดปัจจัยด้วยวิธี MLM แล้วคำนวณค่าคะแนนปัจจัยด้วยวิธี Factor Based Score เนื่องจากตัวแบบที่ 8 สร้างเป็นตัวแปรปัจจัยร่วมได้ 3 ปัจจัย จาก 17 ตัวแปรเดิม และสร้างตัวแบบด้วยวิธีการวิเคราะห์ถดถอยโลจิสติกแบบทวิ ร่วมกับตัวแปรเชิงกลุ่มอีก 3 ตัวแปร สร้างตัวแบบได้ $\log_e \left(\frac{P}{1-P} \right) = 1.675 + 0.322(F_2) + 1.335(D_3)$ ให้ค่าความถูกต้องของการพยากรณ์ร้อยละ 90.3 และค่าอำนาจของการพยากรณ์ $R^2_N = 0.472$ แต่ตัวแบบดังกล่าวมิได้ให้การพยากรณ์สูงสุด เนื่องจากตัวแบบที่ 1 ให้ค่าความถูกต้องของการพยากรณ์ร้อยละ 91 และ $R^2_N = 0.529$ แต่มีการใช้จำนวนตัวแปรอิสระเดิมมากถึง 42 ตัวแปร ดังนั้นจึงเลือกตัวแบบที่ 8 ที่มีความสามารถในการอธิบายความสัมพันธ์ของตัวแปรอิสระที่ผ่านการคัดเลือก ซึ่งประกอบด้วย 2 ตัวแปร คือ ตัวแปรการเกิดฝนของวันก่อนหน้า (D_3) และตัวแปรปัจจัยร่วมที่ 2 (F_2) ที่สร้างจากตัวแปรเดิม 5 ตัวแปร ประกอบด้วย 1) AVG_RH_10_18000_FT (X243) 2) PRECIPITABLE_WATER_

TOTAL (X214) 3) PRECIPITABLE_WATER_SFC500 (X213) 4) MEAN_RH_10000_15000_FT (X246) และ 5) AVG_RH_0_10000_FT (X242) ซึ่งทั้งหมดเป็นตัวแปรในกลุ่มความชื้น จากข้อมูล GPCM ที่สร้างเป็นตัวแทนที่มีความเหมาะสมที่สุดในการคาดการณ์โอกาสการเกิดฝน ภาควัดวันออกเฉียงเหนือ ของประเทศไทย ในการศึกษา

ข้อเสนอแนะ

เนื่องจากข้อมูลที่ศึกษามีปริมาณน้อย (277 รายการ) มีตัวแปรอิสระจำนวนมาก (48 ตัวแปร) และยังขาดความต่อเนื่องของข้อมูล จึงไม่สามารถวิเคราะห์แบบอนุกรมเวลาได้ ดังนั้นในการศึกษาต่อไปมีข้อเสนอแนะดังนี้

1. หากมีข้อมูลครบถ้วนสมบูรณ์ตามรายการตัวแปรที่มีความสำคัญต่อการเกิดฝนในภาคตะวันออกเฉียงเหนือ และมีการบันทึกข้อมูลประจำวันอย่างต่อเนื่อง ควรทำการวิเคราะห์ด้วยแบบทางสถิติในเชิงอนุกรมเวลาโดยใช้ Structural equation modeling with time series (SEM) หรือ Markov model หรือ Neural network ดูการเปลี่ยนแปลงที่เกิดขึ้นระหว่างตัวแปรต่าง ๆ อย่างต่อเนื่อง

2. ถ้าเป็นไปได้ควรทำการวิเคราะห์ปัจจัย และสร้างแบบจำลองพยากรณ์โอกาสการเกิดฝนภาคตะวันออกเฉียงเหนือ แยกตามฤดูกาล โดยเฉพาะในฤดูแล้งของภาคตะวันออกเฉียงเหนือ ของประเทศไทย

เอกสารและสิ่งอ้างอิง

กรมพัฒนาที่ดิน. 2551. ความแห้งแล้งซ้ำซากสู่ภาวะการณ์เป็นทะเลทรายของประเทศไทย. สารานุกรมจากศูนย์สนเทศภาคเหนือ. แหล่งที่มา: http://library.com.ac.th/ntic/knowledge_show.php?docid=20.htm, 5 กรกฎาคม 2552.

กระทรวงเกษตรและสหกรณ์. 2552. ความรู้ในการปฏิบัติการฝนหลวง. ชนิดของ เมฆจำแนกตามลักษณะการเกิดและความสูง. แหล่งที่มา: <http://www.moac.go.th/builder/brradkm/images/Decision%20making.doc>, 22 มีนาคม 2552

..... 2550. แนวทางการปฏิบัติการฝนหลวงช่วยเหลือพื้นที่แล้งซ้ำซาก ของศูนย์ปฏิบัติการฝนหลวงประจำภาค. องค์ความรู้ด้านการปฏิบัติการฝนหลวงของ สำนักฝนหลวงและการบินเกษตรปี พ.ศ.2550. แหล่งที่มา: <http://www.moac.go.th/builder/brradkm/>, 1 พฤษภาคม 2552.

..... ม.ป.ป. เกณฑ์ช่วยการตัดสินใจในการปฏิบัติการฝนหลวง. เอกสาร เผยแพร่. แหล่งที่มา: <http://www.moac.go.th/builder/brradkm/images/5001Decisionmaking.doc>, 5 กรกฎาคม 2552.

เกษม จันทร์แก้ว. 2526. หลักการจัดการลุ่มน้ำ. ภาควิชาอนุรักษ์วิทยา คณะวนศาสตร์ มหาวิทยาลัยเกษตรศาสตร์, กรุงเทพฯ.

คุณฉวี สุขวัฒน์. 2543. การพยากรณ์อากาศด้วยคอมพิวเตอร์. แหล่งที่มา: <http://tmd.motc.go.th/nwp/>, 20 มีนาคม 2552.

ต้นสกุล สานติบุรณ. 2549. อุตุนิยมวิทยา. คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏอุดรธานี.

ทรง กลิ่นประทุม. 2551. **มาตรฐานศูนย์ปฏิบัติการฝนหลวงประจำภาค**. ผลงานวิจัย. แหล่งที่มา:
<http://www.royalrainmaking.thaigov.net>, 5 กรกฎาคม 2552.

นงลักษณ์ วิรัชชัย. 2538. **ความสัมพันธ์เชิงโครงสร้าง (LISREL) : สถิติวิเคราะห์สำหรับการวิเคราะห์**. กรุงเทพฯ : สำนักพิมพ์จุฬาลงกรณ์มหาวิทยาลัย.

ปกรณ์ อินทะเทพ. 2545. **โครงการประชาสัมพันธ์สำหรับการพยากรณ์ฝนตกในเขตกรุงเทพมหานคร**. วิทยานิพนธ์มหาบัณฑิต. มหาวิทยาลัยเกษตรศาสตร์, กรุงเทพฯ.

ฝ่ายทรัพยากรธรรมชาติและสิ่งแวดล้อม สถาบันวิจัยเพื่อการพัฒนาประเทศไทย. 2545. **ทรัพยากรดินและการใช้ที่ดิน**. การจัดลำดับความสำคัญของปัญหาทรัพยากรธรรมชาติและสิ่งแวดล้อม. แหล่งที่มา:
<http://www.thaienvimonitor.net/Concept/priority3.htm>, 5 กรกฎาคม 2552.

พิมลรัตน์ รัตนเพชร. 2547. **การวิเคราะห์การถดถอยที่ตัวแปรตามมีค่าเป็น 2 ลักษณะ ในกรณีที่มีการแปลงข้อมูล**. สาขาสถิติ วิทยานิพนธ์ปริญญาโท. มหาวิทยาลัยเกษตรศาสตร์, กรุงเทพฯ.

มณฑล จิตเอื้ออารีกุล. 2547. **การประมาณปริมาณฝนด้วยดาวเทียมสำหรับกรุงเทพมหานครในช่วงมรสุมตะวันออกเฉียงเหนือ**. วิทยานิพนธ์มหาบัณฑิต. มหาวิทยาลัยเกษตรศาสตร์, กรุงเทพฯ.

ยุทธ ไถยวรรณ. 2551. **วิเคราะห์ข้อมูลวิจัย 4**. ศูนย์สื่อเสริมกรุงเทพ. กรุงเทพฯ : พิมพ์ดี.

รังสรรค์ อาภาศัพท์กุล. 2547. **อุตุนิยมวิทยาเบื้องต้น**. กรุงเทพฯ : สำนักพิมพ์จุฬาลงกรณ์มหาวิทยาลัย.

ลีลี อิงศรีสว่าง. 2550. **โครงการพัฒนาระบบพยากรณ์อากาศและโอกาสความสำเร็จในการปฏิบัติการฝนหลวงภาคตะวันออกเฉียงเหนือ**. สำนักฝนหลวงและการบินเกษตร สำนักปลัดกระทรวงเกษตรและสหกรณ์, กรุงเทพฯ.

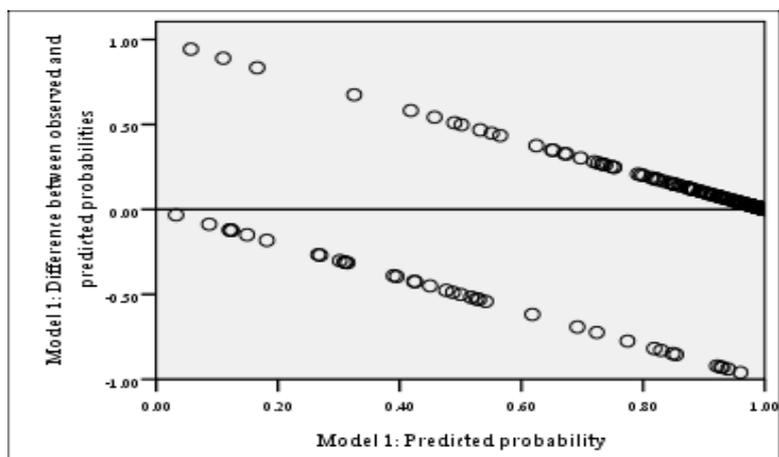
- วรารุช ชันดิยานันท์. 2540. **วิเคราะห์ผลพยากรณ์อากาศโดย GPCM**. สำนักฝนหลวง และการบินเกษตร สำนักปลัดกระทรวงเกษตรและสหกรณ์, กรุงเทพฯ.
- ศิริชัย กาญจนวาสี. 2550. **การวิเคราะห์พหุระดับ**. กรุงเทพฯ : สำนักพิมพ์จุฬาลงกรณ์ มหาวิทยาลัย.
- สถาบันวิจัยเพื่อการพัฒนาประเทศไทย. ม.ป.ป. **ทรัพยากรน้ำ**. เอกสารเผยแพร่. แหล่งที่มา: <http://www.thaienvimonitor.net/Concept/priority2.htm>, 27 กันยายน 2552.
- สมยศ ตั้งเจริญจิตกุล. 2547. **การประยุกต์ใช้โครงข่ายประสาทเทียมในการทำนายระดับน้ำแม่น้ำปิง ณ อำเภอเมือง จังหวัดเชียงใหม่**. วิทยานิพนธ์มหาบัณฑิต. มหาวิทยาลัยเกษตรศาสตร์, กรุงเทพฯ.
- Chatterjee, S. 1991. **Regression analysis by Example**. 2nd ed. A Wiley-Interscience publication. Printed in the United States of America.
- Christensen, L.B. and C.M. Stoup. 1991. **Introduction to statistics for the social and behavioral sciences**. 2nd ed. Pacific Grove, Calif.
- Draper, N.R. and H. Smith. 1998. **Applied Regression Analysis**. 3rd ed. John Willey and Sons, U.S.A.
- Garson, D.G. 2009. **Factor Analysis**. Available Source: <http://Faculty.chass.ncsu.edu/garson/PA765/Factor.htm>, July 6, 2009.
- _____. 2009. **Logistic Regression**. Available Source: <http://Faculty.chass.ncsu.edu/garson/PA765/logistic.htm>, October 5, 2009.
- Hair, J. F. and W.C. Black, 1998. **Multivariate Data Analysis**. 5 ed. Prentice-Hall International, London.

- Hatcher, L. 1998. **A Step-by-Step Approach to Using the SAS System for Factor Analysis and Structural Equation Modeling**, 3rd ed. Cary, NC:SAS Institute. Focus on the CALIS procedure.
- Herve, A. 2003. **Factor Rotations in Factor Analyses**. Factor Analysis. Available Source: <http://www.utdallas.edu/~herve/Abdi-rotations-pretty.pdf>, July 6, 2009.
- Hosmer, D.W. and S. Lemeshow. 2000. **Applied Logistic Regression**. 2 nd ed. John Willey and Sons, Inc. New York.
- Johnson, R.A. and D.W. Wichern, 2002. **Applied Multivariate Statistical Analysis**, 5th ed., Prentice Hall, Upper Saddle River, New Jersey.
- Kutner, M.H., C.J. Nachtsheim, W. Li and J. Neter. 2005. **Applied Linear Statistical Models**. 5th ed. Published by McGraw-Hill Education Printed in Singapore.
- Neter, J., W. Wasserman and M.H. Kutner. 1989. **Applied Linear Regression Models**. 2nd ed. Homewood, Illinois : Richard D. Irwin, Inc., U.S.A.
- Ranacher, A.C. 1995. **Methods of Multivariate Analysis**. John Wiley and Sons, Inc.
- Ravindra, K. and D.N. Naik. 2000. **Multivariate data reduction and discrimination with SAS Software**, 8 ed. Cary, NC: SAS Institute Inc.
- Tabachnick, B.G. and L.S. Fidell, 2007. **Using Multivariate Statistics**. 5th ed. Harper Collins College, New York.
- Williams, R. n.d. **Scalar Measures of Fit : Pseudo R^2 and Information Measures**. Available Source: <http://www.nd.edu/~rwilliam/xsoc73994/L04.pdf>, October 5, 2009.

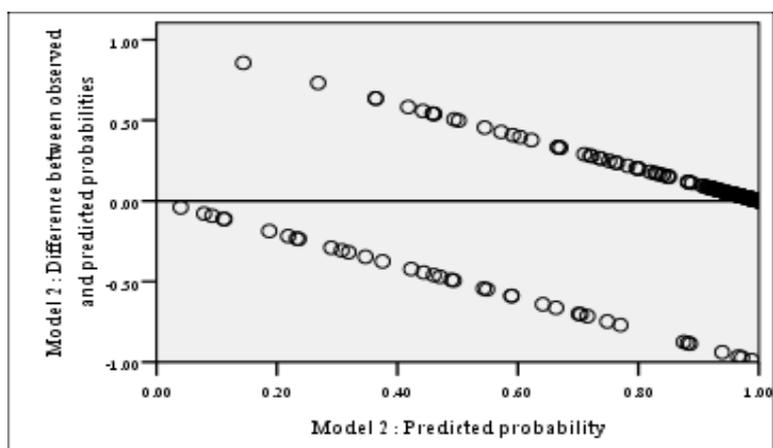
ภาคผนวก

ภาคผนวก ก

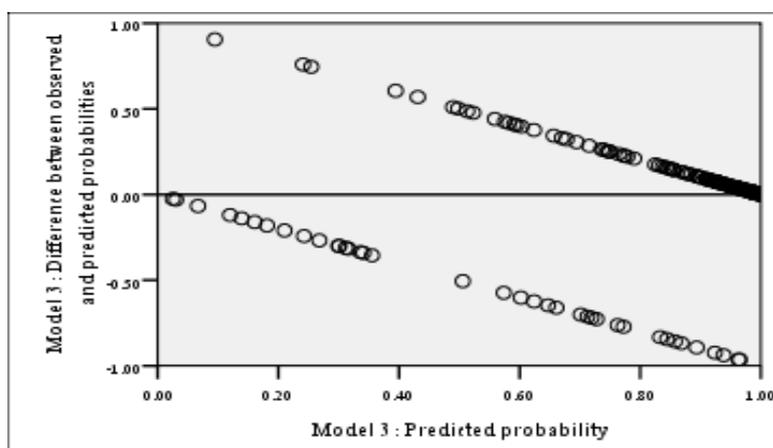
การพล็อตค่า Residual คู่กับ Predicted probability (\hat{Y})
ของแต่ละตัวแบบ



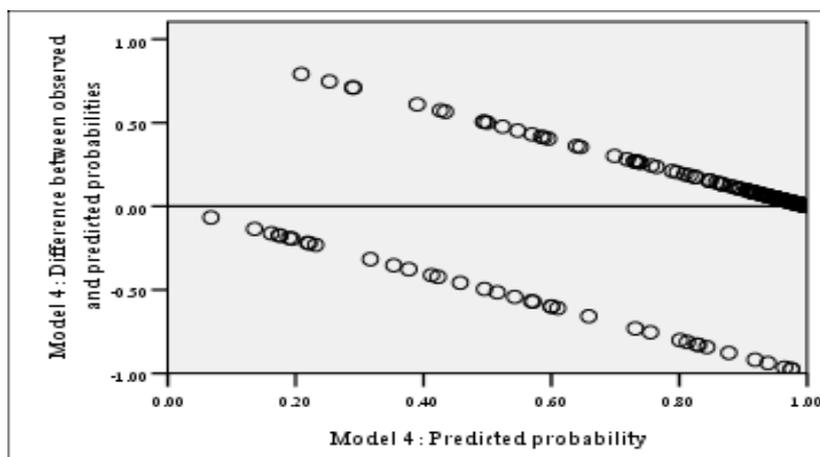
ภาพผนวกที่ ก1 การพล็อตค่า Residual คู่กับ Predicted probability (\hat{Y}) ของตัวแบบที่ 1



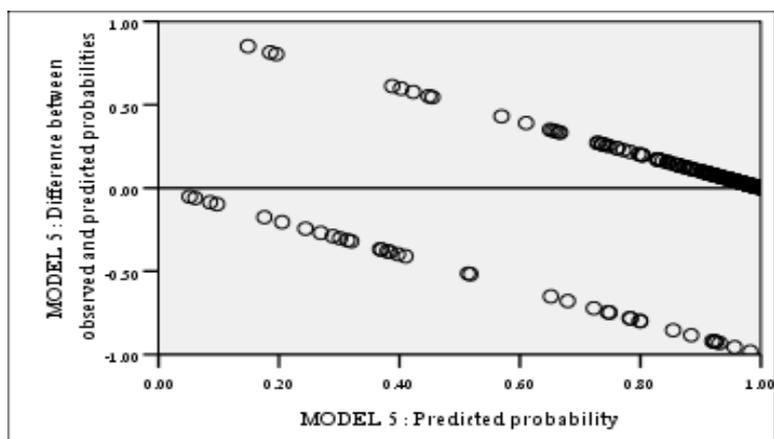
ภาพผนวกที่ ก2 การพล็อตค่า Residual คู่กับ Predicted probability (\hat{Y}) ของตัวแบบที่ 2



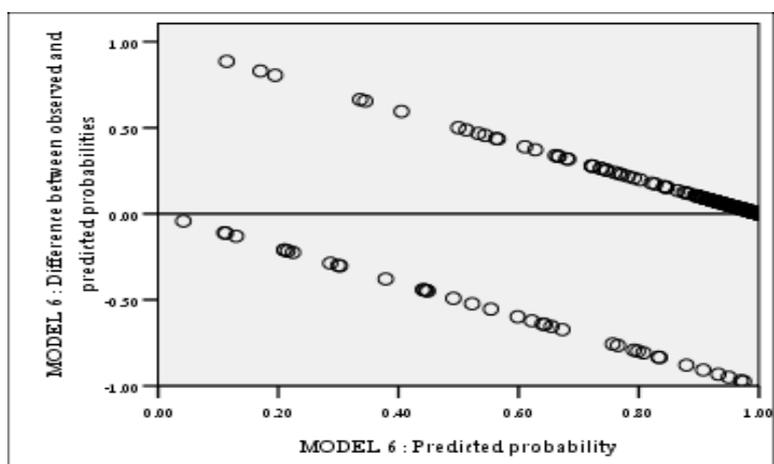
ภาพผนวกที่ ก3 การพล็อตค่า Residual คู่กับ Predicted probability (\hat{Y}) ของตัวแบบที่ 3



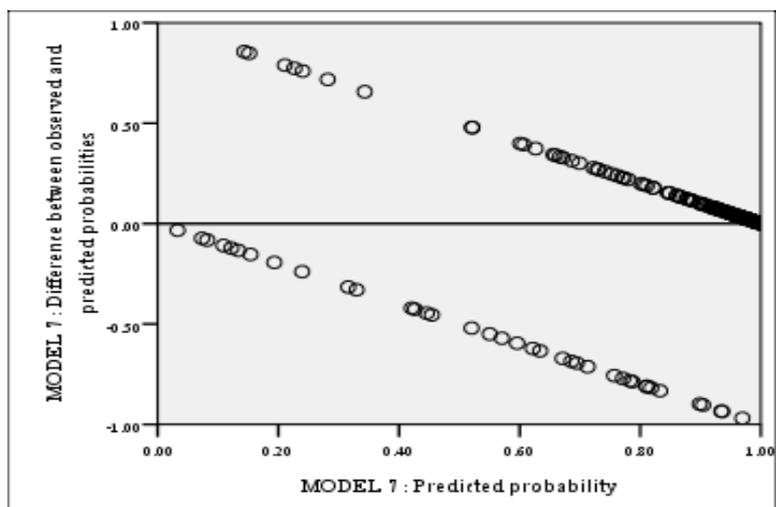
ภาพผนวกที่ ๓ การพล็อตค่า Residual คู่กับ Predicted probability (\hat{Y}) ของตัวแบบที่ 4



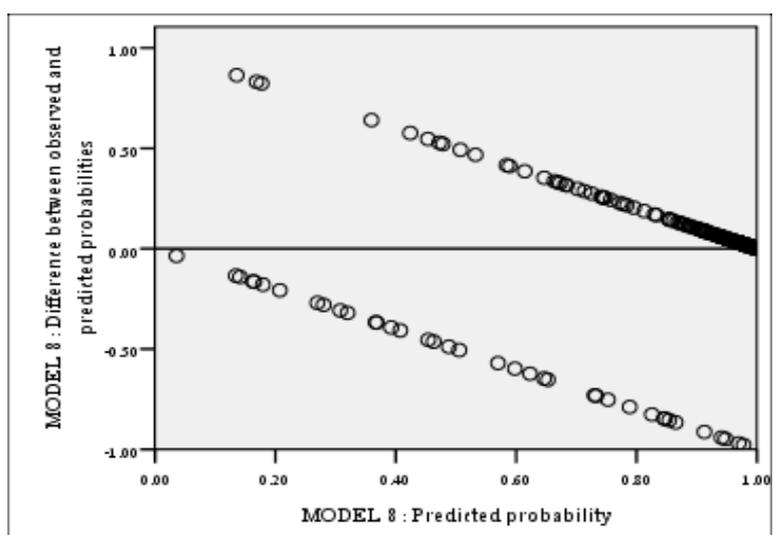
ภาพผนวกที่ ๔ การพล็อตค่า Residual คู่กับ Predicted probability (\hat{Y}) ของตัวแบบที่ 5



ภาพผนวกที่ ๕ การพล็อตค่า Residual คู่กับ Predicted probability (\hat{Y}) ของตัวแบบที่ 6



ภาพผนวกที่ ก7 การพล็อตค่า Residual คู่กับ Predicted probability (\hat{Y}) ของตัวแบบที่ 7



ภาพผนวกที่ ก8 การพล็อตค่า Residual คู่กับ Predicted probability (\hat{Y}) ของตัวแบบที่ 8

ประวัติการศึกษา และการทำงาน

ชื่อ –นามสกุล	นางสาวคณพรพร พันธ์พานิชย์
วัน เดือน ปี ที่เกิด	13 ตุลาคม พ.ศ.2523
สถานที่เกิด	อำเภอเมือง จังหวัดอุดรธานี
ประวัติการศึกษา	ระดับปริญญาตรี คณะวิทยาศาสตร์ วิชาสถิติ มหาวิทยาลัยมหาสารคาม
ตำแหน่งหน้าที่การงานปัจจุบัน	นักชีวสถิติ หน่วยวิจัยกองอายุรกรรม
สถานที่ทำงานปัจจุบัน	โรงพยาบาลพระมงกุฎเกล้า (รพ.ร.๖)
ผลงานดีเด่นและรางวัลทางวิชาการ	-
ทุนการศึกษาที่ได้รับ	-