ตัวแบบทางสถิติในการศึกษาภาวะการมีงานทำของประชากรไทย
Statistical Modeling to Study Employment Status of Thai Population

อำภาวรรณ์ จันทร์คง
Amphawan Chankhong

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญา
วิทยาศาสตรมหาบัณฑิต สาขาวิชาสถิติ
มหาวิทยาลัยสงขลานครินทร์
A Thesis Submitted in Partial Fulfillment of the Requirements for the
Degree of Master of Science in Statistics
Prince of Songkla University
2563
ลิขสิทธิ์ของมหาวิทยาลัยสงขลานครินทร์

ตัวแบบทางสถิติในการศึกษาภาวะการมีงานทำของประชากรไทย
Statistical Modeling to Study Employment Status of Thai Population

อำภาวรรณ์ จันทร์คง
Amphawan Chankhong

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญา
วิทยาศาสตรมหาบัณฑิต สาขาวิชาสถิติ
มหาวิทยาลัยสงขลานครินทร์
A Thesis Submitted in Partial Fulfillment of the Requirements for the
Degree of Master of Science in Statistics
Prince of Songkla University
2563
ลิขสิทธิ์ของมหาวิทยาลัยสงขลานครินทร์

ชื่อวิทยานิพนธ์ ตัวแบบทางสถิติในการศึกษาภาวะการมีงานทำของประชากรไทย
ผู้เขียน นางสาวอำภาวรรณ์ จันทร์คง
สาขาวิชา สถิติ

อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

.....
(ผู้ช่วยศาสตราจารย์ ดร.ใกล้รุ่ง สามารถ)

คณะกรรมการสอบ

.....ประธานกรรมการ
(รองศาสตราจารย์ ดร.วราฤทธิ์ พานิชกิจโกศลกุล)

.....กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.ธรรมรัตน์ พนิชยากุล)

บัณฑิตวิทยาลัย มหาวิทยาลัยสงขลานครินทร์ อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษา ตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาสถิติ

.....
(ศาสตราจารย์ ดร.ดำรงศักดิ์ ฟ้ารุ่งสาง)
คณบดีบัณฑิตวิทยาลัย

ขอรับรองว่า ผลงานวิจัยนี้มาจากการศึกษาวิจัยของนักศึกษาเอง ได้แสดงความขอบคุณบุคคลที่มีส่วนช่วยเหลือแล้ว

ลงชื่อ.....
(ผู้ช่วยศาสตราจารย์ ดร.ใกล้รุ่ง สามารถ)
อาจารย์ที่ปรึกษาวิทยานิพนธ์

ลงชื่อ.....
(นางสาวอำภาวรรณ์ จันทร์คง)
นักศึกษา

ขอรับรองว่า ผลงานวิจัยนี้ไม่เคยเป็นส่วนหนึ่งในการอนุมัติปริญญาในระดับใดมาก่อน และไม่ได้ถูกใช้ในการยื่นขออนุมัติปริญญาในขณะนี้

ลงชื่อ.....
(นางสาวอำภาวรรณ์ จันทร์คง)
นักศึกษา

ชื่อวิทยานิพนธ์ ตัวแบบทางสถิติในการศึกษาภาวะการมีงานทำของประชากรไทย
ผู้เขียน นางสาวอำภาวรรณ์ จันทร์คง
สาขาวิชา สถิติ
ปีการศึกษา 2563

บทคัดย่อ

สถานะการจ้างงานและรายได้ของคนในประเทศ จะเป็นตัวบ่งชี้ว่าเศรษฐกิจของประเทศนั้นเป็นไปในทิศทางใด การศึกษานี้มีวัตถุประสงค์ เพื่อหาตัวแบบที่เหมาะสมในการทำนายสภาวะการมีงานทำของประชากรไทย ตลอดจนทำนายรายได้ของประชากรไทย โดยใช้ตัวแบบการถดถอยลอจิสติก และตัวแบบการถดถอยเชิงเส้นพหุคูณ ผลการวิเคราะห์พบว่า เพศชาย (Odds ratio = 2.187) ที่อาศัยอยู่ในภาคกลางของประเทศไทย (Odds ratio = 1.231) และสำเร็จการศึกษาระดับบัณฑิตศึกษา (Odds ratio = 4.516) มีโอกาสได้งานสูงกว่าเมื่อเทียบกับคนทั่วไป ในขณะที่ผู้ชายที่สำเร็จการศึกษาระดับบัณฑิตศึกษา ประกอบอาชีพข้าราชการ เสมียน หรือช่างเทคนิค จะมีรายได้เฉลี่ยสูงกว่าคนทั่วไป

Thesis Title Statistical Modeling to Study Employment Status of Thai Population
Author Miss Amphawan Chankhong
Major Program Statistics
Academic Year 2563

ABSTRACT

Employment status and income of people in the country are indicators of how the economy of that country is going. The objective of this study is to fit an appropriate model to predict the employment status of the Thai population as well as to predict the income of the Thai population by using the logistic regression and multiple regression models. The analysis results showed that male (Odds ratio = 2.187) who live in central Thailand (Odds ratio = 1.231) and have post graduate studies (Odds ratio = 4.516) have a higher chance of getting a job compared to the general people. While the male who have post graduate studies, being government employee and having occupation (Public servant, Clerk and Technician), have a higher average income than general people.

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จเรียบร้อยสมบูรณ์ เนื่องจากได้รับความกรุณาอย่างดียิ่งจากผู้ช่วยศาสตราจารย์ ดร.ใกล้รุ่ง สามารถ อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่กรุณาให้คำปรึกษาแนะนำ แก้ไขข้อบกพร่องต่าง ๆ ตลอดจนให้ข้อคิดอันเป็นประโยชน์ต่อการวิจัยเป็นอย่างยิ่ง ทั้งยังให้กำลังใจ และติดตามสอบถามเกี่ยวกับการเรียนด้วยความห่วงใยแก่ผู้วิจัยเป็นอย่างดีมาโดยตลอด ผู้วิจัยขอกราบขอบพระคุณเป็นอย่างสูงมา ณ โอกาสนี้

ขอกราบขอบคุณ ผู้ช่วยศาสตราจารย์ ดร.ธรรมรัตน์ พนิชยากุล จากภาควิชาคณิตศาสตร์และสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยสงขลานครินทร์ วิทยาเขตหาดใหญ่ และรองศาสตราจารย์ ดร.วราฤทธิ์ พานิชกิจโกศลกุล จากภาควิชาคณิตศาสตร์และสถิติ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์ (ศูนย์รังสิต) ที่ได้กรุณารับเป็นกรรมการสอบวิทยานิพนธ์ และได้ให้คำแนะนำช่วยให้วิทยานิพนธ์ฉบับนี้สมบูรณ์ยิ่งขึ้น

ขอกราบขอบพระคุณคณาจารย์ภาควิชาคณิตศาสตร์และสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยสงขลานครินทร์ วิทยาเขตหาดใหญ่ ที่ได้กรุณาประสิทธิ์ประสาทความรู้ทางด้านคณิตศาสตร์และสถิติ ทำให้ผู้วิจัยมีความรู้ความสามารถ และสามารถนำความรู้ที่ได้ไปประยุกต์ใช้ให้เกิดประโยชน์สูงสุด

ขอขอบคุณผู้สนับสนุนทุนการศึกษา “ผลการเรียนดีเด่นเข้าศึกษาในระดับบัณฑิตศึกษา” ที่ให้โอกาสแก่ผู้วิจัยได้เข้าศึกษาต่อระดับปริญญาโท หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาสถิติ ขอขอบคุณมา ณ โอกาสนี้ด้วย

ความสำเร็จและคุณค่าของงานวิจัยครั้งนี้ ผู้วิจัยขอมอบให้แก่ นายอำมร จันทร์คง นางวันดี จันทร์คง นางสาวอาภัสรา จันทร์คง และนางสาวอมาวดี จันทร์คง ผู้เป็นบิดา มารดา พี่สาว และน้องสาวของผู้วิจัย ขอกราบขอบพระคุณเป็นอย่างสูงที่ให้กำลังใจ ให้คำปรึกษา ช่วยเหลือและส่งเสริมสนับสนุนด้านการศึกษาด้วยดีมาโดยตลอดจนประสบความสำเร็จในปัจจุบัน ขอขอบคุณเพื่อน ๆ ของผู้วิจัยทุกคนที่ให้ความช่วยเหลือในการแก้ไขข้อบกพร่องของวิทยานิพนธ์ฉบับนี้ คอยแนะแนวทางแก้ปัญหา คอยให้กำลังใจ และให้ความช่วยเหลือตลอดมา และขอขอบพระคุณผู้ที่เกี่ยวข้องทุกท่านที่ไม่ได้ระบุมา ณ ที่นี้ ที่กรุณาให้ความช่วยเหลือแก่ผู้วิจัยมาโดยตลอด

อำภาวรรณ์ จันทร์คง

สารบัญ

บทคัดย่อ 5
ABSTRACT 6
กิตติกรรมประกาศ 7
สารบัญ 8
รายการตาราง 10
รายการภาพประกอบ 12
บทที่ 1 บทนำ 1
1.1 ความสำคัญและที่มาของการวิจัย 1
1.2 วัตถุประสงค์ 2
1.3 ผลที่คาดว่าจะได้รับ 2
1.4 ขอบเขตงานวิจัย 2
1.4.1 ขอบเขตข้อมูลที่ศึกษา 2
1.4.2 ขอบเขตประชากรและตัวอย่าง 2
1.5 นิยามศัพท์เฉพาะ 3
บทที่ 2 เอกสารและงานวิจัยที่เกี่ยวข้อง 4
2.1 แนวคิดและทฤษฎีเกี่ยวกับการทำงาน 4
2.2 การวิเคราะห์การถดถอยลอจิสติก (Logistic regression analysis) 7
2.3 การวิเคราะห์การถดถอย (Regression Analysis) 13
2.4 งานวิจัยที่เกี่ยวข้อง 20
2.5 กรอบแนวคิด 22
บทที่ 3 วิธีดำเนินงานวิจัย 23
3.1 ข้อมูลที่ใช้ในการศึกษา 23
3.2 ตัวแปรที่ใช้ในการศึกษา 23
3.3 การนิยามและกำหนดค่าตัวแปร 24
3.4 การวิเคราะห์ข้อมูล 25
3.4.1 การวิเคราะห์ข้อมูลด้วยสถิติเชิงพรรณนา (Descriptive statistics) 25
3.4.2 การวิเคราะห์การถดถอยลอจิสติก (Logistic regression analysis) 25
3.4.3 การวิเคราะห์การถดถอยเชิงเส้นพหุคูณ (Multiple Linear Regression Analysis) 26
3.5 แผนภาพขั้นตอนการวิเคราะห์ข้อมูล 27
บทที่ 4 ผลการวิเคราะห์ข้อมูล 28

4.1 ผลการวิเคราะห์ข้อมูลเบื้องต้น....28
4.2 การวิเคราะห์การถดถอยลอจิสติกของสถานะการจ้างงานของประชากรไทย....30
4.3 การวิเคราะห์การถดถอยเชิงเส้นพหุคูณของรายได้ของประชากรไทย....36
4.3.1 ผลการวิเคราะห์ข้อมูลเบื้องต้น....36
4.3.2 การวิเคราะห์การถดถอยเชิงเส้นพหุคูณของรายได้ของประชากรไทย....38
บทที่ 5 สรุปผลการวิเคราะห์ อภิปราย และข้อเสนอแนะ....29
5.1 สรุปผลการวิจัย....29
5.1.1 สรุปข้อมูลเบื้องต้น....29
5.1.2 การวิเคราะห์การถดถอยลอจิสติกของสถานะการจ้างงานของประชากรไทย....51
5.1.3 การวิเคราะห์การถดถอยเชิงเส้นพหุคูณของรายได้ของประชากรไทย....52
5.2 อภิปรายผลการวิจัย....54
5.3 ข้อเสนอแนะ....55
บรรณานุกรม....56
ภาคผนวก 1 ตัวอย่างข้อมูล....58
ภาคผนวก 2 คำสั่ง R และผลวิเคราะห์....60
ภาคผนวก 3 ตัวอย่างแบบสอบถาม....68
ภาคผนวก 4 การตีพิมพ์เผยแพร่ผลงาน....74
ประวัติผู้เขียน....84

รายการตาราง

ตารางที่ 2.1 รูปแบบการถดถอยที่ไม่เป็นเส้นตรง (แต่มีความแปรปรวนของความคลาดเคลื่อนคงที่) และการแปลงข้อมูล 16
ตารางที่ 2.2 การแปลงบอกซ์-คอกซ์ 17
ตารางที่ 3.1 การนิยามและกำหนดค่าตัวแปร 24
ตารางที่ 4.1 ผลการวิเคราะห์ข้อมูลเบื้องต้นของผู้ตอบแบบสอบถามจำนวน 172,241 คน 28
ตารางที่ 4.2 Deviance Residual และ Degrees of Freedom และแสดงค่าสถิติ AIC ของตัวแบบ (4.2) 31
ตารางที่ 4.3 ผลการวิเคราะห์การตรวจสอบการแจกแจงปรกติของค่าคลาดเคลื่อนสุ่มโดย Kolmogorov-Smirnov test ของตัวแบบ (4.2) 32
ตารางที่ 4.4 ผลการวิเคราะห์การตรวจสอบการความแปรปรวนของค่าคลาดเคลื่อนสุ่มโดย Bartlett test of homogeneity of variances ของตัวแบบ (4.2) 32
ตารางที่ 4.5 ค่าประมาณพารามิเตอร์จากการวิเคราะห์การถดถอยลอจิสติกจากตัวแบบ (4.2) 34
ตารางที่ 4.6 การวิเคราะห์การจำแนกกลุ่มของตัวแบบ (4.2) 35
ตารางที่ 4.7 ผลการวิเคราะห์ข้อมูลเบื้องต้นของผู้มีรายได้จำนวน 47,933 คน 36
ตารางที่ 4.8 ค่าสัมประสิทธิ์การตัดสินใจ (R^2) ค่าความคลาดเคลื่อนมาตรฐานของส่วนเหลือ (Residual Standard Error) ค่าสถิติ AIC และค่าสถิติ BIC ของตัวแบบ (4.4) 39
ตารางที่ 4.9 ผลการวิเคราะห์การตรวจสอบการแจกแจงปรกติของค่าคลาดเคลื่อนสุ่มโดย Kolmogorov-Smirnov test ของตัวแบบ (4.4) 40
ตารางที่ 4.10 ผลการวิเคราะห์การตรวจสอบการความแปรปรวนของค่าคลาดเคลื่อนสุ่มโดย Bartlett test of homogeneity of variances ของตัวแบบ (4.4) 40
ตารางที่ 4.11 ค่าสัมประสิทธิ์การตัดสินใจ (R^2) ค่าความคลาดเคลื่อนมาตรฐานของส่วนเหลือ (Residual Standard Error) ค่าสถิติ AIC และค่าสถิติ BIC ของตัวแบบ (4.5) 42
ตารางที่ 4.12 ผลการวิเคราะห์การตรวจสอบการแจกแจงปรกติของค่าคลาดเคลื่อนสุ่มโดย Kolmogorov-Smirnov test ของตัวแบบ (4.5) 42
ตารางที่ 4.13 ผลการวิเคราะห์การตรวจสอบการความแปรปรวนของค่าคลาดเคลื่อนสุ่มโดย Bartlett test of homogeneity of variances ของตัวแบบ (4.5) 43
ตารางที่ 4.14 ค่าสัมประสิทธิ์การตัดสินใจ (R^2) ค่าความคลาดเคลื่อนมาตรฐานของส่วนเหลือ (Residual Standard Error) ค่าสถิติ AIC และค่าสถิติ BIC ของตัวแบบ (4.6) 44
ตารางที่ 4.15 ผลการวิเคราะห์การตรวจสอบการแจกแจงปรกติของค่าคลาดเคลื่อนสุ่มโดย Kolmogorov-Smirnov test ของตัวแบบ (4.6) 45

ตารางที่ 4.16 ผลการวิเคราะห์การตรวจสอบการความแปรปรวนของค่าคลาดเคลื่อนสุ่มโดย Bartlett test of homogeneity of variances ของตัวแบบ (4.6).....46
ตารางที่ 4.17 เปรียบเทียบตัวแบบ y_2 , $\ln(y_2)$ และ $1/y_2$ 47
ตารางที่ 4.18 ค่าประมาณพารามิเตอร์จากการวิเคราะห์การถดถอยพหุคูณจากตัวแบบ (4.5).....47

รายการภาพประกอบ

รูป 2.1 กราฟฟังก์ชันการถดถอยลอจิสติก 11
รูป 2.2 กรอบแนวคิด 22
รูป 3.1 แผนภาพขั้นตอนการวิเคราะห์ข้อมูล 27
รูป 4.1 Normal Q-Q plot และแผนภาพการกระจายของส่วนเหลือ (Residuals) ของตัวแบบ (4.2) 32
รูป 4.2 แผนภาพ Half-Normal Plot ของตัวแบบ (4.2) 33
รูป 4.3 Normal Q-Q plot และแผนภาพการกระจายของส่วนเหลือ (Residuals) ของตัวแบบ (4.4) 40
รูป 4.4 การแปลงข้อมูลด้วยบอกซ์-คอกซ์ (Box-Cox transformation) ของตัวแบบ (4.4) 41
รูป 4.5 Normal Q-Q plot และแผนภาพการกระจายของส่วนเหลือ (Residuals) ของตัวแบบ (4.5) 43
รูป 4.6 การแปลงข้อมูลด้วยบอกซ์-คอกซ์ (Box-Cox transformation) 45
รูป 4.7 Normal Q-Q plot และแผนภาพการกระจายของส่วนเหลือ (Residuals) ของตัวแบบ (4.6) 46

บทที่ 1
บทนำ

1.1 ความสำคัญและที่มาของการวิจัย

ปัจจุบันการว่างงานถือเป็นปัญหาลำดับต้น ๆ และเป็นต้นเหตุของปัญหาสังคมอื่น ๆ เนื่องจากการทำงานและการมีงานทำเป็นที่มาของรายได้ ที่กล่าวว่าการว่างงานเป็นปัญหาสังคมนั้น เนื่องจากว่าการมีงานทำมีความหมายต่อชีวิตมนุษย์นอกเหนือไปจากการหาเงินหรือการมีรายได้เพื่อดำรงชีพ แต่การทำงานยังเป็นตัวกำหนดพฤติกรรมหลาย ๆ อย่างของคน ตลอดจนสถานภาพทางสังคมและศักดิ์ศรีอันเกิดจากความแตกต่างระหว่างคนมีงานทำกับคนว่างงาน ภาวะการมีงานทำจะสะท้อนให้เห็นถึงความอยู่ดีมีสุขของคนในสังคม และสภาวะเศรษฐกิจของประเทศว่าเป็นเช่นไร รวมทั้งจะเป็นตัวบ่งบอกถึงความเจริญเติบโตทางเศรษฐกิจว่าประเทศมีสภาวะเศรษฐกิจที่ดีหรือมีสภาวะเศรษฐกิจที่ตกต่ำ

ผลการสำรวจภาวะการทำงานของประชากรเดือนธันวาคม พ.ศ. 2562 พบว่าผู้มีอายุตั้งแต่ 15 ปีขึ้นไป จำนวน 56.7 ล้านคน เป็นผู้ที่อยู่ในกำลังแรงงานหรือผู้ที่พร้อมจะทำงาน 38.21 ล้านคน ซึ่งประกอบด้วยผู้มีงานทำ 37.6 ล้านคน ผู้ว่างงาน 3.7 แสนคน และผู้ที่รอฤดูกาล 1.8 แสนคน ส่วนผู้ที่อยู่นอกกำลังแรงงานหรือผู้ที่ไม่พร้อมทำงาน จำนวน 18.49 ล้านคน ได้แก่ แม่บ้าน นักเรียน คนชรา เป็นต้น

สำหรับผู้ที่มีงานทำ 37.6 ล้านคน ประกอบด้วยผู้มีงานทำในภาคเกษตรกรรม 12.69 ล้านคน และนอกภาคเกษตรกรรม 24.97 ล้านคน เมื่อเปรียบเทียบกับเดือนธันวาคม พ.ศ. 2561 พบว่าผู้มีงานทำลดลง 2.1 แสนคน (จาก 37.87 ล้านคน เป็น 37.6 ล้านคน) จำนวนการว่างงานในเดือนธันวาคม พ.ศ. 2562 มีทั้งสิ้น 3.7 แสนคน หรือคิดเป็นอัตราการว่างงานร้อยละ 1 เมื่อเปรียบเทียบกับช่วงเวลาเดียวกันของปี 2561 จำนวนผู้ว่างงานเพิ่มขึ้น 1.8 หมื่นคน เมื่อพิจารณาอัตราการว่างงานจำแนกตามเพศพบว่า เพศชายมีอัตราการว่างงานสูงกว่าเพศหญิง โดยเพศชายมีอัตราการว่างงานร้อยละ 1.1 และเพศหญิงมีอัตราการว่างงานร้อยละ 0.8 ในขณะที่อัตราการว่างงานของผู้ที่สำเร็จการศึกษาในระดับอุดมศึกษามีจำนวน 1.5 แสนคน รองลงมาเป็นระดับมัธยมศึกษาตอนปลายหรือเทียบเท่า 7.6 หมื่นคน ประถมศึกษา 6.6 หมื่นคน ระดับมัธยมศึกษาตอนต้น 5.5 หมื่นคน และไม่มีการศึกษา 2.1 หมื่นคน เมื่อเปรียบเทียบจำนวนผู้ว่างงานในเดือนธันวาคม พ.ศ. 2561 และ พ.ศ. 2562 จำแนกตามภูมิภาคพบว่า ภาคตะวันออกเฉียงเหนือมีจำนวนผู้ว่างงานเพิ่มขึ้น 2.7 หมื่นคน ภาคกลางและภาคเหนือเพิ่มขึ้นเท่ากันคือ 1.0 หมื่นคน ในขณะที่กรุงเทพมหานครลดลง 2.7 หมื่นคน และภาคใต้ลดลง 2.0 พันคน (สำนักงานสถิติแห่งชาติ, 2562)

นอกจากนี้อาจมีปัจจัยหลายอย่างที่ส่งผลต่อสถานะการจ้างงานและรายได้ เช่น ปัจจัยด้านประชากร การศึกษา สถานภาพการสมรส ฯลฯ การว่างงานเป็นสิ่งที่น่าสนใจเนื่องจากเกี่ยวข้องกับภาวะทางเศรษฐกิจและสังคมที่สำคัญและเป็นสาเหตุของปัญหาสังคมอื่น ๆ (Cadden & Arnett, 2015)

ผู้วิจัยจึงได้ศึกษาเรื่อง “ตัวแบบทางสถิติในการศึกษาภาวะการมีงานทำของประชากรไทย” ซึ่งจะทำให้ทราบว่า ปัจจุบันประชากรไทยมีภาวะการทำงานเป็นอย่างไร และมีปัจจัยใดที่มีอิทธิพลต่อการทำงานของประชากรไทย เพื่อสร้างตัวแบบที่เหมาะสมในการทำนายภาวะการมีงานทำของประชากรไทย รวมทั้งศึกษาปัจจัยที่มีอิทธิพลต่อรายได้ของประชากรไทย และสร้างตัวแบบที่เหมาะสมทำนายรายได้ของประชากรไทย ซึ่งอาจทำให้การแก้ปัญหาคุณภาพชีวิตของคนไทยเป็นไปอย่างถูกทิศทาง อีกทั้งผลการศึกษาในครั้งนี้หน่วยงานที่สนใจสามารถนำไปใช้ประโยชน์ต่อไปได้

1.2 วัตถุประสงค์

1. เพื่อศึกษาภาวะการทำงานของประชากรไทย
2. เพื่อศึกษาปัจจัยที่มีผลต่อภาวะการมีงานทำของประชากรไทย
3. เพื่อหาตัวแบบที่เหมาะสมทำนายภาวะการมีงานทำของประชากรไทย
4. เพื่อหาตัวแบบที่เหมาะสมทำนายรายได้ของประชากรไทย

1.3 ผลที่คาดว่าจะได้รับ

1. ทราบถึงสถานะการปัจจุบันเกี่ยวกับภาวะการทำงานของประชากรไทย
2. ทราบถึงปัจจัยที่มีผลต่อภาวะการมีงานทำของประชากรไทย
3. ทราบถึงปัจจัยที่มีผลต่อรายได้ของประชากรไทย
4. ทำให้ได้ตัวแบบที่เหมาะสมในการทำนายภาวะการมีงานทำของประชากรไทย
5. เพื่อเป็นแนวทางให้แก่หน่วยงานที่เกี่ยวข้องนำไปวางแผนเพื่อพัฒนาและช่วยเหลือในเรื่องการมีงานทำของประชากรไทย

1.4 ขอบเขตงานวิจัย

1.4.1 ขอบเขตข้อมูลที่ศึกษา

การวิจัยในครั้งนี้เป็นการวิจัยเชิงปริมาณ (Quantitative Research) มุ่งศึกษาถึงปัจจัยที่มีผลต่อภาวะการมีงานทำของประชากรไทย โดยใช้ข้อมูลทุติยภูมิจากโครงการสำรวจภาวะการทำงานของประชากรไทยในไตรมาสที่ 4 ปี พ.ศ.2561 ของสำนักงานสถิติแห่งชาติ

1.4.2 ขอบเขตประชากรและตัวอย่าง

ประชากร คือ ประชากรไทย

ตัวอย่าง คือ ตัวอย่างประชากรในการเก็บข้อมูลจากโครงการสำรวจภาวะการทำงานของประชากรไทยในไตรมาสที่ 4 ปี พ.ศ.2561 ของสำนักงานสถิติแห่งชาติ จำนวน 172,241 คน

1.5 นิยามศัพท์เฉพาะ

งาน หมายถึง กิจการที่ทำที่มีลักษณะอย่างหนึ่งอย่างใด ดังต่อไปนี้

1. กิจการที่ทำแล้วได้รับค่าตอบแทนเป็นเงิน หรือสิ่งของ ค่าตอบแทนที่เป็นเงิน อาจจ่ายเป็น รายเดือน รายสัปดาห์/ รายวัน หรือรายชิ้น

2. กิจการที่ทำแล้วได้ผลกำไร หรือส่วนแบ่งเป็นการตอบแทน

3. กิจการที่ทำกับธุรกิจของสมาชิกในครัวเรือน โดยไม่ได้รับค่าจ้างหรือผลกำไรตอบแทนอย่างใด โดยสมาชิกที่ประกอบธุรกิจนั้นจะมีสถานภาพการทำงาน เป็นประกอบธุรกิจส่วนตัวหรือนายจ้าง

ผู้มีงานทำ หมายถึง บุคคลที่มีอายุ 15 ปีขึ้นไป และในระหว่าง 7 วันก่อนวันสัมภาษณ์มีลักษณะอย่างใดอย่างหนึ่ง คือ

1. ได้ทำงานตั้งแต่ 1 ชั่วโมงขึ้นไป โดยได้รับค่าจ้าง เงินเดือน ผลกำไร เงินปันผล ค่าตอบแทนที่เป็นเงินสดหรือสิ่งของ

2. ทำงานอย่างน้อย 1 ชั่วโมง โดยไม่ได้รับค่าจ้างในวิสาหกิจหรือไร่นาของหัวหน้าครัวเรือนหรือสมาชิกในครัวเรือน

3. ไม่ได้ทำงานหรือทำงานน้อยกว่า 1 ชั่วโมง (เป็นผู้ปกติมีงานประจำ) แต่

3.1 ยังคงได้รับค่าตอบแทน ค่าจ้าง ผลประโยชน์อื่นๆ ผลกำไรจากงานหรือธุรกิจในระหว่างที่ไม่ได้ทำงาน

3.2 ไม่ได้รับค่าตอบแทน ค่าจ้างผลประโยชน์อื่นๆ ผลกำไรจากงานหรือธุรกิจในระหว่างที่ไม่ได้ทำงาน แต่ยังมีงานหรือธุรกิจที่จะกลับไปทำ

ผู้ว่างงาน หมายถึง บุคคลที่มีอายุ 15 ปีขึ้นไป และในสัปดาห์แห่งการสำรวจมีลักษณะอย่างหนึ่งอย่างใด ดังต่อไปนี้

1. ไม่ได้ทำงานและไม่มีงานประจำ แต่ได้มีการหางาน สมัครงานหรือรอการบรรจุในระหว่าง 30 วันก่อนวันสัมภาษณ์

2. ไม่ได้ทำงานและไม่มีงานประจำ และไม่ได้หางานทำในระหว่าง 30 วันก่อนวันสัมภาษณ์ แต่พร้อมที่จะทำงานในสัปดาห์แห่งการสำรวจ

บทที่ 2
เอกสารและงานวิจัยที่เกี่ยวข้อง

การดำเนินการวิจัยเรื่อง “ตัวแบบทางสถิติในการศึกษาภาวะการมีงานทำของประชากรไทย” โดยใช้ข้อมูลทุติยภูมิจากโครงการสำรวจภาวะการทำงานของประชากรไทย ในปี พ.ศ. 2561 ของสำนักงานสถิติแห่งชาติ เพื่อให้การศึกษาครั้งนี้บรรลุจุดประสงค์ ผู้วิจัยได้ทบทวนเอกสารและงานวิจัยที่เกี่ยวข้องเพื่อมาสนับสนุนงานวิจัย ดังนี้

2.1 แนวคิดและทฤษฎีเกี่ยวกับการทำงาน

2.2 การวิเคราะห์การถดถอยลอจิสติก

2.3 การวิเคราะห์การถดถอย

2.4 งานวิจัยที่เกี่ยวข้อง

2.5 กรอบแนวคิด

2.1 แนวคิดและทฤษฎีเกี่ยวกับการทำงาน

แนวคิดและคำนิยามของครัวเรือน (โครงการรายได้และการกระจายรายได้ของครัวเรือน, 2545)

1. ครัวเรือนส่วนบุคคล ประกอบด้วย

1.1 ครัวเรือนหนึ่งคน คือ บุคคลคนเดียวซึ่งหุงหาอาหารและจัดหาสิ่งอุปโภคบริโภคที่จำเป็นแก่การครองชีพ โดยไม่เกี่ยวกับผู้ใดซึ่งอาจพำนักอยู่ในเคหสถานเดียวกัน

1.2 ครัวเรือนหลายคน คือ ครัวเรือนที่มีบุคคลตั้งแต่ 2 คนขึ้นไป ร่วมกันจัดหาและใช้สิ่งอุปโภคบริโภคที่จำเป็นแก่การครองชีพร่วมกัน บุคคลที่มาอยู่รวมกันในครัวเรือนอาจจะเป็นญาติหรือไม่เป็นญาติกันก็ได้ ซึ่งได้แก่

1.2.1 ครัวเรือนที่ประกอบไปด้วยบุคคลที่มีความสัมพันธ์ฉันท์ญาติมาอยู่ด้วยกัน หรืออยู่แบบครอบครัว

1.2.2 ครัวเรือนที่ประกอบไปด้วยบุคคลที่ไม่มีความสันพันธ์ฉันท์ญาติไม่เกิน 5 คน มาอาศัยอยู่ด้วยกัน ช่วยกันจ่ายค่าที่พัก ให้นับเป็นครัวเรือนส่วนบุคคล 1 ครัวเรือน

1.2.3 ครัวเรือนคนงานที่นายจ้างจัดให้อยู่อาศัยร่วมกันไม่เกิน 5 คน ถือเป็น 1 ครัวเรือนส่วนบุคคล

1.2.4 ครัวเรือนที่มีบุคคลตั้งแต่ 6 คนขึ้นไปมาอยู่ร่วมกัน ซึ่งบุคคลเหล่านั้นต้องมีบุคคลที่เป็นญาติกันตั้งแต่ 4 คน ขึ้นไป หรือเป็นญาติกันไม่เกิน 3 คน แต่มีคนใดคนหนึ่งรับผิดชอบค่าใช้จ่ายของคนทั้งหมด ถือเป็นครัวเรือนส่วนบุคคล 1 ครัวเรือน

หมายเหตุ สำหรับข้อ 2.2 และ 2.3 มีความแตกต่างจากการสำรวจที่ดำเนินการก่อนปี 2541 ซึ่งได้แยกแต่ละคนเป็น 1 ครัวเรือน

2. สมาชิกของครัวเรือน

สมาชิกของครัวเรือนส่วนบุคคล หมายถึง บุคคลซึ่งอาศัยอยู่เป็นประจำในครัวเรือน รวมทั้งคนที่เคยอาศัยอยู่ในครัวเรือน แต่ได้จากไปที่อื่นชั่วคราว โดยไม่มีวัตถุประสงค์จะไปอยู่ประจำที่อื่น และมีลักษณะอย่างใดอย่างหนึ่งดังนี้

2.1 จากไปที่อื่นชั่วคราวไม่เกิน 3 เดือน เช่น ผู้ป่วยที่เข้ารับการรักษาที่โรงพยาบาล หรือผู้ที่ไปบวช (ไม่นับคนในครัวเรือนที่จากไปเพื่อเรียนหนังสือ หรือประกอบอาชีพ ซึ่งมีที่อยู่ประจำที่อื่น รวมทั้งผู้ต้องโทษและผู้ถูกเกณฑ์ทหาร คนไข้ในโรงพยาบาลโรคจิต)

2.2 จากไปเกินกว่า 3 เดือน แต่ไม่มีที่อยู่อาศัยประจำที่อื่น เช่น พนักงานขาย เป็นต้น

2.3 จากไปเพื่อศึกษา อบรมดูงานต่างประเทศไม่เกิน 6 เดือน

2.4 จากไปทำงานและกินอยู่ที่อื่นชั่วคราวไม่เกิน 30 วันนับจากวันไปถึงวันกำหนดกลับ เช่น คนไปตัดอ้อย ฯลฯ นอกจากนี้สมาชิกของครัวเรือน ยังรวมถึง ผู้ที่มาพักอาศัยอยู่ชั่วคราวเกินกว่า 3 เดือน

3. หัวหน้าครัวเรือน

หัวหน้าครัวเรือน หมายถึง บุคคลผู้ซึ่งเป็นที่ยอมรับนับถือของสมาชิกอื่นในครัวเรือน โดยยกย่องให้เป็นหัวหน้า อาจจะเป็นผู้รับผิดชอบทางด้านการเงินและสวัสดิการของครัวเรือนหรือไม่ก็ตาม

4. ผู้ทำงานให้เกิดรายได้และผู้รับเงินรายได้

ผู้ทำงานให้เกิดรายได้ (earners) คือ สมาชิกของครัวเรือนผู้ปฏิบัติงานในเชิงเศรษฐกิจ ได้แก่ ผู้ดำเนินธุรกิจของตนเอง ผู้ทำงานรับจ้าง และผู้ช่วยธุรกิจของครัวเรือนโดยไม่ได้รับค่าจ้าง ไม่นับรวมผู้รับเงินค่าเช่าเงินช่วยเหลือจากบุคคลอื่นนอกครัวเรือน

ผู้รับเงินรายได้ (income receivers) คือ สมาชิกของครัวเรือน ซึ่งมีรายได้จากแหล่งต่างๆ เช่น กำไรจากการประกอบธุรกิจ ค่าจ้างและเงินเดือน ค่าเช่าหรือเงินช่วยเหลือจากบุคคลนอกครัวเรือน ผู้รับเงินรายได้อาจจะเป็นผู้ปฏิบัติงานในเชิงเศรษฐกิจหรือไม่ก็ได้ นอกจากนั้นจะไม่นับรวมผู้ช่วยธุรกิจของครัวเรือนโดยไม่ได้รับค่าจ้าง เป็นผู้รับเงินรายได้

5. รายได้ของครัวเรือน

รายได้ทั้งหมดของครัวเรือน ประกอบด้วย

5.1 ค่าจ้างและเงินเดือน เงินรางวัลบริกร เงินโบนัส เป็นต้น

5.2 กำไรสุทธิจากการประกอบธุรกิจการเกษตรและธุรกิจอื่น ๆ

5.3 รายได้จากทรัพย์สิน เช่น ค่าเช่าที่ดิน ค่าลิขสิทธิ์ ดอกเบี้ย และเงินปันผล

5.4 เงินได้รับเป็นการช่วยเหลือ บำเหน็จ บำนาญ เงินทุนการศึกษา

5.5 รายได้ที่ไม่เป็นตัวเงิน ได้แก่ มูลค่าของสินค้าและบริการที่ได้รับเป็นส่วนหนึ่งของค่าจ้าง เงินเดือน มูลค่าของสินค้าหรืออาหารที่ครัวเรือนผลิตและบริโภคเอง (รวมค่าประเมินค่าเช่าบ้านที่ครัวเรือนเป็นเจ้าของ) หรือได้รับมาโดยไม่ต้องซื้อ

5.6 รายรับที่เป็นตัวเงินอื่น ๆ เช่น เงินได้รับจากการประกันภัยหรือประกันชีวิต เงินรางวัลสลากกินแบ่ง และรายรับอื่น ๆ ในประเภทเดียวกัน

รายได้ประจำ ได้แก่ รายได้ทั้งหมดของครัวเรือน แต่ไม่รวมรายรับที่เป็นตัวเงินอื่น ๆ ในข้อ 5.6

6. ค่าใช้จ่ายของครัวเรือน ประกอบด้วย

6.1 จำนวนเงินที่ครัวเรือนได้ใช้จ่ายเพื่อซื้อสินค้าและบริการต่าง ๆ เพื่อใช้ในการดำรงชีพ

6.2 มูลค่าของสินค้าและบริการที่ได้รับเป็นส่วนหนึ่งของค่าจ้าง เงินเดือน สินค้าหรืออาหารที่ครัวเรือนผลิตและบริโภคเอง (รวมค่าประเมินค่าเช่าบ้านที่ครัวเรือนเป็นเจ้าของและอยู่อาศัยเอง) หรือได้มาโดยไม่ต้องซื้อ

6.3 รายจ่ายอื่น ๆ เช่น ค่าภาษี เงินบริจาค ค่าเบี้ยประกันภัย ค่าสลากกินแบ่ง ดอกเบี้ยจ่าย และ รายจ่ายที่มิใช่เพื่อการบริโภคอื่น ๆ

ค่าใช้จ่ายเพื่อการอุปโภคบริโภค ได้แก่ รายจ่ายทั้งสิ้นของครัวเรือน แต่ไม่รวมรายจ่ายอื่น ๆ ตามข้อ 6.3

ค่าใช้จ่ายของครัวเรือน ไม่รวมถึง ค่าผ่อนชำระบ้าน ที่ดิน และรายจ่ายเพื่อการสะสมทุนอื่น ๆ ตลอดจนรายจ่ายที่จัดว่าเป็นการออม เช่น เบี้ยประกันชีวิตชนิดสะสมทรัพย์ เงินสะสมกองทุนสำรองเลี้ยงชีพ/กบข. เป็นต้น

7. สถานะทางเศรษฐกิจสังคมของครัวเรือน

ในการวิเคราะห์ข้อมูล ได้จัดจำแนกครัวเรือนเป็นกลุ่มตามสถานะทางเศรษฐกิจและสังคม โดยพิจารณาจากแหล่งรายได้ส่วนใหญ่ของครัวเรือน สถานภาพการทำงาน ประเภทของกิจกรรมในเชิงเศรษฐกิจและอาชีพเป็นหลัก โดยได้นำเสนอเป็น 4 กลุ่ม ดังนี้

7.1 ผู้ถือครองทำการเกษตร (รวมการทำประมงทะเล)

7.2 ผู้ดำเนินธุรกิจของตนเอง

7.3 ลูกจ้าง แบ่งออกเป็น 2 กลุ่มย่อย คือ

7.3.1 ผู้ปฏิบัติงานวิชาชีพ นักวิชาการ นักบริหาร เสมียนพนักงาน และพนักงานผู้ให้บริการ

7.3.2 ผู้ปฏิบัติงานในการผลิต และก่อสร้าง คนงานเกษตร คนงานทั่วไป

7.4 ผู้ไม่ได้ปฏิบัติงานในเชิงเศรษฐกิจ หมายถึง บุคคลต่อไปนี้

ก) แม่บ้านหรือผู้ที่ทำงานบ้าน

ข) นักเรียน

ค) ผู้เกษียณอายุ และคนชรา

ง) ผู้ไม่สามารถทำงานได้ เนื่องจากพิการทางร่างกายหรือจิตใจ หรือเนื่องจากเจ็บป่วยเรื้อรัง

จ) คนขอทาน

8. กลุ่มควินไทล์ของครัวเรือนเรียงลำดับตามรายได้ประจำต่อคนต่อเดือน

หมายถึง การจัดจำแนกครัวเรือนออกเป็น 5 กลุ่ม เท่า ๆ กัน (แต่ละกลุ่มมีจำนวนเท่ากับ ร้อยละ 20 ของจำนวนครัวเรือนทั้งหมด) โดยจัดเรียงครัวเรือนตามรายได้ประจำต่อคนต่อเดือนจากน้อยไปหามาก ทั้งนี้กลุ่มควินไทล์ที่ 1 คือกลุ่มที่มีรายได้น้อยที่สุด และควินไทล์ที่ 5 คือกลุ่มที่มีรายได้มากที่สุด

9. หนี้สินของครัวเรือน

หมายถึง มูลค่าหนี้จากการกู้ยืม การเช่าซื้อทรัพย์สินต่าง ๆ เช่น บ้าน ที่ดิน รถ ตลอดจน สิ่งของเครื่องใช้ในครัวเรือน รวมทั้งการซื้อเชื่อสินค้าอุปโภคบริโภคที่ยังไม่ได้ชำระเงินของสมาชิกทุกคนในครัวเรือน ณ วันสุดท้ายของเดือนที่แล้วก่อนการสัมภาษณ์

2.2 การวิเคราะห์การถดถอยลอจิสติก (Logistic regression analysis)

การวิเคราะห์การถดถอยลอจิสติก (Logistic regression analysis) เป็นเทคนิคการวิเคราะห์สถิติเชิงคุณภาพ (Qualitative statistical techniques) การวิเคราะห์การถดถอยลอจิสติกแบ่งเป็น 2 ประเภท คือ (1) การวิเคราะห์การถดถอยลอจิสติกทวิภาค (Binary logistic regression analysis) และ (2) การวิเคราะห์การถดถอยลอจิสติกพหุ (Multinomial logistic regression analysis) การวิเคราะห์การถดถอยลอจิสติกทั้ง 2 ประเภท แตกต่างกันในด้านตัวแปรตาม โดยที่การวิเคราะห์การถดถอยลอจิสติกทวิภาคใช้กับตัวแปรตามที่แบ่งออกเป็น 2 กลุ่มย่อย (Dichotomous variable) มีค่าที่เป็นไปได้ 2 ค่า คือ 0 กับ 1 เช่น กลุ่มที่มีเหตุการณ์กับกลุ่มที่ไม่มีเหตุการณ์ ส่วนการวิเคราะห์ลอจิสติกแบบพหุใช้กับตัวแปรตามที่มีหลายค่ามากกว่า 2 กลุ่ม (Polytomous variable) เช่น โรงพยาบาลมีมาตรฐานการให้บริการสูง ปานกลาง และต่ำ การวิเคราะห์ลอจิสติกมีเป้าหมายก็คือ เพื่อทำนายโอกาสที่จะเกิดเหตุการณ์ที่สนใจโดยอาศัยสมการลอจิสติกที่สร้างขึ้นจากชุดตัวแปรทำนายที่มีข้อมูลเป็นตัวแปรที่มีข้อมูลอยู่ในมาตราช่วง (Interval scale) เป็นอย่างน้อย (กัลยา วานิชยบัญชา, 2549)

2.2.1 วัตถุประสงค์ของการวิเคราะห์การถดถอยลอจิสติก

เพื่อทำนายว่าจะเกิดเหตุการณ์หนึ่งขึ้นหรือไม่หรือมีโอกาสเกิดขึ้นมากน้อยเพียงใด โดยมีการกำหนดค่าตัวแปรตัวหนึ่งหรือหลายตัวที่คาดว่าจะส่งผลต่อการเกิดเหตุการณ์นั้นๆ และในที่สุดก็จะทำให้เราเข้าใจสาเหตุการเกิดเหตุการณ์นั้น ๆ ได้

2.2.2 ตัวแบบการวิเคราะห์การถดถอยลอจิสติก

การประมาณค่าความน่าจะเป็นของการเกิดเหตุการณ์โดยมีตัวแปรอิสระเพียงตัวเดียว ฟังก์ชันลอจิสติกจะแสดงความน่าจะเป็นของเหตุการณ์ดังต่อไปนี้

$$\text{prob(event)} = \frac{1}{1+e^{-(\beta_0+\beta_1 x)}}$$

หรือ

$$\text{prob(event)} = \frac{e^{(\beta_0+\beta_1 x)}}{1+e^{(\beta_0+\beta_1 x)}}$$

โดยที่ β_0 คือ ค่าคงที่ (เมื่อไม่มีอิทธิพลจากตัวแปรอิสระใด)

β_1 คือ ค่าสัมประสิทธิ์ของตัวแปรอิสระ (ประมาณได้จากข้อมูลสังเกต)

x คือ ตัวแปรอิสระ (ตัวแปรทำนาย)

e คือ ลอการิทึมธรรมชาติ (มีค่าประมาณ 2.71828)

ในกรณีที่มีตัวแปรทำนายหลายตัว (n) ฟังก์ชันลอจิสติกจะแสดงในรูปดังนี้

$$\text{prob(event)} = \frac{1}{1+e^{-z}}$$

หรือ

$$\text{prob(event)} = \frac{e^{z}}{1+e^{z}}$$

เมื่อ z เป็น Linear combination ที่อยู่ในรูป

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p$$

ดังนั้น ความน่าจะเป็นของการไม่เกิดเหตุการณ์ ก็คือ

$$\text{prob(no.event)} = 1 - \text{prob(event)}$$

2.2.3 การตรวจสอบตัวแบบการถดถอย (นราทิพย์ จั่นสกุล, 2561)

การตรวจสอบตัวแบบการถดถอยเป็นเทคนิคหนึ่งที่ใช้สำหรับการสืบหา (detecting) หรือตรวจสอบ (checking) ปัญหาที่เกิดขึ้นในการวิเคราะห์การถดถอย เนื่องจากเราไม่สามารถมั่นใจว่า ตัวแบบการถดถอยที่ได้นั้นเหมาะสมกับข้อมูลหรือไม่ ดังนั้นจึงมีความจำเป็นที่ต้องมีการตรวจสอบตัวแบบ ซึ่งประกอบไปด้วย

1. ตรวจสอบข้อสมมติของตัวแบบ (Model Assumption)

เป็นข้อสมมติเกี่ยวกับค่าคลาดเคลื่อนสุ่ม (random error) และการตรวจสอบค่านอกเกณฑ์ (outliers) มี 4 ข้อดังนี้

1. ค่าคลาดเคลื่อนสุ่มจะมีค่าเฉลี่ย $\mathrm{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$
2. ค่าคลาดเคลื่อนสุ่มต้องมีการแจกแจงปรกติ
3. ค่าคลาดเคลื่อนสุ่มจะมีความแปรปรวนคงที่
4. ค่าคลาดเคลื่อนสุ่ม $\boldsymbol{\varepsilon}$ จะต้องเป็นอิสระต่อกัน

2. ตรวจสอบโดยใช้แผนภาพ Half-Normal Plot (จิรภัทร ภู่ขวัญทอง, 2550)

การวินิจฉัยตัวแบบโดยใช้ Half-Normal Plot with simulated envelope สำหรับ Normal regression model ซึ่งเสนอโดย Atkinson (1985) เป็นการตรวจสอบการแจกแจงของส่วนเหลือ (Residual) โดยการนำค่าสัมบูรณ์ของ Pearson residual จากตัวแบบที่ได้ มาวาดกราฟกับ Half-Normal score $\phi^{-1}\left(\frac{\mathrm{i}+\mathrm{n}}{\mathrm{n}}\right)$ โดยที่ ϕ คือ ฟังก์ชันการแจกแจงสะสมของการแจกแจงปรกติมาตรฐาน ต่อมา Demetrio and Hinde (1997) ได้ปรับเป็น $\phi^{-1}\left(\frac{\mathrm{i}+\mathrm{n}-0.125}{2\mathrm{n}+0.5}\right)$ ซึ่งให้ค่าที่ใกล้เคียงกับค่าคาดหมายของสถิติอันดับมากขึ้น ถ้าค่าของ Pearson residual ทุกค่าอยู่ใน envelope แสดงว่า Pearson residual นั้นมีการแจกแจงปรกติ Williams (1987) ได้เสนอการตรวจสอบความถูกต้องของตัวแบบเชิงสถิติโดยใช้ Half-Normal Plot with simulated envelope โดยใช้ค่าสัมบูรณ์ของ Standardized Deviance Residual ของตัวแบบที่ถูกเรียงลำดับ (จากน้อยไปมาก) เขายังกล่าวว่า Half-Normal Plot with simulated envelope สามารถใช้ตรวจสอบได้ทั้งข้อสมมติเบื้องต้นของตัวแบบและ Underdispersion หรือ Overdispersion ได้ด้วย การตรวจสอบความถูกต้องของตัวแบบโดย Half-Normal Plot with simulated envelope มีขั้นตอนการสร้างดังนี้

1. สร้างตัวแบบและคำนวณค่าสัมบูรณ์ของ Standardized Deviance residual แทนด้วย $\mathrm{r}_{\mathrm{Di}}^{\bullet}$ เรียงลำดับจากน้อยไปมาก สมมติแทนด้วย $\mathrm{d}_{(\mathrm{i})}$

2. จำลองตัวแปรตาม 19 ชุด (แนะนำให้จำลองตัวแปรตาม 19 ชุด หรือประมาณ 5% ของโอกาสที่จะทำให้ค่า $\left|\mathrm{d}_{(\mathrm{i})}\right|$ ที่มากที่สุดจากชุดข้อมูลเดิม ยังอยู่ภายใต้เส้นประของค่าเฉลี่ยของ Standardized Deviance residual) ภายใต้ข้อสมมติว่าตัวแบบในข้อ 1. เหมาะสม

3. สร้างตัวแบบโดยใช้ตัวแปรตามจากข้อ 2. ใช้ตัวแปรอิสระชุดเดียวกับตัวแบบที่สร้างในข้อ 1. คำนวณค่าสัมบูรณ์ของ $r_{Di}^{\bullet}$ เรียงลำดับจากน้อยไปมากแล้วแทนด้วย $d_{k(i)}^{*}$ โดยที่ $k = 1, 2, ..., 19$; $i = 1, 2, ..., n$

4. คำนวณค่าต่ำสุด ค่าสูงสุด และค่าเฉลี่ยของ $d_{k(i)}^{*}$

5. พล็อตค่า $d_{(i)}$ ค่าต่ำสุด ค่าสูงสุด และค่าเฉลี่ยของ $d_{k(i)}^{*}$ กับค่า Half-Normal score

การตรวจสอบความถูกต้องของตัวแบบโดย Half-Normal Plot with simulated envelope พิจารณากราฟ Half-Normal Plot ของตัวแบบ ถ้าค่า Standardized Deviance residual ที่ได้จากการสร้างตัวแบบทุกค่า อยู่ภายใต้เส้นประของค่าเฉลี่ยของ Standardized Deviance residual ที่ได้จากการจำลองข้อมูล จะได้ว่าตัวแบบนั้นเป็นตัวแบบที่เหมาะสมกับข้อมูลมากที่สุด

2.2.4 การเลือกตัวแบบที่เหมาะสม (นราทิพย์ จั่นสกุล, 2561)

ในการพิจารณาตัวแบบอาจจะมีตัวแบบหลายตัวแบบที่เหมาะสมกับข้อมูลและมีประสิทธิภาพในการทำนายหรือประมาณค่าใกล้เคียงกัน ผู้ทำการวิเคราะห์จึงต้องพิจารณาว่าตัวแบบใดเหมาะสมกับข้อมูลมากที่สุด ซึ่งสามารถพิจารณาได้จากค่าสถิติต่อไปนี้

1. การเปรียบเทียบค่า Deviance Residual และ Degrees of Freedom

การเปรียบเทียบอัตราส่วนระหว่าง Deviance Residual และ Degrees of Freedom เป็นเกณฑ์ในการพิจารณาว่าเกิด Underdispersion หรือ Overdispersion ใน Binomial counts หรือไม่

- ถ้า Deviance Residual และ Degrees of Freedom มีค่าใกล้เคียงกันมาก แสดงว่าตัวแบบนั้นเหมาะสมกับข้อมูล
- ถ้า Deviance Residual มีค่ามากกว่าหรือเท่ากับ 2 เท่าของ Degrees of Freedom ถือว่าตัวแบบนั้นไม่เหมาะสมกับข้อมูล

2. สถิติ AIC (นราทิพย์ และทัศนา, 2560)

สถิติ AIC ย่อมาจาก Akaike Information Citerion เป็นสถิติที่เสนอโดย Akaike,H. ในปี 1974 ซึ่งแนวคิดของการสร้างสถิติ AIC ขึ้น คือ การปรับค่าประมาณภาวะน่าจะเป็นสูงสุดด้วยจำนวนพารามิเตอร์ในตัวแบบที่ต้องการประมาณ คือ

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip} + \varepsilon_i$$

โดยที่ $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \boldsymbol{\sigma}^2)$ ดังนั้น ฟังก์ชันภาวะน่าจะเป็นสำหรับ $y_1, y_2, ..., y_n$ คือ

$$L=\prod_{i=1}^{n}\frac{1}{\left(2\pi\sigma^2\right)^{1/2}}\exp\left(-\frac{\varepsilon_i^2}{2\sigma^2}\right)$$
$$=\prod_{i=1}^{n}\frac{1}{\left(2\pi\sigma^2\right)^{1/2}}\exp\left(-\frac{1}{2\sigma^2}\left[y_i-\beta_0-\beta_1x_{i1}-\beta_2x_{i2}-\ldots-\beta_px_{ip}\right]^2\right)$$
$$=\frac{1}{\left(2\pi\sigma^2\right)^{n/2}}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left[y_i-\beta_0-\beta_1x_{i1}-\beta_2x_{i2}-\ldots-\beta_px_{ip}\right]^2\right)$$

และลอการิทึมฐานธรรมชาติของฟังก์ชันภาวะน่าจะเป็น คือ

$$LogL=-\frac{n}{2}\log\left(2\pi\right)-\frac{n}{2}\log\left(\sigma^2\right)-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left[y_i-\beta_0-\beta_1x_{i1}-\beta_2x_{i2}-\ldots-\beta_px_{ip}\right]^2$$

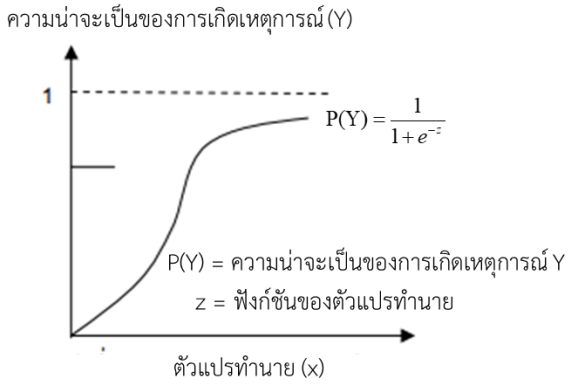
แล้วสถิติ

$$AIC=-2\times\widehat{\log L}+2(\text{จำนวนพารามิเตอร์ในตัวแบบ})$$

เมื่อ $\widehat{\log L}$ คือ LogL ที่แทนค่า σ^2 ด้วย MSE และ β_j ด้วย $\hat{\beta}_j$
ดังนั้น จำนวนพารามิเตอร์ใน LogL จึงเท่ากับ $p+2$
โดยตัวแบบที่ให้ค่า AIC น้อยที่สุดจะเป็นตัวแบบที่ถูกเลือก

2.2.5 อัตราส่วนออดส์ (Odds Ratio)

Odds คือ อัตราส่วนความน่าจะเป็นของการเกิดเหตุการณ์ที่สนใจ $P(Y=1)$ กับความน่าจะเป็นที่จะไม่เกิดเหตุการณ์ที่สนใจ $P(Y=0)$



รูป 2.1 กราฟฟังก์ชันการถดถอยลอจิสติก

สมมติให้ $P(Y=1)$ = ความน่าจะเป็นของการเกิดเหตุการณ์ที่สนใจ

$P(Y=0)$ = ความน่าจะเป็นของการไม่เกิดเหตุการณ์ที่สนใจ

Z = Linear combination ของตัวแปรทำนายหรือ $(\beta_0+\beta_1x_1+\beta_2x_2+...+\beta_px_p)$

ดังนั้น
$$P(Y=1)=\frac{1}{1+e^{-z}}=\frac{e^z}{1+e^z} \quad \textbf{(2.1)}$$

$$P(Y=0)=1-P(Y=1)=1-\frac{1}{1+e^{-z}}=\frac{1}{1+e^z} \quad \textbf{(2.2)}$$

จากสมการที่ (2.1) และ (2.2) สามารถหาค่า Odds ratio ของเหตุการณ์ที่สนใจ ได้ดังนี้

$$\frac{P(Y=1)}{1-P(Y=1)}=e^z \quad \textbf{(2.3)}$$

สมการที่ (2.3) แปลงให้อยู่ในรูปของสมการถดถอยเชิงเส้น (linear regression model) โดยการใส่ Natural logarithm ทั้ง 2 ข้างของสมการ จะได้ตัวแบบในรูปของ log(odds) ดังนี้

$$\log\left(\frac{P(Y=1)}{1-P(Y=1)}\right)=\log\left(e^z\right)=Z=\beta_0+\beta_1x_1+\beta_2x_2+...+\beta_px_p$$

2.2.6 การแปลผลด้วยค่า Odds ratio

Odds ratio (OR) คือ อัตราการเปลี่ยนแปลงของความน่าจะเป็นหรือโอกาสที่จะเกิดเหตุการณ์ที่สนใจเมื่อตัวแปรอิสระเปลี่ยนไป 1 หน่วย

- ถ้าค่า Odds ratio = 1 แสดงว่า การเปลี่ยนแปลงของ x ไม่มีผลต่อ Y
- ถ้าค่า Odds ratio > 1 แสดงว่า เมื่อ x เปลี่ยนไป 1 หน่วย ทำให้โอกาสการเกิดเหตุการณ์เพิ่มขึ้น(หน่วยเป็นเท่าตัว) เมื่อเทียบกับค่าเดิมของ x
- ถ้าค่า Odds ratio < 1 แสดงว่า เมื่อ x เปลี่ยนไป 1 หน่วย ทำให้โอกาสการเกิดเหตุการณ์ลดลง(หน่วยเป็นเท่าตัว) เมื่อเทียบกับค่าเดิมของ x
- กรณีค่า Odds ratio < 1 นิยมอ่านผลเป็นร้อยละการเปลี่ยนแปลงของค่า Odd คำนวณจาก (1 - Odds ratio)*100

2.3 การวิเคราะห์การถดถอย (Regression Analysis)

การวิเคราะห์การถดถอย (Regression Analysis) เป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรตั้งแต่สองตัวขึ้นไป โดยพิจารณาถึงรูปแบบที่เป็นไปได้ของความสัมพันธ์ และเป้าหมายของการวิเคราะห์ คือ หาสมการ ตัวแบบ หรือฟังก์ชันทางคณิตศาสตร์ที่จะใช้เป็นตัวแทนของความสัมพันธ์ระหว่างตัวแปรตั้งแต่สองตัวขึ้นไปนั้น ทั้งนี้เพื่อที่จะใช้ประมาณค่าหรือทำนายค่าของตัวแปรหนึ่งจากค่าที่กำหนดให้ของตัวแปรอื่นๆ ที่เกี่ยวข้องกัน เรียกตัวแปรที่ต้องการประมาณหรือทำนายนั้นว่าตัวแปรตาม (Dependent Variable) และเรียกตัวแปรอื่นๆ ที่ส่งผลต่อตัวแปรตามว่าตัวแปรอิสระ (Independent Variable)

การวิเคราะห์การถดถอยเหมาะกับตัวแปรตามที่มีระดับการวัดมาตราอันตรภาค (Interval Scale) หรือมาตราอัตราส่วน (Ratio Scale) แต่ตัวแปรอิสระนั้น อาจมีระดับการวัดได้ทุกมาตรา แต่ถ้ามีระดับการวัดมาตรานามบัญญัติ (Nominal Scale) หรือมาตราอันดับ (Ordinal Scale) จะต้องทำการแปลงตัวแปรอิสระนั้น ให้เป็นตัวแปรหุ่น (Dummy Variable) ซึ่งมีค่าเป็น 0 หรือ 1 เท่านั้น (นราทิพย์ และทัศนา, 2560)

2.3.1 ประเภทของการวิเคราะห์การถดถอย

การวิเคราะห์การถดถอยแบ่งออกได้ 2 ประเภท ดังนี้

การถดถอยเชิงเส้นเชิงเดียว (Simple Linear Regression) เป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรสองตัว ซึ่งจะประกอบด้วยตัวแปรตาม Y จำนวน 1 ตัว และมีตัวแปรอิสระ x จำนวน 1 ตัว การพิจารณาลักษณะแนวโน้มของความสัมพันธ์ระหว่างตัวแปรทั้งสองนี้สามารถพิจารณาได้จากแผนภาพการกระจาย ว่ามีแนวโน้มเป็นเส้นตรงหรือเส้นโค้ง ถ้าแนวโน้มของความสัมพันธ์มีลักษณะเป็นเส้นตรง สมการแสดงความสัมพันธ์ระหว่างตัวแปร x และตัวแปร Y จะอยู่ในรูปของสมการเส้นตรง และเรียกการถดถอยนี้ว่า การถดถอยเส้นตรง (Straight Line Regression) หรือการถดกอยเชิงเส้นเชิงเดียว (Simple Linear Regression) แต่ถ้าแนวโน้มของความสัมพันธ์ มีลักษณะเชิงเส้นโค้ง (Curvilinear) ก็จำเป็นต้องหาสมการเส้นโค้งให้กับความสัมพันธ์นั้น ถ้าแนวโน้มของความสัมพันธ์ระหว่างตัวแปรสองตัวมีลักษณะเป็นเส้นตรง และประกอบด้วยตัวแปรอิสระ x เพียงตัวเดียว สามารถเขียนตัวแบบการถดถอยเชิงเส้นเชิงเดียว ได้ดังนี้

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

และตัวแบบการถดถอยเชิงเส้นเชิงเดียวของประชากรขนาด N คือ

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i ; i = 1, 2, \ldots, N$$

โดยที่ Y_i คือ ตัวแปรสุ่มที่สังเกตค่า (ข้อมูล) ได้ (Observable Random Variable)

β_0 และ β_1 คือ พารามิเตอร์ของตัวแบบ หรือสัมประสิทธิ์การถดถอย (Regression Coefficients)

x_i คือ ตัวแปรไม่สุ่ม (Non-Random Variable) หรือ ตัวแปรตรึง (Fixed Variable)

ε_i คือ พจน์ค่าคลาดเคลื่อนสุ่ม (Random Error Term) ที่มีค่าเฉลี่ย $E(\varepsilon_i)=0$ และ ความแปรปรวน $Var(\varepsilon_i)=\sigma^2$,ε_i และ ε_j เป็นอิสระซึ่งกันและกัน นั่นคือ ความแปรปรวนร่วม $cov(\varepsilon_i,\varepsilon_j)=0$ สำหรับทุกๆค่าของ $i, j; i \neq j$

การวิเคราะห์การถดถอยเชิงเส้นพหุคูณ (Multiple Linear Regression Analysis) เป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรตาม Y จำนวน 1 ตัว กับตัวแปรอิสระมากกว่าหนึ่งตัว ถ้าความสัมพันธ์ระหว่างตัวแปรตามและตัวแปรอิสระดังกล่าวมีลักษณะเชิงเส้นในพารามิเตอร์แล้ว เราจะเรียกการวิเคราะห์การถดถอยพหุคูณนั้นว่า การถดถอยเชิงเส้นพหุคูณ (Multiple Linear Regression)

<u>กำหนดให้</u> ตัวแบบการถดถอยเกี่ยวข้องกับตัวแปรสุ่ม Y ซึ่งเป็นตัวแปรตาม และตัวแปรอิสระ p ตัว ได้แก่ $x_1, x_2, \ldots, x_p$ ทำนองเดียวกันกับการถดถอยเชิงเส้นเชิงเดียว ที่มี Y_i คือ ตัวแปรสุ่มตัวที่ i และ $x_{i1}, x_{i2}, \ldots, x_{ip}$ เป็นชุดของตัวแปรอิสระที่สอดคล้องกับ Y_i ดังนั้น ตัวแบบการถดถอยเชิงเส้นพหุคูณ คือ

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + \varepsilon_i; i = 1, 2, \ldots, N$$

โดยที่ Y_i คือ ตัวแปรสุ่ม

$\beta_0, \beta_1, \beta_2, \ldots, \beta_p$ คือ สัมประสิทธิ์การถดถอยซึ่งเป็นพารามิเตอร์ของตัวแบบ

$x_{i1}, x_{i2}, \ldots, x_{ip}$ คือ ตัวแปรที่เป็นอิสระต่อกัน และมีค่าคงตัว

ε_i คือ พจน์ค่าคลาดเคลื่อนสุ่ม

2.3.2 ความหมายของพารามิเตอร์

สัมประสิทธิ์การถดถอยสำหรับตัวแบบการถดถอยเชิงเส้นเชิงเดียว เรียกพารามิเตอร์ β_0 และ β_1 ในตัวแบบการถดถอยว่า สัมประสิทธิ์การถดถอย (Regression Coefficients) โดยที่ β_0 คือ ส่วนตัดแกน (Intercept) หมายถึง ค่าเฉลี่ยของการแจกแจงความน่าจะเป็นของ Y_i

ที่ $x_i = 0$ สำหรับ β_1 คือ ความชัน (Slope) ของเส้นการถดถอย จะหมายถึง การเปลี่ยนแปลงในค่าเฉลี่ยของ Y_i เมื่อ x_i เพิ่มขึ้นหนึ่งหน่วย

- ถ้า β_1 มีค่าเท่ากับ 0 แสดงว่า ตัวแปรตามและตัวแปรอิสระไม่มีความสัมพันธ์เชิงเส้นต่อกัน
- ถ้า β_1 มีค่ามากกว่า 0 แสดงว่าตัวแปรตามและตัวแปรอิสระมีความสัมพันธ์เชิงเส้นต่อกัน และเป็นความสัมพันธ์ทางบวก
- ถ้า β_1 มีค่าน้อยกว่า 0 แสดงว่าตัวแปรตามและตัวแปรอิสระมีความสัมพันธ์เชิงเส้นต่อกัน และเป็นความสัมพันธ์ทางลบ

สัมประสิทธิ์การถดถอยสำหรับตัวแบบการถดถอยพหุคูณ β_j , $j = 1, 2, \ldots, p$ เรียกว่า สัมประสิทธิ์การถดถอยบางส่วน (partial regression coefficient) หมายถึง การเปลี่ยนแปลงค่าเฉลี่ยของตัวแปรตาม Y เมื่อตัวแปรอิสระ x_j เพิ่มขึ้น 1 หน่วย โดยที่กำหนดให้ตัวแปรอิสระอื่นๆ มีค่าคงที่

2.3.3 การแปลงข้อมูล (Data Transformation)

เป็นการดำเนินการแปลงข้อมูลของตัวแปรตาม ตัวแปรอิสระ หรือของทั้งสองตัวแปร เพื่อให้ได้รูปแบบการถดถอยที่เป็นไปตามข้อสมมติพื้นฐานและง่ายต่อการประมาณค่าพารามิเตอร์ เมื่อตรวจพบว่ารูปแบบการถดถอย $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ หรือ $E(Y_i) = \beta_0 + \beta_1 x_i$ ไม่เหมาะสมกับข้อมูล วิธีการแก้ไขที่ทำได้ในเบื้องต้น มี 2 วิธี ได้แก่

1. ไม่ใช้รูปแบบการถดถอย $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ และหารูปแบบการถดถอยที่เหมาะสมกว่า
2. ทำการแปลงข้อมูลเพื่อให้รูปแบบ $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ เหมาะสมกับข้อมูล

การแก้ไขวิธีแรกจะได้รูปแบบที่ซับซ้อนและมองเห็นภาพของความสัมพันธ์ได้ แต่การประมาณค่าพารามิเตอร์จะยุ่งยาก ส่วนวิธีการแปลงข้อมูลจะทำให้การประมาณค่าพารามิเตอร์ง่ายกว่า ซึ่งในงานวิจัยเล่มนี้จะใช้วิธีที่ 2 แบ่งออกเป็น 2 แบบดังนี้

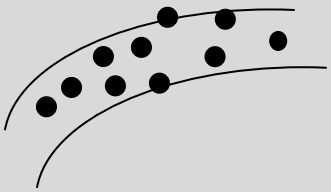
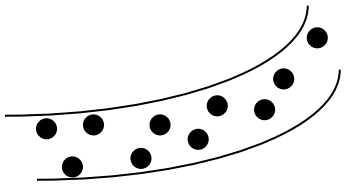
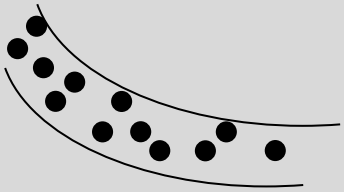
การแปลงข้อมูลสำหรับความสัมพันธ์ไม่เป็นเส้นตรงแต่ค่าคลาดเคลื่อนสุ่มมีความแปรปรวนคงที่

ในกรณีที่ค่าคลาดเคลื่อนมีการแจกแจงปรกติและมีความแปรปรวนคงที่ แต่ความสัมพันธ์ระหว่าง x และ y ไม่เป็นเส้นตรง การแปลงข้อมูลจะดำเนินการที่ตัวแปร x เหตุผลที่ไม่แปลงตัวแปร y เช่น $y' = \ln(y)$ หรือ $y' = \sqrt{y}$ เพราะอาจทำให้ค่าคลาดเคลื่อนมี

คุณสมบัติไม่เป็นไปตามข้อสมมติเบื้องต้น นั่นคือ อาจมีการแจกแจงเป็นแบบอื่นที่ไม่ใช่การแจกแจงปรกติ หรืออาจมีความแปรปรวนไม่คงที่

การตรวจสอบว่าจำเป็นต้องแปลงข้อมูลตัวแปรหรือไม่ ดำเนินการโดยวาดแผนภาพการกระจายระหว่าง y และ x ถ้าลักษณะของความสัมพันธ์หรือการถดถอยเป็นเส้นโค้ง ให้พิจารณาว่า เป็นเส้นโค้งรูปแบบใด แล้วเลือกวิธีการแปลงข้อมูลที่เหมาะสมกับรูปแบบนั้น ดังแสดงในตาราง 2.1 ซึ่งแสดงรูปแบบการถดถอย และการแปลงข้อมูล x เป็น x' อย่างง่าย การแปลงข้อมูลลักษณะนี้อาจทำให้ความสัมพันธ์ระหว่าง y และ x' เป็นเส้นตรงได้โดยไม่ส่งผลกระทบต่อการแจกแจงของ Y เมื่อแปลงข้อมูลแต่ละครั้ง ให้สร้างตัวแบบการถดถอยจากข้อมูลที่แปลงแล้ว และตรวจสอบด้วยการวาดกราฟของส่วนเหลือ เพื่อดูว่าตัวแบบถดถอยที่ได้จากการแปลงข้อมูลเป็นไปตามที่ต้องการหรือไม่

ตารางที่ 2.1 รูปแบบการถดถอยที่ไม่เป็นเส้นตรง (แต่มีความแปรปรวนของความคลาดเคลื่อนคงที่) และการแปลงข้อมูล

รูปแบบความสัมพันธ์ระหว่าง y และ x	**การแปลงข้อมูลของ x**
(ก) 	$x' = \ln(x),\ x' = \sqrt{x}$
(ข) 	$x' = x^2, x' = \exp(x)$
(ค) 	$x' = \frac{1}{x}, x' = \exp(-x)$

การแปลงข้อมูลสำหรับความสัมพันธ์ไม่เป็นเส้นตรงและค่าคลาดเคลื่อนมีความแปรปรวนไม่คงที่

ในกรณีที่ความแปรปรวนของค่าคลาดเคลื่อนสุ่มไม่คงที่หรือการแจกแจงไม่เป็นแบบปรกติ หรือเกิดขึ้นทั้งสองกรณีการแก้ไขทำได้โดยการแปลงข้อมูลของตัวแปรตาม Y เนื่องจากข้อสมมติทั้งสองจะเกี่ยวข้องกับตัวแปรตาม Y การแปลงข้อมูลของ Y นอกจากจะทำให้ค่า

คลาดเคลื่อนมีความแปรปรวนคงที่แล้ว ยังอาจช่วยทำให้รูปแบบความสัมพันธ์จากเดิมเป็นเส้นโค้งเปลี่ยนมาเป็นเส้นตรงได้ บางครั้งอาจจำเป็นต้องแปลงข้อมูลของ Y และ x ไปพร้อม ๆ กัน

การแปลงข้อมูลไม่มีลักษณะตายตัว ขึ้นอยู่กับประสบการณ์ของผู้วิจัยบางครั้งอาจต้องลองผิดลองถูก และตรวจสอบจนกระทั่งได้ตัวแบบการถดถอยที่เหมาะสม วิธีการแปลงข้อมูลวิธีหนึ่งที่เป็นที่นิยมในการวิเคราะห์การถดถอยเชิงเส้นก็คือ การแปลงบอกซ์-คอกซ์ (Box-Cox transformation) ซึ่งเสนอโดย Box และ Cox ในปี 1964 วิธีนี้ช่วยในการแก้ปัญหาที่ค่าคลาดเคลื่อนสุ่มมีการแจกแจงลักษณะเบ้ (skewed) ความแปรปรวนไม่เท่ากัน และความสัมพันธ์ไม่เป็นเส้นตรง การแปลงบอกซ์-คอกซ์ มีสูตรในการแปลงข้อมูล Y ดังนี้

$$y(\lambda) = \begin{cases} \dfrac{y^{\lambda}-1}{\lambda} & ,\lambda \neq 0 \\ \log(y) & ,\lambda = 0 \end{cases}$$

เมื่อ y เป็นค่าของตัวแปรตามที่ $y > 0$ และ λ คือพารามิเตอร์ของการแปลง ที่ค่าต่าง ๆ ของ λ จะได้การแปลงข้อมูลเป็นชุด ดังนี้

ตารางที่ 2.2 การแปลงบอกซ์-คอกซ์

λ	y' ของ บอกซ์ – คอกซ์	y' ในทางปฏิบัติ
–2	$\dfrac{1-y^2}{2}$	$\dfrac{1}{y^2}$
–1	$1-y^{-1}$	$\dfrac{1}{y}$
–0.5	$2\left(1-y^{-\frac{1}{2}}\right)$	$\dfrac{1}{\sqrt{y}}$
0	$\ln y$	$\ln y$
0.5	$2\left(y^{\frac{1}{2}}-1\right)$	$\sqrt{y}$
1	$y-1$	y
2	$\dfrac{y^2-1}{2}$	y^2

เกณฑ์ในการกำหนด λ ที่เหมาะสมสำหรับการแปลงข้อมูล y คือ เลือกค่า λ ที่ทำให้ SSE มีค่าน้อยที่สุด หรือลอการิทึมของฟังก์ชันภาวะน่าจะเป็นมีค่ามากที่สุด หรือพิจารณาจากช่วงความเชื่อมั่น 95% ของ λ ซึ่งค่า λ ที่เหมาะสมจะตกอยู่ในช่วงความเชื่อมั่น 95% นั่นเอง

2.3.4 การตรวจสอบตัวแบบการถดถอยเชิงเส้น

การตรวจสอบตัวแบบการถดถอยเป็นเทคนิคหนึ่งที่ใช้สำหรับการสืบหา (detecting) หรือตรวจสอบ (checking) ปัญหาที่เกิดขึ้นในการวิเคราะห์การถดถอย เนื่องจากเราไม่สามารถมั่นใจว่า ตัวแบบการถดถอยที่ได้นั้นเหมาะสมกับข้อมูลหรือไม่ ดังนั้นจึงมีความจำเป็นต้องมีการตรวจสอบ ซึ่งจะตรวจสอบข้อสมมติของตัวแบบ (model assumption) เป็นข้อสมมติเกี่ยวกับค่าคลาดเคลื่อนสุ่ม (random error) และการตรวจสอบค่านอกเกณฑ์ (outliers) มี 4 ข้อดังนี้

1. ค่าคลาดเคลื่อนสุ่มจะมีค่าเฉลี่ย $\mathrm{E}(\boldsymbol{\varepsilon})=\mathbf{0}$
2. ค่าคลาดเคลื่อนสุ่มต้องมีการแจกแจงปรกติ
3. ค่าคลาดเคลื่อนสุ่มจะมีความแปรปรวนมีค่าคงที่
4. ค่าคลาดเคลื่อนสุ่ม $\boldsymbol{\varepsilon}$ จะต้องเป็นอิสระต่อกัน

2.3.5 การเลือกตัวแบบที่เหมาะสม

ในการพิจารณาตัวแบบการถดถอยเชิงเส้นพหุคูณสำหรับข้อมูลจากตัวอย่างนั้น อาจจะมีตัวแบบหลายตัวแบบที่เหมาะสมกับข้อมูลและมีประสิทธิภาพในการทำนายหรือประมาณค่าใกล้เคียงกัน ผู้ทำการวิเคราะห์จึงต้องพิจารณาว่าตัวแบบใดเหมาะสมกับข้อมูลมากที่สุด ซึ่งสามารถพิจารณาได้จากค่าสถิติต่อไปนี้

1. สัมประสิทธิ์การตัดสินใจพหุคูณ R^2

สัมประสิทธิ์การตัดสินใจพหุคูณ (Multiple Coefficient of Determination) หรือ R^2 คือสัดส่วนของความแปรผันทั้งหมดในตัวแปรตามที่สามารถอธิบายได้ด้วยตัวแปรอิสระในตัวแบบ ดังนั้นจึงสามารถใช้ R^2 เป็นตัววัดความสามารถของตัวแบบว่าเหมาะสมกับข้อมูลมากน้อยเพียงใดได้ นั่นคือเกณฑ์ R^2 นี้จะต้องเลือกตัวแบบที่ให้ค่า R^2 สูงสุด ซึ่งสามารถคำนวณได้จากสูตร

$$\mathrm{R}^2=\frac{\mathrm{SSR}}{\mathrm{SST}}=\frac{\mathrm{SST}-\mathrm{SSE}}{\mathrm{SST}}=1-\frac{\mathrm{SSE}}{\mathrm{SST}}$$

โดยที่ $\mathrm{SST}=\sum_{i=1}^{n}(y_i-\bar{y})^2$, $\mathrm{SSR}=\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2$ และ $\mathrm{SSE}=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$

เนื่องจาก $0\le \mathrm{SSE}\le \mathrm{SST}$ ดังนั้น R^2 จึงมีค่าตั้งแต่ 0 ถึง 1 หรือ $0\le \mathrm{R}^2\le 1$ ถ้า

- $\mathrm{R}^2=1$ แสดงว่า $\mathrm{SSE}=0$ หมายความว่า ทุก ๆ จุดของค่าสังเกตจะอยู่บนเส้นการถดถอยประมาณ นั่นคือความผันแปรทั้งหมดใน Y สามารถอธิบายได้ด้วยเส้นการถดถอย ซึ่งในที่นี้หมายถึงมีความสัมพันธ์กันในลักษณะเชิงเส้น
- $\mathrm{R}^2=0$ แสดงว่า ความผันแปรทั้งหมดใน Y ไม่ได้เนื่องมาจาก x
- R^2 มีค่าเข้าใกล้ 0 แสดงว่า Y และ x มีความสัมพันธ์ในลักษณะเชิงเส้นน้อย
- R^2 มีค่าเข้าใกล้ 1 แสดงว่า Y และ x มีความสัมพันธ์ในลักษณะเชิงเส้นมาก

แต่สิ่งที่จะต้องระวังก็คือ การเพิ่มตัวแปรอิสระเข้าไปในตัวแบบนั้นไม่ทำให้ R^2 ลดลง และตัวแบบที่มีตัวแปรอิสระทุกตัวในตัวแบบมักจะให้ค่า R^2 มากที่สุด ดังนั้นจึงไม่ควรใช้เกณฑ์นี้เพียงอย่างเดียวในการพิจารณาตัวแบบที่เหมาะสมที่สุด โดยทั่วไปเราจะใช้ $R^2 \times 100$ เป็นตัวบอกถึงเปอร์เซ็นต์ของความสัมพันธ์ในลักษณะเชิงเส้นระหว่าง Y และ x

2. ค่าความคลาดเคลื่อนมาตรฐานของส่วนเหลือ (Residual Standard Error)

การคำนวณหาตัวแบบการถดถอยก็เพื่อประมาณค่าของสัมประสิทธิ์ของการถดถอย เพื่อประมาณค่าเฉลี่ยของตัวแปรตามเมื่อกำหนดค่าของตัวแปรอิสระ และทำนายค่าใหม่ของตัวแปรตาม ในกรณีที่เป็นการประมาณค่าแบบช่วงนั้น ค่าประมาณจะขึ้นอยู่กับค่าคลาดเคลื่อนมาตรฐานของส่วนเหลือ (Residual Standard Error หรือ $\sqrt{MSE}$) เพื่อให้ช่วงความเชื่อมั่นแคบ แสดงว่าต้องการให้ $\sqrt{MSE}$ มีค่าน้อย นั่นคือตัวแบบที่เหมาะสมได้แก่ตัวแบบที่ R^2 มีค่ามาก และ $\sqrt{MSE}$ มีค่าน้อย

สำหรับการเปรียบเทียบตัวแบบการถดถอย เกณฑ์ $\sqrt{MSE}$ จะดีกว่าเกณฑ์ R^2 เพราะว่าการเพิ่มตัวแปรอิสระใดๆ ก็ตามเข้าไปในตัวแบบก็จะเพิ่มค่า R^2 แต่ถ้าเพิ่มตัวแปรอิสระที่ไม่มีความสำคัญต่อตัวแปรอิสระแล้ว อาจทำให้ค่าของ $\sqrt{MSE}$ ไม่ลดลง ตัวแบบที่ดีคือ ตัวแบบที่มีค่า $\sqrt{MSE}$ น้อยๆ ซึ่งสามารถคำนวณได้จากสูตร

$$\sqrt{MSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}$$

โดยที่ y_i คือ ค่าสังเกตของข้อมูลตัวที่ i

$\hat{y}_i$ คือ ค่าพยากรณ์ของข้อมูลตัวที่ i

n คือ จำนวนข้อมูล

3. สถิติ AIC

รายละเอียดแสดงดังหน้าที่ 10

4. สถิติ BIC

สถิติ BIC ย่อมาจาก Baysian Information Criterion ซึ่งเสนอโดย Schwarz, G. ในปี 1978 มีแนวคิดเหมือนกับ AIC แต่มีการนำจำนวนค่าสังเกตหรือขนาดตัวอย่าง (n) เข้ามาพิจารณาด้วย โดยที่

$$BIC = -2 \times \widehat{\log L} + 2(\text{จำนวนพารามิเตอร์ในตัวแบบ})$$

โดยตัวแบบที่ให้ค่า BIC น้อยที่สุดจะเป็นตัวแบบที่ถูกเลือก

2.4 งานวิจัยที่เกี่ยวข้อง

การศึกษาครั้งนี้ผู้วิจัยได้ทบทวนเอกสาร วรรณกรรม และงานวิจัยที่เกี่ยวข้องเพื่อมาสนับสนุนงานวิจัย โดยทำการศึกษางานวิจัยของผู้เชี่ยวชาญหลาย ๆ ท่าน ดังนี้

ผกากรอง เทพรักษ์ (2557) ได้ศึกษาปัจจัยที่ส่งผลต่อการมีงานทำ อัตราการว่างงาน พฤติกรรมการหางานทำตลอดจนปัญหา อุปสรรค และข้อเสนอแนะในการเตรียมตัวสมัครงาน ของบัณฑิต ประชากร คือ บัณฑิตมหาวิทยาลัยเทคโนโลยีราชมงคลศรีวิชัย วิทยาเขตสงขลา ที่จบการศึกษาในปีการศึกษา 2550-2554 จำนวน 7,593 คน ทำการสุ่มตัวอย่างแบบแบ่งชั้นภูมิ โดยใช้หน่วยคณะและสาขาเป็นชั้นภูมิ ได้กลุ่มตัวอย่างจำนวน 390 คน วิเคราะห์ความสัมพันธ์โดยใช้ไคสแควร์ ผลการศึกษาพบว่า 1. ปัจจัยที่ส่งผลต่อการมีงานทำของบัณฑิตได้แยกเป็น 2 ส่วน คือ ปัจจัยส่วนบุคคล และปัจจัยพฤติกรรมการหางานทำ โดยปัจจัยส่วนบุคคลที่พบว่าส่งผลต่อการมีงานทำของบัณฑิต ได้แก่ อาชีพของบิดา รูปร่างหน้าตา ความเป็นผู้นำ ความเชื่อมั่นในตนเอง ความกล้าแสดงออก และประสบการณ์การทำงาน ของบัณฑิต ส่วนปัจจัยด้านพฤติกรรมการหางานทำที่ส่งผลต่อการมีงานทำ ได้แก่ การเตรียมตัวด้านคอมพิวเตอร์การหาข้อมูลข่าวสารเกี่ยวกับตำแหน่งงาน 2. บัณฑิตเพศชายมีงานทำประมาณ 77% ส่วนบัณฑิตเพศหญิงมีงานทำประมาณ 82% ในภาพรวมพบว่าบัณฑิตมีงานทำประมาณ 80% โดยทำงานตรงวุฒิประมาณร้อยละ 68% บัณฑิตที่มีงานทำร้อยละ 50% ได้รับเงินเดือนน้อยกว่าวุฒิ ซึ่งในทางเศรษฐศาสตร์ถือว่าการจบออกไปทำงานโดยยอมรับเงินเดือนที่ต่ำกว่าวุฒิการศึกษาถือว่าเป็นการว่างงานแฝง ดังนั้นการศึกษาครั้งนี้สะท้อนให้เห็นว่าบัณฑิตที่มีงานทำและได้เงินเดือนตรงตามวุฒิประมาณครึ่งหนึ่งของบัณฑิตที่จบออกไปเท่านั้น 3. พฤติกรรมการหางานทำของบัณฑิต พบว่า บัณฑิตส่วนใหญ่หาข่าวสารเกี่ยวกับตำแหน่งงานว่างจากสื่ออินเตอร์เน็ต รองลงมา คือ สื่อหนังสือพิมพ์ โดยจำนวนแหล่งงานที่สมัครเฉลี่ยประมาณคนละ 2-3 แหล่ง นอกจากนี้ยังพบว่า มีบัณฑิตเตรียมตัวก่อนการสมัครงานด้านภาษาต่างประเทศประมาณ 89% ด้านคอมพิวเตอร์ประมาณ 77% สิ่งที่บัณฑิตคำนึงถึงในการสมัครงานมากที่สุดคือ คุณวุฒิและสาขาวิชาที่สำเร็จการศึกษา รองลงมาคือ ชื่อเสียงของบริษัท และการมีคนที่รู้จักทำอยู่ก่อน ตามลำดับ

วรรณวิมล จงจรวยสกุล (2555) ได้ศึกษาปัจจัยที่มีความสัมพันธ์กับการมีงานทำและการประกอบอาชีพของบัณฑิตวิทยาลัยราชพฤกษ์ ปีการศึกษา 2554- 2555 จำนวน 1,209 คน จากผู้สำเร็จการศึกษา จำนวน 1,245 คน คิดเป็นร้อยละ 97.10 เครื่องมือในการวิจัย ได้แก่ แบบสอบถามภาวะการมีงานทำ นำข้อมูลที่ได้มาวิเคราะห์ด้วยสถิติเชิงพรรณนา ได้แก่ การแจกแจงความถี่ และค่าร้อยละ ผลการศึกษาพบว่า บัณฑิตมีงานทำก่อนเข้าศึกษาร้อยละ 27.6 มีงานทำระหว่างศึกษาร้อยละ 32.2 และมีงานทำหลังสำเร็จการศึกษาร้อยละ 40.2 หน่วยงานที่ทำเป็นบริษัท/องค์กรธุรกิจเอกชนมากที่สุดร้อยละ 56.1 เงินเดือนที่ได้รับส่วนใหญ่คือ 10,001-12,000 บาท คิดเป็นร้อยละ 26.2 สำหรับบัณฑิตที่ยังไม่ได้งานทำ สาเหตุที่ยังไม่ได้ทำงานมากที่สุด พบว่ารอฟังคำตอบจากหน่วยงานร้อยละ 61.4 ส่วนปัญหาที่พบในการทำงานหลังสำเร็จการศึกษามากที่สุดได้แก่ การหางานที่ถูกใจไม่ได้ร้อยละ 36

พรทวี เถื่อนคำแสน และคณะ (2559) ได้ศึกษาปัจจัยด้านประชากรศาสตร์ของพนักงานระดับปฏิบัติการในนิคมอุตสาหกรรมเหมราช อีสเทิร์นซีบอร์ด จังหวัดระยอง มีวัตถุประสงค์

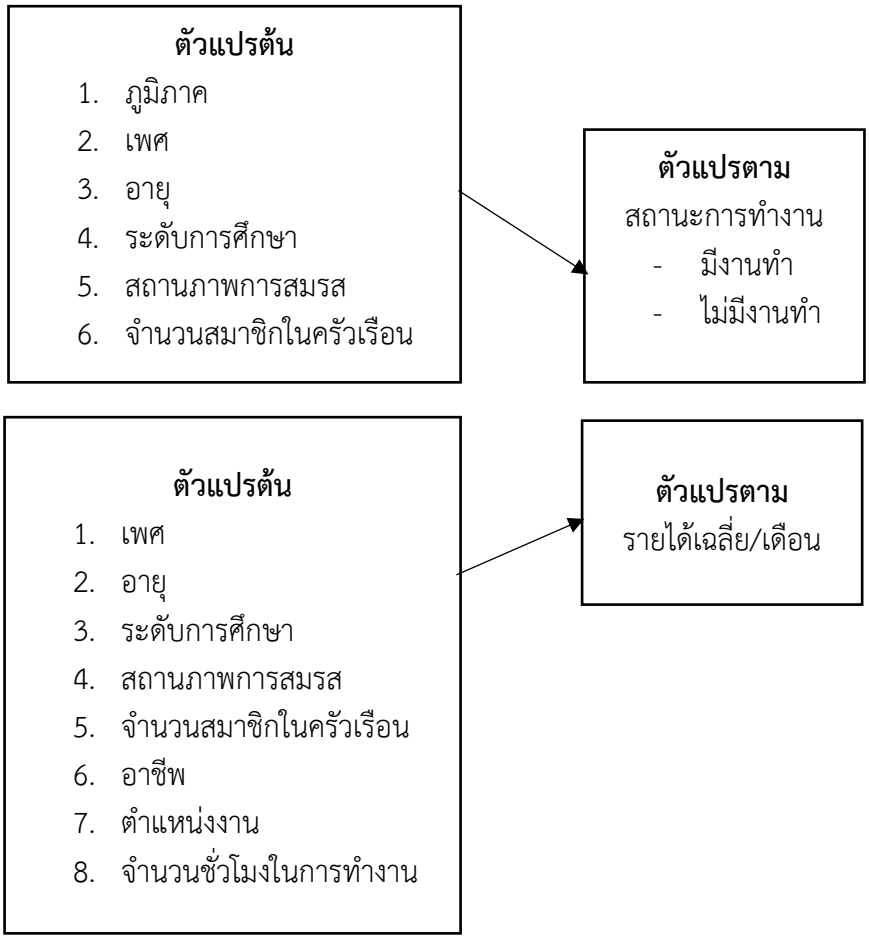
เพื่อเปรียบเทียบปัจจัยด้านประชากรศาสตร์ ได้แก่ เพศ อายุ ระดับการศึกษา สถานภาพสมรส ระดับรายได้ต่อเดือน และอายุการทำงาน ที่แตกต่างกันมีผลต่อการตัดสินใจเลือกทำงานของพนักงานระดับปฏิบัติการในนิคมอุตสาหกรรมเหมราช อีสเทิร์นซีบอร์ด จังหวัดระยองแตกต่างกันหรือไม่ เก็บรวบรวมข้อมูลโดยใช้แบบสอบถามจากกลุ่มตัวอย่างจากพนักงานระดับปฏิบัติการที่ทำงานในนิคมอุตสาหกรรมเหมราช อีสเทิร์นซีบอร์ด จังหวัดระยอง จำนวน 400 คน สุ่มตัวอย่างโดยไม่ใช้ความน่าจะเป็น เลือกกลุ่มตัวอย่างแบบโควตาตามสัดส่วนของกลุ่มอุตสาหกรรมจำนวน 10 บริษัท บริษัทละ 40 ชุด และใช้วิธีเลือกสุ่มตัวอย่างแบบตามสะดวก สถิติที่ใช้ในการวิเคราะห์ข้อมูล ได้แก่ ร้อยละ ความถี่ ส่วนเบี่ยงเบนมาตรฐาน และนำข้อมูลที่ได้จากแบบสอบถามมาวิเคราะห์โดยใช้ T-Test และ F-Test (One way ANOVA) ผลการศึกษาพบว่า พนักงานเป็นเพศชาย ร้อยละ 77.75 มีอายุระหว่าง 20-30 ปี ร้อยละ 69.50 มีระดับการศึกษาอนุปริญญาหรือเทียบเท่า ร้อยละ 47.50 มีสถานภาพสมรสแต่งงาน ร้อยละ 51.75 มีรายได้ต่อเดือน 15,001 – 25,000 บาท ร้อยละ 60.75 และมีอายุการทำงาน 4-5 ปี ร้อยละ 40.50 ส่วนการเปรียบเทียบปัจจัยด้านประชากรศาสตร์ที่แตกต่างกัน มีผลต่อการตัดสินใจเลือกทำงานที่แตกต่างกันหรือไม่ พบว่า เพศ อายุ สถานภาพสมรส ระดับรายได้ต่อเดือน และอายุการทำงานที่แตกต่างกัน มีผลต่อการตัดสินใจเลือกทำงานที่ไม่แตกต่างกัน มีเพียงด้านระดับการศึกษาที่แตกต่างกันที่มีผลต่อการตัดสินใจเลือกทำงานที่แตกต่างกัน

องอาจ รุกขวัฒนกุล (2558) ได้ศึกษาปัจจัยที่มีอิทธิพลต่อการเปลี่ยนแปลงรายได้ของประชาชนในพื้นที่การพัฒนาเขตเศรษฐกิจพิเศษ หนองคาย มุกดาหาร แม่สอด มีวัตถุประสงค์เพื่อศึกษาผลกระทบต่อรายได้ของคนในพื้นที่จากการพัฒนาเขตเศรษฐกิจพิเศษ โดยใช้ข้อมูลรายได้ครัวเรือนจากสำมะโนประชากร และข้อมูลการสำรวจภาวะแรงงานของสำนักงานสถิติแห่งชาติ ตั้งแต่ปี พ.ศ. 2548 ถึงปี พ.ศ. 2557 นำมาวิเคราะห์ด้วยวิธีการวิเคราะห์การถดถอยพหุคูณ (Multiple Regression Analysis) ผลจากการศึกษา พบว่าปัจจัยที่มีผลต่อการเปลี่ยนแปลงรายได้ของครัวเรือนในพื้นที่เขตเศรษฐกิจพิเศษของหนองคาย มุกดาหาร และแม่สอด คือ จำนวนสมาชิกของครัวเรือน อายุ การศึกษา จำนวนชั่วโมงทำงาน อาชีพ และปีประกาศเขตเศรษฐกิจพิเศษ

จากการศึกษางานวิจัยข้างต้น ผกากรอง เทพรักษ์ (2557) กล่าวว่าปัจจัยที่มีผลต่อการมีงานทำของบัณฑิต ได้แก่ อาชีพ รูปร่างหน้าตา ความเป็นผู้นำ ความมั่นใจในตนเอง และการกล้าแสดงออก ในขณะที่พรทวี เถื่อนคำแสน (2559) กล่าวว่าปัจจัยที่มีผลต่อสภาพการทำงาน ได้แก่ เพศ อายุ สถานภาพสมรส และรายได้ต่อเดือน มีเพียงระดับการศึกษาที่แตกต่างกันเท่านั้นที่มีผลต่อการตัดสินใจเลือกงานที่แตกต่างกัน องอาจ รุกขวัฒนกุล (2558) เผยว่าปัจจัยที่มีผลต่อการเปลี่ยนแปลงรายได้ของครัวเรือน ได้แก่ จำนวนสมาชิกในครัวเรือน อายุ การศึกษา อาชีพ และจำนวนชั่วโมงทำงาน

นอกจากนี้วรรณวิมล จงจรวยสกุล (2555) พบว่าผู้คนส่วนใหญ่จะมีเงินเดือนอยู่ที่ 10,001-12,000 บาท ขณะที่พรทวี เถื่อนคำแสน และคณะ (2559) พบว่ากลุ่มตัวอย่างของพวกเขาส่วนใหญ่มีรายได้เฉลี่ยต่อเดือนอยู่ที่ 15,001 – 25,000 บาท ทางด้านขององอาจ รุกขวัฒนกุล (2558) ที่ศึกษาปัจจัยที่ส่งผลต่อรายได้ พบว่าปัจจัยที่มีผลต่อรายได้ของกลุ่มตัวอย่างนั้น ประกอบด้วยจำนวนสมาชิกของครัวเรือน อายุ การศึกษา จำนวนชั่วโมงทำงาน และอาชีพ ซึ่งตัวแปรที่ได้กล่าวมาข้างต้นส่วนหนึ่งเป็นตัวแปรที่ผู้วิจัยได้คัดเลือกเพื่อจะนำมาวิเคราะห์ในการศึกษาครั้งนี้ด้วย

2.5 กรอบแนวคิด



รูป 2.2 กรอบแนวคิด

บทที่ 3
วิธีดำเนินงานวิจัย

การดำเนินการวิจัยเรื่อง “ตัวแบบทางสถิติในการศึกษาภาวะการมีงานทำของประชากรไทย” โดยใช้ข้อมูลทุติยภูมิจากโครงการสำรวจภาวะการทำงานของประชากรไทย ในปี พ.ศ.2561 ของสำนักงานสถิติแห่งชาติ เพื่อให้การศึกษาครั้งนี้บรรลุจุดประสงค์ ผู้วิจัยจึงได้กำหนดวิธีการดำเนินการวิจัย ซึ่งมีรายละเอียดและขั้นตอนในการวิจัย ดังนี้

3.1 ข้อมูลที่ใช้ในการศึกษา
3.2 ตัวแปรที่ใช้ในการศึกษา
3.3 การนิยามและกำหนดค่าตัวแปร
3.4 ขั้นตอนในการดำเนินการศึกษา
3.5 การวิเคราะห์ข้อมูล

3.1 ข้อมูลที่ใช้ในการศึกษา

ข้อมูลที่ใช้ในการศึกษาครั้งนี้ คือ ข้อมูลจากผู้ตอบแบบสอบถามซึ่งเป็นข้อมูลทุติยภูมิจากโครงการสำรวจภาวะการทำงานของประชากรไทย ไตรมาสที่ 4 ประจำปี พ.ศ. 2561 ของสำนักงานสถิติแห่งชาติจำนวน 172,241 คน และข้อมูลจากผู้ตอบแบบสอบถามที่ระบุรายได้ต่อเดือนจำนวน 47,933 คน

3.2 ตัวแปรที่ใช้ในการศึกษา

1. ตัวแปรต้น

r	แทน ภูมิภาค
s	แทน เพศ
m	แทน สถานภาพการสมรส
e	แทน ระดับการศึกษา
o	แทน อาชีพ
t	แทน ตำแหน่งงาน
x_1	แทน อายุ
x_2	แทน จำนวนสมาชิกในครัวเรือน
x_3	แทน จำนวนชั่วโมงในการทำงาน

2. ตัวแปรตาม

y_1 แทน สถานะการทำงาน

y_2 แทน รายได้เฉลี่ย/เดือน

3.3 การนิยามและกำหนดค่าตัวแปร

ตารางที่ 3.1 การนิยามและกำหนดค่าตัวแปร

ตัวแปร	ค่าของตัวแปร	ประเภทของตัวแปร	ชนิดของตัวแปร
สถานะการทำงาน (y_1)	0 = ไม่มีงานทำ 1 = มีงานทำ	ตัวแปรเชิงคุณภาพ	ตัวแปรตาม
รายได้เฉลี่ย/เดือน (y_2)		ตัวแปรเชิงปริมาณ	
ภูมิภาค (r)	0 = กรุงเทพมหานคร 1 = ภาคใต้ 2 = ภาคกลาง 3 = ภาคเหนือ 4 = ภาค ตะวันออกเฉียงเหนือ	ตัวแปรเชิงคุณภาพ	ตัวแปรอิสระ
เพศ (s)	0 = หญิง 1 = ชาย	ตัวแปรเชิงคุณภาพ	
สถานภาพการสมรส (m)	0 = โสด 1 = สมรส 2 = ม่าย 3 = หย่าร้าง	ตัวแปรเชิงคุณภาพ	
ระดับการศึกษา (e)	0 = ไม่ได้รับการศึกษา 1 = ประถมศึกษาหรือต่ำกว่า 2 = มัธยมศึกษา 3 = ปริญญาตรีหรือเทียบเท่า 4 = สูงกว่าปริญญาตรี	ตัวแปรเชิงคุณภาพ	

ตารางที่ 3.1 การนิยามและกำหนดค่าตัวแปร (ต่อ)

ตัวแปร	ค่าของตัวแปร	ประเภทของตัวแปร	ชนิดของตัวแปร
อาชีพ (0)	0 = ข้าราชการ เสมียน และช่างเทคนิค 1 = พนักงานบริการ และผู้จำหน่าย ผลิตภัณฑ์ 2 = เกษตรและประมง 3 = ช่างฝีมือและผู้ควบคุมเครื่องจักร	ตัวแปรเชิงคุณภาพ	ตัวแปรอิสระ
ตำแหน่งงาน (t)	0 = ลูกจ้างรัฐบาล 1 = ลูกจ้างเอกชน 2 = ลูกจ้างทั่วไป	ตัวแปรเชิงคุณภาพ	
อายุ (X_1)		ตัวแปรเชิงปริมาณ	
จำนวนสมาชิกในครัวเรือน (X_2)		ตัวแปรเชิงปริมาณ	
จำนวนชั่วโมงในการทำงาน (X_3)		ตัวแปรเชิงปริมาณ	

หมายเหตุ : ข้อมูลในตารางที่ 3.1 เป็นข้อมูลของบุคคลที่มีอายุ 15 ปีขึ้นไป ของทุกครัวเรือนที่ตอบแบบสอบถาม

3.4 การวิเคราะห์ข้อมูล

ในการวิเคราะห์ข้อมูล ผู้วิจัยนำข้อมูลที่ได้มาวิเคราะห์โดยใช้โปรแกรม SPSS for Windows version 20.0 และโปรแกรม R-Studio กำหนดระดับนัยสำคัญอยู่ที่ 0.05 โดยมีขั้นตอนวิเคราะห์ดังต่อไปนี้

3.4.1 การวิเคราะห์ข้อมูลด้วยสถิติเชิงพรรณนา (Descriptive statistics)

1. ความถี่ (Frequency) ร้อยละ (Percentage) ของสถานะการจ้างงาน ภูมิภาค เพศ สถานภาพการสมรส และระดับการศึกษา

2. ค่าเฉลี่ย (Mean) ส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation) ค่าต่ำสุด (Minimum) และค่าสูงสุด (Maximum) ของอายุ และจำนวนสมาชิกในครัวเรือน

3.4.2 การวิเคราะห์การถดถอยลอจิสติก (Logistic regression analysis)

ใช้การวิเคราะห์การถดถอยลอจิสติก เพื่อทำนายโอกาสการจ้างงานของประชากรไทย โดยมีขั้นตอนในการวิเคราะห์ ดังนี้

1. สร้างตัวแบบการถดถอยลอจิสติกที่สามารถทำนายโอกาสการจ้างงานของประชากรไทย

2. ตรวจสอบตัวแบบการประมาณ โดยทำการตรวจสอบข้อสมมติเบื้องต้นเกี่ยวกับค่าคลาดเคลื่อนสุ่ม และใช้แผนภาพ Half-Normal Plot ควบคู่กัน

3. เลือกตัวแบบที่เหมาะสมโดยการเปรียบเทียบค่า Deviance Residual และ Degrees of Freedom และค่าสถิติ AIC

4. แปลผลค่า odds ratio และค่าการวิเคราะห์การจำแนกกลุ่ม

3.4.3 การวิเคราะห์การถดถอยเชิงเส้นพหุคูณ (Multiple Linear Regression Analysis)

ใช้การวิเคราะห์การถดถอยเชิงเส้นพหุคูณ เพื่อทำนายรายได้ต่อเดือนของประชากรไทย โดยมีขั้นตอนในการวิเคราะห์ ดังนี้

1. วิเคราะห์ข้อมูลด้วยสถิติเชิงพรรณนา ความถี่ ร้อยละ ของเพศ สถานภาพการสมรส ระดับการศึกษา อาชีพ และตำแหน่งงาน รวมทั้งหาค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ค่าต่ำสุด และค่าสูงสุด ของอายุ จำนวนสมาชิกในครัวเรือน และจำนวนชั่วโมงในการทำงาน

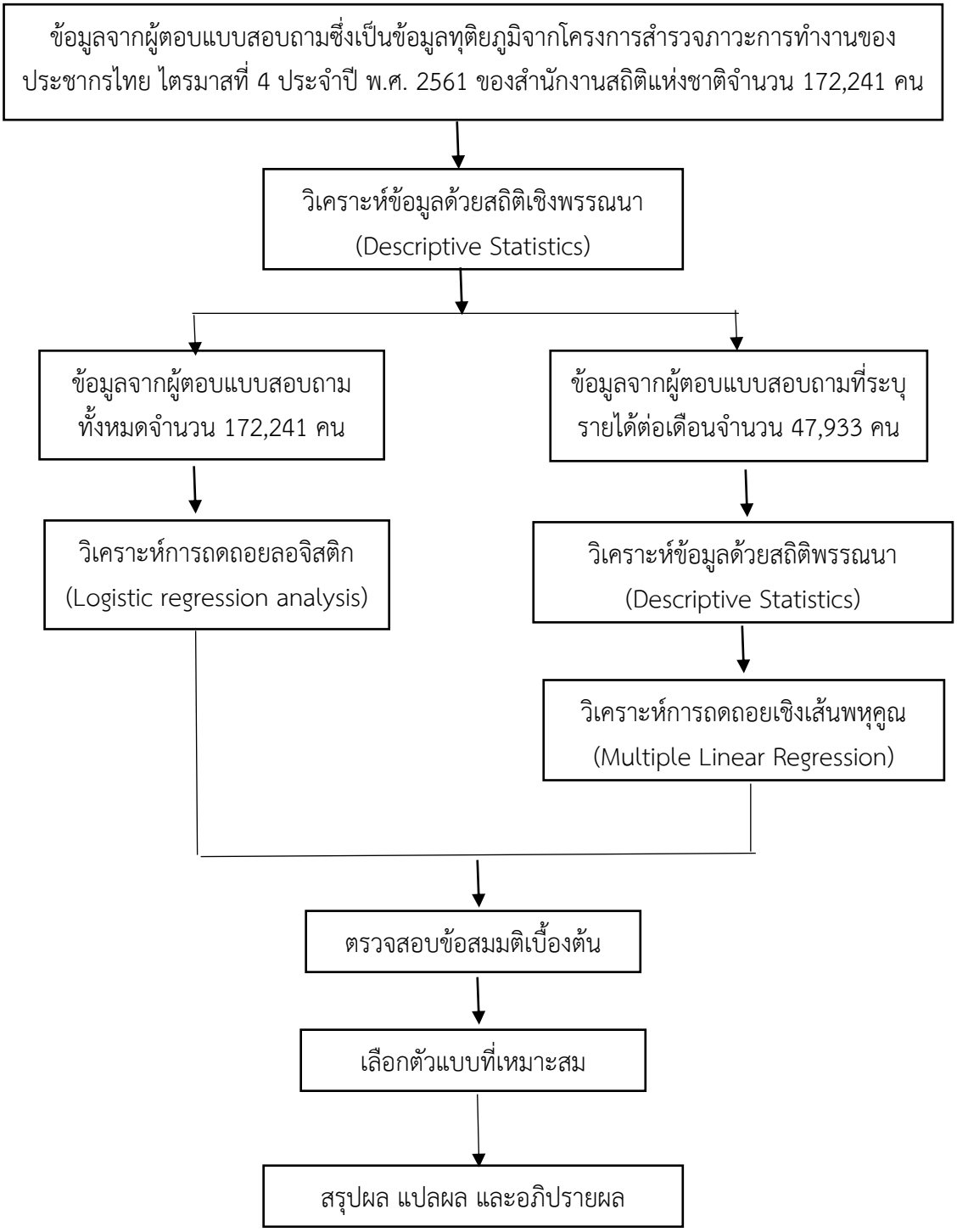
2. สร้างตัวแบบการถดถอยเชิงเส้นพหุคูณที่สามารถทำนายรายได้ต่อเดือนของประชากรไทย

3. ตรวจสอบตัวแบบการประมาณ โดยทำการตรวจสอบข้อสมมติเบื้องต้นเกี่ยวกับค่าคลาดเคลื่อนสุ่ม หากตัวแบบการประมาณไม่เป็นไปตามข้อสมมติเบื้องต้น สามารถแก้ปัญหาได้โดยการแปลงข้อมูลด้วยบอกซ์-คอกซ์ (Box-Cox transformation) และสร้างตัวแบบการประมาณใหม่อีกครั้ง

4. เลือกตัวแบบที่เหมาะสมโดยการเปรียบเทียบค่าสัมประสิทธิ์การตัดสินใจพหุคูณ (R^2) ค่าความคลาดเคลื่อนมาตรฐานของส่วนเหลือ (Residual Standard Error) ค่าสถิติ AIC และค่าสถิติ BIC

5. แปลผลค่าพารามิเตอร์ของตัวแบบการถดถอย

3.5 แผนภาพขั้นตอนการวิเคราะห์ข้อมูล



รูป 3.1 แผนภาพขั้นตอนการวิเคราะห์ข้อมูล

บทที่ 4
ผลการวิเคราะห์ข้อมูล

การดำเนินการวิจัยเรื่อง “ตัวแบบทางสถิติในการศึกษาภาวะการมีงานทำของประชากรไทย” โดยใช้ข้อมูลทุติยภูมิจากโครงการสำรวจภาวะการทำงานของประชากรไทย ในปี พ.ศ. 2561 ของสำนักงานสถิติแห่งชาติ มีวัตถุประสงค์เพื่อศึกษาการทำงานของประชากรไทย ศึกษาปัจจัยที่มีผลต่อภาวะการมีงานทำของประชากรไทย หาตัวแบบที่เหมาะสมทำนายภาวะการมีงานทำของประชากรไทย และหาตัวแบบที่เหมาะสมทำนายรายได้ของประชากรไทย ในจะการนำเสนอผลการวิเคราะห์ข้อมูลผู้วิจัยได้แบ่งเป็น 3 ส่วน ดังนี้

4.1 ผลการวิเคราะห์ข้อมูลเบื้องต้น โดยใช้ความถี่ และร้อยละ นำเสนอด้วยตารางแจกแจงความถี่

4.2 ผลการวิเคราะห์จากการถดถอยลอจิสติก (Logistic regression analysis) เพื่อหาปัจจัยที่สามารถทำนายโอกาสการจ้างงานของประชากรไทย โดยจะนำเสนอตัวแบบ ขั้นตอนการตรวจสอบตัวแบบ การเลือกตัวแบบ และการแปลผลจากตัวแบบ

4.3 ผลการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ (Multiple Linear Regression Analysis) เพื่อทำนายรายได้ต่อเดือนของประชากรไทย โดยจะนำเสนอตัวแบบ ขั้นตอนการตรวจสอบตัวแบบ การเลือกตัวแบบ และการแปลผลจากตัวแบบ

4.1 ผลการวิเคราะห์ข้อมูลเบื้องต้น

ผลการวิเคราะห์ข้อมูลเบื้องต้น โดยใช้ความถี่ และร้อยละ กับข้อมูลสถานะการจ้างงาน ภูมิภาค เพศ สถานภาพการสมรส และระดับการศึกษา รวมทั้งหาค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ค่าต่ำสุด และค่าสูงสุดของอายุ และจำนวนสมาชิกในครัวเรือน นำเสนอด้วยตารางแจกแจงความถี่ ดังต่อไปนี้

ตารางที่ 4.1 ผลการวิเคราะห์ข้อมูลเบื้องต้นของผู้ตอบแบบสอบถามจำนวน 172,241 คน

ตัวแปร	ความถี่ (คน)	ร้อยละ
ส่วนที่ 1 ผลการวิเคราะห์ข้อมูลเบื้องต้นของตัวแปรเชิงคุณภาพ		
สถานะการทำงาน (y_1)		
• มีงานทำ	112,511	65.3
• ไม่มีงานทำ	59,730	34.7

ตารางที่ 4.1 ผลการวิเคราะห์ข้อมูลเบื้องต้นของผู้ตอบแบบสอบถามจำนวน 172,241 คน (ต่อ)

ตัวแปร	ความถี่ (คน)	ร้อยละ
ส่วนที่ 1 ผลการวิเคราะห์ข้อมูลเบื้องต้นของตัวแปรเชิงคุณภาพ		
ภูมิภาค (r)		
• ภาคใต้	29,582	17.2
• ภาคกลาง	51,561	29.9
• ภาคเหนือ	36,919	21.4
• ภาคตะวันออกเฉียงเหนือ	44,765	26.0
• กรุงเทพมหานคร	9,414	5.5
เพศ (s)		
• ชาย	80,163	46.5
• หญิง	92,078	53.5
สถานภาพการสมรส (m)		
• โสด	36,729	21.3
• สมรส	109,674	63.7
• ม่าย	17,472	10.1
• หย่าร้าง	8,366	4.9
ระดับการศึกษา (e)		
• ประถมศึกษาหรือต่ำกว่า	84,618	49.1
• มัธยมศึกษา	51,544	29.9
• ปริญญาตรีหรือเทียบเท่า	25,008	14.5
• สูงกว่าปริญญาตรี	2,698	1.6
• ไม่ได้รับการศึกษา	8,373	4.9
รวม	172,241	100.0

ส่วนที่ 2 ผลการวิเคราะห์ข้อมูลเบื้องต้นของตัวแปรเชิงปริมาณ				
ตัวแปร	**ค่าเฉลี่ย**	**SD**	**ค่าต่ำสุด**	**ค่าสูงสุด**
อายุ (x_1)	47.61	17.743	15	98
จำนวนสมาชิกในครัวเรือน (x_2)	3.57	1.834	1	18

จากตารางที่ 4.1 ผลการวิเคราะห์ข้อมูลเบื้องต้นของผู้ตอบแบบสอบถามจำนวน 172,241 คน พบว่า คนส่วนใหญ่มีงานทำจำนวน 112,511 คน หรือคิดเป็นร้อยละ 65.3 และคนที่ไม่มีงานทำจำนวน 59,730 คน หรือคิดเป็นร้อยละ 34.7 ในจำนวนนี้มีคนที่อาศัยอยู่กรุงเทพมหานคร

จำนวน 9,414 คน (ร้อยละ 5.5) ภาคใต้จำนวน 29,582 คน (ร้อยละ 17.2) ภาคกลางจำนวน 51,561 คน (ร้อยละ 29.9) ภาคเหนือจำนวน 36,919 คน (ร้อยละ 21.4) และภาคตะวันออกเฉียงเหนือจำนวน 44,765 คน (ร้อยละ 26)

โดยแบ่งเป็นเพศชายจำนวน 80,163 คน (ร้อยละ 46.5) และเพศหญิงจำนวน 92,078 คน (ร้อยละ 53.5) มีอายุเฉลี่ย 47.61 ปี เป็นผู้ที่มีสถานภาพโสดจำนวน 36,729 คน (ร้อยละ 21.3) ผู้ที่มีสถานภาพสมรสจำนวน 109,674 คน (ร้อยละ 63.7) ผู้ที่มีสถานภาพเป็นม่ายจำนวน 17,472 คน (ร้อยละ 10.1) และเป็นผู้ที่มีสถานภาพหย่าร้างจำนวน 8,366 คน (ร้อยละ 4.9)

จำแนกตามระดับการศึกษา พบว่า เป็นผู้ที่จบชั้นประถมศึกษาหรือต่ำกว่าจำนวน 84,618 คน (ร้อยละ 49.1) ผู้ที่จบมัธยมศึกษาจำนวน 51,544 คน (ร้อยละ 29.9) ผู้ที่จบปริญญาตรีหรือเทียบเท่าจำนวน 25,008 คน (ร้อยละ 14.5) ผู้ที่จบสูงกว่าปริญญาตรีจำนวน 2,698 คน (ร้อยละ 1.6) และผู้ที่ไม่ได้รับการศึกษาจำนวน 8,373 คน (ร้อยละ 4.9) โดยมีจำนวนสมาชิกในครัวเรือนเฉลี่ยอยู่ที่ 3.57 คน

4.2 การวิเคราะห์การถดถอยลอจิสติกของสถานะการจ้างงานของประชากรไทย

จากการนำข้อมูลจากโครงการสำรวจภาวะการทำงานของประชากรไทย ในปี พ.ศ. 2561 ของสำนักงานสถิติแห่งชาติ จำนวน 172,241 คน เพื่อนำมาวิเคราะห์โอกาสในการจ้างงานของประชากรไทย สามารถเขียนตัวแบบของโอกาสในการจ้างงาน ได้ดังนี้

$$P(Y_1 = 1) = \frac{1}{1+e^{-z}} \qquad \textbf{(4.1)}$$

เมื่อ $P(Y_1 = 1)$ คือ ความน่าจะเป็นที่ประชากรไทยจะมีงานทำ

และ $z = \beta_0 + \beta_1 r_1 + \beta_2 r_2 + \beta_3 r_3 + \beta_4 r_4 + \beta_5 s_1 + \beta_6 m_1 + \beta_7 m_2 + \beta_8 m_3$
$+\beta_9 e_1 + \beta_{10} e_2 + \beta_{11} e_3 + \beta_{12} e_4 + \beta_{13} x_1 + \beta_{14} x_1$

โดยที่ $\beta_0, \beta_1, \beta_2, \ldots, \beta_{14}$ คือ สัมประสิทธิ์การถดถอยซึ่งเป็นพารามิเตอร์ของตัวแบบ

r_1 คือ ภาคกลางของประเทศไทย

r_2 คือ ภาคเหนือของประเทศไทย

r_3 คือ ภาคตะวันออกเฉียงเหนือของประเทศไทย

r_4 คือ ภาคใต้ของประเทศไทย

s_1 คือ เพศชาย

m_1 คือ สถานภาพสมรส

m_2 คือ สถานภาพม่าย

m_3 คือ สถานภาพหย่าร้าง

e_1 คือ ประถมศึกษาหรือต่ำกว่า

e_2 คือ มัธยมศึกษา

e_3 คือ ปริญญาตรีหรือเทียบเท่า

e_4 คือ สูงกว่าปริญญาตรี

x_1 คือ อายุของประชากรไทย

x_2 คือ จำนวนสมาชิกในครัวเรือนของประชากรไทย

สร้างตัวแบบการถดถอยลอจิสติกเพื่อทำนายโอกาสที่จะมีงานทำของประชากรไทย พร้อมทั้งตรวจสอบตัวแบบการประมาณ โดยทำการตรวจสอบข้อสมมติเบื้องต้นเกี่ยวกับค่าคลาดเคลื่อนสุ่ม (Random Error) และใช้แผนภาพ Half-Normal Plot รวมทั้งเปรียบเทียบค่า Deviance Residual และ Degrees of Freedom และแสดงค่าสถิติ AIC โดยสามารถเขียนตัวแบบการประมาณ ได้ดังนี้

$$p(y_1 = 1) = \frac{1}{1+e^{-z}} \qquad \textbf{(4.2)}$$

เมื่อ $p(y_1 = 1)$ คือ ค่าประมาณของความน่าจะเป็นที่ประชากรไทยจะมีงานทำ

และ $z = 1.01 + 0.208r_1 + 0.181r_2 + 0.14r_3 + 0.172r_4 + 0.782s_1 + 1.766m_1 + 0.671m_2$
$+1.82m_3 + 0.325e_1 - 0.002e_2 + 1.043e_3 + 1.508e_4 - 0.043x_1 - 0.103x_1$

จากตัวแบบ (4.2) สามารถแสดงค่า Deviance Residual และ Degrees of Freedom ที่ได้จากตัวแบบ และคำนวณค่าสถิติ AIC ใส่ลงในตารางได้ดังนี้

ตารางที่ 4.2 Deviance Residual และ Degrees of Freedom และแสดงค่าสถิติ AIC ของตัวแบบ (4.2)

Deviance Residual	Degrees of Freedom	AIC
190,580	172,226	190,610

อภิปรายผลจากตาราง (4.2) การเปรียบเทียบอัตราส่วนระหว่าง Deviance Residual และ Degrees of Freedom มีค่าใกล้เคียงกัน จึงสรุปได้ว่าตัวแบบเหมาะสมกับข้อมูล และมีค่า AIC เท่ากับ 190,610

การตรวจสอบข้อสมมติเบื้องต้นของตัวแบบ (4.2)

- ตรวจสอบการแจกแจงปรกติของค่าคลาดเคลื่อนสุ่มของตัวแบบการถดถอย โดยใช้การทดสอบ Kolmogorov-Smirnov test

สมมติฐานการทดสอบ คือ

$\mathbf{H_0}$: ค่าคลาดเคลื่อนสุ่มของตัวแบบการถดถอยประมาณมีการแจกแจงปรกติ

$\mathbf{H_1}$: ค่าคลาดเคลื่อนสุ่มของตัวแบบการถดถอยประมาณมีการแจกแจงแบบอื่น

ตารางที่ 4.3 ผลการวิเคราะห์การตรวจสอบการแจกแจงปรกติของค่าคลาดเคลื่อนสุ่มโดย Kolmogorov-Smirnov test ของตัวแบบ (4.2)

ค่าสถิติ	p-value
0.65322	0.000

- ตรวจสอบความแปรปรวนของค่าคลาดเคลื่อนสุ่มของตัวแบบการถดถอย โดยใช้การทดสอบ Bartlett test of homogeneity of variances

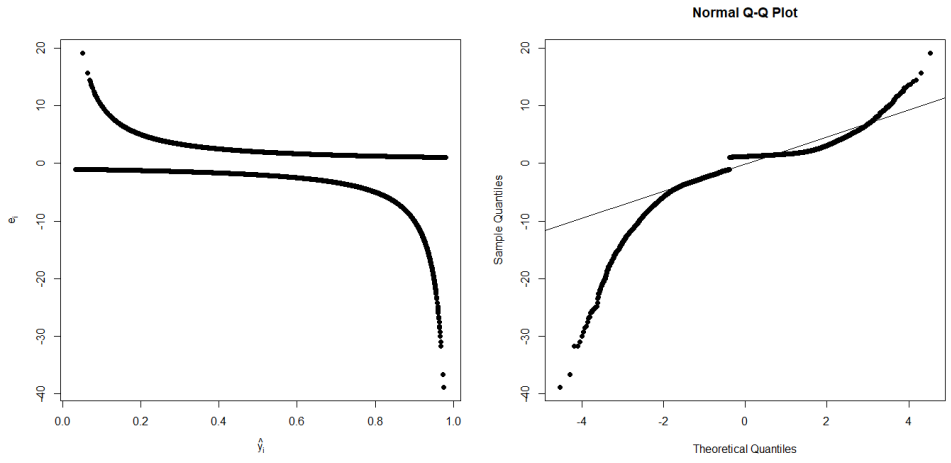
สมมติฐานการทดสอบ คือ

$\mathbf{H_0}$: ค่าคลาดเคลื่อนสุ่มของตัวแบบการถดถอยประมาณมีความแปรปรวนคงที่

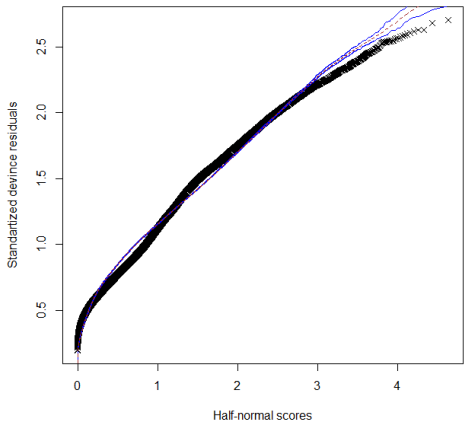
$\mathbf{H_1}$: ค่าคลาดเคลื่อนสุ่มของตัวแบบการถดถอยประมาณมีความแปรปรวนไม่คงที่

ตารางที่ 4.4 ผลการวิเคราะห์การตรวจสอบการความแปรปรวนของค่าคลาดเคลื่อนสุ่มโดย Bartlett test of homogeneity of variances ของตัวแบบ (4.2)

ค่าสถิติ Bartlett's K-squared	p-value
8,447.8	0.000



รูป 4.1 Normal Q-Q plot และแผนภาพการกระจายของส่วนเหลือ (Residuals) ของตัวแบบ (4.2)



รูป 4.2 แผนภาพ Half-Normal Plot ของตัวแบบ (4.2)

อภิปรายผลจากข้อสมมติเบื้องต้นของตัวแบบการถดถอยเชิงเส้น ค่าคลาดเคลื่อนสุ่มจะต้องมีการแจกแจงปรกติและมีความแปรปรวนคงที่ โดยการทดสอบการแจกแจงปรกตินั้นจะพิจารณาจากแผนภาพ Normal Q-Q plot ของส่วนเหลือ และ Kolmogorov-Smirnov test พบว่า ค่าคลาดเคลื่อนสุ่มจากตัวแบบการถดถอยประมาณ ไม่กระจายตัวรอบเส้นปรกติโดยเฉพาะหางและปลายของกราฟมีลักษณะกระจายตัวออกห่างจากเส้นปรกติ และจากการทดสอบโดยใช้สถิติทดสอบ Kolmogorov-Smirnov test ให้ค่าสถิติ D = 0.65322, ค่า p-value = 0.000 ซึ่งมีค่าน้อยกว่าระดับนัยสำคัญ 0.05 จึงสรุปได้ว่าค่าคลาดเคลื่อนสุ่มมีการแจกแจงแบบอื่นที่ไม่ใช่การแจกแจงปรกติ ส่วนการตรวจความแปรปรวนของค่าคลาดเคลื่อนสุ่ม พิจารณาจาก Bartlett test of homogeneity of variances ให้ค่าสถิติ Bartlett's K-squared = 8,447.8, ค่า p-value = 0.000 ซึ่งมีค่าน้อยกว่าระดับนัยสำคัญ 0.05 จึงสรุปได้ว่าค่าคลาดเคลื่อนสุ่มมีความแปรปรวนไม่คงที่ และจากแผนภาพ Half-Normal Plot พบว่ามีค่า Standardized Deviance residual ที่ได้จากการสร้างตัวแบบหลายค่าอยู่นอกเส้น Upper envelope และ Lower envelope

แม้ว่าตัวแบบการประมาณจะไม่ผ่านข้อสมมติเบื้องต้น ซึ่งอาจเกิดจากค่านอกเกณฑ์ (outliers) ของตัวแปรอิสระที่ส่งผลกระทบต่อตัวแปรตาม แต่จากการทดสอบความมีนัยสำคัญของตัวแปรอิสระ พบว่าตัวแปรอิสระทุกตัวส่งผลต่อโอกาสในการมีงานทำอย่างมีนัยสำคัญ อีกทั้งเมื่อได้ทำการคัดเลือกตัวแปรโดยใช้วิธีการถดถอยทีละขั้น (stepwise regression) ยังคงได้ผลลัพธ์ดังตัวแบบ (4.2) หรือแม้กระทั้งผู้วิจัยได้ทดลองสุ่มข้อมูลส่วนหนึ่งมาจากข้อมูลทั้งหมด โดยสุ่มให้ได้ผู้มีงานทำและผู้ที่ไม่มีงานทำในสัดส่วนที่เท่ากัน เนื่องด้วยผู้วิจัยคิดว่าที่ตัวแบบไม่ผ่านข้อสมมติเบื้องต้นอาจเกิดมาจากสัดส่วนของผู้มีงานทำและไม่มีงานทำมีความแตกต่างกันเกินไป ผู้วิจัยจึงได้ดำเนินการสุ่มข้อมูลขึ้นใหม่ดังกล่าว จากนั้นผู้วิจัยได้นำข้อมูลชุดนี้มาวิเคราะห์เพื่อหาตัวแบบและนำมาเปรียบเทียบกับตัวแบบ (4.2) โดยผลการวิเคราะห์พบว่าตัวแบบที่ได้นั้นไม่ผ่านข้อสมมติเบื้องต้นเช่นเดียวกันกับตัวแบบ (4.2) และยังให้ค่า AIC ที่สูงกว่าตัวแบบ (4.2) อีกด้วย ผู้วิจัยจึงตัดสินใจใช้ตัวแบบ (4.2) มาใช้ทำนายโอกาสในการมีงานทำของประชากรไทย

จากการวิเคราะห์การถดถอยลอจิสติกเพื่อทำนายโอกาสในการมีงานทำของประชากรไทยพบว่า ภูมิภาค (r), เพศ (s), สถานภาพการสมรส (m), ระดับการศึกษา (e), อายุ (x_1) และจำนวนสมาชิกในครัวเรือน (x_2) เป็นปัจจัยที่ส่งผลต่อโอกาสในการจ้างงานของประชากรไทย (y_1) รายละเอียดค่าประมาณพารามิเตอร์ต่าง ๆ ของตัวแบบแสดงได้ดังตารางที่ 4.5

ตารางที่ 4.5 ค่าประมาณพารามิเตอร์จากการวิเคราะห์การถดถอยลอจิสติกจากตัวแบบ (4.2)

ตัวแปร	B	p-value	Odds ratio Exp(B)	95% C.I.For Exp(B)	
				Lower	Upper
ภูมิภาค (r)					
• กรุงเทพมหานคร	(ref)	.000	-	-	-
• ภาคกลาง	.208	.000	1.231	1.170	1.296
• ภาคเหนือ	.181	.000	1.198	1.137	1.263
• ภาคตะวันออกเฉียงเหนือ	.140	.000	1.151	1.093	1.212
• ภาคใต้	.172	.000	1.187	1.125	1.253
เพศ (s)					
• หญิง	(ref)	.000	-	-	-
• ชาย	.782	.000	2.187	2.138	2.237
ระดับการศึกษา (e)					
• ไม่ได้รับการศึกษา	(ref)	.000	-	-	-
• ประถมศึกษาหรือต่ำกว่า	.325	.000	1.385	1.314	1.459
• มัธยมศึกษา	-.002	.933	.998	.944	1.054
• ปริญญาตรีหรือเทียบเท่า	1.043	.000	2.839	2.671	3.016
• สูงกว่าปริญญาตรี	1.508	.000	4.516	3.999	5.099
สถานภาพการสมรส (m)					
• โสด	(ref)	.000	-	-	-
• สมรส	1.766	.000	5.848	5.657	6.046
• ม่าย	.671	.000	1.956	1.856	2.061
• หย่าร้าง	1.820	.000	6.173	5.814	6.554
อายุ (x_1)	-.043	.000	.958	.957	.959
จำนวนสมาชิกในครัวเรือน (x_2)	-.103	.000	.902	.896	.907
Constant	1.010	.000	2.746	-	-

หมายเหตุ : ตัวแปรที่ใช้ในการอ้างอิง คือ ภูมิภาค (กรุงเทพมหานคร), เพศ (หญิง), สถานภาพการสมรส (โสด) และระดับการศึกษา (ไม่ได้รับการศึกษา)

จากตารางที่ 4.5 สามารถสรุปผลจากค่า odds ratio ได้ดังนี้

- Odds ratio(r) เมื่อเปรียบเทียบโอกาสในการจ้างงานของผู้ที่ทำงานในกรุงเทพมหานครกับผู้ที่ทำงานในภูมิภาคอื่น ๆ พบว่าผู้ที่ทำงานในภาคกลาง (Odds ratio = 1.231), ภาคเหนือ (Odds ratio = 1.198), ภาคตะวันออกเฉียงเหนือ (Odds ratio = 1.151) และภาคใต้ (Odds ratio = 1.187) มีโอกาสที่จะได้งานทำสูงกว่าคนที่อาศัยอยู่ในกรุงเทพมหานคร
- Odds ratio(s) เมื่อเปรียบเทียบโอกาสในการจ้างงานของเพศหญิงกับเพศชาย พบว่าเพศชายมีโอกาสที่จะได้งานทำสูงกว่าเพศหญิงมากกว่า 2 เท่า (Odds ratio = 2.187)
- Odds ratio(m) เมื่อเปรียบเทียบโอกาสในการจ้างงานของคนโสดกับคนกลุ่มอื่น พบว่าคนโสดมีโอกาสการจ้างงานน้อยที่สุดเมื่อเทียบกับคนอื่น ๆ ขณะที่ผู้ที่ผ่านการหย่าร้าง (Odds ratio = 6.173) และผู้ที่สมรสแล้ว (Odds ratio = 5.848) มีโอกาสการจ้างงานสูงมากกว่าคนกลุ่มอื่น ๆ ตามด้วยกลุ่มที่เป็นม่าย (Odds ratio = 1.956) ตามลำดับ
- Odds ratio(e) เมื่อเปรียบเทียบโอกาสในการจ้างงานตามระดับการศึกษา พบว่าผู้ที่ไม่ได้รับการศึกษามีโอกาสการจ้างงานไม่ต่างจากผู้ที่จบชั้นมัธยม (Odds ratio = .998) โดยผู้ที่มีโอกาสการจ้างงานมากที่สุดคือ ผู้ที่มีระดับการศึกษาสูงกว่าปริญญาตรี (Odds ratio = 4.516) รองลงมาคือ ผู้ที่จบปริญญาตรีหรือเทียบเท่า (Odds ratio = 2.839) และผู้ที่จบประถมศึกษาหรือต่ำกว่า (Odds ratio = 1.385) ตามลำดับ
- Odds ratio(x_1) = .958 หมายความว่า ถ้าหากมีอายุเพิ่มขึ้น 1 ปี (โดยตัวแปรอื่นคงที่) โอกาสการจ้างงานจะลดลงร้อยละ 4.2
- Odds ratio(x_2) = .902 หมายความว่า ถ้าจำนวนสมาชิกในครัวเรือนเพิ่มขึ้น 1 คน (โดยตัวแปรอื่นคงที่) โอกาสการจ้างงานจะลดลงร้อยละ 9.8

ตารางที่ 4.6 การวิเคราะห์การจำแนกกลุ่มของตัวแบบ (4.2)

ตัวแบบ	Observed		Predicted		
			สถานะการจ้างงาน		Percentage Correct
			ไม่มีงานทำ	มีงานทำ	
(4.2)	สถานะการจ้างงาน	ไม่มีงานทำ	22,491	37,239	37.7
		มีงานทำ	11,948	100,563	89.4
	Overall Percentage				**71.4**

หมายเหตุ : $p(y_1 = 1) \geq 0.05$ หมายความว่า มีงานทำ

$p(y_1 = 1) < 0.05$ หมายความว่า ไม่มีงานทำ

จากตารางที่ 4.6 การวิเคราะห์การจำแนกกลุ่ม พบว่าตัวแบบ (4.2) สามารถทำนายผู้มีงานทำได้ถูกต้อง 89.4% หรือ 100,563 คน จาก 112,511 คน และสามารถทำนายผู้ไม่มีงานทำได้ถูกต้อง 37.7% หรือ 22,491 คน จาก 59,730 คน โดยภาพรวมแล้วตัวแบบสามารถทำนายได้ถูกต้อง 71.4%

4.3 การวิเคราะห์การถดถอยเชิงเส้นพหุคูณของรายได้ของประชากรไทย

4.3.1 ผลการวิเคราะห์ข้อมูลเบื้องต้น

ผลการวิเคราะห์ข้อมูลเบื้องต้นโดยใช้ความถี่ และร้อยละของเพศ สถานภาพการสมรส ระดับการศึกษา อาชีพ และตำแหน่งงาน รวมทั้งหาค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ค่าต่ำสุดและค่าสูงสุดของอายุ จำนวนสมาชิกในครัวเรือน และจำนวนชั่วโมงในการทำงาน นำเสนอด้วยตารางแจกแจงความถี่ ดังต่อไปนี้

ตารางที่ 4.7 ผลการวิเคราะห์ข้อมูลเบื้องต้นของผู้มีรายได้จำนวน 47,933 คน

ตัวแปร	ความถี่ (คน)	ร้อยละ
ส่วนที่ 1 ผลการวิเคราะห์ข้อมูลเบื้องต้นของตัวแปรเชิงคุณภาพ		
เพศ (s)		
• ชาย	25,811	53.8
• หญิง	22,122	46.2
สถานภาพการสมรส (m)		
• โสด	12,391	25.9
• สมรส	30,766	64.2
• ม่าย	1,489	3.1
• หย่าร้าง	3,287	6.9
ระดับการศึกษา (e)		
• ประถมศึกษาหรือต่ำกว่า	14,796	30.9
• มัธยมศึกษา	15,760	32.9
• ปริญญาตรีหรือเทียบเท่า	13,281	27.7
• สูงกว่าปริญญาตรี	1,909	4.0
• ไม่ได้รับการศึกษา	2,187	4.6

ตารางที่ 4.7 ผลการวิเคราะห์ข้อมูลเบื้องต้นของผู้มีรายได้จำนวน 47,933 คน (ต่อ)

ตัวแปร	ความถี่ (คน)	ร้อยละ
ส่วนที่ 1 ผลการวิเคราะห์ข้อมูลเบื้องต้นของตัวแปรเชิงคุณภาพ		
อาชีพ (o)		
• ข้าราชการ เสมียนและช่างเทคนิค	15,117	31.5
• พนักงานบริการ และผู้จำหน่ายผลิตภัณฑ์	6,923	14.4
• เกษตรและประมง	2,942	6.1
• ช่างฝีมือ และผู้ควบคุมเครื่องจักร	22,951	47.9
ตำแหน่งงาน (t)		
• ลูกจ้างรัฐบาล	12,326	25.7
• ลูกจ้างเอกชน	32,348	67.5
• ลูกจ้างทั่วไป	3,259	6.8
รวม	47,933	100.0

ส่วนที่ 2 ผลการวิเคราะห์ข้อมูลเบื้องต้นของตัวแปรเชิงปริมาณ

ตัวแปร	ค่าเฉลี่ย	SD	ค่าต่ำสุด	ค่าสูงสุด
รายได้เฉลี่ย/เดือน (y_2)	13,616.17	11,531.472	200	250,000
อายุ (x_1)	39.98	12.016	15	84
จำนวนสมาชิกในครัวเรือน (x_2)	3.53	1.915	1	18
จำนวนชั่วโมงในการทำงาน (x_3)	43.88	9.433	2	98

จากตารางที่ 4.7 ผลการวิเคราะห์ข้อมูลเบื้องต้นของผู้มีรายได้ต่อเดือนจำนวน 47,933 คน พบว่ามีรายได้เฉลี่ยต่อเดือนประมาณ 13,616.17 บาท อายุเฉลี่ยของคนกลุ่มนี้อยู่ที่ 39.98 ปี โดยแบ่งเป็นเพศชายจำนวน 25,811 คน (ร้อยละ 53.8) และเพศหญิงจำนวน 22,122 คน (ร้อยละ 46.2) เป็นผู้ที่มีสถานภาพโสดจำนวน 12,391 คน (ร้อยละ 25.9) ผู้ที่มีสถานภาพสมรสจำนวน 30,766 คน (ร้อยละ 64.2) ผู้ที่มีสถานภาพเป็นม่ายจำนวน 1,489 คน (ร้อยละ 3.1) และเป็นผู้ที่มีสถานภาพหย่าร้างจำนวน 3,287 คน (ร้อยละ 6.9) โดยมีจำนวนสมาชิกในครัวเรือนเฉลี่ยอยู่ที่ 3.53 คน

จำแนกตามระดับการศึกษา พบว่าเป็นผู้ที่จบชั้นประถมศึกษาหรือต่ำกว่าจำนวน 14,796 คน (ร้อยละ 30.9) ผู้ที่จบมัธยมศึกษาจำนวน 15,760 คน (ร้อยละ 32.9) ผู้ที่จบปริญญาตรีหรือเทียบเท่าจำนวน 13,281 คน (ร้อยละ 27.7) ผู้ที่จบสูงกว่าปริญญาตรีจำนวน 1,909 คน (ร้อยละ 4.0) และผู้ที่ไม่ได้รับการศึกษาจำนวน 2,187 คน (ร้อยละ 4.6)

จำแนกตามการประกอบอาชีพ พบว่าประกอบอาชีพเป็นข้าราชการ เสมียน และช่างเทคนิคจำนวน 15,117 คน (ร้อยละ 31.5) ประกอบอาชีพเป็นพนักงานบริการ และผู้จำหน่ายผลิตภัณฑ์จำนวน 6,923 คน (ร้อยละ 14.4) ประกอบอาชีพเกษตร และประมงจำนวน 2,942 คน

(ร้อยละ 6.1) และประกอบอาชีพเป็นช่างฝีมือ และผู้ควบคุมเครื่องจักรจำนวน 22,951 คน (ร้อยละ 47.9) แบ่งเป็นเป็นลูกจ้างรัฐบาลจำนวน 12,326 คน (ร้อยละ 25.7) ลูกจ้างเอกชนจำนวน 32,348 คน (ร้อยละ 67.5) และลูกจ้างทั่วไปจำนวน 3,259 คน (ร้อยละ 6.8) โดยมีจำนวนชั่วโมงในการทำงานเฉลี่ยอยู่ที่ 43.88 ชั่วโมง/สัปดาห์

4.3.2 การวิเคราะห์การถดถอยเชิงเส้นพหุคูณของรายได้ของประชากรไทย

จากการนำข้อมูลจากโครงการสำรวจภาวะการทำงานของประชากรไทย ในปี พ.ศ. 2561 ของสำนักงานสถิติแห่งชาติ จำนวน 47,933 คน เพื่อนำมาวิเคราะห์การถดถอยเชิงเส้นพหุคูณของรายได้ประชากรไทย สามารถเขียนตัวแบบของรายได้ต่อเดือน ได้ดังนี้

$$\begin{aligned} Y_2 &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 s_1 + \beta_5 m_1 + \beta_6 m_2 + \beta_7 m_3 + \beta_8 e_1 \\ &\quad + \beta_9 e_2 + \beta_{10} e_3 + \beta_{11} e_4 + \beta_{12} o_1 + \beta_{13} o_2 + \beta_{14} o_3 + \beta_{15} t_1 + \beta_{16} t_2 \end{aligned} \qquad \textbf{(4.3)}$$

โดยที่ Y_2 คือ รายได้ต่อเดือนของประชากรไทย

$\beta_0, \beta_1, \beta_2, \ldots, \beta_{16}$ คือ สัมประสิทธิ์การถดถอยซึ่งเป็นพารามิเตอร์ของตัวแบบ

x_1 คือ อายุของประชากรไทย

x_2 คือ จำนวนสมาชิกในครัวเรือนของประชากรไทย

x_3 คือ จำนวนชั่วโมงในการทำงานของประชากรไทย

s_1 คือ เพศหญิง

m_1 คือ สถานภาพสมรส

m_2 คือ สถานภาพเป็นม่าย

m_3 คือ สถานภาพหย่าร้าง

e_1 คือ มัธยมศึกษา

e_2 คือ ปริญญาตรีหรือเทียบเท่า

e_3 คือ สูงกว่าปริญญาตรี

e_4 คือ ไม่ได้รับการศึกษา

o_1 คือ ประกอบอาชีพเป็นพนักงานบริการ และผู้จำหน่ายผลิตภัณฑ์

o_2 คือ ประกอบอาชีพเกษตรและประมง

o_3 คือ ประกอบอาชีพช่างฝีมือและผู้ควบคุมเครื่องจักร

t_1 คือ ลูกจ้างเอกชน

t_2 คือ ลูกจ้างทั่วไป

สร้างตัวแบบการถดถอยเชิงเส้นพหุคูณเพื่อทำนายรายได้ต่อเดือนของประชากรไทย พร้อมทั้งตรวจสอบตัวแบบการประมาณ โดยทำการตรวจสอบข้อสมมติเบื้องต้นเกี่ยวกับค่าคลาดเคลื่อนสุ่ม (Random Error) และแสดงค่าสัมประสิทธิ์การตัดสินใจพหุคูณ (R^2) ค่าความคลาดเคลื่อนมาตรฐานของส่วนเหลือ (Residual Standard Error) ค่าสถิติ AIC และค่าสถิติ BIC โดยสามารถเขียนตัวแบบการประมาณ ได้ดังนี้

$$\begin{aligned} y_2 = {} & 2,805.27 + 207.86x_1 - 343.27x_2 + 83.65x_3 - 2,242.52s_1 \\ & + 1,156.91m_1 - 1,065.43m_2 - 323.71m_3 + 3,316.59e_1 \\ & + 10,391.48e_2 + 23,760.43e_3 + 1,044.25e_4 - 4,746.17o_1 \\ & - 7,562.85o_2 - 5,578.16o_3 - 738.32t_1 - 3,720.12t_2 \end{aligned} \qquad \textbf{(4.4)}$$

จากตัวแบบ (4.4) สามารถเขียนค่าสัมประสิทธิ์การตัดสินใจ (R^2) ที่ได้จากตัวแบบและคำนวณค่าความคลาดเคลื่อนมาตรฐานของส่วนเหลือ (Residual Standard Error) ค่าสถิติ AIC และค่าสถิติ BIC ใส่ลงในตารางได้ดังนี้

ตารางที่ 4.8 ค่าสัมประสิทธิ์การตัดสินใจ (R^2) ค่าความคลาดเคลื่อนมาตรฐานของส่วนเหลือ (Residual Standard Error) ค่าสถิติ AIC และค่าสถิติ BIC ของตัวแบบ (4.4)

Model	R^2	RMSE	AIC	BIC
y_2	0.495	8,193.132	999,917.4	1,000,075

อภิปรายผลสมการที่ (4.4) และตารางที่ 4.8 ผลการวิเคราะห์การถดถอยเชิงเส้นพหุคูณของรายได้ประชากรไทย พบว่าเพศ (s), สถานภาพการสมรส (m), ระดับการศึกษา (e), อาชีพ (o), ตำแหน่งงาน (t), อายุ (x_1), จำนวนสมาชิกในครัวเรือน (x_2) และจำนวนชั่วโมงในการทำงาน (x_3) เป็นปัจจัยที่ส่งผลต่อรายได้เฉลี่ยต่อเดือนของประชากรไทย (y_2) โดยมีค่าสัมประสิทธิ์การตัดสินใจ (R^2) เท่ากับ 0.495 หมายความว่า ตัวแปรอิสระเหล่านี้สามารถอธิบายความผันแปรของรายได้ประชากรไทย 49.5% โดยมีค่าความคลาดเคลื่อนมาตรฐานของส่วนเหลือ (Residual Standard Error) เท่ากับ 8,193.132 ค่าสถิติ AIC เท่ากับ 999,917.4 และค่าสถิติ BIC เท่ากับ 1,000,075

การตรวจสอบข้อสมมติเบื้องต้นของตัวแบบ (4.4)

- ตรวจสอบการแจกแจงปรกติของค่าคลาดเคลื่อนสุ่มของตัวแบบการถดถอย โดยใช้การทดสอบ Kolmogorov-Smirnov test

สมมติฐานการทดสอบ คือ

$\mathbf{H_0}$: ค่าคลาดเคลื่อนสุ่มของตัวแบบการถดถอยประมาณมีการแจกแจงปรกติ

$\mathbf{H_1}$: ค่าคลาดเคลื่อนสุ่มของตัวแบบการถดถอยประมาณมีการแจกแจงแบบอื่น

ตารางที่ 4.9 ผลการวิเคราะห์การตรวจสอบการแจกแจงปรกติของค่าคลาดเคลื่อนสุ่มโดย Kolmogorov-Smirnov test ของตัวแบบ (4.4)

ค่าสถิติ	p-value
0.76968	0.000

- ตรวจสอบความแปรปรวนของค่าคลาดเคลื่อนสุ่มของตัวแบบการถดถอย โดยใช้การทดสอบ Bartlett test of homogeneity of variances

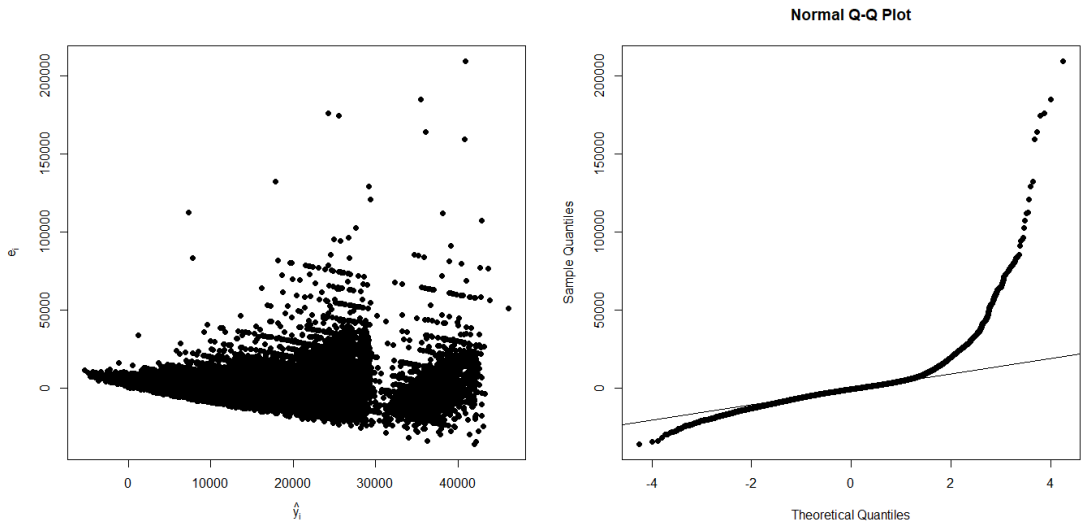
สมมติฐานการทดสอบ คือ

$\mathbf{H_0}$: ค่าคลาดเคลื่อนสุ่มของตัวแบบการถดถอยประมาณมีความแปรปรวนคงที่

$\mathbf{H_1}$: ค่าคลาดเคลื่อนสุ่มของตัวแบบการถดถอยประมาณมีความแปรปรวนไม่คงที่

ตารางที่ 4.10 ผลการวิเคราะห์การตรวจสอบการความแปรปรวนของค่าคลาดเคลื่อนสุ่มโดย Bartlett test of homogeneity of variances ของตัวแบบ (4.4)

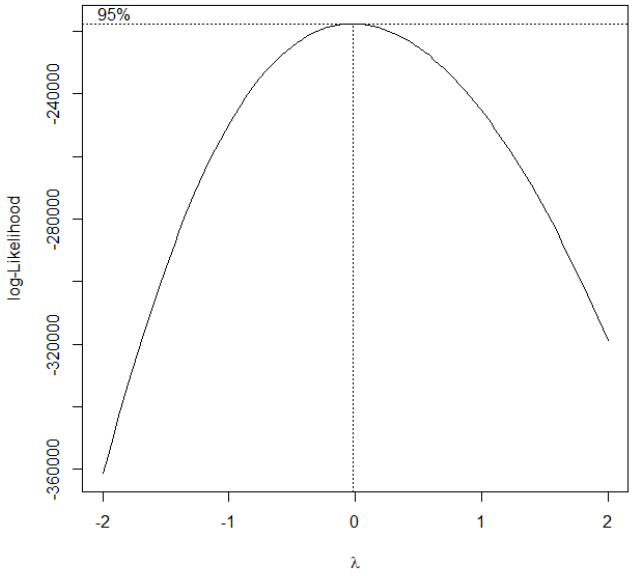
ค่าสถิติ Bartlett's K-squared	p-value
34,029	0.000



รูป 4.3 Normal Q-Q plot และแผนภาพการกระจายของส่วนเหลือ (Residuals) ของตัวแบบ (4.4)

อภิปรายผลจากข้อสมมติเบื้องต้นของตัวแบบการถดถอยเชิงเส้น ค่าคลาดเคลื่อนสุ่มจะต้องมีการแจกแจงปรกติและมีความแปรปรวนคงที่ โดยการทดสอบการแจกแจงปรกตินั้นจะพิจารณาจากแผนภาพ Normal Q-Q plot ของส่วนเหลือ และ Kolmogorov-Smirnov test พบว่า ค่าคลาดเคลื่อนสุ่มจากตัวแบบการถดถอยประมาณ ไม่กระจายตัวรอบเส้นปรกติโดยเฉพาะหางและปลายของกราฟมีลักษณะกระจายตัวออกห่างจากเส้นปรกติมาก และจากการทดสอบโดยใช้สถิติทดสอบ Kolmogorov-Smirnov test ให้ค่าสถิติ D = 0.76968, ค่า p-value = 0.000 ซึ่งมีค่าน้อย

กว่าระดับนัยสำคัญ 0.05 จึงสรุปได้ว่าค่าคลาดเคลื่อนสุ่มมีการแจกแจงแบบอื่นที่ไม่ใช่การแจกแจงแบบปรกติ ส่วนการตรวจความแปรปรวนของค่าคลาดเคลื่อนสุ่ม พิจารณาจาก Bartlett test of homogeneity of variances ให้ค่าสถิติ Bartlett's K-squared = 34029, ค่า p-value = 0.000 ซึ่งมีค่าน้อยกว่าระดับนัยสำคัญ 0.05 จึงสรุปได้ว่าค่าคลาดเคลื่อนสุ่มมีความแปรปรวนไม่คงที่ จากผลข้างต้นพบว่าตัวแบบ (4.4) ไม่ผ่านข้อตกลงเบื้องต้นของตัวแบบการถดถอยเชิงเส้น ผู้วิจัยจึงเลือกวิธีแก้ปัญหาด้วยการแปลงข้อมูลรายได้เฉลี่ยต่อเดือนของประชากรไทย (y_2) ด้วยบอกซ์-คอกซ์ (Box-Cox transformation) โดยแสดงดังรูป 4.4



รูป 4.4 การแปลงข้อมูลด้วยบอกซ์-คอกซ์ (Box-Cox transformation) ของตัวแบบ (4.4)

อภิปรายผลจากรูป 4.4 การแปลงข้อมูลด้วยบอกซ์-คอกซ์ (Box-Cox transformation) พบว่าค่าประมาณของพารามิเตอร์การแปลง (λ) ที่ทำให้ค่า log likelihood ของการประมาณค่ามีค่ามากที่สุด มีค่าเข้าใกล้ 0 ดังนั้นจึงทำการแปลงรายได้เฉลี่ยต่อเดือนของประชากรไทย (y_2) ให้เป็น $\ln(y_2)$ สามารถเขียนตัวแบบการถดถอยหลังจากทำการแปลงค่า y_2 ให้เป็น $\ln(y_2)$ ได้ดังนี้

$$\begin{aligned}\ln(y_2) = {} & 8.492 + 0.008x_1 - 0.021x_2 + 0.015x_3 - 0.148s_1 + 0.076m_1 \\ & -0.098m_2 + 0.012m_3 + 0.213e_1 + 0.648e_2 + 1.155e_3 - 0.012e_4 \\ & -0.27o_1 - 0.622o_2 - 0.329o_3 - 0.154t_1 - 0.498t_2 \end{aligned} \quad \textbf{(4.5)}$$

จากตัวแบบ (4.5) สามารถแสดงค่าสัมประสิทธิ์การตัดสินใจ (R^2) ที่ได้จากตัวแบบ และคำนวณค่าความคลาดเคลื่อนมาตรฐานของส่วนเหลือ (Residual Standard Error) ค่าสถิติ AIC และค่าสถิติ BIC ใส่ลงในตารางได้ดังนี้

ตารางที่ 4.11 ค่าสัมประสิทธิ์การตัดสินใจ (R^2) ค่าความคลาดเคลื่อนมาตรฐานของส่วนเหลือ (Residual Standard Error) ค่าสถิติ AIC และค่าสถิติ BIC ของตัวแบบ (4.5)

Model	R^2	RMSE	AIC	BIC
$\ln(y_2)$	0.5876	0.4275	54,606.44	54,764.44

อภิปรายผลตัวแบบ (4.5) และตารางที่ 4.11 จากการวิเคราะห์การถดถอยเชิงเส้นพหุคูณของรายได้ประชากรไทย พบว่าเพศ (s), สถานภาพการสมรส (m), ระดับการศึกษา (e), อาชีพ (o), ตำแหน่งงาน (t), อายุ (x_1), จำนวนสมาชิกในครัวเรือน (x_2) และจำนวนชั่วโมงในการทำงาน (x_3) เป็นปัจจัยที่ส่งผลต่อรายได้เฉลี่ยต่อเดือนของประชากรไทย (y_2) โดยมีค่าสัมประสิทธิ์การตัดสินใจ (R^2) เท่ากับ 0.5876 หมายความว่า ตัวแปรอิสระสามารถอธิบายความผันแปรของรายได้ประชากรไทย 58.76% โดยมีค่าความคลาดเคลื่อนมาตรฐานของส่วนเหลือ (Residual Standard Error) เท่ากับ 0.4275 ค่าสถิติ AIC เท่ากับ 54,606.44 และค่าสถิติ BIC เท่ากับ 54,764.44

การตรวจสอบข้อสมมติเบื้องต้นของตัวแบบ (4.5)

- ตรวจสอบการแจกแจงปรกติของค่าคลาดเคลื่อนสุ่มของตัวแบบการถดถอย โดยใช้การทดสอบ Kolmogorov-Smirnov test

สมมติฐานการทดสอบ คือ

$\mathbf{H_0}$: ค่าคลาดเคลื่อนสุ่มของตัวแบบการถดถอยประมาณมีการแจกแจงปรกติ

$\mathbf{H_1}$: ค่าคลาดเคลื่อนสุ่มของตัวแบบการถดถอยประมาณมีการแจกแจงแบบอื่น

ตารางที่ 4.12 ผลการวิเคราะห์การตรวจสอบการแจกแจงปรกติของค่าคลาดเคลื่อนสุ่มโดย Kolmogorov-Smirnov test ของตัวแบบ (4.5)

ค่าสถิติ	p-value
1	0.000

- ตรวจสอบความแปรปรวนของค่าคลาดเคลื่อนสุ่มของตัวแบบการถดถอย โดยใช้การทดสอบ Bartlett test of homogeneity of variances

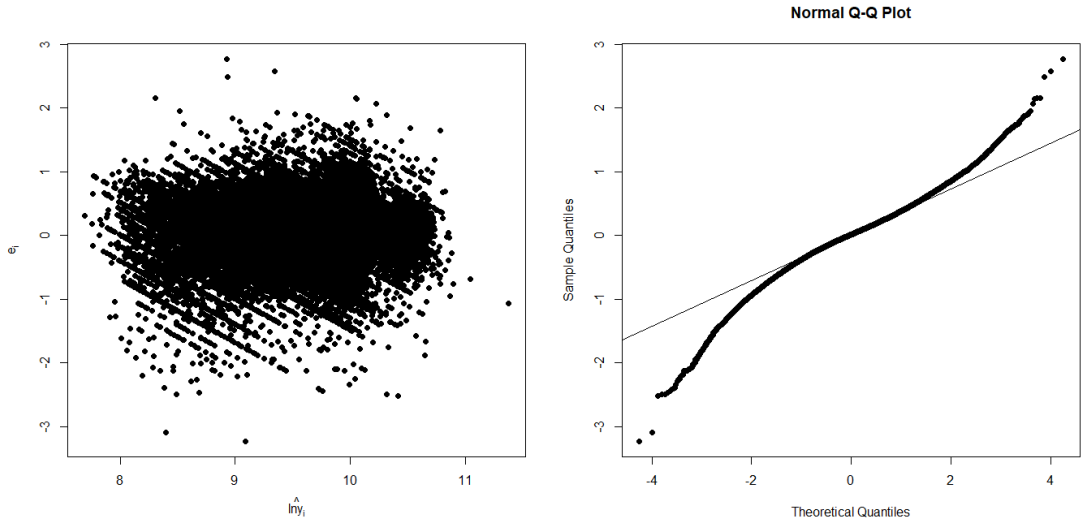
สมมติฐานการทดสอบ คือ

$\mathbf{H_0}$: ค่าคลาดเคลื่อนสุ่มของตัวแบบการถดถอยประมาณมีความแปรปรวนคงที่

$\mathbf{H_1}$: ค่าคลาดเคลื่อนสุ่มของตัวแบบการถดถอยประมาณมีความแปรปรวนไม่คงที่

ตารางที่ 4.13 ผลการวิเคราะห์การตรวจสอบการความแปรปรวนของค่าคลาดเคลื่อนสุ่มโดย Bartlett test of homogeneity of variances ของตัวแบบ (4.5)

ค่าสถิติ Bartlett's K-squared	p-value
1,884	0.000



รูป 4.5 Normal Q-Q plot และแผนภาพการกระจายของส่วนเหลือ (Residuals) ของตัวแบบ (4.5)

อภิปรายผล จากข้อสมมติเบื้องต้นของตัวแบบการถดถอยเชิงเส้น ค่าคลาดเคลื่อนสุ่มจะต้องมีการแจกแจงปรกติและมีความแปรปรวนคงที่ โดยการทดสอบการแจกแจงปรกตินั้นจะพิจารณาจากแผนภาพ Normal Q-Q plot ของส่วนเหลือ และ Kolmogorov-Smirnov test พบว่า ค่าคลาดเคลื่อนสุ่มจากตัวแบบการถดถอยประมาณ ไม่กระจายตัวรอบเส้นปรกติโดยเฉพาะหางและปลายของกราฟมีลักษณะกระจายตัวออกห่างจากเส้นปรกติ และจากการทดสอบโดยใช้สถิติทดสอบ Kolmogorov-Smirnov test ให้ค่าสถิติ D = 1, ค่า p-value = 0.000 ซึ่งมีค่าน้อยกว่าระดับนัยสำคัญ 0.05 จึงสรุปได้ว่าค่าคลาดเคลื่อนสุ่มมีการแจกแจงแบบอื่นที่ไม่ใช่การแจกแจงแบบปรกติ ส่วนการตรวจความแปรปรวนของค่าคลาดเคลื่อนสุ่ม พิจารณาจาก Bartlett test of homogeneity of variances ให้ค่าสถิติ Bartlett's K-squared = 1,884 , ค่า p-value = 0.000 ซึ่งมีค่าน้อยกว่าระดับนัยสำคัญ 0.05 จึงสรุปได้ว่าค่าคลาดเคลื่อนเชิงมีความแปรปรวนไม่คงที่

จากการตรวจสอบข้อสมมติเบื้องต้นของตัวแบบ $\ln(y_2)$ พบว่าได้ผลลัพธ์ไม่ต่างจากตัวแบบ y_2 โดยทั้งสองตัวแบบไม่ผ่านข้อสมมติเบื้องต้นของตัวแบบการถดถอยเชิงเส้น แม้ว่าตัวแบบ $\ln(y_2)$ จะใช้บอกซ์-คอกซ์ (Box-Cox transformation) เพื่อแก้ปัญหาแล้ว อีกทั้งยังได้ทำการคัดเลือกตัวแปรโดยใช้วิธีการถดถอยทีละขั้น (stepwise regression) แต่ยังคงได้ผลลัพธ์ดังเดิม เนื่องจากข้อมูลรายได้เฉลี่ยต่อเดือนของประชากรไทย (y_2) มีค่านอกเกณฑ์ (outliers) เป็นจำนวนมากจึงส่งผลให้ตัวแบบการประมาณทั้ง 2 ตัวแบบไม่ผ่านข้อสมมติเบื้องต้นของการถดถอยเชิงเส้น

ผู้วิจัยจึงใช้วิธีการจัดการข้อมูลโดยการเลือกข้อมูลรายได้เฉลี่ยต่อเดือนของประชากรไทย (y_2) ที่ผู้วิจัยคาดว่าเป็นช่วงรายได้ที่เหมาะสมกับประชากรไทย โดยใช้เกณฑ์การเลือกจากการทบทวนวรรณกรรม ซึ่งวรรณวิมล จงจรวยสกุล (2555) พบว่าผู้คนส่วนใหญ่จะมีเงินเดือนอยู่ที่ 10,001-12,000 บาท ขณะที่พรทวี เถื่อนคำแสน และคณะ (2559) พบว่าส่วนใหญ่มีรายได้เฉลี่ยต่อเดือนอยู่ที่ 15,001–25,000 บาท จากผลสรุปข้างต้น ผู้วิจัยจึงเลือกช่วงของรายได้เฉลี่ยต่อเดือนที่เหมาะสมกับประชากรไทย คือ 10,001-25,000 บาท เพื่อตัดค่านอกเกณฑ์ (outliers) ออกไป โดยจะนำข้อมูลชุดนี้มาวิเคราะห์เพื่อหาตัวแบบการประมาณและนำมาเปรียบเทียบตัวแบบ (4.4) และ (4.5)

จากการสร้างตัวแบบการถดถอยเชิงเส้นพหุคูณเพื่อทำนายรายได้ต่อเดือนของประชากรไทยโดยใช้ข้อมูลที่ตัดค่านอกเกณฑ์ (outliers) ออกไป และทำการคัดเลือกตัวแปรโดยใช้วิธีการถดถอยทีละขั้น (stepwise regression) อีกทั้งยังผ่านการแปลงข้อมูลด้วยบอกซ์-คอกซ์ (Box-Cox transformation) โดยค่าประมาณของพารามิเตอร์การแปลง (λ) ที่ทำให้ค่า log likelihood ของการประมาณค่ามีค่ามากที่สุด มีค่าเข้าใกล้ -1 จึงทำการแปลงรายได้เฉลี่ยต่อเดือนของประชากรไทย (y_2) ให้เป็น $1/y_2$ โดยสามารถเขียนตัวแบบการประมาณ ได้ดังนี้

$$
\begin{aligned}
1/y_2 = {} & 11{,}105.07 + 82.82x_1 - 106.43x_2 + 10.62x_3 - 711.09s_1 \\
& + 247.01m_1 - 217.52m_2 - 427.53m_3 + 1{,}354.01e_1 \\
& + 3{,}617.75e_2 + 7{,}271.54e_3 + 566.95e_4 - 1{,}879.77o_1 \\
& - 2{,}569.87o_2 - 2{,}042.73o_3 - 686.22t_1 - 1{,}872.03t_2
\end{aligned}
\qquad \textbf{(4.6)}
$$

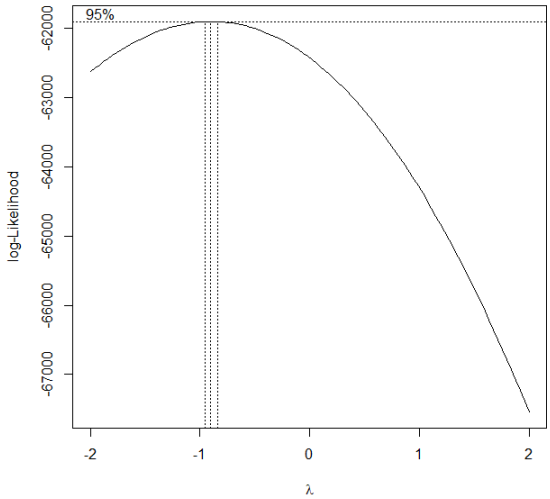
จากตัวแบบ (4.6) สามารถเขียนค่าสัมประสิทธิ์การตัดสินใจ (R^2) ที่ได้จากตัวแบบและคำนวณค่าความคลาดเคลื่อนมาตรฐานของส่วนเหลือ (Residual Standard Error) ค่าสถิติ AIC และค่าสถิติ BIC ใส่ลงในตารางได้ดังนี้

ตารางที่ 4.14 ค่าสัมประสิทธิ์การตัดสินใจ (R^2) ค่าความคลาดเคลื่อนมาตรฐานของส่วนเหลือ (Residual Standard Error) ค่าสถิติ AIC และค่าสถิติ BIC ของตัวแบบ (4.6)

Model	R^2	RMSE	AIC	BIC
$1/y_2$	0.3112	3,572.942	350,477.4	350,618.1

อภิปรายผลตัวแบบ (4.6) และตารางที่ 4.14 จากการวิเคราะห์การถดถอยเชิงเส้นพหุคูณของรายได้ประชากรไทย พบว่าเพศ (s), สถานภาพการสมรส (m), ระดับการศึกษา (e), อาชีพ (o), ตำแหน่งงาน (t), อายุ (x_1), จำนวนสมาชิกในครัวเรือน (x_2) และจำนวนชั่วโมงในการทำงาน (x_3) เป็นปัจจัยที่ส่งผลต่อรายได้เฉลี่ยต่อเดือนของประชากรไทย (y_2) โดยมีค่าสัมประสิทธิ์

การตัดสินใจ (R^2) เท่ากับ 0.3112 หมายความว่า ตัวแปรอิสระสามารถอธิบายความผันแปรของรายได้ประชากรไทย 31.12% โดยมีค่าความคลาดเคลื่อนมาตรฐานของส่วนเหลือ (Residual Standard Error) เท่ากับ 3,572.942 ค่าสถิติ AIC เท่ากับ 350,477.4 และค่าสถิติ BIC เท่ากับ 350,618.1



รูป 4.6 การแปลงข้อมูลด้วยบอกซ์-คอกซ์ (Box-Cox transformation)

การตรวจสอบข้อสมมติเบื้องต้นของตัวแบบ (4.6)

- ตรวจสอบการแจกแจงปรกติของค่าคลาดเคลื่อนสุ่มของตัวแบบการถดถอย โดยใช้การทดสอบ Kolmogorov-Smirnov test

สมมติฐานการทดสอบ คือ

$\mathbf{H_0}$: ค่าคลาดเคลื่อนสุ่มของตัวแบบการถดถอยประมาณมีการแจกแจงปรกติ

$\mathbf{H_1}$: ค่าคลาดเคลื่อนสุ่มของตัวแบบการถดถอยประมาณมีการแจกแจงแบบอื่น

ตารางที่ 4.15 ผลการวิเคราะห์การตรวจสอบการแจกแจงปรกติของค่าคลาดเคลื่อนสุ่มโดย Kolmogorov-Smirnov test ของตัวแบบ (4.6)

ค่าสถิติ	p-value
0.99085	0.000

- ตรวจสอบความแปรปรวนของค่าคลาดเคลื่อนสุ่มของตัวแบบการถดถอย โดยใช้การทดสอบ Bartlett test of homogeneity of variances

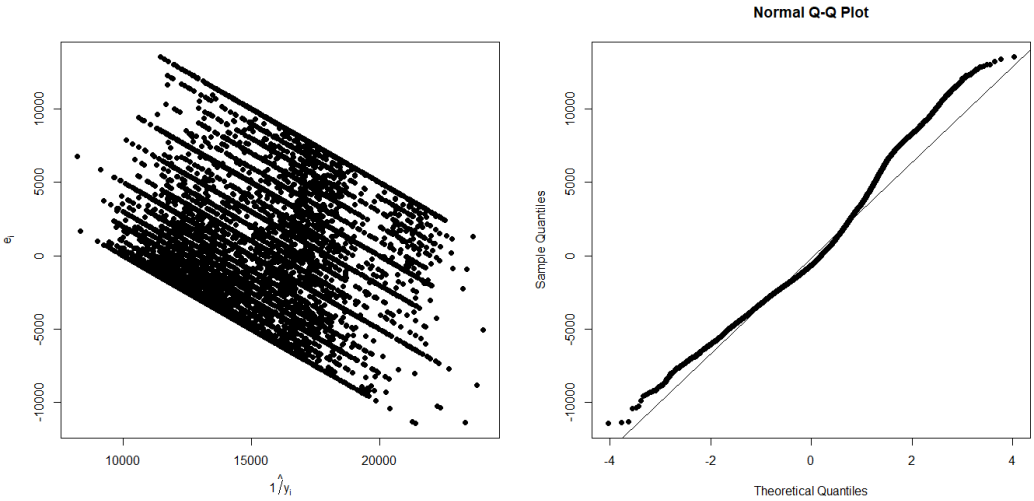
สมมติฐานการทดสอบ คือ

$\mathbf{H_0}$: ค่าคลาดเคลื่อนสุ่มของตัวแบบการถดถอยประมาณมีความแปรปรวนคงที่

$\mathbf{H_1}$: ค่าคลาดเคลื่อนสุ่มของตัวแบบการถดถอยประมาณมีความแปรปรวนไม่คงที่

ตารางที่ 4.16 ผลการวิเคราะห์การตรวจสอบการความแปรปรวนของค่าคลาดเคลื่อนสุ่มโดย Bartlett test of homogeneity of variances ของตัวแบบ (4.6)

ค่าสถิติ Bartlett's K-squared	p-value
1,475.2	0.000



รูป 4.7 Normal Q-Q plot และแผนภาพการกระจายของส่วนเหลือ (Residuals) ของตัวแบบ (4.6)

อภิปรายผล จากข้อสมมติเบื้องต้นของตัวแบบการถดถอยเชิงเส้น ค่าคลาดเคลื่อนสุ่มจะต้องมีการแจกแจงปรกติและมีความแปรปรวนคงที่ โดยการทดสอบการแจกแจงปรกตินั้นจะพิจารณาจากแผนภาพ Normal Q-Q plot ของส่วนเหลือ และ Kolmogorov-Smirnov test พบว่า ค่าคลาดเคลื่อนสุ่มจากตัวแบบการถดถอยประมาณ ไม่กระจายตัวรอบเส้นปรกติโดยเฉพาะหางและปลายของกราฟมีลักษณะกระจายตัวออกห่างจากเส้นปรกติ และจากการทดสอบโดยใช้สถิติทดสอบ Kolmogorov-Smirnov test ให้ค่าสถิติ D = 0.99085, ค่า p-value = 0.000 ซึ่งมีค่าน้อยกว่าระดับนัยสำคัญ 0.05 จึงสรุปได้ว่าค่าคลาดเคลื่อนสุ่มมีการแจกแจงแบบอื่นที่ไม่ใช่การแจกแจงแบบปรกติ ส่วนการตรวจความแปรปรวนของค่าคลาดเคลื่อนสุ่ม พิจารณาจาก Bartlett test of homogeneity of variances ให้ค่าสถิติ Bartlett's K-squared = 1,475.2, ค่า p-value = 0.000 ซึ่งมีค่าน้อยกว่าระดับนัยสำคัญ 0.05 จึงสรุปได้ว่าค่าคลาดเคลื่อนเชิงมีความแปรปรวนไม่คงที่

เปรียบเทียบตัวแบบ y_2 , $\ln(y_2)$ และ $1/y_2$ เพื่อหาตัวแบบการประมาณที่เหมาะสมกับข้อมูลที่สุดเพื่อใช้ในการทำนายรายได้เฉลี่ยต่อเดือนของประชากรไทย

ตารางที่ 4.17 เปรียบเทียบตัวแบบ y_2 , $\ln(y_2)$ และ $1/y_2$

Model	R^2	RMSE	AIC	BIC
y_2	0.4950	8,193.132	999,917.4	1,000,075
$\ln(y_2)$	0.5876	0.4275	54,606.44	54,764.44
$1/y_2$	0.3112	3,572.942	350,477.4	350,618.1

จากตารางที่ 4.17 เมื่อเปรียบเทียบตัวแบบทั้ง 3 ตัวแบบ พบว่าตัวแบบ $\ln(y_2)$ มีค่าสัมประสิทธิ์การตัดสินใจพหุคูณ (R^2) ที่สูงกว่าตัวแบบอื่น ๆ และให้ค่าความคลาดเคลื่อนมาตรฐานของส่วนเหลือ (RMSE) ค่า AIC และค่า BIC ที่น้อยกว่าตัวแบบอื่น ๆ เช่นเดียวกัน ดังนั้นผู้วิจัยจึงเลือกตัวแบบ $\ln(y_2)$ ในการทำนายรายได้เฉลี่ยต่อเดือนของประชากรไทย รายละเอียดค่าประมาณพารามิเตอร์ต่าง ๆ ของตัวแบบแสดงได้ดังตารางที่ 4.18

ตารางที่ 4.18 ค่าประมาณพารามิเตอร์จากการวิเคราะห์การถดถอยพหุคูณจากตัวแบบ (4.5)

ตัวแปร	**B**	**Std. Error**	**t**	**p-value**
Constant	8.492	.016	519.098	.000
อายุ (x_1)	.008	.000	38.173	.000
จำนวนสมาชิกในครัวเรือน (x_2)	-.021	.001	-20.133	.000
จำนวนชั่วโมงในการทำงาน (x_3)	.015	.000	63.053	.000
เพศหญิง (s_1)	-.148	.004	-36.008	.000
สถานภาพสมรส (m_1)	.076	.005	14.944	.000
สถานภาพเป็นม่าย (m_2)	-.098	.013	-7.747	.000
สถานภาพหย่าร้าง (m_3)	.012	.009	1.369	.171
มัธยมศึกษา (e_1)	.213	.005	39.036	.000
ปริญญาตรีหรือเทียบเท่า (e_2)	.648	.007	89.032	.000
สูงกว่าปริญญาตรี (e_3)	1.155	.012	94.595	.000
ไม่ได้รับการศึกษา (e_4)	-.012	.009	-1.160	.246

ตารางที่ 4.18 ค่าประมาณพารามิเตอร์จากการวิเคราะห์การถดถอยพหุคูณจากตัวแบบ (4.5) (ต่อ)

ตัวแปร	B	Std. Error	t	p-value
อาชีพพนักงานบริการ และผู้จำหน่ายผลิตภัณฑ์ (O_1)	-.270	.007	-36.979	.000
อาชีพเกษตรและประมง (O_2)	-.622	.010	-59.351	.000
อาชีพช่างฝีมือ และผู้ควบคุมเครื่องจักร (O_3)	-.329	.007	-48.581	.000
ลูกจ้างเอกชน (t_1)	-.154	.006	-26.186	.000
ลูกจ้างทั่วไป (t_2)	-.498	.009	-51.714	.000

หมายเหตุ : การแปลความหมายของความค่าสัมประสิทธิ์จะต้องเปลี่ยนค่าสัมประสิทธิ์ให้อยู่ในรูปของฟังก์ชั่นเลขชี้กำลัง (exponential) และตัวแปรหุ่นที่ใช้ในการอ้างอิง คือ เพศ (ชาย), สถานภาพการสมรส (โสด), ระดับการศึกษา (ประถมศึกษาหรือต่ำกว่า), อาชีพ (ข้าราชการ เสมียน และช่างเทคนิค), และตำแหน่งงาน (ลูกจ้างรัฐบาล)

จากตารางที่ 4.18 สามารถสรุปผลจากค่าสัมประสิทธิ์การถดถอย ได้ดังนี้

$b_1 = .008$ หมายความว่า ค่าเฉลี่ยรายได้ต่อเดือนของประชากรไทยจะ “เพิ่มขึ้น” 1.01 บาท/เดือน ถ้ามีอายุเพิ่มขึ้น 1 ปี โดยกำหนดตัวแปรอิสระที่เหลือมีค่าคงที่

$b_2 = -.021$ หมายความว่า ค่าเฉลี่ยรายได้ต่อเดือนของประชากรไทยจะ “ลดลง” 0.98 บาท/เดือน ถ้าจำนวนสมาชิกในครัวเรือนเพิ่มขึ้น 1 คน โดยกำหนดตัวแปรอิสระที่เหลือมีค่าคงที่

$b_3 = .015$ หมายความว่า ค่าเฉลี่ยรายได้ต่อเดือนของประชากรไทยจะ “เพิ่มขึ้น” 1.02 บาท/เดือน ถ้าจำนวนชั่วโมงในการทำงานเพิ่มขึ้น 1 ชั่วโมงต่อสัปดาห์ โดยกำหนดตัวแปรอิสระที่เหลือมีค่าคงที่

$b_4 = -.148$ หมายความว่า เพศหญิงจะมีค่าเฉลี่ยรายได้ต่อเดือน “น้อยกว่า” เพศชาย 0.86 บาท/เดือน โดยกำหนดตัวแปรอิสระที่เหลือมีค่าคงที่

$b_5 = .076$ หมายความว่า ผู้ที่มีสถานภาพสมรสจะมีค่าเฉลี่ยรายได้ต่อเดือน “มากกว่า” ผู้ที่มีสถานภาพโสด 1.08 บาท/เดือน โดยกำหนดตัวแปรอิสระที่เหลือมีค่าคงที่

$b_6 = -.098$ หมายความว่า ผู้ที่มีสถานภาพเป็นม่ายจะมีค่าเฉลี่ยรายได้ต่อเดือน “น้อยกว่า” ผู้ที่มีสถานภาพโสด 0.91 บาท/เดือน โดยกำหนดตัวแปรอิสระที่เหลือมีค่าคงที่

$b_7 = .012$ หมายความว่า ผู้ที่มีสถานภาพเป็นหย่าร้างจะมีค่าเฉลี่ยรายได้ต่อเดือน “มากกว่า” ผู้ที่มีสถานภาพโสด 1.02 บาท/เดือน โดยกำหนดตัวแปรอิสระที่เหลือมีค่าคงที่

$b_8 = .213$ หมายความว่า ผู้ที่มีระดับการศึกษาชั้นมัธยมศึกษาจะมีค่าเฉลี่ยรายได้ต่อเดือน “มากกว่า” ผู้ที่มีระดับการศึกษาชั้นประถมศึกษาหรือต่ำกว่า 1.24 บาท/เดือน โดยกำหนดตัวแปรอิสระที่เหลือมีค่าคงที่

$b_9 = .648$ หมายความว่า ผู้ที่มีระดับการศึกษาปริญญาตรีหรือเทียบเท่าจะมีค่าเฉลี่ยรายได้ต่อเดือน “มากกว่า” ผู้ที่มีระดับการศึกษาชั้นประถมศึกษาหรือต่ำกว่า 1.91 บาท/เดือน โดยกำหนดตัวแปรอิสระที่เหลือมีค่าคงที่

$b_{10} = 1.155$ หมายความว่า ผู้ที่มีระดับการศึกษาสูงกว่าปริญญาตรีจะมีค่าเฉลี่ยรายได้ต่อเดือน “มากกว่า” ผู้ที่มีระดับการศึกษาชั้นประถมศึกษาหรือต่ำกว่า 3.17 บาท/เดือน โดยกำหนดตัวแปรอิสระที่เหลือมีค่าคงที่

$b_{11} = -.012$ หมายความว่า ผู้ที่ไม่ได้รับการศึกษาจะมีค่าเฉลี่ยรายได้ต่อเดือน “น้อยกว่า” ผู้ที่มีระดับการศึกษาชั้นประถมศึกษาหรือต่ำกว่า 0.99 บาท/เดือน โดยกำหนดตัวแปรอิสระที่เหลือมีค่าคงที่

$b_{12} = -.270$ หมายความว่า ผู้ที่มีอาชีพเป็นพนักงานบริการ และผู้จำหน่ายผลิตภัณฑ์จะมีค่าเฉลี่ยรายได้ต่อเดือน “น้อยกว่า” ผู้ที่มีอาชีพเป็นข้าราชการ เสมียน และช่างเทคนิค 0.76 บาท/เดือน โดยกำหนดตัวแปรอิสระที่เหลือมีค่าคงที่

$b_{13} = -.622$ หมายความว่า ผู้ที่ทำอาชีพเกษตร และประมงจะมีค่าเฉลี่ยรายได้ต่อเดือน “น้อยกว่า” ผู้ที่มีอาชีพเป็นข้าราชการ เสมียน และช่างเทคนิค 0.54 บาท/เดือน โดยกำหนดตัวแปรอิสระที่เหลือมีค่าคงที่

$b_{14} = -.329$ หมายความว่า ผู้ที่มีอาชีพเป็นช่างฝีมือ และผู้ควบคุมเครื่องจักรจะมีค่าเฉลี่ยรายได้ต่อเดือน “น้อยกว่า” ผู้ที่มีอาชีพเป็นข้าราชการ เสมียน และช่างเทคนิค 0.72 บาท/เดือน โดยกำหนดตัวแปรอิสระที่เหลือมีค่าคงที่

$b_{15} = -.154$ หมายความว่า ผู้ที่เป็นลูกจ้างเอกชนจะมีค่าเฉลี่ยรายได้ต่อเดือน “น้อยกว่า” ผู้ที่ลูกจ้างรัฐบาล 0.86 บาท/เดือน โดยกำหนดตัวแปรอิสระที่เหลือมีค่าคงที่

$b_{16} = -.498$ หมายความว่า ผู้ที่เป็นลูกจ้างทั่วไปจะมีค่าเฉลี่ยรายได้ต่อเดือน “น้อยกว่า” ผู้ที่ลูกจ้างรัฐบาล 0.61 บาท/เดือน โดยกำหนดตัวแปรอิสระที่เหลือมีค่าคงที่

บทที่ 5
สรุปผลการวิเคราะห์ อภิปราย และข้อเสนอแนะ

การศึกษาวิจัยเรื่อง “ตัวแบบทางสถิติในการศึกษาภาวะการมีงานทำของประชากรไทย” มี**วัตถุประสงค์**เพื่อศึกษาภาวะการมีงานทำของประชากรไทย ศึกษาปัจจัยที่มีผลต่อภาวะการมีงานทำของประชากรไทย หาตัวแบบที่เหมาะสมทำนายภาวะการมีงานทำของประชากรไทย และหาตัวแบบที่เหมาะสมทำนายรายได้ของประชากรไทย โดยใช้**ข้อมูลทุติยภูมิ**จากโครงการสำรวจภาวะการทำงานของประชากรไทย ในปี พ.ศ.2561 ของสำนักงานสถิติแห่งชาติ **ประชากร**ที่ใช้ในการศึกษาคือ ประชากรไทย **ตัวอย่าง**ที่ใช้ในการศึกษาคือ ตัวอย่างประชากรในการเก็บข้อมูลจากโครงการสำรวจภาวะการทำงานของประชากรไทยในปี พ.ศ.2561 จำนวน 172,241 คน และผู้มีรายได้ต่อเดือน จำนวน 47,933 คน **โปรแกรมวิเคราะห์ข้อมูล**ใช้ SPSS for Windows version 20.0 และโปรแกรม R Studio **สถิติที่ใช้ในการวิเคราะห์** ได้แก่ สถิติพรรณนา (Descriptive statistics) การวิเคราะห์ถดถอยลอจิสติก (Logistic regression analysis) และการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ (Multiple Linear Regression Analysis)

5.1 สรุปผลการวิจัย

5.1.1 สรุปข้อมูลเบื้องต้น

ผลการวิเคราะห์ข้อมูลเบื้องต้นของผู้ตอบแบบสอบถามจำนวน 172,241 คน พบว่าคนส่วนใหญ่มีงานทำจำนวน 112,511 คน หรือคิดเป็นร้อยละ 65.3 และคนที่ไม่มีงานทำจำนวน 59,730 คน หรือคิดเป็นร้อยละ 34.7 ในจำนวนนี้ส่วนใหญ่อาศัยอยู่ภาคกลางจำนวน 51,561 คน (ร้อยละ 29.9) รองลงมาคือ ภาคตะวันออกเฉียงเหนือจำนวน 44,.765 คน (ร้อยละ 26) ภาคเหนือจำนวน 36,919 คน (ร้อยละ 21.4) ภาคใต้จำนวน 29,582 คน (ร้อยละ 17.2) และกรุงเทพมหานครจำนวน 9,414 คน (ร้อยละ 5.5) ตามลำดับ

โดยแบ่งเป็นเพศชายจำนวน 80,163 คน (ร้อยละ 46.5) และเพศหญิงจำนวน 92,078 คน (ร้อยละ 53.5) มีอายุเฉลี่ย 47.61 ปี ส่วนใหญ่เป็นผู้ที่มีสถานภาพสมรสจำนวน 109,674 คน (ร้อยละ 63.7) รองลงมาคือ ผู้ที่มีสถานภาพโสดจำนวน 36,729 คน (ร้อยละ 21.3) ผู้ที่มีสถานภาพเป็นม่ายจำนวน 17,472 คน (ร้อยละ 10.1) และเป็นผู้ที่มีสถานภาพหย่าร้างจำนวน 8,366 คน (ร้อยละ 4.9) ตามลำดับ

จำแนกตามระดับการศึกษา พบว่าส่วนใหญ่เป็นผู้ที่จบชั้นประถมศึกษาหรือต่ำกว่าจำนวน 84,618 คน (ร้อยละ 49.1) รองลงมาคือ ผู้ที่จบมัธยมศึกษาจำนวน 51,544 คน (ร้อยละ 29.9) ผู้ที่จบปริญญาตรีหรือเทียบเท่าจำนวน 25,008 คน (ร้อยละ 14.5) ผู้ที่ไม่ได้รับการศึกษาจำนวน

8,373 คน (ร้อยละ 4.9) และผู้ที่จบสูงกว่าปริญญาตรีจำนวน 2,698 คน (ร้อยละ 1.6) ตามลำดับ โดยมีจำนวนสมาชิกในครัวเรือนเฉลี่ยอยู่ที่ 3.57 คน

5.1.2 การวิเคราะห์การถดถอยลอจิสติกของสถานะการจ้างงานของประชากรไทย

ผลการวิเคราะห์การถดถอยลอจิสติกเพื่อทำนายโอกาสในการจ้างงานของประชากรไทย พบว่าภูมิภาค (r), เพศ (s), สถานภาพการสมรส (m), ระดับการศึกษา (e), อายุ (x_1) และจำนวนสมาชิกในครัวเรือน (x_2) เป็นปัจจัยที่ส่งผลต่อโอกาสในการจ้างงานของประชากรไทย (y_1) โดยอัตราส่วนระหว่าง Deviance Residual และ Degrees of Freedom มีค่ามีค่าใกล้เคียงกัน จึงสรุปได้ว่าตัวแบบเหมาะสมกับข้อมูล และมีค่า AIC เท่ากับ 190,610 ซึ่งจากผลการวิเคราะห์การจำแนกกลุ่ม พบว่าตัวแบบสามารถทำนายผู้มีงานทำได้ถูกต้อง 89.4% และสามารถทำนายผู้ไม่มีงานทำได้ถูกต้อง 37.7% และโดยภาพรวมแล้วตัวแบบสามารถทำนายได้ถูกต้อง 71.4% สามารถเขียนตัวแบบการประมาณ ได้ดังนี้

$$p(y_1 = 1) = \frac{1}{1 + e^{-z}}$$

เมื่อ $p(y_1 = 1)$ คือ ค่าประมาณของความน่าจะเป็นที่ประชากรไทยจะมีงานทำ

และ $z = 1.01 + 0.208r_1 + 0.181r_2 + 0.14r_3 + 0.172r_4 + 0.782s_1 + 1.766m_1 + 0.671m_2$
$+1.82m_3 + 0.325e_1 - 0.002e_2 + 1.043e_3 + 1.508e_4 - 0.043x_1 - 0.103x_1$

จากการแปลผลจากค่า odds ratio พบว่าเมื่อเปรียบเทียบโอกาสในการจ้างงานของผู้ที่ทำงานในกรุงเทพมหานครกับผู้ที่ทำงานในภูมิภาคอื่น ๆ พบว่าผู้ที่ทำงานในภาคกลาง (Odds ratio = 1.231), ภาคเหนือ (Odds ratio = 1.198), ภาคตะวันออกเฉียงเหนือ (Odds ratio = 1.151) และภาคใต้ (Odds ratio = 1.187) มีโอกาสที่จะได้งานทำสูงกว่าคนที่อาศัยอยู่ในกรุงเทพมหานคร เมื่อเปรียบเทียบโอกาสในการจ้างงานของเพศหญิงกับเพศชาย พบว่าเพศชายมีโอกาสที่จะได้งานทำสูงกว่าเพศหญิงมากกว่า 2 เท่า (Odds ratio = 2.187) เมื่อเปรียบเทียบโอกาสในการจ้างงานของคนโสดกับคนกลุ่มอื่น พบว่าคนโสดมีโอกาสการจ้างงานน้อยที่สุดเมื่อเทียบกับคนอื่น ๆ ขณะที่ผู้ที่ผ่านการหย่าร้าง (Odds ratio = 6.173) และผู้ที่สมรสแล้ว (Odds ratio = 5.848) มีโอกาสการจ้างงานสูงมากกว่าคนกลุ่มอื่น ๆ ตามด้วยกลุ่มที่เป็นม่าย (Odds ratio = 1.956) ตามลำดับ เมื่อเปรียบเทียบโอกาสในการจ้างงานตามระดับการศึกษา พบว่าผู้ที่ไม่ได้รับการศึกษามีโอกาสการจ้างงานไม่ต่างจากผู้ที่จบชั้นมัธยม (Odds ratio = .998) โดยผู้ที่มีโอกาสการจ้างงานมากที่สุดคือ ผู้ที่มีระดับการศึกษาสูงกว่าปริญญาตรี (Odds ratio = 4.516) รองลงมาคือ ผู้ที่จบปริญญาตรีหรือเทียบเท่า (Odds ratio = 2.839) และผู้ที่จบประถมศึกษาหรือต่ำกว่า (Odds ratio = 1.385) ตามลำดับ นอกจากนี้ถ้าหากมี

อายุเพิ่มขึ้น 1 ปี (โดยตัวแปรอื่นคงที่) โอกาสการจ้างงานจะลดลงร้อยละ 4.2 และถ้าจำนวนสมาชิกในครัวเรือนเพิ่มขึ้น 1 คน (โดยตัวแปรอื่นคงที่) โอกาสการจ้างงานจะลดลงร้อยละ 9.8

5.1.3 การวิเคราะห์การถดถอยเชิงเส้นพหุคูณของรายได้ของประชากรไทย

ผลการวิเคราะห์ข้อมูลเบื้องต้นของผู้มีรายได้ต่อเดือนจำนวน 47,933 คน พบว่ามีรายได้เฉลี่ยต่อเดือนประมาณ 13,616.17 บาท อายุเฉลี่ยของคนกลุ่มนี้อยู่ที่ 39.98 ปี โดยแบ่งเป็นเพศชายจำนวน 25,811 คน (ร้อยละ 53.8) และเพศหญิงจำนวน 22,122 คน (ร้อยละ 46.2) ส่วนใหญ่เป็นผู้ที่มีสถานภาพสมรสจำนวน 30,766 คน (ร้อยละ 64.2) รองลงมาคือ ผู้ที่มีสถานภาพโสดจำนวน 12,391 คน (ร้อยละ 25.9) ผู้ที่มีสถานภาพหย่าร้างจำนวน 3,287 คน (ร้อยละ 6.9) และผู้ที่มีสถานภาพเป็นม่ายจำนวน 1,489 คน (ร้อยละ 3.1) ตามลำดับ โดยมีจำนวนสมาชิกในครัวเรือนเฉลี่ยอยู่ที่ 3.53 คน

จำแนกตามระดับการศึกษา พบว่าส่วนใหญ่เป็นผู้ที่จบมัธยมศึกษาจำนวน 15,760 คน (ร้อยละ 32.9) รองลงมาคือ ผู้ที่จบชั้นประถมศึกษาหรือต่ำกว่าจำนวน 14,796 คน (ร้อยละ 30.9) ผู้ที่จบปริญญาตรีหรือเทียบเท่าจำนวน 13,281 คน (ร้อยละ 27.7) ผู้ที่ไม่ได้รับการศึกษาจำนวน 2,187 คน (ร้อยละ 4.6) และผู้ที่จบสูงกว่าปริญญาตรีจำนวน 1,909 คน (ร้อยละ 4.0) ตามลำดับ

จำแนกตามการประกอบอาชีพ พบว่าส่วนใหญ่ประกอบอาชีพเป็นช่างฝีมือ และผู้ควบคุมเครื่องจักรจำนวน 22,951 คน (ร้อยละ 47.9) รองลงมาคือ ประกอบอาชีพเป็นข้าราชการเสมียน และช่างเทคนิคจำนวน 15,117 คน (ร้อยละ 31.5) ประกอบอาชีพเป็นพนักงานบริการและผู้จำหน่ายผลิตภัณฑ์จำนวน 6,923 คน (ร้อยละ 14.4) และประกอบอาชีพเกษตรและประมงจำนวน 2,942 คน (ร้อยละ 6.1) ตามลำดับ แบ่งเป็นเป็นลูกจ้างเอกชนจำนวน 32,348 คน (ร้อยละ 67.5) ลูกจ้างรัฐบาลจำนวน 12,326 คน (ร้อยละ 25.7) และลูกจ้างทั่วไปจำนวน 3,259 คน (ร้อยละ 6.8) โดยมีจำนวนชั่วโมงในการทำงานเฉลี่ยอยู่ที่ 43.88 ชั่วโมง/สัปดาห์

ผลการวิเคราะห์การถดถอยเชิงเส้นพหุคูณของรายได้ประชากรไทย พบว่าตัวแบบที่เหมาะสมในการทำนายรายได้เฉลี่ยต่อเดือนของประชากรไทยคือ ตัวแบบ $\ln(y_2)$ โดยปัจจัยที่ส่งผลต่อรายได้เฉลี่ยต่อเดือนของประชากรไทย (y_2) ประกอบด้วย เพศ (s), สถานภาพการสมรส (m), ระดับการศึกษา (e), อาชีพ (o), ตำแหน่งงาน (t), อายุ (x_1), จำนวนสมาชิกในครัวเรือน (x_2) และจำนวนชั่วโมงในการทำงาน (x_3) โดยมีค่าสัมประสิทธิ์การตัดสินใจ (R^2) เท่ากับ 0.5876 หมายความว่า ตัวแปรอิสระสามารถอธิบายความผันแปรของรายได้ประชากรไทย 58.76% โดยมีค่าความคลาดเคลื่อนมาตรฐานของส่วนเหลือ (Residual Standard Error) เท่ากับ 0.4275 ค่าสถิติ AIC เท่ากับ 54,606.44 และค่าสถิติ BIC เท่ากับ 54,764.44 สามารถเขียนตัวแบบการประมาณรายได้เฉลี่ยต่อเดือนของประชากรไทย ได้ดังนี้

$$\begin{aligned}\ln(y_2) = {} & 8.492 + 0.008x_1 - 0.021x_2 + 0.015x_3 - 0.148s_1 + 0.076m_1 \\ & -0.098m_2 + 0.012m_3 + 0.213e_1 + 0.648e_2 + 1.155e_3 - 0.012e_4\end{aligned}$$

$$-0.27o_1 - 0.622o_2 - 0.329o_3 - 0.154t_1 - 0.498t_2$$

โดยที่

$b_1 = .008$ หมายความว่า ค่าเฉลี่ยรายได้ต่อเดือนของประชากรไทยจะ “เพิ่มขึ้น” 1.01 บาท/เดือน ถ้ามีอายุเพิ่มขึ้น 1 ปี โดยกำหนดตัวแปรอิสระที่เหลือมีค่าคงที่

$b_2 = -.021$ หมายความว่า ค่าเฉลี่ยรายได้ต่อเดือนของประชากรไทยจะ “ลดลง” 0.98 บาท/เดือน ถ้าจำนวนสมาชิกในครัวเรือนเพิ่มขึ้น 1 คน โดยกำหนดตัวแปรอิสระที่เหลือมีค่าคงที่

$b_3 = .015$ หมายความว่า ค่าเฉลี่ยรายได้ต่อเดือนของประชากรไทยจะ “เพิ่มขึ้น” 1.02 บาท/เดือน ถ้าจำนวนชั่วโมงในการทำงานเพิ่มขึ้น 1 ชั่วโมงต่อสัปดาห์ โดยกำหนดตัวแปรอิสระที่เหลือมีค่าคงที่

$b_4 = -.148$ หมายความว่า เพศหญิงจะมีค่าเฉลี่ยรายได้ต่อเดือน “น้อยกว่า” เพศชาย 0.86 บาท/เดือน โดยกำหนดตัวแปรอิสระที่เหลือมีค่าคงที่

$b_5 = .076$ หมายความว่า ผู้ที่มีสถานภาพสมรสจะมีค่าเฉลี่ยรายได้ต่อเดือน “มากกว่า” ผู้ที่มีสถานภาพโสด 1.08 บาท/เดือน โดยกำหนดตัวแปรอิสระที่เหลือมีค่าคงที่

$b_6 = -.098$ หมายความว่า ผู้ที่มีสถานภาพเป็นม่ายจะมีค่าเฉลี่ยรายได้ต่อเดือน “น้อยกว่า” ผู้ที่มีสถานภาพโสด 0.91 บาท/เดือน โดยกำหนดตัวแปรอิสระที่เหลือมีค่าคงที่

$b_7 = .012$ หมายความว่า ผู้ที่มีสถานภาพเป็นหย่าร้างจะมีค่าเฉลี่ยรายได้ต่อเดือน “มากกว่า” ผู้ที่มีสถานภาพโสด 1.02 บาท/เดือน โดยกำหนดตัวแปรอิสระที่เหลือมีค่าคงที่

$b_8 = .213$ หมายความว่า ผู้ที่มีระดับการศึกษาชั้นมัธยมศึกษาจะมีค่าเฉลี่ยรายได้ต่อเดือน “มากกว่า” ผู้ที่มีระดับการศึกษาชั้นประถมศึกษาหรือต่ำกว่า 1.24 บาท/เดือน โดยกำหนดตัวแปรอิสระที่เหลือมีค่าคงที่

$b_9 = .648$ หมายความว่า ผู้ที่มีระดับการศึกษาปริญญาตรีหรือเทียบเท่าจะมีค่าเฉลี่ยรายได้ต่อเดือน “มากกว่า” ผู้ที่มีระดับการศึกษาชั้นประถมศึกษาหรือต่ำกว่า 1.91 บาท/เดือน โดยกำหนดตัวแปรอิสระที่เหลือมีค่าคงที่

$b_{10} = 1.155$ หมายความว่า ผู้ที่มีระดับการศึกษาสูงกว่าปริญญาตรีจะมีค่าเฉลี่ยรายได้ต่อเดือน “มากกว่า” ผู้ที่มีระดับการศึกษาชั้นประถมศึกษาหรือต่ำกว่า 3.17 บาท/เดือน โดยกำหนดตัวแปรอิสระที่เหลือมีค่าคงที่

$b_{11} = -.012$ หมายความว่า ผู้ที่ไม่ได้รับการศึกษาจะมีค่าเฉลี่ยรายได้ต่อเดือน “น้อยกว่า” ผู้ที่มีระดับการศึกษาชั้นประถมศึกษาหรือต่ำกว่า 0.99 บาท/เดือน โดยกำหนดตัวแปรอิสระที่เหลือมีค่าคงที่

$b_{12} = -.270$ หมายความว่า ผู้ที่มีอาชีพเป็นพนักงานบริการ และผู้จำหน่ายผลิตภัณฑ์จะมีค่าเฉลี่ยรายได้ต่อเดือน “น้อยกว่า” ผู้ที่มีอาชีพเป็นข้าราชการ เสมียน และช่างเทคนิค 0.76 บาท/เดือน โดยกำหนดตัวแปรอิสระที่เหลือมีค่าคงที่

$b_{13} = -.622$ หมายความว่า ผู้ที่ทำอาชีพเกษตร และประมงจะมีค่าเฉลี่ยรายได้ต่อเดือน “น้อยกว่า” ผู้ที่มีอาชีพเป็นข้าราชการ เสมียน และช่างเทคนิค 0.54 บาท/เดือน โดยกำหนดตัวแปรอิสระที่เหลือมีค่าคงที่

$b_{14} = -.329$ หมายความว่า ผู้ที่มีอาชีพเป็นช่างฝีมือ และผู้ควบคุมเครื่องจักรจะมีค่าเฉลี่ยรายได้ต่อเดือน “น้อยกว่า” ผู้ที่มีอาชีพเป็นข้าราชการ เสมียน และช่างเทคนิค 0.72 บาท/เดือน โดยกำหนดตัวแปรอิสระที่เหลือมีค่าคงที่

$b_{15} = -.154$ หมายความว่า ผู้ที่เป็นลูกจ้างเอกชนจะมีค่าเฉลี่ยรายได้ต่อเดือน “น้อยกว่า” ผู้ที่ลูกจ้างรัฐบาล 0.86 บาท/เดือน โดยกำหนดตัวแปรอิสระที่เหลือมีค่าคงที่

$b_{16} = -.498$ หมายความว่า ผู้ที่เป็นลูกจ้างทั่วไปจะมีค่าเฉลี่ยรายได้ต่อเดือน “น้อยกว่า” ผู้ที่ลูกจ้างรัฐบาล 0.61 บาท/เดือน โดยกำหนดตัวแปรอิสระที่เหลือมีค่าคงที่

5.2 อภิปรายผลการวิจัย

การวิจัยครั้งนี้เป็นการศึกษาปัจจัยที่สามารถทำนายสถานะการจ้างงานของประชากรไทย โดยใช้ข้อมูลทุติยภูมิจากโครงการสำรวจภาวะการทำงานของประชากรไทย ในปี พ.ศ.2561 ของสำนักงานสถิติแห่งชาติ และหาตัวแบบที่ใช้ทำนายรายได้เฉลี่ยของประชากรไทย ผู้วิจัยจึงได้ทำการศึกษาและทบทวนวรรณกรรมและทฤษฎีที่เกี่ยวข้อง และตัดสินใจเลือกใช้ตัวแบบการถดถอยลอจิสติกเพื่อทำนายโอกาสการจ้างงานของประชากรไทย และใช้ตัวแบบการถดถอยพหุคูณเพื่อทำนายรายได้เฉลี่ยของประชากรไทย

จากผลการวิเคราะห์การทำนายโอกาสการจ้างงานของประชากรไทย พบว่าโอกาสการจ้างงานของผู้ที่ทำงานในกรุงเทพมหานคร น้อยกว่าผู้ที่ทำงานในภูมิภาคอื่น ๆ โดยผู้ที่ทำงานในภาคกลางของประเทศไทยจะมีโอกาสการจ้างงานสูงสุด (1.231 เท่า เมื่อเทียบกับผู้ที่ทำงานในกรุงเทพมหานคร) เห็นได้ชัดว่าอาจเป็นเพราะกรุงเทพมหานครเป็นเมืองที่มีประชากรมากที่สุดในประเทศไทย ดังนั้นการหางานจึงมีความสามารถในการแข่งขันสูงกว่าเมื่อเทียบกับภูมิภาคอื่น ๆ ขณะที่ผู้ชายมีโอกาสได้งานมากกว่าผู้หญิงถึง 2.187 เท่า ซึ่งผลลัพธ์ไม่ได้เหนือความคาดหมายแต่อย่าง

ใด เมื่อพิจารณาตามระดับการศึกษา พบว่าโอกาสในการได้งานทำจะสูงขึ้นสำหรับผู้ที่มีการศึกษาสูง ผู้ที่มีระดับการศึกษาสูงกว่าปริญญาตรีมีโอกาสได้งานสูงกว่าผู้ที่ไม่ได้รับการศึกษาถึง 4.516 เท่า ซึ่งแสดงให้เห็นว่าระดับการศึกษามีส่วนสำคัญอย่างมากในการได้งานทำ สอดคล้องกับพรทวี เถื่อนคำแสน (2559) นอกจากนี้ผู้ที่มีสถานะโสดจะมีโอกาสการจ้างงานน้อยเมื่อเทียบกับกลุ่มอื่น ๆ อย่างไรก็ตามผลลัพธ์นี้มีความคลุมเครือเนื่องจากบุคคลในกลุ่มตัวอย่างอาจเป็นคนรุ่นใหม่ที่เพิ่งสำเร็จการศึกษาและยังไม่ได้งาน จึงไม่ได้หมายความว่าพวกเขาว่างงานเพราะเป็นโสด อายุที่เพิ่มขึ้น และจำนวนสมาชิกในครอบครัวที่เพิ่มขึ้น จะทำให้โอกาสการจ้างงานน้อยลง

จากผลการวิเคราะห์การทำนายรายได้เฉลี่ยของประชากรไทย พบว่าอายุที่เพิ่มขึ้นและชั่วโมงทำงานที่เพิ่มขึ้นส่งผลต่อรายได้เฉลี่ยที่เพิ่มขึ้นสอดคล้องกับองอาจ รุกขวัฒนกุล (2558) ในขณะที่จำนวนสมาชิกในครัวเรือนเพิ่มขึ้นจะส่งผลให้รายได้เฉลี่ยลดลง เพศชายมีรายได้เฉลี่ยสูงกว่าเพศหญิง ในขณะที่ผู้ที่มีครอบครัวแล้ว ผู้ที่จบการศึกษาระดับบัณฑิตศึกษา ผู้ที่มีอาชีพ (ข้าราชการ พนักงาน และช่างเทคนิค) และผู้ที่เป็นพนักงานของรัฐจะมีรายได้เฉลี่ยสูงที่สุด

จากผลข้างต้นที่กล่าวมาจึงสามารถสรุปได้ว่า เพศชายที่อาศัยอยู่ในภาคกลางของประเทศไทย และมีระดับการศึกษาสูงกว่าปริญญาตรี จะมีโอกาสมีงานทำมากกว่ากลุ่มคนอื่น ๆ ในขณะที่เพศชายวัยกลางคนที่มีระดับการศึกษาสูงกว่าปริญญาตรี เป็นพนักงานของรัฐ และทำงานในกลุ่มข้าราชการ เจ้าหน้าที่ธุรการ หรือเจ้าหน้าที่เทคนิค จะมีรายได้เฉลี่ยสูงกว่าคนกลุ่มอื่น ๆ

5.3 ข้อเสนอแนะ

การวิจัยครั้งนี้ได้รับการอนุเคราะห์ข้อมูลจากสำนักงานสถิติแห่งชาติ โดยข้อมูลมีการเก็บรวบรวมไว้อยู่ก่อนแล้ว เป็นที่น่าเสียดายที่ผู้วิจัยไม่สามารถควบคุมจำนวนคำตอบในแบบสำรวจได้ โดยในบางตัวแปร เช่น รายได้ต่อเดือน ซึ่งเป็นตัวแปรที่ผู้วิจัยสนใจที่จะวิเคราะห์ แต่ผู้มีงานทำที่เปิดเผยรายได้ต่อเดือนนั้นมีน้อยมาก โดยมีสัดส่วนประมาณ 50% ของผู้มีงานทำทั้งหมด หากเป็นไปได้ควรรวบรวมข้อมูลทั้งหมดให้ได้มากที่สุดเพื่อผลการวิเคราะห์ที่ดีขึ้น

เนื่องจากการวิจัยครั้งนี้มีเป้าหมายมุ่งเน้นไปที่ปัจจัยที่จะสามารถเพิ่มโอกาสในการได้งาน และต้องการรับรู้ว่าปัจจัยใดที่จะช่วยให้รายได้เฉลี่ยต่อเดือนเพิ่มขึ้น ซึ่งถ้ามองจากผลลัพธ์ของการวิจัยครั้งนี้ที่เห็นได้ชัดเลย คือ เพศชายจะมีโอกาสได้งานมากกว่าเพศหญิง ผู้ที่ประกอบอาชีพข้าราชการ เสมียร และช่างเทคนิคจะมีรายได้เฉลี่ยต่อเดือนสูงกว่าคนอื่น ๆ ผู้ที่มีระดับการศึกษาสูง ๆ จะมีโอกาสได้งานมากกว่าคนอื่น ๆ หรือแม้กระทั่งมีรายได้เฉลี่ยที่สูงขึ้นด้วย ถ้ามองกลับกันจะเห็นว่าผู้หญิง หรือผู้ที่มีการศึกษาระดับกลาง ๆ จะเสียโอกาสจากปัจจัยดังกล่าวค่อนข้างชัดเจน ดังนั้นหน่วยงานที่เกี่ยวข้องสามารถนำผลการวิเคราะห์จากงานวิจัยชิ้นนี้อาจนำไปสู่การช่วยเหลือกลุ่มคนเหล่านี้ให้มีคุณภาพชีวิตที่ดีขึ้นได้

นอกจากนี้ผู้วิจัยยังได้ทำการวิเคราะห์ปัจจัยที่มีผลต่อภาวะการมีงานทำของประชากรไทยและหาตัวแบบเพื่อทำนายรายได้ของประชากรไทยโดยแยกเป็นตัวแบบระดับภูมิภาค แต่เนื่องจากตัวแบบแต่ละภูมิภาคนั้นให้ผลลัพธ์ที่ไม่แตกต่างกัน ผู้วิจัยจึงเห็นว่าตัวแบบที่ได้นำเสนอไปเพียงพอต่อการสรุปผลการวิจัย

บรรณานุกรม

กัลยา วานิชย์บัญชา. (2549). *สถิติสำหรับงานวิจัย* (พิมพ์ครั้งที่ 2). กรุงเทพฯ: โรงพิมพ์จุฬาลงกรณ์มหาวิทยาลัย ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย.

โครงการรายได้และการกระจายรายได้ของครัวเรือน. (2545). *รายงานเชิงวิเคราะห์เรื่องรายได้และการกระจายรายได้ของครัวเรือน พ.ศ. 2545 ระดับจังหวัด*.
ค้นจาก http://service.nso.go.th/nso/knowledge/knowledge09/revenue.html

จิรภัทร ภู่ขวัญทอง. (2550). *กระบวนการวิเคราะห์ข้อมูลทวินามแบบลบที่มีความแปรปรวนเป็นฟังก์ชันพหุนามกำลังสองของค่าเฉลี่ย.* (วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต). มหาวิทยาลัยสงขลานครินทร์, สงขลา

ชลิดา ตระกูลสุนทร. (2557). *สถิติเพื่อการวิจัย.* เอกสารประกอบการเรียนการสอนรายวิชาสถิติเพื่อการวิจัย, นครปฐม: คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏนครปฐม.

นราทิพย์ จั่นสกุล, และทัศนา วิกยานนท์. (2560). *การวิเคราะห์การถดถอยโดยใช้โปรแกรม R* (พิมพ์ครั้งที่ 2). สงขลา: โรงพิมพ์ดิจิตอล หน่วยโสตทัศนศึกษา คณะวิทยาศาสตร์ มหาวิทยาลัยสงขลานครินทร์.

นราทิพย์ จั่นสกุล. (2561). *Analysis of Advanced Statistical Model*. เอกสารประกอบการเรียนการสอนรายวิชาการวิเคราะห์ตัวแบบเชิงสถิติขั้นสูง, สงขลา: คณะวิทยาศาสตร์ มหาวิทยาลัยสงขลานครินทร์.

ผกากรอง เทพรักษ์. (2557). พฤติกรรมการหางานทำที่ส่งผลต่อการมีงานทำของบัณฑิต มหาวิทยาลัยเทคโนโลยีราชมงคลศรีวิชัย วิทยาเขตสงขลา. *วารสารวิจัยทางการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ*, 8(2), 154-160.
ค้นจาก http://ejournals.swu.ac.th/index.php/jre/article/view/6809/6411

วรรณวิมล จงจรวยสกุล. (2555). *ภาวะการมีงานทำของผู้สำเร็จการศึกษาวิทยาลัยราชพฤกษ์ ปีการศึกษา 2554-2555*. นนทบุรี: วิทยาลัยราชพฤกษ์.

พรทวี เถื่อนคำแสน และคณะ. (2559). ปัจจัยที่มีผลต่อแรงจูงใจในการตัดสินใจเลือกทำงานของพนักงานระดับปฏิบัติการในนิคมอุตสาหกรรมเหมราช อีสเทิร์นซีบอร์ด จังหวัดระยอง. *วารสารบริหารธุรกิจเทคโนโลยีมหานคร*, 13(1), 1-23.
ค้นจาก https://www.tci-thaijo.org/index.php/journalmbsmut/article/view/77268

สำนักงานสถิติแห่งชาติ. (2562). สรุปผลการสำรวจภาวะการทำงานของประชากรเดือนธันวาคม พ.ศ. 2562. NSO Publication. กรุงเทพมหานคร.

องอาจ รุกขวัฒนกุล. (2558). ปัจจัยที่มีอิทธิพลต่อการเปลี่ยนแปลงรายได้ของประชาชนในพื้นที่การพัฒนา เขตเศรษฐกิจพิเศษ หนองคาย มุกดาหาร แม่สอด. *วารสารการเงินการคลัง*, 29(91). 69-75. ค้นจาก

http://ethesisarchive.library.tu.ac.th/thesis/2015/TU_2015_5704010700_4720_2858.pdf

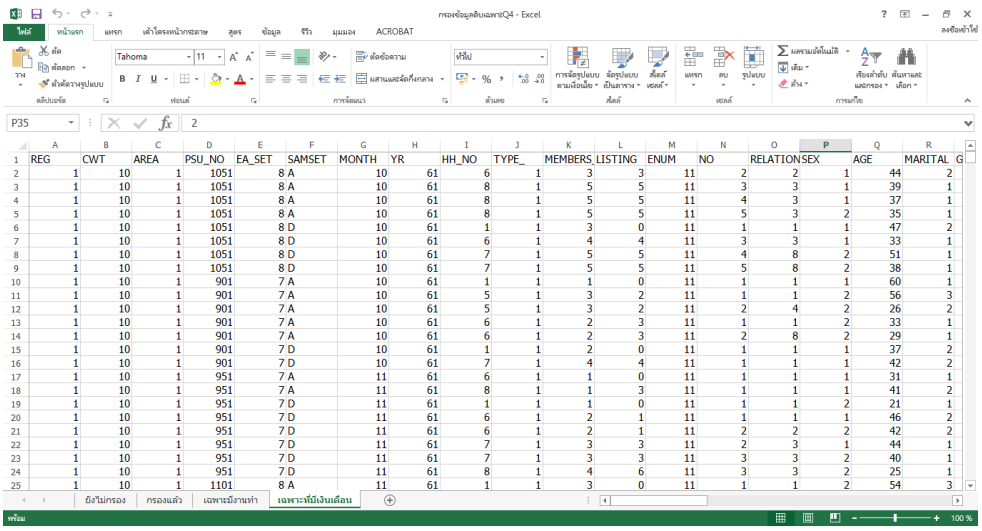
Atkinson, A. (1985). *Plots, Tranformations and Regression*. An introduction to Grahical Methods of Diagnostic Regression Analysis, Oxford: Clarendon Press.

Cadden, M. & Arnett, P. (2015). Factors Associated with Employment Status in Individuals with Multiple Sclerosis. *International Journal of MS Care*, 17(6), 284-291.

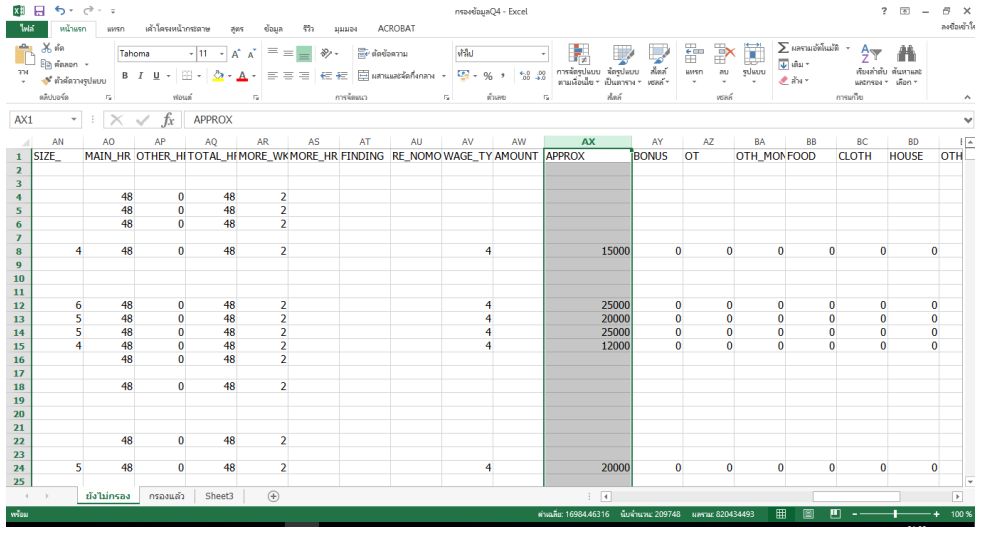
Demetrio, C.G.B. & Hinde, J.P. (1997). Half-normal plots and overdispersion. *GLIM Newsletter*, n. 27, p. 19-26.

William, D.A. (1987). Generalize linear model diagnostics using the deviance and single case deletions, *Applied Statistics*, 36(2), 181-191.

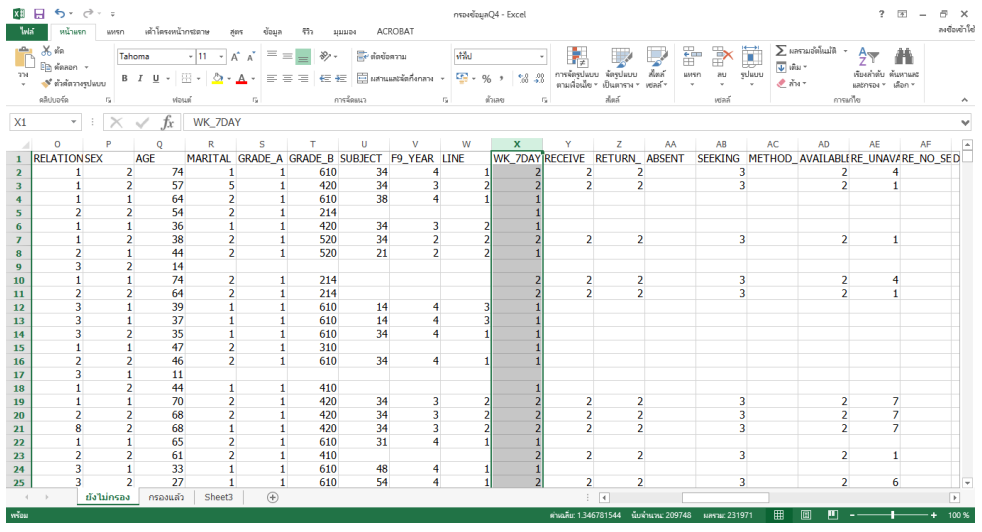
ภาคผนวก 1
ตัวอย่างข้อมูล



REG	CWT	AREA	PSU_NO	EA_SET	SAMSET	MONTH	YR	HH_NO	TYPE_	MEMBERS	LISTING	ENUM	NO	RELATION	SEX	AGE	MARITAL	G
1	10	1	1051	8	A	10	61	6	1	3	3	11	2	2	1	44	2	
1	10	1	1051	8	A	10	61	8	1	5	5	11	3	3	1	39	1	
1	10	1	1051	8	A	10	61	8	1	5	5	11	4	3	1	37	1	
1	10	1	1051	8	A	10	61	8	1	5	5	11	5	3	2	35	1	
1	10	1	1051	8	D	10	61	1	1	3	0	11	1	1	1	47	2	
1	10	1	1051	8	D	10	61	6	1	4	4	11	3	3	1	33	1	
1	10	1	1051	8	D	10	61	7	1	5	5	11	4	8	2	51	1	
1	10	1	1051	8	D	10	61	7	1	5	5	11	5	8	2	38	1	
1	10	1	901	7	A	10	61	1	1	1	0	11	1	1	1	60	1	
1	10	1	901	7	A	10	61	5	1	3	2	11	1	1	2	56	3	
1	10	1	901	7	A	10	61	5	1	3	2	11	2	4	2	26	2	
1	10	1	901	7	A	10	61	6	1	2	3	11	1	1	2	33	1	
1	10	1	901	7	A	10	61	6	1	2	3	11	2	8	2	29	1	
1	10	1	901	7	D	10	61	1	1	2	0	11	1	1	1	37	2	
1	10	1	901	7	D	10	61	7	1	4	4	11	1	1	1	42	2	
1	10	1	951	7	A	11	61	6	1	1	0	11	1	1	1	31	1	
1	10	1	951	7	A	11	61	8	1	1	3	11	1	1	1	41	2	
1	10	1	951	7	D	11	61	1	1	1	0	11	1	1	2	21	1	
1	10	1	951	7	D	11	61	6	1	2	1	11	1	1	1	46	2	
1	10	1	951	7	D	11	61	6	1	2	1	11	2	2	2	42	2	
1	10	1	951	7	D	11	61	7	1	3	3	11	2	3	1	44	1	
1	10	1	951	7	D	11	61	7	1	3	3	11	3	3	2	40	1	
1	10	1	951	7	D	11	61	8	1	4	6	11	3	3	2	25	1	
1	10	1	1101	8	A	11	61	1	1	3	0	11	1	1	2	54	3	



SIZE_	MAIN_HR	OTHER_HI	TOTAL_HI	MORE_WK	MORE_HR	FINDING	RE_NOMO	WAGE_TY	AMOUNT	APPROX	BONUS	OT	OTH_MON	FOOD	CLOTH	HOUSE	OTH
	48	0	48	2													
	48	0	48	2													
	48	0	48	2													
4	48	0	48	2				4		15000	0	0	0	0	0	0	
6	48	0	48	2				4		25000	0	0	0	0	0	0	
5	48	0	48	2				4		20000	0	0	0	0	0	0	
5	48	0	48	2				4		25000	0	0	0	0	0	0	
4	48	0	48	2				4		12000	0	0	0	0	0	0	
	48	0	48	2													
	48	0	48	2													
	48	0	48	2													
5	48	0	48	2				4		20000	0	0	0	0	0	0	



RELATION	SEX	AGE	MARITAL	GRADE_A	GRADE_B	SUBJECT	F9_YEAR	LINE	WK_7DAY	RECEIVE	RETURN_	ABSENT	SEEKING	METHOD_	AVAILABL	RE_UNAV	RE_NO_SE	D
1	2	74	1	1	610	34	4	1	2	2	2		3		2	4		
1	2	57	5	1	420	34	3	2	2	2	2		3		2	1		
1	1	64	2	1	610	38	4	1	1									
2	2	54	2	1	214				1									
1	1	36	1	1	420	34	3	2	1									
1	2	38	2	1	520	34	2	2	2	2	2		3		2	1		
2	1	44	2	1	520	21	2	2	1									
3	2	14																
1	1	74	2	1	214				2	2	2		3		2	4		
2	2	64	2	1	214				2	2	2		3		2	1		
3	1	39	1	1	610	14	4	3	1									
3	1	37	1	1	610	14	4	3	1									
3	2	35	1	1	610	34	4	1	1									
1	1	47	2	1	310				1									
2	2	46	2	1	610	34	4	1	1									
3	1	11																
1	2	44	1	1	410				1									
1	1	70	2	1	420	34	3	2	2	2	2		3		2	7		
2	2	68	2	1	420	34	3	2	2	2	2		3		2	7		
8	2	68	1	1	420	34	3	2	2	2	2		3		2	7		
1	1	65	2	1	610	31	4	1	1									
2	2	61	2	1	410				2	2	2		3		2	1		
3	1	33	1	1	610	48	4	1	1									
3	2	27	1	1	610	54	4	1	2	2	2		3		2	6		

ภาคผนวก 2
คำสั่ง R และผลวิเคราะห์

❖ ผลการวิเคราะห์ข้อมูลเบื้องต้น

ภูมิภาค

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ภาคใต้	29582	17.2	17.2	17.2
	ภาคกลาง	51561	29.9	29.9	47.1
	ภาคเหนือ	36919	21.4	21.4	68.5
	ภาคตะวันออกเฉียงเหนือ	44765	26.0	26.0	94.5
	กรุงเทพมหานคร	9414	5.5	5.5	100.0
	Total	172241	100.0	100.0	

เพศ

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ชาย	80163	46.5	46.5	46.5
	หญิง	92078	53.5	53.5	100.0
	Total	172241	100.0	100.0	

สถานภาพการสมรส

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	หย่าร้าง/แยกกันอยู่	8366	4.9	4.9	4.9
	สมรส/อยู่ด้วยกัน	109674	63.7	63.7	68.5
	ม่าย	17472	10.1	10.1	78.7
	โสด	36729	21.3	21.3	100.0
	Total	172241	100.0	100.0	

ระดับการศึกษา

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ประถมศึกษา/ต่ำกว่า	84618	49.1	49.1	49.1
	มัธยมศึกษา	51544	29.9	29.9	79.1
	ปริญญาตรี/เทียบเท่า	25008	14.5	14.5	93.6
	สูงกว่าปริญญาตรี	2698	1.6	1.6	95.1
	ไม่ได้รับการศึกษา	8373	4.9	4.9	100.0
	Total	172241	100.0	100.0	

การจ้างงาน

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ไม่มีงานทำ	59730	34.7	34.7	34.7
	มีงานทำ	112511	65.3	65.3	100.0
	Total	172241	100.0	100.0	

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
จำนวนสมาชิกในครัวเรือน	172241	1	18	3.57	1.834
อายุ	172241	15	98	47.61	17.743
Valid N (listwise)	172241				

❖ การวิเคราะห์การถดถอยลอจิสติกของสถานะการจ้างงานของประชากรไทย

- คำสั่ง R ที่ใช้ในการวิเคราะห์การถดถอยลอจิสติก

```
setwd("D:/Thesis")
data_logis <- read.csv("DataAnalysisInR1.csv",h=T)
attach(data_logis)
names(data_logis)
fr <- factor(r)
fs <- factor(s)
fm <- factor(m)
fe <- factor(e)
#สร้างโมเดล
model_logis <- glm(y1~fr+fs+fm+fe+x1+x2,family = binomial)
summary(model_logis)
#ข้อตกลงเบื้องต้น
windows()
par(mfrow=c(1,2))
plot(model_logis$fitted.values,model_logis$residuals,pch=16,ylab = expression(e[i]),xlab
=expression(hat(y[i])) )
qqnorm(model_logis$residuals,pch=16)
qqline(model_logis$residuals)
#ทดสอบการแจกแจงของข้อมูล
ks.test(model_logis$residuals,model_logis$fitted.values)
#ทดสอบความแปรปรวน
fv <- model_logis$fitted.values
summary(fv)
fv1 <- ifelse(fv <= summary(fv)[2],1,fv)
fv1 <- ifelse(fv > summary(fv)[2]&fv <= summary(fv)[3],2,fv1)
fv1 <- ifelse(fv > summary(fv)[3]&fv <= summary(fv)[5],3,fv1)
fv1 <- ifelse(fv > summary(fv)[5],4,fv1)
table(fv1)
fv.group <- factor(fv1)
bartlett.test(model_logis$residuals~fv.group)
#สร้างกราฟ Half-Normal Plot
windows()
hnp.binorm(model_logis)
#คำนวณค่า AIC
AIC(model_logis)
```

- ฟังก์ชั่นการสร้างแผนภาพ Half-Normal Plot

```
hnp.binorm <- function(model.bin)
{
  p <- round(fitted(model.bin),4)
  ni <- model.bin$model[,1]
  mu <-ni*p ; q <- 1-p
  n <- length(mu)
  r.sim <- matrix(0,n,19)
  rmax <- 0 ; rmin <- 0 ; rme <- 0
  sres <- sort(abs(rstandard(model.bin)))

  for(i in 1:19) {
    ys <- rbinom(n,ni,p)
    model.bin$model[,1] <- ys
    #model.bin$model[[1]][,2] <- ni-ys
    ys.glm <- glm(model.bin$model,family = binomial)
    rs <- sort(abs(rstandard(ys.glm)))
    r.sim[,i] <- rs
  }

  for (i in 1:nrow(r.sim)) {
    rmin[i] <- min(r.sim[i,])
    rme[i] <- mean(r.sim[i,])
    rmax[i] <- max(r.sim[i,])
  }

  i <- 1:n
  nd<-qnorm((i+n-0.125)/(2*n+0.5))
  plot(nd,sres,pch=4,xlab = "Half-normal scores",ylab = "Standartized devince residuals")
  ylim=c(min(rmin),max(rmax))
  lines(nd,rmin,col="blue")
  lines(nd,rmax,col="blue")
  lines(nd,rme,col="brown",lty="dashed")
}
```

- ค่าสัมประสิทธิ์และค่าพารามิเตอร์จากการวิเคราะห์การถดถอยลอจิสติกของสถานะการจ้างงานของประชากรไทย

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	REG			72.438	4	.000	
	REG(1)	.172	.028	38.900	1	.000	1.187
	REG(2)	.208	.026	64.199	1	.000	1.231
	REG(3)	.181	.027	45.256	1	.000	1.198
	REG(4)	.140	.026	28.390	1	.000	1.151
	MEMBERS	-.103	.003	1099.148	1	.000	.902
	SEX(1)	.782	.012	4530.444	1	.000	2.187
	AGE	-.043	.000	8629.714	1	.000	.958
	MARITAL			14579.752	3	.000	
	MARITAL(1)	1.820	.031	3543.144	1	.000	6.173
	MARITAL(2)	1.766	.017	10803.478	1	.000	5.848
	MARITAL(3)	.671	.027	634.373	1	.000	1.956
	GRADE_B			3310.593	4	.000	
	GRADE_B(1)	.325	.027	149.869	1	.000	1.385
	GRADE_B(2)	-.002	.028	.007	1	.933	.998
	GRADE_B(3)	1.043	.031	1133.687	1	.000	2.839
	GRADE_B(4)	1.508	.062	590.861	1	.000	4.516
	Constant	1.010	.043	551.745	1	.000	2.746

a. Variable(s) entered on step 1: REG, MEMBERS, SEX, AGE, MARITAL, GRADE_B.

Classification Table[a]

Observed			Predicted		
			WK_7DAY		Percentage Correct
			ไม่มีงานทำ	มีงานทำ	
Step 1	WK_7DAY	ไม่มีงานทำ	22491	37239	37.7
		มีงานทำ	11948	100563	89.4
	Overall Percentage				71.4

❖ การวิเคราะห์การถดถอยเชิงเส้นพหุคูณของรายได้ของประชากรไทย

- ผลการวิเคราะห์ข้อมูลเบื้องต้น

เพศ

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ชาย	25811	53.8	53.8	53.8
	หญิง	22122	46.2	46.2	100.0
	Total	47933	100.0	100.0	

สถานภาพการสมรส

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	โสด	12391	25.9	25.9	25.9
	สมรส/อยู่ด้วยกัน	30766	64.2	64.2	90.0
	ม่าย	1489	3.1	3.1	93.1
	หย่าร้าง/แยกกันอยู่	3287	6.9	6.9	100.0
	Total	47933	100.0	100.0	

ระดับการศึกษา

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ประถมศึกษา/ต่ำกว่า	14796	30.9	30.9	30.9
	มัธยมศึกษา	15760	32.9	32.9	63.7
	ปริญญาตรี/เทียบเท่า	13281	27.7	27.7	91.5
	สูงกว่าปริญญาตรี/เทียบเท่า	1909	4.0	4.0	95.4
	ไม่ได้รับการศึกษา	2187	4.6	4.6	100.0
	Total	47933	100.0	100.0	

อาชีพ

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ข้าราชการ/เสมียน/ช่างเทคนิค	15117	31.5	31.5	31.5
	พนักงานบริการ/ผู้จัดจำหน่ายผลิตภัณฑ์	6923	14.4	14.4	46.0
	เกษตรกร/ประมง	2942	6.1	6.1	52.1
	ช่างฝีมือ/ผู้ควบคุมเครื่องจักร	22951	47.9	47.9	100.0
	Total	47933	100.0	100.0	

ตำแหน่งงาน

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ลูกจ้างรัฐบาล	12326	25.7	25.7	25.7
	ลูกจ้างเอกชน	32348	67.5	67.5	93.2
	ลูกจ้างทั่วไป	3259	6.8	6.8	100.0
	Total	47933	100.0	100.0	

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
จำนวนชั่วโมงในการทำงาน	47933	2	98	43.88	9.433
จำนวนสมาชิกในครัวเรือน	47933	1	18	3.53	1.915
รายได้ต่อเดือน	47933	200	250000	13616.17	11531.472
อายุ	47933	15	84	39.98	12.016
Valid N (listwise)	47933				

- คำสั่งที่ใช้ในการวิเคราะห์การถดถอยเชิงเส้นพหุคูณของรายได้ของประชากรไทย

```
setwd("D:/Thesis")
data_regress <- read.csv("DataAnalysisInR1Regression.csv",h=T)
attach(data_regress)
names(data_regress)
fs <- factor(s)
fm <- factor(m)
fe <- factor(e)
fo <- factor(o)
ft <- factor(t)
#สร้างโมเดล
model_regress <- lm(y2~fs+fm+fe+fo+ft+x1+x2+x3)
summary(model_regress)
#ข้อตกลงเบื้องต้น
windows()
par(mfrow=c(1,2))
plot(model_regress$fitted.values,model_regress$residuals,pch=16,ylab =
expression(e[i]),xlab =expression(hat(y[i])) )
qqnorm(model_regress$residuals,pch=16)
qqline(model_regress$residuals)
#ทดสอบการแจกแจงของข้อมูล
ks.test(model_regress$residuals,model_regress$fitted.values)
#ทดสอบความแปรปรวน
fv <- model_regress$fitted.values
summary(fv)
fv1 <- ifelse(fv <= summary(fv)[2],1,fv)
fv1 <- ifelse(fv > summary(fv)[2]&fv <= summary(fv)[3],2,fv1)
fv1 <- ifelse(fv > summary(fv)[3]&fv <= summary(fv)[5],3,fv1)
fv1 <- ifelse(fv > summary(fv)[5],4,fv1)
table(fv1)
fv.group <- factor(fv1)
bartlett.test(model_regress$residuals~fv.group)
#คำนวณค่า AIC,BIC
AIC(model_regress)
```

```
BIC(model_regress)
#คำนวณค่า RMSE
sqrt((sum((y2-model_regress$fitted.values)^2))/length(y2))
#########################
#แปลงข้อมูล
windows()
boxcox(model_regress)
ly <- log(y2)
#สร้างโมเดลของ log(y2)
model_regress_ly <- lm(ly~fs+fm+fe+fo+ft+x1+x2+x3)
summary(model_regress_ly)
#ข้อตกลงเบื้องต้น
windows()
par(mfrow=c(1,2))
plot(model_regress_ly$fitted.values,model_regress_ly$residuals,pch=16,ylab =
expression(e[i]),xlab =expression(hat(lny[i])) )
qqnorm(model_regress_ly$residuals,pch=16)
qqline(model_regress_ly$residuals)
#ทดสอบการแจกแจงของข้อมูล
ks.test(model_regress_ly$residuals,model_regress_ly$fitted.values
#ทดสอบความแปรปรวน
fv_2 <- model_regress_ly$fitted.values
summary(fv_2)
fv1_2 <- ifelse(fv_2 <= summary(fv_2)[2],1,fv_2)
fv1_2 <- ifelse(fv_2 > summary(fv_2)[2]&fv_2 <= summary(fv_2)[3],2,fv1_2)
fv1_2 <- ifelse(fv_2 > summary(fv_2)[3]&fv_2 <= summary(fv_2)[5],3,fv1_2)
fv1_2 <- ifelse(fv_2 > summary(fv_2)[5],4,fv1_2)
table(fv1_2)
fv_2.group <- factor(fv1_2)
bartlett.test(model_regress_ly$residuals~fv_2.group)
#คำนวณค่า AIC,BIC
AIC(model_regress_ly)
BIC(model_regress_ly)
#คำนวณค่า RMSE
sqrt((sum((ly-model_regress_ly$fitted.values)^2))/length(ly))
```

ภาคผนวก 3
ตัวอย่างแบบสอบถาม

ลับ

สำนักงานสถิติแห่งชาติ

สรง. 3
(แบบแจงนับ)

การสำรวจภาวะการทำงานของประชากร

ไตรมาสที่ 4 (ต.ค. - ธ.ค.) 2561

REG CWT 1-2

AMP TMB 4-7

8-11

AREA EA

VIL

PSU_NO

EA_SET SAMSET 20

MONTH_YR

HH_NO TYPE 22

MEMBERS

LISTING 160

1. ภาค..... จังหวัด.....

2. อำเภอ/เขต..... ตำบล/แขวง.....

3. บ้านเลขที่..... ถนน..... ตรอก/ซอย.....

4. ในเขตเทศบาล EA.....

นอกเขตเทศบาล EA.....หมู่ที่..... ชื่อหมู่บ้าน.....

5. ลำดับที่ EA ตัวอย่าง.....

6. ชุด EA ตัวอย่าง..... ชุดครัวเรือนตัวอย่าง.....

เดือน..... พ.ศ.....

7. ลำดับที่ครัวเรือนตัวอย่าง.....ประเภท(ส่วนบุคคล.....1 กลุ่มบุคคล(ระบุ)..... 2)

8. จำนวนสมาชิกในครัวเรือน ขั้นแจงนับ..... คน

9. จำนวนสมาชิกในครัวเรือน ขั้นนับจด..... คน

10. ชื่อผู้ตอบสัมภาษณ์..... เป็นสมาชิกในครัวเรือน.....1

ไม่เป็นสมาชิกในครัวเรือน.....2

WHO 31-32

ENUM

11. ผลการแจงนับครัวเรือนตัวอย่างนี้ (บันทึกรหัส)

ได้ข้อมูล		รหัส	ไม่ได้ข้อมูล		รหัส
ขั้นนับจด	ขั้นแจงนับ		ขั้นนับจด	ขั้นแจงนับ	
1. เป็นครัวเรือนตัวอย่าง			เป็นครัวเรือนตัวอย่าง	ไปสามครั้งไม่พบ	
1.1 มีครัวเรือนอาศัยอยู่	แจงนับได้	11		ผู้ตอบสัมภาษณ์	21
	รื้อถอน ไฟไหม้	12		ไม่ให้ความร่วมมือ	22
	เป็นที่ว่าง	13		หาบ้านไม่พบ	23
2. ไม่เป็นครัวเรือนตัวอย่าง (ครัวเรือนใหม่อยู่แทนครัวเรือนเดิมที่เป็นตัวอย่าง)	แจงนับได้	14		อื่น ๆ (ระบุ)	24

รหัสพนักงานแจงนับ

ID CODE1 200-206

รหัสผู้ตรวจสอบ

ID CODE2 207-21

2

ตอนที่ 1 ลักษณะทั่วไปของสมาชิกในครัวเรือน

ลำดับที่	ถามทุกคนในครัวเรือน				ถามเฉพาะผู้มีอายุ 15 ปีขึ้นไป
	ชื่อ-นามสกุล	ความเกี่ยวพันกับหัวหน้าครัวเรือน บันทึกรหัส หัวหน้าครัวเรือน.....1 ภรรยาหรือสามี.....2 บุตรที่ยังไม่สมรส.....3 บุตรที่สมรสแล้ว.....4 บุตรเขย บุตรสะใภ้.....5 บุตรของบุตร.....6 พ่อ แม่ พ่อตา แม่ยาย...7 ญาติอื่น ๆ .....8 ผู้อาศัย และคนรับใช้.....9 สมาชิกครัวเรือนกลุ่มบุคคล....0	เพศ บันทึกรหัส ชาย.....1 หญิง.....2	อายุ (บันทึกอายุเต็มปี ถ้าต่ำกว่า 1 ปี บันทึก " 00 ") ถ้าบันทึก 00 - 14 ปี ให้ข้ามไปถาม ตอนที่ 6	สถานภาพสมรส บันทึกรหัส โสด.....1 สมรส.....2 ม่าย.....3 หย่า.....4 แยกกันอยู่.....5 เคยสมรสแต่ไม่ทราบ สถานภาพ.....6
ก1	ก2	ก3	ก4	ก5	ก6

หมายเหตุ :

ตอนที่ 2 การศึกษา ถามเฉพาะผู้มีอายุตั้งแต่ 15 ปีขึ้นไป				ตอนที่ 3 การทำงา ถามเฉพาะผู้มีอายุตั้งแต่ 15		
" ... (ชื่อ) ... ขณะนี้กำลังเรียนชั้นใด" บันทึก ชั้น ปี สถาบันที่กำลังเรียนโดยละเอียด **ถ้าไม่เรียน** บันทึก **"ไม่เรียน"**	... (ชื่อ) ...จบการศึกษาสูงสุดระดับใด **บันทึก ชั้น ปี และประกาศนียบัตร หรือปริญญาที่สำเร็จ มาโดยละเอียด** **ถ้าจบ**มหาวิทยาลัย ฝึกหัดครู (สายวิชาการศึกษา) หรือคณะวิชาศึกษา ให้บันทึกสาขาวิชา คณะหรือแผนกและชื่อสถาบันที่เรียนจบ **ถ้าไม่เคยเรียน** บันทึก "ไม่เคย" **ถ้าจบ**มหาวิทยาลัย ฝึกหัดครู (สายวิชาการศึกษา) หรือคณะวิชาศึกษา ถามต่อไป **ถ้าจบการศึกษาระดับอื่น** ข้ามไปถามสตอมภ์ F11	ถามเฉพาะผู้ที่จบ "มหาวิทยาลัย" "ฝึกหัดครู" (สายวิชาการ) หรือ "อาชีวศึกษา" "ม.ปลาย" ในสตอมภ์ F8		" ระหว่าง 7 วัน ก่อนวันสัมภาษณ์ ... (ชื่อ) ... ได้ทำงานที่ ก่อให้เกิดรายได้/ เงินเดือน/ค่าจ้าง/ ผลกำไร/ผลตอบแทน หรือได้ทำงาน ในฐานะช่วยธุรกิจ ของครัวเรือน โดยไม่ได้รับค่าจ้าง หรือไม่ " บันทึกรหัสทำงาน.....1 (ข้ามไปถามคาชีพ สตอมภ์ F24) ไม่ทำงาน.....2 (ถามต่อไป)	ผู้บันทึก รหัส 2 ในสตอมภ์ F11 **"ถึงแม้ว่า ...(ชื่อ) ... ไม่ได้ทำงานระหว่าง 7 วันก่อนวันสัมภาษณ์ ยังได้รับค่าจ้าง/ เงินเดือนหรือผลประโยชน์อื่น ๆ หรือผลกำไรจากงาน หรือธุรกิจในระหว่างที่ไม่ได้ทำงานหรือไม่"** บันทึกรหัสได้รับ.....1 (ข้ามไปถามคาชีพ สตอมภ์ F24) ไม่ได้รับ.....2 (ถามต่อไป)	ผู้บันทึก รหัส 2 ในสตอมภ์ F12 **"... (ชื่อ) ... มีงาน/ธุรกิจกิจการที่จะกลับไปทำ หรือไม่"** (ผู้ช่วยธุรกิจฯ ให้ถือว่าไม่มีงาน/ธุรกิจหรือกิจการที่จะกลับไปทำ ให้บันทึกรหัส 2) บันทึกรหัสมี.....1 (ถามต่อไป) ไม่มี.....2 (ข้ามไปถามสตอมภ์ F15)
			จำนวนปีของหลักสูตร	**" ... (ชื่อ) ... การศึกษาสูงสุดที่สำเร็จอยู่ในสายวิชาใด"** บันทึกรหัสสายวิชาการ (สามัญ)....1 สายอาชีวศึกษา.....2 สายวิชาการศึกษา.....3		
F7	F8		F9	F10	F11	F12	F13
	ระดับการศึกษา	**สาขาวิชา**					

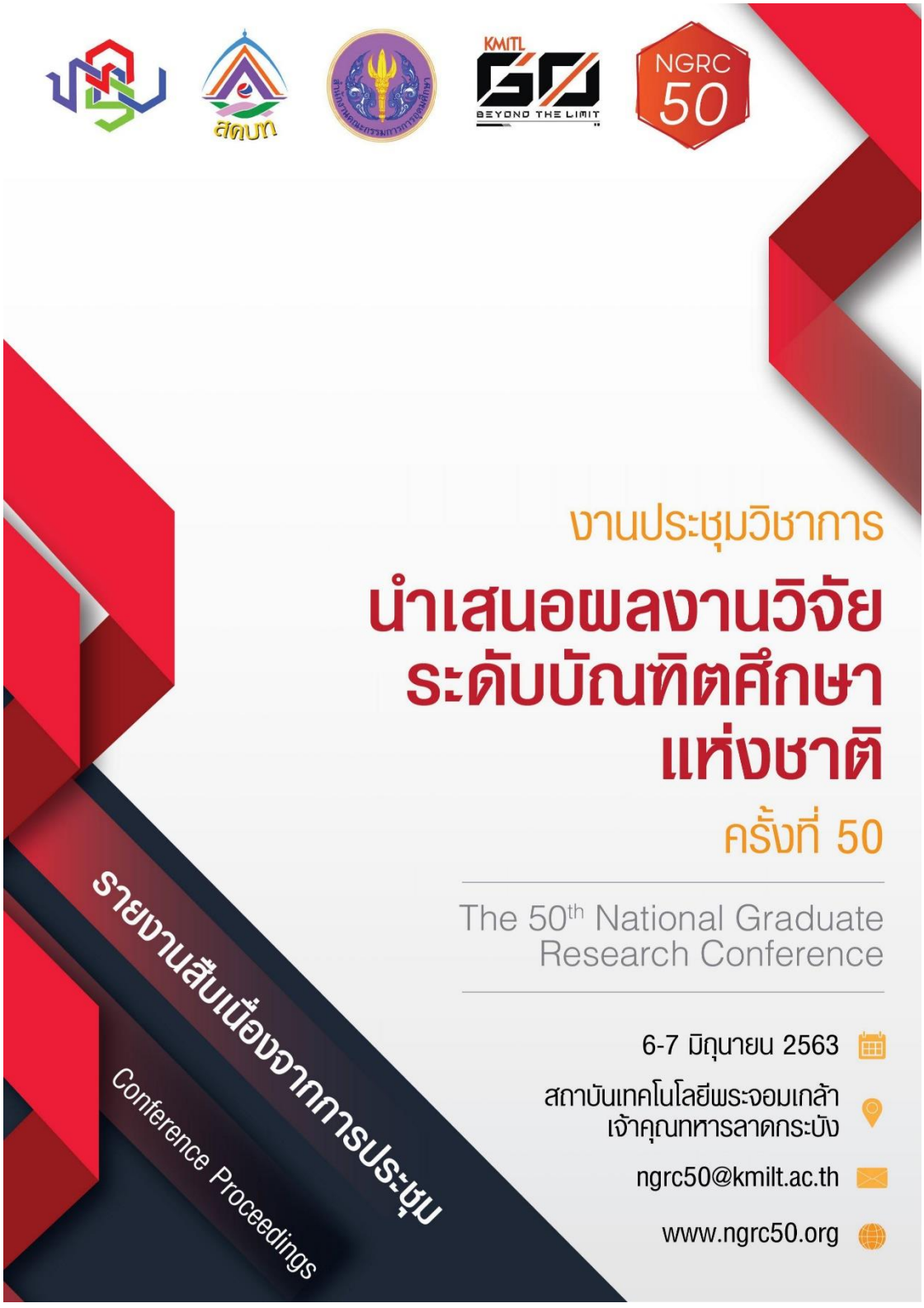
4

ตอนที่ 3 การทำงาน ถามเฉพาะผู้มีอายุตั้งแต่ 15 ปีขึ้นไป (ต่อ)

ผู้บันทึก รหัส 1 ในสดมภ์ F13	ผู้บันทึก รหัส 2 ในสดมภ์ F13	ผู้หางานทำระหว่าง 7 วัน ก่อนวันสัมภาษณ์ รหัส 1 ในสดมภ์ F15	ผู้บันทึก รหัส 2 หรือ 3 ในสดมภ์ F15	ผู้ไม่พร้อมทำงาน รหัส 2 ในสดมภ์ F17	ผู้พร้อมทำงาน รหัส 1 ในสดมภ์ F17	ผู้หางานทำ/พร้อมทำงาน ระหว่าง 7 วันก่อนวันสัมภาษณ์ (ผู้ว่างงาน) รหัส 1 ในสดมภ์ F15 หรือ F17	
" ... ชื่อ ... หยุดทำงานมานานเท่าใด"	"ก่อนวันสัมภาษณ์ ... (ชื่อ) ... ได้หางานทำ/สมัครงานบ้างหรือไม่ "	" ... (ชื่อ) ...ได้หางานทำ/สมัครงาน อย่างไร " (บันทึกวิธีการที่ใช้มากที่สุด)	" ระหว่าง 7 วัน ก่อนวันสัมภาษณ์ ... (ชื่อ) ...พร้อมทำงานหรือไม่"	" เหตุผลที่ ... (ชื่อ) ... ไม่พร้อมทำงาน"	"ทำไม ... (ชื่อ) ... จึงไม่หางานทำ/ไม่หาใน 7 วัน"	" ... (ชื่อ) ... หางาน/พร้อมทำงานมานานเท่าใด "	"... (ชื่อ) ... เคยทำงานมาก่อนหรือไม่"
บันทึกรหัส น้อยกว่า/เท่ากับ 2 เดือน.....1 มากกว่า 2 เดือน..2 (ข้ามไปถามอาชีพ สดมภ์ 1/24)	บันทึกรหัส ทำใน 7 วัน.....1 (ถามต่อไป) ทำใน 8-30 วัน..2 ไม่ทำงานทำ.....3 (บันทึกรหัส 2,3 ข้ามไปถาม สดมภ์ 1/17)	บันทึกรหัส อ่านหนังสือพิมพ์/นิตยสาร.....1 วิทยุ/โทรทัศน์.....2 เพื่อน/ญาติหางานให้.....3 ทบ/สมัครงานที่หน่วยจัดหางาน.4 ทบ/สมัครงานที่หน่วยงานโดยตรง.....5 ส่งจดหมาย/สมัครงานตามที่ต่าง ๆ .....6 (ข้ามไปถาม สดมภ์ 1/20) อื่น ๆ (ระบุ).....8	บันทึกรหัส พร้อม.....1 (ข้ามไปถาม สดมภ์ 1/19) ไม่พร้อม.....2 (ถามต่อไป)	บันทึกรหัส ทำงานบ้าน.....1 เรียนหนังสือ.....2 รอฤดูกาล.....3 ดูแลเด็กหรือสมาชิกผู้ป่วย พิการ ฯลฯ.....4 จนไม่สามารถทำงานได้.....5 พักผ่อน.....6 (ข้ามไปถามเงื่อนไขการทำงาน....7 อื่น ๆ (ระบุ).....8	บันทึกรหัส เชื่อว่าหางานทำไม่ได้....1 หมดหวังแล้วหาไม่ได้...2 ไม่ทราบว่าจะไปหางานอย่างไร.....3 ไม่สามารถหางานที่เหมาะสมได้.....4 ติดฝึกอาชีพ/ไม่มีเครื่องมือ.....5 อื่น ๆ (ระบุ).....6 (ถามต่อไป)	บันทึกรหัส ต่ำกว่า 1 เดือน.....1 1 - 2.9 เดือน.....2 3 - 5.9 เดือน.....3 6 - 8.9 เดือน.....4 9 - 11.9 เดือน....5 1 - 1.9 ปี.....6 2 - 2.9 ปี.....7 ตั้งแต่ 3 ปีขึ้นไป..8 (ถามต่อไป)	บันทึกรหัส เคย.....1 (ถามต่อไป) ไม่เคย.....2 (ข้ามไปถามตอนที่ 6)
F14	F15	F16	F17	F18	F19	F20	F21
☐	☐	☐	☐	☐	☐	☐	☐
☐	☐	☐	☐	☐	☐	☐	☐
☐	☐	☐	☐	☐	☐	☐	☐
☐	☐	☐	☐	☐	☐	☐	☐
☐	☐	☐	☐	☐	☐	☐	☐

ตอนที่ 4 ความต้องการจะทำงานเพิ่มในระหว่าง 7 วัน ก่อนวันสัมภาษณ์ (ต่อ)			ตอนที่ 5 รายได้ของลูกจ้าง		
ผู้มีงานทำ รหัส 1 ในสดมภ์ F11, F12 รหัส 1 หรือ 2 ในสดมภ์ F14			ค่าจ้าง ผู้มีงานทำที่มีสถานภาพการทำงานเป็นลูกจ้างทุกประเภท (ผู้ที่บันทึกรหัส 1 ในสดมภ์ F11, F12) (รหัส 1 หรือ 2 ในสดมภ์ F14) และมีสถานภาพการทำงานเป็นรหัส 4 หรือ 5 หรือ 6 หรือ 7 ในสดมภ์ F26		
ผู้มีงานทำที่มีเวลาและต้องการทำงานเพิ่ม รหัส 1 ในสดมภ์ F31					
"... (ชื่อ) ... ต้องการทำงานเพิ่มขึ้นสัปดาห์ละกี่ชั่วโมง" (บันทึกจำนวนเป็นชั่วโมง) (ถามต่อไป)	"... (ชื่อ) ... ได้หางานทำ/ติดต่อสมัครงานเพิ่มบ้างหรือไม่" บันทึกรหัส หา.....1 ไม่หา...2 (ถามต่อไป)	"... (ชื่อ) ... ทำไม ... (ชื่อ) ... ไม่ได้ทำงานเพิ่ม" บันทึกรหัส เครื่องจักร/อุปกรณ์ ในการปรับให้ทำงานเสีย.....1 ลดกำลังการผลิต/ลดเวลาทำงาน.....2 ฤดูระหว่างการประท้วง.....3 หางานทำ/เชื่อว่าหางานทำเพิ่มไม่ได้.....4 เจ็บป่วย.....5 อื่น ๆ (ระบุ).....6 (ผู้มีสถานภาพการทำงานเป็นรหัส 4, 5 หรือ 6 ในสดมภ์ F26 ถามต่อไป รหัส 1, 2, 3 หรือ 7 ข้ามไปถามสดมภ์ F34)	"งานที่ ... (ชื่อ) ... ทำอยู่ได้รับค่าจ้างในประเภทใด" บันทึกรหัส เป็นตัวเงิน :- รายชั่วโมง.....1 รายวัน.....2 รายสัปดาห์.....3 รายเดือน.....4 อื่น ๆ (ระบุ).....5 ไม่ทราบ.....6 ไม่เป็นตัวเงิน.....7 (ถ้าไม่ใช่ รหัส 1-3 ข้ามไปถามสดมภ์ F37)	สำหรับผู้บันทึกรหัส 1, 2, 3 ในสดมภ์ F35 "ค่าจ้างที่ได้รับนั้นได้รับกี่บาทต่อชั่วโมงหรือต่อวันหรือต่อสัปดาห์" (บันทึกจำนวนชิดขวา) (ถามต่อไป)	สำหรับผู้บันทึกรหัส 1-7 ในสดมภ์ F35 "ค่าจ้างที่ได้รับประมาณเดือนละเท่าไร" (บันทึกจำนวนชิดขวา) (ถามต่อไป)
F32	F33	F34	F35	F36	F37

ภาคผนวก 4
การตีพิมพ์เผยแพร่ผลงาน


KMITL
BEYOND THE LIMIT
NGRC
50
งานประชุมวิชาการ
นำเสนอผลงานวิจัย
ระดับบัณฑิตศึกษา
แห่งชาติ
ครั้งที่ 50
The 50th National Graduate
Research Conference
รายงานสืบเนื่องจากการประชุม
Conference Proceedings
6-7 มิถุนายน 2563
สถาบันเทคโนโลยีพระจอมเกล้า
เจ้าคุณทหารลาดกระบัง
ngrc50@kmilt.ac.th
www.ngrc50.org

Statistical Modeling to Study Employment Status of Thai Population

Ampawan Chankong[1] and Klairung Samart[1]

Abstract

Employment status and income of people in the country are indicators of how the economy of that country is going. The objective of this study is to fit an appropriate model to predict the employment status of the Thai population as well as to predict the income of the Thai population by using the logistic regression and multiple regression models, respectively. The analysis results showed that male (Odds ratio = 2.187) who live in central Thailand (Odds ratio = 1.231) and have post graduate studies (Odds ratio = 4.516) have a higher chance of getting a job compared to the general people. While the male who have post graduate studies, being government employee and having occupation (Public servant, Clerk and Technician), have a higher average income than general people.

Keyword : Employment status; average income; logistic regression; multiple regression; odds ratio

Introduction

Employment status and income of population are indicators that reflect socioeconomic status, economic growth as well as well-being of people in a country. There might be various factors affecting the employment status and income such as demographic factors, education, marital status, etc. Unemployment is of interest since it is associated with substantial socioeconomic burden and is the cause of other social problems (Cadden & Arnett, 2015).

Tapparak (2014) stated that factors affecting graduate employment are father's occupation, appearance, leadership, self-confidence and courageous expression. While Thuencumsan (2016) stated that factors affecting work conditions are sex, age, marital status, and monthly income. There are only different educational levels that affect the decision to choose a different job. Rukkhawattanakul (2015) revealed that the factors affecting household income change are number of household members, age, education, occupation and number of working hours.

According to the National Statistical Office (2019), in December 2019 the number of Thai people aged 15 years up is 56.70 million in total where 38.21 million are in the labor force, consisting of 37.66 million employed, 3.7 hundred thousand unemployed and 1.8 hundred thousand waiting for the season. When compared to December 2018, the employment has decreased by 2.1 hundred thousand (from 37.87 million to 37.66 million) whereas the unemployment rate has increased by 18 thousand (from 3.49 hundred thousand to 3.67 hundred thousand).

---

[1] Department of Mathematics and Statistics, Faculty of Science, Prince of Songkla University, Songkhla 90110, Thailand.

Considering the unemployment rate by gender, it was found that males had a higher unemployment rate than females, with males having an unemployment rate of 1.1 percent, and females had an unemployment rate of 0.8 percent. While the unemployment rate of those who completed Tertiary education amounts to 1.48 hundred thousand, followed by high school or equivalent to 7.6 ten thousand, primary education is 6.6 ten thousand, lower secondary school level of 5.5 ten thousand and no education 2.1 ten thousand. In addition, when comparing the number of unemployed in December 2018 and 2019 classified by region, the number of unemployed people increased by 2.7 ten thousand in the northeast region, with the middle and northern regions increasing equally to 1.0 ten thousand while Bangkok dropped by 2.7 ten thousand and the southern region reduced by 2.0 thousand.

These attentions have driven us to study the factors affecting employment status as well as the income of Thai population by using appropriate statistical models. The results of this study could be useful for those whom responsible for policy making for better employment and therefore better socioeconomic status and higher economic growth.

Objectives

The objective of this study is to fit an appropriate model to predict the employment status of the Thai population as well as to predict the income of the Thai population.

Materials and Methods

Participants

This research is quantitative research. The secondary data from the fourth quarter of "2018 Thai Labor Status Survey Project" collected by the National Statistical Office was used in the analysis. The total sample to study the factors affecting the employment status was 172,241. However, to study the factors affecting the income of employed population, the sample size decreased to 47,933. The sample size of the latter one was decreased because the focus was only on employed population and there were only 47,933 people provided their income information.

Statistical analyses

The aims of this study were to examine the factors that affect the employment status and the income of Thai population. As the employment status is binary, logistic regression was fitted to the data where region, sex, education level, marital status, age and number of household members were treated as baseline predictors. To examine the factors affecting the income of employed population, multiple regression was employed where sex, education level, marital status, age, number of household members, number of working hours, occupation and job title were treated as baseline predictors. Descriptive and inferential statistics were conducted using SPSS version 20 and R studio with the significance level at 0.05.

Results/ Conclusion

From the descriptive analysis of the initial data, we found that the average age of the overall sample is 47.61 years. Most of them were female 53.5%, completed primary education (≤ grade 8) 49.1%, were married/partnered 63.7%, and were living in the central region of Thailand 29.9%. The proportion of employed people (65.3%) is greater than unemployed (34.7%).

Logistic regression analysis of employment status of Thai population.

The results of the logistic regression analysis to find factors that can predict the employment status of the Thai population are shown in Table 1.

The results displayed in Table 1 reveal that when comparing the employment opportunities of those working in Bangkok with those working in other regions, it is found that those working in Central Thailand (Odds ratio = 1.231), Northern Thailand (Odds ratio = 1.198), North Eastern Thailand (Odds ratio = 1.151) and Southern Thailand (Odds ratio = 1.187) have a chance to get a job higher than those living in Bangkok. Obviously, this is because Bangkok is the most populated city in Thailand and therefore the job seeking is much more competitive compared to the other regions.

Unsurprisingly, when we compare by sex groups, it is found that men are 2.187 times more likely to get a job than women. And when consider by education level we found that the chance of getting a job is higher for people with higher education. The opportunity of employment at post graduate level is 4.516 times higher than those without education, while the chance of getting a job at high school level (Odds ratio = .998, p-value = .933) is not significantly different to those who do not have an education.

Table 1: Estimated coefficients of the selected regression model for the logistic regression analysis of employment status of Thai population

Variable	B	p-value	Odds ratio Exp(B)	95% C.I.For Exp(B)	
				Lower	Upper
Region					
• Bangkok	(ref)	.000	-	-	-
• Central Thailand	.208	.000	1.231	1.170	1.296
• Northern Thailand	.181	.000	1.198	1.137	1.263
• North Eastern Thailand	.140	.000	1.151	1.093	1.212
• Southern Thailand	.172	.000	1.187	1.125	1.253
Sex					
• Female	(ref)	.000	-	-	-
• Male	.782	.000	2.187	2.138	2.237

Education level					
• No education	(ref)	.000	-	-	-
• ≤ 8th Grade	.325	.000	1.385	1.314	1.459
• High school	-.002	.933	.998	.944	1.054
• Bachelor degree/College 4-year degree	1.043	.000	2.839	2.671	3.016
• Post graduate studies	1.508	.000	4.516	3.999	5.099
Marital status					
• Single	(ref)	.000	-	-	-
• Married/partnered	1.766	.000	5.848	5.657	6.046
• Widowed	.671	.000	1.956	1.856	2.061
• Divorced/separated	1.820	.000	6.173	5.814	6.554
Age	-.043	.000	.958	.957	.959
Number of household members	-.103	.000	.902	.896	.907
Constant	1.010	.000	2.746	-	-

Marital status is another important factor to employment opportunity. When comparing the employment opportunity of those who are single with those who are not single, it is found that those who are married (Odds ratio = 5.848), divorced/separated (Odds ratio = 6.173) and widowed (Odds ratio = 1.956) have a chance to get a job higher than those who are single. These results are ambiguous, however, since the single persons in the sample maybe young generation who just graduated at some levels and therefore still haven't yet got a job. This doesn't mean that they are unemployed because they are single.

In addition, age and number of household members also affect employment opportunity. As the age increases by 1 year, the chance of getting a job will decrease by 4.2%. While the number of household members increases by 1 person, the chance of getting a job will also decrease by 9.8%.

Multiple regression analysis of income of Thai population

The best model used to predict the income of the Thai population from stepwise regression and the Box-Cox transformation was found to be in the logarithm form of income as follows:

$$ln(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 s_1 + \beta_5 m_1 + \beta_6 m_2 + \beta_7 m_3 + \beta_8 e_1 + \beta_9 e_2 + \beta_{10} e_3 + \beta_{11} e_4 + \beta_{12} o_1 + \beta_{13} o_2 + \beta_{14} o_3 + \beta_{15} t_1 + \beta_{16} t_2$$

where y : Income of Thai people (baht/month)

x_1 : Age

x_2 : Number of household members

x_3 : Number of working hours

s_1 : Sex-Female

m_1 : Marital status-Married/partnered

m_2 : Marital status-Widowed

m_3 : Marital status-Divorced/separated

e_1 : Education level-High school

e_2 : Education level-Bachelor degree/College 4-year degree

e_3 : Education level-Post graduate studies

e_4 : No education

o_1 : Occupation- Service staff and Product supplier

o_2 : Occupation- Agriculturist and Fisheries

o_3 : Occupation-Craftsman and Machine operator

t_1 : Job title-Private employees

t_2 : Job title-General employees

The respective performances of the proposed models were compared by using the Root Mean Square Error (RMSE) , Akaike Information Citerion (AIC), Baysian Information Criterion (BIC) and Multiple Coefficient of Determination( ), although not reported here, the best model performance is shown in Table 2.

Table 2: The performance of the income predicting model of Thai population

Model	R^2	*RMSE*	*AIC*	*BIC*
$\ln(y)$	0.5876	0.4275	54606.44	54764.44

According to Table 2, the model for predicting income of the Thai population by using the model provides the lowest RMSE, AIC and BIC values, and has the highest values compared to other models analyzed.

Multiple regression analysis results for predicting income of Thai population are illustrated in Table 3.

Table 3: Estimated coefficients, standard errors, t statistics, and p-values of the selected regression model for predicting income of Thai population

Variable	B	Std. Error	t	p-value
Constant	8.492	.016	519.098	.000
x_1	.008	.000	38.173	.000
x_2	-.021	.001	-20.133	.000
x_3	.015	.000	63.053	.000
s_1	-.148	.004	-36.008	.000
m_1	.076	.005	14.944	.000
m_2	-.098	.013	-7.747	.000
m_3	.012	.009	1.369	.171
e_1	.213	.005	39.036	.000
e_2	.648	.007	89.032	.000
e_3	1.155	.012	94.595	.000
e_4	-.012	.009	-1.160	.246
o_1	-.270	.007	-36.979	.000
o_2	-.622	.010	-59.351	.000
o_3	-.329	.007	-48.581	.000
t_1	-.154	.006	-26.186	.000
t_2	-.498	.009	-51.714	.000

Note: The interpretation of the coefficient must change the coefficient to be in the form of exponential. Dummy variable referenced by Sex(Male), Marital status(Single), Education level($\leq$ 8th Grade), Occupation(Public servant, Clerk and Technician) and Job title(Government employees).

From Table 3, the multiple regression model can be written as

$$ln(y) = 8.492 + .008x_1 - .021x_2 + .015x_3 - .148s_1 + .076m_1 - .098m_2 + .012m_3 + .213e_1 + .648e_2 + 1.155e_3 - .012e_4 - .270o_1 - .622o_2 - .329o_3 - .154t_1 - .498t_2 \quad (1)$$

From model (1), we see that as age and working hours increases by one each where other variables are kept constant, the average income will increase by approximately one baht/month. However, if the number of family members increases by one, the average income will decrease by approximately one baht/month. Moreover, we found that females earn less on average per month than males by 0.86 baht/month.

When consider the marital status, the average single people income is less than those who are married /partnered and divorced/separated by 1.08 and 1.01 baht/month respectively, while the average income of a widow is less than single people by 0.91 baht/month.

For the education level, the average income of those with primary education (≤ 8th grade) is less than those with high school, Bachelor degree/College 4-year degree and Post graduate studies by 1.24, 1.91 and 3.17 baht/month, respectively. Although those with primary education have an average income more than those without education, it is not significant.

For the occupation type, public servant, clerk and technician earn an average income higher than service staff and product supplier by 0.76 baht/month, agriculturist and fisheries by 0.54 baht/month and craftsman and machine operator by 0.71 baht/month.

Lastly, government employees earn an average income higher than private employees and general employees by 0.86 and 0.61 baht/month respectively.

Conclusions

The main focus of this study is the factors that affect the employment opportunity, which found that people living in the central region of Thailand have a higher chance of getting a job than other regions, while men have a chance to get a job 2.187 times higher than women. For those who graduate with post graduate studies will have a chance to get a job of 4.516 times higher than those without education, and those who divorced/separated have a chance to get a job of 6.173 times compared to single people. In addition, the increasing age and increasing number of household members also affect the chance of getting a job which will decrease by 4.2% and 9.8% respectively.

From the analysis of factors affecting the income of the Thai population, it is found that the Thai population has an average income of 4875.61 baht / month. Increased age and increased working hours affect the average income increased. While the number of household members increased, it results in a lower average income. Males have a higher income than females. While those who married/partnered, post graduate studies, occupation (Public servant, Clerk and Technician) and government employees will have the highest average income.

Discussion

The result of this study shows that the factors that affect the employment status of Thai population are age, sex and marital status which is consistent with Thuencumsan (2016). It also found that regions, levels of education and number of household members also have an impact on employment status of the Thai population. In addition, the study of the factors affecting the income of the Thai population, it is found that the number of household members, age, education, occupation and the number of working hours affect the income, which is consistent with Rukkhawattanakul (2015), while gender, marital status and job title also affect income.

There is a limitation in this study. As we used the secondary data from the National Statistical Office where the data were already collected, we could not control number of responses in the survey. In some variables such as income of samples, there were a lot of missing values. Therefore, we could only use the provided information in our analysis. If it is possible, all information should be collected as much as possible for better analysis results. Moreover, although the stepwise regression and Box-Cox transformation were performed to find the best fit model, the assumptions of homogeneity and normality were still not met since there are a lot of possible outliers in this dataset.

Acknowledgement

This research was supported by a grant from Prince of Songkla University "Scholarship for outstanding GPA", contract no. SOGPA16/2561.

References

Cadden M. & Arnett P. 2015. Factors Associated with Employment Status in Individuals with Multiple Sclerosis. *Int. J. MS Care*, 17, 284-291.

National Statistical Office. 2019. Report of the Labour Survey December 2019. NSO Publication. Bangkok.

Rukkhawattanakul O. 2015. Factors causing personal income variation in the Nongkhai, Mukdahan and Mae Sot special economic zones, Thailand. Retrieved from http://ethesisarchive.library.tu.ac.th/thesis/2015/TU_2015_5704010700_4720_2858.pdf

Tapparak P. 2014. A study of job seeking behavior that affecting Employment of University Graduates Technology Srivijaya Songkhla Campus. *Journal of Education Research Faculty of Education, Srinakharinwirot University*, 8, 154-160.

Thongchumnum P, Suwanro S. & Choonpradub C. 2008. Demographic Factors Affecting Employment in Pattani and Songkla Provinces of Thailand. *Asian Soc. Sci*., 4, 169-176.

Thuencumsan P. 2016. Factors influencing to the motivation in decision making of operational-level employees in Hemaraj Eastern Seaboard Industrial Estate, Rayong Province. *MUT Journal of Business Administration*, 13, 1-23.