การวิเคราะห์การถดถอยพหุคูณ

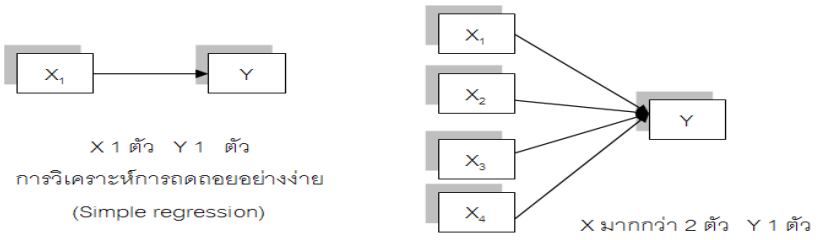
Multiple regression Analysis

สมประสงค์ เสนารัตน์[1]

การวิเคราะห์ถดถอยพหุคูณ (Multiple regression Analysis) เป็นวิธีการวิเคราะห์ข้อมูลเพื่อหาความสัมพันธ์ระหว่างตัวแปรตาม (Y) หรือตัวแปรเกณฑ์ (Criterion Variable) จำนวน 1 ตัว กับตัวแปรอิสระ (X) หรือตัวแปรพยากรณ์ หรือตัวแปรทำนาย (Predictor Variable) ตั้งแต่ 2 ตัวขึ้นไป เป็นเทคนิคทางสถิติที่อาศัยความสัมพันธ์เชิงเส้นตรงระหว่างตัวแปรมาใช้ในการทำนาย โดยเมื่อทราบค่าตัวแปรหนึ่งก็สามารถทำนายอีกตัวแปรหนึ่งได้ ผลการวิเคราะห์ข้อมูลที่ได้อยู่ในรูปของสมการทำนาย สิ่งสำคัญที่ต้องการหาในการวิเคราะห์การถดถอยพหุคูณ คือ สัมประสิทธิ์สหสัมพันธ์พหุคูณ สมการพยากรณ์ในรูปคะแนนดิบ หรือในรูปคะแนนมาตรฐาน หรือทั้งคู่ และความคลาดเคลื่อนมาตรฐานในการพยากรณ์ (บุญชม ศรีสะอาด. 2547:141) บทความนี้จะนำเสนอเป็น 2 ส่วน คือ ส่วนแรกเป็นแนวคิดเกี่ยวกับการวิเคราะห์การถดถอยพหุคูณ และในส่วนที่สองจะเป็นการนำเสนอการวิเคราะห์การถดถอยพหุคูณด้วยโปรแกรม R

กรอบแนวคิดในการวิเคราะห์การถดถอย

การวิเคราะห์การถดถอยอาศัยความสัมพันธ์ระหว่างตัวแปรเป็นหลัก และคำตอบที่ต้องการคือ มีตัวแปรใดบ้างที่สามารถพยากรณ์ตัวแปรเกณฑ์ที่เราสนใจจะศึกษา และตัวแปรใดพยากรณ์ได้มากน้อยกว่ากัน รวมทั้งส่งผลในทางบวกหรือทางลบ ซึ่งการวิจัยในลักษณะนี้จะต้องอาศัยการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง (Review Literature) มาเป็นอย่างดี และสรุปเป็นกรอบแนวคิดในการทำวิจัย และนำไปสร้างเครื่องมือตามกรอบแนวคิดที่กำหนดไว้ สามารถเขียนเป็นกรอบแนวคิดดังภาพประกอบ 1



ภาพประกอบ 1 กรอบแนวคิดการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรเกณฑ์ (Y) กับตัวแปรพยากรณ์ (X)

---

[1] อาจารย์ประจำวิทยาลัยการศึกษา มหาวิทยาลัยราชภัฏร้อยเอ็ด er2_somprasong@windowslive.com

จุดประสงค์ของการวิเคราะห์การถดถอยพหุคูณ

การวิเคราะห์การถดถอยมีจุดประสงค์ คือ เพื่อสร้างสมการพยากรณ์ตัวแปรเกณฑ์ด้วยกลุ่มตัวแปรพยากรณ์

ระดับข้อมูลของตัวแปร

ตัวแปรเกณฑ์ และตัวแปรพยากรณ์ อยู่ในมาตราอันตรภาคชั้น (Interval Scale) หรือมาตราวัดอัตราส่วน (Ratio Scale) ในกรณีที่ตัวแปรพยากรณ์ไม่เป็นไปตามมาตรวัดข้างต้น ให้แปลงข้อมูลเป็นตัวแปรหุ่น (Dummy variable) ก่อนที่จะทำการวิเคราะห์ข้อมูล

ข้อตกลงเบื้องต้นของการวิเคราะห์การถดถอยพหุคูณ

1. Normality ประชากรมีการแจกแจงแบบปกติ (Normality) ตรวจสอบได้โดยการดูกราฟ หรือวิธีการทางสถิติ เช่น เช่นใช้ Kolmogorov-Smirnov Test ในกรณีที่ไม่ทราบค่าเฉลี่ยและความแปรปรวนของประชากร จะใช้ค่าเฉลี่ยและความแปรปรวนของกลุ่มตัวอย่างแทน หรือใช้ Shapiro-Wilk Test ในกรณีที่ทราบหรือไม่ทราบค่าเฉลี่ยและความแปรปรวนของประชากรก็ได้ แต่กลุ่มตัวอย่างต้องมีขนาดไม่เกิน 50 (กัลยา วานิชย์บัญชา. 2546: 220) หรือ Lilliefors Test ซึ่งเป็นวิธีการที่ปรับปรุงมาจากวิธีของ Kolmogorov-Smirnov แต่จะให้ค่าความน่าจะเป็นในการทดสอบน้อยกว่าวิธีของ Kolmogorov-Smirnov (ศิริชัย พงษ์วิชัย. 2551: 170)

2. Linearlity ตัวแปรพยากรณ์กับตัวแปรเกณฑ์มีความสัมพันธ์กันเชิงเส้นตรง (Linearity) ตรวจสอบได้โดยใช้วิธีการทางสถิติ เช่น ดูจากค่าสัมประสิทธิ์สหสัมพันธ์ r_{xy}

3. Homoscedasticity หมายถึง ความแปรปรวนของความคลาดเคลื่อนมีความคงที่ทุกค่าการสังเกต (ทรงศักดิ์ ภูสีอ่อน. 2551: 281) ตรวจสอบได้โดยการดูจากกราฟ หรือใช้วิธีการทางสถิติ เช่น Non-constant Variance Score Test หรือ The Spearman rank-correlation test หรือ The Goldfeld and Quarndt test หรือ White's test

4. ตัวแปรที่นำมาใช้พยากรณ์ต้องไม่มีปัญหาเรื่อง Multicollinearity หมายถึง ตัวแปรที่นำมาใช้พยากรณ์ไม่ควรมีความสัมพันธ์กันสูงเกินไป (ทรงศักดิ์ ภูสีอ่อน. 2551: 280) ตรวจสอบได้ด้วยการดูกราฟ หรือด้วยวิธีการทางสถิติ เช่น ดูจากค่าสัมประสิทธิ์สหสัมพันธ์ r_{xy} หรือดูจากค่า variance inflation factors (VIF)

สถิติที่ใช้ในการวิเคราะห์การถดถอยพหุคูณ

1. สัมประสิทธิ์สหสัมพันธ์ คำนวณด้วยสูตรของเพียร์สัน (r_{xy}) (สมบัติ ทายเรือคำ. 2551: 145)

$$r_{xy} = \frac{n\sum XY - (\sum X)(\sum Y)}{\sqrt{[\, n\sum X^2 - (\sum X)^2 \,]\,[\, n\sum Y^2 - (\sum Y)^2 \,]}}$$

2. ค่า b หาจากสูตร (บุญชม ศรีสะอาด. 2547: 153)

$$b_j = \beta_j \frac{S_Y}{S_j}$$

เมื่อ b_j แทน ค่าน้ำหนักคะแนนหรือสัมประสิทธิ์การถดถอยของตัวพยากรณ์ (ตัวแปรอิสระ) ตัวที่ j ที่ต้องการหาค่าน้ำหนักคะแนน

β_j แทน ค่าน้ำหนักเบต้า ของตัวพยากรณ์ (ตัวแปรอิสระ) ตัวที่ j

S_Y แทน ส่วนเบี่ยงเบนมาตรฐานของตัวเกณฑ์ (ตัวแปรตาม)

S_j แทน ส่วนเบี่ยงเบนมาตรฐานของตัวพยากรณ์ (ตัวแปรอิสระ) ตัวที่ j

3. ค่า β หาจากสูตร (สมบัติ ทายเรือคำ. 2545: 42)

$$\beta_j = b_j \frac{S_j}{S_y}$$

4. สหสัมพันธ์พหุคูณ (Multiple Correlation) แทนด้วย R คำนวณหาค่า R โดยใช้สูตร (สมบัติ ท้ายเรือคำ. 2546: 41)

$$R = \sqrt{\beta_1 r_{1y} + \beta_2 r_{2y} + \ldots + \beta_p r_{py}}$$

5. การทดสอบนัยสำคัญทางสถิติของสัมประสิทธิ์สหสัมพันธ์พหุคูณ (หรือสัมประสิทธิ์การถดถอย) ทดสอบโดยใช้สถิติ F จากสูตร (บุญชม ศรีสะอาด. 2547: 163)

$$F = \frac{R^2/k}{\left(1-R^2\right)/(N-k-1)}$$

เมื่อ F แทน ค่าสถิติที่จะใช้เปรียบเทียบกับค่าวิกฤตจากการแจกแจงแบบ F เพื่อทราบความมีนัยสำคัญของ R

R แทน สัมประสิทธิ์สหสัมพันธ์พหุคูณ

N แทน จำนวนสมาชิกกลุ่มตัวอย่าง

k แทน จำนวนตัวพยากรณ์ (ตัวแปรอิสระ)

6. การทดสอบนัยสำคัญของตัวแปรที่เพิ่มเข้ามาในสมการการถดถอย มีสูตรการทดสอบนัยสำคัญดังนี้ (บุญชม ศรีสะอาด. 2547: 167-168)

$$F = \frac{\left(R^2_{Y.12...I} - R^2_{Y.12...k}\right)/(1-k)}{\left(1-R^2_{Y.12...I}\right)/(N-1-1)}$$

เมื่อ F แทน ค่าสถิติที่จะใช้เปรียบเทียบกับค่าวิกฤตจากการแจกแจงแบบ F เพื่อทราบความมีนัยสำคัญ

$R_{Y.12...k}$ แทน กำลังสองของสัมประสิทธิ์สหสัมพันธ์พหุคูณสำหรับการถดถอยของ Y บนตัวแปร k ตัว

$R_{Y.12...I}$ แทน กำลังสองของสัมประสิทธิ์สหสัมพันธ์พหุคูณสำหรับการถดถอยของ Y บนตัวแปร k ตัว

k แทน จำนวนของตัวพยากรณ์ที่มีจำนวนน้อยกว่า

I แทน จำนวนของตัวพยากรณ์ที่มีจำนวนมากกว่า

7. ความคลาดเคลื่อนมาตรฐานของการพยากรณ์ (Standard errors of estimate) เขียนแทนด้วยตัวย่อ SE_{est} สูตรในการหา SE_{est} ดังนี้ (บุญชม ศรีสะอาด. 2547: 169)

$$SE_{est} = \sqrt{\frac{SS_{res}}{N-k-1}}$$

เมื่อ SE_{est} แทน ความคลาดเคลื่อนมาตรฐานของการพยากรณ์

SE_{res} แทน ผลรวมของกำลังสอง (Sum of Squares) ของส่วนที่เหลือ (ของ Residual) = $\sum d^2$

8. การหาค่าความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์การถดถอย (Standard errors of b coefficients) เขียนแทนด้วยตัวย่อ SE_{bj} สามารถคำนวณได้หลายวิธี สูตรที่นิยมได้แก่ (บุญชม ศรีสะอาด. 2547: 170)

$$SE_{bj} = \sqrt{\frac{SE_{est}^2}{SS_{xj}\left(1-R_j^2\right)}}$$

เมื่อ SE_{bj} แทน ความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์การถดถอย (ของ b)

SE_{est}^2 แทน กำลังสองของความคลาดเคลื่อนมาตรฐานของการพยากรณ์

SS_{xj} แทน ผลรวมกำลังสองของความเบี่ยงเบน (Sum of Squares) ของตัวพยากรณ์ตัวที่ j

R_j^2 แทน กำลังสองสัมประสิทธิ์สหสัมพันธ์พหุคูณระหว่างตัวพยากรณ์ตัวที่ j

9. การทดสอบนัยสำคัญของสัมประสิทธิ์การถดถอย เพื่อทดสอบว่าตัวแปรพยากรณ์แต่ละตัวส่งผลต่อการทำนายตัวเกณฑ์หรือไม่ ทดสอบโดยใช้สูตร (บุญชม ศรีสะอาด. 2547: 171)

$$t_j = \frac{b_j}{SE_{bj}}$$

เมื่อ t_j แทน ค่าสถิติที่จะใช้เปรียบเทียบกับค่าวิกฤตจากการแจกแจงแบบ t เพื่อทดสอบนัยสำคัญทางสถิติ

b_j แทน สัมประสิทธิ์การถดถอยของตัวพยากรณ์ที่ j ที่ต้องการทดสอบนัยสำคัญ

SE_{bj} แทน ความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์การถดถอย

สมการพยากรณ์ในรูปคะแนนดิบ

สมการณ์เชิงเส้นตรงในรูปคะแนนดิบ (บุญชม ศรีสะอาด. 2547: 143)

$$Y' = a + b_1x_1 + b_2x_2 + ... + b_kx_k$$

เมื่อ Y' แทน คะแนนพยากรณ์ของตัวเกณฑ์ (ตัวแปรตาม)

a แทน ค่าคงที่ของสมการพยากรณ์ในรูปแบบคะแนนดิบ

b_1 , b_2,... b_k แทน น้ำหนักคะแนนหรือสัมประสิทธิ์การถดถอยของตัวพยากรณ์ตัวที่ 1 ถึง ตัวที่ k ตามลำดับ

x_1 , x_2 , x_3 แทน คะแนนของตัวพยากรณ์ (ตัวแปรอิสระ) ตัวที่ 1 ถึง ตัวที่ k

k แทน จำนวนตัวพยากรณ์ (ตัวแปรอิสระ)

สมการณ์พยากรณ์ในรูปคะแนนมาตรฐาน

ถ้าต้องการพยากรณ์เกณฑ์ในรูปของคะแนนมาตรฐาน เขียนสมการพยากรณ์ได้ดังนี้ (บุญชม ศรีสะอาด. 2547: 144)

$$Z'_Y = \beta_1 Z_1 + \beta_2 Z_2 + \ldots + \beta_k Z_k$$

เมื่อ Z'_Y แทน คะแนนพยากรณ์ในรูปของคะแนนมาตรฐานของตัวเกณฑ์ (ตัวแปรตาม)

β_1 , $\beta_2 \ldots \beta_k$ แทน สัมประสิทธิ์การถดถอยในรูปของคะแนนมาตรฐานของตัวพยากรณ์ตัวที่ 1 ถึง ตัวที่ k ตามลำดับ

Z_1, $Z_2 .. Z_k$ แทน คะแนนมาตรฐานของตัวพยากรณ์ (ตัวแปรอิสระ) ตัวที่ 1 ถึง ตัวที่ k ตามลำดับ

K แทน จำนวนตัวพยากรณ์

การคัดเลือกตัวแปรเพื่อการพยากรณ์

วิธีคัดเลือกตัวแปรพยากรณ์เข้าสู่สมการถดถอยมีอยู่หลายวิธีด้วยกัน เช่น

1. การวิเคราะห์ถดถอยพหุคูณแบบปกติ (Enter Regression)
2. การวิเคราะห์ถดถอยพหุคูณแบบคัดเลือกออก (Remove Regression)
3. การวิเคราะห์ถดถอยพหุคูณแบบเดินหน้า (Forward Regression)
4. การวิเคราะห์ถดถอยพหุคูณแบบถอยหลัง (Backward Regression)
5. การวิเคราะห์ถดถอยพหุคูณแบบขั้นบันได (Stepwise Regression)

วิธีที่มักจะนำมาใช้กัน คือ วิธีการวิเคราะห์ถดถอยพหุคูณแบบปกติ (Enter Regression) และวิธีการวิเคราะห์ถดถอยพหุคูณแบบขั้นบันได (Stepwise Regression) โดยที่ทั้งสองวิธีตอบสนองความต้องการของผู้ใช้แตกต่างกันไปตามวัตถุประสงค์ของการใช้ประโยชน์ คือ วิธีการวิเคราะห์ถดถอยพหุคูณแบบปกติ ผู้ใช้สามารถเลือกตัวแปรทำนายเข้าสู่สมการได้เอง ส่วนการวิเคราะห์ถดถอยพหุคูณแบบขั้นบันได (Stepwise Regression) นำเสนอสมการตัวแปรพยากรณ์ที่ดีที่สุดเพื่อนำไปใช้ในการทำนาย

ขั้นตอนการวิเคราะห์

1. ตรวจสอบข้อตกลงเบื้องต้น
2. คำนวณค่า r_{xy} ของตัวแปรพยากรณ์กับตัวแปรเกณฑ์
3. คัดเลือกตัวแปรพยากรณ์ที่มีความสัมพันธ์สูงสุดกับตัวแปรเกณฑ์เข้าสมการ และคำนวณค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณ (R)

4. ทดสอบนัยสำคัญของสัมประสิทธิ์สหสัมพันธ์พหุคูณ (R) ว่าทดสอบว่าตัวแปรพยากรณ์ที่เข้าในสมการยังคงอยู่ในสมการต่อไปได้หรือไม่ด้วยสถิติ F

5. หาค่าน้ำหนักความสำคัญของตัวแปรพยากรณ์ (b) หรือ β หรือทั้งสองอย่าง เพื่อนำมาใช้ในการเขียนสมการพยากรณ์ และเปรียบเทียบว่าตัวแปรพยากรณ์ตัวใดพยากรณ์ตัวแปรเกณฑ์ได้ดีกว่า

6. ทดสอบนัยสำคัญของสัมประสิทธิ์การถดถอย เพื่อตรวจสอบว่าตัวแปรพยากรณ์สามารถพยากรณ์ตัวแปรเกณฑ์ได้หรือไม่ ด้วยสถิติ t

7. คำนวณค่าความคลาดเคลื่อนมาตรฐานของตัวแปรพยากรณ์ที่เข้าสมการ (SE_b) และคำนวณค่าความคลาดเคลื่อนมาตรฐานของการพยากรณ์ (SE_{est})

8. คัดเลือกตัวแปรพยากรณ์ที่มีความสัมพันธ์สูงกับตัวแปรเกณฑ์รองลงมาเข้าสมการและทำการทดสอบค่าสัมประสิทธิ์สหสัมพันธ์ที่เปลี่ยนแปลง (R^2 change) ด้วยสถิติ F ถ้า R^2 change ไม่มีนัยสำคัญก็แสดงว่าตัวแปรพยากรณ์ไม่สามารถอยู่ในสมการพยากรณ์ได้ แต่ถ้ามีนัยสำคัญก็ดำเนินการตามข้อ 4, 5, 6 และ 7 และดำเนินการต่อไปจนกว่าจะไม่มีตัวแปรพยากรณ์ใดเข้าในสมการ (การดำเนินการตามข้อ 8 เป็นวิธีการวิเคราะห์ถดถอยพหุคูณแบบขั้นบันได)

การวิเคราะห์การถดถอยพหุคูณด้วยโปรแกรม R

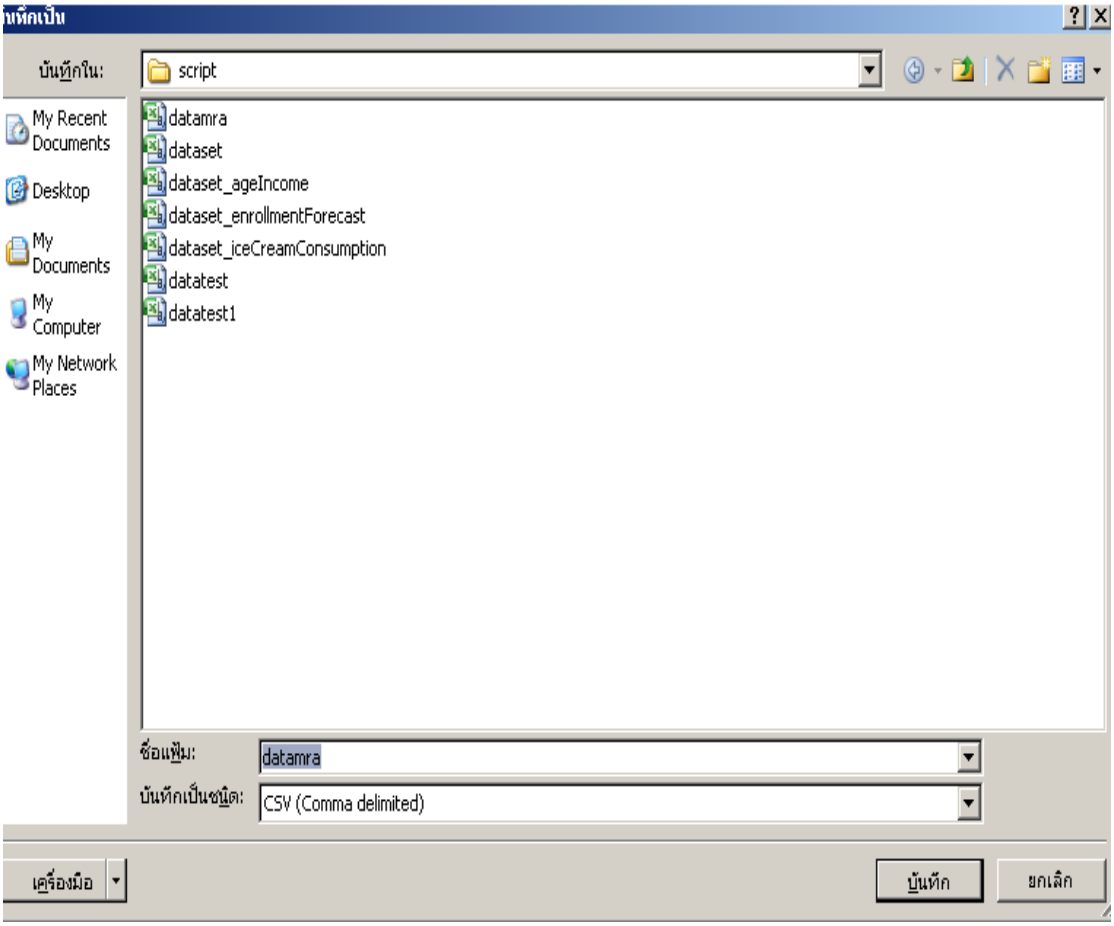
ตัวอย่าง กำหนดให้ตัวแปร X_1 X_2 และ X_3 เป็นตัวแปรพยากรณ์ และตัวแปร Y เป็นตัวแปรเกณฑ์ ข้อมูลจากกลุ่มตัวอย่างจำนวน 29 คน ดังภาพประกอบ 1

	A	B	C	D	E	F
1	ID	Y	X1	X2	X3	
2	1	5501	3240	9552	1923	
3	2	5945	2800	9680	1961	
4	3	6629	2920	9731	1979	
5	4	7556	3000	11666	2030	
6	5	8716	2800	14675	2112	
7	6	9369	2560	15265	2192	
8	7	9920	2600	15484	2235	
9	8	10167	2560	15723	2351	
10	9	11084	2520	16501	2411	
11	10	12504	3080	16890	2475	
12	11	13746	3280	17203	2524	
13	12	13656	3000	17707	2674	
14	13	13850	2960	18108	2833	
15	14	14145	3280	18266	2863	
16	15	14888	4040	19308	2839	
17	16	14991	3680	18224	2898	
18	17	14836	3080	18997	3123	
19	18	14478	2280	19505	3195	
20	19	14539	2600	19800	3239	
21	20	14395	3000	19546	3129	
22	21	14599	2920	19117	3100	
23	22	14969	3680	18774	3008	
24	23	15107	4040	17813	2983	
25	24	14831	3000	17304	3069	
26	25	15081	3520	16756	3151	
27	26	15127	3640	16749	3127	
28	27	15856	3520	16925	3179	
29	28	15938	3120	17231	3207	
30	29	16081	2800	16816	3345	
31						

ภาพประกอบ 1

การวิเคราะห์การถดถอยพหุคูณ มีขั้นตอนการวิเคราะห์ ดังนี้

1. การบันทึกข้อมูล ในตัวอย่างใช้การบันทึกข้อมูลด้วยโปรแกรม Excel โดยบันทึกในรูปแบบไฟล์ .CSV โดยใช้ชื่อไฟล์ datamra เก็บไว้ใน D:/UsedR/script ดังภาพประกอบ 2



ภาพประกอบ 2

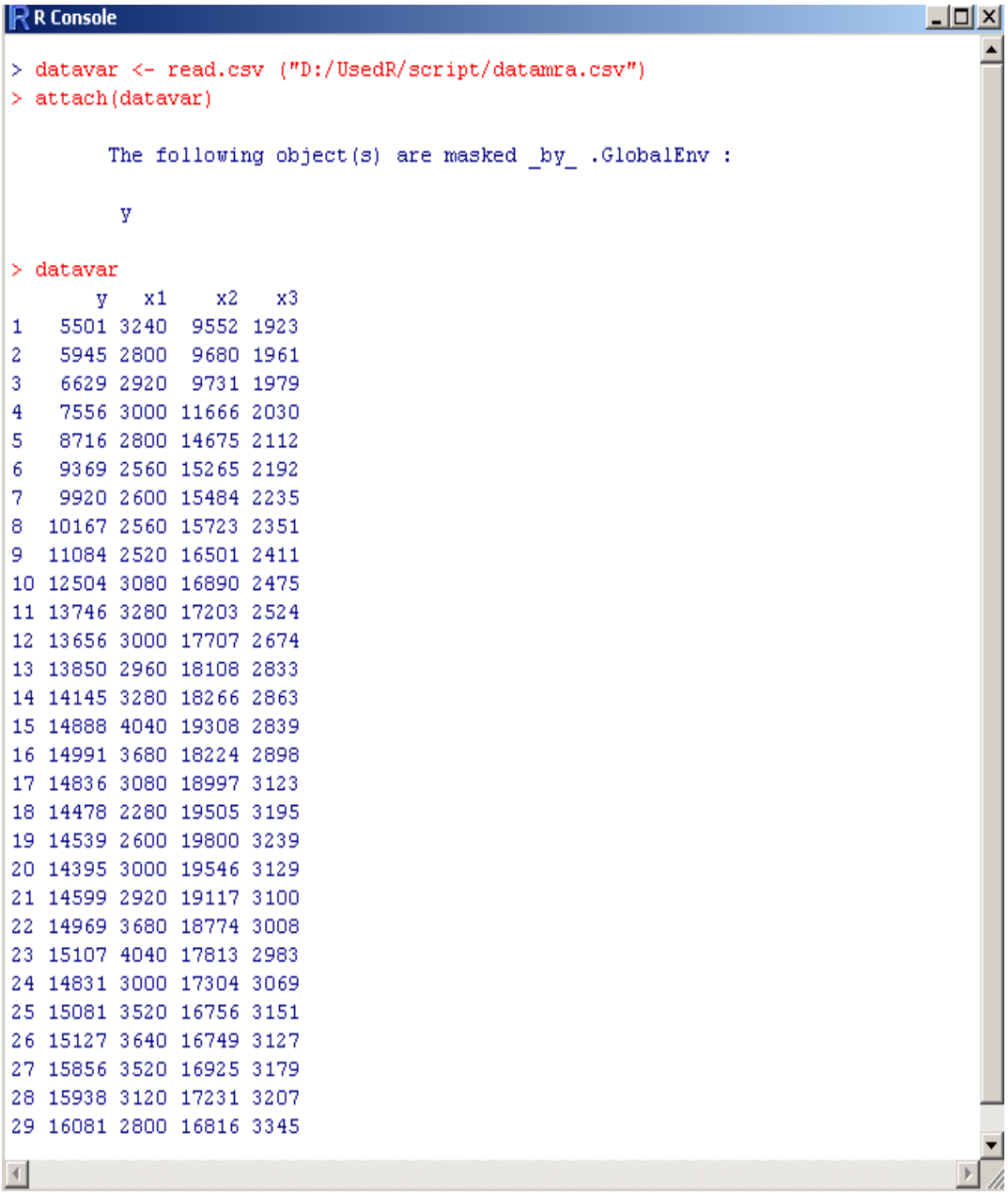
2. นำข้อมูลเข้าโปรแกรม R โดยใช้คำสั่ง

> datavar <- read.csv ("D:/UsedR/script/datamra.csv")

> attach (datavar)

> datavar

เป็นคำสั่งเพื่อให้โปรแกรม R อ่านและรับทราบข้อมูล พร้อมทั้งกำหนดตัวแปรใหม่ในชื่อ “datavar” ซึ่งหมายถึงชื่อแฟ้มข้อมูล จะได้ผลการนำเข้า.ข้อมูลดังภาพประกอบ 3



ภาพประกอบ 3

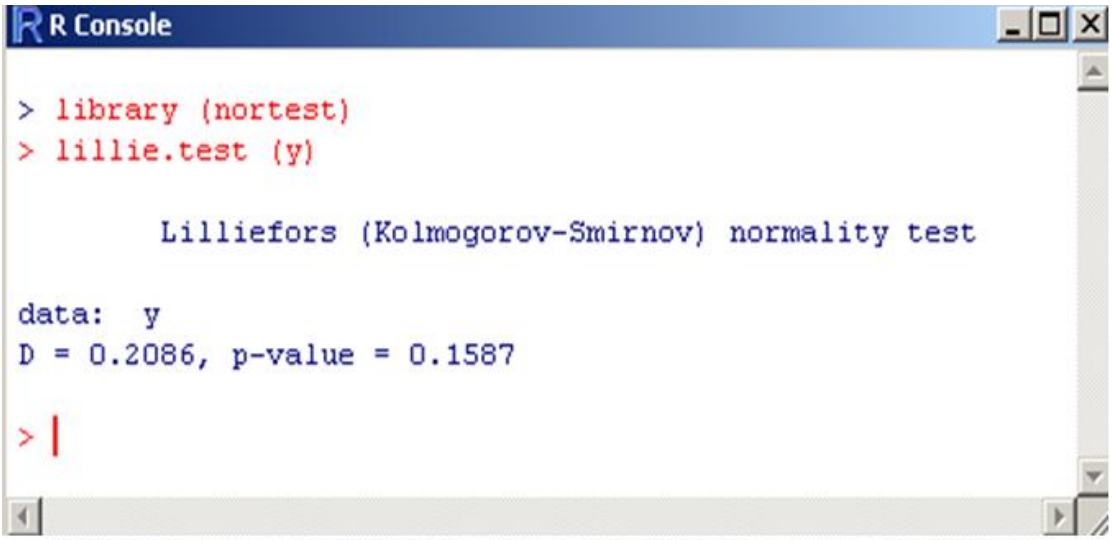
3. การตรวจสอบข้อตกลงเบื้องต้น

3.1 ตรวจสอบว่าการแจกแจงของประชากรเป็นการแจกแจงแบบปกติหรือไม่ ด้วยสถิติ Lilliefors (Kolmogorov-Smirnov) normality test แยกตามกลุ่ม ด้วยคำสั่ง lillie.test ซึ่งก่อนจะใช้คำสั่งนี้จะต้อง Load Packages ที่ชื่อ nortest มาก่อนด้วยคำสั่ง

>library (nortest)

>lillie.test (y)

การใช้คำสั่งและผลลัพธ์ ดังภาพประกอบ 4



ภาพประกอบ 4

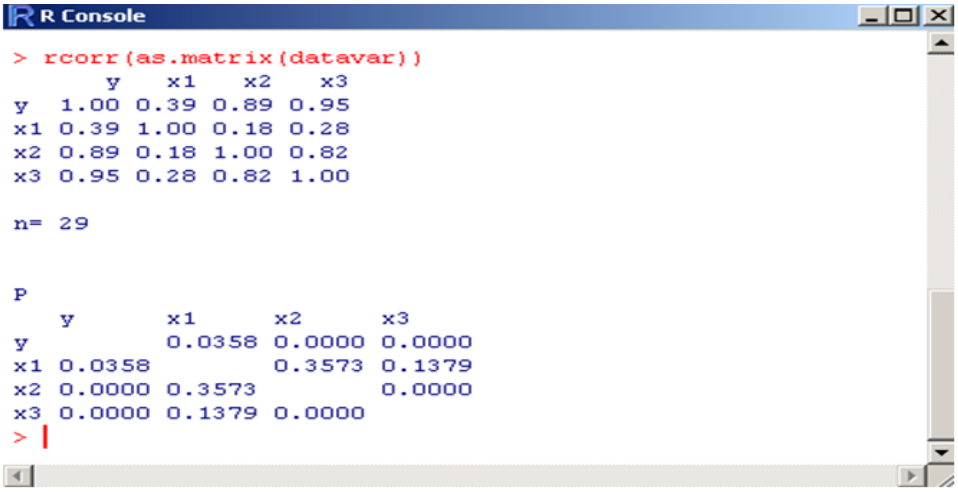
จากภาพประกอบ 4 แสดงว่ากลุ่มตัวอย่างสุ่มมาจากประชากรที่มีการแจกแจงเป็นปกติ (ดูได้จากค่า p-values มีค่ามากกว่า .01; สมมติว่าผู้วิจัยกำหนดระดับนัยสำคัญ .01) เป็นไปตามข้อตกลงเบื้องต้น

3.2 ตรวจสอบความสัมพันธ์กันเชิงเส้นตรง (Linearity) ของตัวแปรพยากรณ์กับตัวแปรเกณฑ์ โดยใช้สหสัมพันธ์แบบเพียร์สัน (r_{xy}) ด้วยคำสั่ง และก่อนใช้งานต้องเรียกใช้ฟังชั่นเสริม Hmisc ก่อน ดังนี้

>library (Hmisc)

>rcorr (as.matrix (datavar))

การใช้คำสั่งและผลลัพธ์ ดังภาพประกอบ 5



ภาพประกอบ 5

จากภาพประกอบ 5 แสดงว่า ตัวแปรพยากรณ์กับตัวแปรเกณฑ์มีความสัมพันธ์กันเชิงเส้นตรงอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 สองตัว คือ (x_2,y) และ (x_3,y) และมีความสัมพันธ์กันเชิงเส้นตรงอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 หนึ่งตัว คือ (x1,y) (ดูได้จากค่า p) เป็นไปตามข้อตกลงเบื้องต้น

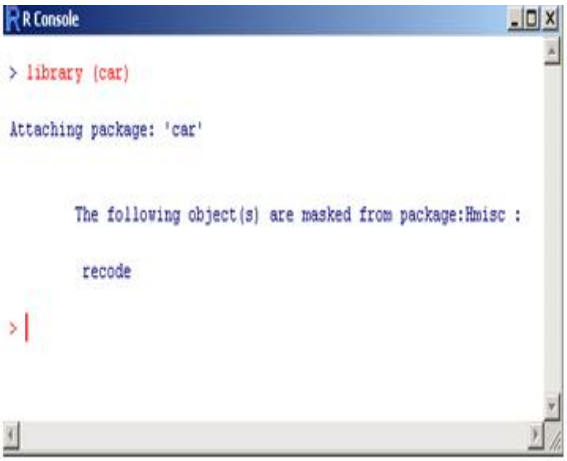
3.3 ตรวจสอบความแปรปรวนของความคลาดเคลื่อนว่ามีความคงที่ทุกค่าการสังเกตหรือไม่ (Homoscedasticity) ตรวจสอบโดยใช้ Non-constant Variance Score Test ด้วยคำสั่ง และก่อนใช้คำสั่งนี้ต้องเรียกใช้ library (car) ก่อนดังนี้

> library (car)

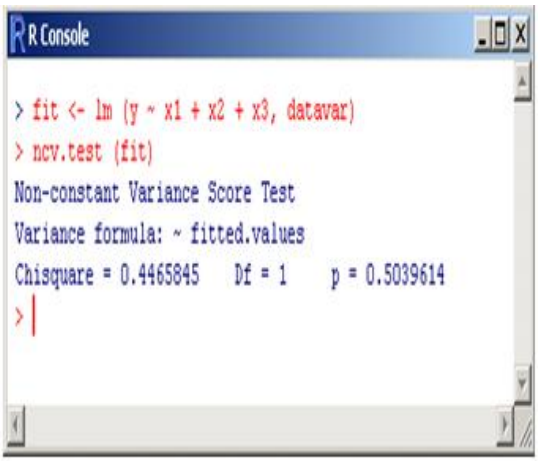
>fit <- lm (y ~ x1 + x2 + x3, datavar)

>ncv.test (fit)

การใช้คำสั่งและผลลัพธ์ ดังภาพประกอบ 6 และ 7



ภาพประกอบ 6



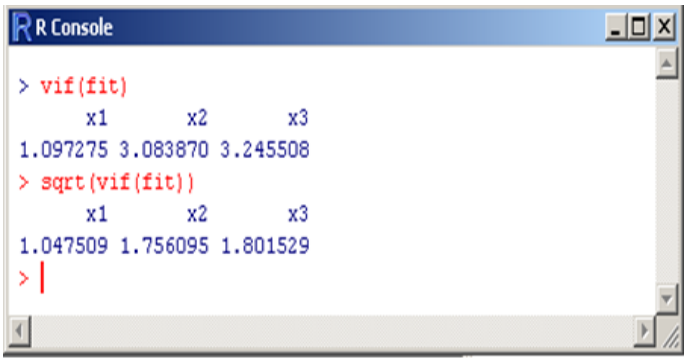
ภาพประกอบ 7

จากภาพประกอบ 7 แสดงว่า ความแปรปรวนของความคลาดเคลื่อนว่ามีความคงที่ทุกค่าการสังเกต (ดูได้จากค่า p มากกว่า .01) เป็นไปตามข้อตกลงเบื้องต้น

3.4 ตรวจสอบ Multicollinearity โดยดูจากค่า variance inflation factors (VIF) โดยใช้คำสั่ง

```
>vif(fit)
>sqrt (vif (fit))
```

การใช้คำสั่งและผลลัพธ์ ดังภาพประกอบ 8



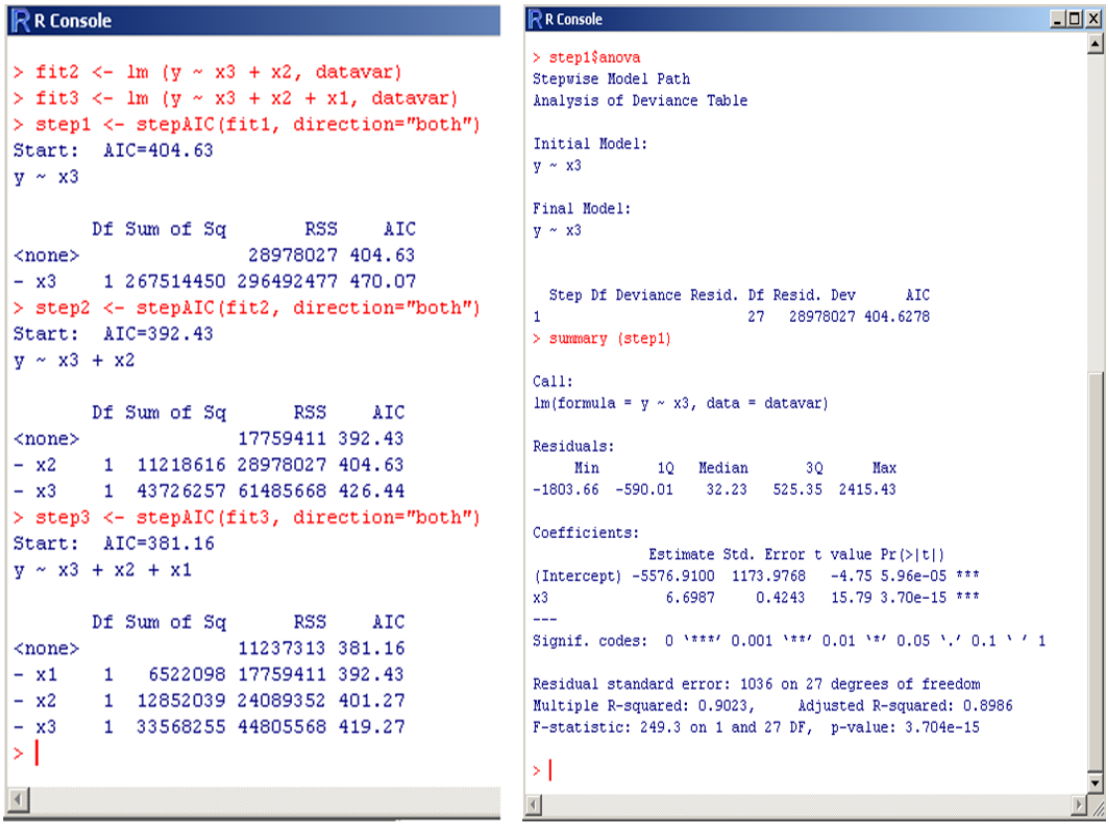
ภาพประกอบ 8

จากภาพประกอบ 8 แสดงว่า ตัวแปรพยากรณ์ไม่มีปัญหา Multicollinearity (ดูได้จากค่า VIF มีค่าน้อยกว่า 4 หรือ รากที่สองของ VIF มีค่าน้อยกว่า 2) เป็นไปตามข้อตกลงเบื้องต้น

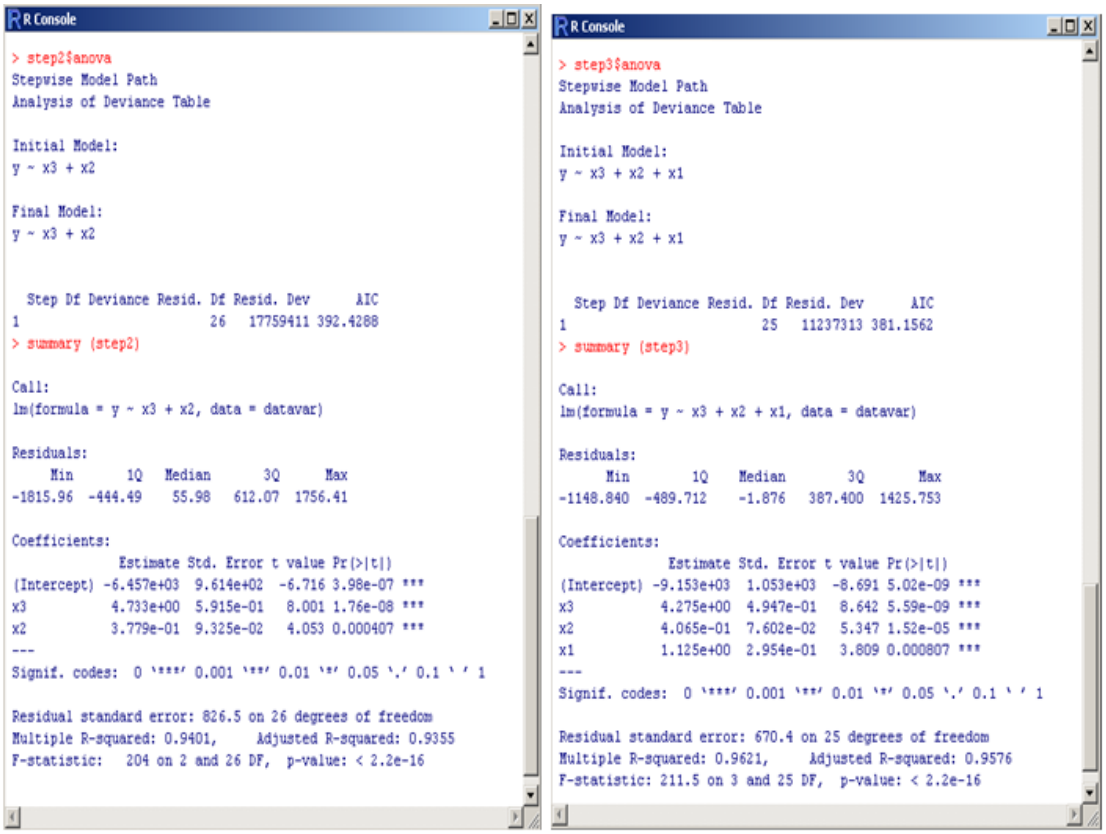
4. การวิเคราะห์การถดถอยพหุคูณ (ในตัวอย่างนี้จะเป็นการวิเคราะห์แบบขั้นบันได (Stepwise Regression) ด้วยคำสั่งดังนี้

```
>library(MASS)
>fit1 <- lm (y ~ x3, datavar)
>fit2 <- lm (y ~ x3 + x2, datavar)
>fit3 <- lm (y ~ x3 + x2 + x1, datavar)
>step1 <- stepAIC (fit1, direction="both")
>step2 <- stepAIC (fit2, direction="both")
>step3 <- stepAIC (fit3, direction="both")
>step1$anova
>summary (step1)
>step2$anova
>summary (step2)
>step3$anova
>summary (step3)
```

การใช้คำสั่งและผลลัพธ์ ดังภาพประกอบ 9 และ10



ภาพประกอบ 9



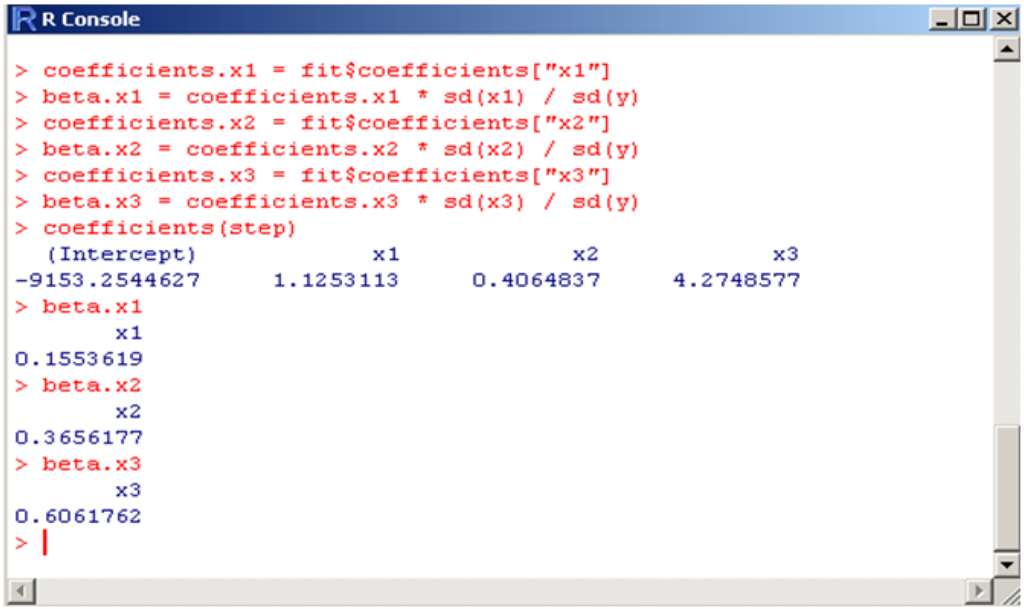
ภาพประกอบ 10

จากภาพประกอบ 9 และ10 เป็นการรายงานค่าน้ำหนักความสำคัญของตัวแปรพยากรณ์ในรูปคะแนนดิบ (b: Estimate) ค่าความคลาดเคลื่อนมาตรฐาน (Std.Error: SE_b) ค่าความคลาดเคลื่อนมาตรฐานในการพยากรณ์ (Residual standard error: SE_{est}) ค่าสัมประสิทธิ์การทำนาย (R^2: R-squared) ค่าสัมประสิทธิ์การทำนายที่ปรับแก้แล้ว (R^2adj: Adjusted R-squared) ค่าสถิติ F ที่ทดสอบนัยสำคัญของสัมประสิทธิ์พหุคูณ (F-statistic) และค่า p-values ในแต่ละโมเดลการวิเคราะห์

5. คำนวณค่า b และเบต้า (β) ด้วยคำสั่งดังนี้

```
>coefficients.x1 = fit$coefficients["x1"]
>beta.x1 = coefficients.x1 * sd(x1) / sd(y)
>coefficients.x2 = fit$coefficients["x2"]
>beta.x2 = coefficients.x2 * sd(x2) / sd(y)
>coefficients.x3 = fit$coefficients["x3"]
>beta.x3 = coefficients.x3 * sd(x3) / sd(y)
>coefficients(step)
>beta.x1
>beta.x2
>beta.x3
```

การใช้คำสั่งและผลลัพธ์ ดังภาพประกอบ 11



ภาพประกอบ 11

จากภาพประกอบ 11 แสดงการคำนวณค่า b และเบต้า (β) โดยค่า b ของ x1 คือ 1.125 b ของ x2 คือ 0.406 และ b ของ x3 คือ 4.275 โดยมีค่าคงที่เป็น –9,153.254 ส่วนค่าเบต้าของ x1 คือ 0.155 ค่าเบต้าของ x2 คือ 0.366 และค่าเบต้าของ x3 คือ 0.606

การสร้างตารางนำเสนอและการแปลความหมายผลการวิเคราะห์

ผลการวิเคราะห์ข้อมูลโดยใช้วิธีวิเคราะห์การถดถอยพหุคูณแบบเป็นขั้นบันได ดังตารางที่ 1

ตาราง 1 เมตริกซ์ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปร

ตัวแปร	x1	x2	x3
x2	.180		
x3	.280	.820**	
y	.390*	.890**	.950**

* มีนัยสำคัญทางสถิติที่ระดับ .05

** มีนัยสำคัญทางสถิติที่ระดับ .01

จากตาราง 1 เมตริกซ์ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรพยากรณ์แต่ละตัว มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 หนึ่งค่า โดยเรียงค่าความสัมพันธ์สูงสุดคือ x2 กับ x3 มีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ .820 รองลงมาเป็นความสัมพันธ์ระหว่าง x1 กับ x3 มีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ .280 และความสัมพันธ์ระหว่าง x1 กับ x2 มีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ .180 ตามลำดับ เมื่อพิจารณาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรพยากรณ์กับตัวแปรเกณฑ์ ปรากฏว่ามีค่าสัมประสิทธิ์สหสัมพันธ์ตั้งแต่ .390 ถึง .950 และมีนัยสำคัญทางสถิติที่ระดับ .01 จำนวน 2 คู่ และมีนัยสำคัญทางสถิติที่ระดับ .05 จำนวน 1 คู่ ที่มีความสหสัมพันธ์กับตัวแปรเกณฑ์สูงสุด คือ x3 รองลงมาเป็น x2 และ x1 ตามลำดับ

ตาราง 2 ผลการวิเคราะห์การถดถอยพหุคูณระหว่างตัวแปรพยากรณ์กับตัวแปรเกณฑ์ โดยใช้วิธีวิเคราะห์การถดถอยพหุคูณแบบเป็นขั้นบันได

ตัวพยากรณ์	R	R^2	R^2_{adj}	SE_{est}	F	p
x3	.950	.902	.899	1,036.000	249.300	. 000
x3, x2	.970	.940	.936	826.500	204.000	. 000
x3, x2, x1	.981	.962	.958	670.400	211.500	. 000

R = .981 R^2 = .962 R^2_{Adj} = .958 SE_{est} = 670.400

F = 211.500 a = -9,153.254

จากตาราง 2 พบว่า ผลการวิเคราะห์การถดถอยพหุคูณแบบเพิ่มตัวแปรเป็นขั้น ๆ ปรากฏว่า ตัวแปรที่สามารถพยากรณ์ตัวแปรเกณฑ์ได้อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 คือ x3, x2 และ x1 ค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณ (R) เท่ากับ .981 ค่าสัมประสิทธิ์การพยากรณ์ปรับปรุง (R^2_{Adj}) เท่ากับ .958 ค่าความคลาดเคลื่อนมาตรฐาน (SE_{est}) เท่ากับ 670.400 ตัวแปรทั้ง 3 ตัว สามารถทำนายได้ร้อยละ 95.80

ตาราง 3 ผลการวิเคราะห์การถดถอยพหุคูณ โดยใช้ตัวแปรพยากรณ์ x1, x2 และ x3 พยากรณ์ตัวแปร y

ตัวพยากรณ์	b	β	t	p
Constant	-9,153.254		-8.691	.000
x3	4.275	0.606	8.642	. 000
x2	0.406	0.366	5.347	. 000
x1	1.125	0.155	3.809	. 000

จากตาราง 3 พบว่า ตัวแปร x3 , x2 และ x1 มีค่าสัมประสิทธิ์ของตัวพยากรณ์ในรูปคะแนนดิบ (b) เท่ากับ 4.275, .406, และ 1.125 ตามลำดับ มีค่าสัมประสิทธิ์ของตัวแปรพยากรณ์ในรูปคะแนนมาตรฐาน (β) เท่ากับ .606, .366 และ .155 ตามลำดับ และมีค่าคงที่ของสมการพยากรณ์ในรูปคะแนนดิบ (a) เท่ากับ -9,153.254 โดยเขียนเป็นสมการได้ดังนี้

สมการพยากรณ์ในรูปคะแนนดิบ

$Y' = -9,153.254 + 4.275(x3) + .406(x2) + 1.125(x1)$

สมการพยากรณ์ในรูปคะแนนมาตรฐาน

$Z_{Y'} = .606\ Z_{x3} + .366\ Z_{x2} + .155\ Z_{x1}$

บรรณานุกรม

Charles Geyer. **Statistics 5601 (Geyer. Fall 2003) Examples: Kolmogorov-Smirnov and Lilliefors Tests**. . [Online] from http://www.stat.umn.edu/geyer/old03/5601/examp/kolmogorov.html. Accessed 20 March 2010.

Mark Gardener. **Using R for statistical analyses - Multiple Regression**. [Online] from http://www.gardenersown.co.uk/Education/Lectures/R/regression.htm. accessed 30 march 2010.

Robert I. Kabacoff. (2008). **Multiple (Linear) Regression**. [Online] from http://www.statmethods.net/stats/regression.html. accessed 29 march 2010,

Robert I. Kabacoff. (2008). **Regression Diagnostics**. [Online] from http://www.statmethods.net/stats/rdiagnostics.html. accessed 30 march 2010,

กัลยา วานิชย์บัญชา. (2546). **การใช้ SPSS for Windows ในการวิเคราะห์ข้อมูล**. สำนักงานพิมพ์จุฬาลงกรณ์มหาวิทยาลัย.

บุญชม ศรีสะอาด. (2547). **วิธีการทางสถิติสำหรับการวิจัย เล่ม 1**. พิมพ์ครั้งที่ 4. กรุงเทพฯ : สุวีริยาสาส์น.

ทรงศักดิ์ ภูสีอ่อน.(2551). **การประยุกต์ใช้ SPSS วิเคราะห์ข้อมูลงานวิจัย**. พิมพ์ครั้งที่ 2. กาฬสินธุ์ :ประสานการพิมพ์.

สมบัติ ท้ายเรือคำ (2551). **ระเบียบวิธีวิจัยสำหรับมนุษยศาสตร์และสังคมศาสตร์**. พิมพ์ครั้งที่ 2. กาฬสินธุ์ : ประสานการพิมพ์.

สมบัติ ท้ายเรือคำ. **การวิเคราะห์การถดถอยพหุ (MRA) การวิเคราะห์เส้นทาง (PA) และการวิเคราะห์องค์ประกอบ (FA)**. วารสารการวัดผลการศึกษา มหาวิทยาลัยมหาสารคาม (ปีที่ 8 หน้า 37-46) 2545.

ศิริชัย พงษ์วิชัย. (2551). **การวิเคราะห์ข้อมูลทางสถิติด้วยคอมพิวเตอร์**. พิมพ์ครั้งที่ 19. กรุงเทพฯ. วี.พรินทร์ (1991).