บทที่ ๓

วิธีดำเนินการวิจัยและผลการวิจัย

๓.๑ หลักการดำเนินการวิจัย

๑. ศึกษาคุณสมบัติและทฤษฎีการประมาณค่าสัมประสิทธิ์ความถดถอยระหว่างวิธีกำลังสองน้อยที่สุดและวิธีริดจ์ รีเกรสชั่น ซึ่งได้กล่าวไว้โดยละเอียดในบทที่ ๒

๒. ศึกษาโปรแกรมสำเร็จรูปการคำนวณค่า Inverse ของเมทริกซ์ เพื่อใช้เขียนโปรแกรมโดยใช้ภาษา FORTRAN IV ในการประมาณค่าสัมประสิทธิ์ความถดถอยและค่าสถิติต่าง ๆ ของวิธีที่มีตัวแปรอิสระทุกตัวในตัวแบบและวิธี Ridge Regression

ส่วนการประมาณค่าสัมประสิทธิ์ความถดถอยของวิธี Stepwise Multiple Regression จะใช้โปรแกรมสำเร็จรูป SPSS ในการคำนวณค่าสถิติต่าง ๆ

๓. เนื่องจากคุณสมบัติของข้อมูลที่จะนำมาคำนวณโดยวิธี ริดจ์ รีเกรสชั่น จะได้ผลดีเมื่อตัวแปรอิสระมีสหสัมพันธ์สูง ฉะนั้นในการศึกษาครั้งนี้จึงเลือกข้อมูลทั้งหมด ๔ ชุด ที่ตัวแปรอิสระมีสหสัมพันธ์ระหว่าง -๐.๑๐ -๐.๙๙ มาศึกษาเปรียบเทียบระหว่าง ๒ วิธี ดังกล่าวข้างต้น ลักษณะข้อมูลที่เลือกมามีรายละเอียดดังแสดงในภาคผนวก ก. ดังนี้

ข้อมูลชุดที่ ๑ การวิเคราะห์เพื่อพยากรณ์ความต้องการใช้พลังงานไฟฟ้าในเขตนครหลวง ซึ่งตัวแปรอิสระมีสหสัมพันธ์อยู่ระหว่าง -๐.๘๗ -๐.๙๙

ข้อมูลชุดที่ ๒ การพยากรณ์ปริมาณการใช้โทรศัพท์ในเขตกรุงเทพมหานคร ซึ่งตัวแปรอิสระมีสหสัมพันธ์อยู่ระหว่าง ๐.๘๙-๐.๙๙

ข้อมูลชุดที่ ๓ การคาดการปริมาณเงินฝากธนาคารพาณิชย์ ซึ่งตัวแปรอิสระมีสหสัมพันธ์อยู่ระหว่าง ๐.๔๔-๐.๙๙

ข้อมูลชุดที่ ๔ การศึกษาอุปสงค์ของเครื่องอุปโภคของเครื่องนุ่งห่ม ซึ่งตัวแปรอิสระมีสหสัมพันธ์อยู่ระหว่าง ๐.๙๔-๐.๙๙

ข้อมูลชุดที่ ๔ ปริมาณการบริโภคน้ำตาลทราย ซึ่งตัวแปรอิสระมีสหสัมพันธ์อยู่ระหว่าง -๐.๑๐-๐.๙๙

๔. นำข้อมูลทั้ง ๕ ชุดมาคำนวณหาค่าตัวประมาณพารามิเตอร์ และค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง, F-test, R^2, $\hat{Y}$ และค่าความเอนเอียงกำลังสองด้วยวิธีกำลังสองน้อยที่สุด และวิธีริดจ์ รีเกรสชั่น โดยใช้เครื่องคอมพิวเตอร์ในการคำนวณ

๕. จากข้อ ๔ เปรียบเทียบผลลัพธ์ของตัวประมาณ $\hat{\beta}$, ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง, R^2, $\hat{Y}$ และค่าความเอนเอียงกำลังสองระหว่างวิธีกำลังสองน้อยที่สุดและวิธีริดจ์ รีเกรสชั่น โดยพิจารณาเป็นคู่ ๆ คือ ระหว่างวิธีที่มีตัวแปรอิสระทุกตัวในตัวแบบกับวิธี Ridge Regression และเปรียบเทียบระหว่างวิธี Step-Wise Multiple Regression กับวิธี Ridge Regression

การเปรียบเทียบแต่ละคู่จะใช้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเป็นเกณฑ์ในการเปรียบเทียบและจะทดสอบค่าประมาณของตัวแปรตาม แต่ละวิธีกับค่าจริงของตัวแปรตามว่ามีความแตกต่างกันตามสมมติฐานที่ตั้งไว้หรือไม่

๓.๒ ผลการวิจัย

ผลการประมาณค่าพารามิเตอร์ ระหว่างวิธีกำลังสองน้อยที่สุด และวิธี ริดจ์ รีเกรสชั่น ของแต่ละวิธีได้จากโปรแกรมคอมพิวเตอร์ซึ่งแสดงไว้ในภาคผนวก

การประมาณค่าสัมประสิทธิ์ความถดถอยของพารามิเตอร์ β ของตัวอย่างข้อมูลชุดที่ ๑ - ๕ กำหนดให้

n_i = จำนวนค่าสังเกตของข้อมูลชุดที่ i

เมื่อ i = 1, 2, 3, 4, 5

$r_{x_i x_j}$ = สหสัมพันธ์ของตัวแปรอิสระ X_i และ X_j $i = 1,2,\ldots,I$, $i \neq j$

I = จำนวนตัวแปรอิสระของข้อมูลแต่ละชุด

ตัวอย่างที่ ๑ ข้อมูลชุดที่ ๑ $i = 5$, $r_{x_i x_j}$ อยู่ระหว่าง -๐.๘๙-๐.๙๙ และ $n_1 = 13$ (ภาคผนวก) นำข้อมูลชุดที่ ๑ แทนค่าในสมการ (๒.๑.๒) และ(๒.๒.๕) เพื่อคำนวณค่า $\hat{\beta}$ และ $E[\ L_1^2(0)\ ]$ ผลเป็นดังนี้

ตารางที่ ๒

แสดงการประมาณค่าพารามิเตอร์ β ด้วยวิธีที่มีตัวแปรอิสระทุกตัวในตัวแบบ

ตัวแปรอิสระ	ตัวประมาณ	$E[\ L_1^2(0)\ ]$	R^2
ค่าคงที่	-1154.9868		
X_1	- 2.3627		
X_2	720.1125		
X_3	144.6253	2,375.492	0.9993
X_4	- 15.0423		
X_5	- 28.4925		

จากตารางที่ ๑ ค่า $E[\ L_1^2(0)\ ]$ และ R^2 มีค่าเท่ากับ 2,375.492 และ 0.9993 ตามลำดับ

จากข้อมูลชุดที่ ๑ การประมาณค่าพารามิเตอร์ β ด้วยวิธี Step-Wise Multiple Regression ด้วยโปรแกรมสำเร็จรูป SPSS เพื่อคำนวณค่า $\hat{\beta}$, $E[L_1^2(0)]$, R^2 และ Partial-F ผลเป็นดังนี้

ตารางที่ ๓

แสดงการประมาณค่าพารามิเตอร์ β โดยวิธี Step-Wise Multiple Regression

ตัวแปรอิสระ	ตัวประมาณ $\hat{\beta}$	$E[L_1^2(0)]$	R^2	F-test
ค่าคงที่	-1209.925			
X_3	133.485	1834.7761	0.868	128.749
X_2	624.899			21.985

จากตารางที่ ๓ เมื่อคำนวณตัวประมาณค่า β ด้วยวิธี Step-Wise Multiple Regression มีหลักเกณฑ์ในการตัดตัวแปรอิสระใด ๆ ออกจากสมการในแต่ละขั้นของวิธีนี้ พิจารณาจากค่า Partial-F ของตัวแปรอิสระที่มีอยู่ในสมการของขั้นนั้น ๆ เปรียบเทียบกับค่า $F_\alpha(1, n-K)$ (K หมายถึงจำนวนตัวแปรอิสระที่มีอยู่ทั้งหมดในสมการของขั้นนั้น ๆ) ถ้าค่า Partial-F ของตัวแปรอิสระใดมากกว่า $F_\alpha(1, n-K)$ ก็รวมเอาตัวแปรอิสระตัวนั้นไว้ในสมการแต่ถ้าค่า Partial-F ใดน้อยกว่า $F_\alpha(1, n-K)$ ก็จะตัดตัวแปรอิสระนั้นออกจากสมการของขั้นนั้น ๆ จะพิจารณาเช่นนี้ทุกครั้งที่มีตัวแปรอิสระเพิ่มเข้ามาในสมการใหม่และสิ้นสุดเมื่อไม่มีตัวแปรอิสระใดที่จะเข้าอยู่ในสมการอีกต่อไป สุดท้ายจะได้สมการที่มีตัวแปรอิสระบางตัวที่เหมาะสมอยู่ในสมการ กล่าวคือจากข้อมูลชุดที่ ๒ ตัวแปรอิสระมีทั้งหมด ๕ ตัว ผลลัพธ์ครั้งสุดท้ายจะเหลือตัวแปรอิสระที่เหมาะสมกับตัวแบบเพียง ๒ ตัว คือ X_2 และ X_3 ส่วนตัวแปรอิสระอีก ๓ ตัวจะถูกตัดออกไปจากตัวแบบ

สำหรับการประมาณค่าพารามิเตอร์ β โดยวิธี ริดจ์ รีเกรสชั่น ได้จากการนำข้อมูลชุดที่ ๑ ไปแทนค่าในสมการ (๒.๑.๓) และสมการ (๒.๕.๖) เพื่อคำนวณค่า $\hat{\beta}^*$ และ $E[L_1^2(K)]$ กับค่า $(BIAS)^2$ กำหนดให้ $K > 0$ โดย K มีค่าเพิ่มขึ้นครั้งละ ๐.๐๑ จะคำนวณค่าเหล่านี้ซ้ำหลายรอบและในแต่ละรอบจะตรวจสอบค่า $E[L_1^2(K)]$ และ $E[L_1^2(0)]$ ถ้าพบว่าค่า $E[L_1^2(K)]$ น้อยกว่า $E[L_1^2(0)]$ จะสิ้นสุดการคำนวณและถือว่าค่า K นี้เริ่มจะให้ $E[L_1^2(K)]$ ที่เหมาะสม ส่วน $E[L_1^2(K)]$ **optimum** ได้จาก

การแทนค่า $K = 2\sigma^2/\hat{\beta}'\hat{\beta}$ ในสมการ (๒.๑.๓) และ (๒.๕.๖) ผลจากการคำนวณค่าต่าง ๆ เป็นดังนี้คือ

ตารางที่ ๔

แสดงการประมาณค่าพารามิเตอร์ β โดยวิธีริดจ์ รีเกรสชั่น

รอบที่	K	$E[L_1^2(K)]$	$(BIAS)^2$	R^2
1	0.01	95,848.562	249.94	0.9720
2	0.02	62,211.765	560.10	0.9818
3	0.03	44,926.793	802.65	0.9868
4	0.04	34,708.707	985.81	0.9898
10	0.10	14,317.496	1,489.81	0.9958
20	0.20	7,716.637	1,673.93	0.9977
30	0.30	5,611.828	1,686.40	0.9983
40	0.40	4,551.179	1,656.71	0.9983
50	0.50	3,901.474	1,622.13	0.9983
80	0.80	2,898.688	1,527.95	0.9992
100	1.00	2,561.031	1,488.43	0.9992
101	1.01	1,950.686	1,426.67	0.9992
365	3.65	1,838.870	1,608.24	0.9992

จากตารางที่ ๔ จะสังเกตได้ว่า ณ. รอบที่ ๑๐๑ เมื่อ $K = 1.01$ เริ่มจะให้ $E[L_1^2(K)] < E[L_1^2(0)]$ (จากตารางที่ ๒) เมื่อมีการคำนวณซ้ำเรื่อย ๆ โดยเพิ่มค่า $K > 1.01$ จะพบว่า ณ. รอบที่ ๓๖๕ เมื่อ $K = 3.65$ ค่า $E[L_1^2(K)]$ จะเป็นค่าที่น้อยที่สุดดังแสดงไว้ในตารางข้างบนนี้ ส่วน R^2 นั้น จะมีค่ามากขึ้นเรื่อย ๆ ทุกครั้งที่เพิ่มค่า K และ R^2 จะมีค่าคงที่เมื่อ $K = .80$

ตัวอย่างที่ ๒ ข้อมูลชุดที่ ๒ $i=5$, $r_{x_i x_j}$ อยู่ระหว่าง ๐.๘๙-๐.๙๙ และ $n_1 = 9$ (จากภาคผนวก)

นำข้อมูลชุดที่ ๒ แทนค่าในสมการ (๒.๑.๒), (๒.๒.๕) เพื่อคำนวณค่า $\hat{\beta}$ และ $E[\ L_1^2(0)\ ]$ ผลเป็นดังนี้

ตารางที่ ๕

แสดงการประมาณค่าพารามิเตอร์ β โดยวิธีที่มีตัวแปรอิสระทุกตัวในตัวแบบ

ตัวแปรอิสระ	$\hat{\beta}$	$E[\ L_1^2(0)\ ]$	R^2
ค่าคงที่	-154.1349		
X_1	17.7881		
X_2	- 10.6353	7.8551	0.998
X_3	- 0.6313		
X_4	0.4828		
X_5	4.1744		

จากตารางที่ ๕ ค่า $E[\ L_1^2(0)\ ]$ และ R^2 มีค่าเท่ากับ 7.8551 และ 0.998 ตามลำดับ

จากข้อมูลชุดที่ ๒ การประมาณค่าพารามิเตอร์ β ด้วยวิธี Step-Wise Multiple Regression ได้ค่า $\hat{\beta}$, $E[\ L_1^2(0)\ ]$, R^2 และ Partial-F test ดังนี้

ตารางที่ ๖

แสดงการประมาณค่าพารามิเตอร์ β โดยวิธี Step-Wise Multiple Regression

ตัวแปรอิสระ	$\hat{\beta}$	$E[L_1^2(0)]$	R^2	F-test
ค่าคงที่	-5.0500			
X_5	4.7482	13.8683	0.9901	119.842
X_3	-0.7568			11.126

จากตารางที่ ๖ เมื่อคำนวณตัวประมาณค่า $\hat{\beta}$ ด้วยวิธี Step-Wise Multiple Regression ตัวแปรอิสระมีทั้งหมด ๔ ตัวผลลัพธ์ครั้งสุดท้ายจะเหลือตัวแปรอิสระที่เหมาะสมกับตัวแบบเพียง ๒ ตัว คือ X_3 และ X_5 ส่วนตัวแปรอิสระอีก ๓ ตัวจะถูกตัดออกไปจากตัวแบบ

สำหรับการประมาณค่าพารามิเตอร์ β ด้วยวิธี ริดจ์ รีเกรสชั่น นำข้อมูลชุดที่ ๒ แทนค่าในสมการ (๒.๑.๓) และ (๒.๕.๖) เพื่อคำนวณหาค่า $\hat{\beta}^*$ และ $E[L_1^2(K)]$ กับค่า $(BIAS)^2$ ผลเป็นดังนี้

ตารางที่ ๗

แสดงการประมาณค่าพารามิเตอร์ β โดยวิธีริดจ์ รีเกรสชั่น

รอบที่	K	$E[\,L_1^2(K)\,]$	$(BIAS)^2$	R^2
1	0.01	156.7940	3.9376	0.9677
2	0.02	88.4508	4.2257	0.9677
5	0.05	45.5494	3.8347	0.9906
10	0.10	24.3290	2.8711	0.9949
15	0.15	15.5775	2.1223	0.9968
20	0.20	10.9660	1.5661	0.9977
25	0.25	8.1986	1.4440	0.9983
26	0.26	7.7731	1.0722	0.9988
40	0.40	4.2124	0.3394	0.9991

จากตารางที่ ๗ จะสังเกตได้ว่า ณ. รอบที่ ๒๖ เมื่อ K = 0.26 เริ่มจะให้ค่า $E[\,L_1^2(K)\,] < E[\,L_1^2(0)\,]$ (ตารางที่ ๕) และเมื่อมีการเพิ่มค่า K เรื่อย ๆ จาก ๐.๒๖ จะพบว่า ณ. รอบที่ ๔๐ เมื่อ K = 0.40 ค่า $E[\,L_1^2(K)\,]$ จะเป็นค่าที่น้อยที่สุดดังแสดงไว้ในตารางข้างบนนี้ ส่วน R^2 นั้น จะมีค่ามากขึ้นเรื่อย ๆ ทุกครั้งที่เพิ่มค่า K

ตัวอย่างที่ ๓ ข้อมูลชุดที่ ๓ $i = 5$, $r_{x_i x_j}$ อยู่ระหว่าง ๐.๔๔-๐.๙๙ และ $n_1 = 10$ (จากภาคผนวก)

นำข้อมูลชุดที่ ๓ แทนค่าในสมการ (๒.๑.๒) และ (๒.๒.๕) เพื่อคำนวณค่า $\hat{\beta}$ และ $E[\ L_1^2(0)\ ]$ ผลเป็นดังนี้

ตารางที่ ๘

แสดงการประมาณค่าพารามิเตอร์ β โดยวิธีที่มีตัวแปรอิสระทุกตัวในตัวแบบ

ตัวแปรอิสระ	$\hat{\beta}$	$E[\ L_1^2(0)\ ]$	R^2
ค่าคงที่	18,158.9961		
X_1	- 964.8763		
X_2	1.3132		
X_3	- 0.1786	13,011,884.00	0.9962
X_4	2.3977		
X_5	- 23.3318		

จากตารางที่ ๘ ค่า $E[\ L_1^2(0)\ ]$ และ R^2 มีค่าเท่ากับ 13,011,884.00 และ 0.9962 ตามลำดับ

จากข้อมูลชุดที่ ๓ การประมาณค่าพารามิเตอร์ β ด้วยวิธี Step-Wise Multiple Regression ได้ค่า $\hat{\beta}$, $E[\ L_1^2(0)\ ]$, R^2 และ Partial-F ดังนี้

ตารางที่ ๙

แสดงการประมาณค่าพารามิเตอร์ β โดยวิธี Step-Wise Multiple Regression

ตัวแปรอิสระ	$\hat{\beta}$	$E[L_1^2(0)]$	R^2	F-test
ค่าคงที่	225.1452			
X_2	0.7847	1,161,596.800	0.99032	30.010
X_4	1.2797			4.650

ตารางที่ ๙ เมื่อคำนวณตัวประมาณค่า $\hat{\beta}$ ด้วยวิธี Step-Wise Multiple Regression ผลลัพธ์ครั้งสุดท้ายจะเหลือตัวแปรอิสระที่เหมาะสมกับตัวแบบเพียง ๒ ตัว คือ X_2 และ X_4 ส่วนตัวแปรอิสระอีก ๓ ตัวจะถูกตัดออกไปจากตัวแบบ

สำหรับการประมาณค่าพารามิเตอร์ β ด้วยวิธีริดจ์ รีเกรสชั่น ได้นำข้อมูลชุดที่ ๓ แทนค่าในสมการ (๒.๑.๓) และ สมการ (๒.๕.๖) เพื่อคำนวณค่า $\hat{\beta}^*$ และ $E[L_1^2(K)]$ กับค่า $(BIAS)^2$ ผลเป็นดังนี้

<u>ตารางที่ ๑๐</u>

แสดงการประมาณค่าพารามิเตอร์ β โดยวิธีริดจ์ รีเกรสชั่น

รอบที่	K	$E[\ L_1^2(K)\ ]$	$(BIAS)^2$	R^2
1	0.01	325,723,648.00	224.3983	0.9050
2	0.02	138,595,664.00	377.6755	0.9596
3	0.03	77,094,079.99	465.3447	0.9775
4	0.04	49,534,431.99	521.0954	0.9855
5	0.05	34,856,016.00	559.4936	0.9898
6	0.06	26,120,896.00	587.5000	0.9992
7	0.07	20,503,424.00	608.8123	0.9996
8	0.08	16,677,518.00	625.5669	0.9951
9	0.09	13,953,913.00	639.0815	0.9959
10	0.10	11,944,918.86	650.2109	0.9965
11	0.11	10,421,530.00	659.5349	0.9970
⋮	⋮	⋮	⋮	⋮
2072	20.72	95,396.875	769.914	0.9999

จากตารางที่ ๑๐ จะสังเกตได้ว่า ณ. รอบที่ ๑๐ เมื่อ K = 0.10 เริ่มจะให้ค่า $E[\ L_1^2(K)\ ] < E[\ L_1^2(0)\ ]$ (จากตารางที่ ๘) และเมื่อมีการเพิ่มค่า K เรื่อย ๆ จาก 0.10 จะพบว่า ณ. รอบที่ ๒๐๗๒ เมื่อ K = 20.72ค่า $E[\ L_1^2(K)\ ]$ จะเป็นค่าน้อยที่สุดดังแสดงไว้ในตารางข้างบนนี้ ส่วน R^2 จะมีค่ามากขึ้นเรื่อย ๆ ทุกครั้งที่เพิ่มค่า K

ตัวอย่างที่ ๔ ข้อมูลชุดที่ ๔ $i = 8$, $r_{x_i x_j}$ อยู่ระหว่าง ๐.๙๔-๐.๙๙ และ $n_1 = 13$ (จากภาคผนวก)

นำข้อมูลชุดที่ ๔ แทนค่าในสมการ (๒.๑.๒) และ (๒.๒.๕) เพื่อคำนวณค่า และ $E[\ L_1^2(0)\ ]$ ผลเป็นดังนี้

ตารางที่ ๑๑

แสดงการประมาณค่าพารามิเตอร์ β โดยวิธีที่มีตัวแปรอิสระทุกตัวในตัวแบบ

ตัวแปรอิสระ	$\hat{\beta}$	$E[\ L_1^2(0)\ ]$	R^2
ค่าคงที่	2118.5320		
X_1	0.1500		
X_2	- 1.8808		
X_3	- 2.3398		
X_4	- 1.9487	129,889.00	0.9964
X_5	- 0.8551		
X_6	0.0260		
X_7	- 1.7039		
X_8	44.9416		

จากตารางที่ ๑๑ ค่า $E[\ L_1^2(0)\ ]$ และ R^2 มีค่าเท่ากับ 129,889.00 และ 0.9964 ตามลำดับ

จากข้อมูลชุดที่ ๔ การประมาณค่าพารามิเตอร์ β ด้วยวิธี Step-Wise Multiple Regression ได้ค่า $\hat{\beta}$, $E[\ L_1^2(0)\ ]$, R^2 และ Partial F-test ดังนี้

ตารางที่ ๑๒

แสดงการประมาณค่าพารามิเตอร์ $\hat{\beta}$ โดยวิธี Step-Wise Multiple Regression

ตัวแปรอิสระ	$\hat{\beta}$	$E[L_1^2(0)]$	R^2	F-test
ค่าคงที่	4433.114			
X_5	2.373	288,835.7666	0.94398	24.244
X_3	- 2.678			4.327

จากตารางที่ ๑๒ เมื่อคำนวณตัวประมาณ $\hat{\beta}$ ด้วยวิธี Step-Wise Multiple Regression ผลลัพธ์ครั้งสุดท้ายจะเหลือตัวแปรอิสระที่เหมาะสมกับตัวแบบเพียง ๒ ตัว คือ X_5 และ X_3 ส่วนตัวแปรอิสระอีก ๖ ตัวถูกตัดออกไปจากตัวแบบ

สำหรับการประมาณค่าพารามิเตอร์ β ด้วยวิธี ริดจ์ รีเกรสชั่น ได้นำข้อมูลชุดที่ ๔ แทนค่าในสมการ (๒.๑.๓) และ (๒.๕.๖) เพื่อคำนวณค่า $\hat{\beta}^*$ และ $E[L_1^2(K)]$ กับค่า $(BIAS)^2$ ผลเป็นดังนี้

ตารางที่ ๑๓

แสดงการประมาณค่าพารามิเตอร์ β โดยวิธีริดจ์ รีเกรสชั่น

รอบที่	K	$E[L_1^2(K)]$	$(BIAS)^2$	R^2
1	0.01	2,856,436.99	87.1020	0.9205
2	0.02	1,020,149.36	124.41	0.9717
3	0.03	518,796.31	142.32	0.9856
4	0.04	313,335.12	152.76	0.9913
5	0.05	269,581.94	159.59	0.9941
6	0.06	150,003.17	164.39	0.9958
7	0.07	112,655.09	167.96	0.9969
1,600	16.00	279.99	192.09	0.9999

จากตารางที่ ๑๓ จะเห็นได้ว่า ณ. รอบที่ ๗ เมื่อ K = 0.07 เริ่มจะให้ค่า $E[L_1^2(K)] < E[L_1^2(0)]$ (จากตารางที่ ๑๑) และเมื่อมีการเพิ่มค่า K เรื่อย ๆ จาก ๐.๐๗ จะพบว่า ณ. รอบที่ ๑,๖๐๐ เมื่อ K = 16.00 ค่า $E[L_1^2(K)]$ จะเป็นค่าที่น้อยที่สุดดังแสดงไว้ในตารางข้างบนนี้ ส่วน R^2 นั้น จะมีค่ามากขึ้นเรื่อย ๆ ทุกครั้งที่เพิ่มค่า K

<u>ตัวอย่างที่ ๕</u> ข้อมูลชุดที่ ๕ $i = 8$, $r_{x_i x_j}$ อยู่ระหว่าง -๐.๑๐-๐.๙๙ $n_i = 14$ (จากภาคผนวก)

นำข้อมูลชุดที่ ๕ แทนค่าในสมการ (๒.๑.๒) และ (๒.๒.๕) เพื่อคำนวณค่า $\hat{\beta}$ และ $E[\ L_1^2(0)\ ]$ ผลลัพธ์เป็นดังนี้

<u>ตารางที่ ๑๔</u>

แสดงการประมาณค่าพารามิเตอร์ β โดยวิธีที่มีตัวแปรอิสระทุกตัวในตัวแบบ

ตัวแปรอิสระ	$\hat{\beta}$	$E[\ L_1^2(0)\ ]$	R^2
ค่าคงที่	0.0216		
X_1	9.6678		
X_2	0.0085		
X_3	-22.2921		
X_4	- 0.0311	442.649	0.9945
X_5	0.2723		
X_6	- 2.8009		
X_7	-		
X_8	0.9907		

จากตารางที่ ๑๔ ค่า $E[\ L_1^2(0)\ ]$ และ R^2 มีค่าเท่ากับ 442.649 และ 0.9945 ตามลำดับ

จากข้อมูลชุดที่ ๕ การประมาณค่าพารามิเตอร์ β ด้วยวิธี Step-Wise Multiple Regression ได้ค่า $\hat{\beta}$, $E[\ L_1^2(0)\ ]$, R^2 และ Partial F-test ดังนี้

ตารางที่ ๑๕

แสดงการประมาณค่าพารามิเตอร์ β โดยวิธี Step-Wise Multiple Regression

ตัวแปรอิสระ	$\hat{\beta}$	$E[\ L_1^2(0)\ ]$	R^2	F-test
ค่าคงที่	-180.6204			
X_1	27.5649			77.433
X_3	- 25.0161	732.7367	0.95131	4.635
X_6	- 1.1123			2.594

จากตารางที่ ๑๕ เมื่อคำนวณตัวประมาณ $\hat{\beta}$ ด้วยวิธี Step-Wise Multiple Regression ผลลัพธ์ครั้งสุดท้ายจะเหลือตัวแปรอิสระที่เหมาะสมกับตัวแบบเพียง เพียง ๓ ตัว คือ X_1, X_3 และ X_6 ส่วนตัวแปรอิสระอีก ๔ ตัว จะถูกตัดออกไปจากตัวแบบ

สำหรับการประมาณค่าพารามิเตอร์ β ด้วยวิธี ริดจ์ รีเกรสชั่น ได้นำข้อมูลชุดที่ ๕ แทนค่าในสมการ (๒.๑.๓) และ (๒.๕.๖) เพื่อคำนวณค่า $\hat{\beta}^*$ และ $E[\ L_1^2(K)\ ]$ กับค่า $(BIAS)^2$ ผลเป็นดังนี้

ตารางที่ ๑๖

แสดงการประมาณค่าพารามิเตอร์ β โดยวิธีริดจ์ รีเกรสชั่น

รอบที่	K	$E[L_1^2(K)]$	$(BIAS)^2$	R^2
1	0.01	125.7630	1.5×10^{-7}	0.9984
2	0.02	125.2898	6.1×10^{-7}	0.9985
3	0.03	124.8197	1.3×10^{-7}	0.9985
4	0.04	124.3524	2.3×10^{-6}	0.9984
5	0.05	123.8820	3.7×10^{-6}	0.9984
6	0.06	123.4269	0.5×10^{-5}	0.9984
7	0.07	122.9684	0.7×10^{-5}	0.9984
15	1.50	6.6940	0.9×10^{-4}	0.9999

ตารางที่ ๑๖ จะเห็นได้ว่า ณ. รอบที่ ๗ เมื่อ K = 0.01 เริ่มจะให้ $E[L_1^2(K)] < E[L_1^2(0)]$ (จากตารางที่ ๑๔) และเมื่อมีการเพิ่ม K เรื่อย ๆ จาก 0.01 จะพบว่า ณ. รอบที่ ๑๕ เมื่อ K = ๑.๕ ค่า $E[L_1^2(K)]$ จะเป็นค่าที่น้อยที่สุดดังแสดงไว้ในตารางข้างบนนี้ ส่วน R^2 จะมีค่ามากขึ้นเรื่อย ๆ ทุกครั้งที่เพิ่มค่า K

๓.๓ ผลการวิจัยการประมาณค่าพารามิเตอร์ β โดยวิธีริดจ์ รีเกรสชั่น

จากการศึกษาข้อมูลตัวอย่าง ๕ ชุด สามารถสรุปคุณสมบัติต่าง ๆ ที่สำคัญของตัวประมาณ $\hat{\beta}^*$ ได้ดังนี้

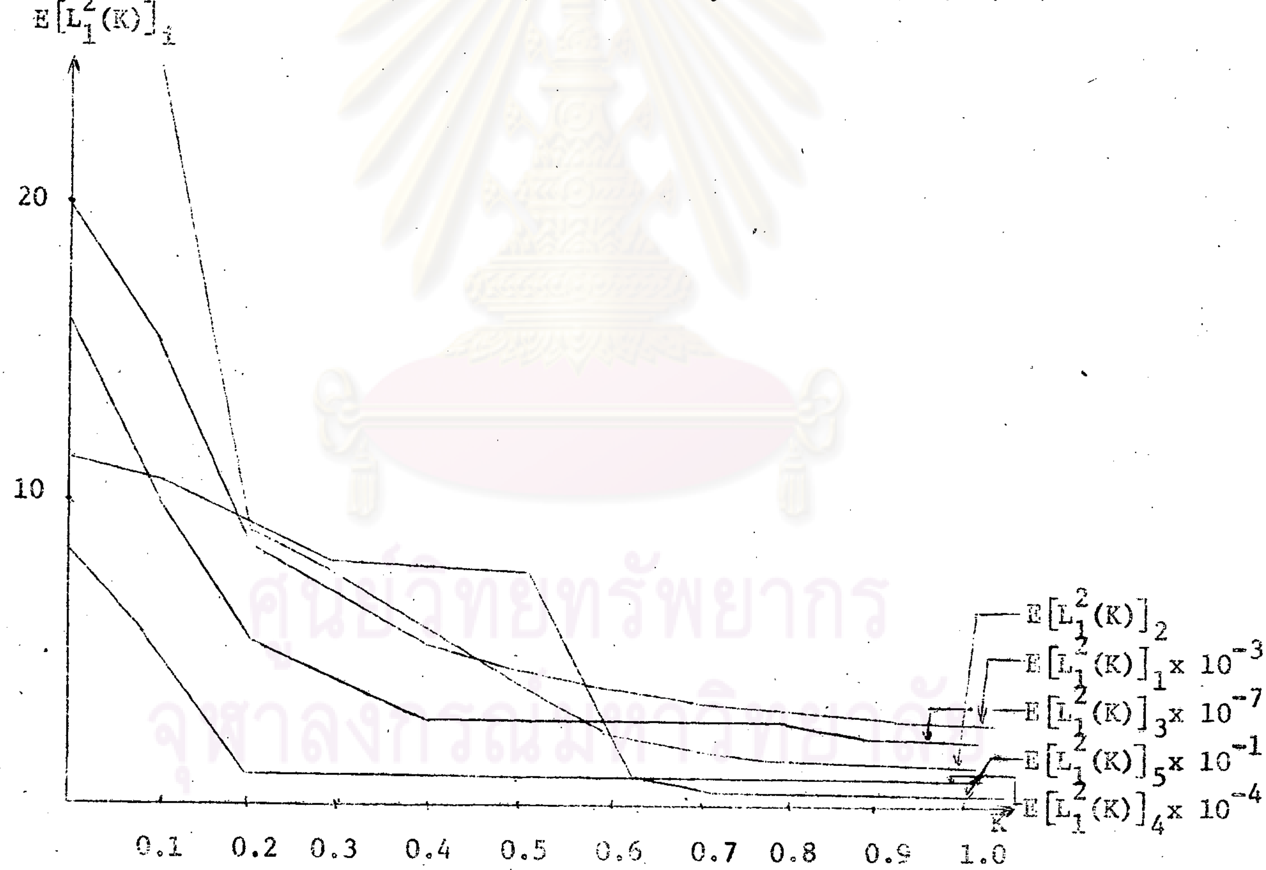
๑. ลักษณะค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของการประมาณค่าพารามิเตอร์ β ที่วิเคราะห์โดยวิธีริดจ์ รีเกรสชั่น

๑.๑ จากการคำนวณค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณค่า $\hat{\beta}^*$ มีค่าเป็น $E[\,L_1^2(K)\,] = \gamma_1(K) + \gamma_2(K)$ เมื่อกำหนดให้ K เป็นค่าคงที่อยู่ระหว่าง (0 - 1] จะเห็นได้ว่าข้อมูลทั้ง ๕ ชุดที่วิเคราะห์โดยวิธีริดจ์ รีเกรสชั่น จะให้ค่า $E[\,L_1^2(K)\,]$ มีลักษณะดังรูปที่ ๖

<u>รูปที่ ๖</u>

แสดงความสัมพันธ์ระหว่าง $E[\,L_1^2(K)]_i$ กับค่า K

ของข้อมูลทั้งหมด ๕ ชุด เมื่อ i = จำนวนชุดของข้อมูลมีค่าเท่ากับ ๑, ๒, ๓, ๔, ๕



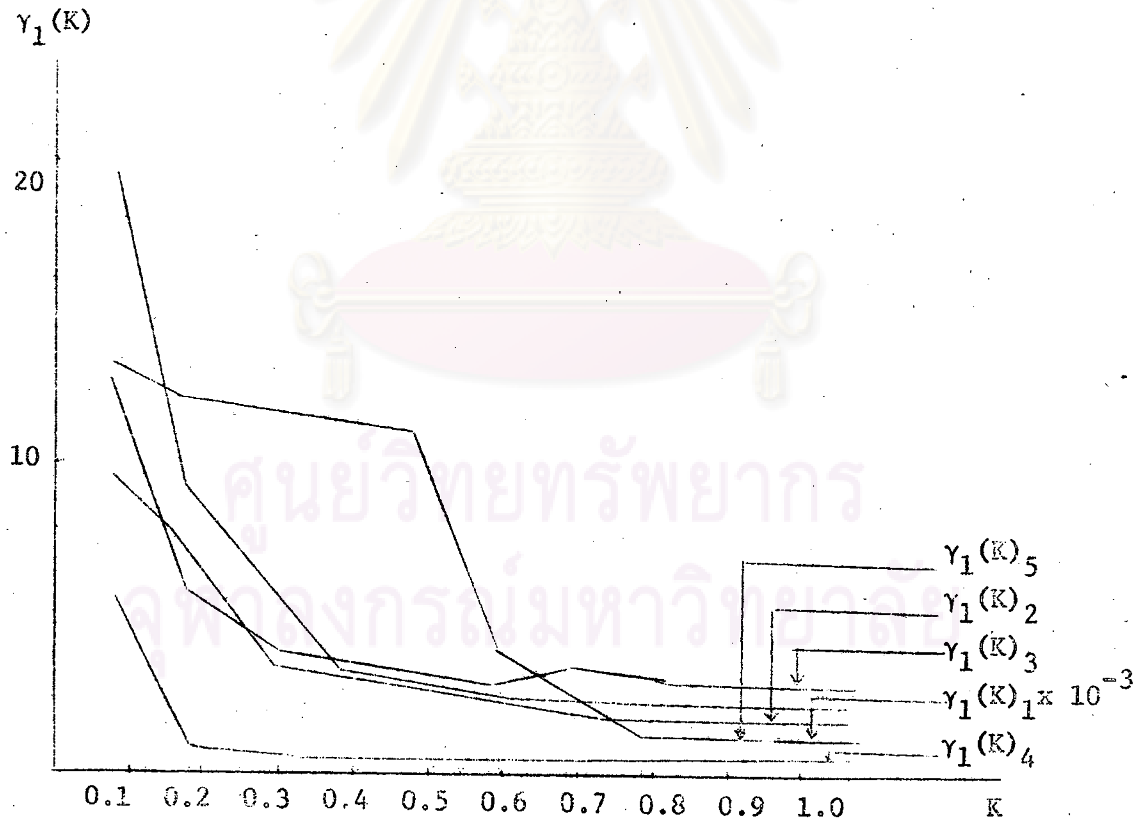
จากกราฟค่า $E[\,L_1^2(K)\,]_i$ และ $E[\,L_1^2(0)\,]_i$ ซึ่งเป็นค่าเฉลี่ยความคลาดเคลื่อนกำลังสองที่ประมาณค่าโดยวิธีริดจ์ รีเกรสชั่น และวิธีกำลังสองน้อยที่สุดตามลำดับ ดังนั้น $E[\,L_1^2(K)\,] = E[\,L_1^2(0)\,]$ เมื่อ $K = 0$ ซึ่งสอดคล้องกับคุณสมบัติของค่าเฉลี่ยความคลาดเคลื่อน

เคลื่อนกำลังสองของ (๒.๕.๖) ของการพิสูจน์คุณสมบัติของ $E[\ L_1^2(K)\ ]$ กล่าวคือ ลักษณะของ $E[\ L_1^2(K)\ ]$ จะเป็น monotonically decreasing ช่วงแรกและจะมีลักษณะเป็นเส้นตรงที่ค่อนข้างจะ Stabilize ในช่วงหลัง ซึ่งมีค่า $K \longrightarrow 1$

๑.๒ เมื่อ K มีค่าเพิ่มขึ้นเรื่อย ๆ โดยกำหนดให้ $K > 0$ ($K = 0.10, 0.20, 0.30, 0.40, \ldots, 0.90, 1.0$) เส้นโค้งของความแปรปรวนของตัวประมาณ β มีลักษณะดังรูปที่ ๗

รูปที่ ๗

แสดงความสัมพันธ์ระหว่างค่าความแปรปรวนของ $\hat{\beta}^*$ กับค่า K เมื่อ $K\epsilon(0,1]$



จากกราฟความแปรปรวนของตัวประมาณ $\hat{\beta}^*$ เมื่อ K มีค่าเพิ่มขึ้นเรื่อย ๆ (โดยที่ $K = 0.10, 0.20, \ldots, 0.90, 1.0$) เส้นโค้งของความแปรปรวนของตัวประมาณ

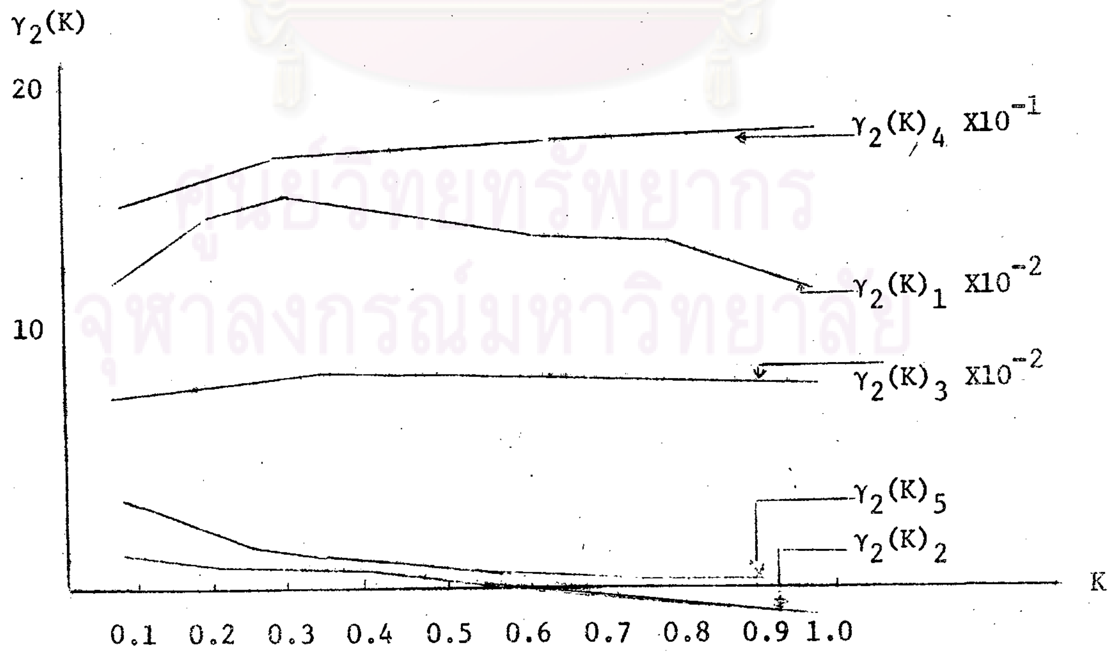
$\hat{\beta}^*[\ \gamma_1(K)\ ]$ จะค่อย ๆ ลดลงเรื่อย ๆ และลักษณะการลดลงแบบ monotonically decreasing ฟังก์ชั่นของ K ซึ่งสอดคล้องกับทฤษฎี 2.5.1 ของการพิสูจน์คุณสมบัติของ $E[\ L_1^2(K)\ ]$

จากรูปที่ ๖ - ๗ สรุปได้ว่าค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง $E[\ L_1^2(K)\ ]$ และค่าความแปรปรวนของ $\hat{\beta}^*$ ของข้อมูลทั้ง ๕ ชุด มีลักษณะของรูปกราฟเหมือนกัน ดังนั้น $E[\ L_1^2(K)\ ]$ หรือ $\gamma_1(K)$ อาจเป็นตัวแทนของการเปลี่ยนแปลงของกันและกันได้ กล่าวคือ $E[\ L_1^2(K)\ ]$ แปรผันตรงกับ $\gamma_1(K)$

๑.๓ เมื่อ K มีค่าอยู่ระหว่าง (0, 1] โดยที่ K = 0.10, 0.20, 0.30, ..., 0.90, 1.00 เส้นโค้งของความเอนเอียงกำลังสอง $\gamma_2(K)$ จะมีการเปลี่ยนแปลงดังรูปที่ ๘

<u>รูปที่ ๘</u>

แสดงความสัมพันธ์ระหว่างความเอนเอียงกำลังสองกับค่า K



จากกราฟรูปที่ ๘ สรุปผลเป็นดังนี้

๑.๓.๑ ความเอนเอียงกำลังสอง $\gamma_2(K)$ ของข้อมูลชุดที่ ๑, ๒ มีลักษณะเป็น monotonically decreasing ฟังก์ชั่นของ K ซึ่งค้านกับทฤษฎี 2.5.2 ของการพิสูจน์คุณสมบัติของ $E[L_1^2(K)]$ เพราะผลเนื่องมาจากข้อมูลทั้ง ๓ นี้มีความแปรปรวนของ $\hat{\beta}$ ที่คำนวณด้วยวิธีกำลังสองน้อยที่สุดนั้นมีค่าไม่สูงมากนัก ดังนั้นเมื่อนำข้อมูลทั้ง ๓ ชุดนี้มาวิเคราะห์โดยวิธีริดจ์ รีเกรสชั่นอาจมีผลไม่ดีต่อตัวประมาณค่า $\hat{\beta}$ ดัง Richard F. Gust and Robert L. Mason ได้กล่าวไว้ว่า "ถ้า σ^2 มีค่าน้อย ตัวประมาณ ที่ประมาณค่าโดยวิธีกำลังสองน้อยที่สุดจะเป็นตัวประมาณค่าที่ดี ถึงแม้จะเกิดสหสัมพันธ์ในตัวแปรอิสระก็ตาม" (จากเรื่องเดิม (๑) หน้า ๖๑๗)

๑.๓.๒ ความเอนเอียงกำลังสอง $\gamma_2(K)$ ของข้อมูลชุดที่ ๓, ๔, ๑๕ มีลักษณะเป็น monotonically increasing ฟังก์ชั่นของ K ซึ่งสอดคล้องกับทฤษฎี 2.5.2 ของการพิสูจน์คุณสมมติของ $E[L_1^2(K)]$

ดังนั้นจะเห็นได้ว่าจากข้อมูลชุดที่ ๓, ๔ ถ้า K มีค่าเข้าใกล้ค่าอนันต์จะมีผลทำให้ $\gamma_2(K)$ มีค่ามากที่สุดคือมีค่าเท่ากับ $\gamma_2(K)$ max

๓.๔ <u>การเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองทั้ง ๒ วิธี</u>

จากตารางที่ ๒ - ๑๖ ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของข้อมูลทั้ง ๕ ชุดที่วิเคราะห์โดยวิธีริดจ์ รีเกรสชั่น และวิธีกำลังสองน้อยที่สุด ส่วนตารางที่ ๑๗ แสดงการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองทั้ง ๒ วิธี

ตารางที่ ๑๗

แสดงการเปรียบเทียบค่า $E[L_1^2(0)]$ และ $E[L_1^2(K)]$ ระหว่างวิธีกำลังสองน้อยที่สุดและวิธีริดจ์ รีเกรสชั่น ตามลำดับ และแสดงเปอร์เซนต์ของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองที่ลดลงเมื่อเปรียบเทียบระหว่าง ๒ วิธี

ข้อมูลชุดที่	วิธีกำลังสองน้อยที่สุด		วิธีริดจ์ รีเกรสชั่น	
	ค่า $E[L_1^2(0)]$ จากวิธีที่มีตัวแปรอิสระทุกตัวในตัวแบบ	ค่า $E[L_1^2(0)]$ จากวิธี Step-Wise Multiple Reg.	ณ. ค่า K เริ่มให้ $E[L_1^2(K)] < E[L_1^2(0)]$	ณ. ค่า $K = 2\sigma^2/\hat{\beta}'\hat{\beta}$ ให้ค่า $E[L_1^2(K)]$
1	2,375.4920	1,834.7761	K = 1.01 1,950.686 (17.88%; 06.32^{+}%)	K = 3.65 1,838.870 (22.59%; $2.23 \times 10^{-3+}$%)
2	7.8551	13.8683	K = 0.26 7.7731 (0.01%; 43.95%)	K = 0.40 4.2124 (46.37%; 69.63%)
3	13,011,884.00	11,161,596.800	K = 0.11 10,421,530.00 (19.91%; 6.63%)	K = 20.72 95,396.875 (99.26%; 99.15%)
4	129,889.00	288,835.7666	K = 0.07 112,655.09 (13.27%; 60.99%)	K = 16.00 279.99 (99.78%; 99.90%)
5	442.649	732.7367	K = 0.01 125.7630 (71.59%; 82.84%)	K = 1.50 6.6940 (98.98 ; : 99.08%)

(% ; %) หมายถึง จำนวนเปอร์เซนต์ของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองที่ลดลงเมื่อเปรียบเทียบระหว่างวิธีที่มีตัวแปรอิสระทุกตัวในตัวแบบ และวิธี Step-Wise Multiple Regression ณ. ค่า K ที่ต่างกัน

จากตารางที่ ๑๗ แสดงให้เห็นว่าเมื่อนำข้อมูลทั้ง ๕ ชุดที่ตัวแปรอิสระมีสหสัมพันธ์ระหว่าง -๐.๐๑-๐.๙๙ (จากภาคผนวก) มาคำนวณหาค่าตัวประมาณ $\hat{\beta}$ และค่าเฉลี่ยความคลาดเคลื่อนกำลังสองโดยการเปรียบเทียบการประมาณค่าสัมประสิทธิ์ความถดถอยระหว่างวิธีกำลังสองน้อยที่สุด และวิธีริดจ์ รีเกรสชั่น จากวิธีกำลังสองน้อยที่สุดจะคำนวณค่าสถิติต่าง ๆ ด้วยวิธีที่มีตัวแปรอิสระทุกตัวในตัวแบบ และวิธี Step-Wise Multiple Regression ส่วนวิธี Ridge Regression จะคำนวณด้วยวิธี Ordinary Ridge Regression และเมื่อนำข้อมูลทั้ง ๕ ชุดมาศึกษาเปรียบเทียบผลปรากฏดังนี้ คือ ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง ($E[\, L_1^2(K) \,]$)น้อยกว่าค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง ($E[\, L_1^2(0) \,]$) เมื่อคำนวณด้วยวิธีริดจ์ รีเกรสชั่น และวิธีกำลังสองน้อยที่สุดตามลำดับ และที่ $K > 0$ คือ $K = 2\sigma^2/\hat{\beta}'\hat{\beta}$ สามารถนำไปใช้กับข้อมูลได้ทุกกรณีที่ให้คุณสมบัติ $E[\, L_1^2(K) \,] < E[\, L_1^2(0) \,]$

เปอร์เซนต์การลดลงของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง ($E[\, L_1^2(K) \,]$) เมื่อเปรียบเทียบกับค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง ($E[\, L_1^2(0) \,]$) อธิบายได้ดังนี้

ข้อมูลชุดที่ ๑ คำนวณตัวประมาณ $\hat{\beta}^*$ ด้วยวิธีริดจ์ รีเกรสชั่น เมื่อ $K = 1.01$ เริ่มจะมีผลทำให้ $E[\, L_1^2(K) \,] < E[\, L_1^2(0) \,]$ ประมาณ ๑๗.๘๘% และ ๖.๓๒$^+$% เมื่อเปรียบเทียบกับวิธีที่มีตัวแปรอิสระทุกตัวในตัวแบบ และวิธี Step-Wise Multiple Regression ตามลำดับ ส่วน $K = 3.65$ จะให้ค่า $E[\, L_1^2(K) \,] < E[L_1^2(0)]$ ประมาณ ๒๒.๕๙% และ $2.23 \times 10^{-3+}$% เมื่อเปรียบเทียบกับวิธีที่มีตัวแปรอิสระทุกตัวในตัวแบบและวิธี Step-Wise Multiple Regression ตามลำดับ ในทำนองเดียวกันข้อมูลชุดที่ ๒ - ๕ สามารถอธิบายการเปลี่ยนแปลงค่าเฉลี่ยความคลาดเคลื่อนกำลังสองระหว่างทั้ง ๒ วิธีได้เหมือนกัน โดยพิจารณาเปอร์เซนต์ที่ $E[\, L_1^2(K) \,] < E[\, L_1^2(0) \,]$ ดังแสดงค่าเหล่านี้ในตารางที่ ๑๗

๓.๕ <u>ความสัมพันธ์ระหว่างตัวประมาณ $\hat{\beta}^*$ และ K ซึ่งเรียกว่าริดจ์เทรซผลเป็นดังนี้</u>

จากรูปที่ ๙ อาจแสดงความสัมพันธ์ระหว่าง $\hat{\beta}^*$ และ K กล่าวคือ เมื่อ K มีค่าเพิ่มขึ้นจะมีผลทำให้ $\hat{\beta}_i^*$ แต่ละค่าลดลง ดังนั้น อาจสรุปได้ว่า $\hat{\beta}^*$ แปรผกผันกับค่า K เมื่อ K > 0 แสดงผลด้วยกราฟได้ดังนี้

<u>รูปที่ ๙</u>

แสดงความสัมพันธ์ระหว่าง $\hat{\beta}^*$ กับค่า K (Ridge Trace) เมื่อ $K \epsilon (0, 1]$

(แสดงริดจ์เทรซของข้อมูลชุดที่ ๑)

จากรูปที่ ๙ แสดงริดจ์เทรซของข้อมูลชุดที่ ๑ กล่าวคือเมื่อเพิ่ม K มากกว่าศูนย์เพิ่มขึ้นเรื่อย ๆ พบว่า ค่า $\hat{\beta}_i^*$ เริ่มจะลดลง และที่ K = 1.01 จะมีผลเริ่มทำให้ $E[\,L_1^2(K)\,] < E[\,L_1^2(0)\,]$ นั่นคือเส้นโค้งของ $\hat{\beta}^*$ เริ่มจะคงที่ (Stabilize) และคงที่เมื่อ K มีค่าเหมาะสมที่สุด $K = 2\sigma^2/\hat{\beta}'\hat{\beta}$

๓.๖ เปรียบเทียบค่าตัวประมาณพารามิเตอร์ β เมื่อคำนวณด้วยวิธีทั้ง ๒

ตารางที่ ๑๘

แสดงการเปรียบเทียบค่าตัวประมาณพารามิเตอร์ β เมื่อคำนวณด้วยวิธีทั้ง ๒

ข้อมูลชุดที่	ตัวแปรอิสระ	ตัวประมาณค่าพารามิเตอร์ β			
		วิธีกำลังสองน้อยที่สุด		วิธีริดจ์ รีเกรสชั่น	
		วิธีที่มีตัวแปรอิสระทุกตัวในตัวแบบ	วิธี Step-Wise Multiple Regression	ณ. K เริ่ม optimum	ณ. K optimum
				K = 1.01	K = 3.65
	ค่าคงที่	-1154.9845	-1209.9250	-21.78463	-12.2544
	X_1	-2.3627	-	- 9.25428	- 9.5444
	X_2	720.1128	624.899	0.44326	3.0052
1	X_3	144.6253	133.485	143.29104	133.8502
	X_4	-15.6423	-	8.06139	17.7484
	X_5	-28.4925	-	34.10857	41.9313
				K = 0.26	K = 0.40
	ค่าคงที่	-154.1349	-5.0500	-0.6268	-0.40038
	X_1	17.7881	-	2.8856	2.1971
	X_2	-10.6353	-	-1.8594	-1.2167
2	X_3	-0.3613	-0.7568	-1.8591	-1.5325
	X_4	0.4828	-	2.2026	1.8003
	X_5	4.1744	4.7482	4.1875	4.2878
				K = 0.11	K = 20.72
	ค่าคงที่	18158.0000	255.1452	1355.9931	19.4291
	X_1	-964.8763	-	-1580.5365	-311.7366
	X_2	1.3132	0.7847	1.0486	1.0109
3	X_3	-0.1786	-	-0.1002	-0.1078

ข้อมูลชุดที่	ตัวแปรอิสระ	ตัวประมาณค่าพารามิเตอร์ β			
		วิธีกำลังสองน้อยที่สุด		วิธีริดจ์ รีเกรสชั่น	
		วิธีที่มีตัวแปรอิสระทุกตัวในตัวแบบ	วิธี Step-Wise Multiple* Regression	ณ. K เริ่ม optimum	ณ. K optimum
	X_4	2.3997	4.2797	2.0116	1.6410
	X_5	-23.3318	-	7.1184	9.7358
				K = 0.07	K = 16.07
	ค่าคงที่	2118.5320	4433.114	137.1548	$0.0115\ X10^{-5}$
	X_1	0.1500	-	0.1516	$0.0844\ X10^{-1}$
	X_2	-1.8808	-	-0.7996	$0.0116\ X10^{-2}$
4	X_3	-2.3398	-2.678	-3.1839	$-0.0193\ X10^{-2}$
	X_4	-1.9487	-	-1.9946	$-0.0798\ X10^{-3}$
	X_5	-0.8551	2.373	-0.7446	$0.0675\ X10^{-2}$
	X_6	0.0260	-	-0.0261	$0.07849\ X10^{-1}$
	X_7	-1.7039	-	-1.8339	$0.01314\ X10^{-3}$
	X_8	44.9416	-	40.9435	$0.0421\ X10^{-3}$
				K = 0.01	K = 1.5
	ค่าคงที่	0.0216	-180.6204	-2.2178	-0.1279
	X_1	9.6678	27.5649	9.6712	5.3599
	X_2	0.0085	-	0.0085	0.0054
5	X_3	-22.2921	-25.0161	-22.2501	-4.6131
	X_4	-0.0311	-	-0.0311	-0.0241
	X_5	0.2723	-	0.2724	0.3744
	X_6	-2.8009	1.1123	-2.8025	-3.6269
	X_7	-	-	-	-
	X_8	0.9907	-	0.9909	1.3717

หมายเหตุ ณ. K เริ่ม optimum หมายถึงค่า K ที่เริ่มให้ค่า $E\left[L_1^2(K)\right] < E\left[L_1^2(0)\right]$

จากตารางที่ ๑๘ สามารถสรุปได้ว่า ณ. $K > 0$ มีผลให้ $E[L_1^2(K)]$ น้อยกว่า $E[L_1^2(0)]$ และพบว่าค่าประมาณพารามิเตอร์ β คือ $\hat{\beta}$ และ $\hat{\beta}^*$ ที่ประมาณค่าโดยวิธีกำลังสองน้อยที่สุดและวิธีริดจ์ รีเกรสชั่น ตามลำดับ ให้ค่าไม่แตกต่างกัน เมื่อนำค่าตัวประมาณค่า $\hat{\beta}$, $\hat{\beta}^*$ แทนค่าในสมการประมาณค่าตัวแปรตามจะได้ค่า $\hat{Y}$ ดังตารางที่ ๑๙

ตารางที่ ๑๙

แสดงค่าประมาณของตัวแปรตามที่เกิดจากการประมาณค่าระหว่างวิธีกำลังสองน้อยที่สุดและวิธีริดจ์ รีเกรสชั่น

ข้อมูลชุดที่	ค่าจริงของตัวแปรตาม (Y_i)	ค่าประมาณของตัวแปรตาม ($\hat{Y}$)			
		วิธีกำลังสองน้อยที่สุด		วิธีริดจ์ รีเกรสชั่น	
		วิธีที่มีตัวแปรอิสระทุกตัวในตัวแบบ	วิธี Step-Wise Multiple Regression	ณ. K เริ่ม optimum	ณ. K optimum
1	177.0340	153.2694	149.1603	103.3300	94.2172
	288.6678	285.8633	296.9717	239.9034	235.2062
	427.7908	453.7769	462.8882	453.2959	454.1821
	612.9983	652.7661	649.4702	718.2881	724.9648
	816.5131	833.5576	834.7734	984.8696	991.9636
	1,118.5339	1,067.0027	1,049.2400	1,075.2810	1,080.8417
	1,382.6035	1,404.2319	1,404.2520	1,408.7173	1,420.4138
	1,695.0153	1,667.2988	1,669.8960	1,630.1724	1,640.9746
	2,001.9974	2,027.6165	2,037.9390	2,016.6440	2,019.0832
	2,405.9039	2,340.4656	2,342.2080	2,365.6057	2,362.6545
	2,822.0935	2,861.6602	2,839.1580	2,862.7097	2,850.2795
	2,926.6287	2,884.0891	2,894.6520	2,924.9295	2,919.8403
	3,296.3448	3,325.5298	3,336.5620	3,226.3176	3,217.3713
2	35.064	35.757	39.649	39.629	38.553
	39.155	39.922	36.136	36.247	35.444
	42.060	38.726	37.369	37.393	37.493
	43.861	44.484	44.065	44.099	45.145
	48.170	49.224	51.268	51.121	53.855
	56.395	55.942	57.059	56.853	56.274
	66.384	66.647	66.794	66.922	65.181
	105.554	107.985	105.997	106.044	105.676
	127.221	125.104	125.744	125.221	126.238
3	21,126.10	22,475.79	20,311.39	22,334.01	20,390.30
	24,672.10	24,845.42	23,516.32	24,684.52	23,643.49
	28,224.30	25,680.30	25,935.80	25,766.08	25,941.79
	32,514.30	30,116.49	30,425.04	29,972.93	31,099.18
	28,698.90	33,997.65	36,167.97	34,480.11	35,747.96
	48,405.50	47,027.92	46,505.43	46,742.90	47,039.61
	58,744.70	59,267.16	60,388.90	59,246.24	59,463.71

ตารางที่ ๑๙ (ต่อ)

ข้อมูลชุดที่	ค่าจริงของตัวแปรตาม (Y_i)	ค่าประมาณของตัวแปรตาม ($\hat{Y}$)			
		วิธีกำลังสองน้อยที่สุด		วิธีริดจ์ รีเกรสชั่น	
		วิธีที่มีตัวแปรอิสระทุกตัวในตัวแบบ	วิธี Step-Wise Multiple Regression	ณ. K เริ่ม optimum	ณ. K optimum
	74,603.30	72,139.69	74,986.81	72,785.50	72,942.81
	87,698.40	88,946.75	88,897.06	89,828.13	87,912.84
	107,171.70	107,425.063	105,223.80	108,008.50	107,304.84
4	3,045	2,985.06	2,910.99	2,886.71	2,732.29
	3,809	4,022.60	4,588.35	4,015.31	3,263.27
	4,600	4,667.65	4,298.51	4,691.82	3,922.43
	4,615	4,609.08	4,230.21	4,633.95	4,266.72
	4,089	3,249.26	4,171.95	3,300.25	4,171.26
	4,161	3,715.86	4,198.52	3,764.28	4,607.70
	4,890	5,098.90	4,522.96	5,094.28	4,888.96
	5,392	5,848.37	6,118.32	5,864.14	5,819.82
	7,818	7,846.37	7,228.95	7,848.69	6,522.42
	9,190	8,859.18	8,501.00	8,822.85	7,385.11
	8,125	8,158.31	8,112.79	8,150.64	6,736.03
	8,274	8,245.91	8,591.13	8,208.55	8,227.95
	8,469	8,622.65	8,869.26	8,552.12	9,041.85
5	125.077	100.505	82.774	100.528	122.685
	81.945	96.845	99.596	96.879	125.229
	117.854	119.135	122.305	119.186	139.393
	223.067	218.705	207.508	218.665	201.894
	184.317	209.589	223.039	209.569	192.684
	233.726	211.922	251.768	211.922	195.401
	237.304	243.047	266.256	243.081	241.454
	324.459	326.223	301.570	326.265	336.584
	357.559	354.091	337.583	254.113	329.705
	364.680	358.810	338.825	358.858	361.384
	312.302	307.509	323.037	307.540	296.968
	362.316	351.014	372.848	351.073	343.518
	382.298	389.232	373.894	389.318	403.711
	400.000	397.209	400.896	395.254	404.535

จากตารางที่ ๑๙ นำค่าประมาณของตัวแปรตามมาทดสอบหาความแตกต่างกับค่าจริงและในการศึกษาวิจัยครั้งนี้ได้แยกการทดสอบค่าประมาณของตัวแปรตามเป็น ๒ กรณี

๑. นำค่าประมาณของตัวแปรตามที่คำนวณได้จากวิธีกำลังสองน้อยที่สุดและวิธีริดจ์ รีเกรสชั่น มาทดสอบว่ามีความคลาดเคลื่อนจากค่าสังเกตของตัวแปรตามอย่างมีนัยสำคัญทางสถิติหรือไม่ โดยตั้งสมมติฐานดังนี้คือ

$$H_o : Y = \hat{Y}_i$$

$$H_1 : Y \neq \hat{Y}_i$$

เมื่อ i หมายถึงวิธีการประมาณค่าพารามิเตอร์ β ของแต่ละวิธี และผลจากการทดสอบข้อมูลของตัวแปรตามและตัวประมาณค่าของตัวแปรตามดังแสดงไว้ในตารางที่ ๒๐

๒. ทดสอบดูว่าค่าประมาณของตัวแปรตาม ณ. K ที่เริ่ม optimum และ K optimum ว่ามีความแตกต่างกันหรือไม่ โดยตั้งสมมติฐานดังนี้คือ

$$H_o : \hat{Y}_{ORR} = \hat{Y}_{RR}$$

$$H_1 : \hat{Y}_{ORR} \neq \hat{Y}_{RR}$$

เมื่อ $\hat{Y}_{ORR}$ = ค่าประมาณของตัวแปรตามที่ประมาณค่าโดยวิธีริดจ์ รีเกรสชั่น ณ. K ที่เริ่ม optimum

$\hat{Y}_{RR}$ = ค่าประมาณของตัวแปรตามที่ประมาณค่าโดยวิธีริดจ์ รีเกรสชั่น ณ. K optimum

การทดสอบความแตกต่างระหว่างค่าประมาณของค่าประมาณของตัวแปรตามของวิธีริดจ์ รีเกรสชั่น ณ. K ที่เริ่ม optimum และ K ที่ optimum สรุปได้ดังตารางที่ ๒๐

ตารางที่ ๒๐

เปรียบเทียบการทดสอบสมมติฐานของค่าประมาณของตัวแปรตามที่ประมาณค่าจากวิธีกำลังสองน้อยที่สุดและวิธีริดจ์ รีเกรสชั่น กับค่าสังเกตของตัวแปรตามสำหรับข้อมูลชุดที่ ๑-๔

ตัวแปรตาม (Y)	ค่าประมาณจาก วิธีที่มีตัวแปรอิสระทุกตัวในตัวแบบ	ค่าประมาณจาก วิธี Step-Wise Multiple Regression	ค่าประมาณจาก วิธีริดจ์ รีเกรสชั่น ณ.K เริ่ม optimum	ค่าประมาณจาก วิธีริดจ์ รีเกรสชั่น ณ. K optimum
ค่าสังเกต	ยอมรับ H_o	ยอมรับ H_o	ยอมรับ H_o	ยอมรับ H_o
ค่าประมาณจากที่มีตัวแปรอิสระทุกตัวในตัวแบบ		ยอมรับ H_o	ยอมรับ H_o	ยอมรับ H_o
ค่าประมาณจากวิธี Step-Wise Multiple Regression			ยอมรับ H_o	ยอมรับ H_o
ค่าประมาณจากวิธีริดจ์ รีเกรสชั่น ณ. K เริ่ม optimum				ยอมรับ H_o

จากตารางที่ ๒๐ เมื่อทดสอบค่าประมาณของตัวแปรตามที่ประมาณค่าจากวิธีกำลังสองน้อยที่สุด และวิธีริดจ์ รีเกรสชั่น กับค่าจริงของตัวแปรตามของข้อมูลทั้ง ๕ ชุด สรุปได้ดังนี้

๑. ค่าประมาณของตัวแปรตามที่ประมาณค่าโดยวิธีที่มีตัวแปรอิสระทุกตัวในตัวแบบ วิธี Step-Wise Multiple Regression วิธีริดจ์ รีเกรสชั่น ณ. K เริ่ม optimum และ ณ. K optimum เมื่อนำมาทดสอบกับค่าสังเกตของตัวแปรตามปรากฏว่าให้ค่าประมาณโดยวิธีดังกล่าวเหล่านี้ไม่แตกต่างจากค่าสังเกตของตัวแปรตาม

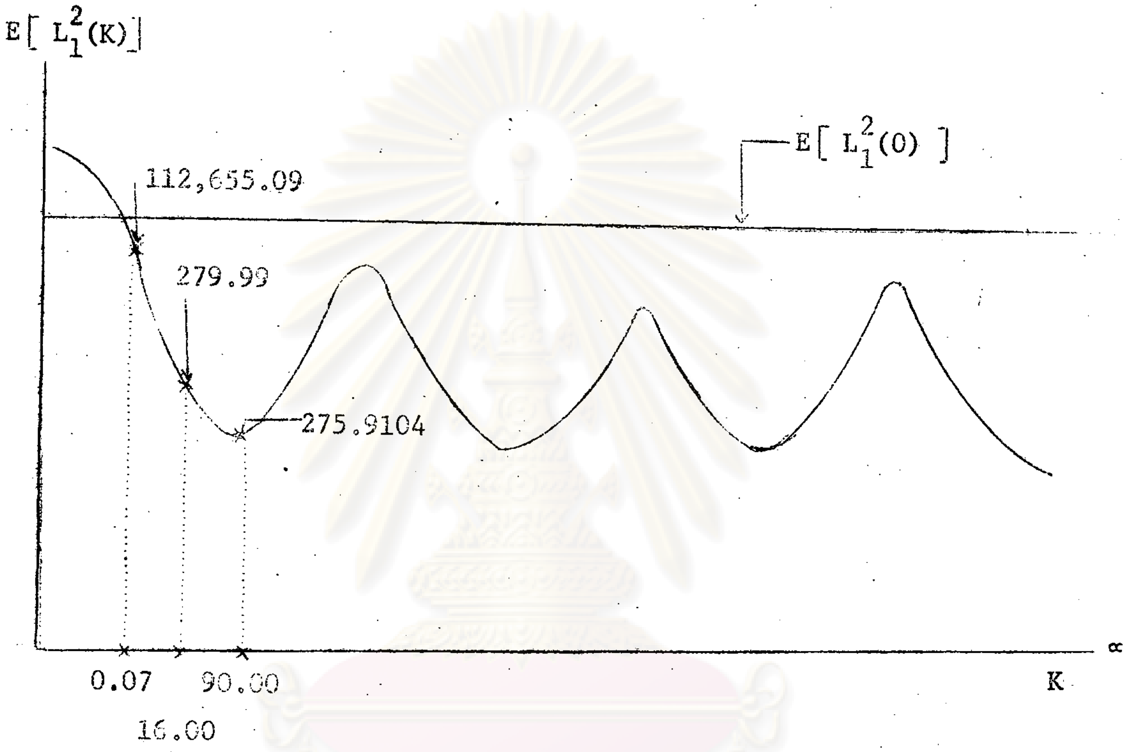
๒. ค่าประมาณของตัวแปรตามที่ประมาณค่าโดยวิธีริดจ์ รีเกรสชั่น ณ. K เริ่ม optimum และ K optimum ให้ค่าประมาณของตัวแปรตามไม่แตกต่างกันอย่างมีนัยสำคัญ

๓.๗ ความสัมพันธ์ระหว่างค่า $E[L_1^2(0)]$ และ $E[L_1^2(K)]$ เมื่อ $K \varepsilon (0, \infty]$

เนื่องจากในการวิจัยครั้งนี้มุ่งที่จะหาวิธีการหรือแนวทางที่เหมาะสมที่จะนำไปใช้กับข้อมูลในทางปฏิบัติในการวิเคราะห์สัมประสิทธิ์ความถดถอยในกรณีที่เกิดสหสัมพันธ์ในตัวแปรอิสระเท่านั้น ดังนั้น ในการศึกษาวิจัยครั้งนี้จะแสดงให้เห็นการเปลี่ยนแปลงการกระจายค่าของ $E[L_1^2(K)]$ ให้เห็นจริงเพียงข้อมูลชุดที่ ๔ โดยที่มีการเปลี่ยนแปลงค่า K เพิ่มขึ้นเรื่อย ๆ ดังแสดงในรูปที่ ๑๐

รูปที่ ๑๐

แสดงการเปลี่ยนแปลงค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของข้อมูลชุดที่ ๔



จากข้อมูลชุดที่ ๔ เมื่อ K > 0 และเปลี่ยนแปลงค่า K โดยเพิ่มครั้งละ ๐.๐๑ เมื่อคำนวณตัวประมาณ $\hat{\beta}^*$ จะมีผลทำให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีลักษณะเป็น increasing และ decreasing สลับกันไป ดังนั้นเมื่อเพิ่มค่า K มากขึ้นเรื่อย ๆ ค่า $E[L_1^2(K)]$ มีลักษณะดังรูปข้างบน เมื่อนำข้อมูลทั้ง ๕ ชุด มาคำนวณหาค่าประมาณของพารามิเตอร์ ด้วยวิธีกำลังสองน้อยที่สุด และวิธีริดจ์ รีเกรสชั่น ได้ข้อสรุปต่าง ๆ ดังตารางที่ ๒๑

ตารางที่ ๒๑

แสดงคุณสมบัติของการประมาณค่า ด้วยวิธีกำลังสองน้อยที่สุดและวิธีริดจ์ รีเกรสชั่น เมื่อนำไปใช้กับข้อมูลทั่วไปในทางปฏิบัติ

คุณสมบัติ	วิธี		
	ที่มีตัวแปรอิสระทุกตัวในตัวแบบ	Step-Wise Multiple Regression	Ridge Regression
๑ Minimium Variance			√
๒ ค่าประมาณไม่เอนเอียง	√	√	
๓ ประหยัดเวลาในการคำนวณและเลือกตัวแปร		√	√
๔ คำนวณง่าย	√	√	
๕ ข้อมูลมีสหสัมพันธ์ในตัวแปรอิสระและความแปรปรวนสูง			√
๖ ความเหมาะสมในการเลือกใช้กรณีทั่วไป		√	
๗ ค่าประมาณของตัวแปรอิสระเหมือนกัน	√	√	√
๘ ค่าสมบูรณ์ของค่าประมาณ β มีคุณภาพไม่ดีควรใช้			√
๙ ข้อมูลมีลักษณะ nearly orthogonal	√		
๑๐ ไม่ประหยัดเวลาในการคำนวณโดยทั่วไป			√
๑๑ สามารถควบคุมค่าเฉลี่ยความคลาดเคลื่อนกำลังสองให้ลดลง			√

ข้อสรุปที่ได้จากตารางที่ ๒๑ มีดังนี้

๑. สำหรับการประมาณค่าพารามิเตอร์ β พบว่าวิธีริดจ์ รีเกรสชั่น เป็นวิธีที่ให้ค่าความแปรปรวนต่ำสุดเมื่อเปรียบเทียบกับวิธีกำลังสองน้อยที่สุด

๒. ตัวประมาณค่า $\hat{\beta}^*$ เป็นตัวประมาณที่เอนเอียง

๓. เมื่อมีตัวแปรอิสระจำนวนมากควรใช้วิธี Step-Wise Multiple Regression และวิธีริดจ์ รีเกรสชั่น ประมาณค่า β เพราะสะดวกในการคำนวณและไม่ต้องเสียเวลาในการพิจารณาว่าจะตัดตัวแปรอิสระใดออกจากตัวแบบ

๔. การประมาณค่าพารามิเตอร์ β ด้วยวิธีกำลังสองน้อยที่สุด คือวิธีที่มีตัวแปรอิสระทุกตัวในตัวแบบ และวิธี Step-Wise Multiple Regression เป็นวิธีที่คำนวณง่ายกว่าวิธีริดจ์ รีเกรสชั่น และเหมาะที่จะนำไปใช้ในการประมาณค่าพารามิเตอร์ β กับข้อมูลทั่วไป เว้นแต่กรณีที่ข้อมูลมีสหสัมพันธ์ในตัวแปรอิสระสูงและให้ค่าเฉลี่ยความคลาดเคลื่อนสูง ควรใช้วิธีริดจ์ รีเกรสชั่น จะมีผลทำให้ค่าเฉลี่ยความคลาดเคลื่อนลดน้อยลง

๕. กรณีที่ข้อมูลมีลักษณะเข้าใกล้ orthogonal ควรใช้วิธีที่มีตัวแปรอิสระทุกตัวในตัวแบบ แต่กรณีข้อมูลเกิด ill-condition (nonorthogonal) ควรใช้วิธี Step-Wise Multiple Regression และวิธีริดจ์ รีเกรสชั่น แทน แต่ถ้าต้องการให้ค่าเฉลี่ยความคลาดเคลื่อนลดน้อยลงควรใช้วิธีริดจ์ รีเกรสชั่น จะให้ผลดีกว่าวิธีทั้ง ๒

๖. กรณีที่ค่าสมบูรณ์ของตัวประมาณ $\hat{\beta}$ มีค่ามากควรใช้วิธีริดจ์ รีเกรสชั่น แทนวิธีกำลังสองน้อยที่สุด เพราะจะมีผลทำให้ค่าสมบูรณ์ $\hat{\beta}^*$ ลดลง

๗. การประมาณค่าตัวแปรตามจากวิธีทั้ง ๓ คือ วิธีที่มีตัวแปรอิสระทุกตัวในตัวแบบ วิธี Step-Wise Multiple Regression และวิธีริดจ์ รีเกรสชั่น จะสังเกตได้จากตารางที่ ๒๐ (เมื่อทดสอบค่า Y และ $\hat{Y}$ ของแต่ละวิธี) พบว่าค่าประมาณของตัวแปรตามที่ได้จากวิธีทั้ง ๓ นี้มีค่าไม่แตกต่างจากค่าสังเกตของตัวแปรตาม