การประมาณค่าสูญหายในการวิเคราะห์การถดถอยพหุคูณ โดยวิธีการถดถอยพหุคูณ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 วิธีการถดถอยสโตแคสติก และวิธีการถดถอยสโตแคสติกใกล้สุด k ตัว ด้วยค่าถ่วงน้ำหนักที่เท่ากัน

MISSING VALUE IMPUTATION IN MULTIPLE REGRESSSION ANALYSIS WITH MULTIPLE REGRESSION, REGRESSION-RATIO-Q1, REGRESSION-RATIO-Q3, STOCHASTIC REGRESSION AND K-NEAREST STOCHASTIC REGRESSION WITH EQUIVALENT WEIGHTED

จัสมิน อิซซาเบลล่า เซเยอร์ส
ภูษิต เกิดมั่งมี
วนัชฌา ภูมาวงศ์

ปัญหาพิเศษนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาสตรบัณฑิต (สถิติประยุกต์)
ภาควิชาสถิติ คณะวิทยาศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ปีการศึกษา 2565

MISSING VALUE IMPUTATION IN MULTIPLE REGRESSSION ANALYSIS WITH MULTIPLE REGRESSION, REGRESSION-RATIO-Q1, REGRESSION-RATIO-Q3, STOCHASTIC REGRESSION AND K-NEAREST STOCHASTIC REGRESSION WITH EQUIVALENT WEIGHTED

JASMIN ISABELLA SAYERS
PUSIT KERDMANGMEE
WANUTCHA POOMAWONG

A SPECIAL PROBLEM SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR
THE DEGREE OF BACHELOR OF SCIENCE IN APPLIED STATISTICS
DEPARTMENT OF STATISTICS, SCHOOL OF SCIENCE
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG
ACADEMIC YEAR 2022

หัวข้อปัญหาพิเศษ การประมาณค่าสูญหายในการวิเคราะห์การถดถอยพหุคูณ โดยวิธีการถดถอยพหุคูณ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 วิธีการถดถอยสโตแคสติก และวิธีการถดถอยสโตแคสติกใกล้สุด k ตัว ด้วยค่าถ่วงน้ำหนักที่เท่ากัน

Missing Value Imputation in Multiple Regression Analysis with Multiple Regression, Regression-Ratio-Q1, Regression-Ratio-Q3, Stochastic Regression and K-Nearest Stochastic Regression with Equivalent Weighted

ชื่อนักศึกษา นางสาวจัสมิน อิซซาเบลล่า เซเยอร์ส รหัสนักศึกษา 62050755
นายภูษิต เกิดมั่งมี รหัสนักศึกษา 62050813
นางสาววนัชฌา ภูมาวงศ์ รหัสนักศึกษา 62050826

ปริญญา วิทยาศาสตรบัณฑิต (สถิติประยุกต์)

ภาควิชา สถิติ

ปีการศึกษา 2565

อาจารย์ที่ปรึกษา รศ.สายชล สินสมบูรณ์ทอง

คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง (สจล.) อนุมัติให้ปัญหาพิเศษนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาตรบัณฑิต (สถิติ) ประจำปีการศึกษา 2565

คณะกรรมการสอบ	ลายมือชื่อ
รศ.ดร.อัชฌา อระวีพร ประธานกรรมการ	
ผศ.ดร.ยุวดี กล่อมวิเศษ กรรมการ	
รศ.สายชล สินสมบูรณ์ทอง กรรมการและอาจารย์ที่ปรึกษา	

หัวข้อปัญหาพิเศษ	การประมาณค่าสูญหายในการวิเคราะห์การถดถอยพหุคูณ โดยวิธีการถดถอยพหุคูณ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 วิธีการถดถอยสโตแคสติก และวิธีการถดถอยสโตแคสติกใกล้สุด k ตัว ด้วยค่าถ่วงน้ำหนักที่เท่ากัน
ชื่อนักศึกษา	นางสาวจัสมิน อิซซาเบลล่า เซเยอร์ส รหัสนักศึกษา 62050755
	นายภูษิต เกิดมั่งมี รหัสนักศึกษา 62050813
	นางสาววนัชฌา ภูมาวงศ์ รหัสนักศึกษา 62050826
ปริญญา	วิทยาศาสตรบัณฑิต (สถิติประยุกต์)
ภาควิชา	สถิติ
คณะ	วิทยาศาสตร์
มหาวิทยาลัย	สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง (สจล.)
ปีการศึกษา	2565
อาจารย์ที่ปรึกษา	รศ.สายชล สินสมบูรณ์ทอง

บทคัดย่อ

งานวิจัยนี้มีจุดประสงค์เพื่อศึกษาเปรียบเทียบวิธีการประมาณค่าสูญหายในตัวแบบการถดถอยพหุคูณ ด้วยวิธีการประมาณ 5 วิธี คือ วิธีการถดถอยพหุคูณ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1, วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 วิธีการถดถอยสโตแคสติก และวิธีการถดถอยสโตแคสติกใกล้สุด k ตัว ด้วยค่าถ่วงน้ำหนักที่เท่ากัน เมื่อขนาดตัวอย่าง ความแปรปรวนของความคลาดเคลื่อน และร้อยละของจำนวนข้อมูลสูญหายแตกต่างกัน เกณฑ์ที่ใช้ในการพิจารณาคือค่าคลาดเคลื่อนกำลังสองเฉลี่ย กำหนดให้ขนาดตัวอย่างขนาดเล็กเท่ากับ 20, 40 ขนาดปานกลางเท่ากับ 60, 80 และขนาดใหญ่เท่ากับ 100, 120 มีร้อยละข้อมูลสูญหาย 5, 10, 15 และ 20 ซึ่งเป็นการสูญหายอย่างสุ่ม โดยค่าความแปรปรวนของความคลาดเคลื่อนที่นำไปสร้างตัวแบบในแต่ละกรณีจะเท่ากับ 1, 3, 5, 7 และ 9 ผลการวิจัยสรุปได้ว่า ทุกค่าร้อยละของจำนวนข้อมูลสูญหายและทุกค่าความแปรปรวนของความคลาดเคลื่อน กรณีขนาดตัวอย่างเล็ก ขนาดตัวอย่างปานกลางและขนาดตัวอย่างใหญ่ ส่วนใหญ่วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 มีประสิทธิภาพมากที่สุด รองลงมาคือวิธีการถดถอยพหุคูณ

คำสำคัญ : วิธีการถดถอยพหุคูณ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 วิธีการถดถอยสโตแคสติก วิธีการถดถอยสโตแคสติกใกล้สุด k ตัว

Title	Missing Value Imputation in Multiple Regression Analysis with Multiple Regression, Regression-Ratio-Q1, Regression-Ratio-Q3, Stochastic Regression and K-Nearest Stochastic Regression with Equivalent Weighted
Students	Miss Jasmin Isabella Sayers Student ID 62050755
	Mr. Pusit Kerdmangmee Student ID 62050813
	Miss Wanutcha Poomawong Student ID 62050826
Degree	Bachelor of Science (Applied Statistics)
Department	Statistics
School	Science
University	King Mongkut's Institute of Technology Ladkrabang (KMITL)
Academic Year	2022
Advisor	Assoc.Prof.Saichon Sinsomboonthong

Abstract

The objective of this research is to compare the missing value imputation method in the multiple regression model with five methods as followed : the multiple regression, regression-ratio-Q1 , regression-ratio-Q3 , stochastic regression and k-nearest stochastic regression with equivalent weighted methods. The sample sizes, variances of the error and the missing percentages of the missing value are different value. The criterion will be considered by the mean square error. The small sample size is 20 and 40, medium sample size is 60 and 80 and large sample size is 100 and 120. The missing percentages of values are 5, 10, 15 and 20 using missing at random. The variance of the error used to create the model in each case are 1, 3, 5, 7 and 9. The Research result can be concluded that missing percentages of values and every variances of the error also small, medium and large sample sizes, the most of the regression-ratio Q1 method is the maximum efficiency, followed by the multiple regression method.

Keywords : Multiple Regression, Regression-Ratio-Q1 Regression-Ratio-Q3, Stochastic Regression, K-Nearest Stochastic Regression with Equivalent Weighted

กิตติกรรมประกาศ

ปัญหาพิเศษนี้สำเร็จลุล่วงไปได้ด้วยดี เนื่องด้วยได้รับความกรุณาจาก รศ.สายชล สินสมบูรณ์ทอง อาจารย์ที่ปรึกษาปัญหาพิเศษ ผู้ให้คำแนะนำ ข้อเสนอแนะ คำปรึกษา เอื้อเฟื้อเอกสารต่าง ๆ ที่ใช้เป็นแนวทางในการวิเคราะห์ข้อมูลและตรวจทานแก้ไขความถูกต้องตลอดจนติดตามผลงานทุก ๆ ขั้นตอนของการดำเนินงานในการทำปัญหาพิเศษนี้จนกระทั่งเสร็จสมบูรณ์ จึงขอขอบพระคุณด้วยความเคารพอย่างสูงไว้ ณ โอกาสนี้ด้วย

ขอขอบพระคุณ รศ.ดร.อัชฌา อระวีพร และผศ.ดร.ยุวดี กล่อมวิเศษ คณะกรรมการปัญหาพิเศษที่ให้คำปรึกษา และคำแนะนำที่ทำให้ปัญหาพิเศษนี้มีความสมบูรณ์ยิ่งขึ้น

ขอขอบคุณคณาจารย์สาขาวิชาสถิติประยุกต์ทุกท่านที่ได้ประสิทธิ์ประสาทวิชาความรู้ และช่วยเหลือให้คำแนะนำในเรื่องต่าง ๆ มาโดยตลอด

สุดท้ายนี้ ขอขอบพระคุณบิดามารดา ที่คอยให้กำลังใจและสนับสนุนผู้จัดทำปัญหาพิเศษมาโดยตลอด และขอขอบคุณเพื่อน ๆ พี่น้อง ทุกคนที่ให้คำปรึกษา ช่วยเหลือการทำงานมาโดยตลอดจนปัญหาพิเศษนี้สำเร็จลุล่วงมาได้ด้วยดี

จัสมิน อิซซาเบลล่า เซเยอร์ส
ภูษิต เกิดมั่งมี
วนัชฌา ภูมาวงศ์

สารบัญ

สารบัญ (ต่อ)

หน้า

สารบัญตาราง

ตารางที่ **หน้า**

สารบัญตาราง (ต่อ)

ตารางที่ **หน้า**

สารบัญรูป

สารบัญรูป (ต่อ)

รูปที่ หน้า

สารบัญรูป (ต่อ)

สารบัญรูป (ต่อ)

รูปที่ หน้า

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

การถดถอยพหุคูณ (Multiple Regression) เป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรอิสระ (Independent Variable) หลายตัวกับตัวแปรตาม (Dependent Variable) 1 ตัว เพื่อศึกษาว่ามีตัวแปรอิสระตัวใดบ้างที่ร่วมกันประมาณหรืออธิบายการผันแปรของตัวแปรตามได้ มีรูปแบบสมการถดถอยคือ $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_k X_{ki} + \varepsilon_i$ หรือตัวแบบการถดถอยในรูปแบบเมทริกซ์ $\underline{Y} = X\underline{\beta} + \underline{\varepsilon}$ โดยการประมาณค่าพารามิเตอร์มักเกิดปัญหา ซึ่งปัญหาที่ผู้วิจัยพบบ่อยคือข้อมูลที่รวบรวมมาได้นั้นมีจำนวนข้อมูลไม่ครบถ้วนสมบูรณ์ ข้อมูลมีเพียงบางส่วนหรือมีข้อมูลสูญหาย (Missing Data) เครื่องมือการจัดเก็บหรือการถ่ายโอนข้อมูลเกิดความผิดพลาด หรือเกิดจากข้อจำกัดของเทคโนโลยีที่ใช้งาน ทำให้เกิดข้อมูลสูญหายซึ่งส่งผลให้ไม่สามารถใช้ประโยชน์จากข้อมูลชุดนั้นได้เต็มที่ (รัตติกาล, 2555) โดยปกติแล้วการวิเคราะห์การถดถอยพหุคูณจะใช้ข้อมูลหรือค่าสังเกตที่มีความครบถ้วนสมบูรณ์ทุกตัวแปรมาใช้ในการพิจารณา แต่ถ้าหากข้อมูลที่รวบรวมมาได้นั้นมีบางค่าสูญหายไป ไม่ว่าจะด้วยสาเหตุใดก็ตาม ซึ่งเราไม่สามารถไปตามเก็บเพิ่มเติมได้ ทำให้ข้อมูลของตัวอย่างไม่สมบูรณ์ เกิดปัญหาในการวิเคราะห์ ซึ่งด้วยสาเหตุนี้ผู้วิจัยบางคนอาจแก้ปัญหาโดยการตัดค่าสังเกตชุดนั้นทิ้งไป และวิเคราะห์ข้อมูลเท่าที่มีอยู่ วิธีดังกล่าวนี้ไม่เหมาะสมเนื่องจากข้อมูลที่เหลืออยู่อาจมีจำนวนน้อยกว่าที่วางแผนไว้มาก เป็นผลทำให้สูญเสียประสิทธิภาพทางสถิติไปได้มากและอาจทำให้ค่าประมาณที่ได้มีความเอนเอียงหรือมีความคลาดเคลื่อนมาตรฐานสูง โดยเฉพาะอย่างยิ่งเมื่อหน่วยที่ทิ้งไปแตกต่างจากที่เหลืออยู่มาก (ประชุม, 2552) ทำให้การประมาณค่าขาดความน่าเชื่อถือได้ ดังนั้นการประมาณค่าของข้อมูลสูญหายจึงมีความสำคัญมาก

จากการศึกษาของ Sujitta et al. (2006) ได้ศึกษาการประมาณค่าสูญหายในการวิเคราะห์การถดถอยพหุคูณ ใช้วิธีการในการประมาณค่าสูญหายของตัวแปรตาม 3 วิธี คือ วิธีการถดถอย วิธี EM Algorithm และวิธี Pairwise Deleted ขนาดตัวอย่างที่ใช้คือ 20, 100 และ 150 ค่าสูญหายคือ 10%, 30% และ 50% ความแปรปรวนของความคลาดเคลื่อนคือ 25, 50, 75 และ 100 โดยใช้ความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error, MSE) ในการประมาณค่าสูญหาย จากการศึกษาพบว่าวิธีการถดถอยเป็นวิธีที่ดีที่สุด ซึ่งมีค่าความถี่สูงสุดของค่าความคลาดเคลื่อนกำลังสองเฉลี่ยจากทำซ้ำทั้งหมด 500 ครั้ง มีประสิทธิภาพสูงกว่าวิธีอื่นๆ ในหลายสถานการณ์

รัตติกาล (2555) ได้ศึกษาการประมาณค่าข้อมูลสูญหายในการวิเคราะห์การถดถอยพหุคูณ โดยนำเสนอวิธีประมาณค่าสูญหายของตัวแปรอิสระ 4 วิธี คือ วิธีอัตราส่วนควอร์ไทล์ที่ 1 (Ratio Quartile 1)

วิธีอัตราส่วนควอร์ไทล์ที่ 3 (Ratio Quartile 3) วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 (Regression-Ratio Quartile 1) และวิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 (Regression-Ratio Quartile 3) โดยการเปรียบเทียบประสิทธิภาพของ 4 วิธี จะใช้รากของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Square Error, RMSE) และเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error, MAPE) จากผลการศึกษาสรุปได้ว่าวิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 และวิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 เป็นวิธีที่มีประสิทธิภาพสูงกว่าวิธีอื่น

รัตติกาลและพาชิตชนัต (2558) ได้ศึกษาการประมาณค่าสูญหายในการวิเคราะห์การถดถอยพหุคูณเช่นเดียวกับงานวิจัยของรัตติกาล (2555) โดยเพิ่มวิธีการเปรียบเทียบอีก 2 วิธี คือ การประมาณค่าสูญหายด้วยค่าเฉลี่ย (Mean Imputation) และการประมาณค่าสูญหายด้วยการถดถอย (Regression Imputation) โดยการเปรียบเทียบประสิทธิภาพของ 6 วิธี จะใช้เปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เฉลี่ย จากผลการศึกษาพบว่าวิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 และวิธีการประมาณค่าสูญหายด้วยการถดถอยเป็นวิธีที่มีประสิทธิภาพสูงกว่าวิธีอื่นๆ

เรืองลักษณ์และคณะ (2560) ได้ศึกษาการเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าสูญหายสำหรับการวิเคราะห์การถดถอยพหุคูณด้วยตัวแปรตามมีการสูญหายอย่างสุ่ม (Missing at Random) โดยวิธีการประมาณค่าสูญหายแบ่งออกเป็น 2 แบบ คือ วิธีการประมาณค่าสูญหายแบบเดี่ยว 4 วิธี ได้แก่ วิธีการถดถอย (Regression Imputation) วิธีการถดถอยสโตแคสติก (Stochastic Regression Imputation) วิธีเพื่อนบ้านใกล้สุด k ตัว (K-Nearest Neighbor) และวิธี EM Algorithm ส่วนวิธีการประมาณค่าสูญหายแบบร่วม 2 วิธี ได้แก่ วิธีการถดถอยใกล้สุด k ตัว ด้วยค่าถ่วงน้ำหนักที่เท่ากัน (K-Nearest Regression Imputation with Equivalent Weighted) และวิธีการถดถอยสโตแคสติกใกล้สุด k ตัว ด้วยค่าถ่วงน้ำหนักที่เท่ากัน (K-Nearest Stochastic Regression Imputation with Equivalent Weighted) โดยการเปรียบเทียบประสิทธิภาพของทั้ง 6 วิธี จะใช้ความคลาดเคลื่อนกำลังสองเฉลี่ย จากผลการศึกษาสรุปได้ว่าวิธีการถดถอยสโตแคสติกและวิธีการถดถอยสโตแคสติกใกล้สุด k ตัว ด้วยค่าถ่วงน้ำหนักที่เท่ากันมีประสิทธิภาพสูงกว่าวิธีอื่นๆ ในหลายสถานการณ์

Marcelino et al. (2022) ได้ทำการศึกษาวิธีการในการประมาณค่าสูญหายทั้งหมด 6 วิธี คือ วิธีต้นไม้ตัดสินใจ (Decision Tree) วิธีป่าสุ่ม (Random Forest) วิธีอะดาบูส (Adaboost) วิธีเพื่อนบ้านใกล้สุด k ตัว (K-Nearest Neighbor) วิธีซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) และวิธีโครงข่ายประสาทเทียม (Artificial Neural Network) โดยใช้เปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เฉลี่ย ซึ่งวิธีที่ดีที่สุดในการประมาณค่าสูญหายคือ วิธีเพื่อนบ้านใกล้สุด k ตัว ซึ่งมีเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เฉลี่ยน้อยกว่าวิธีอื่นๆ ในปีเดียวกันนั้น Muhammad and Klairung (2022) ได้ศึกษาวิธีการในการประมาณค่าสูญหาย 7 วิธี ได้แก่ Mean Imputation, Hot Deck Imputation, K-Nearest

Neighbors Imputation (KNN), Stochastic Regression Imputation, Namely Hot Deck and KNN Imputation with Equivalent Weight (HKEW), Hot Deck and Stochastic Regression Imputation with Equivalent Weight (HSEW) และ Mean and Stochastic Regression Imputation with Equivalent Weight (MSEW) ซึ่งวิธีที่ดีที่สุดในการประมาณค่าสูญหายคือ วิธีการถดถอยสโตแคสติก (Stochastic Regression Imputation) เมื่อประมาณค่าสูญหายด้วยความคลาดเคลื่อนกำลังสองเฉลี่ยจะมีผลลัพธ์ที่ดีในทุกขนาดตัวอย่าง และทุกเปอร์เซ็นต์ของค่าสูญหาย

จากการศึกษางานวิจัยที่กล่าวมาข้างต้น ผู้วิจัยสนใจที่จะศึกษาการประมาณค่าสูญหายในการวิเคราะห์การถดถอยพหุคูณ โดยวิธีการถดถอยพหุคูณจากงานวิจัยการศึกษาของ Sujitta et al. (2006) และรัตติกาลและพาชิตชนัต (2558) วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 จากการศึกษาของรัตติกาล (2555) ส่วนวิธีการถดถอยสโตแคสติกและวิธีการถดถอยสโตแคสติกใกล้สุด k ตัว ด้วยค่าถ่วงน้ำหนักที่เท่ากันจากการศึกษาของเรืองลักษณ์และคณะ (2560) และ Muhammad and Klairung (2022) เนื่องจากวิธีที่กล่าวมาข้างต้นเป็นวิธีที่ให้ประสิทธิภาพที่ดี โดยคณะผู้วิจัยจะพิจารณาจากความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error, MSE) เพื่อหาประสิทธิภาพที่เหมาะสมที่สุดและนำไปใช้ตัดสินใจในการเลือกใช้ในแต่ละสถานการณ์ที่มีขนาดตัวอย่าง ความแปรปรวนของความคลาดเคลื่อน และร้อยละของจำนวนข้อมูลสูญหายแตกต่างกัน

1.2 วัตถุประสงค์ของการวิจัย

เพื่อศึกษาเปรียบเทียบวิธีการประมาณค่าสูญหายในตัวแบบการถดถอยพหุคูณ ด้วยวิธีการประมาณ 5 วิธี คือ วิธีการถดถอยพหุคูณ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 วิธีการถดถอยสโตแคสติก และวิธีการถดถอยสโตแคสติกใกล้สุด k ตัว ด้วยค่าถ่วงน้ำหนักที่เท่ากัน เมื่อขนาดตัวอย่าง ความแปรปรวนของความคลาดเคลื่อน และร้อยละของจำนวนข้อมูลสูญหายแตกต่างกัน

1.3 ขอบเขตการศึกษาของการวิจัย

กำหนดขอบเขตการศึกษาดังนี้

1.3.1 ตัวแบบที่ใช้คือ ตัวแบบการถดถอยพหุคูณ (Multiple Regression Model) เป็นดังนี้

$$\underline{Y} = X\underline{\beta} + \underline{\varepsilon}$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}_{n\times 1} = \begin{bmatrix} 1 & X_{11} & X_{12} \\ 1 & X_{21} & X_{22} \\ \vdots & \vdots & \vdots \\ 1 & X_{n1} & X_{n2} \end{bmatrix}_{n\times 3} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}_{3\times 1} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}_{n\times 1} \sim N(0,\sigma^2)$$

หรือเขียนในรูปสมการได้ดังนี้

$$Y_1 = \beta_0 + \beta_1 X_{11} + \beta_2 X_{12} + \varepsilon_1$$
$$Y_2 = \beta_0 + \beta_1 X_{21} + \beta_2 X_{22} + \varepsilon_2$$
$$\vdots$$
$$Y_n = \beta_0 + \beta_1 X_{n1} + \beta_2 X_{n2} + \varepsilon_n$$

เมื่อ $\underline{Y}$ คือ เวกเตอร์ของตัวแปรตามซึ่งมีขนาด $n \times 1$

x คือ เมทริกซ์ของตัวแปรอิสระซึ่งมีขนาด $n \times 3$

$\underline{\beta}$ คือ เวกเตอร์ของพารามิเตอร์ของตัวแบบซึ่งมีขนาด 3×1

$\underline{\varepsilon}$ คือ เวกเตอร์ของความคลาดเคลื่อนซึ่งมีขนาด $n \times 1$ ภายใต้ข้อกำหนด $\underline{\varepsilon} \sim N_n(0, \sigma^2 I_n)$

I_n คือ เมทริกซ์เอกลักษณ์ซึ่งมีขนาด $n \times n$

n คือ ขนาดตัวอย่าง

k คือ จำนวนตัวแปรอิสระในตัวแบบ ในที่นี้มี 2 ตัว คือ X_1 และ X_2

1.3.2 ค่าพารามิเตอร์คือ $\beta_0 = 0.5, \beta_1 = 1$ และ $\beta_2 = -0.3$

1.3.3 ความคลาดเคลื่อนมีการแจกแจงปรกติ $(\varepsilon_i \sim N(0, \sigma^2))$ โดยมีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวน σ^2 เท่ากับ 1, 3, 5, 7 และ 9 (รัตติกาล, 2555)

1.3.4 ตัวแปรอิสระ X_2 มีค่าสูญหายร้อยละ 5, 10, 15 และ 20 ซึ่งเป็นการสูญหายอย่างสุ่ม (Missing at Random) เป็นลักษณะของข้อมูลสูญหายที่ค่าสูญหายเกิดขึ้นอย่างสุ่ม แต่จะมีความเกี่ยวข้องกับตัวแปรอิสระบางตัวหรือกลุ่มตัวอย่างย่อยบางลักษณะ โดยการสูญหายของค่าสังเกตมีความสัมพันธ์กับตัวแปรอิสระอื่นๆ ที่ทราบค่า (รัตติกาล, 2555)

1.3.5 กำหนดขนาดตัวอย่างที่ใช้ ตัวอย่างขนาดเล็กคือ 20 และ 40 ตัวอย่างขนาดปานกลางคือ 60 และ 80 ตัวอย่างขนาดใหญ่คือ 100 และ 120 (รัตติกาล, 2555)

1.3.6 ตัวแปรอิสระไม่มีความสัมพันธ์เชิงเส้นพหุ (No Multicollinearity) กล่าวคือ ตัวแปรอิสระ X_1 และ X_2 ไม่มีความสัมพันธ์กัน (จิราวัลย์, 2558; Montgomery et al., 2012)

1.3.7 ประมาณค่าสูญหายทั้งหมด 5 วิธี ได้แก่ วิธีการถดถอยพหุคูณ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 วิธีการถดถอยสโตแคสติก และวิธีการถดถอยสโตแคสติกใกล้สุด k ตัว ด้วยค่าถ่วงน้ำหนักที่เท่ากัน

1.3.8 ใช้โปรแกรม R Studio V.4.2.1 จำลองการทดลองแต่ละสถานการณ์ ในแต่ละการสุ่มตัวอย่างจะกระทำซ้ำ 1,000 รอบ (รัตติกาล, 2555)

1.3.9 เกณฑ์ที่ใช้เปรียบเทียบประสิทธิภาพของการประมาณค่าสูญหายในรูปแบบการถดถอยพหุคูณภายใต้สถานการณ์ต่าง ๆ ที่กำหนดขึ้น พิจารณาจากค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean

Square Error ; MSE) ของตัวประมาณค่าพารามิเตอร์ โดยวิธีที่ให้ค่า MSE ต่ำที่สุดจะเป็นวิธีที่มีประสิทธิภาพดีที่สุด (Sujitta et al., 2006; รัตติกาล, 2555; เรืองลักษณ์และคณะ, 2560; Muhammad and Klairung, 2022)

1.3.10 วิเคราะห์และสรุปผลการวิจัย

1.4 ประโยชน์ที่คาดว่าจะได้รับ

ทำให้ทราบประสิทธิภาพของแต่ละวิธีที่ศึกษา เพื่อนำไปใช้ตัดสินใจในการประมาณค่าสูญหายอย่างสุ่มของตัวแบบการถดถอยพหุคูณที่มีข้อมูลในลักษณะแตกต่างกัน

1.5 นิยามศัพท์

การถดถอยพหุคูณ (Multiple Regression) คือ การศึกษาความสัมพันธ์ระหว่างตัวแปรอิสระหลายตัวกับตัวแปรตาม 1 ตัว เพื่อศึกษาว่ามีตัวแปรอิสระตัวใดบ้างที่ร่วมกันทำนายหรืออธิบายการผันแปรของตัวแปรตามได้ (นงลักษณ์, 2542)

ข้อมูลสูญหาย (Missing Data) คือ ค่าสังเกตที่ต้องการทราบค่า แต่ไม่สามารถทราบค่าได้ ข้อมูลสูญหายสามารถส่งผลกระทบต่อการวิจัยทั้งในส่วนของการวิเคราะห์และการสรุปผลตีความ โดยที่ระดับความรุนแรงของผลกระทบนี้ขึ้นอยู่กับองค์ประกอบจากหลายส่วน แต่ที่สำคัญคือขนาดของข้อมูลสูญหาย ประเภทของข้อมูลสูญหายที่เกิดขึ้นและวิธีจัดการกับข้อมูลสูญหาย (ทัตดา, 2554)

วิธีการแทนค่าการถดถอย (Regression Imputation) คือ วิธีที่ดำเนินการสร้างสมการถดถอยเพื่อทำนายค่าข้อมูลสูญหายจากข้อมูลที่สมบูรณ์ โดยกำหนดให้ตัวแปรที่ต้องการประมาณค่าข้อมูลสูญหายเป็นตัวแปรที่มีข้อมูลไม่สมบูรณ์ (รัตติกาลและพาชิตชนัต, 2558)

วิธีการแทนค่าการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 (Regression-Ratio Quartile1 - RRQ1 Imputation) คือ การนำประโยชน์จากค่าควอร์ไทล์ที่ 1 และสมการถดถอยมาประยุกต์ใช้ในการประมาณค่าข้อมูลสูญหาย (รัตติกาล, 2555)

วิธีการแทนค่าการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 (Regression-Ratio Quartile3 - RRQ3 Imputation) คือ การนำประโยชน์จากค่าควอร์ไทล์ที่ 3 และสมการถดถอยมาประยุกต์ใช้ในการประมาณค่าข้อมูลสูญหาย (รัตติกาล, 2555)

วิธีการแทนค่าการถดถอยสโตแคสติก (Stochastic Regression Imputation) คือ วิธีประมาณค่าข้อมูลสูญหายด้วยวิธีการถดถอย โดยสมการถดถอยจะมีค่าประมาณความคลาดเคลื่อนมาเป็นตัวช่วยในการประมาณค่าข้อมูลสูญหาย โดยได้มาจากการสุ่มด้วยวิธีมอนติคาร์โล (Monte Carlo Method) (ทัตดา, 2554)

วิธีเพื่อนบ้านใกล้สุด k ตัว (K-Nearest Neighbor หรือ KNN) คือ วิธีการประมาณค่าข้อมูลที่มีประสิทธิภาพและได้รับความนิยมในการใช้เป็นอย่างมาก เนื่องจากเป็นวิธีการที่ง่ายและสามารถนำไปประยุกต์ใช้งานได้หลากหลาย เช่น การจำแนกข้อมูล (Classification) และการแทนที่ค่าข้อมูลสูญหาย (Missing Values Imputation) (พัชนา, 2561)

วิธีการแทนค่าการถดถอยสโตแคสติกใกล้สุด k ตัว ด้วยค่าถ่วงน้ำหนักที่เท่ากัน (K-Nearest Stochastic Regression Imputation with Equivalent Weighted) คือ วิธีการประยุกต์ใช้ค่าถ่วงน้ำหนักที่เท่ากัน (Equivalent Weighted) ที่ได้จากวิธีการประมาณค่าข้อมูลสูญหาย 2 วิธี ได้แก่ วิธีเพื่อนบ้านใกล้สุด k ตัว และวิธีการถดถอยสโตแคสติก (เรืองลักษณ์และคณะ, 2560)

ควอร์ไทล์ (Quartiles) คือ การวัดตำแหน่งของข้อมูลที่แบ่งข้อมูลทั้งหมดออกเป็น 4 ส่วนเท่าๆกัน เมื่อเรียงข้อมูลจากน้อยไปหามาก ค่าที่แบ่งข้อมูลออกเป็น 4 ส่วน มี 3 ค่า คือควอร์ไทล์ที่ 1 (Q1) ควอร์ไทล์ที่ 2 (Q2) และควอร์ไทล์ที่ 3 (Q3) ตามลำดับ (Wilks, 1959)

วิธีมอนติคาร์โล (Monte Carlo Method) คือ การใช้การสุ่มตัวอย่างสุ่มซ้ำเพื่อสร้างข้อมูลที่จำลองตามตัวแบบคณิตศาสตร์ โดยมาจากการวิเคราะห์ทางสถิติ (Williamson, 2013)

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในบทนี้ คณะผู้วิจัยได้นำเสนอเนื้อหาที่เน้นถึงทฤษฎีและงานวิจัยที่เกี่ยวข้องโดยมีรายละเอียดของเนื้อหาประกอบด้วยหัวข้อย่อย 9 หัวข้อ ดังนี้

2.1 วิธีกำลังสองน้อยที่สุด (Least Squares Method : LS)

วิธีการหาสัมประสิทธิ์การถดถอยโดยวิธีกำลังสองน้อยที่สุด คือ การหาค่าประมาณของพารามิเตอร์ที่ทำให้ผลบวกกำลังสองของผลต่างระหว่างค่าสังเกตได้และค่าคาดหวังของตัวแปรมีค่าต่ำที่สุด (สายชล, 2554)

จากสมการความสัมพันธ์ระหว่างตัวแปรตาม Y และตัวแปรอิสระ X (Montgomery and Peck, 2001) คือ

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

เมื่อ Y_i คือ ตัวแปรตาม ค่าที่ $i = 1,2,\ldots,n$

X_{ki} คือ ตัวแปรอิสระ ตัวที่ k ค่าที่ i

β_k คือ ค่าพารามิเตอร์ของตัวแบบ

ε_i คือ ค่าความคลาดเคลื่อน ค่าที่ i ซึ่งมี $E(\varepsilon_i) = 0$ และ $V(\varepsilon_i) = \sigma^2$

k คือ จำนวนตัวแปรอิสระในตัวแบบ

หรือตัวแบบการถดถอยเชิงเส้นในรูปเมทริกซ์ เขียนได้ดังนี้

$$\underline{Y} = X\underline{\beta} + \underline{\varepsilon}$$

โดยที่

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1k} \\ 1 & X_{21} & X_{22} & \cdots & X_{2k} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_0 \\ \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

เมื่อ $\underline{Y}$ คือ เวกเตอร์ของตัวแปรตามขนาด $n\times 1$

X คือ เมทริกซ์ของตัวแปรอิสระขนาด $n\times(k+1)$

$\underline{\beta}$ คือ เวกเตอร์ของพารามิเตอร์ขนาด $(k+1)\times 1$

n คือ ขนาดตัวอย่าง

k คือ จำนวนตัวแปรอิสระ

$\underline{\varepsilon}$ คือ เวกเตอร์ของความคลาดเคลื่อนขนาด $n\times 1$

โดยที่ $E(\underline{\varepsilon})=0$ และ $\sigma^2(\underline{\varepsilon})=\sigma^2 I_n$ ดังนั้น

$$\begin{aligned} E(\underline{Y}) &= E(X\underline{\beta}+\underline{\varepsilon}) \\ &= E(X\underline{\beta})+E(\underline{\varepsilon}) \\ &= X\underline{\beta} \end{aligned}$$

โดยทั่วไปเมื่อมีข้อมูลครบถ้วน วิธีกำลังสองน้อยที่สุดในการประมาณค่าสัมประสิทธิ์ของการถดถอย จะทำให้ผลบวกกำลังสองของความคลาดเคลื่อน (Sum of Square of Error - SSE) มีค่าน้อยที่สุด นั่นคือ

$$\begin{aligned} SSE &= \underline{\varepsilon}'\underline{\varepsilon} \\ &= \underline{Y}'\underline{Y}-2\hat{\underline{\beta}}'X'\underline{Y}+\hat{\underline{\beta}}'X'X\hat{\underline{\beta}} \end{aligned}$$

การหาค่ากำลังสองน้อยที่สุดของผลบวกกำลังสองของความคลาดเคลื่อน ทำได้โดยหาอนุพันธ์เทียบกับ $\hat{\beta}$ แล้วกำหนดให้เท่ากับศูนย์ ดังนี้

$$\frac{\partial}{\partial\hat{\underline{\beta}}}(\underline{Y}'\underline{Y}-2\hat{\underline{\beta}}'X'\underline{Y}+\hat{\underline{\beta}}'X'X\hat{\underline{\beta}})=0$$

$$\hat{\underline{\beta}}=(X'X)^{-1}X'\underline{Y}$$

ดังนั้น $\hat{\beta}$ เป็นตัวประมาณแบบกำลังสองน้อยที่สุดของ $\underline{\beta}$

$$\begin{aligned} E(\hat{\underline{\beta}}) &= (X'X)^{-1}XE'(\underline{Y}) \\ &= (X'X)^{-1}X'X\underline{\beta} \\ &= \underline{\beta} \end{aligned}$$

และ $\underline{\hat{\beta}}$ เป็นตัวประมาณที่ไม่เอนเอียงของ $\underline{\beta}$
ดังนั้น สมการถดถอยที่ใช้ทำนาย คือ

$$\underline{\hat{Y}} = X\underline{\hat{\beta}}$$

โดยที่ $E(\underline{\hat{\beta}}) = \underline{\beta},\ V(\underline{\hat{\beta}}) = (X'X)^{-1}\sigma^2$

2.2 ประเภทของข้อมูลสูญหาย

การจำแนกประเภทของข้อมูลสูญหายนั้นเป็นสิ่งจำเป็นอย่างยิ่ง หากเลือกใช้วิธีจัดการกับข้อมูลสูญหายที่ไม่เหมาะสมย่อมส่งผลกระทบ ทำให้เกิดการบิดเบือนต่อผลการวิเคราะห์ได้ ซึ่งโดยทั่วไปมักจำแนกข้อมูลสูญหายออกเป็น 3 ประเภท (Little and Rubin, 1987) คือ

1. Missing completely at random (MCAR) เป็นลักษณะของข้อมูลสูญหายที่เกิดขึ้นอย่างสุ่มจากค่าสังเกตทั้งหมด นั่นคือข้อมูลที่ไม่สูญหายไม่ขึ้นอยู่กับค่าใดค่าหนึ่ง ข้อมูลสูญหายเป็นอิสระกัน
2. Missing at random (MAR) เป็นลักษณะของข้อมูลสูญหายซึ่งไม่ได้เกิดขึ้นอย่างสุ่มจากค่าสังเกตทั้งหมด แต่เกิดขึ้นอย่างสุ่มภายในบางส่วนหรือบางกลุ่มของค่าสังเกต นั่นคือ ค่าของข้อมูลสูญหายขึ้นอยู่กับตัวแปรบางตัวอื่นๆ ในฐานข้อมูลซึ่งไม่ได้เป็นตัวแปรที่เกิดข้อมูลสูญหาย
3. Not missing at random (NMAR) เป็นลักษณะของข้อมูลสูญหายซึ่งไม่ได้เกิดขึ้นอย่างสุ่ม โดยค่าของข้อมูลสูญหายขึ้นอยู่กับค่าของข้อมูลสมบูรณ์ในตัวแปรเดียวกัน รวมถึงตัวแปรอื่นๆ ด้วย หรือในบางกรณีค่าของข้อมูลสูญหายอาจไม่ขึ้นอยู่กับตัวแปรใดๆ ในฐานข้อมูลเลย แต่ขึ้นอยู่กับตัวแปรอื่นที่ไม่ได้ถูกเก็บรวบรวมไว้ในการศึกษาครั้งนั้น

2.3 ตัวประมาณอัตราส่วน (Ratio Estimators)

ตัวประมาณอัตราส่วนเป็นวิธีที่ใช้ประมาณค่า μ_Y โดยอาศัยความสัมพันธ์ระหว่างตัวแปรช่วย X กับตัวแปรตาม Y ภายใต้เงื่อนไขว่าค่าเฉลี่ยประชากร μ_x เป็นพารามิเตอร์ที่ทราบค่า ซึ่งนิยามภายใต้การสุ่มตัวอย่างอย่างง่ายได้ดังนี้ (Al-Omari et al., 2009)

$$\mu_{YSRS} = \mu_X \left(\frac{\bar{Y}_{SRS}}{\bar{X}_{SRS}} \right)$$

เมื่อ $\hat{\mu}_{YSRS}$ คือ ค่าประมาณอัตราส่วน

$\bar{X}_{SRS} = \frac{1}{n}\sum_{i=1}^{n} X_i$ และ $\bar{Y}_{SRS} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ คือค่าเฉลี่ยตัวอย่างของตัวแปรอิสระ X_2 และตัวแปรตาม Y ตามลำดับ

วิธีประมาณค่าอัตราส่วนโดยใช้ควอร์ไทล์ที่ 1 และ 3

Al-Omari et al. ในปี ค.ศ. 2009 นำเสนอตัวประมาณแบบอัตราส่วนเพื่อใช้ในการประมาณค่าเฉลี่ยประชากรของตัวแปรตาม Y โดยนำประโยชน์จากค่าควอร์ไทล์ที่ 1 และ 3 ของตัวแปร X มาประยุกต์ใช้ในตัวประมาณ ตัวประมาณที่ได้นำเสนอนั้นจะเป็นการศึกษาภายใต้การสุ่มตัวอย่างอย่างง่าย ซึ่งมีรายละเอียดดังนี้

$$\hat{\mu}_{YSRS1} = \overline{Y}_{SRS}\left(\frac{\mu_X + q_1}{\overline{X}_{SRS} + q_1}\right) \tag{2.1}$$

$$\hat{\mu}_{YSRS3} = \overline{Y}_{SRS}\left(\frac{\mu_X + q_3}{\overline{X}_{SRS} + q_3}\right) \tag{2.2}$$

เมื่อ $\hat{\mu}_{YSRS1}$ คือ ค่าประมาณค่าเฉลี่ยโดยใช้ควอร์ไทล์ที่ 1

$\hat{\mu}_{YSRS3}$ คือ ค่าประมาณค่าเฉลี่ยโดยใช้ควอร์ไทล์ที่ 3

μ_X คือ ค่าเฉลี่ยประชากรของตัวแปรอิสระ X_2

$\overline{X}_{SRS}$ และ $\overline{Y}_{SRS}$ คือ ค่าเฉลี่ยตัวอย่างของตัวแปรอิสระ X_2 และตัวแปรตาม Y ตามลำดับ

q_1 และ q_3 คือ ค่าควอร์ไทล์ที่ 1 และ 3 ของประชากรของตัวแปรอิสระ X_2 ตามลำดับ

ตัวประมาณค่าข้อมูลสูญหายที่ รัตติกาล (2555) นำเสนอ

ตัวประมาณอัตราส่วนของ Al-Omari et al. (2009) เป็นตัวประมาณค่าเฉลี่ยของประชากรโดยอาศัยควอร์ไทล์ที่ 1 และ 3 ดังแสดงในสมการที่ (2.1) และ (2.2) นอกจากนี้ยังเป็นตัวประมาณที่มีความเอนเอียงต่ำมาก $(Bias \cong 0)$ อย่างไรก็ตาม ตัวประมาณนี้ใช้สำหรับข้อมูลสมบูรณ์และมีเงื่อนไขว่าทราบค่าเฉลี่ยประชากรที่แท้จริงของตัวแปรอิสระ X ในกรณีศึกษานี้เป็นกรณีที่มีข้อมูลสูญหายและโดยทั่วไปในการวิเคราะห์การถดถอยพหุคูณ เราจะไม่ทราบค่าเฉลี่ยประชากรที่แท้จริงของตัวแปรอิสระ X อย่างไรก็ตาม ในการประมาณค่าสูญหาย ผู้วิจัยสนใจใช้การประมาณค่าวิธีอัตราส่วน เนื่องจากเป็นวิธีประมาณที่ใช้ตัวแปรช่วย (X) โดยอาศัยความสัมพันธ์ระหว่างตัวแปร Y และ X ซึ่งมีลักษณะใกล้เคียงกับการวิเคราะห์การถดถอยพหุคูณ

จากสูตรในสมการ (2.1) และ (2.2) รัตติกาล (2555) ได้นำค่า $\overline{X}_{in}$ มาแทนค่า μ_X เนื่องจากในความเป็นจริงการหาค่าเฉลี่ยของประชากรนั้นสามารถทำได้ยาก จึงใช้ค่าเฉลี่ยของตัวอย่าง ($\overline{X}_{in}$) แทนค่าเฉลี่ยของประชากร (μ_X) และใช้ค่าเฉลี่ยของตัวอย่าง $\overline{X}_{ir}$ แทนค่าเฉลี่ยของตัวอย่าง $\overline{X}_{SRS}$ ทำให้ได้ตัวประมาณค่าข้อมูลสูญหาย 2 ตัว ดังสมการ (2.3) และ (2.4) ในที่นี้จะเรียกว่า การแทนค่าสูญหายอัตราส่วนควอร์ไทล์ที่ 1 (Ratio Quartile1 : RQ1) และการแทนค่าสูญหายอัตราส่วนควอร์ไทล์ที่ 3 (Ratio Quartile3 : RQ3) ซึ่งมีสูตรดังนี้

$$Y_{RQ1} = Y_j = \overline{Y}\prod_{i=1}^{k}\left(\frac{\overline{X}_{in} + q_1}{\overline{X}_{ir} + q_1}\right) \quad (2.3)$$

$$Y_{RQ3} = Y_j = \overline{Y}\prod_{i=1}^{k}\left(\frac{\overline{X}_{in} + q_3}{\overline{X}_{ir} + q_3}\right) \quad (2.4)$$

เมื่อ Y_{RQ1} คือ ค่าประมาณอัตราส่วนโดยใช้ควอร์ไทล์ที่ 1

Y_{RQ3} คือ ค่าประมาณอัตราส่วนโดยใช้ควอร์ไทล์ที่ 3

$\overline{X}_{in}$ คือ ค่าเฉลี่ยตัวอย่างที่สมบูรณ์ของตัวแปรอิสระ X_2

$\overline{X}_{ir}$ คือ ค่าเฉลี่ยตัวอย่างที่ตัดค่าสูญหายออกของตัวแปรอิสระ X_2

$\overline{Y}$ คือ ค่าเฉลี่ยตัวแปรตาม Y ที่สมบูรณ์

อย่างไรก็ตาม การใช้ตัวประมาณค่าเพียงค่าเดียว $(\overline{Y})$ แทนค่าสูญหายทุกตัวอาจไม่เหมาะสม เนื่องจากการใช้ค่าคงที่แทนค่าสูญหายทุกตัว ทำให้ความแปรปรวนของตัวประมาณมีค่าต่ำกว่าที่ควรจะเป็น (Underestimate) ดังนั้นรัตติกาล (2555) จึงนำค่าประมาณจากวิธีการถดถอยคือ $\hat{Y}_{reg}$ มาแทนใน $\overline{Y}$ ในสมการ (2.3) และ (2.4) จะได้สมการ (2.5) และ (2.6) ซึ่งเรียกว่า การแทนค่าสูญหายการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 (Regression-Ratio Quartile1 Imputation : RRQ1) และ การแทนค่าสูญหายการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 (Regression-Ratio Quartile3 Imputation : RRQ3) ดังนี้

$$Y_{RRQ1} = Y_j = Y_{reg}\prod_{i=1}^{k}\left(\frac{\overline{X}_{in} + q_1}{\overline{X}_{ir} + q_1}\right) \quad (2.5)$$

$$Y_{RRQ3} = Y_j = Y_{reg}\prod_{i=1}^{k}\left(\frac{\overline{X}_{in} + q_3}{\overline{X}_{ir} + q_3}\right) \quad (2.6)$$

เมื่อ Y_j คือ ค่าประมาณตัวแปรตาม Y

$\overline{X}_{in}$ คือ ค่าเฉลี่ยตัวอย่างที่สมบูรณ์ของตัวแปรอิสระ X_2

$\overline{X}_{ir}$ คือ ค่าเฉลี่ยตัวอย่างที่ตัดค่าสูญหายออกของตัวแปรอิสระ X_2

Y_{reg} คือ ค่าเฉลี่ยตัวแปรตาม Y ที่ประมาณได้จากวิธีการถดถอย

q_1 และ q_3 คือ ค่าควอร์ไทล์ที่ 1 และ 3 ของตัวอย่างของตัวแปรอิสระ X_2 ที่มีขนาด n ตามลำดับ

k คือ จำนวนตัวแปรที่มีค่าข้อมูลสูญหายในที่นี้คือตัวแปรอิสระ X_2 $, i = 1, 2, 3, ..., r$
$j = r+1, r+2, ..., n$

การแทนค่าประมาณ $\overline{Y}$ ด้วยค่าประมาณจากวิธีการถดถอย $(\hat{Y}_{reg})$ เนื่องจากค่าเฉลี่ยเป็นค่ากลางหรือที่เรียกว่า การวัดแนวโน้มเข้าสู่ส่วนกลางของข้อมูล (Measure of Tendency) ซึ่งเป็นการนำค่าของข้อมูลทุกค่ามาคำนวณค่าเฉลี่ย ด้วยสาเหตุนี้ค่าเฉลี่ยที่ได้จากการคำนวณ $(\overline{Y})$ อาจสูงหรือต่ำมากเกินไป เนื่องจากอิทธิพลของค่าสังเกตบางค่าที่สูงหรือต่ำกว่าปกติ (outlier) จึงทำให้เกิดความ

คลาดเคลื่อนของการประมาณค่าสูง ดังนั้นจึงใช้ค่าประมาณจากวิธีการถดถอย ($\hat{Y}_{reg}$) มาแทน เนื่องจากเป็นค่าที่ต้องอาศัยความสัมพันธ์ระหว่างตัวแปรอิสระที่ถูกกำหนดขึ้นกับตัวแปรตาม

2.4 วิธีการประมาณค่าข้อมูลสูญหาย

กำหนดให้ข้อมูลของตัวแปรอิสระ X_1, X_2 และตัวแปรตาม Y โดยมีข้อมูลทั้งหมดชุดละ 20 ค่า ดังแสดงในตารางที่ 2.1

ตารางที่ 2.1 ข้อมูล X_1, X_2 และ Y

ค่าสังเกตที่	X_1	X_2	Y
1	7	560	16.68
2	3	220	11.50
3	3	340	12.03
4	4	80	14.88
5	6	150	13.75
6	7	330	18.11
7	2	110	8.00
8	7	210	17.83
9	30	1460	79.24
10	5	605	21.50
11	16	688	40.33
12	10	215	21.00
13	4	255	13.50
14	6	462	19.75
15	9	448	24.00
16	10	776	29.00
17	6	200	15.35
18	7	132	19.00
19	3	36	9.50
20	17	770	35.10

ที่มา Montgomery and Peck (2001)

จากตารางที่ 2.1 แสดงข้อมูล X_1, X_2 และ Y จำนวนอย่างละ 20 ค่า

การคำนวณด้วยโปรแกรมสำเร็จรูปอาร์ (R Studio)

2.4.1 คำนวณจำนวนข้อมูลสูญหายจากค่าร้อยละของข้อมูลสูญหาย

จำนวนข้อมูลทั้งหมด 20 ค่า สูญหายร้อยละ 5

```
N=20
PMD = 5
NMD = function(N,PMD) { (N)*(PMD)/100 }
NMD(N,PMD)
```

```
[1] 1
```

รูปที่ 2.1 จำนวนข้อมูลสูญหาย

จากรูปที่ 2.1 แสดงค่าจำนวนข้อมูลสูญหายคือ 1 ค่า

2.4.2 การสุ่มตำแหน่งข้อมูลสูญหาย

```
N_MD = NMD(N,PMD)
PM=sample(N,size=N_MD)
PM
```

```
[1] 5
```

รูปที่ 2.2 ตำแหน่งข้อมูลสูญหาย

จากรูปที่ 2.2 แสดงค่าข้อมูลสูญหายอยู่ในตำแหน่งที่ 5

2.5 วิธีการถดถอยพหุคูณ (Multiple Regression Imputation)

2.5.1 การคำนวณด้วยเครื่องคิดเลข

2.5.1.1 คำนวณข้อมูลสูญหายด้วยวิธีค่าเฉลี่ย (Series Mean) มีสูตรในการคำนวณ ดังนี้

$$\hat{X} = \frac{\sum_{i=1}^{k} X_i}{k}$$

$$= \frac{560+220+340+80+330+110+210+1460+605+688+215+255+462+448+776+200+132+36+770}{19}$$

$$= 415.6316$$

ภายหลังจากที่ทำการประมาณค่าข้อมูลสูญหายวิธีการถดถอยพหุคูณด้วยวิธีค่าเฉลี่ย โดยค่าประมาณที่ได้คือ 415.6316 ค่าที่ได้แสดงในตารางที่ 2.2 ดังนี้

ตารางที่ 2.2 ข้อมูลหลังจากแทนที่ค่าสูญหายด้วยวิธีค่าเฉลี่ย (Series Mean)

ค่าสังเกตที่	X_1	X_2	Y
1	7	560	16.68
2	3	220	11.50
3	3	340	12.03
4	4	80	14.88
5	6	**415.6316**	19.75
6	7	330	18.11
7	2	110	8.00
8	7	210	17.83
9	30	1460	79.24
10	5	605	21.50
11	16	688	40.33
12	10	215	21.00
13	4	255	13.50
14	6	462	19.75
15	9	448	24.00
16	10	776	29.00
17	6	200	15.35
18	7	132	19.00
19	3	36	9.50
20	17	770	35.10

จากตารางที่ 2.2 แสดงค่าข้อมูลที่ประมาณค่าสูญหายด้วยวิธีค่าเฉลี่ย (Series Mean) มีค่าเท่ากับ 415.6316 แทนลงไปในตำแหน่งที่ 5 ของตัวแปรอิสระ X_2 จากค่าเดิมเท่ากับ 150

2.5.1.2 หาตัวแบบการถดถอยพหุคูณ ขั้นแรกเขียนเมทริกซ์ X และเวกเตอร์ $\underline{Y}$ ในรูป

$$X = \begin{bmatrix} 1 & 7 & 560 \\ 1 & 3 & 220 \\ \vdots & \vdots & \vdots \\ 1 & 17 & 770 \end{bmatrix}, \quad \underline{Y} = \begin{bmatrix} 16.68 \\ 11.50 \\ \vdots \\ 35.10 \end{bmatrix}$$

เมทริกซ์ $X'X$ คือ

$$X'X = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 7 & 3 & \cdots & 17 \\ 560 & 220 & \cdots & 770 \end{bmatrix} \begin{bmatrix} 1 & 7 & 560 \\ 1 & 3 & 220 \\ \vdots & \vdots & \vdots \\ 1 & 17 & 770 \end{bmatrix}$$

$$= \begin{bmatrix} 20 & 162 & 8,312.63158 \\ 162 & 2,118 & 103,302.789 \\ 8,312.63158 & 103,302.789 & 5,572,012.609 \end{bmatrix}$$

$$(X'X)^{-1} = \begin{bmatrix} 20 & 162 & 8,312.63158 \\ 162 & 2,118 & 103,302.789 \\ 8,312.63158 & 103,302.789 & 5,572,012.609 \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} 0.137136 & -0.005333 & -0.000106 \\ -0.005333 & 0.005138 & -0.0000873 \\ -0.000106 & -0.0000873 & 0.00000196 \end{bmatrix}$$

และเวกเตอร์ $X'\underline{Y}$ คือ

$$X'\underline{Y} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 7 & 3 & \cdots & 17 \\ 560 & 220 & \cdots & 770 \end{bmatrix} \begin{bmatrix} 16.68 \\ 11.50 \\ \vdots \\ 35.10 \end{bmatrix}$$

$$= \begin{bmatrix} 440.05 \\ 5,465.73 \\ 273,196.9 \end{bmatrix}$$

ตัวประมาณกำลังสองน้อยสุดของ $\underline{\beta}$ คือ

$$\underline{\hat{\beta}} = (X'X)^{-1} X'\underline{Y}$$

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} 0.137136 & -0.005333 & -0.000106 \\ -0.005333 & 0.005138 & -0.0000873 \\ -0.000106 & -0.0000873 & 0.00000196 \end{bmatrix} \begin{bmatrix} 440.05 \\ 5,465.73 \\ 273,196.9 \end{bmatrix}$$

$$= \begin{bmatrix} 2.316763 \\ 1.886012 \\ 0.010608 \end{bmatrix}$$

ตัวแบบกำลังสองน้อยสุดคือ

$$\hat{Y}_i = 2.316763 + 1.886012X_1 + 0.010608X_2$$

เช่น

$$\hat{Y}_1 = 2.316763 + 1.886012(7) + 0.010608(560) = 21.459353$$
$$\hat{Y}_2 = 2.316763 + 1.886012(3) + 0.010608(220) = 10.308569$$
$$\hat{Y}_3 = 2.316763 + 1.886012(3) + 0.010608(340) = 11.581535$$
$$\hat{Y}_4 = 2.316763 + 1.886012(4) + 0.010608(80) = 10.709455$$
$$\hat{Y}_5 = 2.316763 + 1.886012(6) + 0.010608(415.6316) = 18.041874$$
$$\vdots$$
$$\hat{Y}_{20} = 2.316763 + 1.886012(17) + 0.010608(770) = 42.547164$$

ตารางที่ 2.3 ข้อมูล X_1, X_2 และ $\widehat{Y}_i$ หลังจากแทนที่ค่าสูญหายของวิธีการถดถอยพหุคูณ

ค่าสังเกตที่	X_1	X_2	$\hat{Y}_i$
1	7	560	21.459353
2	3	220	10.308569
3	3	340	11.581535
4	4	80	10.709455
5	6	**415.6316**	**18.041874**
6	7	330	19.019503
7	2	110	7.2556723
8	7	210	17.746538
9	30	1460	74.384874
10	5	605	18.164691
11	16	688	39.791293
12	10	215	23.457614
13	4	255	12.565863
14	6	462	18.533753
15	9	448	24.043277
16	10	776	29.408727
17	6	200	15.754445
18	7	132	16.91911
19	3	36	8.3566892
20	17	770	42.547164

จากตารางที่ 2.3 แสดงค่าข้อมูล X_1, X_2 ที่หลังจากแทนที่ค่าสูญหายด้วยวิธีค่าเฉลี่ย (Series Mean) และ $\widehat{Y}_i$ ไปหาค่าใหม่

2.5.1.3 คำนวณค่า MSE

ตารางที่ 2.4 ค่า $\boldsymbol{Y_i}$ เดิมและค่าทำนาย $\hat{Y}_i$ ใหม่และส่วนเหลือ $\boldsymbol{e_i}$ จากตัวแบบของวิธีการถดถอยพหุคูณ

$\boldsymbol{Y_i}$	$\hat{Y}_i$	$e_i = Y_i - \hat{Y}_i$	e_i^2
16.68	21.459353	-4.7793532	22.8422174
11.50	10.308569	1.1914306	1.41950698
12.03	11.581535	0.4484653	0.20112112
14.88	10.709455	4.1705447	17.3934431
13.75	**18.041874**	-4.2918744	18.420186
18.11	19.019503	-0.909503	0.82719569
8.00	7.2556723	0.7443277	0.55402377
17.83	17.746538	0.0834624	0.00696596
79.24	74.384874	4.8551264	23.5722527
21.50	18.164691	3.3353091	11.1242869
40.33	39.791293	0.5387074	0.29020567
21.00	23.457614	-2.4576144	6.03986863
13.50	12.565863	0.9341369	0.87261175
19.75	18.533753	1.2162473	1.47925752
24.00	24.043277	-0.0432766	0.00187287
29.00	29.408727	-0.4087274	0.1670581
15.35	15.754445	-0.404445	0.16357577
19.00	16.91911	2.0808898	4.33010249
9.50	8.3566892	1.1433108	1.30715968
35.10	42.547164	-7.4471644	55.4602581
รวม			166.47317

จากตารางที่ 2.4 แสดงค่าข้อมูล $\boldsymbol{Y_i}$ และ $\hat{Y}_i$ ของวิธีการถดถอยพหุคูณ ที่ได้หลังจากการประมาณค่าสูญหายไปหา MSE

คำนวณค่า MSE ได้จากสูตร

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - Y_i)^2$$

$$= \frac{22.8422174 + 1.41950698 + 0.20112112 + ... + 55.4602581}{20}$$

$$= \frac{166.47317}{20}$$

$$= 8.323659$$

จะได้ MSE เท่ากับ 8.323659

2.5.2 การคำนวณด้วยโปรแกรมสำเร็จรูปอาร์

2.5.2.1 กำหนดค่าข้อมูล X_1, X_2 ในรูปแบบเมทริกซ์

```
Xi = c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1)
X1 = c(7,3,3,4,6,7,2,7,30,5,16,10,4,6,9,10,6,7,3,17)
X2 = c(560,220,340,80,150,330,110,210,1460,605,688,215,255,462,448,776,200,132,36,770)
Data_X = c(Xi, X1, X2)
matrix_X_20 <- matrix(Data_X,nrow = 20,ncol = 3)
matrix_X_20
```

```
      [,1] [,2] [,3]
 [1,]    1    7  560
 [2,]    1    3  220
 [3,]    1    3  340
 [4,]    1    4   80
 [5,]    1    6  150
 [6,]    1    7  330
 [7,]    1    2  110
 [8,]    1    7  210
 [9,]    1   30 1460
[10,]    1    5  605
[11,]    1   16  688
[12,]    1   10  215
[13,]    1    4  255
[14,]    1    6  462
[15,]    1    9  448
[16,]    1   10  776
[17,]    1    6  200
[18,]    1    7  132
[19,]    1    3   36
[20,]    1   17  770
```

รูปที่ 2.3 ตัวอย่างข้อมูล X_1, X_2 ในรูปแบบเมทริกซ์ของวิธีการถดถอยพหุคูณ

จากรูปที่ 2.3 แสดงค่าข้อมูล X_1, X_2 ในรูปแบบเมทริกซ์

2.5.2.2 กำหนดค่าข้อมูล Y ในรูปแบบเมทริกซ์

```
Data_Y =
c(16.68,11.50,12.03,14.88,13.75,18.11,8.00,17.83,79.24,21.50,40.33,21.00,13.50,19.75,24.00,
29.00,15.35,19.00,9.50,35.10)
matrix_Y_20 <- matrix(Data_Y,nrow = 20,ncol = 1)
matrix_Y_20
```

```
       [,1]
 [1,] 16.68
 [2,] 11.50
 [3,] 12.03
 [4,] 14.88
 [5,] 13.75
 [6,] 18.11
 [7,]  8.00
 [8,] 17.83
 [9,] 79.24
[10,] 21.50
[11,] 40.33
[12,] 21.00
[13,] 13.50
[14,] 19.75
[15,] 24.00
[16,] 29.00
[17,] 15.35
[18,] 19.00
[19,]  9.50
[20,] 35.10
```

รูปที่ 2.4 ตัวอย่างข้อมูล Y ในรูปแบบเมทริกซ์ของวิธีการถดถอยพหุคูณ

จากรูปที่ 2.4 แสดงค่าข้อมูล Y ในรูปแบบเมทริกซ์

2.5.2.3 ประมาณค่าข้อมูลสูญหายด้วยวิธีการถดถอยพหุคูณโดยวิธีค่าเฉลี่ย

```
New_Estimate = (sum(X2)-(matrix_X_20[5,3]))/(N-1)
New_Estimate
```

```
[1] 415.6316
```

รูปที่ 2.5 ค่าประมาณข้อมูลสูญหายด้วยวิธีการถดถอยพหุคูณโดยวิธีค่าเฉลี่ย

จากรูปที่ 2.5 แสดงค่าประมาณข้อมูลสูญหายด้วยวิธีการถดถอยพหุคูณโดยวิธีค่าเฉลี่ยมีค่าเท่ากับ 415.63156

2.5.2.4 แทนค่าข้อมูลสูญหายที่ประมาณได้ลงในข้อมูล

X2_New=replace(X2,list= PM_20,values = New_Estimate)

X2_New

```
 [1]  560.0000  220.0000  340.0000   80.0000  415.6316  330.0000  110.0000  210.0000 1460.0000  605.0000  688.0000  215.0000  255.0000
[14]  462.0000  448.0000  776.0000  200.0000  132.0000   36.0000  770.0000
```

รูปที่ 2.6 ข้อมูลที่หลังจากการแทนที่ค่าข้อมูลสูญหายของวิธีการถดถอยพหุคูณ

จากรูปที่ 2.6 แสดงค่าประมาณข้อมูลสูญหายที่ได้ไปแทนในตำแหน่งที่ 5 ของตัวแปรอิสระ X_2 จากจากค่าเดิมเท่ากับ 150 เปลี่ยนเป็น 415.63156

2.5.2.5 กำหนดค่าข้อมูลหลังจากแทนค่าข้อมูลสูญหาย X_1, X_2 ในรูปแบบเมทริกซ์

Data_X_New = c(Xi, X1, X2_New)

matrix_X_20_New <- matrix(Data_X_New,nrow = 20,ncol = 3)

matrix_X_20_New

```
      [,1] [,2]      [,3]
 [1,]    1    7  560.0000
 [2,]    1    3  220.0000
 [3,]    1    3  340.0000
 [4,]    1    4   80.0000
 [5,]    1    6  415.6316
 [6,]    1    7  330.0000
 [7,]    1    2  110.0000
 [8,]    1    7  210.0000
 [9,]    1   30 1460.0000
[10,]    1    5  605.0000
[11,]    1   16  688.0000
[12,]    1   10  215.0000
[13,]    1    4  255.0000
[14,]    1    6  462.0000
[15,]    1    9  448.0000
[16,]    1   10  776.0000
[17,]    1    6  200.0000
[18,]    1    7  132.0000
[19,]    1    3   36.0000
[20,]    1   17  770.0000
```

รูปที่ 2.7 ข้อมูล X_1, X_2 หลังจากแทนค่าข้อมูลสูญหายในรูปแบบเมทริกซ์ของวิธีการถดถอยพหุคูณ

จากรูปที่ 2.7 แสดงค่าข้อมูล X_1, X_2 หลังจากแทนค่าข้อมูลสูญหายในรูปแบบเมทริกซ์จำนวน 20 ค่า

2.5.2.6 หาค่า X' ของข้อมูล X_1, X_2 ในรูปแบบเมทริกซ์

Transpose_X_New = t(matrix_X_20_New)

Transpose_X_New

```
     [,1] [,2] [,3] [,4]     [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14] [,15] [,16] [,17] [,18] [,19] [,20]
[1,]    1    1    1    1   1.0000    1    1    1    1     1     1     1     1     1     1     1     1     1     1     1
[2,]    7    3    3    4   6.0000    7    2    7   30     5    16    10     4     6     9    10     6     7     3    17
[3,]  560  220  340   80 415.6316  330  110  210 1460   605   688   215   255   462   448   776   200   132    36   770
```

รูปที่ 2.8 ค่า X' ของข้อมูล X_1, X_2 ในรูปแบบเมทริกซ์ของวิธีการถดถอยพหุคูณ

จากรูปที่ 2.8 แสดงค่าข้อมูล X_1, X_2 หลังจากแทนค่าข้อมูลสูญหายไปหาค่า X'

2.5.2.7 หาค่า $X'X$ ของข้อมูล X_1, X_2 ในรูปแบบเมทริกซ์

TransposeX_X_New = Transpose_X_New%*%matrix_X_20_New

TransposeX_X_New

```
         [,1]     [,2]        [,3]
[1,]   20.000    162.0    8312.632
[2,]  162.000   2118.0  103302.789
[3,] 8312.632 103302.8 5572012.609
```

รูปที่ 2.9 ค่า $X'X$ ของข้อมูล X_1, X_2 ในรูปแบบเมทริกซ์ของวิธีการถดถอยพหุคูณ

จากรูปที่ 2.9 แสดงค่าข้อมูล X_1, X_2 หลังจากแทนค่าข้อมูลสูญหายไปหาค่า $X'X$

2.5.2.8 หาค่า $(X'X)^{-1}$ ของข้อมูล X_1, X_2 ในรูปแบบเมทริกซ์

Inverse_TransposeX_X_New = solve(TransposeX_X_New)

Inverse_TransposeX_X_New

```
              [,1]          [,2]          [,3]
[1,]  0.1371360831 -5.333237e-03 -1.057111e-04
[2,] -0.0053332368  5.138149e-03 -8.730272e-05
[3,] -0.0001057111 -8.730272e-05  1.955730e-06
```

รูปที่ 2.10 ค่า $(X'X)^{-1}$ ของข้อมูล X_1, X_2 ในรูปแบบเมทริกซ์ของวิธีการถดถอยพหุคูณ

จากรูปที่ 2.10 แสดงค่าข้อมูล X_1, X_2 หลังจากแทนค่าข้อมูลสูญหายไปหาค่า $(X'X)^{-1}$

2.5.2.9 หาค่า $X'Y$ ของข้อมูล X_1, X_2 และ $\underline{Y}$ ในรูปแบบเมทริกซ์

```
TransposeX_Y_New = Transpose_X_New%*%matrix_Y_20
TransposeX_Y_New
```

```
          [,1]
[1,]    440.05
[2,]   5465.73
[3,] 273196.87
```

รูปที่ 2.11 ค่า $X'\underline{Y}$ ของข้อมูล X_1, X_2 และ Y ในรูปแบบเมทริกซ์ของวิธีการถดถอยพหุคูณ

จากรูปที่ 2.11 แสดงค่าข้อมูล X_1, X_2 และ Y หลังจากแทนค่าข้อมูลสูญหายไปหาค่า $X'\underline{Y}$

2.5.2.10 หาค่า $\underline{\hat{\beta}}$ ของข้อมูล X_1, X_2 และ $\underline{Y}$

```
B_New = Inverse_TransposeX_X_New%*%TransposeX_Y_New
B_New
```

```
           [,1]
[1,] 2.31676310
[2,] 1.88601222
[3,] 0.01060804
```

รูปที่ 2.12 ค่า $\underline{\hat{\beta}}$ ของข้อมูล X_1, X_2 และ $\underline{Y}$ ของวิธีการถดถอยพหุคูณ

จากรูปที่ 2.12 แสดงค่าข้อมูล X_1, X_2 หลังจากแทนค่าข้อมูลสูญหายและ $\underline{Y}$ ไปหาค่า $\underline{\hat{\beta}}$

2.5.2.11 หาตัวแบบของข้อมูล X_1, X_2 และ $\underline{Y}$

```
cat("Yhat=",B_New[1,1],"+",B_New[2,1],"X1+",B_New[3,1],"X2")
```

```
Yhat= 2.316763 + 1.886012 X1+ 0.01060804 X2
```

รูปที่ 2.13 ตัวแบบของข้อมูล X_1, X_2 และ $\underline{Y}$ ของวิธีการถดถอยพหุคูณ

จากรูปที่ 2.13 แสดงค่าข้อมูล X_1, X_2 หลังจากแทนค่าข้อมูลสูญหายและ $\underline{Y}$ ไปหาตัวแบบ

2.5.2.12 หาค่า $\underline{\hat{Y}}$ จากตัวแบบ

```
Yhat_20_New = function(B_New,X1,X2_New) {
B_New[1,1]+(B_New[2,1]*X1)+(B_New[3,1]*X2_New)
}
Yhat_20_New(B_New,X1,X2_New)
```

```
 [1] 21.459353 10.308569 11.581535 10.709455 18.041874 19.019503  7.255672 17.746538 74.384873 18.164691 39.791293 23.457615 12.565863
[14] 18.533753 24.043277 29.408727 15.754445 16.919110  8.356689 42.547164
```

รูปที่ 2.14 ข้อมูล $\underline{\hat{Y}}$ จากตัวแบบของวิธีการถดถอยพหุคูณ

จากรูปที่ 2.14 แสดงค่าข้อมูล $\underline{\hat{Y}}$ ที่ได้ จากตัวแบบของข้อมูล X_1, X_2 หลังจากแทนค่าข้อมูลสูญหายและ $\underline{Y}$

2.5.2.13 หาค่า $\underline{\hat{Y}}$ จากตัวแบบในรูปแบบเวกเตอร์

```
matrix_Yhat_20_New <- matrix(Yhat_20_New(B_New,X1,X2_New),nrow = 20,ncol = 1)
matrix_Yhat_20_New
```

```
           [,1]
 [1,] 21.459353
 [2,] 10.308569
 [3,] 11.581535
 [4,] 10.709455
 [5,] 18.041874
 [6,] 19.019503
 [7,]  7.255672
 [8,] 17.746538
 [9,] 74.384873
[10,] 18.164691
[11,] 39.791293
[12,] 23.457615
[13,] 12.565863
[14,] 18.533753
[15,] 24.043277
[16,] 29.408727
[17,] 15.754445
[18,] 16.919110
[19,]  8.356689
[20,] 42.547164
```

รูปที่ 2.15 ข้อมูล $\underline{\hat{Y}}$ จากตัวแบบในรูปแบบเวกเตอร์ของวิธีการถดถอยพหุคูณ

จากรูปที่ 2.15 แสดงค่าข้อมูล $\underline{\hat{Y}}$ ที่ได้ จากตัวแบบของข้อมูล X_1, X_2 หลังจากแทนค่าข้อมูลสูญหายและ $\underline{\hat{Y}}$ ในรูปแบบเวกเตอร์

2.5.2.14 หาค่า MSE

```
MSE_20 = function(N,matrix_Y_20,matrix_Yhat_20_New) {
(1/N)*(sum((matrix_Y_20-matrix_Yhat_20_New)^2))
}
MSE_20(N,matrix_Y_20,matrix_Yhat_20_New)
```

```
[1] 8.323659
```

รูปที่ 2.16 ค่า MSE ของวิธีการถดถอยพหุคูณ

จากรูปที่ 2.16 แสดงว่าค่า MSE ที่ได้จากข้อมูล $\underline{Y}$ และ $\hat{\underline{Y}}$ เท่ากับ 8.323659

2.6 วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 (Regression-Ratio Quartile1 Imputation : RRQ1)

2.6.1 การคำนวณด้วยเครื่องคิดเลข

ตารางที่ 2.5 ข้อมูล X_1, X_2 และ Y

ค่าสังเกตที่	X_1	X_2
1	7	560
2	3	220
3	3	340
4	4	80
5	6	150
6	7	330
7	2	110
8	7	210
9	30	1460
10	5	605
11	16	688
12	10	215
13	4	255
14	6	462
15	9	448
16	10	776
17	6	200
18	7	132
19	3	36
20	17	770

จากตารางที่ 2.5 แสดงค่าข้อมูล X_1, X_2 และ Y

2.6.1.1 คำนวณหาค่าควอร์ไทล์ที่ 1 โดยใช้สูตร ดังนี้

คำนวณหาตำแหน่ง q_1

$$q_k = \frac{k(n+1)}{4}$$

$$q_1 = \frac{1}{4}(20+1)$$

$$= 5.25$$

$$\approx 5$$

ตารางที่ 2.6 การเรียงค่า X_2 จากน้อยไปหามาก

ค่าสังเกตที่	X_2	X_2 ในการเรียงลำดับจากน้อยไปมาก
1	560	36
2	220	80
3	340	110
4	80	132
5	150	**150**
6	330	200
7	110	210
8	210	215
9	1460	220
10	605	255
11	688	330
12	215	340
13	255	448
14	462	462
15	448	560
16	776	605
17	200	688
18	132	770
19	36	776
20	770	1460

จากตารางที่ 2.6 แสดงค่าข้อมูล X_2 โดยเรียงค่า X_2 จากน้อยไปหามาก ดังนั้น q_1 เป็นข้อมูลที่อยู่ที่ตำแหน่งที่ 5 ของตัวแปรอิสระ X_2 มีค่าเท่ากับ 150

2.5.1.2 คำนวณหาค่าเฉลี่ยตัวอย่างที่สมบูรณ์และค่าเฉลี่ยตัวอย่างที่ตัดค่าสูญหายออกไปของตัวแปรอิสระ X_2

$$\overline{X}_{in} = \frac{560+220+340+80+150+330+110+210+1460+605+688+215+255+462+448+776+200+132+36+770}{20}$$

$$= 402.35$$

เมื่อ $\overline{X}_{in}$ คือ ค่าเฉลี่ยตัวอย่างที่สมบูรณ์ของตัวแปรอิสระ X_2

$$\overline{X}_{ir} = \frac{560+220+340+80+330+110+210+1460+605+688+215+255+462+448+776+200+132+36+770}{19}$$

$$= 415.6316$$

เมื่อ $\overline{X}_{ir}$ คือ ค่าเฉลี่ยตัวอย่างที่ตัดค่าสูญหายออกไปของตัวแปรอิสระ X_2

2.6.1.2 คำนวณหาค่าทำนาย $\hat{Y}_{RRQ1}$

ตารางที่ 2.7 ข้อมูล X_1, X_2 และ $\hat{Y}_i$ จากตารางที่ 2.3 หลังจากแทนที่ค่าสูญหายของวิธีการถดถอยพหุคูณ

X_1	X_2	$\hat{Y}_i$
7	560	21.459353
3	220	10.308569
3	340	11.581535
4	80	10.709455
6	**415.6316**	**18.041874**
7	330	19.019503
2	110	7.2556723
7	210	17.746538
30	1460	74.384874
5	605	18.164691
16	688	39.791293
10	215	23.457614
4	255	12.565863
6	462	18.533753
9	448	24.043277
10	776	29.408727
6	200	15.754445
7	132	16.91911
3	36	8.3566892
17	770	42.547164

จากตารางที่ 2.7 แสดงว่านำข้อมูลที่ประมาณค่าสูญหายด้วยวิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 มีค่าเท่ากับ 17.6182307 แทนลงไปในตำแหน่งที่ 5 ของตัวแปร $\hat{Y}$ ที่ได้มากจากวิธีการถดถอยพหุคูณจากค่าเดิมเท่ากับ 18.041874

$$Y_{RRQ1} = Y_j = Y_{reg}\left(\frac{\overline{X}_{in} + q_1}{\overline{X}_{ir} + q_1}\right)$$

$$= 18.041874\left(\frac{402.35+150}{415.6316+150}\right)$$

$$= 17.6182307$$

ตารางที่ 2.8 ค่า Y_i เดิมและค่าทำนายได้ $\hat{Y}_i$ ใหม่และส่วนเหลือ e_i จากตัวแบบของวิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1

Y_i	$\hat{Y}_i$	$e_i = Y_i - \hat{Y}_i$	e_i^2
16.68	21.459353	-4.7793532	22.8422174
11.50	10.308569	1.1914306	1.41950698
12.03	11.581535	0.4484653	0.20112112
14.88	10.709455	4.1705447	17.3934431
13.75	**17.6182307**	-3.8682307	14.96320875
18.11	19.019503	-0.909503	0.82719569
8.00	7.2556723	0.7443277	0.55402377
17.83	17.746538	0.0834624	0.00696596
79.24	74.384874	4.8551264	23.5722527
21.50	18.164691	3.3353091	11.1242869
40.33	39.791293	0.5387074	0.29020567
21.00	23.457614	-2.4576144	6.03986863
13.50	12.565863	0.9341369	0.87261175
19.75	18.533753	1.2162473	1.47925752
24.00	24.043277	-0.0432766	0.00187287
29.00	29.408727	-0.4087274	0.1670581
15.35	15.754445	-0.404445	0.16357577
19.00	16.91911	2.0808898	4.33010249
9.50	8.3566892	1.1433108	1.30715968
35.10	42.547164	-7.4471644	55.4602581
รวม			163.016193

จากตารางที่ 2.8 แสดงว่านำข้อมูล Y_i และ $\hat{Y}_i$ ของวิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 ที่ได้หลังจากการประมาณค่าสูญหายไปหา MSE

2.6.1.3 คำนวณค่า MSE ได้จากสูตร

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - Y_i)^2$$

$$= \frac{22.8422174 + 1.41950698 + 0.20112112 + \ldots + 55.4602581}{20}$$

$$= \frac{163.016193}{20}$$

$$= 8.150811$$

จะได้ MSE เท่ากับ 8.150811

2.6.2 การคำนวณด้วยโปรแกรมสำเร็จรูปอาร์

2.6.2.1 หาตำแหน่งค่าควอร์ไทล์ที่ 1 ของตัวแปรอิสระ X_2

```
R1=1
RRQ1=function(N,R1) {
(R1*(N+1))/4
}
RRQ1(N,R1)
```

```
[1] 5.25
```

รูปที่ 2.17 ตำแหน่งค่าควอร์ไทล์ที่ 1 ของตัวแปรอิสระ X_2 ของวิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1

จากรูปที่ 2.17 แสดงค่าตำแหน่งค่าควอร์ไทล์ที่ 1 ของตัวแปรอิสระ X_2 มีค่าเท่ากับ 5.25 หรือตำแหน่งที่ 5

2.6.2.2 เรียงลำดับข้อมูลตัวแปรอิสระ X_2

```
P_RRQ1=RRQ1(N,R1)
X2_Sort=sort(X2)
X2_Sort
```

```
[1]   36   80  110  132  150  200  210  215  220  255  330  340  448  462  560  605  688  770  776 1460
```

รูปที่ 2.18 การเรียงลำดับข้อมูลตัวแปรอิสระ X_2 ของวิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1

จากรูปที่ 2.18 แสดงค่าเรียงลำดับข้อมูลตัวแปรอิสระ X_2

2.6.2.3 หาค่าควอร์ไทล์ที่ 1 ของตัวแปรอิสระ X_2

```
matrix_X2_Sort <- matrix(X2_Sort,nrow = N,ncol = 1)
X2_RRQ1 = matrix_X2_Sort[P_RRQ1,1]
X2_RRQ1
```

```
[1] 150
```

รูปที่ 2.19 ค่าควอร์ไทล์ที่ 1 ของตัวแปรอิสระ X_2 ของวิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1

จากรูปที่ 2.19 แสดงค่าควอร์ไทล์ที่ 1 ของตัวแปรอิสระ X_2 ตำแหน่งที่ 5 มีค่าเท่ากับ 150

2.6.2.4 ประมาณค่าสูญหายของตัวแปร $\hat{Y}$ ด้วยวิธีสมการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1

```
Data_Y_REG_New=Y_REG_New(B_REG_New,X1,Data_X2_REG_New)
matrix_Y_REG_New <- matrix(Data_Y_REG_New,nrow = N,ncol = 1)
New_Estimate_RRQ1 = (matrix_Y_REG_New[PM,1])*((mean(X2)+X2_RRQ1)/(((sum(X2)-
(matrix_X_REG[PM,3]))/(N-N_MD))+X2_RRQ1))
New_Estimate_RRQ1
```

```
[1] 17.61823
```

รูปที่ 2.20 ค่าประมาณของตัวแปร $\hat{Y}$ ด้วยวิธีสมการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1

จากรูปที่ 2.20 แสดงค่าประมาณของตัวแปร $\hat{Y}$ ด้วยวิธีสมการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 มีค่าเท่ากับ 17.61823

2.6.2.5 ข้อมูลตัวแปร $\hat{Y}$ หลังจากการประมาณค่าสูญหายด้วยวิธีสมการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1

```
Data_Y_RRQ1_New=replace(Data_Y_REG_New,list= PM,values = New_Estimate_RRQ1)
Data_Y_RRQ1_New
```

```
[1] 21.459353 10.308569 11.581535 10.709455 17.618234 19.019503  7.255672 17.746538 74.384873 18.164691 39.791293 23.457615 12.565863
[14] 18.533753 24.043277 29.408727 15.754445 16.919110  8.356689 42.547164
```

รูปที่ 2.21 ค่าของตัวแปร $\hat{Y}$ หลังจากการประมาณค่าสูญหายด้วยวิธีสมการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1

จากรูปที่ 2.21 แสดงค่าประมาณข้อมูลสูญหายที่ได้ไปแทนในตำแหน่งที่ 5 ของตัวแปร $\hat{Y}$ จากค่าเดิมเท่ากับ 18.041874 เปลี่ยนเป็น 17.61823

2.6.2.6 ข้อมูลตัวแปร $\hat{Y}$ หลังจากการประมาณค่าสูญหายด้วยวิธีสมการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 ในรูปแบบเมทริกซ์

```
matrix_Y_RRQ1_New <- matrix(Data_Y_RRQ1_New,nrow = N,ncol = 1)
matrix_Y_RRQ1_New
```

```
          [,1]
 [1,] 21.459353
 [2,] 10.308569
 [3,] 11.581535
 [4,] 10.709455
 [5,] 17.618234
 [6,] 19.019503
 [7,]  7.255672
 [8,] 17.746538
 [9,] 74.384873
[10,] 18.164691
[11,] 39.791293
[12,] 23.457615
[13,] 12.565863
[14,] 18.533753
[15,] 24.043277
[16,] 29.408727
[17,] 15.754445
[18,] 16.919110
[19,]  8.356689
[20,] 42.547164
```

รูปที่ 2.22 ข้อมูล $\hat{Y}$ หลังจากแทนที่ค่าสูญหายในรูปแบบเมทริกซ์ของวิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1

จากรูปที่ 2.22 แสดงค่าข้อมูล $\hat{Y}$ หลังจากแทนค่าข้อมูลสูญหายในรูปแบบเมทริกซ์

2.6.2.7 หาค่า MSE

```
MSE_RRQ1 = function(N,matrix_Y_REG,matrix_Y_RRQ1_New) {
(1/N)*(sum((matrix_Y_REG-matrix_Y_RRQ1_New)^2))}
MSE_RRQ1(N,matrix_Y_REG,matrix_Y_RRQ1_New)
```

```
[1] 8.150811
```

รูปที่ 2.23 ค่า MSE ของวิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1

จากรูปที่ 2.23 แสดงค่า MSE ที่ได้จากข้อมูล Y และ $\hat{Y}$ เท่ากับ 8.150811

2.7 วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 (Regression-Ratio Quartile3 Imputation : RRQ3)

2.7.1 การคำนวณด้วยเครื่องคิดเลข

ตารางที่ 2.9 ข้อมูล X_1, X_2 และ Y

ค่าสังเกตที่	X_1	X_2	Y
1	7	560	16.68
2	3	220	11.50
3	3	340	12.03
4	4	80	14.88
5	6	150	13.75
6	7	330	18.11
7	2	110	8.00
8	7	210	17.83
9	30	1460	79.24
10	5	605	21.50
11	16	688	40.33
12	10	215	21.00
13	4	255	13.50
14	6	462	19.75
15	9	448	24.00
16	10	776	29.00
17	6	200	15.35
18	7	132	19.00
19	3	36	9.50
20	17	770	35.10

จากตารางที่ 2.8 แสดงวข้อมูล X_1, X_2 และ Y

2.7.1.1 คำนวณหาค่าควอร์ไทล์ที่ 3 โดยใช้สูตร ดังนี้

คำนวณหาตำแหน่ง q_3

$$q_k = \frac{k(n+1)}{4}$$

$$q_1 = \frac{3}{4}(20+1)$$

$$=15.75 \quad \approx 15$$

ตารางที่ 2.10 เรียงค่า X_2 จากน้อยไปหามาก

ค่าสังเกตที่	X_2	X_2 ในการเรียงลำดับจากน้อยไปมาก
1	560	36
2	220	80
3	340	110
4	80	132
5	150	150
6	330	200
7	110	210
8	210	215
9	1460	220
10	605	255
11	688	330
12	215	340
13	255	448
14	462	462
15	448	**560**
16	776	605
17	200	688
18	132	770
19	36	776
20	770	1460

จากตารางที่ 2.10 แสดงค่าข้อมูล X_2 โดยเรียงค่า X_2 จากน้อยไปหามาก

ดังนั้น q_3 เป็นข้อมูลที่อยู่ที่ตำแหน่งที่ 15 ของตัวแปรอิสระ X_2 มีค่าเท่ากับ 560

2.7.1.2 คำนวณหาค่าเฉลี่ยตัวอย่างที่สมบูรณ์และค่าเฉลี่ยตัวอย่างที่ตัดค่าสูญหายออกไปของตัวแปรอิสระ X_2

$$\overline{X}_{in} = \frac{560+220+340+80+150+330+110+210+1460+605+688+215+255+462+448+776+200+132+36+770}{20}$$

$$= 402.35$$

เมื่อ $\overline{X}_{in}$ คือ ค่าเฉลี่ยตัวอย่างที่สมบูรณ์ของตัวแปรอิสระ X_2

$$\bar{X}_{ir} = \frac{560+220+340+80+330+110+210+1460+605+688+215+255+462+448+776+200+132+36+770}{19}$$

$$= 415.6316$$

เมื่อ $\overline{X}_{ir}$ คือ ค่าเฉลี่ยตัวอย่างที่ตัดค่าสูญหายออกไปของตัวแปรอิสระ X_2

2.7.1.3 คำนวณหาค่า $\hat{Y}_{RRQ3}$

ตารางที่ 2.11 ข้อมูล X_1, X_2 และ $\hat{Y}_i$ จากตารางที่ 2.3 หลังจากแทนที่ค่าสูญหายของวิธีการถดถอยพหุคูณ

X_1	X_2	$\hat{Y}_i$
7	560	21.459353
3	220	10.308569
3	340	11.581535
4	80	10.709455
6	**415.6316**	**18.041874**
7	330	19.019503
2	110	7.2556723
7	210	17.746538
30	1460	74.384874
5	605	18.164691
16	688	39.791293
10	215	23.457614
4	255	12.565863
6	462	18.533753
9	448	24.043277
10	776	29.408727
6	200	15.754445
7	132	16.91911
3	36	8.3566892
17	770	42.547164

จากตารางที่ 2.11 แสดงว่านำข้อมูลที่ประมาณค่าสูญหายด้วยวิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 มีค่าเท่ากับ 17.7962639 แทนลงไปในตำแหน่งที่ 5 ของตัวแปร $\hat{Y}$ ที่ได้มากจากวิธีการถดถอยพหุคูณ จากค่าเดิมเท่ากับ 18.041874

$$Y_{RRQ3} = Y_j = Y_{reg}\left(\frac{\overline{X}_{in} + q_3}{\overline{X}_{ir} + q_3}\right)$$

$$= 18.041874\left(\frac{402.35 + 560}{415.6316 + 560}\right)$$

$$= 17.7962639$$

ตารางที่ 2.12 ค่า Y_i เก่าและค่าทำนายได้ $\widehat{Y}_i$ ใหม่และส่วนเหลือ e_i จากตัวแบบนี้ของวิธีการถดถอย-อัตราส่วน ควอร์ไทล์ที่ 3

Y_i	$\hat{Y}_i$	$e_i = Y_i - \hat{Y}_i$	e_i^2
16.68	21.459353	-4.7793532	22.8422174
11.50	10.308569	1.1914306	1.41950698
12.03	11.581535	0.4484653	0.20112112
14.88	10.709455	4.1705447	17.3934431
13.75	**17.7962639**	-4.0462639	16.37225155
18.11	19.019503	-0.909503	0.82719569
8.00	7.2556723	0.7443277	0.55402377
17.83	17.746538	0.0834624	0.00696596
79.24	74.384874	4.8551264	23.5722527
21.50	18.164691	3.3353091	11.1242869
40.33	39.791293	0.5387074	0.29020567
21.00	23.457614	-2.4576144	6.03986863
13.50	12.565863	0.9341369	0.87261175
19.75	18.533753	1.2162473	1.47925752
24.00	24.043277	-0.0432766	0.00187287
29.00	29.408727	-0.4087274	0.1670581
15.35	15.754445	-0.404445	0.16357577
19.00	16.91911	2.0808898	4.33010249
9.50	8.3566892	1.1433108	1.30715968
35.10	42.547164	-7.4471644	55.4602581
รวม			164.425236

จากตารางที่ 2.8 แสดงว่านำข้อมูล Y_i และ $\hat{Y}_i$ ของวิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 ที่ได้หลังจากการประมาณค่าสูญหายไปหา MSE

2.7.1.4 คำนวณหาค่า MSE ได้จากสูตร

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - Y_i)^2$$

$$= \frac{22.8422174 + 1.41950698 + 0.20112112 + ... + 55.4602581}{20}$$

$$= \frac{164.425236}{20}$$

$$= 8.221262$$

จะได้ MSE เท่ากับ 8.221262

2.7.2 การคำนวณด้วยโปรแกรมสำเร็จรูปอาร์

2.7.2.1 หาตำแหน่งค่าควอร์ไทล์ที่ 3 ของตัวแปรอิสระ X_2

```
R3=3
RRQ3=function(N,R3) {
(R3*(N+1))/4
}
RRQ3(N,R3)
```

```
[1] 15.75
```

รูปที่ 2.24 ตำแหน่งค่าควอร์ไทล์ที่ 3 ของตัวแปรอิสระ X_2 ของวิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3

จากรูปที่ 2.24 แสดงค่าตำแหน่งค่าควอร์ไทล์ที่ 1 ของตัวแปรอิสระ X_2 มีค่าเท่ากับ 15.75 หรือตำแหน่งที่ 15

2.7.2.2 เรียงลำดับข้อมูลตัวแปรอิสระ X_2

```
P_RRQ3=RRQ3(N,R3)
X2_Sort=sort(X2)
X2_Sor
```

```
[1]   36   80  110  132  150  200  210  215  220  255  330  340  448  462  560  605  688  770  776 1460
```

รูปที่ 2.25 การเรียงลำดับข้อมูลตัวแปรอิสระ X_2 ของวิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3

จากรูปที่ 2.25 แสดงค่าเรียงลำดับข้อมูลตัวแปรอิสระ X_2

2.7.2.3 ชุดคำสั่งการหาค่าควอร์ไทล์ที่ 3 ของตัวแปรอิสระ X_2

```
matrix_X2_Sort <- matrix(X2_Sort,nrow = N,ncol = 1)
X2_RRQ3 = matrix_X2_Sort[P_RRQ3,1]
X2_RRQ3
```

```
[1] 560
```

รูปที่ 2.26 ค่าควอร์ไทล์ที่ 3 ของตัวแปรอิสระ X_2 ของวิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3

จากรูปที่ 2.26 แสดงค่าควอร์ไทล์ที่ 3 ของตัวแปรอิสระ X_2 ตำแหน่งที่ 5 มีค่าเท่ากับ 560

2.7.2.4 ประมาณค่าสูญหายของตัวแปร $\hat{Y}$ ด้วยวิธีสมการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3

```
Data_Y_REG_New=Y_REG_New(B_REG_New,X1,Data_X2_REG_New)
matrix_Y_REG_New <- matrix(Data_Y_REG_New,nrow = N,ncol = 1)
New_Estimate_RRQ3 = (matrix_Y_REG_New[PM,1])*((mean(X2)+X2_RRQ3)/(((sum(X2)-
(matrix_X_REG[PM,3]))/(N-N_MD))+X2_RRQ3))
New_Estimate_RRQ3
```

```
[1] 17.79626
```

รูปที่ 2.27 ค่าประมาณของตัวแปร $\hat{Y}$ ด้วยวิธีสมการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3

จากรูปที่ 2.27 แสดงค่าประมาณของตัวแปร $\hat{Y}$ ด้วยวิธีสมการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 มีค่าเท่ากับ 17.79626

2.7.2.5 ข้อมูลตัวแปร $\hat{Y}$ หลังจากการประมาณค่าสูญหายด้วยวิธีสมการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3

```
Data_Y_RRQ3_New=replace(Data_Y_REG_New,list= PM,values = New_Estimate_RRQ3)
Data_Y_RRQ3_New
```

```
 [1] 21.459353 10.308569 11.581535 10.709455 17.796265 1[illegible].019503  7.255672 17.746538 74.384873 18.164691 39.791293 23.457615 12.565863
[14] 18.533753 24.043277 29.408727 15.754445 [illegible] [illegible].356689 42.547164
```

รูปที่ 2.28 ค่าของตัวแปร $\hat{Y}$ หลังจากการประมาณค่าสูญหายด้วยวิธีสมการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3

จากรูปที่ 2.28 แสดงค่าประมาณข้อมูลสูญหายที่ได้ไปแทนในตำแหน่งที่ 5 ของตัวแปร $\hat{Y}$ จาก จากค่าเดิมเท่ากับ 18.041874 เปลี่ยนเป็น 17.79626

2.7.2.6 ข้อมูลตัวแปรYหลังจากการประมาณค่าสูญหายด้วยวิธีสมการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 ในรูปแบบเมทริกซ์

```
matrix_Y_RRQ3_New <- matrix(Data_Y_RRQ3_New,nrow = N,ncol = 1)
matrix_Y_RRQ3_New
```

```
           [,1]
 [1,] 21.459353
 [2,] 10.308569
 [3,] 11.581535
 [4,] 10.709455
 [5,] 17.796265
 [6,] 19.019503
 [7,]  7.255672
 [8,] 17.746538
 [9,] 74.384873
[10,] 18.164691
[11,] 39.791293
[12,] 23.457615
[13,] 12.565863
[14,] 18.533753
[15,] 24.043277
[16,] 29.408727
[17,] 15.754445
[18,] 16.919110
[19,]  8.356689
[20,] 42.547164
```

รูปที่ 2.29 ข้อมูล $\hat{Y}$ หลังจากแทนที่ค่าสูญหายในรูปแบบเมทริกซ์ ของวิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3

จากรูปที่ 2.29 แสดงค่าข้อมูล $\hat{Y}$ หลังจากแทนค่าข้อมูลสูญหายในรูปแบบเมทริกซ์

2.7.2.7 หาค่า MSE

```
MSE_RRQ3 = function(N,matrix_Y_REG,matrix_Y_RRQ3_New) {
(1/N)*(sum((matrix_Y_REG-matrix_Y_RRQ3_New)^2)) }
MSE_RRQ3(N,matrix_Y_REG,matrix_Y_RRQ3_New)
```

```
[1] 8.221262
```

รูปที่ 2.30 ค่า MSE ของวิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3

จากรูปที่ 2.23 แสดงค่า MSE ที่ได้จากข้อมูล Y และ $\hat{Y}$ เท่ากับ 8.221262

2.8 วิธีการถดถอยสโตแคสติก (Stochastic Regression Imputation)

2.8.1 การคำนวณด้วยเครื่องคิดเลข

2.8.1.1 จากตารางที่ได้จากการประมาณค่าสูญหายวิธีการถดถอยพหุคูณ

ตารางที่ 2.13 ข้อมูล X_1, X_2 และ $\hat{Y}_i$ จากตาราง 2.3 หลังจากแทนที่ค่าสูญหายของวิธีการถดถอยพหุคูณ

ค่าสังเกตที่	X_1	X_2	$\hat{Y}_i$
1	7	560	21.459353
2	3	220	10.308569
3	3	340	11.581535
4	4	80	10.709455
5	6	**415.6316**	**18.041874**
6	7	330	19.019503
7	2	110	7.2556723
8	7	210	17.746538
9	30	1460	74.384874
10	5	605	18.164691
11	16	688	39.791293
12	10	215	23.457614
13	4	255	12.565863
14	6	462	18.533753
15	9	448	24.043277
16	10	776	29.408727
17	6	200	15.754445
18	7	132	16.91911
19	3	36	8.3566892
20	17	770	42.547164

จากตารางที่ 2.13 แสดงว่าข้อมูล X_1, X_2 และ $\hat{Y}_i$ หลังจากแทนที่ค่าสูญหายของวิธีการถดถอยพหุคูณ

2.8.1.2 ทำการสุ่มค่าประมาณของค่าคลาดเคลื่อน Z_i โดยมีการแจกแจงแบบปรกติ ค่าเฉลี่ยเท่ากับ 0 ค่าความแปรปรวนเท่ากับ 1 ที่ได้มาจากวิธีสุ่มมอนติคาร์โล

ตารางที่ 2.14 ข้อมูล Z_i ที่ได้มาจากวิธีสุ่มมอนติคาร์โล

ค่าสังเกตที่	Z_i
1	-0.086874214
2	0.248555013
3	-0.110815039
4	0.281550561
5	-0.262335522
6	-0.184077965
7	-0.215726884
8	-0.344050868
9	-0.023414820
10	0.005479278
11	-0.348057271
12	0.375499961
13	-0.251669407
14	0.384892370
15	0.394134042
16	-0.001775015
17	-0.087754925
18	-0.044802408
19	0.364800404
20	-0.257476502

จากตารางที่ 2.14 แสดงค่าคลาดเคลื่อน Z_i โดยมีการแจกแจงแบบปรกติ ค่าเฉลี่ยเท่ากับ 0 ค่าความแปรปรวนเท่ากับ 1 ที่ได้มาจากวิธีสุ่มมอนติคาร์โล

2.8.1.3 นำ $\hat{Y}_i$ หลังจากแทนที่ค่าสูญหายด้วยวิธีการถดถอยพหุคูณมาบวกกับเทอม Z_i ที่ได้มาจากวิธีสุ่มมอนติคาร์โล

ตารางที่ 2.15 ข้อมูล $\hat{Y}_{REG}$, Z_i และ $\hat{Y}_{SRI} = \hat{Y}_{REG} + Z_i$ หลังจากแทนที่ค่าสูญหายของวิธีการถดถอยสโตแคสติก

$\hat{Y}_{REG}$	Z_i	$\hat{Y}_{SRI} = \hat{Y}_{REG} + Z_i$
21.459353	-0.086874214	21.372479
10.308569	0.248555013	10.557124
11.581535	-0.110815039	11.470720
10.709455	0.281550561	10.991006
18.041874	-0.262335522	17.779539
19.019503	-0.184077965	18.835425
7.2556723	-0.215726884	7.039945
17.746538	-0.344050868	17.402487
74.384874	-0.023414820	74.361459
18.164691	0.005479278	18.170170
39.791293	-0.348057271	39.443235
23.457614	0.375499961	23.833115
12.565863	-0.251669407	12.314194
18.533753	0.384892370	18.918645
24.043277	0.394134042	24.437411
29.408727	-0.001775015	29.406952
15.754445	-0.087754925	15.666690
16.91911	-0.044802408	16.874308
8.3566892	0.364800404	8.721490
42.547164	-0.257476502	42.289688

จากตารางที่ 2.15 แสดงค่า $\hat{Y}$ ของวิธีการถดถอยสโตแคสติกโดยนำค่า $\hat{Y}_{REG} + Z_i$

ตารางที่ 2.16 ค่า Y_i เดิมและค่าทำนายได้ $\hat{Y}_i$ ใหม่และส่วนเหลือ e_i จากตัวแบบของวิธีการถดถอยสโตแคสติก

Y_i	$\hat{Y}_{SRI} = \hat{Y}_{REG} + Z_i$	$e_i = Y_i - \hat{Y}_{SRI}$	e_i^2
16.68	21.372479	-4.692479	22.0193592
11.50	10.557124	0.942876	0.88901515
12.03	11.470720	0.55928	0.31279412
14.88	10.991006	3.888994	15.1242743
13.75	17.779539	-4.029539	16.2371846
18.11	18.835425	-0.725425	0.52624143
8.00	7.039945	0.960055	0.9217056
17.83	17.402487	0.427513	0.18276737
79.24	74.361459	4.878541	23.8001623
21.50	18.170170	3.32983	11.0877678
40.33	39.443235	0.886765	0.78635217
21.00	23.833115	-2.833115	8.0265406
13.50	12.314194	1.185806	1.40613587
19.75	18.918645	0.831355	0.69115114
24.00	24.437411	-0.437411	0.19132838
29.00	29.406952	-0.406952	0.16560993
15.35	15.666690	-0.31669	0.10029256
19.00	16.874308	2.125692	4.51856648
9.50	8.721490	0.77851	0.60607782
35.10	42.289688	-7.189688	51.6916135
รวม			159.28494

จากตารางที่ 2.16 แสดงว่านำข้อมูล Y_i และ $\hat{Y}_i$ ของวิธีการถดถอยสโตแคสติกที่ได้หลังจากการประมาณค่าสูญหายไปหา MSE

2.8.1.4 คำนวณค่า MSE ได้จากสูตร

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - Y_i)^2$$

$$= \frac{22.0193592 + 0.88901515 + 0.31279412 + ... + 51.6916135}{20}$$

$$= \frac{159.28494}{20}$$

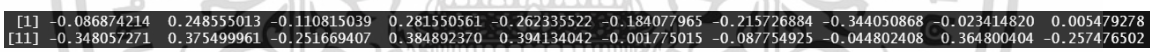
$$= 7.96424702$$

จะได้ MSE เท่ากับ 7.96424702

2.8.2 การคำนวณด้วยโปรแกรมสำเร็จรูปอาร์

2.8.2.1 สุ่มค่าประมาณของค่าคลาดเคลื่อน Z_i โดยมีการแจกแจงแบบปรกติ ค่าเฉลี่ยเท่ากับ 0 ค่าความแปรปรวนเท่ากับ 1 ที่ได้มาจากวิธีสุ่มมอนติคาร์โล

```
Zi = replicate(n=N, rnorm(n=N, mean = 0, sd = Data_e))
Data_Zi=apply(Zi,MARGIN = 2, FUN = mean)
Data_Zi
```



```
 [1] -0.086874214  0.248555013 -0.110815039  0.281550561 -0.262335522 -0.184077965 -0.215726884 -0.344050868 -0.023414820  0.005479278
[11] -0.348057271  0.375499961 -0.251669407  0.384892370  0.394134042 -0.001775015 -0.087754925 -0.044802408  0.364800404 -0.257476502
```

รูปที่ 2.31 ค่าที่ได้จากการสุ่มมอนติคาร์โลของวิธีการถดถอยสโตแคสติก

จากรูปที่ 2.31 แสดงค่าคลาดเคลื่อน Z_i โดยมีการแจกแจงแบบปรกติ ค่าเฉลี่ยเท่ากับ 0 ค่าความแปรปรวนเท่ากับ 1 ที่ได้มาจากวิธีสุ่มมอนติคาร์โล

2.8.2.2 การสุ่มมอนติคาร์โลโดยมีค่าเฉลี่ยเท่ากับ 0 ความแปรปรวนเท่ากับ 1 ในรูปแบบเมทริกซ์

```
matrix_Zi_SRI_New <- matrix(Data_Zi,nrow = N,ncol = 1)
matrix_Zi_SRI_New
```

```
          [,1]
 [1,] -0.086874214
 [2,]  0.248555013
 [3,] -0.110815039
 [4,]  0.281550561
 [5,] -0.262335522
 [6,] -0.184077965
 [7,] -0.215726884
 [8,] -0.344050868
 [9,] -0.023414820
[10,]  0.005479278
[11,] -0.348057271
[12,]  0.375499961
[13,] -0.251669407
[14,]  0.384892370
[15,]  0.394134042
[16,] -0.001775015
[17,] -0.087754925
[18,] -0.044802408
[19,]  0.364800404
[20,] -0.257476502
```

รูปที่ 2.32 ค่าที่ได้จากการสุ่มมอนติคาร์โลในรูปแบบเมทริกซ์ของวิธีการถดถอยสโตแคสติก

จากรูปที่ 2.32 แสดงค่าคลาดเคลื่อน Z_i โดยมีการแจกแจงแบบปรกติ ค่าเฉลี่ยเท่ากับ 0 ค่าความแปรปรวนเท่ากับ 1 ที่ได้มาจากวิธีสุ่มมอนติคาร์โลในรูปแบบเมทริกซ์

2.8.2.3 หาค่า $\hat{Y}$ หลังการประมาณค่าสูญหายจากตัวแบบในรูปแบบเมทริกซ์

Data_Y_SRI_New = Data_Y_REG_New+matrix_Zi_SRI_New

Data_Y_SRI_New

```
          [,1]
 [1,] 21.372479
 [2,] 10.557124
 [3,] 11.470720
 [4,] 10.991006
 [5,] 17.779539
 [6,] 18.835425
 [7,]  7.039945
 [8,] 17.402487
 [9,] 74.361459
[10,] 18.170170
[11,] 39.443235
[12,] 23.833115
[13,] 12.314194
[14,] 18.918645
[15,] 24.437411
[16,] 29.406952
[17,] 15.666690
[18,] 16.874308
[19,]  8.721490
[20,] 42.289688
```

รูปที่ 2.33 ข้อมูล $\hat{Y}$ หลังประมาณค่าสูญหายจากตัวแบบในรูปแบบเมทริกซ์ของวิธีการถดถอยสโตแคสติก

จากรูปที่ 2.33 แสดงค่า $\hat{Y}$ ของวิธีการถดถอยสโตแคสติกโดยนำค่า $\hat{Y}_{REG} + Z_i$ ในรูปแบบเมทริกซ์

2.8.2.4 หาค่า MSE

```
MSE_SRI = function(N,matrix_Y_REG,Data_Y_SRI_New) {
(1/N)*(sum((matrix_Y_REG-Data_Y_SRI_New)^2))
}
MSE_SRI(N,matrix_Y_REG,Data_Y_SRI_New)
```

รูปที่ 2.34 ค่า MSE ของวิธีการถดถอยสโตแคสติก

จากรูปที่ 2.23 แสดงค่า MSE ที่ได้จากข้อมูล Y และ $\hat{Y}$ เท่ากับ 7.964247

2.9 วิธีการถดถอยสโตแคสติกใกล้สุด k ตัว ด้วยค่าถ่วงน้ำหนักที่เท่ากัน (Stochastic Regression and K-Nearest Stochastic Regression with Equivalent Weighted)

2.9.1 การคำนวณด้วยเครื่องคิดเลข

วิธีนี้เป็นวิธีที่นำค่า $\hat{Y}_i$ หลังจากแทนที่ค่าสูญหายของวิธีการถดถอยสโตแคสติกและค่า $\hat{Y}_i$ หลังจากแทนที่ค่าสูญหายของวิธีเพื่อนบ้านใกล้สุด k ตัวมาทำการถ่วงน้ำหนักที่เท่ากัน จึงจะได้ค่า $\hat{Y}_i$ ของวิธีการถดถอยสโตแคสติกใกล้สุด k ตัว ด้วยค่าถ่วงน้ำหนักที่เท่ากัน

2.9.1.1 ค่า $\hat{Y}_i$ หลังจากแทนที่ค่าสูญหายของวิธีการถดถอยสโตแคสติก

ตารางที่ 2.17 ค่า $\hat{Y}_i$ หลังจากแทนที่ค่าสูญหายของวิธีการถดถอยสโตแคสติก

ค่าสังเกตที่	$\hat{Y}_i$
1	21.372479
2	10.557124
3	11.470720
4	10.991006
5	17.779539
6	18.835425
7	7.039945
8	17.402487
9	74.361459
10	18.170170
11	39.443235
12	23.833115
13	12.314194
14	18.918645
15	24.437411
16	29.406952
17	15.666690
18	16.874308
19	8.721490
20	42.289688

จากตารางที่ 2.17 แสดงค่า $\hat{Y}_i$ หลังจากแทนที่ค่าสูญหายของวิธีการถดถอยสโตแคสติก

2.9.1.2 ค่า $\hat{Y}_i$ หลังจากแทนที่ค่าสูญหายของวิธีเพื่อนบ้านใกล้สุด k ตัว

การประมาณค่าข้อมูลสูญหายด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว มีขั้นตอนการดำเนินการดังนี้

1) กำหนดค่า k

จากสูตร

$$\begin{aligned} \boldsymbol{k} &= \sqrt{\boldsymbol{n}} \\ &= \sqrt{20} \\ &= 4.4721 \\ &\approx 5 \end{aligned}$$

2) คำนวณหาระยะห่างระหว่างข้อมูลตัวอย่างที่สนใจกับข้อมูลอื่นๆ ทุกตัวด้วยวิธีระยะห่างยุคลิด (Euclidean Distance)

$$dist(X_i, X_j) = \sqrt{\sum_{k=1}^{n}(X_{i,k} - X_{j,k})^2}$$

	ลำดับ
$dist(R_5, R_1) = \sqrt{(6-7)^2 + (13.75-16.68)^2} = 3.095949$	4
$dist(R_5, R_2) = \sqrt{(6-3)^2 + (13.75-11.5)^2} = 3.75$	6
$dist(R_5, R_3) = \sqrt{(6-3)^2 + (13.75-12.03)^2} = 3.458092$	5
$dist(R_5, R_4) = \sqrt{(6-4)^2 + (13.75-14.88)^2} = 2.29715$	3
$dist(R_5, R_6) = \sqrt{(6-7)^2 + (13.75-18.11)^2} = 4.473209$	8
$dist(R_5, R_7) = \sqrt{(6-2)^2 + (13.75-8)^2} = 7.004463$	12
$dist(R_5, R_8) = \sqrt{(6-7)^2 + (13.75-17.83)^2} = 4.200762$	7
$dist(R_5, R_9) = \sqrt{(6-30)^2 + (13.75-79.24)^2} = 69.749123$	19
$dist(R_5, R_{10}) = \sqrt{(6-5)^2 + (13.75-21.5)^2} = 7.81425$	13
$dist(R_5, R_{11}) = \sqrt{(6-16)^2 + (13.75-40.33)^2} = 28.39888$	18
$dist(R_5, R_{12}) = \sqrt{(6-10)^2 + (13.75-21)^2} = 8.280248$	14
$dist(R_5, R_{13}) = \sqrt{(6-4)^2 + (13.75-13.5)^2} = 2.015564$	2
$dist(R_5, R_{14}) = \sqrt{(6-6)^2 + (13.75-19.75)^2} = 6$	11
$dist(R_5, R_{15}) = \sqrt{(6-9)^2 + (13.75-24)^2} = 10.680005$	15
$dist(R_5, R_{16}) = \sqrt{(6-10)^2 + (13.75-29)^2} = 15.765865$	16

$$dist(R_5, R_{17}) = \sqrt{(6-6)^2 + (13.75-15.35)^2} = 1.6$$ 1

$$dist(R_5, R_{18}) = \sqrt{(6-7)^2 + (13.75-19)^2} = 5.34439$$ 10

$$dist(R_5, R_{19}) = \sqrt{(6-3)^2 + (13.75-9.5)^2} = 5.202163$$ 9

$$dist(R_5, R_{20}) = \sqrt{(6-17)^2 + (13.75-35.1)^2} = 24.017129$$ 17

3) เรียงลำดับระยะห่างระหว่างจุด โดยพิจารณาจากข้อมูลที่ใกล้ที่สุดจำนวน k ตัว ตามที่กำหนดในการประมาณค่าข้อมูลสูญหาย ทำการประมาณค่าสูญหายโดยการหาค่าเฉลี่ยจากข้อมูลที่ใกล้ที่สุด k ตัว จากตัวแปรเดียวกันกับตำแหน่งที่มีการสูญหาย มีสูตรการคำนวณดังนี้

ค่าสูญหาย คือ $X_i = \frac{\sum_{i=1}^{k} X_i}{k}$

ตารางที่ 2.18 ค่าระยะห่างระหว่างจุดของ X_2

	X_2	ระยะห่าง	ลำดับ
R_{17}	200	1.6	1
R_{13}	255	2.015564437	2
R_4	80	2.297150409	3
R_1	560	3.095948966	4
R_3	340	3.458091959	5

จากตารางที่ 2.18 แสดงค่าระยะห่างระหว่างจุดของ X_2

$$X_5 = \frac{200+255+80+560+340}{5}$$

$$= 287$$

ตารางที่ 2.19 ข้อมูล X_1, X_2 ที่ได้จากการประมาณค่าข้อมูลสูญหายด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว และ Y_i

ค่าสังเกตที่	X_1	X_2	Y_i
1	7	560	16.68
2	3	220	11.50
3	3	340	12.03
4	4	80	14.88
5	6	**287**	13.75
6	7	330	18.11
7	2	110	8.00
8	7	210	17.83
9	30	1460	79.24
10	5	605	21.50
11	16	688	40.33
12	10	215	21.00
13	4	255	13.50
14	6	462	19.75
15	9	448	24.00
16	10	776	29.00
17	6	200	15.35
18	7	132	19.00
19	3	36	9.50
20	17	770	35.10

จากตารางที่ 2.19 แสดงค่าข้อมูล X_1, X_2 ที่ได้จากการประมาณค่าข้อมูลสูญหายด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว จากค่าเดิมเท่ากับ 150 เปลี่ยนเป็น 287 และ Y_i

4) ในการหาตัวแบบการถดถอยพหุคูณ ขั้นแรกเขียนเมทริกซ์ X และเวกเตอร์ $\underline{Y}$ ในรูปแบบ

$$X = \begin{bmatrix} 1 & 7 & 560 \\ 1 & 3 & 220 \\ \vdots & \vdots & \vdots \\ 1 & 17 & 770 \end{bmatrix}, \quad \underline{Y} = \begin{bmatrix} 16.68 \\ 11.50 \\ \vdots \\ 35.10 \end{bmatrix}$$

เมทริกซ์ $X'X$ คือ

$$X'X = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 7 & 3 & \cdots & 17 \\ 560 & 220 & \cdots & 770 \end{bmatrix} \begin{bmatrix} 1 & 7 & 560 \\ 1 & 3 & 220 \\ \vdots & \vdots & \vdots \\ 1 & 17 & 770 \end{bmatrix}$$

$$= \begin{bmatrix} 20 & 162 & 8{,}184 \\ 162 & 2{,}118 & 102{,}531 \\ 8{,}184 & 102{,}531 & 5{,}481{,}632 \end{bmatrix}$$

$$(X'X)^{-1} = \begin{bmatrix} 20 & 162 & 8{,}312.63158 \\ 162 & 2{,}118 & 103{,}302.789 \\ 8{,}312.63158 & 103{,}302.789 & 5{,}572{,}012.609 \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} 0.1354324 & -0.006035671 & -0.0000893 \\ -0.006035671 & 0.005264 & -0.0000894 \\ -0.0000893 & -0.0000894 & 0.00000199 \end{bmatrix}$$

และเวกเตอร์ $X'Y$ คือ

$$X'\underline{Y} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 7 & 3 & \cdots & 17 \\ 560 & 220 & \cdots & 770 \end{bmatrix} \begin{bmatrix} 16.68 \\ 11.50 \\ \vdots \\ 35.10 \end{bmatrix}$$

$$= \begin{bmatrix} 440.05 \\ 5{,}465.73 \\ 271{,}428.19 \end{bmatrix}$$

ตัวประมาณกำลังสองน้อยสุดของ $\underline{\beta}$ คือ

$$\hat{\underline{\beta}} = (X'X)^{-1}X'\underline{Y}$$

$$\begin{bmatrix}\hat{\beta}_0\\ \hat{\beta}_1\\ \hat{\beta}_2\end{bmatrix} = \begin{bmatrix}0.1354324 & -0.006035671 & -0.0000893\\ -0.006035671 & 0.005264 & -0.0000894\\ -0.0000893 & -0.0000894 & 0.00000199\end{bmatrix}\begin{bmatrix}440.05\\ 5,465.73\\ 271,428.19\end{bmatrix}$$

$$= \begin{bmatrix}2.367874\\ 1.836464\\ 0.011631\end{bmatrix}$$

ตัวแบบกำลังสองน้อยสุดโดยวิธีเพื่อนบ้านใกล้สุด k ตัว คือ

$$\hat{Y}_i^* = 2.367874 + 1.836464X_1 + 0.01163006X_2$$

เช่น

$$\hat{Y}_1^* = 2.367874 + 1.836464(7) + 0.01163006(560) = 21.55439$$
$$\hat{Y}_2^* = 2.367874 + 1.836464(3) + 0.01163006(220) = 10.49657$$
$$\hat{Y}_3^* = 2.367874 + 1.836464(3) + 0.01163006(340) = 11.65120$$
$$\hat{Y}_4^* = 2.367874 + 1.836464(4) + 0.01163006(80) = 10.81759$$
$$\hat{Y}_5^* = 2.367874 + 1.836464(6) + 0.01163006(287) = 17.2521$$
$$\vdots$$
$$\hat{Y}_{20}^* = 2.367874 + 1.836464(17) + 0.01163006(770) = 42.41653$$

ตารางที่ 2.20 ข้อมูล X_1, X_2 และ $\hat{Y}_i$ หลังจากแทนที่ค่าสูญหายของวิธีเพื่อนบ้านใกล้สุด k ตัว

ค่าสังเกตที่	X_1	X_2	$\hat{Y}_i$
1	7	560	21.55439
2	3	220	10.49657
3	3	340	11.65120
4	4	80	10.81759
5	6	**287**	17.2521
6	7	330	18.94833
7	2	110	7.18006
8	7	210	17.53402
9	30	1460	74.40201
10	5	605	18.37846
11	16	688	39.59822
12	10	215	23.53311
13	4	255	12.49687
14	6	462	18.83933
15	9	448	24.27200
16	10	776	29.58243
17	6	200	15.68974
18	7	132	16.81634
19	3	36	8.50873
20	17	770	42.41653

จากตารางที่ 2.20 แสดงค่าข้อมูล X_1, X_2 ที่ได้จากการประมาณค่าข้อมูลสูญหายด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว จากค่าเดิมเท่ากับ 150 เปลี่ยนเป็น 287 และ $\hat{Y}_i$ หลังจากหาตัวแบบใหม่

ทำการถ่วงน้ำหนักด้วยวิธีการให้น้ำหนักที่เท่ากันให้กับค่าประมาณที่ได้จากวิธีการประมาณค่าสูญหายโดยวิธีเพื่อนบ้านใกล้สุด k ตัว และวิธีการถดถอยสโตแคสติก ดังนี้

$$\overline{Y}_i^* = W_M(\hat{Y}_i^* + Y_i) \quad ;i = r+1,\ldots,n$$

โดยที่ $W_M = \dfrac{1}{M}$

เมื่อ $\overline{Y}_i^*$ คือ ค่าประมาณที่ได้จากการถ่วงน้ำหนักของวิธีการประมาณค่าสูญหายโดยวิธีเพื่อนบ้านใกล้สุด k ตัวและวิธีการถดถอยสโตแคสติก

$\hat{Y}_i^*$ คือ ค่าประมาณที่ได้จากวิธีการประมาณค่าสูญหายโดยวิธีการถดถอยสโตแคสติก

$\hat{Y}_i$ คือ ค่าประมาณที่ได้จากวิธีการประมาณค่าสูญหายโดยวิธีเพื่อนบ้านใกล้สุด k ตัว

W_M คือ ค่าถ่วงน้ำหนักที่เท่ากันของวิธีการประมาณค่าสูญหาย

M คือ จำนวนวิธีการประมาณค่าสูญหายด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว และวิธีการถดถอยสโตแคสติก ดังนั้นจะได้ $M = 2$

r คือ ลำดับที่ของวิธีการประมาณค่าสูญหาย

ตารางที่ 2.21 ค่าคลาดเคลื่อนกำลังสองเฉลี่ยของวิธีการถดถอยสโตแคสติกใกล้สุด k ตัว ด้วยค่าถ่วงน้ำหนักที่เท่ากัน

ค่าสังเกตที่	Y_i	$\hat{Y}_i^*$	Y_i	$\overline{Y}_i^*$	e_i	e_i^2
1	16.68	21.37248	21.55439	21.4634345	-4.7834345	22.8812456
2	11.50	10.55712	10.49657	10.526847	0.973153	0.94702676
3	12.03	11.47072	11.65120	11.56096	0.46904	0.21999852
4	14.88	10.99101	10.81759	10.904298	3.975702	15.8062064
5	13.75	17.77954	17.2521	17.5158195	-3.7658195	14.1813965
6	18.11	18.83543	18.94833	18.8918775	-0.7818775	0.61133243
7	8.00	7.03995	7.18006	7.1100025	0.8899975	0.79209555
8	17.83	17.40249	17.53402	17.4682535	0.3617465	0.13086053
9	79.24	74.36146	74.40201	74.3817345	4.8582655	23.6027437
10	21.50	18.17017	18.37846	18.274315	3.225685	10.4050437
11	40.33	39.44324	39.59822	39.5207275	0.8092725	0.65492198
12	21.00	23.83312	23.53311	23.6831125	-2.6831125	7.19909269
13	13.50	12.31410	12.49687	12.405532	1.094468	1.1978602
14	19.75	18.91865	18.83933	18.8789875	0.8710125	0.75866278
15	24.00	24.43741	24.27200	24.3547055	-0.3547055	0.12581599
16	29.00	29.40695	29.58243	29.49469	-0.49469	0.2447182
17	15.35	15.66669	15.68974	15.678215	-0.328215	0.10772509
18	19.00	16.87431	16.81634	16.845324	2.154676	4.64262866
19	9.50	8.721490	8.50873	8.61511	0.88489	0.78303031
20	35.10	42.28969	42.41653	42.353109	-7.253109	52.6075902
รวม						157.899996

จากตารางที่ 2.21 แสดงว่านำข้อมูล Y_i และ $\hat{Y}_i$ ของวิธีการถดถอยสโตแคสติกใกล้สุด k ตัว ด้วยค่าถ่วงน้ำหนักที่เท่ากันที่ได้หลังจากการประมาณค่าสูญหายไปหา MSE

2.9.1.3 คำนวณค่า MSE ได้จากสูตร

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - Y_i)^2$$

$$= \frac{22.8812456 + 0.94702676 + 0.21999852 + \ldots + 52.6075902}{20}$$

$$= \frac{157.899996}{20}$$

$$= 7.844484$$

จะได้ MSE เท่ากับ 7.844484

2.9.2 การคำนวณด้วยโปรแกรมสำเร็จรูปอาร์

2.9.2.1 ข้อมูลตัวแปร Y จากวิธีการถดถอยสโตแคสติก

Data_Y_SRI_New

```
          [,1]
 [1,] 21.372479
 [2,] 10.557124
 [3,] 11.470720
 [4,] 10.991006
 [5,] 17.779539
 [6,] 18.835425
 [7,]  7.039945
 [8,] 17.402487
 [9,] 74.361459
[10,] 18.170170
[11,] 39.443235
[12,] 23.833115
[13,] 12.314194
[14,] 18.918645
[15,] 24.437411
[16,] 29.406952
[17,] 15.666690
[18,] 16.874308
[19,]  8.721490
[20,] 42.289688
```

รูปที่ 2.35 ข้อมูล Y ของวิธีการถดถอยสโตแคสติก

จากรูปที่ 2.35 แสดงค่า $\hat{Y}_i$ หลังจากแทนที่ค่าสูญหายของวิธีการถดถอยสโตแคสติก

2.9.2.2 คำนวณระยะห่างระหว่างข้อมูลด้วยวิธีระยะห่างยุคลิดของวิธีการประมาณค่าสูญหายเพื่อนบ้านใกล้สุด k ตัว

```
EX_KNN_X1 = matrix_X_REG[PM,2]
EX_KNN_Y = matrix_Y_REG[PM,1]
KNN_EX <- data.frame(X1,X2,Data_Y)
KNN_EX$"EX_KNN" <- sqrt(((EX_KNN_X1-X1)^2)+((EX_KNN_Y-Data_Y)^2))
KNN_EX
```

	X1	X2	Data_Y	EX_KNN
1	7	560	16.68	3.095949
2	3	220	11.50	3.750000
3	3	340	12.03	3.458092
4	4	80	14.88	2.297150
5	6	150	13.75	0.000000
6	7	330	18.11	4.473209
7	2	110	8.00	7.004463
8	7	210	17.83	4.200762
9	30	1460	79.24	69.749123
10	5	605	21.50	7.814250
11	16	688	40.33	28.398880
12	10	215	21.00	8.280248
13	4	255	13.50	2.015564
14	6	462	19.75	6.000000
15	9	448	24.00	10.680005
16	10	776	29.00	15.765865
17	6	200	15.35	1.600000
18	7	132	19.00	5.344390
19	3	36	9.50	5.202163
20	17	770	35.10	24.017129

รูปที่ 2.36 ข้อมูลระยะห่างที่คำนวณด้วยวิธีระยะห่างยุคลิด

จากรูปที่ 2.36 แสดงค่าข้อมูลระยะห่างที่คำนวณด้วยวิธีระยะห่างยุคลิด

2.9.2.3 เรียงลำดับระยะห่างที่คำนวณมาจากวิธีระยะห่างยุคลิด

```
Sort_KNN <- KNN_EX[order(KNN_EX$"EX_KNN"), ]
Sort_KNN
```

```
   X1   X2 Data_Y    EX_KNN
5   6  150  13.75  0.000000
17  6  200  15.35  1.600000
13  4  255  13.50  2.015564
4   4   80  14.88  2.297150
1   7  560  16.68  3.095949
3   3  340  12.03  3.458092
2   3  220  11.50  3.750000
8   7  210  17.83  4.200762
6   7  330  18.11  4.473209
19  3   36   9.50  5.202163
18  7  132  19.00  5.344390
14  6  462  19.75  6.000000
7   2  110   8.00  7.004463
10  5  605  21.50  7.814250
12 10  215  21.00  8.280248
15  9  448  24.00 10.680005
16 10  776  29.00 15.765865
20 17  770  35.10 24.017129
11 16  688  40.33 28.398880
9  30 1460  79.24 69.749123
```

รูปที่ 2.37 ข้อมูลที่เรียงลำดับระยะห่างจากวิธีระยะห่างยุคลิด

จากรูปที่ 2.36 แสดงค่าข้อมูลที่เรียงลำดับระยะห่างจากวิธีระยะห่างยุคลิด

2.9.2.4 ประมาณค่าสูญหายด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว

K=k+1

New_Estimate_KNN=mean(Sort_KNN[, "X2"][2:K])

New_Estimate_KNN

```
[1] 287
```

รูปที่ 2.38 ค่าประมาณข้อมูลสูญหายด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว

จากรูปที่ 2.38 แสดงค่าประมาณข้อมูลสูญหายด้วยวิธีเพื่อนบ้านใกล้สุด k ตัวมีค่าเท่ากับ 287

2.9.2.5 แทนค่าข้อมูลสูญหายที่ประมาณได้ลงในข้อมูล

Data_X2_KNN_New=replace(X2,list= PM,values = New_Estimate_KNN)

Data_X2_KNN_New

```
[1]  560  220  340   80  287  330  110  210 1460  605  688  215  255  462  448  776  200  132   36  770
```

รูปที่ 2.39 ข้อมูลที่หลังจากการแทนที่ค่าข้อมูลสูญหายของวิธีการประมาณค่าสูญหายเพื่อนบ้านใกล้สุด k ตัว

จากรูปที่ 2.39 แสดงค่าประมาณข้อมูลสูญหายที่ได้ไปแทนในตำแหน่งที่ 5 ของตัวแปรอิสระ X_2 จากค่าเดิมเท่ากับ 150 เปลี่ยนเป็น 287

2.9.2.6 กำหนดค่าข้อมูลหลังจากแทนค่าข้อมูลสูญหาย ในรูปแบบเมทริกซ์

Data_X_KNN_New = c(Xi, X1, Data_X2_KNN_New)

matrix_X_KNN_New <- matrix(Data_X_KNN_New,nrow = N,ncol = 3)

matrix_X_KNN_New

	[,1]	[,2]	[,3]
[1,]	1	7	560
[2,]	1	3	220
[3,]	1	3	340
[4,]	1	4	80
[5,]	1	6	287
[6,]	1	7	330
[7,]	1	2	110
[8,]	1	7	210
[9,]	1	30	1460
[10,]	1	5	605
[11,]	1	16	688
[12,]	1	10	215
[13,]	1	4	255
[14,]	1	6	462
[15,]	1	9	448
[16,]	1	10	776
[17,]	1	6	200
[18,]	1	7	132
[19,]	1	3	36
[20,]	1	17	770

รูปที่ 2.40 ข้อมูลหลังจากแทนค่าข้อมูลสูญหายในรูปแบบเมทริกซ์ของวิธีการประมาณค่าสูญหายเพื่อนบ้านใกล้สุด k ตัว

จากรูปที่ 2.40 แสดงค่าข้อมูล X_1, X_2 หลังจากแทนค่าข้อมูลสูญหายในรูปแบบเมทริกซ์

2.9.2.7 หาค่า X' ของข้อมูล X_1, X_2 ในรูปแบบเมทริกซ์

Transpose_X_KNN_New = t(matrix_X_KNN_New)

Transpose_X_KNN_New

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]	[,11]	[,12]	[,13]	[,14]	[,15]	[,16]	[,17]	[,18]	[,19]	[,20]
[1,]	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
[2,]	7	3	3	4	6	7	2	7	30	5	16	10	4	6	9	10	6	7	3	17
[3,]	560	220	340	80	287	330	110	210	1460	605	688	215	255	462	448	776	200	132	36	770

รูปที่ 2.41 ค่า X' ของข้อมูล X_1, X_2 ในรูปแบบเมทริกซ์

จากรูปที่ 2.41 แสดงค่าข้อมูล X_1, X_2 หลังจากแทนค่าข้อมูลสูญหายไปหาค่า X' ของวิธีการประมาณค่าสูญหายเพื่อนบ้านใกล้สุด k ตัว

2.9.2.8 หาค่า $X'X$ ของข้อมูล X_1, X_2 ในรูปแบบเมทริกซ์

TransposeX_X_KNN_New = Transpose_X_KNN_New%*%matrix_X_KNN_New

TransposeX_X_KNN_New

```
     [,1]   [,2]    [,3]
[1,]   20    162    8184
[2,]  162   2118  102531
[3,] 8184 102531 5481632
```

รูปที่ 2.42 ค่า $X'X$ ของข้อมูล X_1, X_2 ในรูปแบบเมทริกซ์ของวิธีการประมาณค่าสูญหายเพื่อนบ้านใกล้สุด k ตัว

จากรูปที่ 2.42 แสดงค่าข้อมูล X_1, X_2 หลังจากแทนค่าข้อมูลสูญหายไปหาค่า $X'X$

2.9.2.9 หาค่า $(X'X)^{-1}$ ของข้อมูล X_1, X_2 ในรูปแบบเมทริกซ์

Inverse_TransposeX_X_KNN_New = solve(TransposeX_X_KNN_New)

Inverse_TransposeX_X_KNN_New

```
              [,1]          [,2]          [,3]
[1,]  1.354324e-01 -6.035671e-03 -8.930472e-05
[2,] -6.035671e-03  5.263681e-03 -8.944318e-05
[3,] -8.930472e-05 -8.944318e-05  1.988745e-06
```

รูปที่ 2.43 ค่า $(X'X)^{-1}$ ของข้อมูล X_1, X_2 ในรูปแบบเมทริกซ์ของวิธีการประมาณค่าสูญหายเพื่อนบ้านใกล้สุด k ตัว

จากรูปที่ 2.43 แสดงค่าข้อมูล X_1, X_2 หลังจากแทนค่าข้อมูลสูญหายไปหาค่า $(X'X)^{-1}$

2.9.2.10 หาค่า $X'Y$ ของข้อมูล $X'Y$ และ Y ในรูปแบบเมทริกซ์

TransposeX_Y_KNN_New = Transpose_X_KNN_New%*%matrix_Y_REG

TransposeX_Y_KNN_New

```
          [,1]
[1,]    440.05
[2,]   5465.73
[3,] 271428.19
```

รูปที่ 2.44 ค่า $X'Y$ ของข้อมูล $X'Y$ และ Y ในรูปแบบเมทริกซ์ของวิธีการประมาณค่าสูญหายเพื่อนบ้านใกล้สุด k ตัว

จากรูปที่ 2.44 แสดงค่าข้อมูล X_1, X_2 และ Y หลังจากแทนค่าข้อมูลสูญหายไปหาค่า $X'Y$ ของวิธีการประมาณค่าสูญหายเพื่อนบ้านใกล้สุด k ตัว

2.9.2.11 หาค่า $\hat{\beta}$ ของข้อมูล X_1, X_2 และ Y

```
B_KNN_New = Inverse_TransposeX_X_KNN_New%*%TransposeX_Y_KNN_New
B_KNN_New
```

```
           [,1]
[1,] 2.36787351
[2,] 1.83646438
[3,] 0.01163066
```

รูปที่ 2.45 ค่า $\hat{\beta}$ ของข้อมูล X_1, X_2 และ Y ของวิธีการประมาณค่าสูญหายเพื่อนบ้านใกล้สุด k ตัว

จากรูปที่ 2.45 แสดงค่าข้อมูล X_1, X_2 หลังจากแทนค่าข้อมูลสูญหายและ Y ไปหาค่า $\hat{\beta}$

2.9.2.12 หาตัวแบบของข้อมูล X_1, X_2 และ Y

```
cat("Yhat=",B_KNN_New[1,1],"+",B_KNN_New[2,1],"X1+",B_KNN_New[3,1],"X2")
```

```
Yhat= 2.367874 + 1.836464 X1+ 0.01163066 X2
```

รูปที่ 2.46 ตัวแบบของข้อมูล X_1, X_2 และ Y ของวิธีการประมาณค่าสูญหายเพื่อนบ้านใกล้สุด k ตัว

จากรูปที่ 2.46 แสดงค่าข้อมูล X_1, X_2 หลังจากแทนค่าข้อมูลสูญหายและ Y ไปหาตัวแบบของวิธีการประมาณค่าสูญหายเพื่อนบ้านใกล้สุด k ตัว

2.9.2.13 ประมาณค่าข้อมูลสูญหายด้วยวิธีการถดถอยสโตแคสติกใกล้สุด k ตัวด้วยค่าถ่วงน้ำหนักที่เท่ากัน

```
Y_KNN_New = function(B_KNN_New,X1,Data_X2_KNN_New) {
B_KNN_New[1,1]+(B_KNN_New[2,1]*X1)+(B_KNN_New[3,1]*Data_X2_KNN_New)
}
Y_KNN=Y_KNN_New(B_KNN_New,X1,Data_X2_KNN_New)
Y_KSEW = function(Y_KNN,Data_Y_SRI_New) {
(Data_Y_SRI_New+Y_KNN)/2
}
Y_KSEW(Y_KNN,Data_Y_SRI_New)
```

```
      [,1]
 [1,] 21.55439
 [2,] 10.49657
 [3,] 11.65120
 [4,] 10.81759
 [5,] 17.25210
 [6,] 18.94833
 [7,]  7.18006
 [8,] 17.53402
 [9,] 74.40201
[10,] 18.37846
[11,] 39.59822
[12,] 23.53311
[13,] 12.49687
[14,] 18.83933
[15,] 24.27200
[16,] 29.58243
[17,] 15.68974
[18,] 16.81634
[19,]  8.50873
[20,] 42.41653
```

รูปที่ 2.47 ค่าประมาณข้อมูลสูญหายด้วยวิธีการถดถอยสโตแคสติกใกล้สุด k ตัวด้วยค่าถ่วงน้ำหนักที่เท่ากัน

จากรูปที่ 2.47 แสดงค่าข้อมูล $\hat{Y}$ ของวิธีการถดถอยสโตแคสติกและ $\hat{Y}$ ของวิธีเพื่อนบ้านใกล้สุด k ตัวไปถ่วงน้ำหนักที่เท่ากัน

2.9.2.14 หาค่า MSE

```
matrix_Y_KSEW_New <- matrix(Y_KSEW(Y_KNN,Data_Y_SRI_New),nrow = N,ncol = 1)
MSE_KSEW = function(N,matrix_Y_REG,matrix_Y_KSEW_New) {
(1/N)*(sum((matrix_Y_REG-matrix_Y_KSEW_New)^2))
}
MSE_KSEW(N,matrix_Y_REG,matrix_Y_KSEW_New)
```

```
[1] 7.844484
```

รูปที่ 2.48 ค่า MSE ของวิธีการถดถอยสโตแคสติกใกล้สุด k ตัวด้วยค่าถ่วงน้ำหนักที่เท่ากัน

จากรูปที่ 2.48 แสดงค่า MSE ที่ได้จากข้อมูล Y และ $\hat{Y}$ เท่ากับ 7.844484

ตารางที่ 2.22 ค่าคลาดเคลื่อนกำลังสองเฉลี่ยของแต่ละวิธีประมาณค่าสูญหาย

วิธีประมาณค่าสูญหาย	MSE
RI	8.323659
RRQ1	8.150811
RRQ3	8.221262
SRI	7.964247
KSEW	7.844484

จากตารางที่ 2.22 แสดงค่าคลาดเคลื่อนกำลังสองเฉลี่ยของแต่ละวิธีประมาณค่าสูญหาย

RI หมายถึง วิธีการถดถอยพหุคูณ มีค่า MSE เท่ากับ 8.323659

RRQ1 หมายถึง วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 มีค่า MSE เท่ากับ 8.150811

RRQ3 หมายถึง วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 มีค่า MSE เท่ากับ 8.221262

SRI หมายถึง วิธีการถดถอยสโตแคสติก มีค่า MSE เท่ากับ 7.964247

KSEW หมายถึง วิธีการถดถอยสโตแคสติกใกล้สุด k ตัว ด้วยค่าถ่วงน้ำหนักที่เท่ากัน มีค่า MSE เท่ากับ 7.844484

บทที่ 3
วิธีการดำเนินงานวิจัย

การวิจัยครั้งนี้จะดำเนินตามขั้นตอนโดยแบ่งออกเป็นขั้นตอนดังนี้

3.1 สร้างข้อมูลประชากรของตัวแปรอิสระ X_1, X_2 และค่าความคลาดเคลื่อน ε ที่มีขนาดเท่ากับ 100,000 ค่า โดยให้ข้อมูลมีการแจกแจงปรกติ (Normal Distribution) ซึ่งมีฟังก์ชันความหนาแน่นความน่าจะเป็น (Probability Density Function) คือ

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, -\infty < x < \infty, -\infty < \mu < \infty, \sigma^2 > 0$$

เมื่อ ค่าคาดหมายคือ $E(X) = \mu$ และความแปรปรวนคือ $Var(X) = \sigma^2$

โดยการสร้างข้อมูลประชากรในครั้งนี้ กำหนดให้ตัวแปรอิสระ X_1 มีพารามิเตอร์ $\mu = 3, \sigma^2 = 2.25$ ตัวแปรอิสระ X_2 มีพารามิเตอร์ $\mu = 5, \sigma^2 = 4$ และความแปรปรวนของความคลาดเคลื่อนมีพารามิเตอร์ $\mu = 0, \sigma^2 = 1, 3, 5, 7$ และ 9 (รัตติกาล, 2555)

3.2 ทำการสุ่มตัวอย่างแบบไม่ใส่คืน (Sampling Without Replacement) ของตัวแปรอิสระ X_1, X_2 และค่าความคลาดเคลื่อน ε จาก 100,000 ค่า ด้วยวิธีการสุ่มตัวอย่างอย่างง่าย (Simple Random Sampling) ที่มีขนาดตัวอย่างเท่ากับ 20, 40, 60, 80, 100 และ 120 (รัตติกาล, 2555)

3.3 กำหนดค่าพารามิเตอร์ระยะตัดแกน Y คือ $\beta_0 = 0.5$ และสัมประสิทธิ์การถดถอย คือ $\beta_1 = 1$ และ $\beta_0 = -0.3$

9.4 ทดสอบความสัมพันธ์เชิงเส้นพหุ (Multicollinearity) ระหว่างตัวแปรอิสระ กล่าวคือ ตัวแปรอิสระ X_1 และ X_2 ไม่มีความสัมพันธ์กัน โดยใช้ค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน (Pearson Correlation Coefficient) ถ้าค่าสัมประสิทธิ์สหสัมพันธ์มีค่ามากกว่าหรือเท่ากับ 0.6 ให้กลับไปทำการสุ่มตัวอย่างในขั้นตอนที่ 9.2 ใหม่อีกครั้ง แล้วนำมาทดสอบความสัมพันธ์เชิงเส้นพหุด้วยค่าสัมประสิทธิ์สหสัมพันธ์ จนกระทั่งค่าสัมประสิทธิ์สหสัมพันธ์มีค่าน้อยกว่า 0.6 แสดงว่าตัวแปรอิสระ X_1 และ X_2 ไม่มีความสัมพันธ์กัน จึงทำการสร้างตัวแปรตาม Y ในขั้นตอนที่ 9.5 ต่อไป (Glass and Hopkins, 1984; Howell, 2007)

3.5 สร้างตัวแปรตามที่มีความสัมพันธ์เชิงเส้นกับตัวแปรอิสระ ระยะตัดแกน Y สัมประสิทธิ์การถดถอยและความคลาดเคลื่อน โดยใช้รูปแบบความสัมพันธ์เชิงเส้นดังนี้

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i \qquad ; i = 1, 2, \ldots, N$$

เมื่อ Y_i คือ ตัวแปรตาม

X_{1i}, X_{2i} คือ ตัวแปรอิสระตัวที่ 1 และ 2 ตามลำดับ

β_0 คือ ระยะตัดแกน Y หรือค่าของ Y เมื่อ $X_{1i}=0, X_{2i}=0$

β_1, β_2 คือ สัมประสิทธิ์การถดถอยหรือความชันของเส้นตรง

ε_i คือ ค่าความคลาดเคลื่อน มีพารามิเตอร์ $\mu=0, \sigma^2=1,2,3,4$ และ 5 (รัตติกาล, 2555)

3.6 คำนวณจำนวนข้อมูลสูญหายและสุ่มตำแหน่งข้อมูลสูญหายของตัวแปรอิสระ

3.6.1 คำนวณหาจำนวนข้อมูลสูญหายของตัวแปรอิสระดังนี้

จำนวนข้อมูลสูญหาย = ขนาดตัวอย่าง x ร้อยละของการสูญหาย

ถ้าค่าที่คำนวณได้เป็นเลขทศนิยม จะใช้เลขจำนวนเต็มที่น้อยที่สุดที่มีค่ามากกว่าค่านั้น

3.6.2 ทำการสุ่มตำแหน่งข้อมูลสูญหายของตัวแปรอิสระ X_2 โดยกำหนดให้เป็นการสูญหายอย่างสุ่ม และกำหนดให้ร้อยละข้อมูลสูญหายคือ 5, 10, 15 และ 20

ตารางที่ 3.1 จำนวนข้อมูลสูญหายและจำนวนข้อมูลปกติที่มีข้อมูลสูญหายร้อยละ 5, 10, 15 และ 20 ตามตัวอย่างขนาดเล็ก ขนาดปานกลาง และขนาดใหญ่

ขนาดตัวอย่าง (n)	ร้อยละของข้อมูลสูญหาย							
	5		10		15		20	
	จำนวนข้อมูลสูญหาย	จำนวนข้อมูลปกติ	จำนวนข้อมูลสูญหาย	จำนวนข้อมูลปกติ	จำนวนข้อมูลสูญหาย	จำนวนข้อมูลปกติ	จำนวนข้อมูลสูญหาย	จำนวนข้อมูลปกติ
ตัวอย่างขนาดเล็ก								
20	1	19	2	18	3	17	4	16
40	2	38	4	36	6	34	8	32
ตัวอย่างขนาดปานกลาง								
60	3	57	6	54	9	51	12	48
80	4	76	8	72	12	68	16	64
ตัวอย่างขนาดใหญ่								
100	5	95	10	90	15	85	20	80
120	6	114	12	108	18	102	24	96

3.7 ทำการประมาณข้อมูลสูญหายของข้อมูลตัวแปรอิสระทั้ง 5 วิธีคือ

3.7.1 วิธีการถดถอยพหุคูณ (Multiple Regression Imputation)

วิธีการถดถอยพหุคูณเป็นวิธีการประมาณค่าสูญหายตัวแปรตาม Y จากข้อมูลสูญหายของตัวแปรอิสระ X_2 โดยทำการประมาณสมการถดถอยของตัวแปรตาม Y โดยใช้ข้อมูลที่ไม่สูญหาย $X_i, Y_i;\ i = 1, \ldots, r$ วิธีนี้เสนอโดย Buck ในปี ค.ศ.1960 โดยกำหนดให้สมการถดถอยของตัวแปรตาม Y เป็นดังนี้ (ทัตดา, 2554)

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i \qquad ; \quad \mathrm{i} = 1,\ 2, \ldots,\ N$$

เมื่อ Y_i คือ ตัวแปรตาม

X_{1i}, X_{2i} คือ ตัวแปรอิสระตัวที่ 1 และ 2 ตามลำดับ

β_0 คือ ระยะตัดแกน Y หรือค่าของ Y เมื่อ $X_{1i} = 0$ และ $X_{2i} = 0$

β_1, β_2 คือ สัมประสิทธิ์การถดถอยหรือความชันของเส้นตรง

ε_i คือ ค่าความคลาดเคลื่อน โดยกำหนดค่าความคลาดเคลื่อนมีพารามิเตอร์ μ = 0, $\sigma^2 = 1, 3, 5, 7$ และ 9

ข้อมูลหลังจากแทนที่ค่าข้อมูลสูญหายด้วยวิธีค่าเฉลี่ย (Series mean) คำนวณได้ดังนี้ (ฐณัฐ, 2559)

$$\hat{X}_i = \frac{\sum_{i=1}^{k} X_i}{k}$$

เมื่อ X_i คือ ข้อมูลที่ไม่สูญหาย

k คือ จำนวนข้อมูลที่ไม่สูญหาย

$\hat{X}$ คือ ค่าประมาณข้อมูลที่สูญหายด้วยค่าเฉลี่ยของข้อมูลที่ไม่สูญหาย

จึงจะได้ค่าประมาณของตัวแปรตาม Y_i จากการสุ่มของตัวแปรอิสระ X_{1i}, X_{2i} ด้วยการคำนวณวิธีการถดถอยพหุคูณ ก่อนที่จะประมาณข้อมูลสูญหายด้วยวิธีการถดถอยพหุคูณ ดังแสดงในตารางที่ 3.2

ตารางที่ 3.2 ข้อมูลค่าประมาณของตัวแปรตาม Y_i จากการสุ่มของตัวแปรอิสระ X_{1i}, X_{2i} ด้วยการคำนวณวิธีการถดถอยพหุคูณ ก่อนที่จะประมาณข้อมูลสูญหายด้วยวิธีการถดถอยพหุคูณ

X_{1i}	X_{2i}	Y_i
X_{11}	X_{21}	Y_1
X_{12}	X_{22}	Y_2
X_{13}	X_{23}	Y_3
$\vdots$	$\vdots$	$\vdots$
X_{1r}	X_{2r}	Y_r

ลักษณะของข้อมูลหลังจากแทนที่ค่าข้อมูลสูญหายด้วยวิธีค่าเฉลี่ย (Series mean) ดังนี้ (ฐณัฐ, 2559)

$$\hat{X}_i = \frac{\sum_{i=1}^{k} X_i}{k}$$

เมื่อ X_i คือ ข้อมูลที่ไม่สูญหาย

k คือ จำนวนข้อมูลที่ไม่สูญหาย

$\hat{X}$ คือ ค่าประมาณข้อมูลที่สูญหายด้วยค่าเฉลี่ยของข้อมูลที่ไม่สูญหาย

ข้อมูลหลังจากการประมาณข้อมูลสูญหายด้วยวิธีการประมาณข้อมูลสูญหายโดยวิธีค่าเฉลี่ย (Series mean) ดังแสดงในตารางที่ 3.3

ตารางที่ 3.3 ข้อมูลหลังจากการประมาณข้อมูลสูญหายด้วยค่าประมาณจากการคำนวณด้วยวิธีการถดถอยพหุคูณ

X_{1i}	X_{2i}	$X_{2i,new}$	$Y_{i,new}$
X_{11}	X_{21}	X_{21}	$Y_{1,new}$
X_{12}	X_{22}	X_{22}	$Y_{2,new}$
X_{13}	X_{23}	X_{23}	$Y_{3,new}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
			$Y_{r,new}$ } ข้อมูลหลังจากแทนที่ค่าข้อมูลสูญหาย
X_{1r}	X_{2r}	X_{2r}	
$X_{1,r+1}$	} ข้อมูลสูญหาย	$X_{2,r+1,new}$	$Y_{r+1,new}$
$\vdots$		$\vdots$	$\vdots$
X_{1n}		$X_{2n,new}$	$Y_{n,new}$

นำข้อมูลตัวแปรอิสระ X_{1i} และตัวแปรอิสระ X_{2i} ที่แทนค่าสูญหายด้วยวิธีค่าเฉลี่ย (Series mean) มาประมาณค่าของตัวแปรตาม Y โดยวิธีการถดถอยพหุคูณ
ดังนั้นค่าประมาณของตัวแปรตาม Y คือ

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i}$$

เมื่อ $\hat{Y}_i$ คือ ค่าประมาณของตัวแปรตาม Y ของหน่วยตัวอย่างที่ $i, i = 1, \ldots, r$

$\hat{\beta}_0$ คือ ค่าระยะตัดแกน Y

$\hat{\beta}_1, \hat{\beta}_2$ คือ ค่าประมาณของสัมประสิทธิ์ของการถดถอย

X_{1i}, X_{2i} คือ ค่าของตัวแปรอิสระ X ที่ไม่สูญหายของหน่วยตัวอย่างที่ $i, i = 1, \ldots, r$

3.7.2 วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 (Regression-Ratio Quartile1 Imputation : RRQ1)

วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีการประมาณค่าสูญหายด้วยอัตราส่วน ซึ่งประยุกต์โดยการใช้ค่า $\hat{Y}_{reg}$ ซึ่งเป็นค่า $\hat{Y}$ ที่ได้จากการประมาณค่าสูญหายด้วยวิธีการถดถอยมาแทนค่า $\bar{Y}$ และใช้ประโยชน์จากควอร์ไทล์ที่ 1 ของตัวแปร $X_1;1,2$ (รัตติกาล, 2555)

3.7.3 วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 (Regression-Ratio Quartile3 Imputation : RRQ3)

วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 เป็นวิธีการประมาณค่าสูญหายด้วยอัตราส่วน ซึ่งประยุกต์โดยการใช้ค่า $\hat{Y}_{reg}$ ซึ่งเป็นค่า $\hat{Y}$ ที่ได้จากการประมาณค่าสูญหายด้วยวิธีการถดถอยมาแทนค่า $\bar{Y}$ และใช้ประโยชน์จากควอร์ไทล์ที่ 3 ของตัวแปร $X_i;1,2$ (รัตติกาล, 2555)

การแทนค่าสูญหายการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 (Regression-Ratio Quartile1 Imputation : RRQ1) และ การแทนค่าสูญหายการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 (Regression-Ratio Quartile3 Imputation : RRQ3) เป็นดังนี้

$$Y_{RRQ1} = Y_j = Y_{reg}\left(\frac{\overline{X}_{in} + q_1}{\overline{X}_{ir} + q_1}\right)$$

$$Y_{RRQ3} = Y_j = Y_{reg}\left(\frac{\overline{X}_{in} + q_3}{\overline{X}_{ir} + q_3}\right)$$

เมื่อ $\hat{Y}_j$ คือ ค่าประมาณตัวแปรตาม $Y,\ j = r+1, r+2, \ldots, n$

$\bar{X}_{in}$ คือ ค่าเฉลี่ยตัวอย่างที่สมบูรณ์ของตัวแปรอิสระ X_2

$\bar{X}_{ir}$ คือ ค่าเฉลี่ยตัวอย่างที่ตัดค่าสูญหายออกของตัวแปรอิสระ X_2

Y_{reg} คือ ค่าเฉลี่ยตัวแปรตาม Y ที่ประมาณได้จากวิธีการถดถอย

q_1 และ q_3 คือ ค่าควอร์ไทล์ที่ 1 และ 3 ของตัวอย่างของตัวแปรอิสระ X_2 ที่มีขนาด n

3.7.4 วิธีการถดถอยสโตแคสติก (Stochastic Regression Imputation)

วิธีการถดถอยสโตแคสติกเป็นวิธีการประมาณค่าสูญหายของตัวแปรตาม Y จากข้อมูลสูญหายของตัวแปรอิสระ X_2 โดยทำการประมาณสมการถดถอยของตัวแปรตาม Y โดยใช้ข้อมูลที่ไม่สูญหาย $X_i, Y_i;\ i = 1,\ldots,r$ วิธีนี้จะแตกต่างจากวิธีการถดถอยคือ วิธีการถดถอยสโตแคสติกจะเพิ่มเทอมความคลาดเคลื่อนสุ่ม Z_i เข้ามาเป็นตัวช่วยในการประมาณค่าข้อมูลสูญหาย โดยได้มาจากการสุ่มด้วยวิธีมอนติคาร์โล (Monte Carlo Method) เพื่อประมาณสมการถดถอยของตัวแปรตาม $\left(Y_i^*\right);\ j = r+1,\ldots,n$ วิธีนี้เสนอโดย Buck ในปี ค.ศ.1960 โดยกำหนดให้สมการถดถอยของตัวแปรตาม Y เป็นดังนี้ (Chaimongkol, 2005; ทัตดา, 2554)

$$Y_{SRI} = Y_i^* = Y_{REQ} + Z_i$$

โดยที่
$$Y_{REG} = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}$$

เมื่อ $\hat{Y}_{SRI}, \hat{Y}_i^*$ คือ ค่าประมาณของตัวแปรตาม Y ของหน่วยตัวอย่างที่ i จากการเพิ่มเทอมค่าคลาดเคลื่อนอย่างสุ่ม

$\hat{Y}_{REG}, \hat{Y}_i$ คือ ค่าประมาณของตัวแปรตาม Y ของหน่วยตัวอย่างที่ i จากการเพิ่มเทอมค่าคลาดเคลื่อนอย่างสุ่ม

$\hat{\beta}_0$ คือ ค่าระยะตัดแกน Y

$\hat{\beta}_1, \hat{\beta}_2$ คือ ค่าประมาณของสัมประสิทธิ์ของการถดถอย

X_{1i}, X_{2i} คือ ค่าของตัวแปรอิสระ X ที่ไม่สูญหายของหน่วยตัวอย่างที่ $i, i = 1,\ldots,r$

Z_i คือ ค่าประมาณของค่าคลาดเคลื่อน โดยมีการแจกแจงปรกติ ค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 1, 2, 3, 4 และ 5 ซึ่งค่าประมาณได้มาจากการสุ่มด้วยวิธีมอนติคาร์โล

ข้อมูลหลังจากแทนที่ค่าข้อมูลสูญหายด้วยวิธีค่าเฉลี่ย (Series mean) ดังนี้ (ฐณัฐ, 2559)

$$\hat{X}_i = \frac{\sum_{i=1}^{k} X_i}{k}$$

เมื่อ X_i คือ ข้อมูลที่ไม่สูญหาย

k คือ จำนวนข้อมูลที่ไม่สูญหาย

$\hat{X}$ คือ ค่าประมาณข้อมูลที่สูญหายด้วยค่าเฉลี่ยของข้อมูลที่ไม่สูญหาย

จะได้ค่าประมาณของตัวแปรตาม Y_i จากการสุ่มของตัวแปรอิสระ X_{1i}, X_{2i} ด้วยการคำนวณวิธีการถดถอยพหุคูณ ก่อนที่จะประมาณข้อมูลสูญหายด้วยวิธีการถดถอยสโตแคสติก ดังแสดงในตารางที่ 3.4

ตารางที่ 3.4 ข้อมูลค่าประมาณของตัวแปรตาม Y_i จากการสุ่มของตัวแปรอิสระ X_{1i}, X_{2i} ด้วยการคำนวณวิธีการถดถอยพหุคูณ ก่อนที่จะประมาณข้อมูลสูญหายด้วยวิธีการถดถอยสโตแคสติก

X_{1i}	X_{2i}	Y_i^*
X_{11}	X_{21}	$Y_1^* = Y_1$
X_{12}	X_{22}	$Y_2^* = Y_2$
X_{13}	X_{23}	$Y_3^* = Y_3$
$\vdots$	$\vdots$	$\vdots$
X_{1r}	X_{2r}	$Y_r^* = Y_r$

ลักษณะของข้อมูลหลังจากแทนที่ค่าข้อมูลสูญหายด้วยวิธีค่าเฉลี่ย (Series mean) ดังนี้ (ฐณัฐ, 2559)

$$\hat{X}_i = \frac{\sum_{i=1}^{k} X_i}{k}$$

เมื่อ X_i คือ ข้อมูลที่ไม่สูญหาย

k คือ จำนวนข้อมูลที่ไม่สูญหาย

$\hat{X}$ คือ ค่าประมาณข้อมูลที่สูญหายด้วยค่าเฉลี่ยของข้อมูลที่ไม่สูญหาย

ข้อมูลหลังจากการประมาณข้อมูลสูญหายด้วยวิธีการประมาณข้อมูลสูญหายโดยวิธีค่าเฉลี่ย (Series mean) ดังแสดงในตารางที่ 3.5

ตารางที่ 3.5 ข้อมูลหลังจากการประมาณข้อมูลสูญหายด้วยค่าประมาณจากการคำนวณด้วยวิธีการถดถอยสโตแคสติก

X_{1i}	X_{2i}	$X_{2i,new}$	Z_i	$Y^*_{i,new}$
X_{11} X_{12} X_{13} $\vdots$ X_{1r}	X_{21} X_{22} X_{23} $\vdots$ X_{2r}	X_{21} X_{22} X_{23} $\vdots$ X_{2r}	Z_1 Z_2 Z_3 $\vdots$ Z_r	$Y^*_{1,new}$ $Y^*_{2,new}$ $Y^*_{3,new}$ $\vdots$ $Y^*_{r,new}$
$X_{1,r+1}$ $\vdots$ X_{1n}	} ข้อมูลสูญหาย	$X_{2,r+1,new}$ $\vdots$ $X_{2n,new}$	Z_{r+1} $\vdots$ Z_n	$Y^*_{r+1,new}$ } ข้อมูลหลังจากแทนที่ค่าข้อมูลสูญหาย $\vdots$ $Y^*_{n,new}$

นำข้อมูลตัวแปรอิสระ X_{1i} และตัวแปรอิสระ X_{2i} ที่แทนค่าสูญหายด้วยวิธีค่าเฉลี่ย (Series mean) มาประมาณค่าของตัวแปรตาม Y โดยวิธีการถดถอยสโตแคสติก

ดังนั้นค่าประมาณของตัวแปรตาม Y คือ

$$Y_i^* = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + Z_i$$

เมื่อ $\hat{Y}_i^*$ คือ ค่าประมาณของตัวแปรตาม Y ของหน่วยตัวอย่างที่ $i, i = 1, \ldots, r$

$\hat{\beta}_0$ คือ ค่าระยะตัดแกน Y

$\hat{\beta}_1, \hat{\beta}_2$ คือ ค่าประมาณของสัมประสิทธิ์ของการถดถอย

X_{1i}, X_{2i} คือ ค่าของตัวแปรอิสระ X ที่ไม่สูญหายของหน่วยตัวอย่างที่ $i, i = 1, \ldots, r$

Z_i คือ ค่าประมาณของค่าคลาดเคลื่อน โดยมีการแจกแจงปรกติ ค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 1, 3, 5, 7 และ 9 ซึ่งค่าประมาณได้มาจากการสุ่มด้วยวิธีมอนติคาร์โล

3.7.5 วิธีการถดถอยสโตแคสติกใกล้สุด k ตัว ด้วยค่าถ่วงน้ำหนักที่เท่ากัน (K-Nearest Stochastic Regression Imputation with Equivalent Weighted)

วิธีการถดถอยสโตแคสติกใกล้สุด k ตัว ด้วยค่าถ่วงน้ำหนักที่เท่ากันเป็นวิธีการประยุกต์ใช้ค่าถ่วงน้ำหนักที่เท่ากัน (Equivalent Weighted) ให้กับค่าประมาณที่ได้จากวิธีการประมาณค่าข้อมูลสูญหาย 2 วิธี ได้แก่ วิธีเพื่อนบ้านใกล้สุด k ตัว และวิธีการถดถอยสโตแคสติก โดยมีขั้นตอนดังนี้

3.7.5.1 ประมาณค่าข้อมูลสูญหายด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว มีขั้นตอนการดำเนินการดังนี้ (Troyanskaya et al., 2001)

1) กำหนดค่า k

จากสูตร $k=\sqrt{n}$ โดยที่ k เป็นจำนวนเต็มคี่ที่มีค่าใกล้เคียงกับ $\sqrt{n}$ มากที่สุด เมื่อ k คือจำนวนข้อมูลที่มีความใกล้เคียงหรือคล้ายคลึงกับข้อมูลที่เราสนใจ ค่า k ควรมีค่าอยู่ในช่วงอยู่ระหว่าง 10-20 เนื่องจากถ้าค่า k มีค่ามากเกินไปจะทำให้ความถูกต้องลดลงและใช้เวลาในการคำนวณเพิ่มขึ้น (Troyanskaya et al., 2001) และ n เป็นจำนวนข้อมูลทั้งหมด

ตารางที่ 3.6 จำนวนของค่า k ในแต่ละขนาดตัวอย่าง

ขนาดตัวอย่าง (n)	ค่า k	
	ค่าที่คำนวณได้	ค่าที่นำไปใช้
ตัวอย่างขนาดเล็ก		
20	4.4721	5
40	6.3246	7
ตัวอย่างขนาดกลาง		
60	7.7459	7
80	8.9443	9
ตัวอย่างขนาดใหญ่		
100	10	11
120	10.9545	11

2) คำนวณหาระยะห่างระหว่างข้อมูลตัวอย่างที่สนใจกับข้อมูลอื่นๆ ทุกตัวด้วยวิธีระยะห่างยุคลิด (Euclidean Distance)

$$dist\left(X_i, X_j\right)=\sqrt{\sum_{k=1}^{n}\left(X_{i,k}-X_{j,k}\right)^2}$$

เมื่อ $dist(X_i, X_j)$ คือ ระยะห่างระหว่างตัวอย่าง X_i กับตัวอย่าง X_j

n คือ จำนวนข้อมูลทั้งหมด

$X_{i,k}$ คือ ค่าที่คำนวณได้จากขนาดตัวอย่างตัวที่ k ของตัวอย่าง X_i

$X_{j,k}$ คือ ค่าที่คำนวณได้จากขนาดตัวอย่างตัวที่ k ของตัวอย่าง X_j

3) เรียงลำดับระยะห่างระหว่างจุด โดยพิจารณาจากข้อมูลที่ใกล้ที่สุดจำนวน k ตัว ตามที่กำหนดในการประมาณค่าข้อมูลสูญหาย ทำการประมาณค่าสูญหายโดยการหาค่าเฉลี่ยจากข้อมูลที่ใกล้ที่สุด k ตัว จากตัวแปรเดียวกันกับตำแหน่งที่มีการสูญหาย มีสูตรการคำนวณดังนี้

ค่าสูญหาย คือ $$X_i = \frac{\sum_{i=1}^{k} X_i}{k}$$

4) นำข้อมูลที่ได้ประมาณค่าสูญหายแล้ว มาทำการประมาณค่าสัมประสิทธิ์การถดถอยใหม่ด้วยวิธีกำลังสองน้อยที่สุดสามัญ (Ordinary Least Square Method) เพื่อหาสมการถดถอยพหุคูณที่ใช้ในการทำนายค่าของตัวแปรตาม Y ดังนี้

$$\widehat{Y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 X_{1i} + \widehat{\beta}_2 X_{2i} + \varepsilon_i$$

จะได้ค่าประมาณของข้อมูลตัวแปรตามจากข้อมูลสูญหายของตัวแปรอิสระ X_2 โดยวิธีเพื่อนบ้านใกล้สุด k ตัว คือ $\hat{Y}_i$ และวิธีการถดถอยสโตแคสติกตามในหัวข้อ 3.7.4 จากข้อมูลที่ทราบค่าคือ $\hat{Y}_i^*$

3.7.5.2 ถ่วงน้ำหนักด้วยวิธีการให้น้ำหนักที่เท่ากันให้กับค่าประมาณที่ได้จากวิธีการประมาณค่าสูญหายโดยวิธีเพื่อนบ้านใกล้สุด k ตัว และวิธีการถดถอยสโตแคสติก ดังนี้

$$\overline{Y}_i^* = W_M\left(\hat{Y}_i^* + Y_i\right) \qquad ; \quad i = r+1,\ldots,n$$

โดยที่ $$W_M = \frac{1}{M}$$

เมื่อ $\overline{Y}_i^*$ คือ ค่าประมาณที่ได้จากการถ่วงน้ำหนักของวิธีการประมาณค่าสูญหายโดยวิธีเพื่อนบ้านใกล้สุด k ตัว และวิธีการถดถอยสโตแคสติก

$\hat{Y}_i^*$ คือ ค่าประมาณที่ได้จากวิธีการประมาณค่าสูญหายโดยวิธีการถดถอยสโตแคสติก

$\hat{Y}_i$ คือ ค่าประมาณที่ได้จากวิธีการประมาณค่าสูญหายโดยวิธีเพื่อนบ้านใกล้สุด k ตัว

W_M คือ ค่าถ่วงน้ำหนักที่เท่ากันของวิธีการประมาณค่าสูญหาย

M คือ จำนวนวิธีการประมาณค่าสูญหายด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว และวิธีการถดถอยสโตแคสติก ดังนั้นจะได้ $M = 2$

r คือ ลำดับที่ของวิธีการประมาณค่าสูญหาย

3.8 เปรียบเทียบวิธีการประมาณค่าสูญหายของตัวแปรอิสระจากค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error, MSE) ทั้ง 5 วิธี ดังนี้ (Sujitta et al., 2006; รัตติกาล, 2555; เรืองลักษณ์และคณะ, 2560; Muhammad and Klairung, 2022)

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - Y_i)^2$$

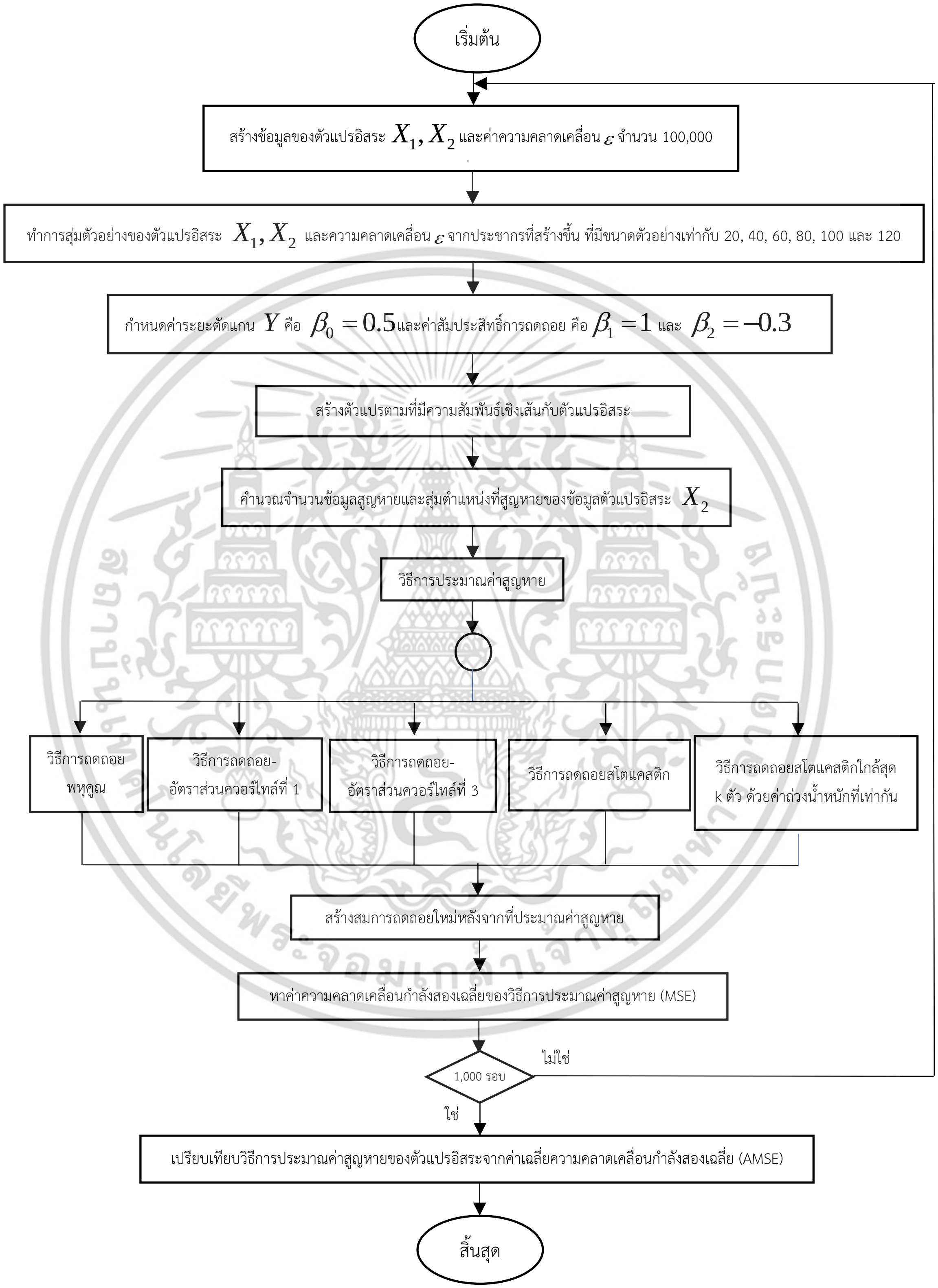
เมื่อ Y_i แทน ค่าจริงของข้อมูลตัวแปรตามตัวที่ i

$\hat{Y}_i$ แทน ค่าทำนายของข้อมูลตัวแปรตามตัวที่ i

MSE แทน ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของค่าทำนายตัวแปรตาม

โดยการกระทำตามขั้นตอนต่างๆ จะเปลี่ยนค่าความแปรปรวนของความคลาดเคลื่อนของข้อมูล 4 ค่า ขนาดตัวอย่าง 6 ค่า และร้อยละข้อมูลสูญหายของตัวแปรอิสระ 4 ค่า จนครบทุกสถานการณ์ จำนวน 1,000 รอบแล้วนำมาหาค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ย (Average Mean Square Error, AMSE)

หมายเหตุ : วิธีที่มีประสิทธิภาพมากที่สุด จะมีค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยน้อยที่สุด



รูปที่ 3.1 วิธีการดำเนินงานการประมาณค่าสูญหายในการวิเคราะห์การถดถอยพหุคูณ

บทที่ 4

ผลการวิเคราะห์ข้อมูล

งานวิจัยครั้งนี้คณะผู้วิจัยนำชุดข้อมูล 120 ชุด โดยมีสถานการณ์ที่แตกต่างกัน โดยกำหนดประชากรของตัวแปรอิสระ X_1 มีพารามิเตอร์ $\mu = 3, \sigma^2 = 2.25$ ตัวแปรอิสระ X_2 มีพารามิเตอร์ $\mu = 5, \sigma^2 = 4$ และความแปรปรวนมีพารามิเตอร์ $\mu = 0, \sigma^2 = 1, 3, 5, 7$ และ 9 โดยทำการสุ่มอย่างละ 100,000 ค่า โดยทำการสุ่มตัวอย่างจากประชากรของแต่ละตัวแปรตามขนาดตัวอย่าง คือขนาดตัวอย่างเล็กเท่ากับ 20 และ 40 ขนาดตัวอย่างปานกลางเท่ากับ 60 และ 80 ขนาดตัวอย่างใหญ่เท่ากับ 100 และ 120 กำหนดค่าพารามิเตอร์ในการประมาณตัวแปรตาม Y คือ $\beta_0 = 0.5, \beta_1 = 1$ และ $\beta_2 = -0.3$ และกำหนดความแปรปรวน σ^2 เท่ากับ 1, 3, 5, 7 และ 9 โดยกำหนดร้อยละข้อมูลสูญหาย เท่ากับ 5, 10, 15 และ 20 มาใช้ในการแทนค่าข้อมูลสูญหาย 5 วิธีคือ วิธีการถดถอยพหุคูณ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 วิธีการถดถอยสโตแคสติกและวิธีการถดถอยสโตแคสติกใกล้สุด k ตัว ด้วยค่าถ่วงน้ำหนักที่เท่ากัน ซึ่งสามารถเปรียบเทียบวิธีการประมาณค่าสูญหายของตัวแปรอิสระจากค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ย (Average Mean Square Error, AMSE) จากการทำจำนวน 1,000 รอบโดยที่วิธีการประมาณค่าสูญหายที่มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยน้อยที่สุด จะมีประสิทธิภาพดีที่สุดในละสถานการณ์

สัญญลักษณ์ RI หมายถึง วิธีการถดถอยพหุคูณ

RRQ1 หมายถึง วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1

RRQ3 หมายถึง วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3

SRI หมายถึง วิธีการถดถอยสโตแคสติก

KSEW หมายถึง วิธีการถดถอยสโตแคสติกใกล้สุด k ตัว ด้วยค่าถ่วงน้ำหนักที่เท่ากัน

ตารางที่ 4.1 ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย เมื่อค่าความแปรปรวนของความคลาดเคลื่อนเท่ากับ 1

ขนาดตัวอย่าง	ค่าสูญหาย	วิธีประมาณค่าสูญหาย				
	(%)	RI	RRQ1	RRQ3	SRI	KSEW
ตัวอย่างขนาดเล็ก						
20	5%	**1.0038**	1.0045	1.0044	1.0498	1.0131
	10%	1.0044	**1.0015**	1.0019	1.0552	1.0153
	15%	1.0226	**1.013**	1.0141	1.0731	1.0316
	20%	1.0381	**1.0265**	1.0269	1.091	1.0433
40	5%	1.0145	**1.0135**	1.0136	1.038	1.0189
	10%	1.0226	**1.0174**	1.0181	1.0494	1.0282
	15%	1.0207	**1.0113**	1.0122	1.0454	1.022
	20%	1.0628	**1.0498**	1.0501	1.0872	1.0569
ตัวอย่างขนาดปานกลาง						
60	5%	1.0137	**1.0122**	1.0125	1.0295	1.0164
	10%	1.0196	**1.014**	1.0148	1.0357	1.0214
	15%	1.041	**1.0322**	1.0329	1.0578	1.0388
	20%	1.062	1.051	**1.0508**	1.0784	1.0535
80	5%	1.0047	**1.0031**	1.0033	1.0178	1.0075
	10%	1.021	**1.0149**	1.0157	1.035	1.0232
	15%	1.0259	**1.0163**	1.0171	1.0378	1.0234
	20%	1.0502	1.0391	**1.0389**	1.063	1.0398
ตัวอย่างขนาดใหญ่						
100	5%	1.0102	**1.0085**	1.0088	1.0194	1.0118
	10%	1.0181	**1.0124**	1.0131	1.029	1.019
	15%	1.0305	**1.0202**	1.021	1.0415	1.0278
	20%	1.0516	1.0391	**1.039**	1.0613	1.0414
120	5%	1.006	**1.0042**	1.0045	1.0141	1.0075
	10%	1.0202	**1.0148**	1.0155	1.0285	1.0199
	15%	1.0331	**1.0238**	1.0245	1.042	1.0296
	20%	1.0522	1.041	**1.0408**	1.0609	1.0409

จากตารางที่ 4.1 เมื่อค่าความแปรปรวนของความคลาดเคลื่อนมีค่าเท่ากับ 1 ขนาดตัวอย่าง 20 ร้อยละข้อมูลสูญหาย 5 วิธีการถดถอยพหุคูณ เป็นวิธีที่มีประสิทธิภาพมากที่สุด ส่วนร้อยละข้อมูลสูญหาย 10, 15 และ 20 วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด

ขนาดตัวอย่าง 40 ร้อยละข้อมูลสูญหาย 5, 10, 15 และ 20 วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด

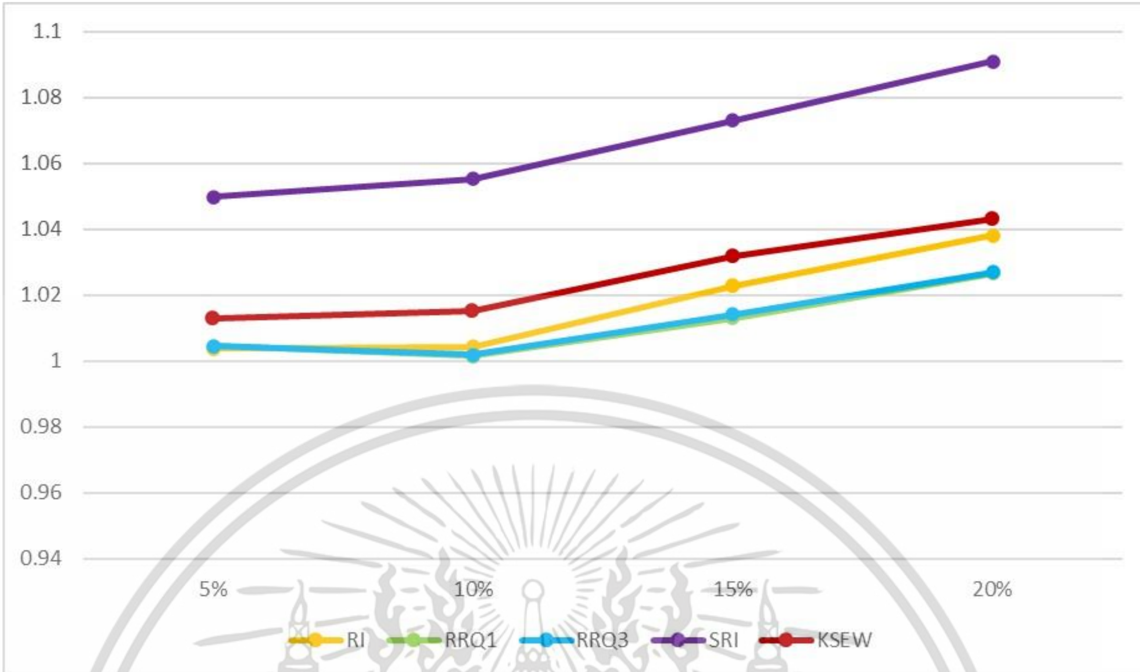
ขนาดตัวอย่าง 60 ร้อยละข้อมูลสูญหาย 5, 10 และ 15 วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด ส่วนร้อยละข้อมูลสูญหาย 20 วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 เป็นวิธีที่มีประสิทธิภาพมากที่สุด

ขนาดตัวอย่าง 80 ร้อยละข้อมูลสูญหาย 5, 10 และ 15 วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด ส่วนร้อยละข้อมูลสูญหาย 20 วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 เป็นวิธีที่มีประสิทธิภาพมากที่สุด

ขนาดตัวอย่าง 100 ร้อยละข้อมูลสูญหาย 5, 10 และ 15 วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด ส่วนร้อยละข้อมูลสูญหาย 20 วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 เป็นวิธีที่มีประสิทธิภาพมากที่สุด

ขนาดตัวอย่าง 120 ร้อยละข้อมูลสูญหาย 5, 10 และ 15 วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด ส่วนร้อยละข้อมูลสูญหาย 20 วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 เป็นวิธีที่มีประสิทธิภาพมากที่สุด

จากตารางที่ 4.1 จะพบว่าตัวอย่างขนาดเล็ก ตัวอย่างขนาดปานกลางและตัวอย่างขนาดใหญ่ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด

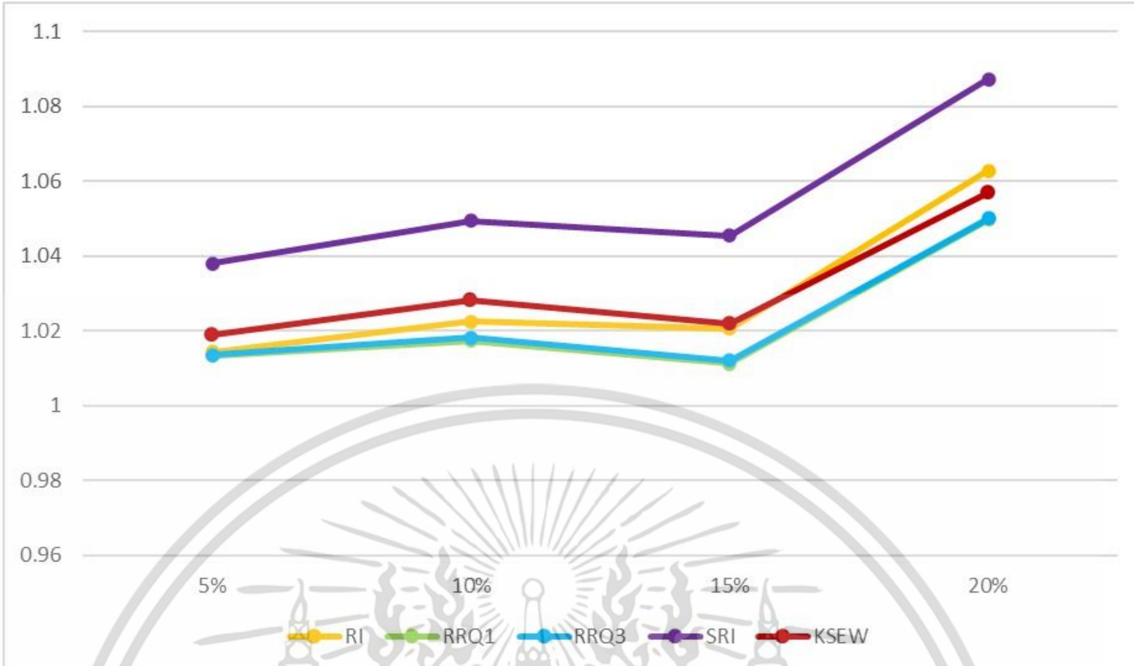


รูปที่ 4.1 การเปรียบเทียบค่า AMSE ที่ขนาดตัวอย่าง 20 ความแปรปรวนของความคลาดเคลื่อน 1

จากรูปที่ 4.1 จะเห็นได้ว่าที่ขนาดตัวอย่าง 20 ความแปรปรวนของความคลาดเคลื่อนเท่ากับ 1 ร้อยละข้อมูลสูญหาย 5 วิธีการถดถอยพหุคูณ เป็นวิธีที่มีประสิทธิภาพมากที่สุด รองลงมาคือ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 ส่วนร้อยละข้อมูลสูญหาย 10, 15 และ 20 วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1เป็นวิธีที่มีประสิทธิภาพมากที่สุด รองลงมาคือ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3

จากรูปที่ 4.1 จะเห็นแนวโน้มว่าทุกร้อยละข้อมูลสูญหาย วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด

จะเห็นแนวโน้มว่า เมื่อร้อยละข้อมูลสูญหายมีค่าเพิ่มมากขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยมีค่าเพิ่มมากขึ้นตามไปด้วย

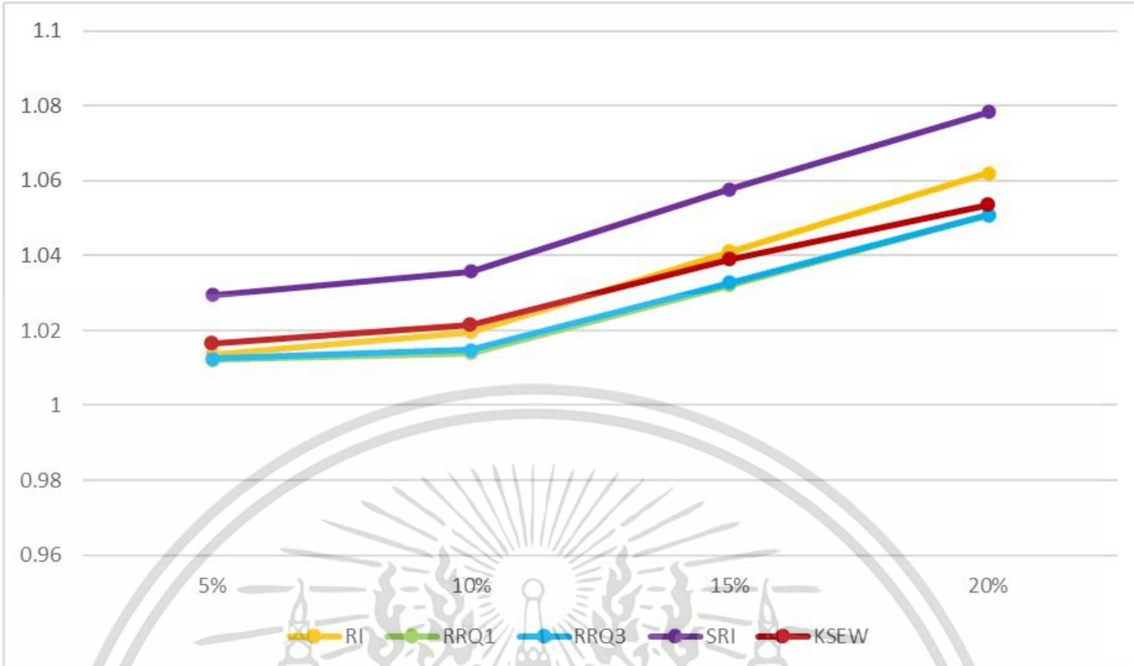


รูปที่ 4.2 การเปรียบเทียบค่า AMSE ที่ขนาดตัวอย่าง 40 ความแปรปรวนของความคลาดเคลื่อน 1

จากรูปที่ 4.2 จะเห็นได้ว่าที่ขนาดตัวอย่าง 40 ความแปรปรวนของความคลาดเคลื่อนเท่ากับ 1 ร้อยละข้อมูลสูญหาย 5, 10, 15 และ 20 วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1เป็นวิธีที่มีประสิทธิภาพมากที่สุด รองลงมาคือ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3

จากรูปที่ 4.2 จะเห็นแนวโน้มว่าทุกร้อยละข้อมูลสูญหาย วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด

จะเห็นแนวโน้มว่า เมื่อร้อยละข้อมูลสูญหายมีค่าเพิ่มมากขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยมีค่าเพิ่มมากขึ้นตามไปด้วย

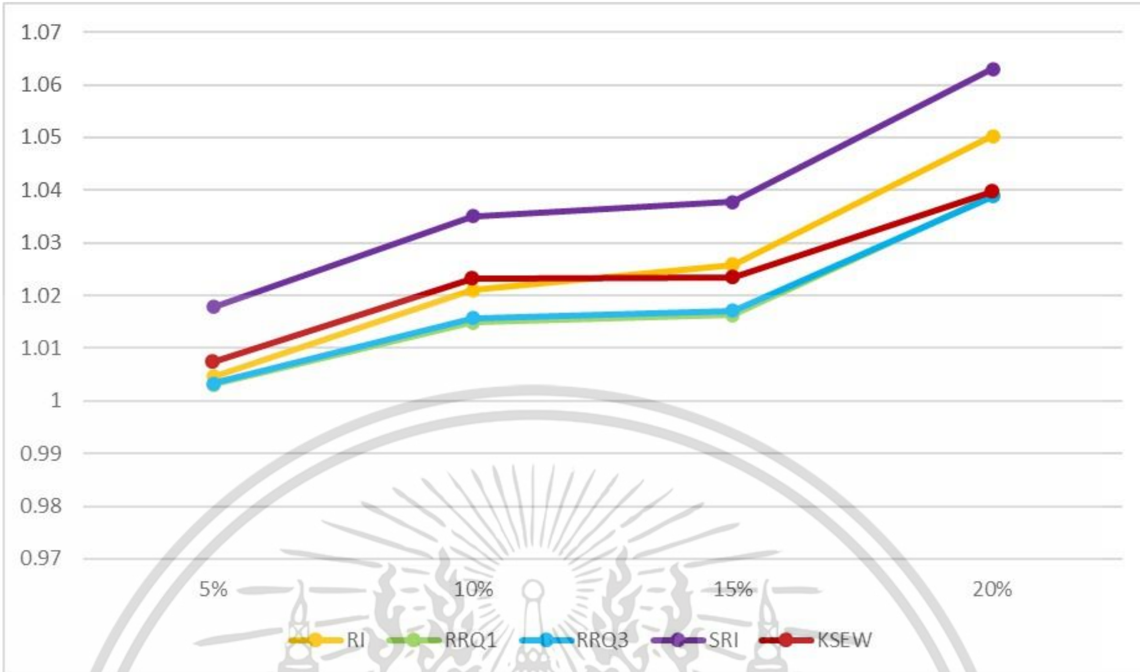


รูปที่ 4.3 การเปรียบเทียบค่า AMSE ที่ขนาดตัวอย่าง 60 ความแปรปรวนของความคลาดเคลื่อน 1

จากรูปที่ 4.3 จะเห็นได้ว่าที่ขนาดตัวอย่าง 60 ความแปรปรวนของความคลาดเคลื่อนเท่ากับ 1 ร้อยละข้อมูลสูญหาย 5, 10 และ 15 วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด รองลงมาคือ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 เมื่อร้อยละข้อมูลสูญหาย 20 วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 เป็นวิธีที่มีประสิทธิภาพมากที่สุด รองลงมาคือ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1

จากรูปที่ 4.3 จะเห็นแนวโน้มว่าทุกร้อยละข้อมูลสูญหาย วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด

จะเห็นแนวโน้มว่า เมื่อร้อยละข้อมูลสูญหายมีค่าเพิ่มมากขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยมีค่าเพิ่มมากขึ้นตามไปด้วย

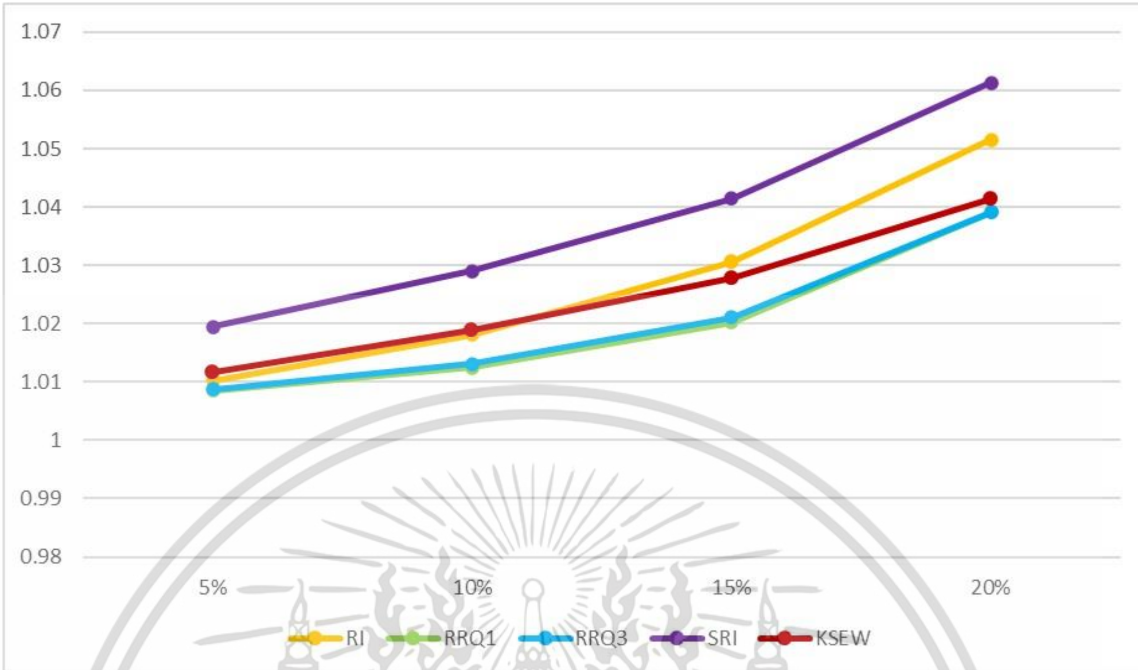


รูปที่ 4.4 การเปรียบเทียบค่า AMSE ที่ขนาดตัวอย่าง 80 ความแปรปรวนของความคลาดเคลื่อน 1

จากรูปที่ 4.4 จะเห็นได้ว่าที่ขนาดตัวอย่าง 80 ความแปรปรวนของความคลาดเคลื่อนเท่ากับ 1 ร้อยละข้อมูลสูญหาย 5, 10 และ 15 วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด รองลงมาคือ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 เมื่อร้อยละข้อมูลสูญหาย 20 วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 เป็นวิธีที่มีประสิทธิภาพมากที่สุด รองลงมาคือ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1

จากรูปที่ 4.4 จะเห็นแนวโน้มว่าทุกร้อยละข้อมูลสูญหาย วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด

จะเห็นแนวโน้มว่า เมื่อร้อยละข้อมูลสูญหายมีค่าเพิ่มมากขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยมีค่าเพิ่มมากขึ้นตามไปด้วย

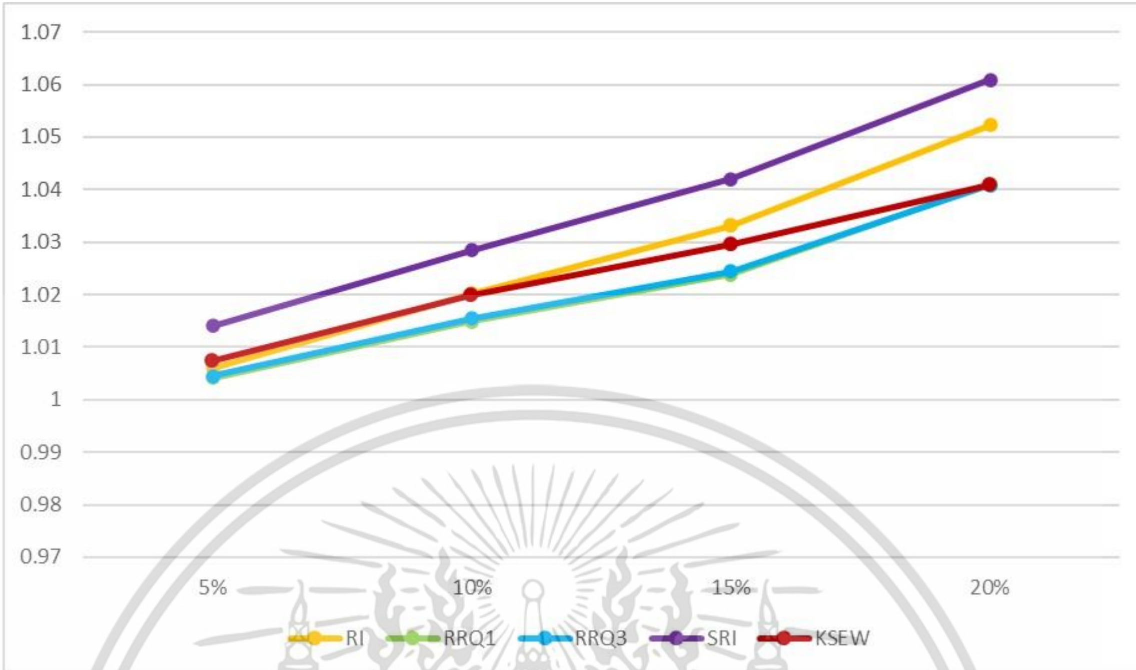


รูปที่ 4.5 การเปรียบเทียบค่า AMSE ที่ขนาดตัวอย่าง 100 ความแปรปรวนของความคลาดเคลื่อน 1

จากรูปที่ 4.5 จะเห็นได้ว่าที่ขนาดตัวอย่าง 100 ความแปรปรวนของความคลาดเคลื่อนเท่ากับ 1 ร้อยละข้อมูลสูญหาย 5, 10 และ 15 วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด รองลงมาคือ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 เมื่อร้อยละข้อมูลสูญหาย 20 วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 เป็นวิธีที่มีประสิทธิภาพมากที่สุด รองลงมาคือ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1

จากรูปที่ 4.5 จะเห็นว่าทุกร้อยละข้อมูลสูญหาย วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด

จะเห็นแนวโน้มว่า เมื่อร้อยละข้อมูลสูญหายมีค่าเพิ่มมากขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยมีค่าเพิ่มมากขึ้นตามไปด้วย



รูปที่ 4.6 การเปรียบเทียบค่า AMSE ที่ขนาดตัวอย่าง 120 ความแปรปรวนของความคลาดเคลื่อน 1

จากรูปที่ 4.6 จะเห็นได้ว่าที่ขนาดตัวอย่าง 120 ความแปรปรวนของความคลาดเคลื่อนเท่ากับ 1 ร้อยละข้อมูลสูญหาย 5, 10 และ 15 วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด รองลงมาคือ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 เมื่อร้อยละข้อมูลสูญหาย 20 วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 เป็นวิธีที่มีประสิทธิภาพมากที่สุด รองลงมาคือ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1

จากรูปที่ 4.6 จะเห็นว่าทุกร้อยละข้อมูลสูญหาย วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด

จะเห็นแนวโน้มว่า เมื่อร้อยละข้อมูลสูญหายมีค่าเพิ่มมากขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยมีค่าเพิ่มมากขึ้นตามไปด้วย

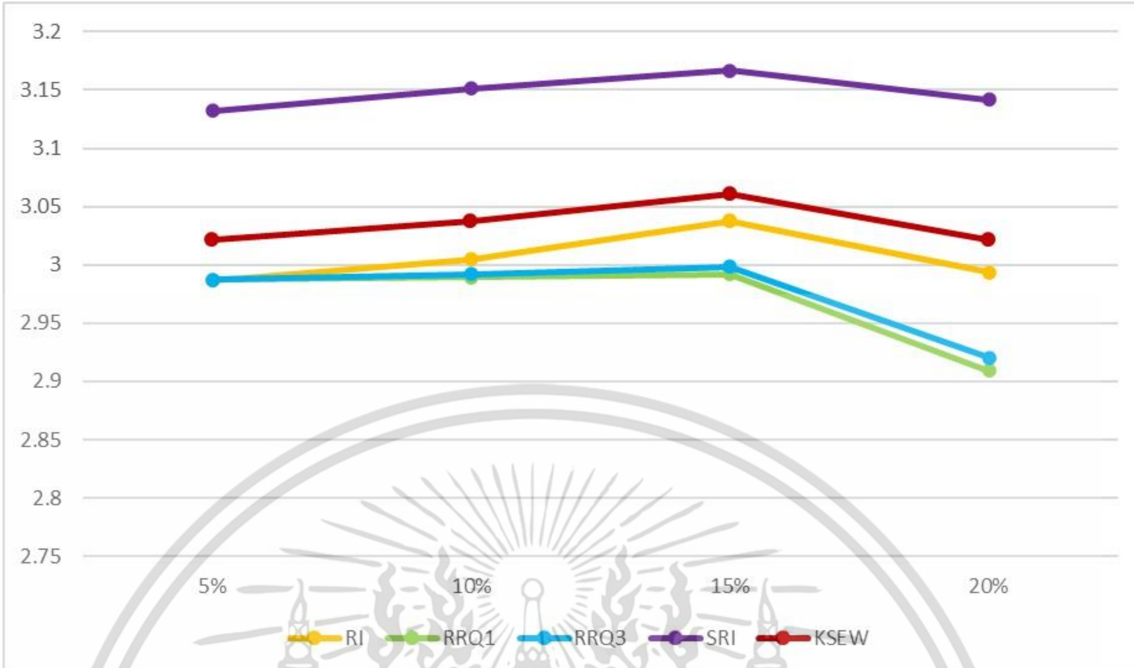
ตารางที่ 4.2 ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย เมื่อค่าความแปรปรวนของความคลาดเคลื่อนเท่ากับ 3

ขนาดตัวอย่าง	ค่าสูญหาย (%)	วิธีประมาณค่าสูญหาย				
		RI	RRQ1	RRQ3	SRI	KSEW
ตัวอย่างขนาดเล็ก						
20	5%	**2.9865**	2.9873	2.9871	3.132	3.0217
	10%	3.0049	**2.9892**	2.9918	3.1512	3.0374
	15%	3.0374	**2.9918**	2.9986	3.1665	3.0606
	20%	2.9933	**2.9091**	2.9205	3.1415	3.0217
40	5%	2.9929	**2.9893**	2.9899	3.0708	3.0126
	10%	3.0063	**2.9843**	2.9878	3.0856	3.0256
	15%	3.0708	**3.0213**	3.0286	3.1512	3.086
	20%	3.0373	**2.9457**	2.9577	3.1218	3.0493
ตัวอย่างขนาดปานกลาง						
60	5%	3.0138	**3.0084**	3.0094	3.0564	3.0211
	10%	3.0205	**2.9956**	2.9995	3.0728	3.0325
	15%	3.0142	**2.9591**	2.9671	3.066	3.0221
	20%	3.0503	**2.9514**	2.9645	3.1034	3.0523
80	5%	3.0226	**3.0167**	3.0177	3.0606	3.0315
	10%	3.0265	**3.0007**	3.0047	3.0618	3.0324
	15%	3.0137	**2.9578**	2.9658	3.0493	3.0159
	20%	3.0537	**2.9565**	2.9692	3.0924	3.0518
ตัวอย่างขนาดใหญ่						
100	5%	3.01	**3.0037**	3.0047	3.0388	3.016
	10%	3.0136	**2.9872**	2.9913	3.0447	3.0199
	15%	3.0275	**2.9702**	2.9784	3.0584	3.0297
	20%	3.0503	**2.9528**	2.9654	3.0799	3.0449
120	5%	2.9876	**2.9814**	2.9824	3.0136	2.994
	10%	3.0139	**2.9875**	2.9916	3.0392	3.0184
	15%	3.0172	**2.9611**	2.9691	3.0399	3.0154
	20%	3.0676	**2.9695**	2.9822	3.0921	3.0585

จากตารางที่ 4.2 เมื่อค่าความแปรปรวนของความคลาดเคลื่อนมีค่าเท่ากับ 3 ขนาดตัวอย่าง 20 ร้อยละข้อมูลสูญหาย 5 วิธีการถดถอยพหุคูณ เป็นวิธีที่มีประสิทธิภาพมากที่สุด ส่วนร้อยละข้อมูลสูญหาย 10, 15 และ 20 วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด

ขนาดตัวอย่าง 40, 60, 80, 100 และ 120 ทุกร้อยละข้อมูลสูญหาย วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด

จากตารางที่ 4.2 จะพบว่าตัวอย่างขนาดเล็ก ตัวอย่างขนาดปานกลางและตัวอย่างขนาดใหญ่ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด

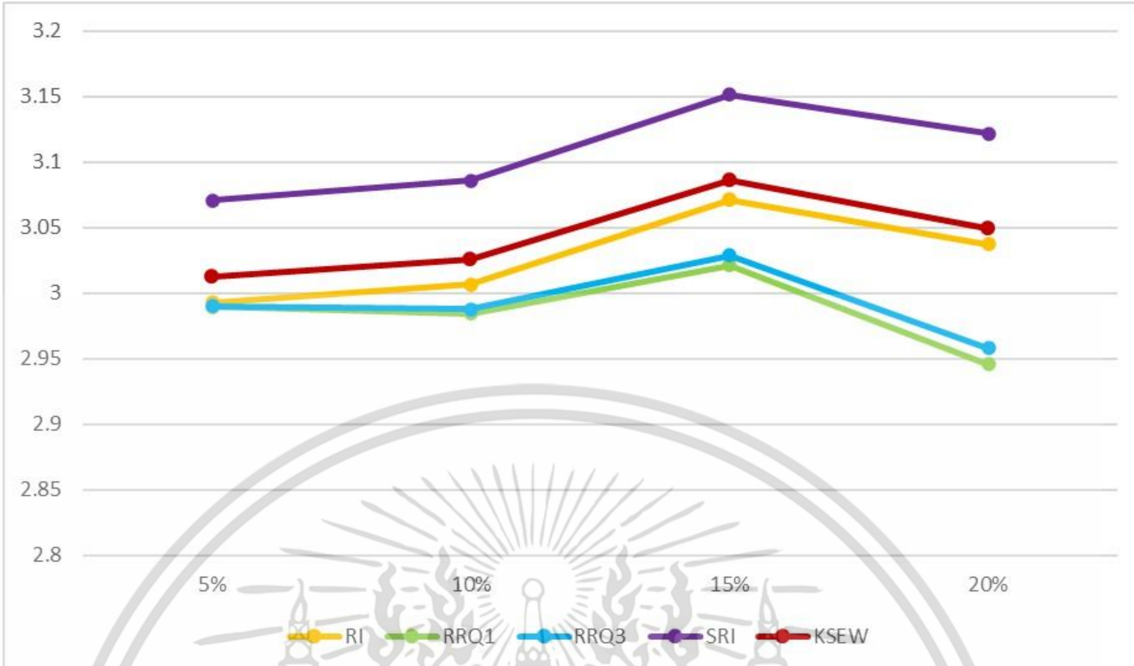


รูปที่ 4.7 การเปรียบเทียบค่า AMSE ที่ขนาดตัวอย่าง 20 ความแปรปรวนของความคลาดเคลื่อน 3

จากรูปที่ 4.7 จะเห็นได้ว่าที่ขนาดตัวอย่าง 20 ความแปรปรวนของความคลาดเคลื่อนเท่ากับ 3 ร้อยละข้อมูลสูญหาย 5 วิธีการถดถอยพหุคูณ เป็นวิธีที่มีประสิทธิภาพมากที่สุด รองลงมาคือ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 ส่วนร้อยละข้อมูลสูญหาย 10, 15 และ 20 วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด รองลงมาคือ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3

จากรูปที่ 4.7 จะเห็นว่าทุกร้อยละข้อมูลสูญหาย วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด

จะเห็นแนวโน้มว่า วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 และ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 เมื่อร้อยละข้อมูลสูญหายมีค่าเพิ่มมากขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยจะมีค่าลดลง ส่วนวิธีการถดถอยพหุคูณ วิธีการถดถอยสโตแคสติก และวิธีการถดถอยสโตแคสติกใกล้สุด k ตัวด้วยค่าถ่วงน้ำหนักที่เท่ากัน เมื่อร้อยละข้อมูลสูญหายมีค่าเพิ่มมากขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยมีค่าเพิ่มมากขึ้นตามไปด้วย

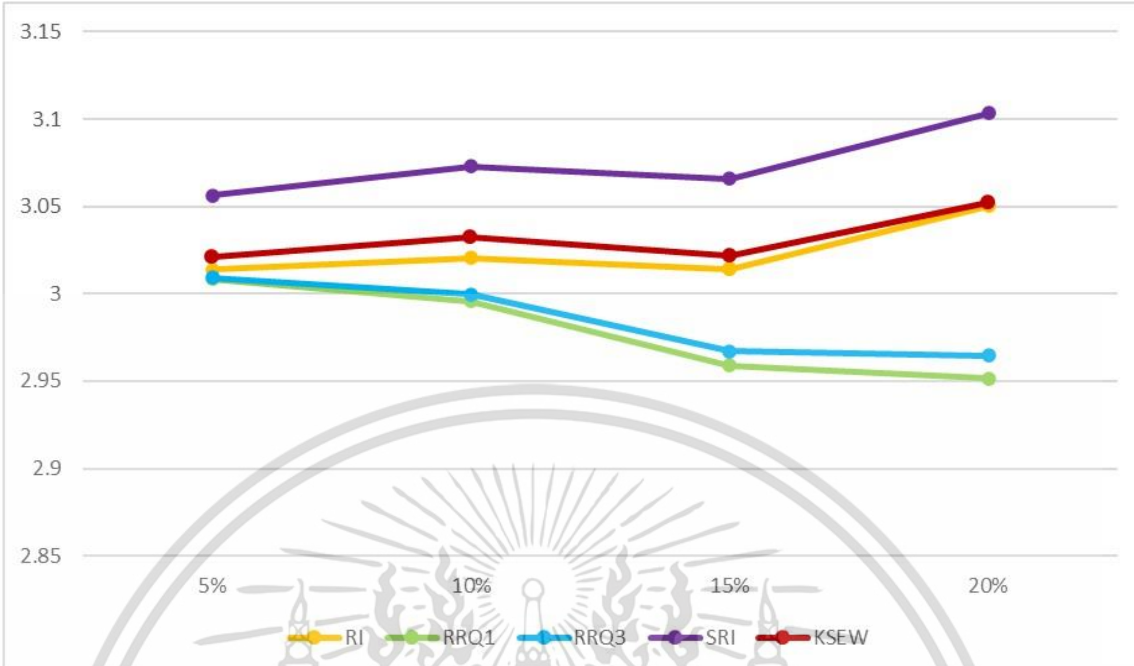


รูปที่ 4.8 การเปรียบเทียบค่า AMSE ที่ขนาดตัวอย่าง 40 ความแปรปรวนของความคลาดเคลื่อน 3

จากรูปที่ 4.8 จะเห็นได้ว่าที่ขนาดตัวอย่าง 40 ความแปรปรวนของความคลาดเคลื่อนเท่ากับ 3 ทุกร้อยละข้อมูลสูญหาย วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด รองลงมาคือ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3

จากรูปที่ 4.8 จะเห็นว่าทุกร้อยละข้อมูลสูญหาย วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด

จะเห็นแนวโน้มว่า วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 และ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 เมื่อร้อยละข้อมูลสูญหายมีค่าเพิ่มมากขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยจะมีค่าลดลง ส่วนวิธีการถดถอยพหุคูณ วิธีการถดถอยสโตแคสติก และวิธีการถดถอยสโตแคสติกใกล้สุด k ตัวด้วยค่าถ่วงน้ำหนักที่เท่ากัน เมื่อร้อยละข้อมูลสูญหายมีค่าเพิ่มมากขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยมีค่าเพิ่มมากขึ้นตามไปด้วย

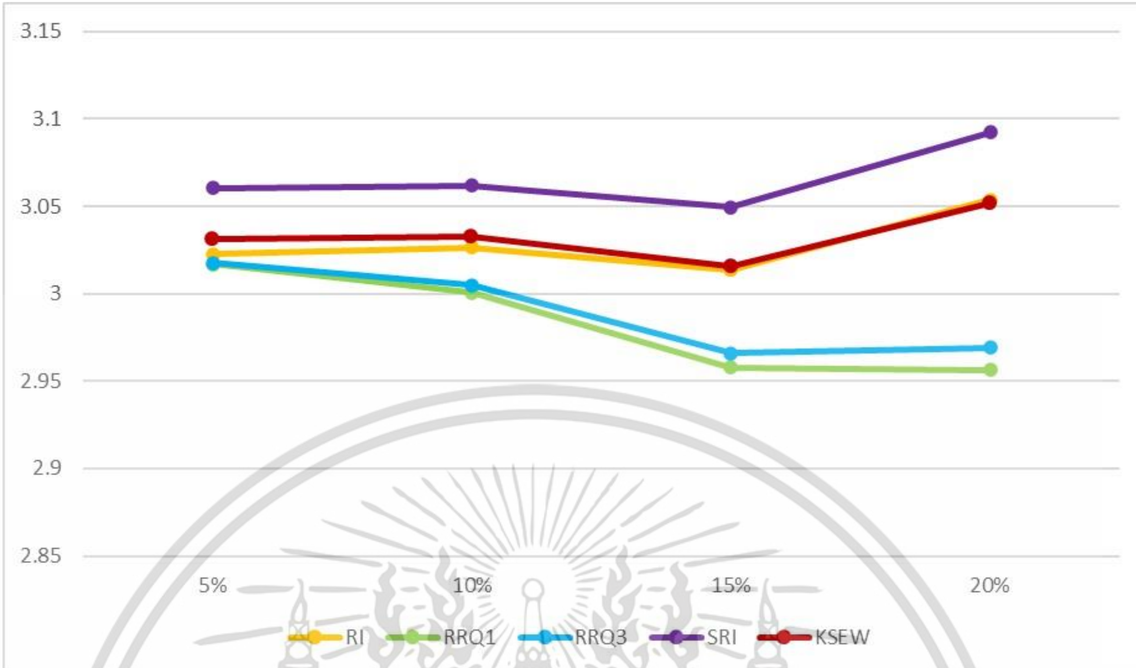


รูปที่ 4.9 การเปรียบเทียบค่า AMSE ที่ขนาดตัวอย่าง 60 ความแปรปรวนของความคลาดเคลื่อน 3

จากรูปที่ 4.9 จะเห็นได้ว่าที่ขนาดตัวอย่าง 60 ความแปรปรวนของความคลาดเคลื่อนเท่ากับ 3 ทุกร้อยละข้อมูลสูญหาย วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด รองลงมาคือ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3

จากรูปที่ 4.9 จะเห็นว่าทุกร้อยละข้อมูลสูญหาย วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด

จะเห็นแนวโน้มว่า ร้อยละข้อมูลสูญหายของวิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 และ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 เมื่อร้อยละข้อมูลสูญหายมีค่าเพิ่มมากขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยจะมีค่าลดลง ส่วนวิธีการถดถอยพหุคูณ วิธีการถดถอยสโตแคสติก และวิธีการถดถอยสโตแคสติกใกล้สุด k ตัวด้วยค่าถ่วงน้ำหนักที่เท่ากัน เมื่อร้อยละข้อมูลสูญหายมีค่าเพิ่มมากขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยจะมีค่าเพิ่มมากขึ้นตามไปด้วย

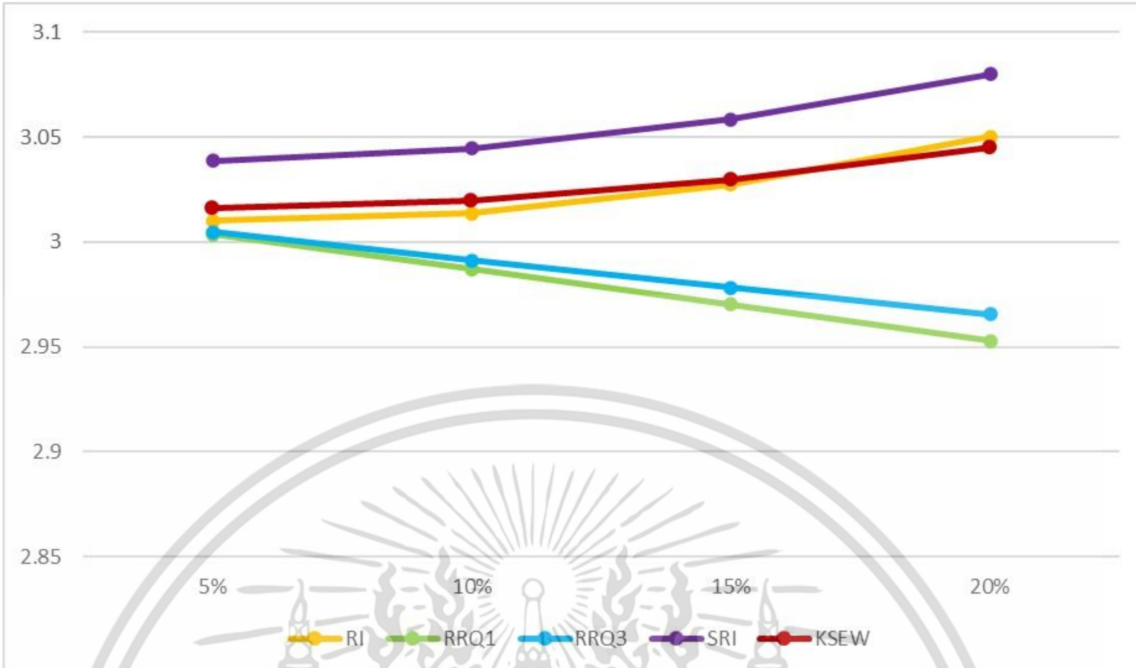


รูปที่ 4.10 การเปรียบเทียบค่า AMSE ที่ขนาดตัวอย่าง 80 ความแปรปรวนของความคลาดเคลื่อน 3

จากรูปที่ 4.10 จะเห็นได้ว่าที่ขนาดตัวอย่าง 80 ความแปรปรวนของความคลาดเคลื่อนเท่ากับ 3 ทุกร้อยละข้อมูลสูญหาย วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด รองลงมาคือ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3

จากรูปที่ 4.10 จะเห็นว่าทุกร้อยละข้อมูลสูญหาย วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด

จะเห็นแนวโน้มว่า ร้อยละข้อมูลสูญหายของวิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 และ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 เมื่อร้อยละข้อมูลสูญหายมีค่าเพิ่มมากขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยจะมีค่าลดลง ส่วนวิธีการถดถอยพหุคูณ วิธีการถดถอยสโตแคสติก และวิธีการถดถอยสโตแคสติกใกล้สุด k ตัวด้วยค่าถ่วงน้ำหนักที่เท่ากัน เมื่อร้อยละข้อมูลสูญหายมีค่าเพิ่มมากขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยจะมีค่าเพิ่มมากขึ้นตามไปด้วย

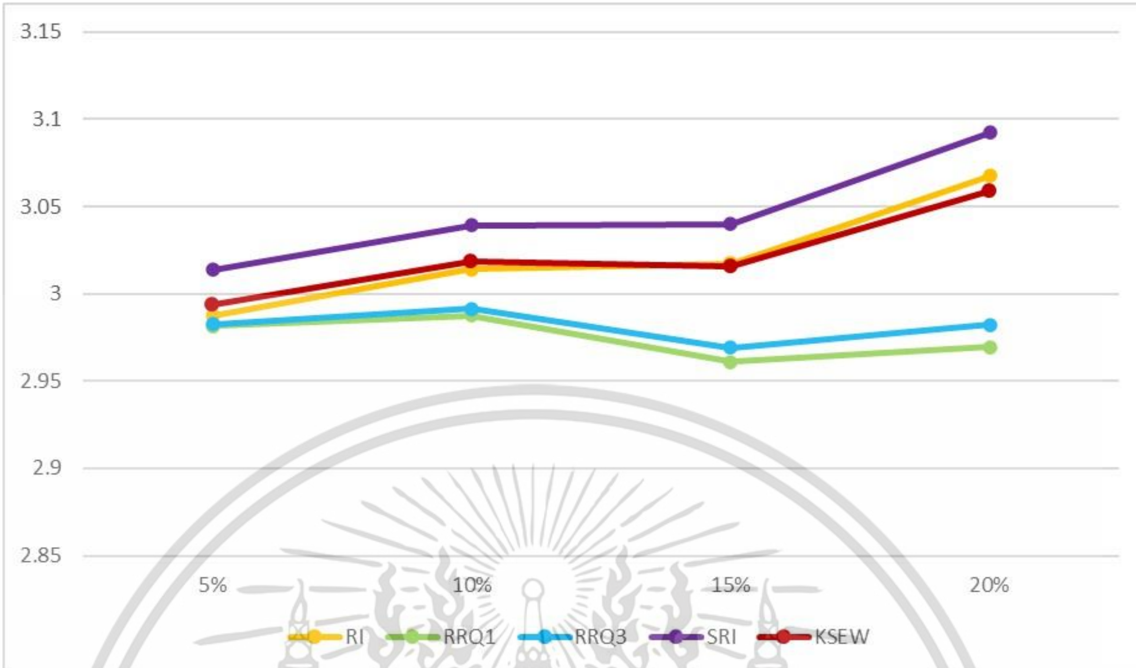


รูปที่ 4.11 การเปรียบเทียบค่า AMSE ที่ขนาดตัวอย่าง 100 ความแปรปรวนของความคลาดเคลื่อน 3

จากรูปที่ 4.11 จะเห็นได้ว่าที่ขนาดตัวอย่าง 100 ความแปรปรวนของความคลาดเคลื่อนเท่ากับ 3 ทุกร้อยละข้อมูลสูญหาย วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด รองลงมาคือ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3

จากรูปที่ 4.11 จะเห็นว่าเมื่อทุกร้อยละข้อมูลสูญหาย วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด

จะเห็นแนวโน้มว่า ร้อยละข้อมูลสูญหายของวิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 และ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 เมื่อร้อยละข้อมูลสูญหายมีค่าเพิ่มมากขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยจะมีค่าลดลง ส่วนวิธีการถดถอยพหุคูณ วิธีการถดถอยสโตแคสติก และวิธีการถดถอยสโตแคสติกใกล้สุด k ตัวด้วยค่าถ่วงน้ำหนักที่เท่ากัน เมื่อร้อยละข้อมูลสูญหายมีค่าเพิ่มมากขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยจะมีค่าเพิ่มมากขึ้นตามไปด้วย



รูปที่ 4.12 การเปรียบเทียบค่า AMSE ที่ขนาดตัวอย่าง 120 ความแปรปรวนของความคลาดเคลื่อน 3

จากรูปที่ 4.12 จะเห็นได้ว่าที่ขนาดตัวอย่าง 120 ความแปรปรวนของความคลาดเคลื่อนเท่ากับ 3 ทุกร้อยละข้อมูลสูญหาย วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด รองลงมาคือ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3

จากรูปที่ 4.12 จะเห็นว่าเมื่อทุกร้อยละข้อมูลสูญหาย วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด

จะเห็นแนวโน้มว่า ร้อยละข้อมูลสูญหายของวิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 และ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 เมื่อร้อยละข้อมูลสูญหายมีค่าเพิ่มมากขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยจะมีค่าลดลง ส่วนวิธีการถดถอยพหุคูณ วิธีการถดถอยสโตแคสติก และวิธีการถดถอยสโตแคสติกใกล้สุด k ตัวด้วยค่าถ่วงน้ำหนักที่เท่ากัน เมื่อร้อยละข้อมูลสูญหายมีค่าเพิ่มมากขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยจะมีค่าเพิ่มมากขึ้นตามไปด้วย

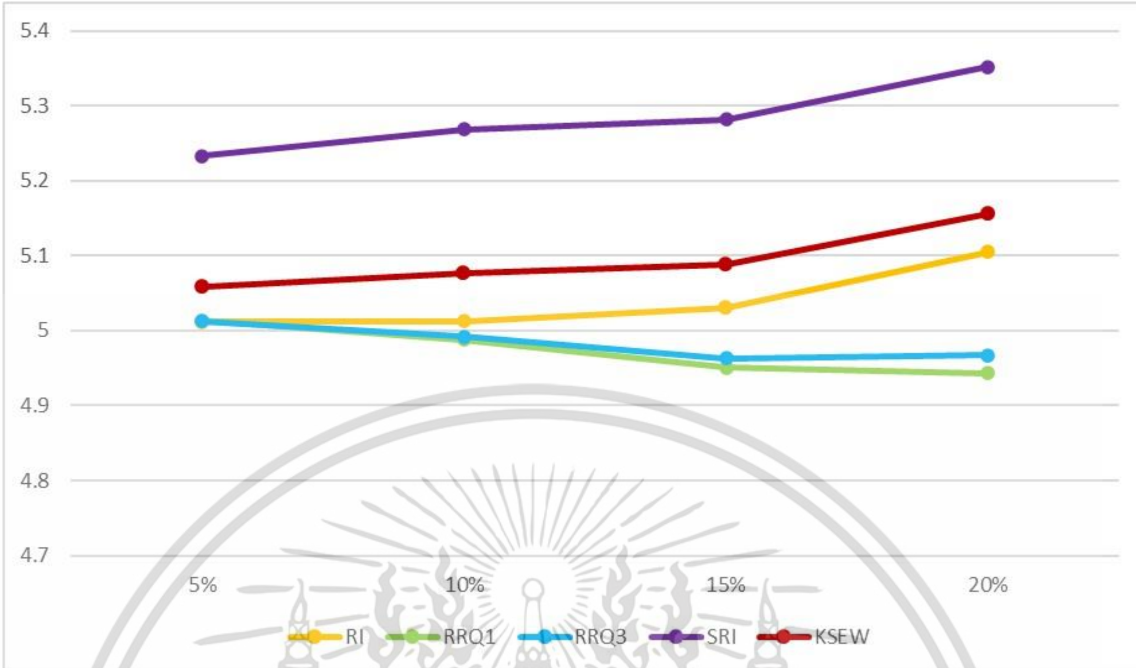
ตารางที่ 4.3 ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย เมื่อค่าความแปรปรวนของความคลาดเคลื่อนเท่ากับ 5

ขนาดตัวอย่าง	ค่าสูญหาย (%)	วิธีประมาณค่าสูญหาย				
		RI	RRQ1	RRQ3	SRI	KSEW
ตัวอย่างขนาดเล็ก						
20	5%	**5.0121**	5.0133	5.0131	5.2331	5.0587
	10%	5.0124	**4.9882**	4.9921	5.2691	5.0764
	15%	5.0311	**4.9505**	4.9631	5.2817	5.0886
	20%	5.1054	**4.9433**	4.967	5.3517	5.1568
40	5%	5.0045	**4.998**	4.999	5.1221	5.0309
	10%	5.0017	**4.9624**	4.9688	5.1133	5.0248
	15%	5.0587	**4.9601**	4.975	5.1751	5.081
	20%	5.0536	**4.8718**	4.8975	5.1798	5.0723
ตัวอย่างขนาดปานกลาง						
60	5%	4.9877	**4.9782**	4.9799	5.0749	5.0097
	10%	4.9945	**4.9505**	4.9576	5.0818	5.0153
	15%	5.0390	**4.9384**	4.9535	5.1323	5.0585
	20%	5.036	**4.8517**	4.8774	5.115	5.043
80	5%	5.0096	**4.9989**	5.0007	5.077	5.027
	10%	5.0201	**4.973**	4.9805	5.0808	5.0333
	15%	5.017	**4.913**	4.9287	5.077	5.0256
	20%	5.0565	**4.8676**	4.8939	5.1126	5.0568
ตัวอย่างขนาดใหญ่						
100	5%	4.9873	**4.9767**	4.9785	5.0429	5.0018
	10%	5.0039	**4.9574**	4.9647	5.0552	5.0149
	15%	5.0576	**4.9521**	4.9679	5.1088	5.0648
	20%	5.0569	**4.8755**	4.9006	5.1116	5.0568
120	5%	5.0178	**5.0062**	5.0081	5.061	5.0283
	10%	5.0341	**4.9862**	4.9938	5.0766	5.0434
	15%	5.0188	**4.9138**	4.9295	5.0583	5.022
	20%	5.068	**4.8842**	4.9097	5.113	5.067

จากตารางที่ 4.3 เมื่อค่าความแปรปรวนของความคลาดเคลื่อนมีค่าเท่ากับ 5 ขนาดตัวอย่าง 20 ร้อยละข้อมูลสูญหาย 5 วิธีการถดถอยพหุคูณ เป็นวิธีที่มีประสิทธิภาพมากที่สุด ส่วนร้อยละข้อมูลสูญหาย 10, 15 และ 20 วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด

ขนาดตัวอย่าง 40, 60, 80, 100 และ 120 ทุกร้อยละข้อมูลสูญหาย วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด

จากตารางที่ 4.3 จะพบว่าทุกขนาดตัวอย่าง วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด

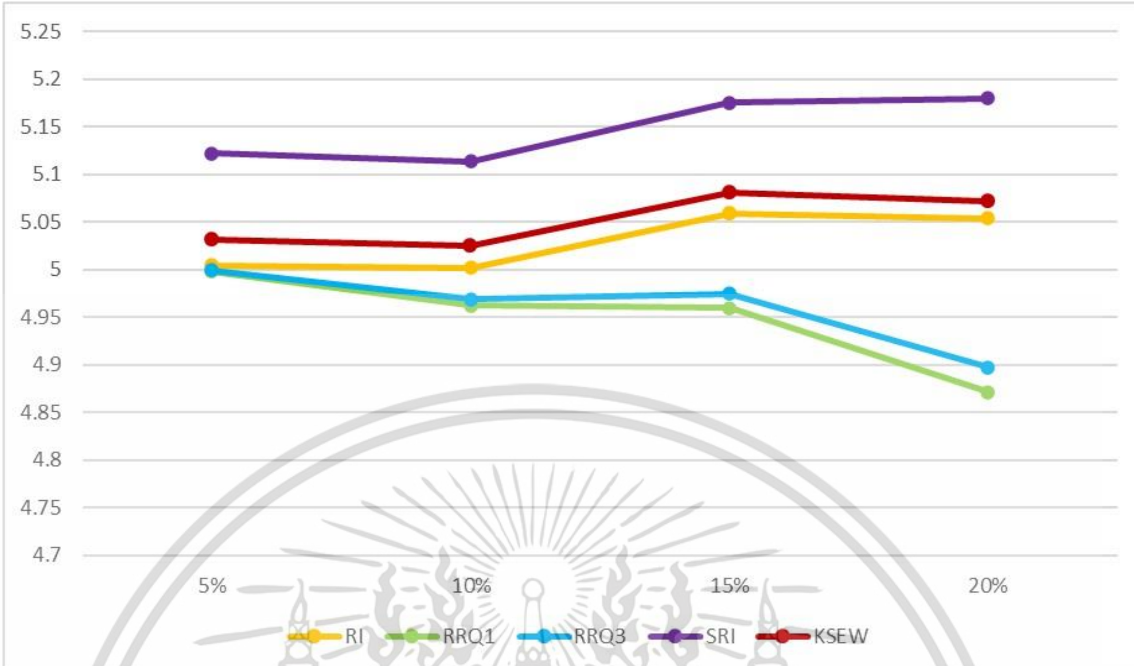


รูปที่ 4.13 การเปรียบเทียบค่า AMSE ที่ขนาดตัวอย่าง 20 ความแปรปรวนของความคลาดเคลื่อน 5

จากรูปที่ 4.13 จะเห็นได้ว่าที่ขนาดตัวอย่าง 20 ความแปรปรวนของความคลาดเคลื่อนเท่ากับ 5 ร้อยละข้อมูลสูญหาย 5 วิธีการถดถอยพหุคูณ เป็นวิธีที่มีประสิทธิภาพมากที่สุด รองลงมาคือ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 ส่วนร้อยละข้อมูลสูญหาย 10, 15 และ 20 วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด รองลงมาคือ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3

จากรูปที่ 4.13 จะเห็นว่าทุกร้อยละข้อมูลสูญหาย วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด

จะเห็นแนวโน้มว่า ร้อยละข้อมูลสูญหายของวิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 และ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 เมื่อร้อยละข้อมูลสูญหายมีค่าเพิ่มมากขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยจะมีค่าลดลง ส่วนวิธีการถดถอยพหุคูณ วิธีการถดถอยสโตแคสติก และวิธีการถดถอยสโตแคสติกใกล้สุด k ตัวด้วยค่าถ่วงน้ำหนักที่เท่ากัน เมื่อร้อยละข้อมูลสูญหายมีค่าเพิ่มมากขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยจะมีค่าเพิ่มมากขึ้นตามไปด้วย

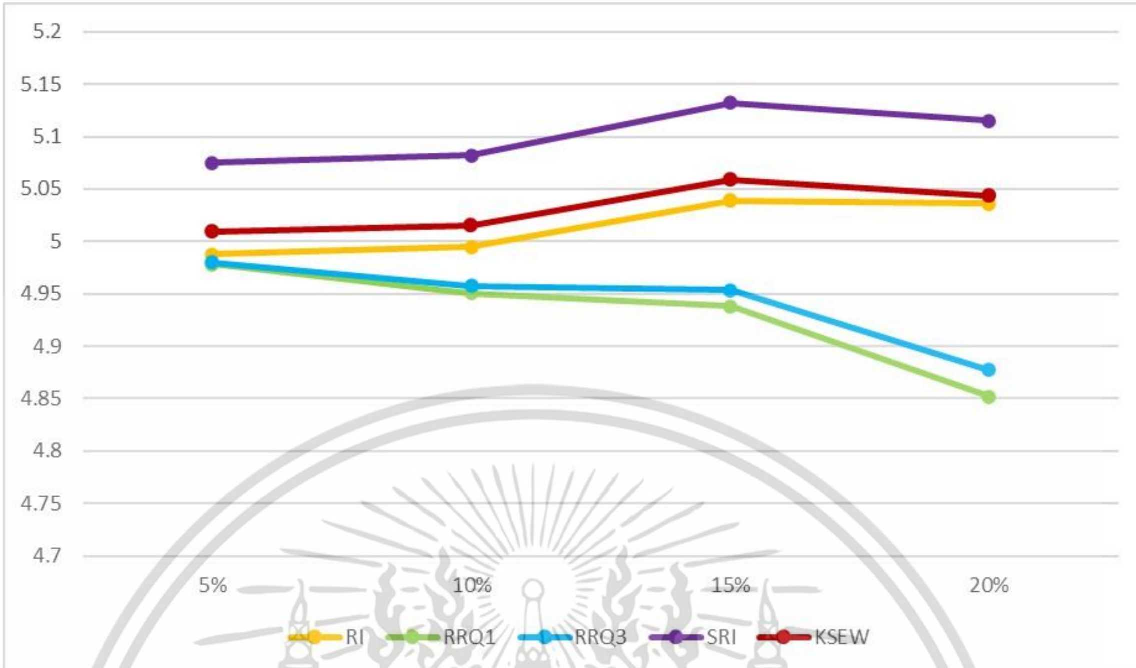


รูปที่ 4.14 การเปรียบเทียบค่า AMSE ที่ขนาดตัวอย่าง 40 ความแปรปรวนของความคลาดเคลื่อน 5

จากรูปที่ 4.14 จะเห็นได้ว่าที่ขนาดตัวอย่าง 40 ความแปรปรวนของความคลาดเคลื่อนเท่ากับ 5 ทุกร้อยละข้อมูลสูญหาย วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด รองลงมาคือ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3

จากรูปที่ 4.14 จะเห็นว่าทุกร้อยละข้อมูลสูญหาย วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด

จะเห็นแนวโน้มว่า ร้อยละข้อมูลสูญหายของวิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 และ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 เมื่อร้อยละข้อมูลสูญหายมีค่าเพิ่มมากขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยจะมีค่าลดลง ส่วนวิธีการถดถอยพหุคูณ วิธีการถดถอยสโตแคสติก และวิธีการถดถอยสโตแคสติกใกล้สุด k ตัวด้วยค่าถ่วงน้ำหนักที่เท่ากัน เมื่อร้อยละข้อมูลสูญหายมีค่าเพิ่มมากขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยจะมีค่าเพิ่มมากขึ้นตามไปด้วย

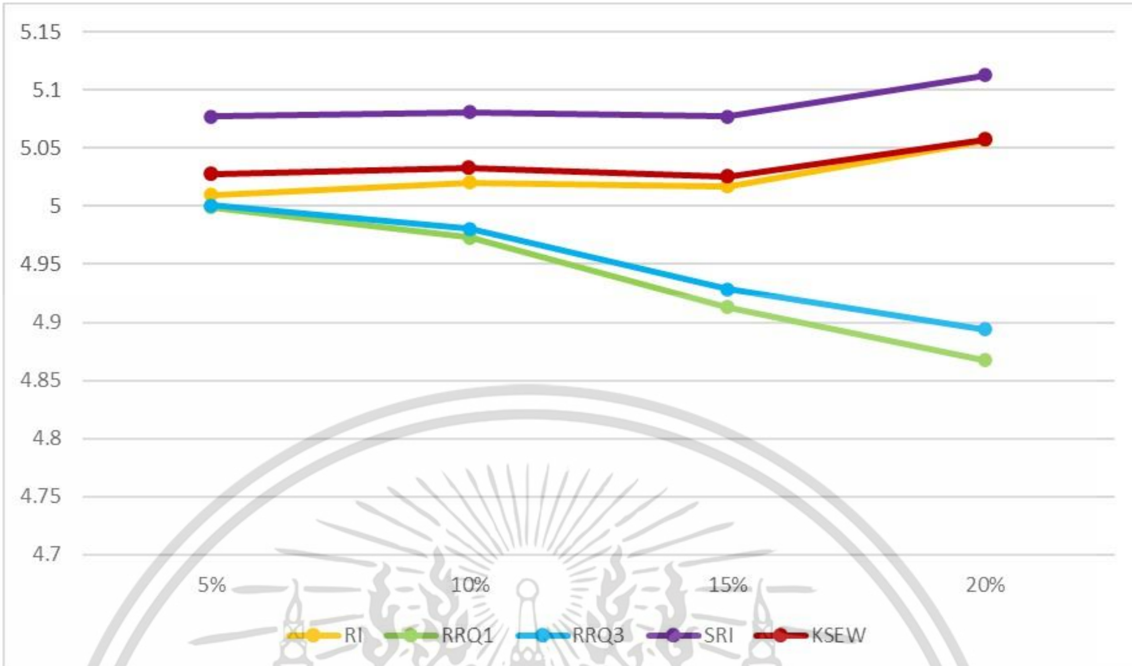


รูปที่ 4.15 การเปรียบเทียบค่า AMSE ที่ขนาดตัวอย่าง 60 ความแปรปรวนของความคลาดเคลื่อน 5

จากรูปที่ 4.15 จะเห็นได้ว่าที่ขนาดตัวอย่าง 60 ความแปรปรวนของความคลาดเคลื่อนเท่ากับ 5 ทุกร้อยละข้อมูลสูญหาย วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด รองลงมาคือ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3

จากรูปที่ 4.15 จะเห็นว่าทุกร้อยละข้อมูลสูญหาย วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด

จะเห็นแนวโน้มว่า ร้อยละข้อมูลสูญหายของวิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 และ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 เมื่อร้อยละข้อมูลสูญหายมีค่าเพิ่มมากขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยจะมีค่าลดลง ส่วนวิธีการถดถอยพหุคูณ วิธีการถดถอยสโตแคสติก และวิธีการถดถอยสโตแคสติกใกล้สุด k ตัวด้วยค่าถ่วงน้ำหนักที่เท่ากัน เมื่อร้อยละข้อมูลสูญหายมีค่าเพิ่มมากขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยจะมีค่าเพิ่มมากขึ้นตามไปด้วย

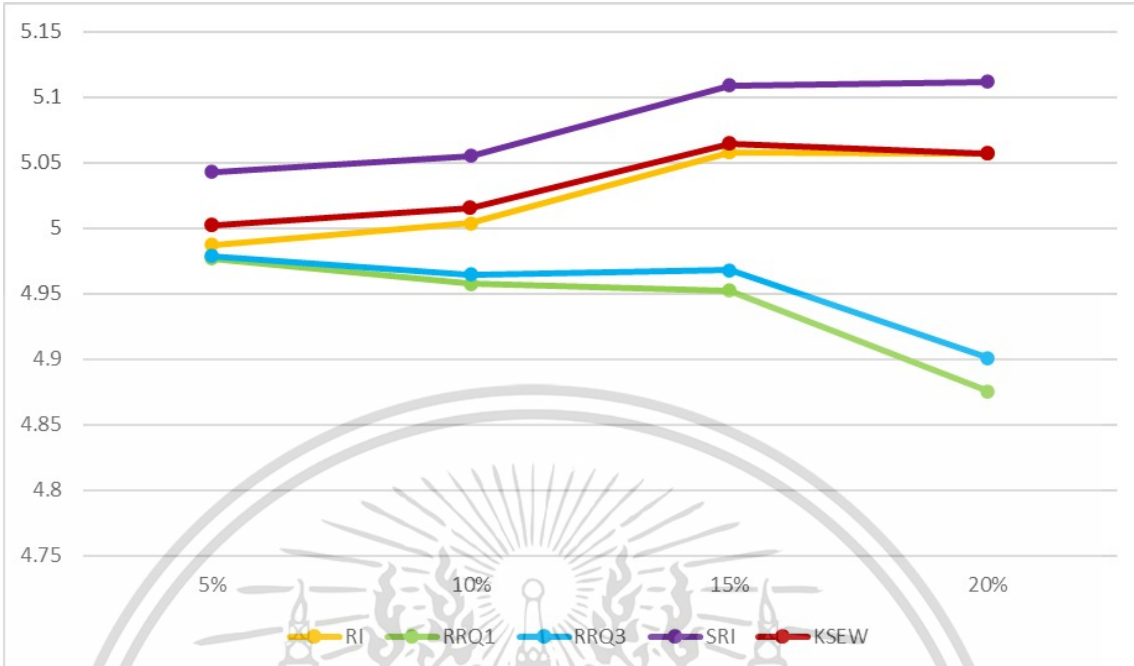


รูปที่ 4.16 การเปรียบเทียบค่า AMSE ที่ขนาดตัวอย่าง 80 ความแปรปรวนของความคลาดเคลื่อน 5

จากรูปที่ 4.16 จะเห็นได้ว่าที่ขนาดตัวอย่าง 80 ความแปรปรวนของความคลาดเคลื่อนเท่ากับ 5 ทุกร้อยละข้อมูลสูญหาย วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด รองลงมาคือ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3

จากรูปที่ 4.16 จะเห็นว่าทุกร้อยละข้อมูลสูญหาย วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด

จะเห็นแนวโน้มว่า ร้อยละข้อมูลสูญหายของวิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 และ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 เมื่อร้อยละข้อมูลสูญหายมีค่าเพิ่มมากขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยจะมีค่าลดลง ส่วนวิธีการถดถอยพหุคูณ วิธีการถดถอยสโตแคสติก และวิธีการถดถอยสโตแคสติกใกล้สุด k ตัวด้วยค่าถ่วงน้ำหนักที่เท่ากัน เมื่อร้อยละข้อมูลสูญหายมีค่าเพิ่มมากขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยจะมีค่าเพิ่มมากขึ้นตามไปด้วย

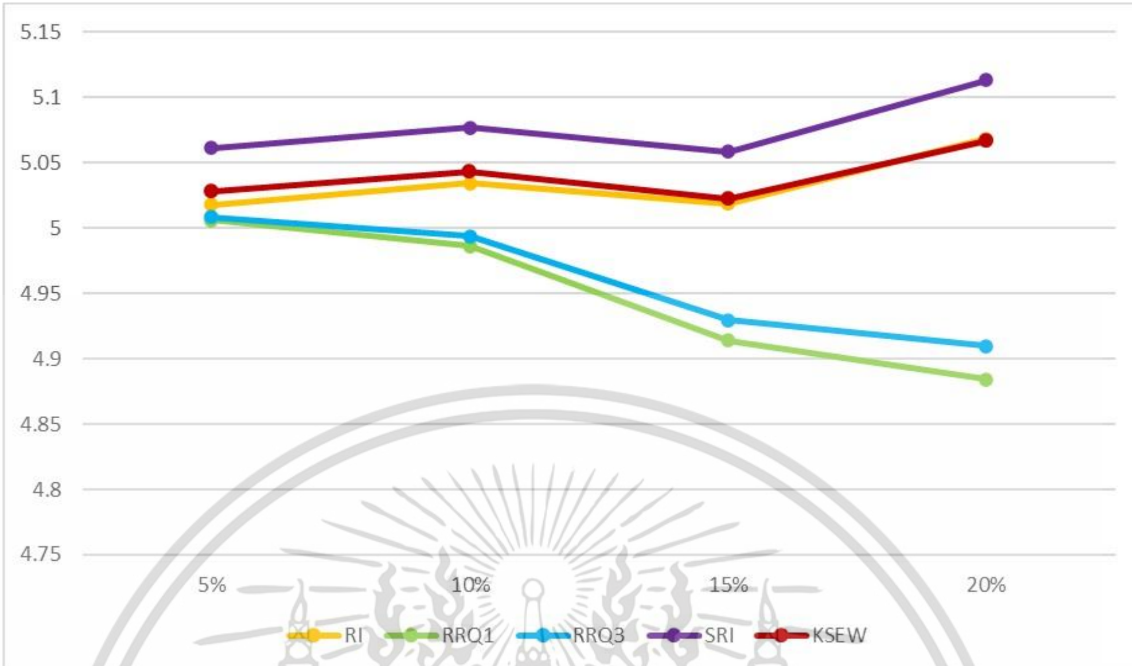


รูปที่ 4.17 การเปรียบเทียบค่า AMSE ที่ขนาดตัวอย่าง 100 ความแปรปรวนของความคลาดเคลื่อน 5

จากรูปที่ 4.17 จะเห็นได้ว่าที่ขนาดตัวอย่าง 100 ความแปรปรวนของความคลาดเคลื่อนเท่ากับ 5 ทุกร้อยละข้อมูลสูญหาย วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด รองลงมาคือ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3

จากรูปที่ 4.17 จะเห็นว่าทุกร้อยละข้อมูลสูญหาย วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด

จะเห็นแนวโน้มว่า ร้อยละข้อมูลสูญหายของวิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 และ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 เมื่อร้อยละข้อมูลสูญหายมีค่าเพิ่มมากขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยจะมีค่าลดลง ส่วนวิธีการถดถอยพหุคูณ วิธีการถดถอยสโตแคสติก และวิธีการถดถอยสโตแคสติกใกล้สุด k ตัวด้วยค่าถ่วงน้ำหนักที่เท่ากัน เมื่อร้อยละข้อมูลสูญหายมีค่าเพิ่มมากขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยจะมีค่าเพิ่มมากขึ้นตามไปด้วย



รูปที่ 4.18 การเปรียบเทียบค่า AMSE ที่ขนาดตัวอย่าง 120 ความแปรปรวนของความคลาดเคลื่อน 5

จากรูปที่ 4.18 จะเห็นได้ว่าที่ขนาดตัวอย่าง 120 ความแปรปรวนของความคลาดเคลื่อนเท่ากับ 5 ทุกร้อยละข้อมูลสูญหาย วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด รองลงมาคือ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3

จากรูปที่ 4.18 จะเห็นว่าทุกร้อยละข้อมูลสูญหาย วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด

จะเห็นแนวโน้มว่า ร้อยละข้อมูลสูญหายของวิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 และ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 เมื่อร้อยละข้อมูลสูญหายมีค่าเพิ่มมากขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยจะมีค่าลดลง ส่วนวิธีการถดถอยพหุคูณ วิธีการถดถอยสโตแคสติก และวิธีการถดถอยสโตแคสติกใกล้สุด k ตัวด้วยค่าถ่วงน้ำหนักที่เท่ากัน เมื่อร้อยละข้อมูลสูญหายมีค่าเพิ่มมากขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยจะมีค่าเพิ่มมากขึ้นตามไปด้วย

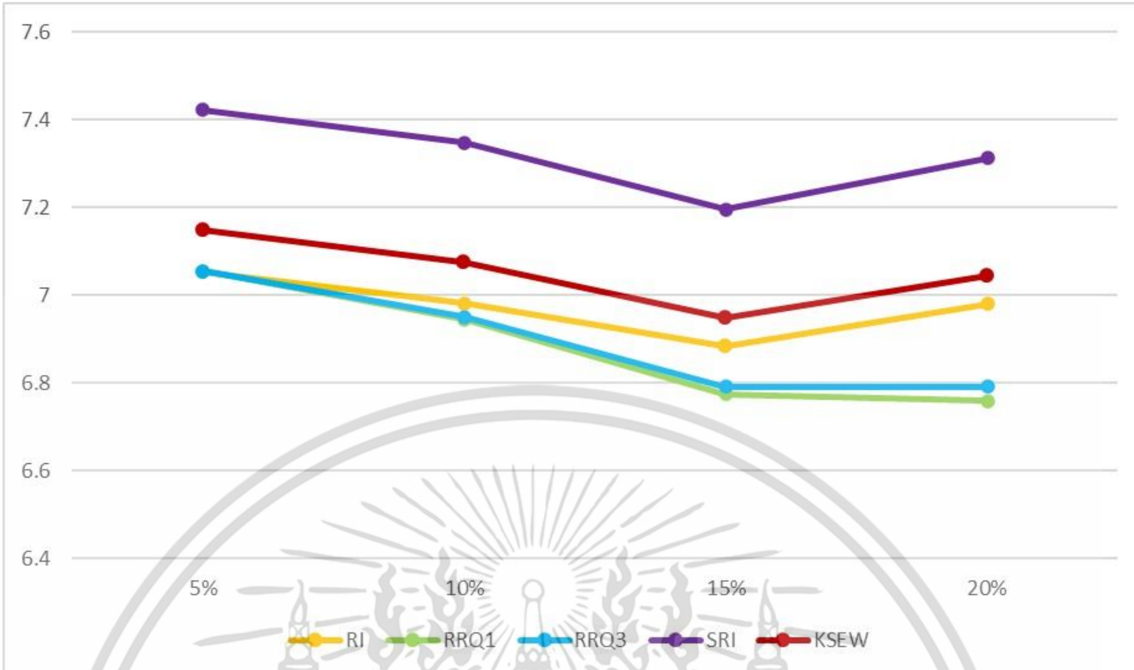
ตารางที่ 4.4 ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย เมื่อค่าความแปรปรวนของความคลาดเคลื่อนเท่ากับ 7

ขนาดตัวอย่าง	ค่าสูญหาย (%)	วิธีประมาณค่าสูญหาย				
		RI	RRQ1	RRQ3	SRI	KSEW
ตัวอย่างขนาดเล็ก						
20	5%	**7.0537**	7.0545	7.0543	7.4233	7.1493
	10%	6.982	**6.9451**	6.9512	7.3478	7.0758
	15%	6.8847	**6.7743**	6.7921	7.1956	6.9495
	20%	6.9796	**6.7587**	6.7919	7.3122	7.0455
40	5%	7.0383	**7.029**	7.0306	7.2063	7.077
	10%	6.9464	**6.8913**	6.9003	7.1249	6.9896
	15%	6.9446	**6.8062**	6.8274	7.107	6.9753
	20%	7.0798	**6.8210**	6.8582	7.2586	7.1133
ตัวอย่างขนาดปานกลาง						
60	5%	7.0678	**7.0543**	7.0566	7.1828	7.095
	10%	7.0654	**7.0022**	7.0124	7.194	7.0993
	15%	7.0421	**6.8942**	6.9168	7.1637	7.0684
	20%	7.1013	**6.8282**	6.8673	7.211	7.1167
80	5%	7.0106	**6.9964**	6.9988	7.0923	7.0294
	10%	7.0119	**6.9479**	6.9581	7.1051	7.0348
	15%	7.0156	**6.8668**	6.8895	7.103	7.031
	20%	7.0215	**6.752**	6.7905	7.1113	7.0321
ตัวอย่างขนาดใหญ่						
100	5%	7.032	**7.0167**	7.0193	7.1055	7.0509
	10%	7.1032	**7.0355**	7.0463	7.1722	7.118
	15%	7.0399	**6.8828**	6.9067	7.1057	7.05
	20%	7.0619	**6.7967**	6.8342	7.1316	7.0649
120	5%	7.0087	**6.9932**	6.9958	7.0645	7.0214
	10%	7.0122	**6.944**	6.9549	7.0737	7.026
	15%	7.0442	**6.89**	6.9134	7.1074	7.0539
	20%	7.0807	**6.808**	6.8467	7.1366	7.0801

จากตารางที่ 4.4 เมื่อค่าความแปรปรวนของความคลาดเคลื่อนมีค่าเท่ากับ 7 ขนาดตัวอย่าง 20 ร้อยละข้อมูลสูญหาย 5 วิธีการถดถอยพหุคูณ เป็นวิธีที่มีประสิทธิภาพมากที่สุด ส่วนร้อยละข้อมูลสูญหาย 10, 15 และ 20 วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด

ขนาดตัวอย่าง 40, 60, 80, 100 และ 120 ทุกร้อยละข้อมูลสูญหาย วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด

จากตารางที่ 4.4 จะพบว่าทุกขนาดตัวอย่าง วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 (RRQ1) เป็นวิธีที่มีประสิทธิภาพมากที่สุด

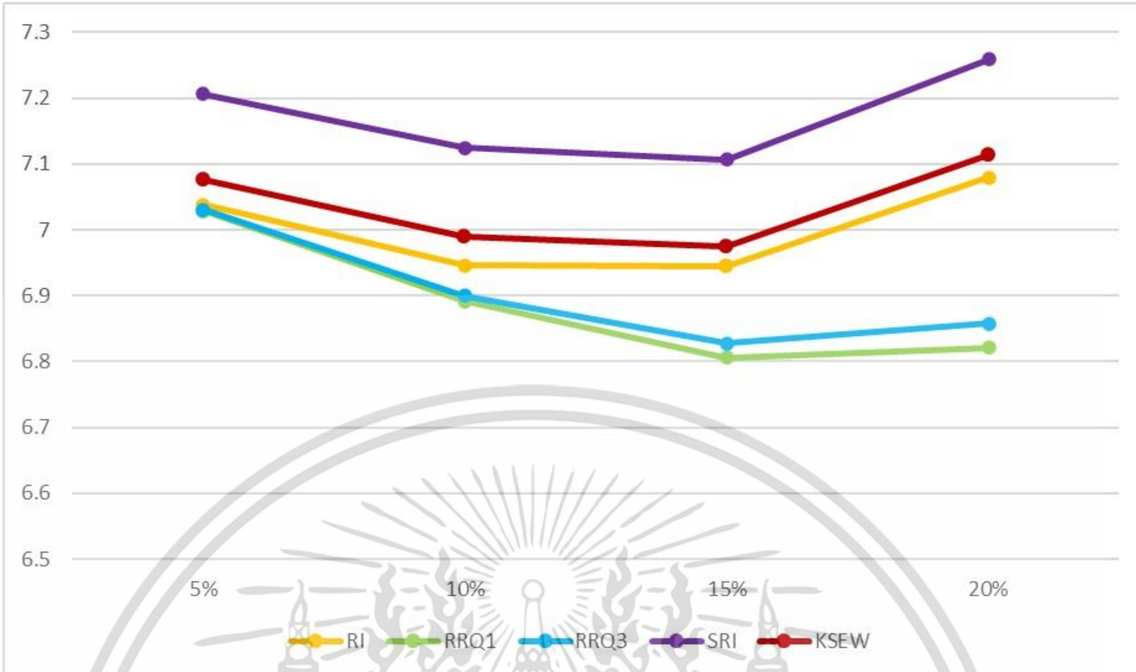


รูปที่ 4.19 การเปรียบเทียบค่า AMSE ที่ขนาดตัวอย่าง 20 ความแปรปรวนของความคลาดเคลื่อน 7

จากรูปที่ 4.19 จะเห็นได้ว่าที่ขนาดตัวอย่าง 20 ความแปรปรวนของความคลาดเคลื่อนเท่ากับ 7 ร้อยละข้อมูลสูญหาย 5 วิธีการถดถอยพหุคูณ เป็นวิธีที่มีประสิทธิภาพมากที่สุด รองลงมาคือ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 ส่วนร้อยละข้อมูลสูญหาย 10, 15 และ 20 วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด รองลงมาคือ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3

จากรูปที่ 4.19 จะเห็นว่าทุกร้อยละข้อมูลสูญหาย วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด

จะเห็นแนวโน้มว่า เมื่อร้อยละข้อมูลสูญหายมีค่าลดลง ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยมีค่าลดลงตามไปด้วย

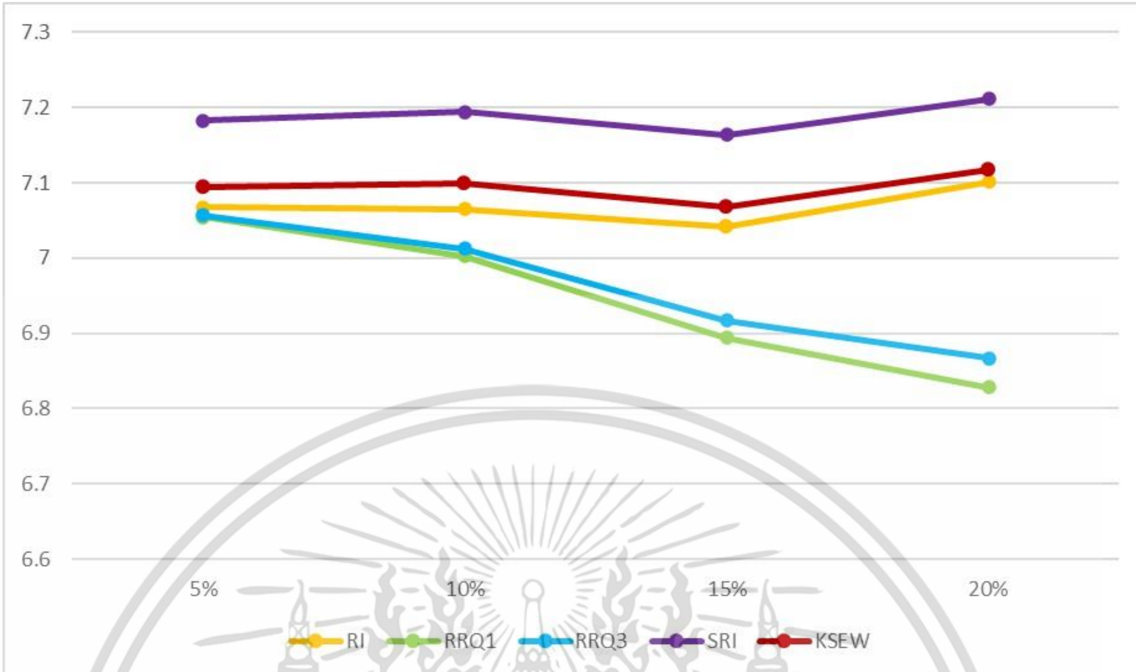


รูปที่ 4.20 การเปรียบเทียบค่า AMSE ที่ขนาดตัวอย่าง 40 ความแปรปรวนของความคลาดเคลื่อน 7

จากรูปที่ 4.20 จะเห็นได้ว่าที่ขนาดตัวอย่าง 40 ความแปรปรวนของความคลาดเคลื่อนเท่ากับ 7 ทุกร้อยละข้อมูลสูญหาย วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด รองลงมาคือ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3

จากรูปที่ 4.20 จะเห็นว่าทุกร้อยละข้อมูลสูญหาย วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด

จะเห็นแนวโน้มว่า ร้อยละข้อมูลสูญหายของวิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 และ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 เมื่อร้อยละข้อมูลสูญหายมีค่าเพิ่มมากขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยจะมีค่าลดลง ส่วนวิธีการถดถอยพหุคูณ วิธีการถดถอยสโตแคสติก และวิธีการถดถอยสโตแคสติกใกล้สุด k ตัวด้วยค่าถ่วงน้ำหนักที่เท่ากัน เมื่อร้อยละข้อมูลสูญหายมีค่าเพิ่มมากขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยจะมีค่าเพิ่มมากขึ้นตามไปด้วย

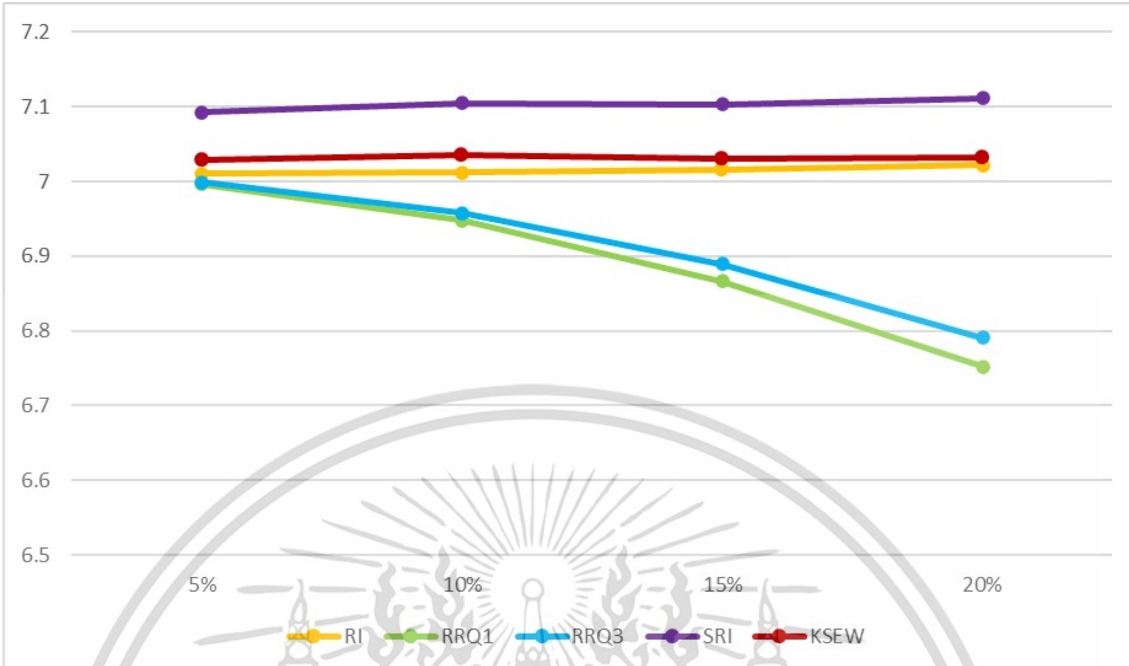


รูปที่ 4.21 การเปรียบเทียบค่า AMSE ที่ขนาดตัวอย่าง 60 ความแปรปรวนของความคลาดเคลื่อน 7

จากรูปที่ 4.21 จะเห็นได้ว่าที่ขนาดตัวอย่าง 60 ความแปรปรวนของความคลาดเคลื่อนเท่ากับ 7 ทุกร้อยละข้อมูลสูญหาย วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด รองลงมาคือ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3

จากรูปที่ 4.21 จะเห็นว่าทุกร้อยละข้อมูลสูญหาย วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด

จะเห็นแนวโน้มว่า ร้อยละข้อมูลสูญหายของวิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 และ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 เมื่อร้อยละข้อมูลสูญหายมีค่าเพิ่มมากขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยจะมีค่าลดลง ส่วนวิธีการถดถอยพหุคูณ วิธีการถดถอยสโตแคสติก และวิธีการถดถอยสโตแคสติกใกล้สุด k ตัวด้วยค่าถ่วงน้ำหนักที่เท่ากัน เมื่อร้อยละข้อมูลสูญหายมีค่าเพิ่มมากขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยจะมีค่าเพิ่มมากขึ้นตามไปด้วย

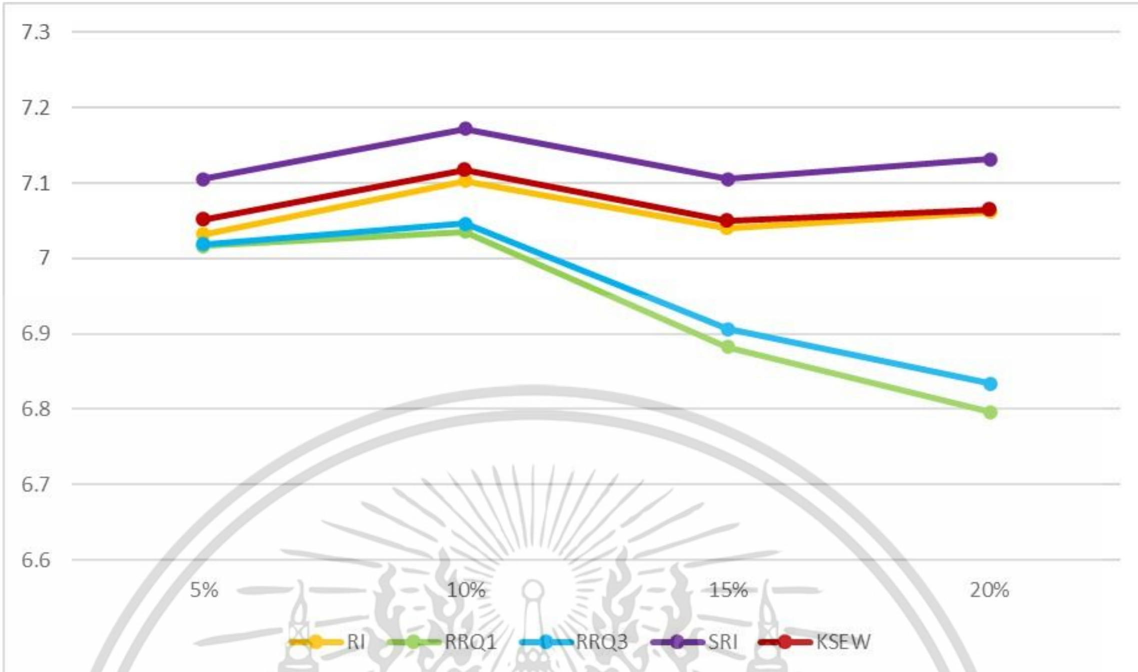


รูปที่ 4.22 การเปรียบเทียบค่า AMSE ที่ขนาดตัวอย่าง 80 ความแปรปรวนของความคลาดเคลื่อน 7

จากรูปที่ 4.22 จะเห็นได้ว่าที่ขนาดตัวอย่าง 80 ความแปรปรวนของความคลาดเคลื่อนเท่ากับ 7 ทุกร้อยละข้อมูลสูญหาย วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด รองลงมาคือ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3

จากรูปที่ 4.22 จะเห็นว่าทุกร้อยละข้อมูลสูญหาย วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด

จะเห็นแนวโน้มว่า ร้อยละข้อมูลสูญหายของวิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 และ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 เมื่อร้อยละข้อมูลสูญหายมีค่าเพิ่มมากขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยจะมีค่าลดลง ส่วนวิธีการถดถอยพหุคูณ วิธีการถดถอยสโตแคสติก และวิธีการถดถอยสโตแคสติกใกล้สุด k ตัวด้วยค่าถ่วงน้ำหนักที่เท่ากัน เมื่อร้อยละข้อมูลสูญหายมีค่าเพิ่มมากขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยจะมีค่าเพิ่มมากขึ้นตามไปด้วย

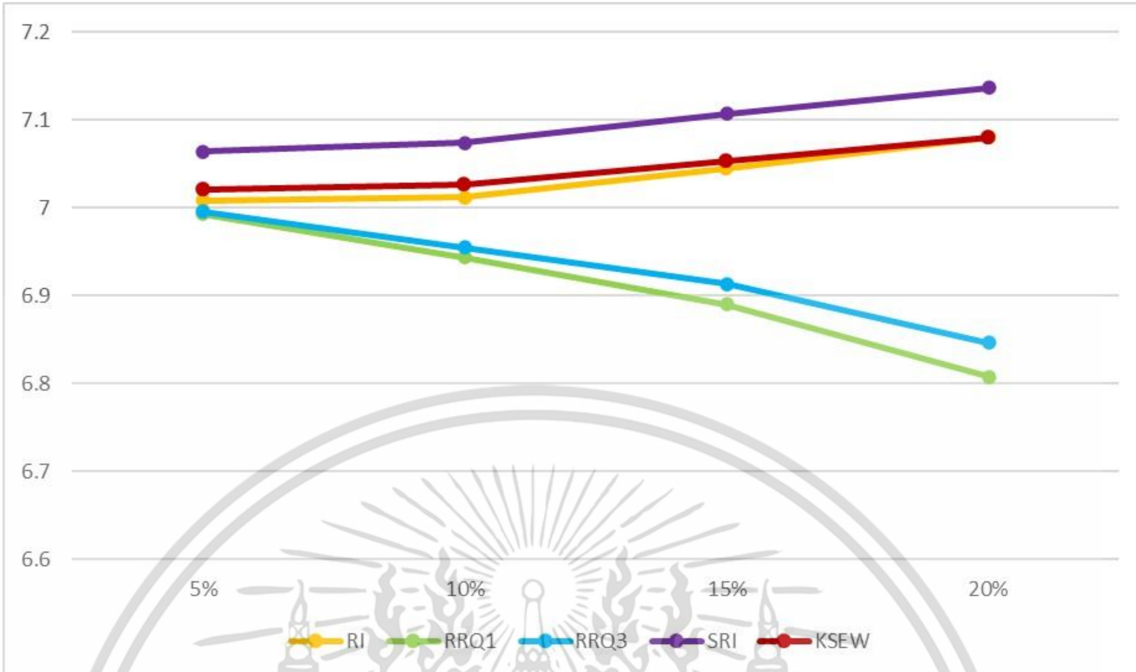


รูปที่ 4.23 การเปรียบเทียบค่า AMSE ที่ขนาดตัวอย่าง 100 ความแปรปรวนของความคลาดเคลื่อน 7

จากรูปที่ 4.23 จะเห็นได้ว่าที่ขนาดตัวอย่าง 100 ความแปรปรวนของความคลาดเคลื่อนเท่ากับ 7 ทุกร้อยละข้อมูลสูญหาย วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด รองลงมาคือ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3

จากรูปที่ 4.23 จะเห็นว่าทุกร้อยละข้อมูลสูญหาย วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด

จะเห็นแนวโน้มว่า ร้อยละข้อมูลสูญหายของวิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 และ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 เมื่อร้อยละข้อมูลสูญหายมีค่าเพิ่มมากขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยจะมีค่าลดลง ส่วนวิธีการถดถอยพหุคูณ วิธีการถดถอยสโตแคสติก และวิธีการถดถอยสโตแคสติกใกล้สุด k ตัวด้วยค่าถ่วงน้ำหนักที่เท่ากัน เมื่อร้อยละข้อมูลสูญหายมีค่าเพิ่มมากขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยจะมีค่าเพิ่มมากขึ้นตามไปด้วย



รูปที่ 4.24 การเปรียบเทียบค่า AMSE ที่ขนาดตัวอย่าง 120 ความแปรปรวนของความคลาดเคลื่อน 7

จากรูปที่ 4.24 จะเห็นได้ว่าที่ขนาดตัวอย่าง 120 ความแปรปรวนของความคลาดเคลื่อนเท่ากับ 7 ทุกร้อยละข้อมูลสูญหาย วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด รองลงมาคือ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3

จากรูปที่ 4.24 จะเห็นว่าทุกร้อยละข้อมูลสูญหาย วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด

จะเห็นแนวโน้มว่า ร้อยละข้อมูลสูญหายของวิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 และ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 เมื่อร้อยละข้อมูลสูญหายมีค่าเพิ่มมากขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยจะมีค่าลดลง ส่วนวิธีการถดถอยพหุคูณ วิธีการถดถอยสโตแคสติก และวิธีการถดถอยสโตแคสติกใกล้สุด k ตัวด้วยค่าถ่วงน้ำหนักที่เท่ากัน เมื่อร้อยละข้อมูลสูญหายมีค่าเพิ่มมากขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยจะมีค่าเพิ่มมากขึ้นตามไปด้วย

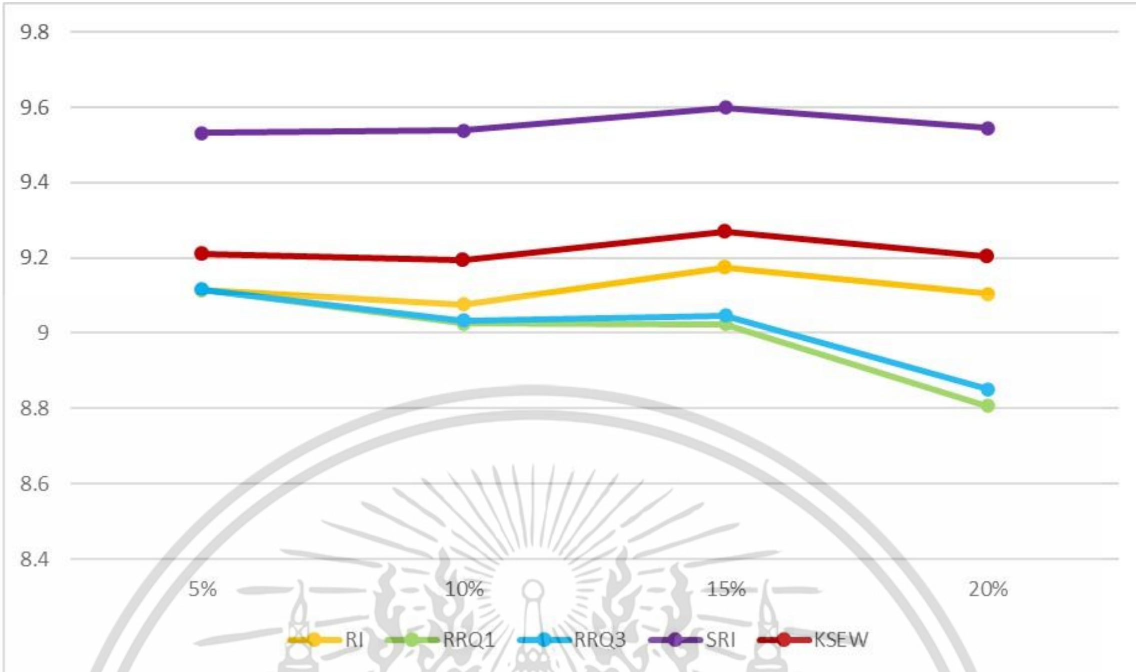
ตารางที่ 4.5 ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย เมื่อค่าความแปรปรวนของความคลาดเคลื่อนเท่ากับ 9

ขนาดตัวอย่าง	ค่าสูญหาย (%)	วิธีประมาณค่าสูญหาย				
		RI	RRQ1	RRQ3	SRI	KSEW
ตัวอย่างขนาดเล็ก						
20	5%	**9.1149**	9.1155	9.1154	9.5324	9.2094
	10%	9.0771	**9.0251**	9.0337	9.5383	9.1931
	15%	9.1741	**9.024**	9.0476	9.5995	9.2706
	20%	9.1047	**8.8066**	8.8512	9.5447	9.2044
40	5%	9.0281	**9.0152**	9.0174	9.2664	9.0896
	10%	8.9846	**8.9144**	8.9258	9.2168	9.0431
	15%	9.0813	**8.8982**	8.9267	9.2742	9.1173
	20%	9.0865	**8.7396**	8.79	9.3255	9.1392
ตัวอย่างขนาดปานกลาง						
60	5%	9.0668	**9.0503**	9.0531	9.2116	9.1015
	10%	9.0361	**8.9562**	8.9691	9.1891	9.0728
	15%	9.0303	**8.8321**	8.8627	9.1845	9.0653
	20%	9.1308	**8.7744**	8.826	9.2852	9.1542
80	5%	9.0004	**8.9816**	8.9848	9.1032	9.0227
	10%	8.9816	**8.8957**	8.9095	9.0888	9.006
	15%	8.9803	**8.7834**	8.8135	9.1061	9.01
	20%	9.0631	**8.7015**	8.7537	9.1789	9.0821
ตัวอย่างขนาดใหญ่						
100	5%	9.0058	**8.9862**	8.9895	9.0998	9.03
	10%	9.0226	**8.9371**	8.9507	9.1147	9.044
	15%	9.0288	**8.8225**	8.854	9.1159	9.0453
	20%	9.1175	**8.7477**	8.801	9.2097	9.1288
120	5%	8.9532	**8.9328**	8.9363	9.0218	8.9686
	10%	9.0014	**8.9116**	8.926	9.0825	9.0204
	15%	8.9765	**8.7726**	8.8036	9.0474	8.9882
	20%	9.0526	**8.69**	8.7421	9.1189	9.0513

จากตารางที่ 4.5 เมื่อค่าความแปรปรวนของความคลาดเคลื่อนมีค่าเท่ากับ 9 ขนาดตัวอย่าง 20 ร้อยละข้อมูลสูญหาย 5 วิธีการถดถอยพหุคูณ เป็นวิธีที่มีประสิทธิภาพมากที่สุด ส่วนร้อยละข้อมูลสูญหาย 10, 15 และ 20 วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด

ขนาดตัวอย่าง 40, 60, 80, 100 และ 120 ทุกร้อยละข้อมูลสูญหาย วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด

จากตารางที่ 4.5 จะพบว่าทุกขนาดตัวอย่าง วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 (RRQ1) เป็นวิธีที่มีประสิทธิภาพมากที่สุด

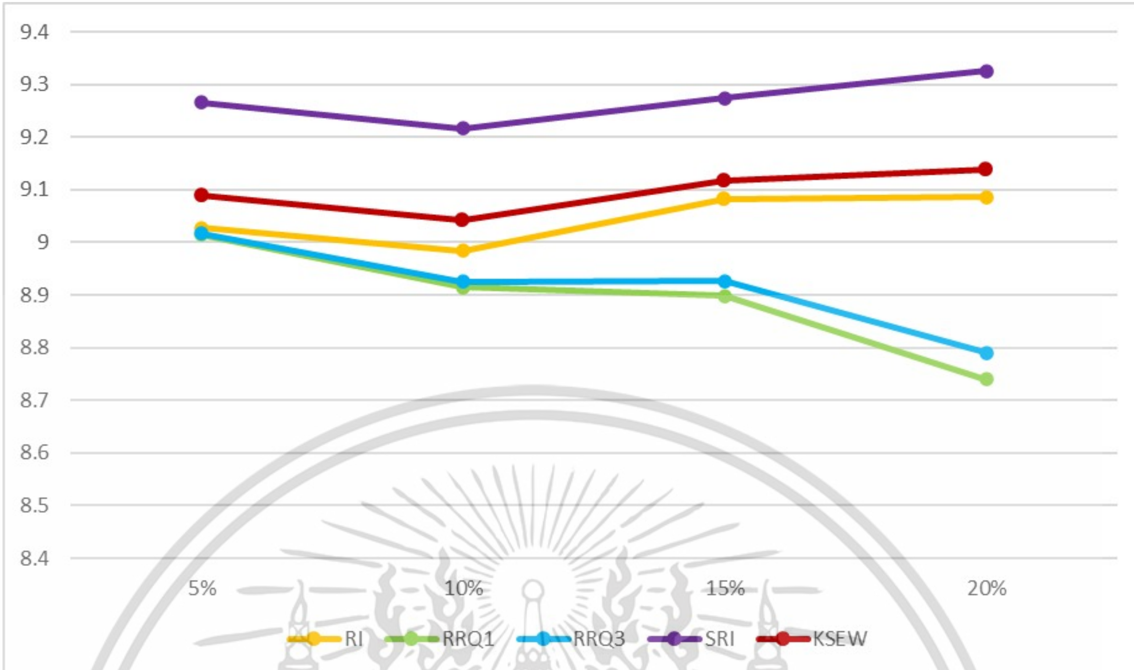


รูปที่ 4.25 การเปรียบเทียบค่า AMSE ที่ขนาดตัวอย่าง 20 ความแปรปรวนของความคลาดเคลื่อน 9

จากรูปที่ 4.25 จะเห็นได้ว่าที่ขนาดตัวอย่าง 20 ความแปรปรวนของความคลาดเคลื่อนเท่ากับ 9 ร้อยละข้อมูลสูญหาย 5 วิธีการถดถอยพหุคูณ เป็นวิธีที่มีประสิทธิภาพมากที่สุด รองลงมาคือ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 ส่วนร้อยละข้อมูลสูญหาย 10, 15 และ 20 วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด รองลงมาคือ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3

จากรูปที่ 4.25 จะเห็นว่าทุกร้อยละข้อมูลสูญหาย วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด

จะเห็นแนวโน้มว่า ร้อยละข้อมูลสูญหายของวิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 และ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 เมื่อร้อยละข้อมูลสูญหายมีค่าเพิ่มมากขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยจะมีค่าลดลง

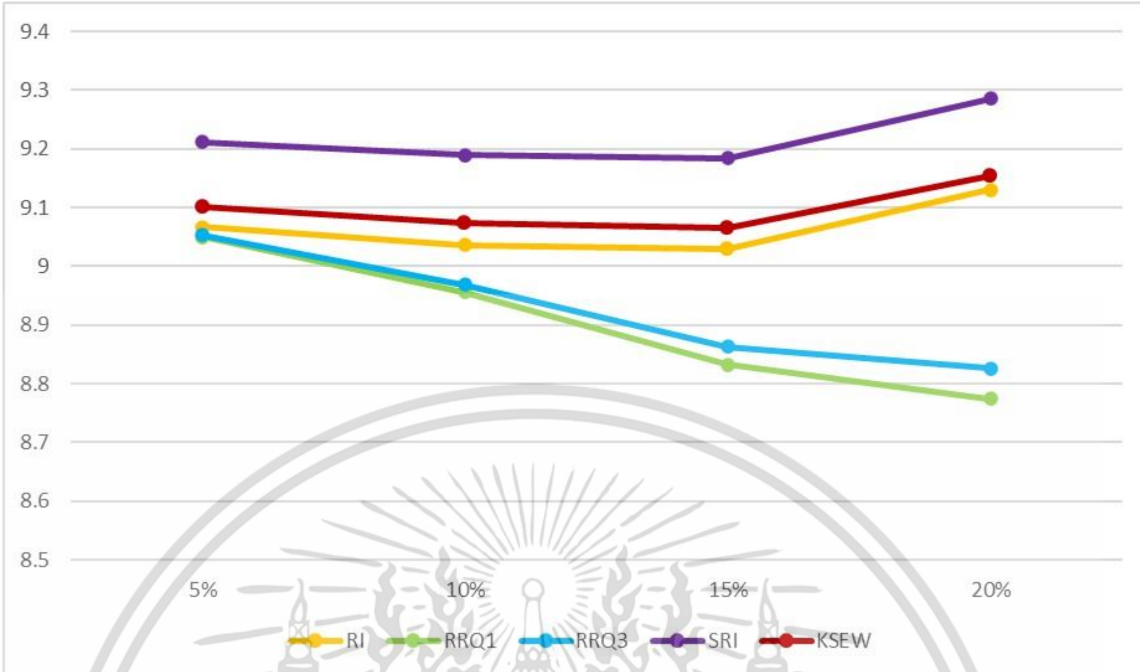


รูปที่ 4.26 การเปรียบเทียบค่า AMSE ที่ขนาดตัวอย่าง 40 ความแปรปรวนของความคลาดเคลื่อน 9

จากรูปที่ 4.26 จะเห็นได้ว่าที่ขนาดตัวอย่าง 40 ความแปรปรวนของความคลาดเคลื่อนเท่ากับ 9 ทุกร้อยละข้อมูลสูญหาย วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด รองลงมาคือ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3

จากรูปที่ 4.26 จะเห็นว่าทุกร้อยละข้อมูลสูญหาย วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด

จะเห็นแนวโน้มว่า ร้อยละข้อมูลสูญหายของวิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 และ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 เมื่อร้อยละข้อมูลสูญหายมีค่าเพิ่มมากขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยจะมีค่าลดลง ส่วนวิธีการถดถอยพหุคูณ วิธีการถดถอยสโตแคสติก และวิธีการถดถอยสโตแคสติกใกล้สุด k ตัวด้วยค่าถ่วงน้ำหนักที่เท่ากัน เมื่อร้อยละข้อมูลสูญหายมีค่าเพิ่มมากขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยจะมีค่าเพิ่มมากขึ้นตามไปด้วย

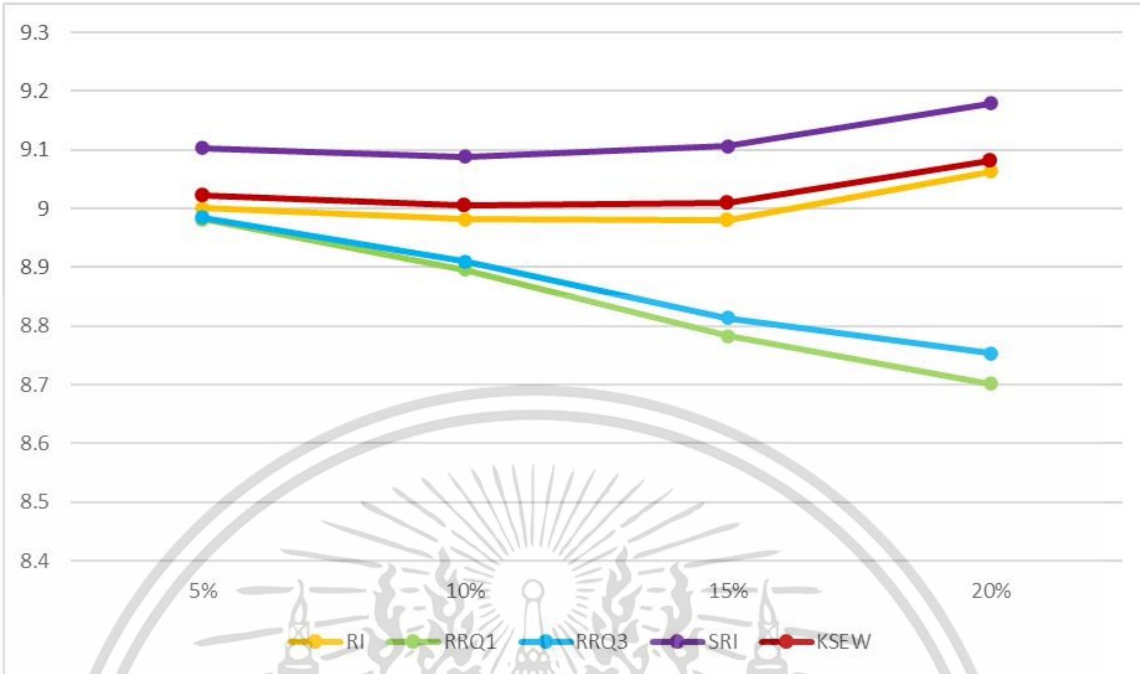


รูปที่ 4.27 การเปรียบเทียบค่า AMSE ที่ขนาดตัวอย่าง 60 ความแปรปรวนของความคลาดเคลื่อน 9

จากรูปที่ 4.27 จะเห็นได้ว่าที่ขนาดตัวอย่าง 60 ความแปรปรวนของความคลาดเคลื่อนเท่ากับ 9 ทุกร้อยละข้อมูลสูญหาย วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด รองลงมาคือ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3

จากรูปที่ 4.27 จะเห็นว่าทุกร้อยละข้อมูลสูญหาย วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด

จะเห็นแนวโน้มว่า ร้อยละข้อมูลสูญหายของวิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 และ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 เมื่อร้อยละข้อมูลสูญหายมีค่าเพิ่มมากขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยจะมีค่าลดลง ส่วนวิธีการถดถอยพหุคูณ วิธีการถดถอยสโตแคสติก และวิธีการถดถอยสโตแคสติกใกล้สุด k ตัวด้วยค่าถ่วงน้ำหนักที่เท่ากัน เมื่อร้อยละข้อมูลสูญหายมีค่าเพิ่มมากขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยจะมีค่าเพิ่มมากขึ้นตามไปด้วย

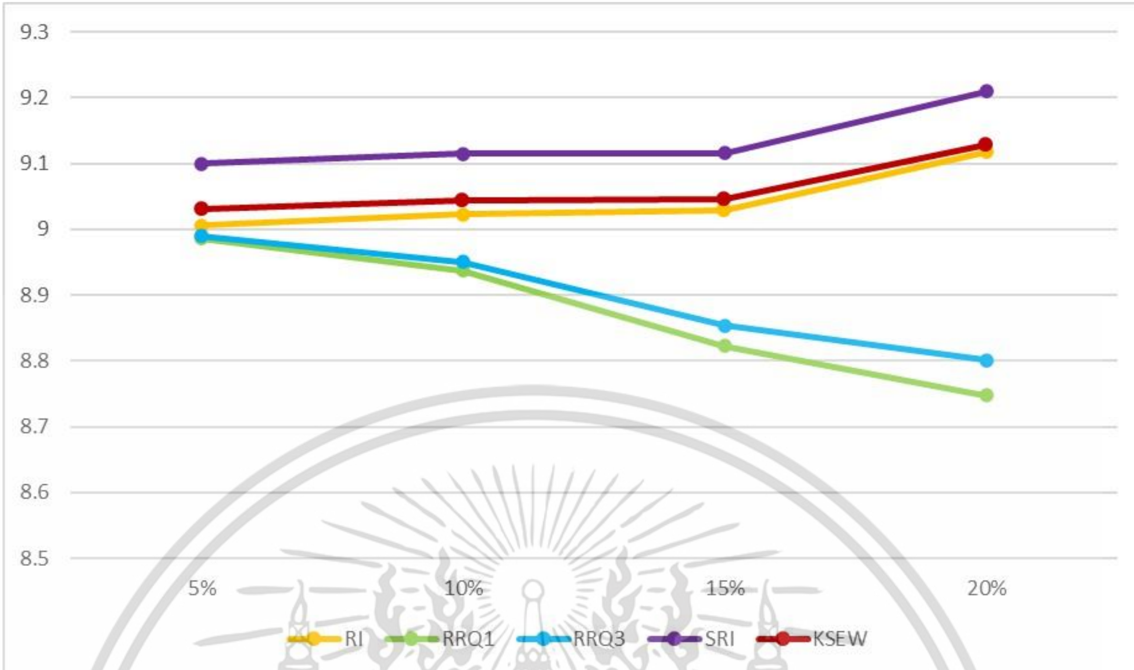


รูปที่ 4.28 การเปรียบเทียบค่า AMSE ที่ขนาดตัวอย่าง 80 ความแปรปรวนของความคลาดเคลื่อน 9

จากรูปที่ 4.28 จะเห็นได้ว่าที่ขนาดตัวอย่าง 80 ความแปรปรวนของความคลาดเคลื่อนเท่ากับ 9 ทุกร้อยละข้อมูลสูญหาย วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด รองลงมาคือ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3

จากรูปที่ 4.28 จะเห็นว่าทุกร้อยละข้อมูลสูญหาย วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด

จะเห็นแนวโน้มว่า ร้อยละข้อมูลสูญหายของวิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 และ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 เมื่อร้อยละข้อมูลสูญหายมีค่าเพิ่มมากขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยจะมีค่าลดลง ส่วนวิธีการถดถอยพหุคูณ วิธีการถดถอยสโตแคสติก และวิธีการถดถอยสโตแคสติกใกล้สุด k ตัวด้วยค่าถ่วงน้ำหนักที่เท่ากัน เมื่อร้อยละข้อมูลสูญหายมีค่าเพิ่มมากขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยจะมีค่าเพิ่มมากขึ้นตามไปด้วย

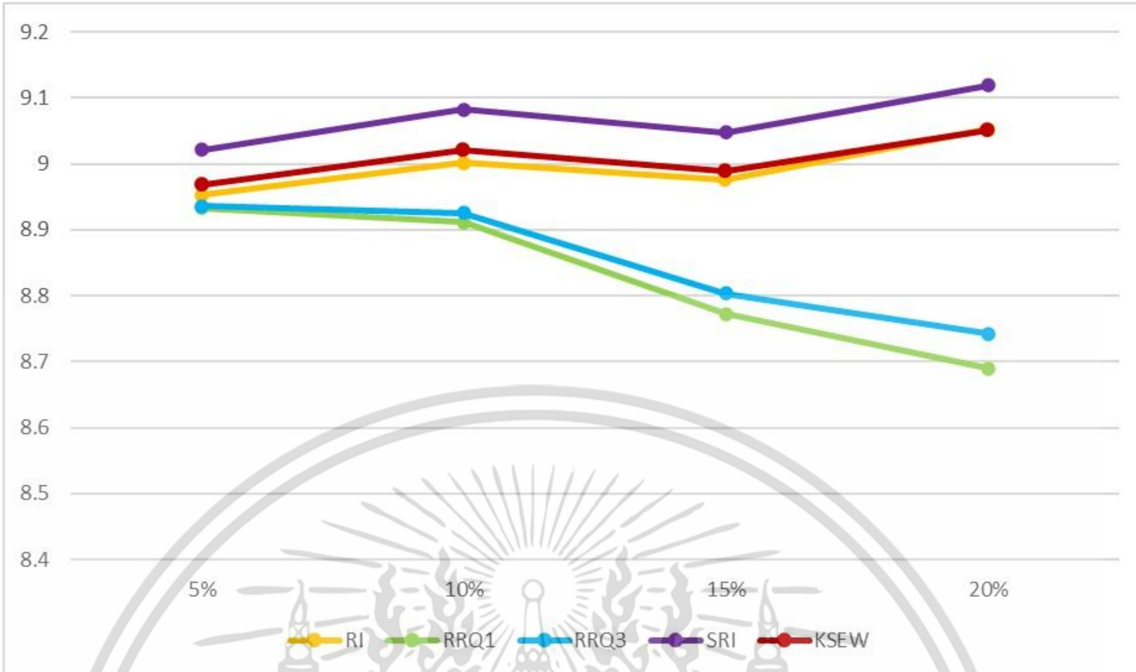


รูปที่ 4.29 การเปรียบเทียบค่า AMSE ที่ขนาดตัวอย่าง 100 ความแปรปรวนของความคลาดเคลื่อน 9

จากรูปที่ 4.29 จะเห็นได้ว่าที่ขนาดตัวอย่าง 100 ความแปรปรวนของความคลาดเคลื่อนเท่ากับ 9 ทุกร้อยละข้อมูลสูญหาย วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด รองลงมาคือ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3

จากรูปที่ 4.29 จะเห็นว่าทุกร้อยละข้อมูลสูญหาย วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด

จะเห็นแนวโน้มว่า ร้อยละข้อมูลสูญหายของวิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 และ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 เมื่อร้อยละข้อมูลสูญหายมีค่าเพิ่มมากขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยจะมีค่าลดลง ส่วนวิธีการถดถอยพหุคูณ วิธีการถดถอยสโตแคสติก และวิธีการถดถอยสโตแคสติกใกล้สุด k ตัวด้วยค่าถ่วงน้ำหนักที่เท่ากัน เมื่อร้อยละข้อมูลสูญหายมีค่าเพิ่มมากขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยจะมีค่าเพิ่มมากขึ้นตามไปด้วย



รูปที่ 4.30 การเปรียบเทียบค่า AMSE ที่ขนาดตัวอย่าง 120 ความแปรปรวนของความคลาดเคลื่อน 9

จากรูปที่ 4.30 จะเห็นได้ว่าที่ขนาดตัวอย่าง 120 ความแปรปรวนของความคลาดเคลื่อนเท่ากับ 9 ทุกร้อยละข้อมูลสูญหาย วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด รองลงมาคือ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3

จากรูปที่ 4.30 จะเห็นว่าทุกร้อยละข้อมูลสูญหาย วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพมากที่สุด

จะเห็นแนวโน้มว่า ร้อยละข้อมูลสูญหายของวิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 และ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 เมื่อร้อยละข้อมูลสูญหายมีค่าเพิ่มมากขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยจะมีค่าลดลง ส่วนวิธีการถดถอยพหุคูณ วิธีการถดถอยสโตแคสติก และวิธีการถดถอยสโตแคสติกใกล้สุด k ตัวด้วยค่าถ่วงน้ำหนักที่เท่ากัน เมื่อร้อยละข้อมูลสูญหายมีค่าเพิ่มมากขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยจะมีค่าเพิ่มมากขึ้นตามไปด้วย

ตารางที่ 4.6 ผลสรุปการประมาณค่าสูญหายในการวิเคราะห์การถดถอยพหุคูณทั้ง 5 วิธี

ขนาดตัวอย่าง	ค่าสูญหาย (%)	ค่าความแปรปรวนของความคลาดเคลื่อน				
		1	3	5	7	9
ตัวอย่างขนาดเล็ก						
20	5%	RI	RI	RI	RI	RI
	10%	RRQ1	RRQ1	RRQ1	RRQ1	RRQ1
	15%	RRQ1	RRQ1	RRQ1	RRQ1	RRQ1
	20%	RRQ1	RRQ1	RRQ1	RRQ1	RRQ1
40	5%	RRQ1	RRQ1	RRQ1	RRQ1	RRQ1
	10%	RRQ1	RRQ1	RRQ1	RRQ1	RRQ1
	15%	RRQ1	RRQ1	RRQ1	RRQ1	RRQ1
	20%	RRQ1	RRQ1	RRQ1	RRQ1	RRQ1
ตัวอย่างขนาดปานกลาง						
60	5%	RRQ1	RRQ1	RRQ1	RRQ1	RRQ1
	10%	RRQ1	RRQ1	RRQ1	RRQ1	RRQ1
	15%	RRQ1	RRQ1	RRQ1	RRQ1	RRQ1
	20%	RRQ3	RRQ1	RRQ1	RRQ1	RRQ1
80	5%	RRQ1	RRQ1	RRQ1	RRQ1	RRQ1
	10%	RRQ1	RRQ1	RRQ1	RRQ1	RRQ1
	15%	RRQ1	RRQ1	RRQ1	RRQ1	RRQ1
	20%	RRQ3	RRQ1	RRQ1	RRQ1	RRQ1
ตัวอย่างขนาดใหญ่						
100	5%	RRQ1	RRQ1	RRQ1	RRQ1	RRQ1
	10%	RRQ1	RRQ1	RRQ1	RRQ1	RRQ1
	15%	RRQ1	RRQ1	RRQ1	RRQ1	RRQ1
	20%	RRQ3	RRQ1	RRQ1	RRQ1	RRQ1
120	5%	RRQ1	RRQ1	RRQ1	RRQ1	RRQ1
	10%	RRQ1	RRQ1	RRQ1	RRQ1	RRQ1
	15%	RRQ1	RRQ1	RRQ1	RRQ1	RRQ1
	20%	RRQ3	RRQ1	RRQ1	RRQ1	RRQ1

หมายเหตุ RI หมายถึง วิธีการถดถอยพหุคูณ

RRQ1 หมายถึง วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1

RRQ3 หมายถึง วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3

SRI หมายถึง วิธีการถดถอยสโตแคสติก

KSEW หมายถึง วิธีการถดถอยสโตแคสติกใกล้สุด k ตัว ด้วยค่าถ่วงน้ำหนักที่เท่ากัน

จากตารางที่ 4.6 จะพบว่าเมื่อตัวอย่างมีขนาดเล็ก ขนาดปานกลางและขนาดใหญ่ พบว่า ในทุกร้อยละค่าสูญหายและทุกค่าความแปรปรวนของความคลาดเคลื่อน วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 มีประสิทธิภาพสูงที่สุด รองลงมาคือวิธีการถดถอยพหุคูณและวิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3

4.2 การอภิปรายผล

จากการศึกษาเปรียบเทียบวิธีการประมาณค่าสูญหายในตัวแบบพหุคูณ เมื่อข้อมูลมีค่าร้อยละค่าสูญหาย ขนาดตัวอย่าง และค่าความแปรปรวนของความคลาดเคลื่อนที่แตกต่างกัน โดยใช้วิธีการประมาณค่าสูญหายทั้งหมด 5 วิธี คือ วิธีการถดถอยพหุคูณ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 วิธีการถดถอยสโตแคสติก และวิธีการถดถอยสโตแคสติกใกล้สุด k ตัว ด้วยค่าถ่วงน้ำหนักที่เท่ากัน เกณฑ์ที่ใช้ในการพิจารณาเพื่อหาวิธีที่มีประสิทธิภาพดีที่สุดคือ ค่าคลาดเคลื่อนกำลังสองเฉลี่ย ผลการวิจัยสรุปได้ว่า เมื่อตัวอย่างมีขนาดเล็กและขนาดปานกลาง พบว่าส่วนใหญ่ในทุกร้อยละค่าสูญหายและทุกค่าความแปรปรวนของความคลาดเคลื่อน วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 มีประสิทธิภาพสูงที่สุด รองลงมาคือวิธีการถดถอยสโตแคสติกใกล้สุด k ตัว ด้วยค่าถ่วงน้ำหนักที่เท่ากัน แต่เมื่อตัวอย่างมีขนาดใหญ่ พบว่าส่วนใหญ่ในทุกร้อยละค่าสูญหายและทุกค่าความแปรปรวนของความคลาดเคลื่อน วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 มีประสิทธิภาพสูงที่สุด รองลงมาคือวิธีการถดถอยสโตแคสติก ซึ่งให้ผลสอดคล้องกับรัตติกาล (2555) พบว่าวิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เป็นวิธีที่มีประสิทธิภาพสูงกว่าวิธีอื่นๆ

ส่วนเมื่อพิจารณาจากค่าความแปรปรวนของความคลาดเคลื่อนมีค่าน้อยๆ ($\varepsilon = 1$) ในทุกขนาดตัวอย่างและทุกร้อยละของข้อมูลสูญหาย ส่วนใหญ่วิธีการถดถอยสโตแคสติกใกล้สุด k ตัว ด้วยค่าถ่วงน้ำหนักที่เท่ากัน จะมีประสิทธิภาพสูงที่สุด แต่เมื่อค่าความแปรปรวนของความคลาดเคลื่อนมีค่ามากขึ้น ($\varepsilon = 3,5,7$ และ 9) ในทุกขนาดตัวอย่างและทุกร้อยละของข้อมูลสูญหาย ส่วนใหญ่วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 จะมีประสิทธิภาพสูงที่สุด ซึ่งให้ผลสอดคล้องกับงานวิจัยของรัตติกาล (2555) พบว่าวิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 มีประสิทธิภาพสูงกว่าวิธีอื่นๆ เมื่อตัวอย่างมีขนาดเล็ก (n=20) ร้อยละของข้อมูลสูญหายมาก (20%) และความแปรปรวนของความคลาดเคลื่อนมีค่าค่อนข้างมาก (σ^2 =1.5, 2)

บทที่ 5
สรุปผลการวิจัย และข้อเสนอแนะ

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาเปรียบเทียบวิธีการประมาณค่าสูญหายในตัวแบบการถดถอยพหุคูณ ด้วยวิธีการประมาณ 5 วิธี คือ วิธีการถดถอยพหุคูณ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 วิธีการถดถอยสโตแคสติก และวิธีการถดถอยสโตแคสติกใกล้สุด k ตัว ด้วยค่าถ่วงน้ำหนักที่เท่ากัน เมื่อขนาดตัวอย่าง ความแปรปรวนของความคลาดเคลื่อน และร้อยละของจำนวนข้อมูลสูญหายแตกต่างกัน โดยศึกษาและเปรียบเทียบในแต่ละสถานการณ์ดังนี้

1. กำหนดขนาดตัวอย่าง คือขนาดตัวอย่างเล็กเท่ากับ 20 และ 40 ขนาดตัวอย่างปานกลางเท่ากับ 60 และ 80 ขนาดตัวอย่างใหญ่เท่ากับ 100 และ 120

2. กำหนดค่าพารามิเตอร์ คือ $\beta_0 = 0.5, \beta_1 = 1$ และ $\beta_2 = -0.3$

3. กำหนดร้อยละข้อมูลสูญหาย เท่ากับ 5, 10, 15 และ 20

4. สุ่มตัวแปรอิสระ X_1 มีพารามิเตอร์ $\mu = 3, \sigma^2 = 2.25$ และตัวแปรอิสระ X_2 มีพารามิเตอร์ $\mu = 5, \sigma^2 = 4$

5. สุ่มค่าความคลาดเคลื่อนมีการแจกแจงปรกติ $(\varepsilon_i \sim N(0,\sigma^2))$ โดยมีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวน σ^2 เท่ากับ 1, 3, 5, 7 และ 9

เกณฑ์ที่ใช้ในการประมาณค่าสูญหายเพื่อหาวิธีในการประมาณค่าสูญหายที่ดีที่สุด จะพิจารณาจากค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square error; MSE) จากจำนวน 1,000 รอบมาหาค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ย โดยใช้โปรแกรม R Studio V.4.2.1 ในการจำลองสถานการณ์

5.1 สรุปผลการวิจัย

จากผลการศึกษาและเปรียบเทียบการประมาณค่าสูญหายทั้ง 5 วิธี โดยพิจารณาจากค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ย ในแต่ละกรณี สำหรับตัวแบบการถดถอยพหุคูณ สามารถสรุปผลการวิเคราะห์ได้ดังนี้

1. กรณีตัวอย่างขนาดเล็ก ขนาดปานกลางและขนาดใหญ่พบว่า ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยของวิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 มีค่าน้อยที่สุด ซึ่งเป็นวิธีที่มีประสิทธิภาพมากที่สุด

2. เมื่อร้อยละข้อมูลสูญหาย 5, 10, 15 และ 20 ในทุกๆร้อยละข้อมูลสูญหายพบว่า วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 มีค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยน้อยที่สุด จึงเป็นวิธีที่มีประสิทธิภาพมากที่สุด

3. สำหรับกรณีความแปรปรวนของความคลาดเคลื่อนเท่ากับ 1, 3, 5, 7 และ 9 พบว่าวิธีที่มีค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยน้อยที่สุดคือ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1

4. ในทุกๆขนาดตัวอย่าง ร้อยละข้อมูลสูญหายและความแปรปรวนของความคลาดเคลื่อน วิธีที่มีประสิทธิภาพมากที่สุดคือ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 ยกเว้นในกรณีที่ขนาดตัวอย่างเล็กเท่ากับ 20 ร้อยละข้อมูลสูญหายเท่ากับ 5 และทุกความแปรปรวนของความคลาดเคลื่อน วิธีการถดถอยพหุคูณ มีประสิทธิภาพมากที่สุด

5.2 ข้อเสนอแนะ

5.2.1 ด้านการนำไปใช้ประโยชน์

1. กรณีตัวแปรอิสระในตัวแบบการถดถอยพหุคูณมีค่าข้อมูลสูญหาย เมื่อต้องการประมาณค่าข้อมูลสูญหายส่วนใหญ่ควรเลือกใช้วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 เนื่องจากให้ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยน้อยที่สุด

2. กรณีตัวแปรอิสระในตัวแบบการถดถอยพหุคูณมีค่าข้อมูลสูญหาย เมื่อต้องการประมาณค่าข้อมูลสูญหาย ในกรณีขนาดตัวอย่างเล็กควรเลือกใช้วิธีการถดถอยพหุคูณ เนื่องจากให้ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยน้อยที่สุด

5.2.2 ด้านการศึกษาวิจัย

1. เพื่อเป็นแนวทางให้แก่ผู้ที่สนใจศึกษาเปรียบเทียบวิธีการประมาณค่าสูญหายในตัวแบบการถดถอยพหุคูณเมื่อขนาดตัวอย่าง ร้อยละข้อมูลสูญหายและความแปรปรวนของความคลาดเคลื่อนมีค่าแตกต่างกัน ด้วยวิธีการประมาณค่าสูญหาย 5 วิธีคือ วิธีการถดถอยพหุคูณ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 วิธีการถดถอยสโตแคสติกและวิธีการถดถอยสโตแคสติกใกล้สุด k ตัว ด้วยค่าถ่วงน้ำหนักที่เท่ากัน อาจใช้วิธีการประมาณค่าสูญหายอื่นๆ เช่น วิธี EM Algorithm วิธี Pairwise Deleted วิธีต้นไม้ตัดสินใจ (Decision Tree) วิธีป่าสุ่ม (Random Forest) และวิธีอะดาบูส (Adaboost)

2. เพื่อเป็นแนวทางให้แก่ผู้ที่สนใจศึกษาเปรียบเทียบวิธีการประมาณค่าสูญหายในตัวแบบการถดถอยพหุคูณเมื่อขนาดตัวอย่าง ร้อยละข้อมูลสูญหายและความแปรปรวนของความคลาดเคลื่อนมีค่าแตกต่างกัน ด้วยวิธีการประมาณค่าสูญหาย 5 วิธีคือ วิธีการถดถอยพหุคูณ วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 วิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 วิธีการถดถอยสโตแคสติกและวิธีการถดถอยสโตแคสติก

ใกล้สุด k ตัว ด้วยค่าถ่วงน้ำหนักที่เท่ากัน อาจใช้ตัวแบบอื่นๆ เช่น การถดถอยเชิงเส้นอย่างง่าย (Simple Linear Regression) และ การถดถอยโลจิสติก (Logistic Regression)

3. เพื่อเพิ่มความหลากหลายของข้อมูลและสถานการณ์ที่แตกต่างกัน ควรเพิ่มตัวแปรอิสระและเปลี่ยนค่าพารามิเตอร์ให้มีความหลากหลายในการศึกษา

เอกสารอ้างอิง

จิราวัลย์ จิตรถเวช. 2558. การวิเคราะห์การถดถอย. กรุงเทพฯ. โครงการส่งเสริมและพัฒนาเอกสารวิชาการ สถาบันบัณฑิตพัฒนบริหารศาสตร์.

ฐณัฐ วงศ์สายเชื้อ. 2559. Replace missing value-การแทนค่าสูญหายในโปรแกรม SPSS. เข้าถึงได้จาก : https://www.youtube.com/watch?v=WzaeJ_HAqtk.

ทัตดา หิรัญพต. 2554. การเปรียบเทียบประสิทธิภาพระหว่างวิธีการประมาณข้อมูลสูญหายด้วยค่าถดถอย และวิธีการประมาณข้อมูลสูญหายด้วยค่าถดถอยแบบสโทแคสติก. วิทยาศาสตรบัณฑิต สาขาวิชาสถิติ. มหาวิทยาลัยบูรพา.

นงลักษณ์ วัชชัย. 2542. โมเดลลิสเรล : สถิติวิเคราะห์สำหรับการวิจัย. พิมพ์ครั้งที่ 3. กรุงเทพฯ. จุฬาลงกรณ์มหาวิทยาลัย.

ประชุม สุวัตถี. 2552. การสำรวจด้วยตัวอย่างการชักตัวอย่างและการวิเคราะห์. กรุงเทพฯ. โครงการส่งเสริมและพัฒนาเอกสารวิชาการ สถาบันบัณฑิตพัฒนบริหารศาสตร์.

ประชุม สุวัตถี. 2554. ทฤษฎีการชักตัวอย่าง. กรุงเทพฯ. โครงการส่งเสริมและพัฒนาเอกสารวิชาการ สถาบันบัณฑิตพัฒนบริหารศาสตร์.

พัชนา สุวรรณแสน. 2561. การจัดการข้อมูลสูญหาย: วิธีเคเนียร์เรสเนเบอร์. วารสารวิจัยวิทยาศาสตร์และเทคโนโลยี. 1: 1-9.

รัตติกาล จอมประพันธ์. 2555. การประมาณค่าสูญหายในการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ. วิทยาศาสตรมหาบัณฑิต สาขาวิชาสถิติ. สถาบันบัณฑิตพัฒนบริหารศาสตร์.

รัตติกาล จอมประพันธ์ และพาชิตชนัต ศิริพาณิช. 2558. การประมาณค่าสูญหายในการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ. วารสารพัฒนบริหารศาสตร์. 55(1): 183-202.

อุษณีย์ วงศ์อมาตย์. 2555. การเปรียบเทียบวิธีการประมาณค่าสูญหายแบบนอนอิกนอร์-เรเบิลในการวิเคราะห์การถดถอยเส้นพหุ. วิทยาศาสตรมหาบัณฑิต. สาขาวิชาสถิติ จุฬาลงกรณ์มหาวิทยาลัย.

เรืองลักษณ์ หลำใจซื่อ, อำไพ ทองธีรภาพ และจุฑาภรณ์ สินสมบูรณ์ทอง. 2560. การเปรียบเทียบวิธีการประมาณค่าสูญหายสำหรับการวิเคราะห์การถดถอยพหุเมื่อตัวแปรตามมีการสูญหายแบบสุ่ม. วารสารวิทยาศาสตร์และเทคโนโลยี. 5: 766-777. doi: 10.14456/tstj.2017.64

เสาวณิต สุขภารังษี. 2564. การใช้เทคนิคการพยากรณ์ร่วมด้วยตัวถ่วงน้ำหนัก. ว.พัฒนาเทคนิคศึกษา. 15(16): 60-65.

เอกสารอ้างอิง (ต่อ)

Al-Omari, A.I., Ibrahim, K. and Jemain, A.A. 2009. New ratio estimators of the mean using simple random sampling and ranked set sampling methods. Revista Investigation Operational. 30(2): 97-108.

Chaimongkol, W. 2005. Three Composite Imputation Methods for Item Nonresponse Estimation in Sample Survey. Ph.D.Thesis. National Institute of Development Administration.

Glass, Gene V. and Hopkins, Kenneth D. 1984. Statistics Method in Education and Psychology. 2nd ed. Englewood Cliffs. New Jersey.

Howell, D. C. 2007. Statistical Methods for Psychology. 6th ed. California. Duxbury Press.

Jonsson, P. and Wohlin, C. 2004. An evaluation of k-nearest neighbor imputation using likert data. International Symposium on Software Metrics (METRICS’04). 1530-1435.

Little, R.J.A. and Rubin, D.B. 1987. Statistical Analysis with Missing Data. New York. Wiley.

Marcelino, C.G., Leite, G.M.C., Celes, P. and Pedreira, C.E. 2022. Missing data analysis in regression. Applied Artificial Intelligence. 36(1): 2157-2182.

Montgomery, D.C. 2012. Introduction to Linear Regession Analysis. 5th ed. New Jersey. Wiley.

Muhammad, A. and Klairung, S. 2022. Imputation for multiple regression with missing heteroscedastic data. Thailand Statistician. 20(1): 1-15.

Sujitta, S., Chancharoen, R., Jaruchat, B., Chanchai, C. and Wilaiwan, N. 2006. A comparison of estimation methods for missing data in multiple linear regression with two independent variables. Thailand Statistician. 4: 13-26.

Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R., Botstein, D. and Altman, R.B. 2001. Missing values estimation methods for DNA microarrays. Bioinformatics. 17(6): 520-525.

Wilks, S.S. 1959. Nonparametric Statistical Inference. New York. Wiley.

Williamson, T. 2013. Calculating Pi using the monte carlo method. The Physics Teacher. 51(8): 468-469.

ภาคผนวก

คำสั่งโปรแกรม R Studio ที่ใช้ในงานวิจัย

****

1. คำสั่งโปรแกรม R Studio สำหรับกำหนดค่าข้อมูลที่ใช้ในการประมาณค่าสูญหาย

****

```
set.seed(5)
N=20
e=1
#ร้อยละข้อมูลสูญหาย
PMD = 5
#ค่า k วิธี KSEW
k=5
m=10
sum_mse_REG <- 0
sum_mse_RRQ1 <- 0
sum_mse_RRQ3 <- 0
sum_mse_SRI <- 0
sum_mse_KSEW <- 0
for (j in 1:m) {
```

****

2. คำสั่งโปรแกรม R Studio สำหรับสร้างข้อมูลประชากรของตัวแปรอิสระ X_1, X_2 และค่าความคลาดเคลื่อน ε

****

```
#สร้างประชากร X1
X1 = rnorm(100000,3,sqrt(1.5))
#สร้างประชากร X2
X2 = rnorm(100000,5,sqrt(2))
#สร้างประชากร E
Data_e = sqrt(e)
E = rnorm(100000,mean=0,sd=Data_e)
```

****

3. คำสั่งโปรแกรม R Studio สำหรับสุ่มตัวอย่างของตัวแปรอิสระ X_1, X_2 และค่าความคลาดเคลื่อน ε

****

```
#สุ่มตัวแปร X1
Sample_X1 = sample(X1, size=N,)
Data_X1=Sample_X1
#สุ่มตัวแปร X2
Sample_X2 = sample(X2, size=N,)
Data_X2=Sample_X2
#สุ่มตัวแปร E
Sample_E = sample(E, size=N,)
Data_E=Sample_E
```

****

4. คำสั่งโปรแกรม R Studio สำหรับทดสอบความสัมพันธ์เชิงเส้นพหุ (Multicollinearity) ระหว่างตัวแปรอิสระ

****

```
# correlation
X_correlation = cbind(Data_X1,Data_X2)
cor(X_correlation)
```

****

5. คำสั่งโปรแกรม R Studio สำหรับสร้างตัวแปรตามที่มีความสัมพันธ์เชิงเส้นกับตัวแปรอิสระ

****

```
#สร้างตัวแปร Y
Y = function(Data_X1,Data_X2, Data_E) {
  0.5+(1*(Data_X1))+(-0.3*(Data_X2)+ Data_E)
}
```

****

6. คำสั่งโปรแกรม R Studio สำหรับคำนวณจำนวนข้อมูลสูญหายและสุ่มตำแหน่งข้อมูลสูญหายของตัวแปรอิสระ

****

```
#คำนวณจำนวนข้อมูลสูญหาย
NMD = function(N,PMD) {
  (N)*(PMD)/100
}
#จำนวนข้อมูลสูญหาย
NMD(N,PMD)
#คำนวณตำแหน่งข้อมูลสูญหาย
N_MD = NMD(N,PMD)
PM=sample(N,size = N_MD)
PM
```

****

7. คำสั่งโปรแกรม R Studio สำหรับการประมาณข้อมูลสูญหายของข้อมูลตัวแปรอิสระด้วยวิธีการถดถอยพหุคูณและหาค่าความคลาดเคลื่อนกำลังสองเฉลี่ย

****

```
#ประมาณค่าสูญหาย_REG
Data_Xi_REG=rep(1,N)
Data_Y_REG = Y(Data_X1,Data_X2, Data_E)
Data_X_REG = c(Data_Xi_REG, Data_X1, Data_X2)
matrix_X_REG <- matrix(Data_X_REG,nrow = N,ncol = 3)
matrix_Y_REG <- matrix(Data_Y_REG,nrow = N,ncol = 1)
#ประมาณค่าสูญหาย
New_Estimate_REG = (sum(Data_X2)-(matrix_X_REG[PM,3]))/(N-N_MD)
Data_X2_REG_New=replace(Data_X2,list= PM,values = New_Estimate_REG)
#สร้าง matrix หลังจากการประมาณค่าสูญหาย
Data_X_REG_New = c(Data_Xi_REG, Data_X1, Data_X2_REG_New)
matrix_X_REG_New <- matrix(Data_X_REG_New,nrow = N,ncol = 3)
#X'_REG_New
Transpose_X_REG_New = t(matrix_X_REG_New)
#X'X_REG_New
TransposeX_X_REG_New = Transpose_X_REG_New%*%matrix_X_REG_New
#X'Y_REG_New
TransposeX_Y_REG_New = Transpose_X_REG_New%*%matrix_Y_REG
#Inverse(X'X)_REG_New
Inverse_TransposeX_X_REG_New = solve(TransposeX_X_REG_New)
#B_REG_New
B_REG_New = Inverse_TransposeX_X_REG_New%*%TransposeX_Y_REG_New
#สมการตัวแบบหลังจากการประมาณค่าสูญหาย
cat("Yhat=",B_REG_New[1,1],"+",B_REG_New[2,1],"X1+",B_REG_New[3,1],"X2+e")
#ค่า Y ที่ประมาณจากตัวแบบหลังจากการประมาณค่าสูญหาย
```

```
Y_REG_New = function(B_REG_New,Data_X1,Data_X2_REG_New,Data_E) {
B_REG_New[1,1]+(B_REG_New[2,1]*Data_X1)+(B_REG_New[3,1]*Data_X2_REG_New)+Data_
E
}
#หาค่า MSE_REG
matrix_Y_REG_New <-
matrix(Y_REG_New(B_REG_New,Data_X1,Data_X2_REG_New,Data_E),nrow = N,ncol = 1)
MSE_REG = function(N,matrix_Y_REG,matrix_Y_REG_New) {
  (1/N)*(sum((matrix_Y_REG-matrix_Y_REG_New)^2))
}
```

****

8. คำสั่งโปรแกรม R Studio สำหรับการประมาณข้อมูลสูญหายของข้อมูลตัวแปรอิสระด้วยวิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 1 และหาค่าความคลาดเคลื่อนกำลังสองเฉลี่ย

****

```
#ประมาณค่าสูญหาย_RRQ1
Data_Y_REG_New=Y_REG_New(B_REG_New,Data_X1,Data_X2_REG_New,Data_E)
matrix_Y_REG_New <- matrix(Data_Y_REG_New,nrow = N,ncol = 1)
#ตำแหน่ง Q1
R1=1
RRQ1=function(N,R1) {
  (R1*(N+1))/4
}
#ข้อมูลตำแหน่งที่ Q1
P_RRQ1=RRQ1(N,R1)
Data_X2_Sort=sort(Data_X2)
matrix_Data_X2_Sort <- matrix(Data_X2_Sort,nrow = N,ncol = 1)
Data_X2_RRQ1 = matrix_Data_X2_Sort[P_RRQ1,1]
New_Estimate_RRQ1 =
(matrix_Y_REG_New[PM,1])*((mean(Data_X2)+Data_X2_RRQ1)/(((sum(Data_X2)-
(matrix_X_REG[PM,3]))/(N-N_MD))+Data_X2_RRQ1))
Data_Y_RRQ1_New=replace(Data_Y_REG_New,list= PM,values = New_Estimate_RRQ1)
#หาค่า MSE_RRQ1
matrix_Y_RRQ1_New <- matrix(Data_Y_RRQ1_New,nrow = N,ncol = 1)
MSE_RRQ1 = function(N,matrix_Y_REG,matrix_Y_RRQ1_New) {
  (1/N)*(sum((matrix_Y_REG-matrix_Y_RRQ1_New)^2))
}
```

****

9. คำสั่งโปรแกรม R Studio สำหรับการประมาณข้อมูลสูญหายของข้อมูลตัวแปรอิสระด้วยวิธีการถดถอย-อัตราส่วนควอร์ไทล์ที่ 3 และหาค่าความคลาดเคลื่อนกำลังสองเฉลี่ย

****

```
#ประมาณค่าสูญหาย_RRQ3
Data_Y_REG_New=Y_REG_New(B_REG_New,Data_X1,Data_X2_REG_New,Data_E)
matrix_Y_REG_New <- matrix(Data_Y_REG_New,nrow = N,ncol = 1)
#ตำแหน่ง Q3
R3=3
RRQ3=function(N,R3) {
  (R3*(N+1))/4
}
#ข้อมูลตำแหน่งที่ Q3
P_RRQ3=RRQ3(N,R3)
Data_X2_Sort=sort(Data_X2)
matrix_X2_Sort <- matrix(Data_X2_Sort,nrow = N,ncol = 1)
Data_X2_RRQ3 = matrix_X2_Sort[P_RRQ3,1]
New_Estimate_RRQ3 =
(matrix_Y_REG_New[PM,1])*((mean(Data_X2)+Data_X2_RRQ3)/(((sum(Data_X2)-
(matrix_X_REG[PM,3]))/(N-N_MD))+Data_X2_RRQ3))
Data_Y_RRQ3_New=replace(Data_Y_REG_New,list= PM,values = New_Estimate_RRQ3)
#หาค่า MSE_RRQ3
matrix_Y_RRQ3_New <- matrix(Data_Y_RRQ3_New,nrow = N,ncol = 1)
MSE_RRQ3 = function(N,matrix_Y_REG,matrix_Y_RRQ3_New) {
  (1/N)*(sum((matrix_Y_REG-matrix_Y_RRQ3_New)^2))
}
```

****

10. คำสั่งโปรแกรม R Studio สำหรับการประมาณข้อมูลสูญหายของข้อมูลตัวแปรอิสระด้วยวิธีการถดถอยสโตแคสติก และหาค่าความคลาดเคลื่อนกำลังสองเฉลี่ย

****

```
#ประมาณค่าสูญหาย_SRI
Zi = replicate(n=N, rnorm(n=N, mean = 0, sd = Data_e))
Data_Zi = apply(X=Zi,MARGIN = 2, FUN = mean)
matrix_Zi_SRI_New <- matrix(Data_Zi,nrow = N,ncol = 1)
Data_Y_SRI_New = Data_Y_REG_New+matrix_Zi_SRI_New
#หาค่า MSE_SRI
MSE_SRI = function(N,matrix_Y_REG,Data_Y_SRI_New) {
  (1/N)*(sum((matrix_Y_REG-Data_Y_SRI_New)^2))
}
```

****

11. คำสั่งโปรแกรม R Studio สำหรับการประมาณข้อมูลสูญหายของข้อมูลตัวแปรอิสระด้วยวิธีการถดถอยสโตแคสติกใกล้สุด k ตัว ด้วยค่าถ่วงน้ำหนักที่เท่ากัน และหาค่าความคลาดเคลื่อนกำลังสองเฉลี่ย

****

```
#ประมาณค่าสูญหาย_KSEW
#SRI
Data_Y_SRI_New
#KNN
EX_KNN_X1 = matrix_X_REG[PM,2]
EX_KNN_Y = matrix_Y_REG[PM,1]
KNN_EX <- data.frame(Data_X1,Data_X2,Data_Y_REG)
KNN_EX$"EX_KNN" <- sqrt(((EX_KNN_X1-Data_X1)^2)+((EX_KNN_Y-Data_Y_REG)^2))
Sort_KNN <- KNN_EX[order(KNN_EX$"EX_KNN"), ]
K=k+1
New_Estimate_KNN=mean(Sort_KNN[, "Data_X2"][2:K])
Data_X2_KNN_New=replace(Data_X2,list= PM,values = New_Estimate_KNN)
#สร้าง matrix หลังจากการประมาณค่าสูญหาย
Data_X_KNN_New = c(Data_Xi_REG, Data_X1, Data_X2_KNN_New)
matrix_X_KNN_New <- matrix(Data_X_KNN_New,nrow = N,ncol = 3)
#X'_New
Transpose_X_KNN_New = t(matrix_X_KNN_New)
#X'X_New
TransposeX_X_KNN_New = Transpose_X_KNN_New%*%matrix_X_KNN_New
#Inverse(X'X)_New
Inverse_TransposeX_X_KNN_New = solve(TransposeX_X_KNN_New)
#X'Y_New
TransposeX_Y_KNN_New = Transpose_X_KNN_New%*%matrix_Y_REG
#B_New
```

```
B_KNN_New = Inverse_TransposeX_X_KNN_New%*%TransposeX_Y_KNN_New
#สมการตัวแบบหลังจากการประมาณค่าสูญหาย
cat("Yhat=",B_KNN_New[1,1],"+",B_KNN_New[2,1],"X1+",B_KNN_New[3,1],"X2+e")
#ค่า Y ที่ประมาณจากตัวแบบหลังจากการประมาณค่าสูญหาย
Y_KNN_New = function(B_KNN_New,Data_X1,Data_X2_KNN_New,Data_E) {
B_KNN_New[1,1]+(B_KNN_New[2,1]*Data_X1)+(B_KNN_New[3,1]*Data_X2_KNN_New)+Dat
a_E
}
Y_KNN=Y_KNN_New(B_KNN_New,Data_X1,Data_X2_KNN_New,Data_E)
Y_KSEW = function(Y_KNN,Data_Y_SRI_New) {
  (Data_Y_SRI_New+Y_KNN)/2
}
Y_KSEW(Y_KNN,Data_Y_SRI_New)
#หาค่า MSE_KSEW
matrix_Y_KSEW_New <- matrix(Y_KSEW(Y_KNN,Data_Y_SRI_New),nrow = N,ncol = 1)
MSE_KSEW = function(N,matrix_Y_REG,matrix_Y_KSEW_New) {
  (1/N)*(sum((matrix_Y_REG-matrix_Y_KSEW_New)^2))
}
```

****

12. คำสั่งโปรแกรม R Studio สำหรับหาค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ย

****

```
cat(c("loop :",j),fill=T)
}
AVR_MSE_REG = sum_mse_REG/m
AVR_MSE_REG
AVR_MSE_RRQ1 = sum_mse_RRQ1/m
AVR_MSE_RRQ1
AVR_MSE_RRQ3 = sum_mse_RRQ3/m
AVR_MSE_RRQ3
AVR_MSE_SRI = sum_mse_SRI/m
AVR_MSE_SRI
AVR_MSE_KSEW = sum_mse_KSEW/m
AVR_MSE_KSEW
```