

ความแม่นยำของการใช้แบบจำลองการถดถอยพหุคูณในการวิเคราะห์ฐานข้อมูลที่ได้จาก

การสุ่มแบบมีเงื่อนไขเพื่อการคาดการณ์ผลการเรียนของนักเรียนนายเรือ

The Accuracy of Multiple Regression Models in Conditional Randomized Database Analysis to Predict Academic Performance of Naval Cadets.

ศศิธร คงอุดมทรัพย์¹ และ พงศิษฐ์ ทวิชพงศ์ธร²

Sasithorn Kongudomthrap¹ and Pongsit Twichpongthorn²

Received : February 10, 2021

Revised : March 15, 2021

Accepted : March 16, 2021

บทคัดย่อ

บทความเชิงวิชาการนี้มีวัตถุประสงค์เพื่อนำเสนอการตรวจสอบความแม่นยำของการใช้แบบจำลองวิเคราะห์การถดถอยเชิงเส้นพหุคูณในการวิเคราะห์ฐานข้อมูลที่ได้จากการสุ่มแบบมีเงื่อนไขด้วยคอมพิวเตอร์ โดยฐานข้อมูลดังกล่าวสร้างขึ้นจากต้นแบบข้อมูลผลการศึกษจริงในรายวิชาความน่าจะเป็นและสถิติสำหรับวิศวกรรมของนักเรียนนายเรือ โรงเรียนนายเรือ ประจำปีการศึกษา 2562 ฐานข้อมูลสุ่มนี้จะบรรจุข้อมูลตัวแปรต้นคือ ปัจจัยที่มีผลต่อผลการเรียนรายวิชาประกอบด้วย ปัจจัยตัวบุคคล และปัจจัยการเรียนการสอน และตัวแปรตามคือ ผลการเรียนรายวิชา (เกรด) เขียนโปรแกรมสร้างฐานข้อมูลสุ่มย้อนหลังจำนวน 30 ปี การศึกษา แล้วทำการทดสอบความแม่นยำของแบบจำลองวิเคราะห์การถดถอยเชิงเส้นพหุคูณ โดยการเขียนโปรแกรมคอมพิวเตอร์จากฐานข้อมูลที่สร้างขึ้น เพื่อสร้างสมการพยากรณ์ผลการเรียนในรายวิชา นำผลที่ได้เปรียบเทียบกับผลการศึกษจริงพบว่า ผลทดสอบความแม่นยำใน 1 ปีการศึกษาใดๆ โดยเฉลี่ยจะอยู่ที่ 84% ถ้าทดสอบกับฐานข้อมูลสุ่มสะสม 3 ปีการศึกษา จะมีความแม่นยำเฉลี่ย 88.4% เมื่อเพิ่มจำนวนฐานข้อมูล

¹ ผู้ช่วยศาสตราจารย์ กองวิชาคณิตศาสตร์ ฝ่ายศึกษา โรงเรียนนายเรือ

Assistant Professor Department of Mathematics, Education Branch, Royal Thai Naval Academy

E-mail : puicmath@gmail.com

² ผู้ช่วยศาสตราจารย์ กองวิชาวิศวกรรมไฟฟ้า ฝ่ายศึกษา โรงเรียนนายเรือ

Assistant Professor Department of Engineering, Education Branch, Royal Thai Naval Academy

E-mail : s@pongsit.com

สุ่มสะสมเป็น 5, 10, 15 และ 20 ปีการศึกษา จะพบว่า ความแม่นยำในการทำนายเฉลี่ยจะอยู่ที่ 90%, 91.67%, 91.67% และ 93% ตามลำดับ และเมื่อใช้ฐานข้อมูลสุ่มทั้ง 30 ปีการศึกษา จะได้สมการพยากรณ์ คือ

$$y = -2.47317 + (0.13134)a + (0.00209)b$$

เมื่อ a แทน ค่าเฉลี่ยในส่วนของปัจจัยตัวบุคคล

b แทน ค่าเฉลี่ยในส่วนของปัจจัยการเรียนการสอน

ซึ่งผลการทดสอบความแม่นยำของสมการพยากรณ์นี้พบว่า สามารถทำนายได้ถูกต้องถึง 96% จากผลการทดสอบความแม่นยำในการทำนายผลการเรียนรายวิชาความน่าจะเป็นและสถิติของนักเรียนนายเรือ โรงเรียนนายเรือ ด้วยการเขียนโปรแกรมคอมพิวเตอร์โดยใช้หลักการของแบบจำลองวิเคราะห์การถดถอย พหุคูณนี้สามารถสรุปได้ว่า เมื่อมีการเพิ่มจำนวนฐานข้อมูลสุ่มมากขึ้นเรื่อยๆ ความแม่นยำในการทำนายก็จะเพิ่มสูงมากขึ้นด้วย

คำสำคัญ: การวิเคราะห์การถดถอยพหุคูณ, ฐานข้อมูลสุ่มด้วยโปรแกรมคอมพิวเตอร์, การเขียนโปรแกรมคอมพิวเตอร์

Abstract

The objective of this academic articles is to show the accuracy of multiple linear regression models as a tool for analyzing randomized databases based on specified conditions. A computer program was designed to create a database using conditional randomization with correlation between academic performance and the factors affecting the performance - personal factors and teaching factors. It is based on the actual study results in a course, probability and statistics for engineering in 2019 Royal Thai Naval Academy. After constructing this randomized databases and getting 30 academic year data out of the designed computer program, the data were analyzed by multiple linear regression models to find the equation used to predict the academic performances. The accuracy of the prediction when using one academic year from the created database to build the predicting equation is approximately 84%. When three years were used, the average accuracy was approximately 88.4%. When the data were used more from 5, 10, 15 and 20 academic years, the forecast accuracy is approximately 90%, 91.67%, 91.67% and 93% respectively. When tested with 30 year randomized databases, the forecast equation is

$$y = -2.47317 + (0.13134)a + (0.00209)b$$

when a is an average of the personal factors and b is an average of teaching factors.

This was able to accurately predict 96% of the students' academic performances. The multiple linear regression models is gaining more predicting accuracy as the databases used to create predicting equation grow.

Keywords: multiple regression, conditional randomized databases, computer program design

1. บทนำ

โลกในยุคปัจจุบันเป็นโลกดิจิทัลที่มีข้อมูลข่าวสารเกิดขึ้นมากมายในชีวิตประจำวัน ข้อมูลส่วนใหญ่อยู่ในรูปแบบของดิจิทัล ข้อดีของดิจิทัลคือสามารถนำไปประมวลผลด้วยเครื่องมือต่าง ๆ ได้อย่างง่ายดายและรวดเร็ว จึงทำให้หน่วยงานทั้งภาครัฐและเอกชนนำระบบดิจิทัลไปใช้ในการดำเนินงานกันอย่างกว้างขวาง สามารถสังเกตได้จากการที่รัฐบาลให้ความสำคัญกับข้อมูลข่าวสารด้วยการพัฒนาขยายโครงข่ายอินเทอร์เน็ตความเร็วสูงทั้งภายในและระหว่างประเทศ เพื่อเพิ่มศักยภาพ และรองรับการสื่อสารที่เพิ่มมากขึ้น [1] แต่ในทางกลับกันจะพบว่าปริมาณข้อมูลข่าวสารที่มีจำนวนมหาศาลและเพิ่มมากขึ้นในทุก ๆ วันนั้น

ได้รับการวิเคราะห์หรือการหาความรู้ใหม่ ๆ จากข้อมูลเหล่านี้มีน้อยมาก โดยเฉพาะอย่างยิ่งข้อมูล หรือระบบฐานข้อมูลที่เกี่ยวข้องกับการศึกษา ซึ่งมีการเก็บรวบรวมข้อมูลเป็นจำนวนมาก หากมีการนำข้อมูลต่าง ๆ มาสำรวจ และนำเข้าสู่กระบวนการวิเคราะห์ข้อมูล ก็จะพบความรู้ใหม่ที่มีค่า ก่อให้เกิดประโยชน์สูงสุดต่อระบบการศึกษาต่อไป

การวิเคราะห์การถดถอยพหุคูณ (Multiple Linear Regression) เป็นวิธีการทางสถิติที่ใช้ศึกษาความสัมพันธ์ระหว่างตัวแปรอิสระ (Independent Variable) กับตัวแปรตาม (Dependent Variable) จะเป็นการศึกษาความสัมพันธ์เชิงเส้นตรง (Linearity) ถ้าศึกษาความสัมพันธ์ระหว่างตัวแปรอิสระหนึ่งตัวกับตัวแปรตามหนึ่งตัวเรียกว่า การวิเคราะห์การถดถอยเชิงเส้นเชิงเดียว หรือการวิเคราะห์การถดถอยเชิงเส้นอย่างง่าย (Simple Linear Regression Analysis) ถ้าตัวแปรอิสระมีมากกว่าหนึ่งตัวกับตัวแปรตามหนึ่งตัวเรียกว่า การวิเคราะห์การถดถอยเชิงเส้นพหุคูณ (Multiple Linear Regression) วัตถุประสงค์ของการใช้การวิเคราะห์การถดถอยคือ นำไปใช้เพื่อศึกษาความสัมพันธ์ระหว่างตัวแปรอิสระกับตัวแปรตาม และศึกษาปัจจัยหรือตัวแปรอิสระที่ร่วมกันทำนาย หรือพยากรณ์ตัวแปรตาม ซึ่งงานวิจัยส่วนใหญ่ที่มีการใช้การวิเคราะห์การถดถอยจะใช้การวิเคราะห์ผ่านโปรแกรม SPSS หรือ Excel ซึ่งเป็นโปรแกรมสำเร็จรูปที่ใช้ทำงานง่าย

การศึกษาปัจจัยที่ส่งผลต่อผลการเรียนรายวิชา หรือผลสัมฤทธิ์ของการเรียนในรายวิชา ในด้านของความหมายและทฤษฎีต่าง ๆ ที่เกี่ยวข้อง พบว่า ในส่วนของความหมายผลการเรียนรายวิชา และปัจจัยที่มีผลต่อผลการเรียนรายวิชานั้น กรมวิชาการกระทรวงศึกษาธิการ [2] และนักวิชาการ เช่น ไพศาล หวังพานิชย์ [3], นงลักษณ์ วิรัชชัย [4], ธวัชพร มะรินทร์ [5], กฤษฎา บุญวัฒน์ [6] เป็นต้น ได้ให้ข้อมูลไว้ทำให้สรุปได้ว่า ผลการเรียนรายวิชาหรือผลการสอบก็คือ ระดับของความรู้ หรือความสามารถที่ได้รับมาจากการเรียนการสอน ซึ่งต้องอาศัยความพยายาม หรือความสามารถเฉพาะตัวของบุคคล ส่วนปัจจัยที่ส่งผลต่อผลการเรียนรายวิชา หรือผลสอบประกอบด้วย ปัจจัยภายในตัวบุคคล และปัจจัยด้านการเรียนการสอน ในการวัดผลการศึกษารายวิชานั้นสามารถวัดได้จากเกรดเฉลี่ย หรือผลคะแนนจากการบ้าน หรืองานที่ครูมอบหมายให้

บทความเชิงวิชาการนี้จะนำเสนอการนำการวิเคราะห์การถดถอยมาใช้หาความสัมพันธ์ของตัวแปรปัจจัยที่มีผลต่อผลการเรียนกับผลการเรียนในอีกรูปแบบหนึ่งที่สอดคล้องกับโลกในยุคดิจิทัล มีการนำเทคโนโลยีสารสนเทศมาประยุกต์ใช้นั่นคือ ใช้การเขียนโปรแกรมวิเคราะห์การถดถอยแทนการใช้โปรแกรมสำเร็จรูป และทดสอบความแม่นยำของการใช้แบบจำลองการถดถอยพหุคูณในการวิเคราะห์ฐานข้อมูลที่ได้จากการสุ่มแบบมีเงื่อนไขเพื่อการคาดการณ์ผลการเรียนของนักเรียนนายเรือ โดยจะกล่าวถึง แนวคิดทฤษฎีที่เกี่ยวข้อง กรอบแนวคิด วัตถุประสงค์ของการนำเสนอ ประชากรกลุ่มตัวอย่าง เครื่องมือที่ใช้ในการวิเคราะห์ข้อมูล การวิเคราะห์ข้อมูล บทสรุป และข้อเสนอแนะในการนำไปใช้ประโยชน์

2. แนวคิด ทฤษฎีที่เกี่ยวข้อง และนิยามคำศัพท์

2.1 การวัดและประเมินผลการศึกษาภาควิทยาการของโรงเรียนนายเรือ [7] ซึ่งได้ทำการระบุเกณฑ์เพื่อใช้เป็นระเบียบในการวัดและประเมินผลการศึกษา ดังนี้ ระหว่างภาคการศึกษา อาจารย์ผู้สอนประจำวิชาเป็นผู้ดำเนินการตามวิธีที่เห็นว่าเหมาะสมในแต่ละกระบวนการวิชาตามหลักการวัดผลการศึกษา และให้มีการวัดผลในแต่ละภาคการศึกษาไม่น้อยกว่า 2 ครั้ง ส่วนปลายภาคการศึกษาเป็นการวัดผลด้วยวิธีการสอบปลายภาคการศึกษา และคิดคะแนนรายบุคคล ดำเนินการโดยคณะกรรมการสอบความรู้ที่โรงเรียนนายเรือแต่งตั้ง การประเมินผลการศึกษาแต่ละวิชาในระบบอักษร แสดงระดับคะแนนโดยมีอักษรประกอบระดับคะแนนต่อ 1 หน่วย ดังนี้

ตารางที่ 1 การวัดและการประเมินผลการสอบความรู้ ใช้ระบบอักษรแสดงระดับคะแนน (Grading System) [7]

อักษรระดับคะแนน	ระดับคะแนนต่อหน่วยกิต	ความเหมาะสม
A	4.0	ดีมาก
B+, B	3.5, 3.0	ดี
C+, C	2.5, 2.0	พอใช้
D+, D	1.5, 1.0	ผ่าน
F	0	ไม่ผ่าน
P	ไม่คิดระดับคะแนน	ผ่านกระบวนการวิชาที่ไม่คิดคะแนนเฉลี่ยสะสม
N	ไม่คิดระดับคะแนน	ไม่ผ่านกระบวนการวิชาที่ไม่คิดคะแนนเฉลี่ยสะสม

2.2 การวิเคราะห์การถดถอย [8] เป็นสถิติที่นิยมใช้ในการหาความสัมพันธ์ของข้อมูลเชิงเป็นเหตุเป็นผล ซึ่งผลการวิเคราะห์ข้อมูลจะสามารถบอกขนาดและทิศทางอิทธิพลของตัวแปรต้นแต่ละตัวที่มีต่อตัวแปรตาม และสร้างสมการพยากรณ์ตัวแปรตามได้เมื่อรู้ค่าตัวแปรต้น คำถามวิจัยที่เหมาะสมกับการใช้สถิติวิเคราะห์การถดถอย จะเป็นคำถามวิจัยเชิงสำรวจขนาดใหญ่ที่มีตัวแปรจำนวนมากทั้งตัวแปรต้น ตัวแปรตาม และตัวแปรแทรกซ้อน เช่น ความสัมพันธ์ระหว่างปัจจัยด้านครอบครัว ปัจจัยด้านนักเรียน และผลสัมฤทธิ์ทางการเรียนของนักเรียน, การพยากรณ์ผลสัมฤทธิ์ทางการเรียนของนักเรียนด้วยตัวแปรปัจจัยด้านครอบครัวและปัจจัยด้านนักเรียน, ผลการเปรียบเทียบขนาดอิทธิพลของตัวแปรปัจจัยด้านครอบครัวและปัจจัยด้านนักเรียนที่มีผลต่อผลสัมฤทธิ์ทางการเรียนของนักเรียน เป็นต้น การวิเคราะห์การถดถอย [9] หรือโมเดลการวิเคราะห์การถดถอย (Regression Analysis Model) ที่แสดงความสัมพันธ์ระหว่างตัวแปรต้น และตัวแปรตาม

มีหลายรูปแบบ เช่น โมเดลการวิเคราะห์การถดถอยพหุคูณ (Multiple Regression Analysis Model), โมเดลการวิเคราะห์การถดถอยพหุนาม (Polynomial Regression Analysis Model), โมเดลการวิเคราะห์การถดถอยพหุคูณที่มีตัวแปรดัมมี่ (Multiple Regression Analysis Model with Dummy Variables) เป็นต้น สำหรับการวิเคราะห์ข้อมูลของงานวิจัยนี้ได้เลือกใช้โมเดลการวิเคราะห์การถดถอยพหุคูณศึกษาความสัมพันธ์ระหว่างตัวแปรต้น กับตัวแปรตาม เมื่อตัวแปรทั้งสองมีความสัมพันธ์แบบเส้นตรง โดยจะได้สมการพยากรณ์ คือ

$$Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k$$

เมื่อ Y คือ ค่าของตัวแปรตาม

$X_1, X_2, \dots, X_k$ คือ ค่าของตัวแปรอิสระแต่ละตัว

k คือ จำนวนตัวแปรอิสระในสมการถดถอย

b_0 คือ ค่าคงที่ (Constant) ของสมการถดถอย

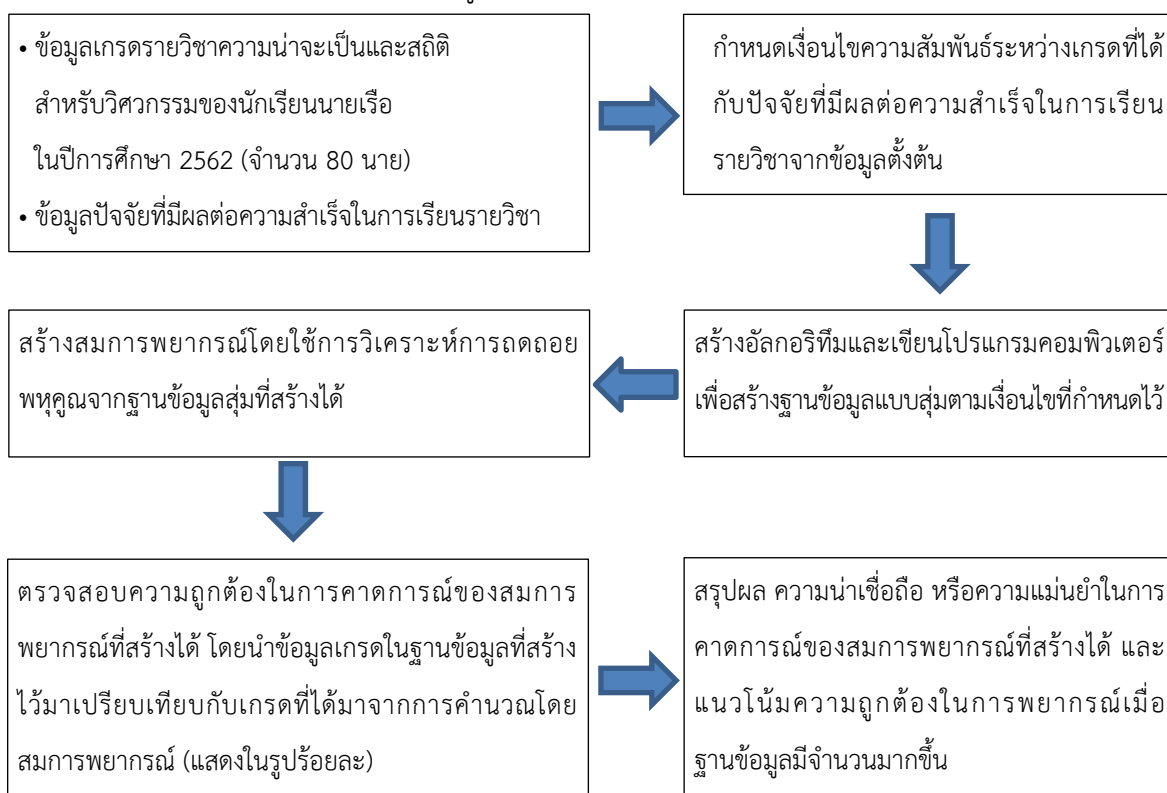
และ $b_1, b_2, \dots, b_k$ คือ ค่าสัมประสิทธิ์การถดถอย (Regression Coefficient) ของตัวแปรอิสระแต่ละตัว โดยที่ค่าสัมประสิทธิ์การถดถอยนี้จะแสดงอัตราการเปลี่ยนแปลงของค่าตัวแปรอิสระแต่ละตัวที่มีต่อค่าตัวแปรตาม

2.3 โปรแกรมที่นิยมนำมาใช้วิเคราะห์ข้อมูลที่นิยมนำใช้มากที่สุดคือ SPSS [10] ซึ่งเป็นโปรแกรมที่ช่วยให้ค้นหาความสัมพันธ์ใหม่ๆ ของข้อมูล และสามารถคาดการณ์สิ่งที่จะเกิดขึ้นต่อไปได้ ทำให้สามารถสร้าง จัดการ และเผยแพร่ข้อมูลเชิงลึกเพื่อการตัดสินใจได้อย่างรวดเร็ว เหมาะในการนำไปใช้งานในองค์กรที่มีข้อมูลเยอะมากๆ แต่ก็มีข้อเสียตรงที่มีความยุ่งยากในการเตรียมข้อมูล และการใช้งาน ประกอบกับในปัจจุบันมีโปรแกรมภาษาต่างๆ มากมายที่มีคำสั่งวิเคราะห์ข้อมูลได้เช่นเดียวกันกับ SPSS ทำให้เกิดความสับสนต่อการใช้งานมากยิ่งขึ้น ในบทความนี้จึงเลือกนำเสนอการวิเคราะห์ข้อมูลโดยใช้วิธีการเขียนโปรแกรมคอมพิวเตอร์เพื่อทดสอบความแม่นยำของการใช้แบบจำลองการถดถอยพหุคูณในการวิเคราะห์ฐานข้อมูลที่ได้จากการสุ่มแบบมีเงื่อนไขเพื่อการคาดการณ์ผลการเรียนของนักเรียนนายเรือ โดยใช้ OOP (Object – Oriented Programming) สำหรับรูปแบบในการพัฒนาโปรแกรม Visual Studio Code สำหรับการเขียนโปรแกรม PHP สำหรับการประมวลผลโปรแกรม HTML CSS และ JavaScript สำหรับการแสดงผลข้อมูล และ MySQL สำหรับการจัดการฐานข้อมูลต่าง ๆ

2.4 นิยามคำศัพท์ “การสุ่มแบบมีเงื่อนไข” ในบทความนี้คือ การตั้งคำตอบเป็นความสัมพันธ์ของข้อมูล แล้วหาคำตอบของข้อมูลโดยใช้แบบจำลองวิเคราะห์การถดถอยพหุคูณเชิงเส้น

3. กรอบแนวคิดของการศึกษา

ลำดับขั้นตอนในการวิเคราะห์ข้อมูล คือ



รูปที่ 1 ลำดับขั้นตอนในการวิเคราะห์ข้อมูล

4. วัตถุประสงค์ของการศึกษา

เพื่อนำเสนอการทดสอบความแม่นยำของหลักการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ โดยใช้การเขียนโปรแกรมคอมพิวเตอร์วิเคราะห์ฐานข้อมูลสุ่มแบบมีเงื่อนไข

4.1 ประชากร และกลุ่มตัวอย่าง

ประชากรที่นำมาใช้ในการศึกษา คือ ฐานข้อมูลสุ่มแบบมีเงื่อนไขจำนวน 30 ปีการศึกษาเป็นตัวแทนของข้อมูลปัจจัยที่มีผลต่อผลการเรียนรายวิชา และผลการเรียนรายวิชาของนักเรียนนายเรือ จำนวน 2,480 นาย ที่สร้างขึ้นจากต้นแบบข้อมูลผลการศึกษาริจริงของนักเรียนนายเรือในรายวิชาความน่าจะเป็นและสถิติสำหรับวิศวกรรม ประจำปีการศึกษา 2562 โรงเรียนนายเรือ

4.2 ตัวแปรที่ใช้ในการศึกษา

4.2.1 ตัวแปรต้น คือ ปัจจัยที่มีผลต่อผลการเรียนรายวิชา ได้แก่ ปัจจัยภายในด้านตัวนักเรียนนายเรือและปัจจัยด้านการศึกษา

4.2.2 ตัวแปรตาม คือ ผลการเรียนรายวิชา (A, B+, B, C+, C, D+, D, F)

4.3 เครื่องมือที่ใช้ในการศึกษา

4.3.1 โปรแกรมคอมพิวเตอร์ที่เขียนโดยใช้โปรแกรมภาษา PHP สำหรับการประมวลผลโปรแกรม HTML CSS และ JavaScript สำหรับการแสดงผลข้อมูล และ MySQL สำหรับการจัดการฐานข้อมูลต่าง ๆ สำหรับการสร้างฐานข้อมูลแบบมีเงื่อนไข

4.3.2 สถิติที่ใช้ในการวิเคราะห์ข้อมูล ได้แก่ การวิเคราะห์สถิติเชิงพรรณนา ประกอบด้วย ค่าเฉลี่ย ร้อยละ และสถิติเชิงอนุมาน ประกอบด้วย การวิเคราะห์การถดถอยเชิงเส้นพหุคูณ

4.4 ลำดับขั้นตอนการศึกษา

ในบทความนี้จะนำเสนอการใช้แบบจำลองวิเคราะห์การถดถอยพหุคูณเป็นเครื่องมือในการวิเคราะห์ฐานข้อมูลเพื่อทำนายผลการศึกษารายวิชาของนักเรียนนายเรือโดยมีลำดับของศึกษาดังนี้

4.4.1 นำข้อมูลผลการศึกษาจริงของนักเรียนนายเรือในรายวิชาความน่าจะเป็นและสถิติสำหรับวิศวกรรม ปีการศึกษา 2562 มาเป็นข้อมูลต้นแบบในการเขียนโปรแกรมคอมพิวเตอร์เพื่อสร้างฐานข้อมูล โดยให้ข้อมูลที่สุ่มได้มีจำนวนนักเรียนที่ได้เกรดเฉลี่ยในแต่ละระดับสอดคล้องกับข้อมูลจริง เช่น ให้แต่ละปีมีนักเรียนที่ได้เกรด A จำนวน 7 คน ได้ C จำนวน 16 คน เป็นต้น

4.4.2 เขียนโปรแกรมคอมพิวเตอร์สร้างฐานข้อมูลในส่วนของปัจจัยที่มีผลต่อผลการศึกษาในรายวิชา โดยกำหนดเงื่อนไขของการสุ่มข้อมูลให้มีความสัมพันธ์กันกับผลการเรียนรายวิชาอย่างชัดเจน เช่น ให้ผู้ที่ได้ A มีคะแนนเฉลี่ยของปัจจัยภายในด้านตัวนักเรียนนายเรือตั้งแต่ร้อยละ 80 ขึ้นไปและด้านการศึกษา ตั้งแต่ร้อยละ 60 ขึ้นไป และให้ผู้ที่ได้ C มีคะแนนเฉลี่ยของปัจจัยภายในด้านตัวนักเรียนนายเรือตั้งแต่ร้อยละ 40 แต่น้อยกว่าร้อยละ 50 และด้านการศึกษา ตั้งแต่ร้อยละ 40 แต่น้อยกว่าร้อยละ 50 เป็นต้น

4.4.3 เขียนโปรแกรมคอมพิวเตอร์สร้างสมการทำนายผลด้วยหลักการของแบบจำลองวิเคราะห์การถดถอยเชิงเส้นพหุคูณจากฐานข้อมูลในข้อ 4.4.2

4.4.4 เขียนโปรแกรมคอมพิวเตอร์คาดการณ์เกรดของนักเรียนนายเรือจากสมการทำนายผล และเปรียบเทียบผลการทำนายที่ได้กับผลการเรียนรายวิชาที่แท้จริงในฐานข้อมูล

4.4.5 เขียนโปรแกรมคำนวณร้อยละความถูกต้องของผลการทำนายที่ได้

4.4.6 ทำการตรวจสอบคุณภาพของเครื่องมือวิจัยโดยการเขียนโปรแกรมคอมพิวเตอร์ทดสอบความถูกต้องของผลการทำนายซ้ำๆ กันในแต่ละเซตข้อมูลหลายๆ รอบเพื่อหาค่าเฉลี่ยของความถูกต้องในการทำนาย

4.5 ผลการศึกษา และการวิเคราะห์ข้อมูล

การทดสอบความแม่นยำของสมการพยากรณ์ที่ได้กับฐานข้อมูลที่มีอยู่ โดยสุ่มข้อมูลในฐานข้อมูลมาทดสอบครั้งละ 1 ปีการศึกษา

เมื่อกำหนดให้ a แทน ค่าเฉลี่ยในส่วนของปัจจัยตัวบุคคล

b แทน ค่าเฉลี่ยในส่วนของปัจจัยการเรียนการสอน

จะได้สมการพยากรณ์ และผลการทดสอบความแม่นยำในการทำนายได้ดังตารางที่ 2

ตารางที่ 2 สมการพยากรณ์ และร้อยละความถูกต้องในแต่ละปีการศึกษา

ปีการศึกษา	สมการพยากรณ์ที่สร้างได้	ร้อยละของความถูกต้อง
2539	$y = -0.67845 + 0.12721a - 0.03293b$	86%
2542	$y = -0.85054 + 0.12509a - 0.02623b$	79%
2546	$y = -0.76448 + 0.12603a - 0.02909b$	85%
2555	$y = -0.73162 + 0.12574a - 0.02996b$	81%
2558	$y = -0.74877 + 0.125517a - 0.02846b$	84%

จากผลการทดสอบที่ได้จะพบว่า เมื่อทดสอบสมการพยากรณ์ที่ได้กับฐานข้อมูลสุ่ม สามารถทำนายเกรดนักเรียนนายเรือได้ถูกต้องประมาณ 83%

ถ้าทำการทดสอบสมการพยากรณ์กับฐานข้อมูลที่มีอยู่ โดยสุ่มข้อมูลที่มีอยู่ในฐานข้อมูลมาทดสอบครั้งละ 3 ปีการศึกษา จะได้สมการพยากรณ์ และผลการทดสอบความแม่นยำในการทำนายได้ดังตารางที่ 3

ตารางที่ 3 สมการพยากรณ์ และร้อยละความถูกต้องใน 3 ปีการศึกษาใดๆ

3 ปีการศึกษาใดๆ	สมการพยากรณ์ที่สร้างได้	ร้อยละของความถูกต้อง
2530 - 2532	$y = -1.56740 + 0.12771a - 0.01306b$	88%
2535 - 2537	$y = -1.37215 + 0.12709a - 0.01667b$	90%
2542 - 2544	$y = -1.54329 + 0.12875a - 0.01551b$	88%
2548 - 2550	$y = -1.39028 + 0.12727a - 0.01674b$	88%
2557 - 2559	$y = -1.41287 + 0.12707a - 0.01617b$	88%

จากผลการทดสอบที่ได้จะพบว่า เมื่อทดสอบสมการพยากรณ์ที่ได้กับฐานข้อมูลสุ่ม สามารถทำนายเกรดนักเรียนนายเรือได้ถูกต้องประมาณ 88.4%

ถ้าทำการทดสอบสมการพยากรณ์กับฐานข้อมูลที่มีอยู่ โดยสุ่มข้อมูลที่มีอยู่ในฐานข้อมูลมาทดสอบครั้งละ 5, 10, 15 และ 20 ปีการศึกษา จะได้สมการพยากรณ์ และผลการทดสอบความแม่นยำในการทำนายได้ดังตารางที่ 4 - 7

ตารางที่ 4 สมการพยากรณ์ และร้อยละความถูกต้องใน 5 ปีการศึกษาใดๆ

5 ปีการศึกษาใดๆ	สมการพยากรณ์ที่สร้างได้	ร้อยละของความถูกต้อง
2534 - 2538	$y = -1.71596 + 0.12857a - 0.01106b$	89%
2543 - 2547	$y = -1.83444 + 0.12923a - 0.00949b$	90%
2537 - 2541	$y = -1.68711 + 0.12919a - 0.01277b$	89%

ตารางที่ 5 สมการพยากรณ์ และร้อยละความถูกต้องใน 10 ปีการศึกษาใดๆ

10 ปีการศึกษาใดๆ	สมการพยากรณ์ที่สร้างได้	ร้อยละของความถูกต้อง
2530 - 2539	$y = - 2.12154 + 0.13018a - 0.00414b$	92%
2541 - 2550	$y = - 2.16439 + 0.13060a - 0.00397b$	92%
2549 - 2558	$y = - 2.13381 + 0.12991a - 0.00365b$	91%

ตารางที่ 6 สมการพยากรณ์ และร้อยละความถูกต้องใน 15 ปีการศึกษาใดๆ

15 ปีการศึกษาใดๆ	สมการพยากรณ์ที่สร้างได้	ร้อยละของความถูกต้อง
2534 - 2548	$y = - 2.26682 + 0.13090a - 0.00202b$	92%
2539 - 2553	$y = - 2.31671 + 0.13101a - 0.00103b$	91%
2543 - 2557	$y = - 2.30317 + 0.13062a - 0.00079b$	92%

ตารางที่ 7 สมการพยากรณ์ และร้อยละความถูกต้องใน 20 ปีการศึกษาใดๆ

20 ปีการศึกษาใดๆ	สมการพยากรณ์ที่สร้างได้	ร้อยละของความถูกต้อง
2535 - 2554	$y = - 2.38109 + 0.13119a + 0.00020b$	95%
2536 - 2555	$y = - 2.37500 + 0.13124a - 0.00005b$	92%
2540 - 2559	$y = - 2.37895 + 0.13101a + 0.00033b$	92%

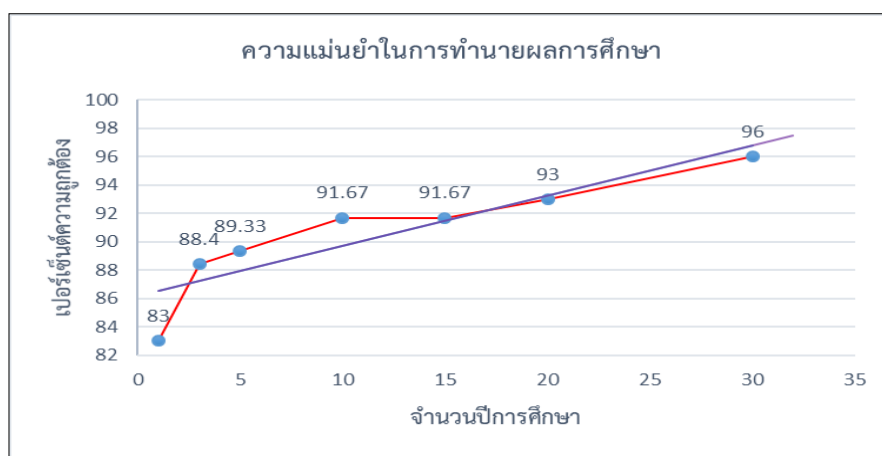
จากผลการทดสอบที่ได้จะพบว่า เมื่อทดสอบสมการพยากรณ์ที่ได้กับฐานข้อมูลสุ่มจำนวน 5, 10 และ 20 ปีการศึกษา จะสามารถทำนายเกรดนักเรียนนายเรือได้ถูกต้องโดยเฉลี่ยประมาณ 89.33%, 91.67%, 91.67% และ 93% ตามลำดับ และเมื่อทำการทดสอบสมการพยากรณ์กับฐานข้อมูลที่มีอยู่ โดยสุ่มข้อมูลที่มีอยู่ในฐานข้อมูลทั้งหมดมาทดสอบ (จำนวน 30 ปีการศึกษา มีข้อมูลทั้งหมด 2,480 ตัวอย่าง) จะได้สมการพยากรณ์ คือ

$$y = - 2.45872 + 0.13137a + 0.00169b$$

ซึ่งเมื่อทำการทดสอบความแม่นยำของสมการพยากรณ์ที่ได้จากแบบจำลองการวิเคราะห์การถดถอยพหุคูณสามารถทำนายผลการเรียนรายวิชาของนักเรียนนายเรือได้ถูกต้องถึง 96%

5. บทสรุป

เมื่อนำผลการทดสอบความแม่นยำในการทำนายผลการศึกษารายวิชาของนักเรียนนายเรือในฐานข้อมูลสุ่มด้วยสมการพยากรณ์ที่ได้จากแบบจำลองการวิเคราะห์การถดถอยพหุคูณมาสร้างกราฟแสดงความสัมพันธ์ระหว่างจำนวนปีการศึกษา หรือจำนวนฐานข้อมูลที่เพิ่มมากขึ้นกับเปอร์เซ็นต์ความถูกต้องในการทำนายผลการศึกษา จะได้กราฟ ดังนี้



รูปที่ 2 ความสัมพันธ์ระหว่างจำนวนปีการศึกษากับเปอร์เซ็นต์ของการทำนายผลการศึกษาได้ถูกต้อง

จากกราฟที่ได้จะพบว่า เมื่อฐานข้อมูลที่ใช้ในการทำนายมีปริมาณเพิ่มมากขึ้น เปอร์เซ็นต์ความถูกต้องในการทำนาย หรือความแม่นยำในการทำนายผลการศึกษารายวิชาของนักเรียนนายเรือก็จะสูงมากขึ้นตามไปด้วย จึงสรุปได้ว่า สมการพยากรณ์ที่ได้จากแบบจำลองการวิเคราะห์การถดถอยพหุคูณเป็นเครื่องมือที่สามารถนำมาใช้ในวิเคราะห์ผลการศึกษารายวิชาของนักเรียนนายเรือได้อย่างแม่นยำ และมีความเหมาะสมที่จะนำไปใช้งานจริง

6. การอภิปรายผลการศึกษา

ผลที่ได้จากการทดลองที่นำเสนอพบว่า แบบจำลองการวิเคราะห์การถดถอยพหุคูณสามารถใช้เป็นเครื่องมือในการทำนายผลการศึกษาในฐานข้อมูลสุ่มแบบมีเงื่อนไขได้ซึ่งสอดคล้องกับหลักการของการวิเคราะห์การถดถอย กล่าวคือ เมื่อฐานข้อมูลสุ่มมีจำนวนเพิ่มมากขึ้น แบบจำลองการวิเคราะห์การถดถอยพหุคูณก็จะมี ความแม่นยำในการทำนายเพิ่มมากขึ้นด้วย แต่อย่างไรก็ตามถ้ามีการเปลี่ยนแปลงของข้อมูลเกิดขึ้น หรือมีการกำหนดเกณฑ์เงื่อนไขช่วงคะแนนของความสัมพันธ์ระหว่างผลการศึกษากับปัจจัยที่มีผลต่อผลการศึกษาแตกต่างไปจากในบทความนี้ ก็จะทำให้เปอร์เซ็นต์ความถูกต้องในการทำนายเปลี่ยนไปด้วย การทดลองที่นำเสนอในบทความนี้เป็นการตรวจสอบคุณภาพเครื่องมือในการวิจัยก่อนการนำไปใช้งานในงานวิจัยอื่นๆ ต่อไป ซึ่งสอดคล้องกับงานวิจัยเรื่อง เครื่องมือที่ใช้ในการวิจัยของ จิตติรัตน์ แสงเลิศอุทัย [11] และเรื่อง การตรวจสอบคุณภาพเครื่องมือวิจัยของอารยา องค์เอี่ยม และพงศ์ธรา วิจิตเวชไพศาล [12]

7. ข้อเสนอแนะ

บทความนี้เป็นการนำเสนอรูปแบบวิธีการในการตรวจสอบความน่าเชื่อถือของเครื่องมือวิจัยแบบจำลองการวิเคราะห์การถดถอยพหุคูณเพื่อให้มั่นใจว่าสามารถนำเครื่องนี้ไปใช้ต่อไปในบริบทของการศึกษาของโรงเรียนนายเรือได้ ซึ่งในการทดสอบเครื่องมือวิจัยนี้จำเป็นต้องใช้ฐานข้อมูลที่มีปริมาณมาก หากเก็บข้อมูล

จากนักเรียนนายเรือโดยตรงอาจต้องใช้เวลานาน เนื่องจากจำนวนนักเรียนนายเรือในแต่ละปีมีจำนวนไม่มาก จึงได้นำฐานข้อมูลที่สร้างจากการเขียนโปรแกรมคอมพิวเตอร์แบบสุ่มค่ามาใช้ จึงอาจทำให้ผลที่ได้มีความคลาดเคลื่อนในการทำนายผลการเรียนจริง หากได้ใช้ฐานข้อมูลจริงที่ได้จากการเก็บข้อมูลจากกลุ่มตัวอย่างหรือประชากรก็จะได้ผลการวิจัยที่น่าเชื่อถือมากยิ่งขึ้น นอกจากนี้การวิเคราะห์ข้อมูลด้วยการใช้ผลการศึกษาในรูปของเกรดรายวิชาอาจจะมีข้อเสียคือคือ เกรดที่ใช้จะมีช่วงกว้าง ผลการวิเคราะห์ที่ได้อาจจะไม่ชัดเจน หากปรับการใช้ผลการเรียนเป็นเปอร์เซ็นต์ อาจจะทำให้สามารถประเมินความแม่นยำของการใช้เครื่องมือได้มากขึ้น

เอกสารอ้างอิง

- [1] ศราวุธ แดงมาก. โลกแห่งดิจิทัล ยุคแห่งข้อมูลข่าวสาร ความมหาศาสตร์ของข้อมูลทางการแพทย์สู่การค้นพบความรู้ที่มีค่าด้วยการทำเหมืองข้อมูล. เวชบัณฑิตศิริราช. ม.ค.-เม.ย. 2560;10(1):29-33.
- [2] กรมวิชาการ กระทรวงศึกษาธิการ. ประมวลคำศัพท์ทางการศึกษา. กรุงเทพฯ: สำนักพิมพ์คุรุสภา; 2515.
- [3] ไพศาล หวังพานิชย์. หลักและวิธีการประเมินโครงการ. วารสารวัดผลการศึกษา. ม.ค.-เม.ย. 2533:23-38.
- [4] นงลักษณ์ วิรัชชัย. การวิจัยหลักสูตรและการเรียนการสอน [วิทยานิพนธ์ปริญญาโทมหาบัณฑิต]. กรุงเทพฯ: มหาวิทยาลัยสุโขทัยธรรมาธิราช; 2553.
- [5] ธวัชพร มะรินทร์. ปัจจัยที่ส่งผลต่อผลสัมฤทธิ์ทางการเรียนของนักเรียนโรงเรียนนายเรือ [วิทยานิพนธ์ปริญญาโทมหาบัณฑิต]. กรุงเทพฯ: สถาบันบัณฑิตพัฒนบริหารศาสตร์; 2559.
- [6] กฤษฎา บุญวัฒน์. ปัจจัยที่มีความสัมพันธ์กับผลสัมฤทธิ์ทางการเรียนของนักเรียนโรงเรียนเตรียมทหาร [ภาคินิพนธ์ปริญญาโทมหาบัณฑิต]. กรุงเทพฯ: สถาบันบัณฑิตพัฒนบริหารศาสตร์; 2541.
- [7] ฝ่ายศึกษา โรงเรียนนายเรือ. หลักสูตรการศึกษาโรงเรียนนายเรือ พ.ศ. 2563 (ฉบับปรับปรุง 2558) [อินเทอร์เน็ต]. สมุทรปราการ: โรงเรียนนายเรือ; 2563. [สืบค้นเมื่อวันที่ 26 ต.ค. 2563]. จาก: www.rtna.ac.th
- [8] ฐณัฐ วงศ์สายเชื้อ. การวิเคราะห์การถดถอยพหุ (Multiple Regression analysis) [อินเทอร์เน็ต]. กรุงเทพฯ: มหาวิทยาลัยเกษตรศาสตร์; 2558. [สืบค้นเมื่อวันที่ 5 พ.ย. 2563]. จาก: http://pirun.ku.ac.th/~fedutnw/pubs/da7010_multiple%20regression19.pdf
- [9] สุทิน ชนะบุญ. การวิเคราะห์การถดถอยพหุคูณ (Multiple Linear Regression) [อินเทอร์เน็ต]. ขอนแก่น: สำนักงานสาธารณสุขจังหวัดขอนแก่น; 2563. [สืบค้นเมื่อวันที่ 5 พ.ย. 2563]. จาก: <http://www.kkpho.go.th/i/index.php/component/attachments/download/1933#>
- [10] 6 โปรแกรม สำหรับใช้วิเคราะห์ข้อมูลในการทำวิจัย [อินเทอร์เน็ต]. กรุงเทพฯ: Advanced Research Group; 2561. [สืบค้นเมื่อวันที่ 10 พ.ย. 2563]. จาก <https://www.arresearch.com/a/44271>

- [11] กัลยา วาณิชย์บัญชา. การวิเคราะห์ข้อมูลด้วย SPSS for Windows. กรุงเทพฯ: โรงพิมพ์ จุฬาลงกรณ์มหาวิทยาลัย; 2542.
- [12] จิตติรัตน์ แสงเลิศอุทัย. เครื่องมือที่ใช้ในการวิจัย. วารสารบัณฑิตศึกษา. ก.ค.-ก.ย. 2558;12 (58):13-24.
- [13] อารยา องค์กรเยี่ยม, พงศ์ธรา วิจิตเวชไพศาล. การตรวจสอบคุณภาพเครื่องมือวิจัย. Thai J Anesthesiol. 2561;44(1):36-42.