บทที่ 3

วิธีการดำเนินการวิจัย

การวิจัยครั้งนี้เป็นการวิจัยเชิงบรรยาย (Descriptive research) โดยใช้ฐานข้อมูลทุติยภูมิ (Secondary Data) จากทะเบียนประวัติของนักศึกษามหาวิทยาลัยอุบลราชธานีปีการศึกษา 2544 - 2546 มีจุดมุ่งหมายเพื่อประยุกต์ใช้เทคนิคคลีเมนไทน์ในการพยากรณ์เกรดเฉลี่ย พร้อมทั้งตรวจสอบผลการพยากรณ์ด้วยเทคนิคคลีเมนไทน์กับเทคนิคการถดถอยพหุคูณ โดยใช้เกณฑ์ในการเปรียบเทียบจากการวัดความคลาดเคลื่อน 3 แบบคือ รากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Square Error หรือ RMSE) ค่ามัธยฐานของค่าสัมบูรณ์ของความคลาดเคลื่อนวัดในรูปร้อยละ (Median Absoluate Percentage Error หรือ MdAPE) และ ร้อยละที่ดีกว่า (Percent Better) ผู้วิจัยนำเสนอรายละเอียดเกี่ยวกับลักษณะแหล่งที่มาของข้อมูล การเก็บรวบรวมข้อมูล ตัวแปรที่ใช้ในการศึกษา และการวิเคราะห์ข้อมูลที่ได้ดำเนินการไว้ ดังนี้

ลักษณะและแหล่งที่มาของข้อมูล

ข้อมูลที่ใช้ในการวิจัยครั้งนี้เป็นข้อมูลจากแหล่งทุติยภูมิ โดยฐานข้อมูล (Database) ที่ใช้เป็นข้อมูลทะเบียนประวัตินักศึกษา ได้แก่ เกรดเฉลี่ยสะสมเมื่อศึกษาอยู่ในชั้นปีที่ 1 เกรดเฉลี่ยสะสมเมื่อสำเร็จการศึกษาระดับมัธยมศึกษาตอนปลาย คะแนนสอบคัดเลือก เพศ ภูมิลำเนา การศึกษาของบิดา การศึกษาของมารดา อาชีพของบิดา อาชีพของมารดา รายได้ต่อเดือนของครอบครัว (บิดามารดารวมกัน) จำนวนพี่น้องในครอบครัว วิธีรับเข้า และสังกัดคณะ โดยใช้ข้อมูลของนักศึกษาที่รับเข้าในปีการศึกษา 2544-2546

แหล่งที่มาของการเก็บรวบรวมข้อมูลสำหรับการวิจัยครั้งนี้ คือข้อมูลทะเบียนประวัติของนักศึกษาที่รับเข้าศึกษาในปีการศึกษา 2544-2546 ของงานทะเบียนประวัติและประมวลผล กองบริการการศึกษา สำนักงานอธิการบดี มหาวิทยาลัยอุบลราชธานี

การเก็บรวบรวมข้อมูล

ในการวิจัยครั้งนี้ ผู้วิจัยบันทึกข้อมูลด้วยตนเอง โดยมีขั้นตอนการรวบรวมข้อมูลดังต่อไปนี้

1. การติดต่อขอความร่วมมือ ผู้วิจัยดำเนินการติดต่อขอความร่วมมือไปยังผู้อำนวยการกองบริการการศึกษา กองบริการการศึกษา สำนักงานอธิการบดี มหาวิทยาลัยอุบลราชธานี เพื่อขอบันทึกข้อมูลทะเบียนประวัติของนักศึกษาที่รับเข้าในปีการศึกษา 2544-2546 แล้วผู้อำนวยการกองบริการการศึกษาดำเนินการประสานงานไปยังเจ้าหน้าที่ฝ่ายทะเบียนประวัติและประมวลผล เพื่อให้ผู้วิจัยได้สำรวจและบันทึกข้อมูลในขั้นตอนต่อไป

2. การสำรวจ และบันทึกข้อมูล โดยสำรวจและบันทึกข้อมูลทะเบียนประวัตินักศึกษา จากงานทะเบียนประวัติและประมวลผล กองบริการการศึกษา มหาวิทยาลัยอุบลราชธานี ได้แก่ รหัสนักศึกษา ชื่อ-สกุล เพศ เลขที่บัตรประจำตัวประชาชน การสังกัดคณะ/สาขาวิชา ประเภทของวิธีรับเข้า คะแนนสอบคัดเลือก เกรดเฉลี่ยสะสมเมื่อสำเร็จการศึกษามัธยมศึกษาตอนปลาย ภูมิลำเนาการศึกษาของบิดามารดา อาชีพของบิดามารดา รายได้ของครอบครัว(บิดามารดารวมกัน) จำนวนพี่น้องในครอบครัว

3. การบรรณาธิกรณ์ (Edit) ข้อมูล ผู้วิจัยตรวจสอบความถูกต้องและความสมบูรณ์ของข้อมูลที่ได้จากการรวบรวมข้อมูลมาทั้งหมด โดยใช้โปรแกรม Microsoft Excel และ Spss for Window

การเก็บรวบรวมข้อมูลทั้ง 3 ขั้นตอนผู้วิจัยใช้เวลาในการดำเนินการทั้งสิ้นประมาณ 2 สัปดาห์ คือในช่วงระหว่างวันที่ 11-25 กุมภาพันธ์ 2548

ตัวแปรที่ใช้ในการวิจัย

ตัวแปรที่ใช้ในการวิจัยครั้งนี้ประกอบด้วยตัวแปรตาม (Dependent Variable) 1 ตัวคือ เกรดเฉลี่ยสะสมนักศึกษาเมื่อศึกษาชั้นปีที่ 1 ในภาคเรียนที่ 2 ของนักศึกษาที่รับเข้าปีการศึกษา 2544-2546 โดยวัดด้วยมาตรอันตรภาค (Interval Scale) และสำหรับตัวแปรอิสระ (Independent Variable) มี 12 ตัววัดด้วยมาตรอันตรภาค (Interval Scale) 2 ตัว ได้แก่ ตัวแปรเกรดเฉลี่ยสะสมเมื่อสำเร็จการศึกษาระดับมัธยมศึกษาตอนปลาย และ ตัวแปรคะแนนสอบคัดเลือก วัดด้วยมาตรอัตราส่วน (Ratio Scale) 2 ตัวได้แก่ ตัวแปรรายได้ต่อเดือนของครอบครัว (บิดามารดารวมกัน) และตัวแปรจำนวนพี่น้องในครอบครัว และวัดด้วยมาตรนามบัญญัติ (Nomial Scale) 8 ตัวได้แก่ ตัวแปรเพศ ตัวแปรตัวแปรภูมิลำเนา ตัวแปรการศึกษาของบิดา ตัวแปรการศึกษาของมารดา ตัวแปรอาชีพของบิดา ตัวแปรอาชีพของมารดา ตัวแปรวิธีรับเข้าและตัวแปรการสังกัดคณะ

สำหรับการพยากรณ์โดยนิวรอลเน็ตเวิร์กนั้น ข้อมูลป้อนเข้า (Input) จะเปรียบเสมือนตัวแปรอิสระ และผลลัพธ์ (Output) จะเปรียบเสมือนเป็นตัวแปรตาม โดยนิวรอล (Neural) ที่จะป้อนค่าข้อมูล

ป้อนเข้าแต่ละตัวจะเป็นตัวแประอิสระแต่ละตัว ดังนั้นในการวิเคราะห์โดยเทคนิคคลีเมนไทน์จะประกอบตัวตัวแปร 2 กลุ่มคือ

1. จำนวนข้อมูลป้อนเข้า (Input) หรือตัวแปรอิสระ 39 ตัวแปร เพื่อใช้ในการเปรียบเทียบกับการพยากรณ์โดยเทคนิคการถดถอยพหุคูณวิธีการคัดเลือกตัวแปรเพื่อการพยากรณ์แบบปกติ (Enter Multiple Regression Analysis)

2. จำนวนข้อมูลป้อนเข้า (Input) หรือตัวแปรอิสระ 12 ตัวแปร เพื่อใช้ในการเปรียบเทียบกับการพยากรณ์โดยเทคนิคการถดถอยพหุคูณวิธีการคัดเลือกตัวแปรเพื่อการพยากรณ์แบบขั้นตอน (Stepwise Multiple Regression Analysis)

โดย 2 กลุ่มใช้ข้อมูลผลลัพธ์ (Output) หรือตัวแปรตาม 1 ตัวแปรซึ่งเป็นตัวเดียวกัน

การวิเคราะห์ข้อมูล

สำหรับขั้นตอนการวิเคราะห์แบ่งเป็น 4 ตอน ตอนที่ 1 การวิเคราะห์ข้อมูลเบื้องต้น ตอนที่ 2 วิเคราะห์ด้วยเทคนิคการถดถอยพหุคูณ และตอนที่ 4 การตรวจสอบผลการพยากรณ์ ใช้โปรแกรม SPSS for Windows และ Microsoft Excel ส่วนในตอนที่ 3 การวิเคราะห์ด้วยเทคนิคคลีเมนไทน์ใช้โปรแกรม Clementine version 7.0 การวิเคราะห์ข้อมูลในแต่ละตอน มีรายละเอียดดังนี้

ตอนที่ 1 การวิเคราะห์ข้อมูลเบื้องต้น

เพื่อศึกษาสถานภาพทั่วไปของนักศึกษา ซึ่งได้แก่ ความถี่ ร้อยละ ค่าเฉลี่ย ค่าสูงสุด ค่าต่ำสุด ค่าความเบ้ ค่าความโด่ง ค่าการกระจายของข้อมูล และค่าเบี่ยงเบนมาตรฐาน ของตัวแปรที่ใช้ศึกษา

ตอนที่ 2 การวิเคราะห์ด้วยเทคนิคการถดถอยพหุคูณ

ใช้โปรแกรม SPSS for Windows โดยขั้นตอนในการวิเคราะห์ มี 2 ขั้นตอนดังนี้

ขั้นตอนที่ 1 นำตัวแปรอิสระที่มีมาตรการวัดเป็นแบบอันตรภาคและอันตราส่วน และนามบัญญัติทุกตัวเข้าไปในสมการ สำหรับตัวแปรอิสระที่มีมาตรการวัดเป็นแบบนามบัญญัติ จะทำการแปลงเป็นตัวแปรหุ่น (Dummy Variables) ดังตารางที่ 3.1 วิเคราะห์ความสัมพันธ์เพียร์สันระหว่างตัวแปรที่ใช้ศึกษา แล้วคัดเลือกตัวแปรที่มีความสำคัญต่อเกรดเฉลี่ยสะสมเมื่อศึกษาอยู่ชั้นปีที่ 1 โดยวิธีการถดถอยพหุแบบปกติและแบบขั้นตอน

ตารางที่ 3.1 การกำหนดตัวแปรหุ่น (Dummy Variables)

ตัวแปร	รหัส	ตัวแปรหุ่น	
เพศ	SEX	1= ชาย	0= หญิง
ภูมิลำเนา	PRO	1= จังหวัดอุบลราชธานี	0= อื่นๆ
การศึกษาของบิดา	EDU_F1	1= ประถมศึกษา	0= อื่นๆ
	EDU_F2	1= มัธยมศึกษา	0= อื่นๆ
	EDU_F3	1= อาชีวศึกษา	0= อื่นๆ
	EDU_F4	1= อนุปริญญา	0= อื่นๆ
	EDU_F5	1= ปริญญาตรี	0= อื่นๆ
	EDU_F6	1= ปริญญาโทขึ้นไป	0= อื่นๆ
	EDU_F7	1= ไม่ได้ศึกษา	0= อื่นๆ
การศึกษาของมารดา	EDU_M1	1= ประถมศึกษา	0= อื่นๆ
	EDU_M2	1= มัธยมศึกษา	0= อื่นๆ
	EDU_M3	1= อาชีวศึกษา	0= อื่นๆ
	EDU_M4	1= อนุปริญญา	0= อื่นๆ
	EDU_M5	1= ปริญญาตรี	0= อื่นๆ
	EDU_M6	1= ปริญญาโทขึ้นไป	0= อื่นๆ
	EDU_M7	1= ไม่ได้ศึกษา	0= อื่นๆ
อาชีพของบิดา	OC_F1	1= ข้าราชการ	0= อื่นๆ
	OC_F2	1= พนักงานรัฐวิสาหกิจ	0= อื่นๆ
	OC_F3	1= ลูกจ้างเอกชน	0= อื่นๆ
	OC_F4	1= เจ้าของกิจการค้าขาย	0= อื่นๆ
	OC_F5	1= เกษตรกรรม/ประมง	0= อื่นๆ
	OC_F6	1= รับจ้าง	0= อื่นๆ
	OC_F7	1= ไม่ประกอบอาชีพ	0= อื่นๆ
	OC_F8	1= ลูกจ้างราชการ	0= อื่นๆ

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 3.1(ต่อ) การกำหนดตัวแปรหุ่น (Dummy Variables)

ตัวแปร	รหัส	ตัวแปรหุ่น	
อาชีพมารดา	OC_M1	1= ข้าราชการ	0= อื่นๆ
	OC_M2	1= พนักงานรัฐวิสาหกิจ	0= อื่นๆ
	OC_M3	1= ลูกจ้างเอกชน	0= อื่นๆ
	OC_M4	1= เจ้าของกิจการค้าขาย	0= อื่นๆ
	OC_M5	1= เกษตรกรรม/ประมง	0= อื่นๆ
	OC_M6	1= รับจ้าง	0= อื่นๆ
	OC_M7	1= ไม่ประกอบอาชีพ	0= อื่นๆ
	OC_M8	1= ลูกจ้างราชการ	0= อื่นๆ
วิธีรับเข้า	AD_1	1= วิธีผ่านระบบกลาง	0= อื่นๆ
	AD_2	1= วิธีรับตรง(โควตา)	0= อื่นๆ
	AD_3	1= วิธีพิเศษ	0= อื่นๆ
สังกัดคณะ	FAC_1	1= คณะวิทยาศาสตร์	0= อื่นๆ
	FAC_2	1= คณะเกษตรศาสตร์	0= อื่นๆ
	FAC_3	1= คณะวิศวกรรมศาสตร์	0= อื่นๆ
	FAC_4	1= คณะศิลปศาสตร์	0= อื่นๆ
	FAC_5	1= คณะเภสัชศาสตร์	0= อื่นๆ
	FAC_6	1= คณะบริหารศาสตร์	

ขั้นตอนที่ 2 ตรวจสอบข้อตกลงเบื้องต้นของการวิเคราะห์การถดถอยพหุคูณ ซึ่งในการวิเคราะห์ถดถอยพหุคูณมีข้อตกลงเบื้องต้นคือ

ข้อที่ 1 ค่าเฉลี่ยของความคลาดเคลื่อนมีค่าเป็นศูนย์

ข้อที่ 2 ความแปรปรวนของความคลาดเคลื่อนมีค่าคงที่

ข้อที่ 3 ความคลาดเคลื่อนเป็นค่าสุ่มที่เป็นอิสระต่อกัน

ข้อที่ 4 ความคลาดเคลื่อนมีการแจกแจงปกติ

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตรวจสอบข้อตกลงเบื้องต้น ดังนี้

1. ตรวจสอบความคลาดเคลื่อนมีความเป็นอิสระต่อกันและมีความแปรปรวนคงที่จากสมการที่ได้ นำตัวแปรอิสระที่ได้รับการคัดเลือกเข้าสมการไปแทนค่าหา $\hat{Y}_i$ และ $e_i = Y_i - \hat{Y}_i$ วิเคราะห์ความผิดปกติของค่าความคลาดเคลื่อนที่ไม่เป็นอิสระโดยการเขียนกราฟ Standardized ของคู่ลำดับ($\hat{Y}_i$, e_i) หรือตรวจสอบจากค่า Durbin Watson test ซึ่งถ้ามีค่าอยู่ระหว่าง 1.5 ถึง 2.5 แสดงว่าความคลาดเคลื่อนมีความเป็นอิสระต่อกัน (กัลยา วานิชย์บัญชา, 2545)

2. ตรวจสอบการแจกแจงปกติของความคลาดเคลื่อน โดยทำ Normal Probability Plot โดยการเขียนกราฟระหว่าง $P\left(e_i \le \frac{i}{n}\right)$ เมื่อ i เป็นลำดับที่ของ e_i เมื่อจัดลำดับแล้วกับ $P\left(Z < \frac{e_i}{\sqrt{MSE}}\right)$ หากพบว่า จุดต่างๆที่ได้เป็นแนวทแยงมุมแสดงว่าความคลาดเคลื่อนมีการแจกแจงปกติ

3. ตรวจสอบการมีความสัมพันธ์กันเองระหว่างตัวแปรอิสระ (Multicollinearity) โดยพิจารณาค่าแฟคเตอร์ความแปรปรวนที่เพิ่มมากขึ้น (Variance Inflation Factor: VIF)

เมื่อ $(VIF)_j$ = Variance Inflation Factor สำหรับ b_j

$= \frac{1}{1-R_j^2}$: j = 1,2,3,...,p-1

R_j^2 = ค่าสัมประสิทธิ์ของตัวกำหนดพหุคูณของสมการถดถอยที่มี X_j เป็นตัวแปรตาม และมีตัวแปร $X_1, X_2, \ldots, Xx_{j-1}, X_{j+1}, \ldots, X_{p-1}$ เป็นตัวแปรอิสระ ถ้า R_j^2 =0 และ $(VIF)_j$=1 แสดงว่าไม่มีปัญหาตัวแปรอิสระมีความสัมพันธ์กันเอง และถ้า R_j^2 มีค่าเข้าใกล้ 1 $(VIF)_j$ มีค่ามากกว่า 1 แสดงว่าเกิดปัญหาตัวแปรอิสระมีความสัมพันธ์กันเอง และถ้า $(VIF)_j$> 10 ถือว่าเป็นปัญหารุนแรงต้องมีการแก้ไข (Marquardt, 1975)

หรือตรวจสอบด้วยค่า Tolerance , Condition Index ค่าสถิติ Tolerance หมายถึง สัดส่วนความแปรปรวนในตัวแปรที่อธิบายไม่ได้ด้วยตัวแปรอื่น ๆ ถ้าค่า Tolerance มีค่าใกล้ 0 แสดงว่าตัวแปรนั้นมีความสัมพันธ์กันสูงกับตัวแปรอื่น ๆ ค่าที่ใกล้ 1 แสดงว่าตัวแปรนั้นมีความสัมพันธ์ต่ำกับตัวแปรอื่นๆ ในการวิเคราะห์ข้อมูลค่าสถิติ VIF มีค่าเท่ากับส่วนกลับของค่า Tolerance จึงแปลความหมายตรงกันข้าม กล่าวคือค่า VIF ที่สูงมาก (ค่าสูงสุดเท่ากับ 10.0) แสดงว่ามีปัญหาภาวะร่วมเส้นตรงพหุสูงมาก

ในกลุ่มตัวแปรอิสระด้วยกัน ส่วนค่าสถิติ Condition Index เป็นค่าสัดส่วนความแปรปรวนซึ่งวัดจากค่าไอเกน (eigen value) เกณฑ์ที่ใช้ตรวจสอบคือ 30 ถ้าตัวแปรใดมีค่า Condition Index เกิน 30 แสดงว่าตัวแปรนั้นมีปัญหาภาวะร่วมเส้นตรงพหุ ค่าสถิติ Condition Indexนี้ใช้ตรวจสอบด้วยกระบวนการสองขั้นตอน ขั้นตอนแรกตรวจดูว่าตัวแปรต้นตัวใดมีค่าเกินกว่าเกณฑ์ ขั้นตอนที่สอง ตรวจสอบสัดส่วนของความแปรปรวนของสัมประสิทธิ์การถดถอย เฉพาะเมื่อตัวแปรอิสระมีค่าค่าสถิติ Condition Index สูงกว่าเกณฑ์ ถ้าค่าสัดส่วนของความแปรปรวนของสัมประสิทธิ์การถดถอยมีค่าสูงกว่า .90 แสดงว่าตัวแปรตัวนั้นมีความสัมพันธ์กับตัวแปรอื่นสูง

สำหรับเทคนิคการถดถอยพหุคูณโดยวิธีการคัดเลือกตัวแปรเพื่อการพยากรณ์แบบปกติ (Enter or All possible Multiple Regression Anaysis) ผู้วิจัยจะคัดเลือกตัวแปรทั้งหมดเข้าในสมการเพื่อต้องการศึกษาถึงปริมาณจำนวนตัวแปรอิสระที่มากซึ่งจะทำให้ค่าสัมประสิทธิ์การพยากรณ์มากด้วยนั้น จะมีผลต่อข้อตกลงเบื้องต้นอย่างไร และผลการพยากรณ์ที่ได้หากฝ่าฝืนข้อตกลงจะได้ผลเป็นอย่างไร ทั้งนี้เพื่อนำไปเป็นข้อมูลในการเปรียบเทียบผลการพยารกรณ์กับเทคนิคคลีเมนไทน์

สำหรับเทคนิคการถดถอยพหุคูณโดยวิธีการคัดเลือกตัวแปรเพื่อการพยากรณ์แบบขั้นตอน (Stepwise Multiple Regression Anaysis) ผู้วิจัยจะให้โปรแกรม Spss ประมวลผลให้โดยจะให้ตัวแปรอิสระทั้งหมด และโปรแกรมจะคำนวณคัดเลือกตัวแปรแบบขั้นตอน ซึ่งผลที่ได้จะเหลือตัวแปรอิสระที่อยู่ในสมการพยากรณ์ที่ดีที่สุดและเป็นไปตามข้อตกลงเบื้องต้น ซึ่งวิธีนี้นอกจากจะได้ผลการวิเคราะห์ที่ไม่ฝ่าฝืนข้อตกลงแล้วตัวแปรอิสระที่เหลือที่ลดลงจะเป็นสิ่งที่ผู้วิจัยสนใจ ที่จะนำไปเปรียบเทียบผลการพยากรณ์โดยเทคนิคของคลีเมนไทน์เมื่อจำนวนตัวแปรอิสระที่จะทำนายน้อย

ตอนที่ 3 วิเคราะห์ข้อมูลด้วยเทคนิคคลีเมนไทน์ โดยโปรแกรม Clementine มีขั้นตอนในการวิเคราะห์ในแต่ละตอนดังนี้

3.1 การวิเคราะห์เทคนิคคลีเมนไทน์ด้วยข้อมูลป้อนเข้า 39นิวรอล

ใช้การวิเคราะห์แบบนิวรอลเน็ตเวิร์กโดยจำนวนข้อมูลป้อนเข้าเหมือนและเท่ากับตัวแปรอิสระจำนวนตัวแปรอิสระที่ใช้การวิเคราะห์การถดถอยพหุคูณวิธีคัดเลือกตัวแปรเพื่อการพยากรณ์แบบปกติ(Enter Multiple Regression)

ขั้นที่ 1 ถ่ายโอนข้อมูลที่ใช้ในการวิเคราะห์จากไฟล์ที่บันทึกโดยโปรแกรม SPSS แล้วกำหนดค่าตัวแปรมาเป็นข้อมูลป้อนเข้า (Input) และข้อมูลผลลัพธ์ (Output) โดยตัวแปรอิสระทั้ง 39 ตัวคือ เกรดเฉลี่ยสะสมเมื่อสำเร็จการศึกษาระดับมัธยมศึกษาตอนปลาย คะแนนสอบคัดเลือก เพศ ภูมิลำเนา การศึกษาของบิดา (ตัวแปรหุ่น 6 ตัวแปร) การศึกษาของมารดา (ตัวแปรหุ่น 6

ตัวแปร) อาชีพของบิดา(ตัวแปรหุ่น 7 ตัวแปร) อาชีพของมารดา (ตัวแปรหุ่น 7 ตัวแปร) รายได้ต่อเดือนของครอบครัว(บิดามารดารวมกัน) จำนวนพี่น้องในครอบครัว วิธีรับเข้า (ตัวแปรหุ่น 2 ตัวแปร) และสังกัดคณะ (ตัวแปรหุ่น 5 ตัวแปร) จะเป็นข้อมูลป้อนเข้า (Input) และตัวแปรตาม 1 ตัวคือเกรดเฉลี่ยสะสมเมื่อศึกษาอยู่ในชั้นปีที่ 1 จะเป็นข้อมูลผลลัพธ์ (Output)

ขั้นที่ 2 การกำหนดค่าเริ่มต้นให้ตัวแบบนิวรอลเน็ตเวิร์กทำงาน ดังนี้

1. แบบของนิวรอลเน็ตเวิร์กประกอบด้วย ชั้นข้อมูลป้อนเข้า (Input Node Layer) ชั้นแอบแฝง (Hidden Node Layer) และชั้นแสดงผลลัพธ์ (Output Node Layer) แบ่งการกำหนดค่าในแต่ละชั้นดังนี้

1.1 ชั้นข้อมูลป้อนเข้า (Input Node Layer) ประกอบด้วยจำนวนข้อมูลป้อนเข้า (Input) 39 นิวรอล หรือ จำนวนตัวแปรอิสระ 39 ตัวแปร

1.2 ชั้นแอบแฝง (Hidden Node Layer) ประกอบด้วยจำนวนนิวรอลเท่ากับ $\sqrt{MN}$ นิวรอล (Masters, 1993) เมื่อ M เป็นจำนวนนิวรอลของชั้นข้อมูลป้อนเข้า ในที่นี้คือจำนวนตัวแปรอิสระ ซึ่งมีค่าเท่ากับ 39 และ N เป็นจำนวนนิวรอลของชั้นแสดงผลลัพธ์ ซึ่งในที่นี้คือจำนวนตัวแปรตาม ซึ่งมีค่าเท่ากับ 1 ดังนั้น ชั้นแอบแฝงจะประกอบด้วยนิวรอลจำนวน $\sqrt{39\times 1} \approx 6$ นิวรอล ซึ่งสูตรนี้เป็นการประมาณขนาดชั้นแอบแฝง แต่การกำหนดจำนวนนิวรอลในชั้นแอบแฝง อาจกำหนดได้โดย

- ถ้ามีจำนวนนิวรอลนำเข้าและนิวรอลแสดงผลลัพธ์น้อยๆ และปัญหามีความซับซ้อน นิวรอลในชั้นแอบแฝงอาจต่ำกว่าความต้องการ
- ถ้าปัญหาเป็นปัญหาที่ง่าย มีจำนวนนิวรอลนำเข้าและ นิวรอลแสดงผลลัพธ์มาก จำนวนนิวรอลในชั้นแอบแฝงน้อยๆ ก็เพียงพอ
- กฎนี้ไม่ควรนำไปใช้กับเน็ตเวิร์กที่มีจำนวนนิวรอลนำเข้าเท่ากับจำนวนนิวรอลแสดงผลลัพธ์ (Autoassociative Networks)

1.3 ชั้นแสดงผลลัพธ์ (Output Node Layer) ประกอบด้วยจำนวนตัวแปรตามนั่นคือ 1 ตัว (1 นิวรอล) ซึ่งแสดงถึง เกรดเฉลี่ยสะสมเมื่อศึกษาอยู่ในชั้นปีที่ 1

2. โมเดลที่ใช้ในการวิจัยจะใช้โมเดลของการเรียนรู้แบบย้อนกลับที่มีการเชื่อมต่อกันอย่างทั่วถึง (Fully Connected Backpropagation)

3. ฟังก์ชันการแปลงค่า ใช้ฟังก์ชันเชิงเส้นสำหรับชั้นข้อมูลป้อนเข้า และชั้นแสดงผลลัพธ์ และใช้ฟังก์ชันซิกมอยด์ (Sigmoid Function) ในการแปลงค่าสำหรับชั้นแอบแฝง

4. กำหนดค่าความคลาดเคลื่อนของค่าผลลัพธ์(Output) ที่ยอมรับได้

(RMS คือ น้อยกว่า 0.0001 (error<1%)

5. กำหนดค่าของสัมประสิทธิ์การเรียนรู้ (η) และค่าโมเมนตัม (α) ตามจำนวนรอบการสอน โดยที่ค่า η ควรเริ่มจากมากไปน้อยและอยู่ในช่วง 0.75-0.25 และ α ควรเริ่มจากน้อยไปมากและอยู่ในช่วง 0.5-0.9 (Paterson, 1996) ดังตารางที่ 3.2

ตารางที่ 3.2 ค่าของสัมประสิทธิ์การเรียนรู้ (η) และค่าโมเมนตัม (α) ตามจำนวนรอบการสอน

ชนิดค่า	จำนวนรอบการสอน(รอบ)				
	5000	10000	15000	20000	30000
η	0.65	0.55	0.45	0.35	0.25
α	0.5	0.6	0.7	0.8	0.9

ขั้นที่ 3 การสอนนิวรอลเน็ตเวิร์ก เมื่อกำหนดรูปแบบของนิวรอลเน็ตเวิร์ก และค่าตัวแปรต่างๆ แล้ว ทำการสอนโดยเริ่มต้นจากค่าน้ำหนักให้กับเน็ตเวิร์ก โดยการสุ่มค่าให้อยู่ในช่วง – 0.1 ถึง 0.1 กำหนดชุดการสอนโดยกำหนดชุดข้อมูลป้อนเข้าและระบุผลลัพธ์ที่ต้องการ แล้วเน็ตเวิร์กจะเริ่มเรียนรู้ด้วยการคำนวณผลลัพธ์จากฟังก์ชันซิกมอยด์ แล้วเปรียบเทียบผลลัพธ์ที่ได้จากการคำนวณกับผลลัพธ์ที่ต้องการ ถ้ามีความแตกต่างกันจะทำการปรับค่าน้ำหนักของเน็ตเวิร์กจะคงที่ ถ้าค่าน้ำหนักของชุดเน็ตเวิร์กคงที่คือ ถ้าน้อยกว่าหรือเท่ากับ 0.0001 ก็จะยุติการสอน (นราพัฒน์ ลิมปนากร, 2545) หรือค่าความคลาดเคลื่อนที่ได้มีค่าน้อยกว่าที่กำหนดไว้หรือจำนวนรอบครบตามจำนวนครั้งที่ได้กำหนดไว้ ก็จะยุติการทำงานเช่นกัน ในที่นี้ผู้วิจัยได้กำหนดจำนวนรอบการเรียนรู้เป็นการยุติการทำงาน

ขั้นที่ 4 คำนวณค่าความแม่นยำการประมาณค่า (Estimated Accuracy) โดยมีสูตรในการคำนวณคือ $0.5 - \left|\frac{Y-\hat{Y}}{S}\right| \times 100$ โดยค่าที่ได้จะเป็นเปอร์เซ็นต์ และทำการคำนวณค่าความสัมพันธ์ที่สำคัญของข้อมูลที่ป้อนเข้าต่อการพยากรณ์ (Relative Importance of Output) โดยมีค่าระหว่าง 0 ถึง 1 ถ้าข้อมูลป้อนเข้าตัวใดมีค่ามากแสดงว่ามีอิทธิพลต่อการพยากรณ์มาก

3.2 การวิเคราะห์เทคนิคคลีเมนไทน์ ด้วยข้อมูลป้อนเข้า 12 นิวรอล

ใช้การวิเคราะห์แบบนิวรอลเน็ตเวิร์กโดยจำนวนข้อมูลป้อนเข้าเหมือนกับจำนวนตัวแปรอิสระที่ใช้การวิเคราะห์การถดถอยโดยวิธีการวิเคราะห์ถดถอยพหุคูณแบบเพิ่มตัวแปรเป็นขั้นตอน โดยวิธีการและขั้นตอนจะเหมือนกับตอนที่ 3.1 แต่เปลี่ยนจำนวนข้อมูลป้อนเข้าให้เป็นตัวแประอิสระที่เหลือที่ได้จากการการวิเคราะห์ถดถอยพหุคูณแบบขั้นตอน เป็น 12 โหนด ดังนี้

ขั้นที่ 1 ถ่ายโอนข้อมูลที่ใช้ในการวิเคราะห์จากไฟล์ที่บันทึกโดยโปรแกรม SPSS แล้วกำหนดค่าตัวแปรมาเป็นข้อมูลป้อนเข้า (Input) และข้อมูลผลลัพธ์ (Output) โดยตัวแปรอิสระทั้ง 12 ตัวคือ เกรดเฉลี่ยสะสมเมื่อสำเร็จการศึกษาระดับมัธยมศึกษาตอนปลาย คะแนนสอบคัดเลือกการศึกษาของบิดา (ตัวแปรหุ่น 2 ตัวแปร คือตัวแปรระดับการศึกษาสูงสุดของบิดาคือปริญญาตรี และตัวแปรระดับการศึกษาสูงสุดของบิดาคือปริญญาตรี) อาชีพของมารดา (ตัวแปรหุ่น 2 ตัวแปร คือตัวแปรมารดาประกอบอาชีพรับราชการและตัวแปรมารดาประกอบอาชีพลูกจ้างราชการ) อาชีพของบิดา (ตัวแปรหุ่น 1 ตัวแปรคือ ตัวแปรบิดาประกอบอาชีพรับราชการ) วิธีรับเข้า (ตัวแปรหุ่น 1 ตัวแปร คือตัวแปรรับนักศึกษาเข้าศึกษาโดยวิธีรับผ่านระบบกลาง) และสังกัดคณะ (ตัวแปรหุ่น 4 ตัวแปรคือตัวแปรการสังกัดคณะเกษตรศาสตร์ ตัวแปรการสังกัดคณะวิศวกรรมศาสตร์ ตัวแปรการสังกัดคณะศิลปศาสตร์ และตัวแปรการสังกัดคณะบริหารศาสตร์) จะเป็นข้อมูลป้อนเข้า (Input) และ ตัวแปรตาม 1 ตัวคือเกรดเฉลี่ยสะสมเมื่อศึกษาอยู่ในชั้นปีที่ 1 จะเป็นข้อมูลผลลัพธ์ (Output)

ขั้นที่ 2 การกำหนดค่าเริ่มต้นให้ตัวแบบนิวรอลเน็ตเวิร์กทำงาน ดังนี้

2.1 แบบของนิวรอลเน็ตเวิร์กประกอบด้วย ชั้นข้อมูลป้อนเข้า (Input Node Layer) ชั้นแอบแฝง (Hidden Node Layer) และชั้นแสดงผลลัพธ์ (Output Node Layer) แบ่งการกำหนดค่าในแต่ละชั้น ดังนี้

2.2.1 ชั้นข้อมูลป้อนเข้า (Input Node Layer) ประกอบด้วยจำนวนข้อมูลป้อนเข้า (Input) 12 นิวรอล หรือ จำนวนตัวแปรอิสระ 12 ตัวแปร

2.2.2 ชั้นแอบแฝง (Hidden Node Layer) ประกอบด้วยจำนวนนิวรอลเท่ากับ $\sqrt{MN}$ นิวรอล (Masters, 1993) เมื่อ M เป็นจำนวนนิวรอลของชั้นข้อมูลป้อนเข้า ในที่นี้คือจำนวนตัวแปรอิสระ ซึ่งมีค่าเท่ากับ 12 และ N เป็นจำนวนนิวรอลของชั้นแสดงผลลัพธ์ ซึ่งในที่นี้คือจำนวนตัวแปรตาม ซึ่งมีค่าเท่ากับ 1 ดังนั้น ชั้นแอบแฝงจะประกอบด้วยนิวรอลจำนวน $\sqrt{12\times1}\approx3$ นิวรอล ซึ่งสูตรนี้เป็นการประมาณขนาดชั้นแอบแฝง หลังจากนี้ดำเนินการเหมือนเดิมกับการวิเคราะห์ใน 3.1 ตั้งแต่ข้อ 2 ในขั้นที่ 2 จนถึงขั้นที่ 4

ตอนที่ 4 การตรวจสอบผลการพยากรณ์

การวิเคราะห์ในตอนนี้เป็นการนำผลจากการพยากรณ์ด้วยเทคนิคคลีเมนไทน์ มาเปรียบเทียบกับผลการพยากรณ์ด้วยเทคนิคการถดถอยพหุคูณ โดยคำนวณความคลาดเคลื่อนในการพยากรณ์ในแต่ละวิธี ด้วยค่าวัดความคลาดเคลื่อน 3 แบบ ได้แก่ รากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Square Error หรือ RMSE) ค่ามัธยฐานของค่าสัมบูรณ์ของความคลาดเคลื่อนวัดในรูปร้อยละ (Median Absoluate Percentage Error หรือ MdAPE) และ ร้อยละที่ดีกว่า (Percent Better) โดยมีขั้นตอนดังนี้

ขั้นที่ 1 คำนวนความคลาดเคลื่อนในการพยากรณ์แต่ละวิธีด้วยความคลาดเคลื่อน 3 แบบ ดังนี้

กำหนดให้ m = วิธีการพยากรณ์

F = ค่าพยากรณ์จากวิธีเดินอย่างสุ่ม

$\hat{Y}$ = ค่าพยากรณ์

Y = ค่าจริง

S = ขนาดอนุกรมเวลา

1. รากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Square Error หรือ RMSE) หมายถึง ค่าวัดความคลาดเคลื่อนของการพยากรณ์ที่วัดจากค่าความคลาดเคลื่อนกำลังสองระหว่างค่าพยากรณ์และค่าจริง ค่าวัดความคลาดเคลื่อนนี้จะไวต่อความคลาดเคลื่อนที่มีขนาดใหญ่ โดยมีเกณฑ์ที่ใช้ในการพิจารณาความเหมาะสม คือค่าวัดความคลาดเคลื่อนที่คำนวณได้ยิ่งต่ำแสดงว่า วิธีการพยากรณ์นั้นมีความคลาดเคลื่อนน้อย ซึ่งมีสูตรในการคำนวณดังนี้

$$RMSE_m = \left(\frac{\sum_{s=1}^{s} \left(\hat{Y} - Y \right)^2}{S} \right)^{1/2}$$

2. ค่ามัธยฐานของค่าสัมบูรณ์ของความคลาดเคลื่อนวัดในรูปร้อยละ (Median Absoluate Percentage Error หรือ MdAPE) หมายถึง ค่าวัดความคลาดเคลื่อนของการพยากรณ์ทีวัดจากค่ามัธยฐานของความคลาดเคลื่อนของการพยากรณ์เทียบกับค่าจริง ค่าวัดความคลาเคลื่อนนี้ เป็นค่าที่ไม่มีหน่วย โดยมีเกณฑ์ที่ใช้ในการพิจารณาความเหมาะสมคือ ค่าวัดความคลาดเคลื่อนที่คำนวณได้ยิ่งต่ำ แสดงว่าวิธีการพยากรณ์นั้นมีความคลาเคลื่อนน้อย ซึ่งมีสูตรในการคำนวณดังนี้

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

$$APE_m = \left| \frac{\hat{Y} - Y}{Y} \right| \times 100$$

เมื่อนำค่าสังเกต APE_m มาเรียงลำดับจะได้

$MdAPE_m = (S+1)/2$ เมื่อ S เป็นจำนวนคี่

$MdAPE_m = (S/2)+1$ เมื่อ S เป็นจำนวนคู่

แบบที่ 3 ร้อยละที่ดีกว่า (Percent Better) หมายถึง ค่าวัดความคลาดเคลื่อนของการพยากรณ์ที่วัดจากการหาผลต่างของความคลาดเคลื่อนจากการพยากรณ์กับการพยากรณ์เชิงสุ่ม โดยไม่คำนึงถึงทิศทางของความคลาดเคลื่อน โดยมีเกณฑ์ที่ใช้ในการพิจารณาความเหมาะสมคือ ค่าวัดความคลาดเคลื่อนที่คำนวณได้ยิ่งสูง แสดงว่าวิธีการพยากรณ์นั้นมีความคลาดเคลื่อนน้อย ซึ่งมีสูตรในการคำนวณดังนี้

$$Percent\ Better_m = \left(\frac{\sum_{s=1}^{s} J_s}{S} \right) x100$$

เมื่อ $J_s = 1$ ถ้า $\left|\hat{Y} - Y\right| < \left|F - Y\right|$

$J_s = 0$ ในกรณีอื่นๆ

ขั้นที่ 2 ภายหลังจากคำนวณผลการพยากรณ์ของแต่ละวิธีแล้วเปรียบเทียบผลการพยากรณ์แต่ละวิธี ถ้าพบว่าวิธีใดมีค่า RMSE MdAPE น้อยที่สุด และมีค่า Percent Better มากที่สุด แสดงว่า วิธีนั้นเป็นวิธีที่มีความคลาดเคลื่อนในการพยากรณ์น้อยที่สุด

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย