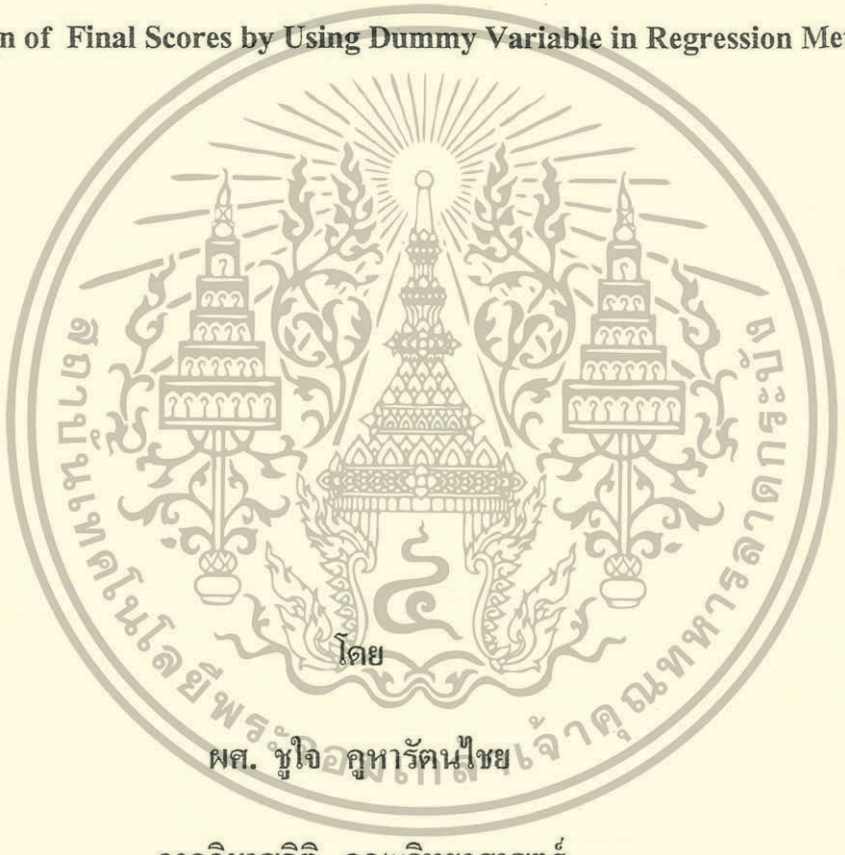
รายงานการวิจัย

เรื่อง

การประมาณคะแนนปลายภาคโดยใช้ตัวแปรดัมมี่ในวิธีการถดถอย

(The Estimation of Final Scores by Using Dummy Variable in Regression Method)



โดย

ผศ. ชูใจ คูหารัตนไชย

ภาควิชาสถิติ คณะวิทยาศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

รายงานการวิจัย

เรื่อง

การประมาณคะแนนปลายภาคโดยใช้ตัวแปรดัมมี่ในวิธีการถดถอย

(The Estimation of Final Scores by Using Dummy Variable in Regression Method)

โดย

ผศ. ชูใจ คูหารัตนไชย

ภาควิชาสถิติ คณะวิทยาศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เลขหมู่.....
เลขทะเบียน 142087
วัน.เดือน.ปี 21 เม.ย. 2559

b 12760507
i.....

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ส่วนที่ 1 รายละเอียดเกี่ยวกับโครงการ

ชื่อโครงการวิจัย(ภาษาไทย) การประมาณคะแนนปลายภาคโดยใช้ตัวแปรดัมมี่ในวิธีการถดถอย

(ภาษาอังกฤษ) The Estimation of Final Scores by Using Dummy Variable in Regression Method

ได้รับทุนอุดหนุนการวิจัยจาก งบรายได้คณะวิทยาศาสตร์ สจล.

ประจำปี 2558 จำนวนเงิน 40,000 บาท

ระยะเวลาทำการวิจัย 1 ปี ตั้งแต่วันที่ 1 ตุลาคม 2557 ถึงวันที่ 30 กันยายน 2558

หน่วยงานและผู้ดำเนินการวิจัยพร้อมหน่วยงานที่สังกัดและเลขหมายโทรศัพท์

ภาควิชาสถิติ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ผศ. ชูใจ ดูหารัตนไชย โทร. 02-3264111 ต่อ 6161

ส่วนที่ 2 บทคัดย่อ

การวิเคราะห์การถดถอยเป็นเทคนิคการพยากรณ์ที่นิยมใช้กันมาก ในการศึกษาความสัมพันธ์ระหว่างตัวแปรอิสระกับตัวแปรตาม ซึ่งตัวแปรดัมมี่ก็ถือเป็นตัวแปรอิสระที่นำมาใช้ในการแบ่งกลุ่มข้อมูลในการวิเคราะห์การถดถอย

ในการทำวิจัยครั้งนี้ มีวัตถุประสงค์ เพื่อศึกษาตัวแปรดัมมี่ที่ใช้ในรูปแบบการถดถอย และความสามารถของตัวแปรดัมมี่ในการประมาณคะแนนสอบปลายภาค โดยการเก็บรวบรวมข้อมูลเกี่ยวกับคะแนนปลายภาค คะแนนกลางภาค เกรดเฉลี่ยสะสม เพศ และสาขาวิชาของนักศึกษาในวิชาสถิติเบื้องต้น

ผลการศึกษาในส่วนของการถดถอยกับตัวแปรดัมมี่สำหรับข้อมูลสองกลุ่มพบว่า การใช้ตัวแปรอิสระเป็นคะแนนสอบกลางภาค และเกรดเฉลี่ยสะสมของนักศึกษา จะมีส่วนในการอธิบายความผันแปรของคะแนนสอบปลายภาคของนักศึกษา ได้ดีกว่าการใช้ตัวแปรอิสระเป็นคะแนนสอบกลางภาคและเพศเป็นตัวแปรดัมมี่ในการแบ่งกลุ่มข้อมูล และในส่วนของการถดถอยกับตัวแปรดัมมี่สำหรับข้อมูลสามกลุ่มพบว่า การใช้ตัวแปรอิสระเป็นคะแนนสอบกลางภาค และเกรดเฉลี่ยสะสมของนักศึกษา จะมีส่วนในการอธิบายความผันแปรของคะแนนสอบปลายภาคของนักศึกษาได้ดีกว่าการใช้ตัวแปรอิสระเป็นคะแนนสอบกลางภาคและสาขาวิชาเป็นตัวแปรดัมมี่ในการแบ่งกลุ่มข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Regression analysis is a forecasting technique popularly used for studying relationships between independent and dependent variables. A dummy variable is regarded as an independent variable used for categorizing groups of data in the regression analysis.

The objective of this research is to study dummy variables used in the regression analysis and their capability in estimating final exam scores by collecting data of final score, midterm score, accumulated score, gender, and field of study of students who enroll for the fundamental statistics subject.

From results of the study on regression analysis of dummy variables for data of two groups, it was found that using midterm score and accumulated score as independent variables can explain variability of students' final score better than using mid-term score and gender as dummy variables for categorizing groups of data. From the regression analysis of dummy variables, it was found that using midterm score and accumulated score as independent variables can explain variability of students' final score better than using midterm score and field of study as dummy variables for categorizing groups of data.

คำสำคัญ (Keywords) : Regression analysis , Dummy variables

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คำนำ

รายงานการวิจัยการประมาณคะแนนปลายภาคโดยใช้ตัวแปรดัมมี่ในวิธีการถดถอย เป็นงานวิจัยโดยศึกษาตัวแปรดัมมี่ในรูปแบบการถดถอย เพื่อนำมาใช้ประมาณคะแนนปลายภาคของนักศึกษา จากผลการวิจัยที่ได้สามารถนำมาใช้เป็นแนวทางในการศึกษาตัวแปรดัมมี่ในการวิเคราะห์การถดถอยได้

ผู้วิจัยหวังเป็นอย่างยิ่งว่า รายงานวิจัยนี้จะเป็นประโยชน์ต่อคณาจารย์ นักศึกษา และผู้ที่สนใจทั่วไป

ผศ. ชูใจ ดูหารัตนไชย

30 กันยายน 2558

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

เรื่อง			หน้า
บทคัดย่อภาษาไทย			ข
บทคัดย่อภาษาอังกฤษ			ค
คำนำ			ง
สารบัญตาราง			ช
บทที่ 1	บทนำ		
	1.1	ความสำคัญของปัญหา	1
	1.2	วัตถุประสงค์ของการวิจัย	2
	1.3	ประโยชน์ที่คาดว่าจะได้รับ	2
	1.4	ขอบเขตของงานวิจัย	2
บทที่ 2	ทฤษฎีที่เกี่ยวข้อง		
	2.1	การวิเคราะห์การถดถอยเชิงเส้นตรงแบบพหุ	4
	2.2	ตัวแปรดัมมี่กับการวิเคราะห์การถดถอยพหุ	14
	2.3	การทดสอบสมมติฐานเกี่ยวกับค่าเฉลี่ยของประชากร 2 กลุ่มที่ไม่เป็นอิสระกัน	20
	2.4	การหาคุณภาพแบบทดสอบ	22
บทที่ 3	การดำเนินงานวิจัย		
	3.1	ข้อมูลที่ใช้ในงานวิจัย	30
	3.2	สถิติที่ใช้ในการวิเคราะห์	30
	3.3	ขั้นตอนในการดำเนินงาน	31
บทที่ 4	ผลการวิเคราะห์ข้อมูล		
	4.1	ผลการวิเคราะห์คุณภาพข้อสอบ	32
	4.2	ผลการวิเคราะห์การถดถอยกับตัวแปรดัมมี่สำหรับข้อมูลสองกลุ่ม	44
	4.3	ผลการวิเคราะห์การถดถอยกับตัวแปรดัมมี่สำหรับข้อมูลสามกลุ่ม	58

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ(ต่อ)

เรื่อง		หน้า
บทที่ 5	สรุปผลการวิเคราะห์และข้อเสนอแนะ	
	5.1 คุณภาพข้อสอบ	64
	5.2 การถดถอยกับตัวแปรดัมมี่สำหรับข้อมูลสองกลุ่ม	65
	5.3 การถดถอยกับตัวแปรดัมมี่สำหรับข้อมูลสามกลุ่ม	67
	5.3 ข้อเสนอแนะ	68
บรรณานุกรม		69

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

เรื่อง		หน้า
ตารางที่ 2.1	แสดงสมมติฐาน ตัวทดสอบสถิติ และช่วงวิกฤติสำหรับการทดสอบแบบ t	19
ตารางที่ 2.2	แสดงสมมติฐาน ตัวทดสอบสถิติ และช่วงวิกฤติสำหรับการทดสอบแบบ F	14
ตารางที่ 2.3	แสดงรูปแบบการถดถอยและสมการถดถอยสำหรับข้อมูลสองกลุ่ม	15
ตารางที่ 2.4	แสดงรูปแบบการถดถอยและสมการถดถอยสำหรับข้อมูลสามกลุ่ม	18
ตารางที่ 2.5	แสดงสมมติฐาน ตัวทดสอบสถิติ และช่วงวิกฤติสำหรับการทดสอบสมมติฐานเกี่ยวกับค่าเฉลี่ยของประชากร 2 กลุ่มที่ไม่เป็นอิสระกัน	21
ตารางที่ 4.1	การเปรียบเทียบความแตกต่างของค่าอำนาจจำแนกรายข้อแบบสัดส่วนและแบบพอยน์ไบซีเรียล โดยใช้ตัวสถิติ t สำหรับค่าเฉลี่ยของประชากร 2 กลุ่มที่ไม่เป็นอิสระกัน	33
ตารางที่ 4.2	คุณภาพข้อสอบกลางภาคแบบรายข้อวิชาสถิติเบื้องต้นสาขาวิชาสถิติประยุกต์ ปีการศึกษา 2555	34
ตารางที่ 4.3	คุณภาพข้อสอบปลายภาคแบบรายข้อ วิชาสถิติเบื้องต้นสาขาวิชาสถิติประยุกต์ ปีการศึกษา 2555	35
ตารางที่ 4.4	คุณภาพข้อสอบกลางภาคแบบรายข้อ วิชาสถิติเบื้องต้นสาขาวิชาสถิติประยุกต์ ปีการศึกษา 2556	36
ตารางที่ 4.5	คุณภาพข้อสอบปลายภาคแบบรายข้อ วิชาสถิติเบื้องต้นสาขาวิชาสถิติประยุกต์ ปีการศึกษา 2556	37
ตารางที่ 4.6	คุณภาพข้อสอบกลางภาคแบบรายข้อ วิชาสถิติเบื้องต้นสาขาวิชาจุลชีววิทยา ปีการศึกษา 2557	39
ตารางที่ 4.7	คุณภาพข้อสอบปลายภาคแบบรายข้อ วิชาสถิติเบื้องต้นสาขาวิชาจุลชีววิทยา ปีการศึกษา 2557	40

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง(ต่อ)

เรื่อง		หน้า
ตารางที่ 4.8	คุณภาพข้อสอบกลางภาคแบบรายข้อ วิชาสถิติเบื้องต้นสาขาวิชาสัตวศาสตร์ ปีการศึกษา 2555	41
ตารางที่ 4.9	คุณภาพข้อสอบปลายภาคแบบรายข้อ วิชาสถิติเบื้องต้นสาขาวิชาสัตวศาสตร์ ปีการศึกษา 2555	42
ตารางที่ 4.10	คุณภาพข้อสอบกลางภาคและปลายภาควิชาสถิติเบื้องต้นจำแนกตามสาขาวิชา	43
ตารางที่ 4.11	รูปแบบการถดถอยกับตัวแปรดัมมี่สำหรับข้อมูลสองกลุ่มของนักศึกษาสาขาวิชาสถิติประยุกต์ ปีการศึกษา 2555	47
ตารางที่ 4.12	รูปแบบการถดถอยกับตัวแปรดัมมี่สำหรับข้อมูลสองกลุ่มของนักศึกษาสาขาวิชาสถิติประยุกต์ ปีการศึกษา 2556	51
ตารางที่ 4.13	รูปแบบการถดถอยกับตัวแปรดัมมี่สำหรับข้อมูลสองกลุ่มของนักศึกษาสาขาวิชาชีวสถิติ ปีการศึกษา 2557	54
ตารางที่ 4.14	รูปแบบการถดถอยกับตัวแปรดัมมี่สำหรับข้อมูลสองกลุ่มของนักศึกษาสาขาวิชาสัตวศาสตร์ ปีการศึกษา 2555	58
ตารางที่ 4.15	รูปแบบการถดถอยกับตัวแปรดัมมี่สำหรับข้อมูลสามกลุ่มของนักศึกษาสาขาวิชาสถิติประยุกต์ สาขาวิชาชีวสถิติ และสาขาวิชาสัตวศาสตร์	62

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1
บทนำ

1.1 ความสำคัญของปัญหา

การวิเคราะห์การถดถอยเป็นเทคนิคการพยากรณ์ ที่นิยมใช้กันอย่างกว้างขวางมากที่สุดในสาขาวิชาต่างๆ โดยใช้ตัวแปรอิสระ (Independent Variable) หรือตัวแปรพยากรณ์ (Predictor) อย่างน้อยหนึ่งตัวไปพยากรณ์ หรืออธิบายความผันแปรของตัวแปรตาม (Dependent Variable) ถ้าการวิเคราะห์การถดถอยประกอบด้วยตัวแปรอิสระหนึ่งตัว ซึ่งวัดเป็นค่าในเชิงปริมาณ เรียกว่า การวิเคราะห์การถดถอยอย่างง่าย (Simple regression analysis) ส่วนกรณีที่มีตัวแปรอิสระมากกว่าหนึ่งตัว จะเรียกว่า การวิเคราะห์การถดถอยแบบพหุ (Multiple regression analysis)

การวิเคราะห์การถดถอยส่วนหนึ่งจะเกี่ยวข้องกับการศึกษาว่า ความผันแปรของตัวแปรตามในรูปแบบการถดถอยต่างๆ เป็นผลเนื่องมาจากปัจจัยใดบ้าง ปัจจัยที่นำมาศึกษาจะแทนด้วยตัวแปรอิสระ ตัวแปรอิสระที่ใช้กันมากมีลักษณะตามชื่อของตัวแปร ได้แก่ ตัวแปรปัจจัย (factor variable) ตัวแปรเวลา (time variable) และตัวแปรดัมมี่ (dummy variable) โดยมีรายละเอียดดังนี้

- ตัวแปรปัจจัย จะแทนปัจจัยต่างๆ ที่มีอิทธิพลต่อตัวแปรตาม เช่น ยอดขายของบริษัทจะมากหรือน้อยจะขึ้นอยู่กับปัจจัยต่างๆ ได้แก่ ยอดโฆษณา ราคาสินค้าที่ขาย และราคาสินค้าของบริษัทคู่แข่ง ผลผลิตข้าวจะมากหรือน้อยขึ้นอยู่กับปัจจัยต่างๆ ได้แก่ ปริมาณน้ำฝน ปริมาณปุ๋ยที่ใช้ และคุณภาพของดิน จำนวนสินค้าที่มีรอยตำหนิที่ผลิตจากโรงงานแห่งหนึ่งจะมากหรือน้อยขึ้นอยู่กับปัจจัยต่างๆ ได้แก่ เครื่องจักรที่ใช้ผลิต ระดับความสามารถของผู้ควบคุมเครื่องจักร เงินเดือนของพนักงานในองค์กรแห่งหนึ่งจะมากหรือน้อยขึ้นอยู่กับ เพศ อายุ ระดับการศึกษาจำนวนวันลาในปีที่ผ่านมา เป็นต้น

- ตัวแปรเวลา เป็นกรณีที่ศึกษาการเคลื่อนไหวของอนุกรมเวลา ซึ่งส่วนใหญ่จะเป็นข้อมูลทางธุรกิจหรือทางเศรษฐกิจ มักจะเป็นการศึกษาแนวโน้มซึ่งเป็นการเคลื่อนไหวของค่าสังเกตตามเวลาในระยะยาว รูปแบบแนวโน้มที่สร้างขึ้นจะกำหนดตัวแปรเวลาเป็นตัวแปรอิสระที่มีค่าเป็นรหัส (coded value) แทนวัน เดือน ไตรมาส หรือปี ที่เก็บรวบรวมข้อมูลมา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- ตัวแปรดัมมี่ เป็นตัวแปรที่สร้างขึ้นเพื่อระบุกลุ่มหรือชุดที่ค่าสังเกตนั้นอยู่ จะกำหนดค่าของตัวแปรดัมมี่เป็น 1 หรือ 0 ซึ่งตัวแปรดัมมี่จะมีค่าเท่ากับ 1 เมื่อค่าสังเกตนั้นอยู่ในกลุ่มที่สนใจ และมีค่าเป็น 0 เมื่อค่าสังเกตนั้นไม่อยู่ในกลุ่มที่สนใจ

ในการเขียนสมการถดถอยเพื่อประมาณค่าตัวแปรตาม ส่วนใหญ่จะให้ความสำคัญกับตัวแปรเชิงปริมาณหรือตัวแปรปัจจัยแล้ว ตัวแปรเชิงคุณภาพหรือตัวแปรดัมมี่ก็มีความสำคัญไม่น้อยไปกว่ากัน แต่ในการวิเคราะห์การถดถอยโดยทั่วไปจะเน้นศึกษาเฉพาะตัวแปรปัจจัย เพราะสามารถวัดค่าออกมาได้ทำให้ง่ายต่อการนำไปวิเคราะห์ และมองข้ามตัวแปรดัมมี่ซึ่งอาจมีอิทธิพลต่อตัวแปรตามเช่นกัน ส่วนตัวแปรดัมมี่ในการวิเคราะห์การถดถอยจะถือเป็นตัวแปรจัดประเภท (Categorical Variable) เพื่อใช้ในการระบุกลุ่มของข้อมูล

จากแนวคิดในการวิเคราะห์การถดถอย ผู้วิจัยจึงสนใจที่จะนำเอาตัวแปรดัมมี่มาใช้ในสมการถดถอยในการประมาณคะแนนสอบปลายภาคของนักศึกษา โดยอาศัยในการเก็บรวบรวมข้อมูลจากคะแนนสอบกลางภาคและเกรดเฉลี่ยสะสม(Grade Point Average หรือ GPA)ของนักศึกษา โดยแบ่งข้อมูลออกเป็นกลุ่มตาม เพศ และสาขาวิชาที่นักศึกษากำลังศึกษา

1.2 วัตถุประสงค์ของการวิจัย

1. เพื่อศึกษาตัวแปรดัมมี่ที่ใช้ในรูปแบบการถดถอย
2. เพื่อศึกษาความสามารถของตัวแปรดัมมี่ในการประมาณคะแนนสอบปลายภาคของนักศึกษา

1.3 ประโยชน์ที่คาดว่าจะได้รับ

1. สามารถใช้เป็นแนวทางในการศึกษาตัวแปรดัมมี่ในรูปแบบการถดถอย
2. ทำให้ทราบถึงความสำคัญของตัวแปรดัมมี่ในการประมาณคะแนนสอบปลายภาคของนักศึกษา
3. นำผลการวิจัยที่ได้มาถ่ายทอดองค์ความรู้ให้กับนักศึกษาได้

1.4 ขอบเขตของงานวิจัย

งานวิจัยครั้งจะทำการศึกษาในการแบ่งกลุ่มข้อมูลโดยอาศัยตัวแปรดัมมี่ในรูปแบบการถดถอย เพื่อนำมาใช้ในการประมาณคะแนนสอบปลายภาคของนักศึกษา จากการเก็บรวบรวมข้อมูลได้จากนักศึกษาสาขาวิชาสถิติประยุกต์ และสาขาวิชาจุลชีววิทยาอุตสาหกรรมหรือจุลชีววิทยาในคณะวิทยาศาสตร์ และนักศึกษาสาขาวิชาสัตวศาสตร์ในคณะเทคโนโลยีการเกษตร เพื่อให้คะแนนสอบปลายภาคและคะแนน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สอบกลางภาคที่ใช้ในการวิเคราะห์ในแต่ละสาขาวิชามีความน่าเชื่อถือ จึงทำการหาคุณภาพของข้อสอบประกอบในการวิเคราะห์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2
ทฤษฎีที่เกี่ยวข้อง

2.1 การวิเคราะห์การถดถอยเชิงเส้นตรงแบบพหุ

2.1.1 รูปแบบการถดถอยเชิงเส้นตรงแบบพหุ

ปัจจัยที่มีผลต่อความผันแปรของตัวแปรตามอาจมีเพียงปัจจัยเดียว หรือมากกว่าหนึ่งปัจจัย เช่นผลผลิตข้าวต่อไร่ นอกจากจะขึ้นอยู่กับปริมาณปุ๋ยที่ใช้ต่อไร่แล้ว ยังจะขึ้นอยู่กับปัจจัยอื่นๆ อีกได้แก่ ความเจริญทางเทคโนโลยีที่แทนด้วยปัจจัยเวลาและปริมาณน้ำในฤดูทำนา ส่วนยอดขายสินค้าของห้างสรรพสินค้านอกจากจะขึ้นอยู่กับยอดโฆษณาแล้วยังจะขึ้นกับปัจจัยอื่นๆ อีกได้แก่ ราคาสินค้าที่กำหนดขึ้น จำนวนห้างสรรพสินค้าที่เป็นคู่แข่งขันและการให้บริการ เป็นต้น

การวิเคราะห์การถดถอยเมื่อมีตัวแปรที่เกี่ยวข้องหลายตัวแปร จะทำการกำหนดให้ตัวแปรหนึ่งที่สนใจเป็นตัวแปรตาม ส่วนตัวแปรที่เหลือแทนปัจจัยที่มีอิทธิพลต่อตัวแปรที่สนใจศึกษาเป็นตัวแปรอิสระ การวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรตามและตัวแปรอิสระจะทำตามรูปแบบการถดถอยที่กำหนดขึ้น เช่น ในกรณีมีตัวแปรอิสระ 3 ตัวแปรได้แก่ X_1 X_2 และ X_3 รูปแบบการถดถอยเชิงเส้นตรงได้แก่

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$$

เรียกรูปแบบการถดถอยนี้ว่า รูปแบบการถดถอยเชิงเส้นตรงแบบพหุ(multiple linear regression model) โดยมีข้อสมมติของรูปแบบดังนี้

1. ε_i มีการแจกแจงแบบปกติที่มีค่าเฉลี่ยเป็น 0 และค่าความแปรปรวนเท่ากับ σ^2 ซึ่งหมายความว่า $\varepsilon_i \sim N(0, \sigma^2)$
2. ε_i และ ε_j สำหรับ $i \neq j$ มีการแจกแจงที่เป็นอิสระกัน จะทำให้ $COV(\varepsilon_i, \varepsilon_j) = 0$ สำหรับ $i \neq j$

จากข้อสมมติของ ε_i สรุปได้ว่าลักษณะการแจกแจงของ Y_i เมื่อ $X_1 = X_{1i}$ $X_2 = X_{2i}$ และ $X_3 = X_{3i}$ เป็นแบบปกติที่เป็นอิสระกัน มีค่าเฉลี่ย $\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i}$ และมีค่าความแปรปรวน σ^2 นั่นหมายถึง $Y_i \sim Nid(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i}, \sigma^2)$ ข้อสมมติอื่นที่นอกเหนือจากข้อสมมติเกี่ยวกับความคลาดเคลื่อน ε_i ได้แก่ รูปแบบเป็นแบบเชิงเส้นตรงของพารามิเตอร์(β_i) และตัวแปรอิสระแต่ละตัวแปรไม่มีความเกี่ยวข้องกัน กรณีที่มีตัวแปรอิสระ k ตัวแปร รูปแบบการถดถอยเชิงเส้นตรงและ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ข้อสมมติของรูปแบบ จะเป็นไปในทำนองเดียวกันกับกรณีตัวแปรอิสระ3ตัวแปรดังกล่าวข้างต้น ในการวิเคราะห์การถดถอยกรณีที่มีตัวแปรอิสระมากกว่าหนึ่งตัวแปร จึงมักจะเขียนรูปแบบการถดถอยให้อยู่ในแบบเมตริกซ์ (matrix form) ซึ่งสามารถเขียนรูปแบบการถดถอยเชิงเส้นตรงแบบเมตริกซ์กรณีมี k ตัวแปรอิสระได้ดังนี้

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

เมื่อ $\underline{Y}$ เป็นเวคเตอร์แถวตั้งขนาด n ที่มีสมาชิกที่ i เป็น Y_i

$\underline{\beta}$ เป็นเวคเตอร์แถวตั้งขนาด (k+1) ที่มีสมาชิกที่ i เป็น β_i

$\underline{\varepsilon}$ เป็นเวคเตอร์แถวตั้งขนาด n ที่มีสมาชิกที่ i เป็น ε_i

และ X เป็นเมตริกซ์ขนาด n×(k+1) โดย

$$\underline{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & X_{11} & \cdots & X_{k1} \\ \vdots & & & \\ 1 & X_{1n} & \cdots & X_{kn} \end{bmatrix}, \quad \underline{\beta} = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_k \end{bmatrix}, \quad \underline{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

2.1.2 สมการถดถอย

การสร้างสมการถดถอยจะเริ่มจากการหาค่าประมาณ b_i ของพารามิเตอร์ β_i ซึ่งเมื่อเขียนตัวประมาณ b_i ของพารามิเตอร์ β_i ในรูปเวคเตอร์เป็น $\underline{b}$ จะหาเวคเตอร์ $\underline{b}$ ได้ด้วยวิธีกำลังสองน้อยที่สุด นั่นคือ หา $\underline{b}$ ที่ทำให้ SSE มีค่าน้อยที่สุด ซึ่ง

$$\begin{aligned} SSE &= \underline{e}'\underline{e} \\ &= (\underline{Y} - X\underline{b})'(\underline{Y} - X\underline{b}) \\ &= (\underline{Y}' - \underline{b}'X')(\underline{Y} - X\underline{b}) \\ &= \underline{Y}'\underline{Y} - 2\underline{b}'X'\underline{Y} + \underline{b}'X'X\underline{b} \end{aligned}$$

โดยเวคเตอร์ $\underline{e}$ เป็นเวคเตอร์ของความคลาดเคลื่อน ซึ่ง $\underline{e} = \left(\underline{Y} - \hat{\underline{Y}}\right) = (\underline{Y} - X\underline{b})$ เวคเตอร์ $\underline{b}$ ที่มีคุณสมบัติดังกล่าวจะหาได้โดยการหาอนุพันธ์ย่อยของ SSE เทียบกับ $\underline{b}$ และกำหนดให้เท่ากับเวคเตอร์ $\underline{0}$ จากสมการดังกล่าวจะเขียนได้เป็นสมการใหม่

$$X'X\underline{b} = X'\underline{Y}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เรียกสมการนี้ว่า สมการปกติ เมื่อนำเอาเมตริกซ์ $(X'X)^{-1}$ คูณทั้งด้านซ้ายและด้านขวาของสมการปกติ จะได้ $\underline{b} = (X'X)^{-1}X'\underline{Y}$ ซึ่ง $(X'X)$ เป็นเมตริกซ์สมมาตรขนาด (k+1)× (k+1) เมื่อ $(X'X)^{-1}$เป็นเมตริกซ์ผกผันของ$(X'X)$ และ $(X'\underline{Y})$เป็นเวคเตอร์แถวตั้งขนาด (k+1) โดย

$$X'X = \begin{bmatrix} n & \sum X_1 & \cdots & \sum X_k \\ \sum X_1 & \sum X_1^2 & \cdots & \sum X_1X_k \\ \vdots & & & \\ \sum X_k & \sum X_1X_k & \cdots & \sum X_k^2 \end{bmatrix} \quad , \quad X'\underline{Y} = \begin{bmatrix} \sum Y \\ \sum X_1Y \\ \vdots \\ \sum X_kY \end{bmatrix}$$

ตัวประมาณ b_i ที่ได้จากวิธีกำลังสองน้อยที่สุดจะเป็นตัวประมาณที่ดี นั่นคือเป็น BLUE (best linear unbiased estimator) ของ β_i กล่าวคือ เป็นตัวประมาณที่เขียนได้ในรูปเชิงเส้นของเวคเตอร์ $\underline{Y}$ ไม่อคติ และมีค่าความแปรปรวนต่ำกว่าค่าความแปรปรวนของตัวประมาณอื่น ที่สามารถเขียนได้ในรูปเชิงเส้นของเวคเตอร์ $\underline{Y}$ และไม่อคติเหมือนกัน จากค่า b_i ที่คำนวณได้จะนำค่า b_i ไปใช้ประโยชน์ในการประมาณค่าของตัวแปรตาม Y ได้ดังนี้

$$\hat{Y}_0 = b_0 + b_1X_{10} + b_2X_{20} + \ldots + b_kX_{k0}$$

ซึ่ง $X_{10}, X_{20}, \ldots, X_{k0}$ เป็นค่าคงที่ และสามารถเขียน $\hat{Y}_0$ ในรูปเมตริกซ์ได้เป็น

$$\hat{Y}_0 = \underline{X}_0'\underline{b}$$

เมื่อ $\underline{X}_0$ เป็นเวคเตอร์แถวตั้งขนาด (k+1) โดย $\underline{X}_0' = (1\ X_{10}\ X_{20}\ \ldots\ X_{k0})$ จะประมาณ σ^2 โดย S^2 ซึ่ง

$$S^2 = \frac{SSE}{n-k-1}$$

2.1.3 การแบ่งส่วนของ SST

การศึกษาความเหมาะสมของรูปแบบการถดถอยนั้น จะพิจารณาค่าสถิติวัดความเหมาะสมบางค่าและการทดสอบสมมติฐาน ค่าสถิติวัดความเหมาะสมของรูปแบบและตัวทดสอบสถิติ F ที่ใช้ในการทดสอบสมมติฐาน จะเขียนได้อยู่ในรูปของผลรวมกำลังสอง(SS) ที่ได้จากการแบ่งส่วนของผลรวมกำลังสองรวม(SST)

จากสมการถดถอยเชิงเส้นตรงแบบพหุ $\hat{Y} = b_0 + b_1X_1 + b_2X_2 + \ldots + b_kX_k$ จะหาค่าประมาณของ Y ที่แต่ละชุดของ X_i ที่กำหนดให้เป็น $\hat{Y}$ จะเขียนผลต่างของค่าจริงที่ i (Y_i) จากค่าเฉลี่ย $\bar{Y}$ ในเทอมของค่าประมาณที่ i $(\hat{Y}_i)$ ได้ดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$Y_i - \bar{Y} = \left(\hat{Y}_i - \bar{Y}\right) + \left(Y_i - \hat{Y}_i\right)$$

เมื่อหาผลรวมกำลังสองทั้งสองข้างจะได้

$$\sum\left(Y_i - \bar{Y}\right)^2 = \sum\left(\hat{Y}_i - \bar{Y}\right)^2 + \sum\left(Y_i - \hat{Y}_i\right)^2$$

หรือ $\quad SST = SSR(X_1 X_2 \ldots X_k) + SSE(X_1 X_2 \ldots X_k)$

เมื่อ $SSR(X_1 X_2 \ldots X_k)$ เป็นผลรวมกำลังสองเนื่องจากการถดถอยกรณีใช้ตัวแปรอิสระ k ตัวแปร ได้แก่ $X_1 X_2 \ldots X_k$ หรือส่วนของ SST ที่อธิบายได้ด้วยตัวแปรอิสระ $X_1 X_2 \ldots X_k$ ตัวอย่างการแบ่งส่วนของ SST เมื่อจำนวนตัวแปรอิสระต่างกัน

เมื่อ $k=1$ $\quad SST = SSR(X_1) + SSE(X_1)$

เมื่อ $k=2$ $\quad SST = SSR(X_1 X_2) + SSE(X_1 X_2)$

เมื่อ $k=3$ $\quad SST = SSR(X_1 X_2 X_3) + SSE(X_1 X_2 X_3)$

จะเห็นได้ว่า ผลรวมกำลังสองไม่ว่าจะมีตัวแปรอิสระกี่ตัวแปรก็สามารถคำนวณในเทอมเมตริกซ์ได้จาก

$$SST = \underline{Y}'\underline{Y} - n\bar{Y}^2$$

$$SSR(X_1 X_2 \ldots X_k) = \underline{b}'X'\underline{Y} - n\bar{Y}^2$$

$$SSE(X_1 X_2 \ldots X_k) = \underline{Y}'\underline{Y} - \underline{b}'X'\underline{Y}$$

เช่น กรณีเมื่อ $k=2$

$$\underline{Y}'\underline{Y} = \sum Y_i^2 \quad , \quad \underline{b}'X'\underline{Y} = \begin{bmatrix} b_0 & b_1 & b_2 \end{bmatrix} \begin{bmatrix} \sum Y_i \\ \sum X_{1i} Y_i \\ \sum X_{2i} Y_i \end{bmatrix} = b_0 \sum Y_i + b_1 \sum X_{1i} Y_i + b_2 \sum X_{2i} Y_i$$

ซึ่งหา $\underline{b}$ ได้จาก $\underline{b} = (X'X)^{-1} X'\underline{Y}$ เป็นต้น

ในกรณีที่สนใจว่า เมื่อกำหนดตัวแปรอิสระ X_1 ในรูปแบบแล้ว การเพิ่มตัวแปรอิสระอีกหนึ่งตัวแปรได้แก่ X_2 เข้าในรูปแบบ จะมีส่วนใน SST เพิ่มมากขึ้นเท่าใด จะวัดจาก $SSR(X_2/X_1)$ ซึ่งเป็นส่วนของ $SSR(X_1 X_2)$ ที่เนื่องจากการเพิ่มตัวแปรอิสระ X_2 เมื่อมีตัวแปรอิสระ X_1 ในรูปแบบแล้ว นั่นคือ

$$SSR(X_2/X_1) = SSR(X_1 X_2) - SSR(X_1)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หลักในการหา SSR ที่เพิ่มขึ้น จากการเพิ่มตัวแปรอิสระตัวหนึ่งหรือหลายตัวเมื่อมีตัวแปรอิสระอื่นอยู่ในรูปแบบแล้วนี้ จะใช้ได้กับรูปแบบการถดถอยที่มีจำนวนตัวแปรอิสระเท่าใดก็ได้ เช่น เมื่อมีตัวแปรอิสระ 4 ตัวแปร X_1 X_2 X_3 X_4 จะหาส่วนของ SST ที่เนื่องจาก X_2 และ X_4 เมื่อมีตัวแปร X_1 และ X_3 อยู่ในรูปแบบแล้วโดย

$$SSR(X_2 X_4 / X_1 X_3) = SSR(X_1 X_2 X_3 X_4) - SSR(X_1 X_3)$$

การแบ่งส่วนของ SSR มี 2 แบบคือ การแบ่งส่วนของ SSR แบบต่อเนื่องและการแบ่งส่วนของ SSR แบบบางส่วน กรณีตัวแปรอิสระ 3 ตัวแปรได้แก่ X_1 X_2 และ X_3 การแบ่งส่วนของ $SSR(X_1 X_2 X_3)$ มีรายละเอียดดังนี้

1. ในการแบ่งส่วนของ SSR แบบต่อเนื่อง แต่ละส่วนที่แบ่งได้นั้นจะต้องนำมารวมกันเท่ากับ $SSR(X_1 X_2 X_3)$ เช่น

$$SSR(X_1 X_2 X_3) = SSR(X_1) + SSR(X_2 / X_1) + SSR(X_3 / X_1 X_2)$$

2. ในการแบ่งส่วนของ SSR แบบบางส่วน แต่ละส่วนที่แบ่งได้นั้นจะนำมารวมกันไม่เท่ากับ $SSR(X_1 X_2 X_3)$ เช่น

$$SSR(X_1 X_2 X_3) \neq SSR(X_1 / X_2 X_3) + SSR(X_2 / X_1 X_3) + SSR(X_3 / X_1 X_2)$$

2.1.4 ค่าวัดประสิทธิภาพของรูปแบบ

เมื่อกำหนดรูปแบบการถดถอยให้กับข้อมูลตัวอย่างแล้ว จะทำการสร้างสมการถดถอยตามรูปแบบการถดถอยนั้น อย่างไรก็ตามรูปแบบการถดถอยที่กำหนดไว้ อาจจะเหมาะสมกับข้อมูลตัวอย่างหรือไม่ก็ได้ ดังนั้น ในการพิจารณาความเหมาะสมของรูปแบบการถดถอย จะได้จากค่าสถิติที่ใช้วัดประสิทธิภาพของรูปแบบและจากการทดสอบสมมติฐาน ค่าวัดประสิทธิภาพของรูปแบบมีอยู่หลายค่า เช่น ผลรวมกำลังสองของความคลาดเคลื่อน(SSE) ค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน(MSE) ค่าสัมประสิทธิ์ตัวกำหนด(R^2) และค่าสัมประสิทธิ์ตัวกำหนดปรับแล้ว (R_a^2) เป็นต้น ค่าสถิติเหล่านี้จะใช้วัดประสิทธิภาพของรูปแบบมีรายละเอียดดังนี้

1. ค่าผลรวมกำลังสองของความคลาดเคลื่อน(Sum Square Error : SSE) เป็นค่าวัดที่ยังไม่มีเกณฑ์แน่นอนว่า รูปแบบที่เหมาะสมจะต้องมีค่า SSE เป็นเท่าใด แต่รูปแบบการถดถอยที่เหมาะสมควรจะเป็นรูปแบบที่มีค่า SSE น้อยที่สุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. ค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน(Mean Square Error : MSE) เป็นค่าวัดที่เป็นฟังก์ชันของ SSE นั่นคือ เป็นค่า SSE ที่ปรับด้วยชั้นแห่งความเป็นอิสระ ซึ่ง $MSE = \frac{SSE}{n-k-1}$ สำหรับรูปแบบการถดถอยที่มีจำนวนตัวแปรอิสระต่างกัน แต่มี SSE เท่ากัน รูปแบบที่มีจำนวนตัวแปรอิสระน้อยกว่าจะให้ค่า MSE ที่ต่ำกว่า ค่า MSE จากรูปแบบที่มีจำนวนตัวแปรอิสระมากกว่า นอกจากการใช้ค่า SSE และ MSE เพื่อวัดประสิทธิภาพของรูปแบบแล้ว ยังมีผู้ใช้ค่ารากที่สองของ MSE หรือ RMSE ซึ่ง $RMSE = \sqrt{MSE}$ ในการพิจารณาความเหมาะสมของรูปแบบ กรณีที่ใช้ค่าMSE หรือRMSE ในการพิจารณารูปแบบที่เหมาะสม รูปแบบที่ให้ค่า MSE หรือ RMSE ต่ำที่สุดจะเป็นรูปแบบที่เหมาะสมที่สุด ซึ่งเป็นทำนองเดียวกับการใช้ค่า SSE

3. ค่าสัมประสิทธิ์ตัวกำหนด(Coefficient of determination : R^2) เป็นค่าสถิติที่ใช้ในการวัดค่าว่า ตัวแปรอิสระที่อยู่ในรูปแบบการถดถอยมีส่วนในการอธิบายความผันแปรรวม $\sum(Y_i - \bar{Y})^2$ มากน้อยเท่าใด รูปแบบที่เหมาะสมที่สุดจะเป็นรูปแบบที่ให้ค่า R^2 สูงสุด ค่า R^2 จะเป็นสัดส่วนของ SSR กับ SST เนื่องจาก $SST = SSR + SSE$ ทำให้ค่า R^2 จึงมีค่าอยู่ระหว่าง 0 ถึง 1

$$R^2 = \frac{SSR}{SST}$$
$$= 1 - \frac{SSE}{SST}$$

โดยทั่วไปจะอธิบายค่าของ R^2 เป็นเปอร์เซนต์แทนการอธิบายด้วยสัดส่วน เช่น สำหรับรูปแบบการถดถอยที่มีตัวแปรอิสระ X_1 และ X_2 ที่มี $R^2 = 0.8921$ จะอธิบายได้ว่าตัวแปรอิสระ X_1 และ X_2 มีส่วนในการอธิบายความผันแปรรวม $\sum(Y_i - \bar{Y})^2$ ได้ดี 89.21 เปอร์เซนต์ เป็นต้น เนื่องจาก R^2 แปรผันกับ SSE ดังนั้นเมื่อ SSE มีค่าน้อย R^2 จะมีค่ามากหรือเข้าใกล้ 1 และเมื่อ SSE มีค่ามาก R^2 จะมีค่าน้อยหรือเข้าใกล้ 0

4. ค่าสัมประสิทธิ์ตัวกำหนดที่ปรับแล้ว (Adjusted coefficient of determination : R_a^2) เป็นค่าสถิติที่นำมาใช้วัดว่าตัวแปรอิสระที่อยู่ในรูปแบบการถดถอย มีส่วนในการอธิบายความแปรปรวน (S_y^2) มากน้อยเพียงใด รูปแบบที่เหมาะสมที่สุดจะเป็นรูปแบบที่ให้ค่า R_a^2 สูงสุด ค่า R_a^2 จะแตกต่างจากค่า R^2 ที่ค่า R_a^2 คำนึงถึงชั้นแห่งความเป็นอิสระของ SSE และ SST นั่นคือ จะพิจารณา MSE แทน SSE และ S_y^2 แทน SST

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$R_a^2 = 1-\frac{SSE/(n-k-1)}{SST/(n-1)}$$

$$= 1-\frac{MSE}{S_y^2}$$

เมื่อ n มีขนาดใหญ่ ค่า R_a^2 จะใกล้เคียงกับค่า R^2 ในการเปรียบเทียบรูปแบบการถดถอยสองรูปแบบที่มีจำนวนตัวแปรอิสระต่างกันแต่มีค่า SSE เท่ากัน รูปแบบที่มีจำนวนตัวแปรอิสระมากจะมีค่า R_a^2 น้อยกว่าค่า R_a^2 ของรูปแบบที่มีจำนวนตัวแปรอิสระน้อย เนื่องจาก R_a^2 แปรผกผันกับ MSE หรือ R_a^2 เป็นฟังก์ชันของ MSE

2.1.5 การทดสอบสมมติฐาน

ในการวิเคราะห์การถดถอยเชิงเส้นตรงแบบพหุที่มีตัวแปรอิสระหลายตัวแปร สมมติฐานที่สนใจจะเกี่ยวข้องกับค่าของพารามิเตอร์ในรูปแบบ เช่น

β_i มีค่าเท่ากับค่าคงที่ค่าใดค่าหนึ่งที่กำหนดหรือไม่ จะเขียนได้เป็น $H_o: \beta_i = c$

$X_0'\underline{\beta}$ มีค่าเท่ากับค่าคงที่ค่าใดค่าหนึ่งที่กำหนดหรือไม่ จะเขียนได้เป็น $H_o: X_0'\underline{\beta} = c$

β_i สำหรับ $i = 1, 2, \ldots, m\ (m \le k)$ มีค่าเท่ากันหรือไม่ จะเขียนได้เป็น

$H_o: \beta_1 = \beta_2 = \ldots = \beta_m = 0$

β_i สำหรับ $i = 1, 2, \ldots, k$ มีค่าเท่ากันหรือไม่ จะเขียนได้เป็น $H_o: \beta_1 = \beta_2 = \ldots = \beta_k = 0$

การทดสอบสมมติฐานดังกล่าวข้างต้น จะใช้การทดสอบแบบ t แบบ F และแบบ F บางส่วน ดังรายละเอียดของการทดสอบต่อไปนี้

1. การทดสอบแบบ t ใช้สำหรับการทดสอบที่สมมติฐานหลักเขียนได้เป็นหนึ่งสมการของพารามิเตอร์ เช่น กรณีตัวแปรอิสระ k ตัว จะทดสอบ H_o ต่างๆ ได้แก่ $\beta_i = 0$ สำหรับ $i = 1, 2, \ldots, k$ $\beta_1 = \beta_2$ หรือ $\beta_1 - \beta_2 = 0$ เป็นต้น และแสดงการทดสอบสมมติฐานแบบ t ในตารางที่ 2.1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 2.1 แสดงสมมติฐาน ตัวทดสอบสถิติ และช่วงวิกฤติสำหรับการทดสอบแบบ t

H_0	H_1	ตัวทดสอบสถิติ	การแจกแจงของตัวทดสอบ	ช่วงวิกฤติ
1. $\beta_i = \beta_{i0}$	$\beta_i \neq \beta_{i0}$ $\beta_i > \beta_{i0}$ $\beta_i < \beta_{i0}$	$t = \frac{b_i - \beta_{i0}}{S_{b_i}}$ เมื่อ $S^2_{b_i} = a_{ii}S^2$	t ที่ชั้นแห่งความเป็นอิสระเท่ากับ n-k-1	$\lvert t\rvert \geq t_{\frac{\alpha}{2},(n-k-1)}$ $t \geq t_{\alpha,(n-k-1)}$ $t \leq -t_{\alpha,(n-k-1)}$
2. $\underline{X}_0'\underline{\beta} = c$	$\underline{X}_0'\underline{\beta} \neq c$ $\underline{X}_0'\underline{\beta} > c$ $\underline{X}_0'\underline{\beta} < c$	$t = \frac{\underline{X}_0'\underline{b} - c}{\sqrt{\underline{X}_0'(X'X)^{-1}\underline{X}_0 S^2}}$	t ที่ชั้นแห่งความเป็นอิสระเท่ากับ n-k-1	$\lvert t\rvert \geq t_{\frac{\alpha}{2},(n-k-1)}$ $t \geq t_{\alpha,(n-k-1)}$ $t \leq -t_{\alpha,(n-k-1)}$

เมื่อค่า a_{ii} เป็นสมาชิกแนวกลางที่ i ของเมตริกซ์ผกผัน $(X'X)^{-1}$

2. การทดสอบแบบ F จะใช้สำหรับการทดสอบสมมติฐานลักษณะต่างๆ ดังนี้

ก. จากรูปแบบเต็ม $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \varepsilon$ ทดสอบ

$H_0: \beta_1 = \beta_2 = \ldots \beta_k = 0$

$H_1: \beta_i$ สำหรับ $i = 1, 2, \ldots, k$ อย่างน้อยหนึ่งค่าไม่เป็น 0

ข. จากรูปแบบเต็ม $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \varepsilon$ ทดสอบ

$H_0: \beta_1 = \beta_2 = \beta_3 = 0$ เมื่อ $k > 3$

$H_1: \beta_1, \beta_2$ และ β_3 อย่างน้อยหนึ่งค่าไม่เป็น 0

ค. จากรูปแบบเต็ม $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \varepsilon$ ทดสอบ

$H_0: \beta_i = 0$

$H_1: \beta_i \neq 0$ สำหรับ $i = 1, 2, \ldots, k$

ตัวทดสอบสถิติ F ที่ใช้ในการทดสอบ จะเป็นผลจากการแบ่งส่วนของผลรวมกำลังสองรวม SST ออกเป็นส่วนๆ ซึ่งการแบ่งส่วนจะขึ้นอยู่กับสมมติฐานที่ต้องการทดสอบ ในที่นี้สมมติตัวแปรอิสระ 3 ตัวแปร สามารถทำการทดสอบสมมติฐานได้ดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1. จากรูปแบบเต็ม $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$ ต้องการทดสอบ

$H_O: \beta_1 = \beta_2 = \beta_3 = 0$

$H_1: \beta_1, \beta_2$ และ β_3 อย่างน้อยหนึ่งค่าไม่เป็น 0

รูปแบบลดรูปตามสมมติฐานหลักเป็น $Y = \beta_0 + \varepsilon$ จะแบ่งส่วน SST ออกเป็น

$$SST = SSR(X_1 X_2 X_3) + SSE(X_1 X_2 X_3)$$

ตัวทดสอบสถิติ F สำหรับการทดสอบนี้ จะเป็นตัวทดสอบสถิติที่เปรียบเทียบ SSR $(X_1X_2X_3)$ และ SSE$(X_1X_2X_3)$ ที่ปรับด้วยชั้นแห่งความเป็นอิสระที่เกี่ยวข้อง ดังนี้

$$F = \frac{SSR(X_1X_2X_3)/3}{SSE(X_1X_2X_3)/(n-4)}$$

$$= \frac{MSR(X_1X_2X_3)}{MSE(X_1X_2X_3)}$$

จะปฏิเสธ H_O ที่ระดับนัยสำคัญ α เมื่อ $F \geq F_{\alpha,(3,n-4)}$

2. จากรูปแบบเต็ม $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$ ต้องการทดสอบ

$H_O: \beta_3 = 0$

$H_1: \beta_3 \neq 0$

รูปแบบลดรูปตามสมมติฐานหลักเป็น $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$ จะแบ่งส่วน SST ออกเป็น

$$SST = SSR(X_1 X_2 X_3) + SSE(X_1 X_2 X_3)$$

จากนั้นจะแบ่ง SSR $(X_1 X_2 X_3)$ ออกเป็น

$$SSR(X_1 X_2 X_3) = SSR(X_3 / X_1 X_2) + SSR(X_1 X_2)$$

ตัวทดสอบสถิติ F เป็นตัวทดสอบสถิติที่ใช้ในการเปรียบเทียบ SSR(X_3/X_1X_2) กับ SSE $(X_1 X_2 X_3)$ ที่ปรับด้วยชั้นแห่งความเป็นอิสระที่เกี่ยวข้อง ดังนี้

$$F = \frac{SSR(X_3 / X_1X_2)}{SSE(X_1X_2X_3)/(n-4)}$$

$$= \frac{MSR(X_3 / X_1X_2)}{MSE(X_1X_2X_3)}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จะปฏิเสธ H_0 ที่ระดับนัยสำคัญ α เมื่อ $F \geq F_{\alpha,(1,n-4)}$

3. จากรูปแบบเต็ม $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$ ต้องการทดสอบ

$H_0 : \beta_2 = \beta_3 = 0$

$H_1 :$ β_2 และ β_3 อย่างน้อยหนึ่งค่าไม่เป็น 0

รูปแบบลดรูปตามสมมติฐานหลักเป็น $Y = \beta_0 + \beta_1 X_1 + \varepsilon$ จะแบ่งส่วน SST ออกเป็น

$$SST = SSR(X_1 X_2 X_3) + SSE(X_1 X_2 X_3)$$

และแบ่ง $SSR(X_1 X_2 X_3)$ ออกได้เป็น

$$SSR(X_1 X_2 X_3) = SSR(X_2 X_3 / X_1) + SSR(X_1)$$

ตัวทดสอบสถิติ F ในการทดสอบครั้งนี้ จะเป็นตัวทดสอบสถิติที่ใช้ในการเปรียบเทียบ $SSR(X_2 X_3 / X_1)$ กับ $SSE(X_1 X_2 X_3)$ ที่ปรับด้วยชั้นแห่งความเป็นอิสระที่เกี่ยวข้อง ดังนี้

$$F = \frac{SSR(X_2X_3/X_1)/2}{SSE(X_1X_2X_3)/(n-4)}$$

$$= \frac{MSR(X_2X_3/X_1)}{MSE(X_1X_2X_3)}$$

จะปฏิเสธ H_0 ที่ระดับนัยสำคัญ α เมื่อ $F \geq F_{\alpha,(2,n-4)}$

ในรูปแบบการถดถอยจะเรียกการทดสอบแบบ F สำหรับการทดสอบทุกตัวแปรอิสระไม่มีส่วนในการอธิบายความผันแปรของ Y ว่า การทดสอบแบบ F รวม (overall F test) และเรียกการทดสอบแบบ F สำหรับการทดสอบตัวแปรอิสระบางตัวไม่มีส่วนในการอธิบายความผันแปรของ Y ว่าการทดสอบแบบ F บางส่วน (partial F test) และสามารถแสดงการทดสอบสมมติฐานแบบ F ในตารางที่ 2.2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 2.2 แสดงสมมติฐาน ตัวทดสอบสถิติ และช่วงวิกฤติสำหรับการทดสอบแบบ F

H_0	H_1	ตัวทดสอบสถิติ	ช่วงวิกฤติ
$\beta_1=\beta_2=\beta_3=0$	β_i (i = 1 , 2 , 3) อย่างน้อยหนึ่งค่าไม่เป็น 0	$F = \frac{SSR(X_1X_2X_3)/3}{SSE(X_1X_2X_3)/(n-4)}$	$F \geq F_{\alpha,(3,n-4)}$
$\beta_3=0$	$\beta_3\neq 0$	$F = \frac{SSR(X_3/X_1X_2)}{SSE(X_1X_2X_3)/(n-4)}$	$F \geq F_{\alpha,(1,n-4)}$
$\beta_2=\beta_3=0$	β_2 และ β_3 อย่างน้อยหนึ่งค่าไม่เป็น 0	$F = \frac{SSR(X_2X_3/X_1)/2}{SSE(X_1X_2X_3)/(n-4)}$	$F \geq F_{\alpha,(2,n-4)}$

2.2 ตัวแปรดัมมี่กับการวิเคราะห์การถดถอยพหุ

ในการวิเคราะห์การถดถอยโดยทั่วไปจะเน้นศึกษาเฉพาะตัวแปรเชิงปริมาณ เพราะสามารถวัดค่าออก มาได้ ทำให้ง่ายต่อการนำไปวิเคราะห์ และมองข้ามตัวแปรเชิงคุณภาพซึ่งก็มีอิทธิพลต่อตัวแปรตามเช่นกัน ซึ่งตัวแปรเชิงคุณภาพจะถือเป็นตัวแปรจัดประเภท (Categorical variable) หรือตัวแปรดัมมี่ (Dummy variable) ดังนั้น ตัวแปรดัมมี่เป็นตัวแปรที่สร้างขึ้นเพื่อระบุกลุ่มหรือชุดที่ค่าสังเกตนั้นอยู่ จะกำหนดค่าของตัวแปรดัมมี่เป็น 1 หรือ 0 ซึ่งตัวแปรดัมมี่จะมีค่าเท่ากับ 1 เมื่อค่าสังเกตนั้นอยู่ในกลุ่มที่สนใจ และมีค่าเป็น 0 เมื่อค่าสังเกตนั้นไม่อยู่ในกลุ่มที่สนใจ กรณีที่แบ่งข้อมูลได้เป็น L กลุ่ม จะสร้างตัวแปรดัมมี่จำนวน L – 1 ตัวแปร จะเห็นว่า เมื่อข้อมูลที่รวบรวมมานั้นมีตัวแปรเชิงคุณภาพรวมอยู่ด้วย ก็สามารถนำเอาตัวแปรดัมมี่มาช่วยในการวิเคราะห์ได้ ในการวิเคราะห์การถดถอยเมื่อมีตัวแปรดัมมี่เป็นตัวแปรอิสระ เราสามารถวิเคราะห์การถดถอยโดยแยกเป็นกรณีต่างๆ ดังนี้

กรณีที่ 1 ตัวแปรดัมมี่สำหรับข้อมูลมีเพียงสองกลุ่ม

เมื่อข้อมูลแยกออกได้เป็นสองกลุ่ม โดยในแต่ละกลุ่ม ตัวแปรอิสระ X และตัวแปรตาม Y มีความสัมพันธ์แบบเส้นตรง จะกำหนดรูปแบบการถดถอยที่มีตัวแปรดัมมี่ D แทนกลุ่ม สิ่งที่สนใจจะพิจารณาจากรูปแบบการถดถอยดังกล่าว ได้แก่ สมการถดถอยเชิงเส้นตรงที่สร้างขึ้นจากข้อมูลสองกลุ่มมีความเหมือนหรือต่างกันอย่างไร ดังนี้

1. เส้นการถดถอยสองเส้นทับกันหรือไม่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. เส้นการถดถอยสองเส้นขนานกันหรือไม่ กล่าวคือ มีจุดที่เส้นการถดถอยตัดแกน Y ต่างกัน แต่มีความลาดชันเดียวกันหรือไม่
3. เส้นการถดถอยสองเส้นมีจุดที่เส้นการถดถอยตัดแกน Y เดียวกัน แต่มีความลาดชันต่างกันหรือไม่
4. เส้นการถดถอยสองเส้นมีจุดที่เส้นการถดถอยตัดแกน Y ต่างกัน และมีค่าความลาดชันต่างกันหรือไม่

ลักษณะของเส้นการถดถอย 4 ลักษณะข้างต้น จะสร้างจากรูปแบบการถดถอย 4 รูปแบบ เมื่อกำหนดตัวแปรดัมมี่ D แทนกลุ่ม โดย

$$D = \begin{cases} 1 & \text{เมื่อค่าสังเกตอยู่ในกลุ่มที่ 1} \\ 0 & \text{เมื่อค่าสังเกตอยู่ในกลุ่มที่ 2} \end{cases}$$

และกำหนดตัวแปร DX เป็นตัวแปรร่วมระหว่างตัวแปรดัมมี่ D และตัวแปรอิสระ X ดังตารางที่ 2.3

ตารางที่ 2.3 แสดงรูปแบบการถดถอยและสมการถดถอยสำหรับข้อมูลสองกลุ่ม

รูปแบบการถดถอย	สมการถดถอย
$Y = \beta_0 + \beta_1 X + \varepsilon$	$\hat{Y} = b_0 + b_1 X$ สำหรับทั้งสองกลุ่ม
$Y = \beta_0 + \beta_1 X + \beta_2 D + \varepsilon$	$\hat{Y} = \begin{cases} (b_0 + b_2) + b_1 X & \text{สำหรับ } D = 1 \\ b_0 + b_1 X & \text{สำหรับ } D = 0 \end{cases}$
$Y = \beta_0 + \beta_1 X + \beta_3 DX + \varepsilon$	$\hat{Y} = \begin{cases} b_0 + (b_1 + b_3)X & \text{สำหรับ } D = 1 \\ b_0 + b_1 X & \text{สำหรับ } D = 0 \end{cases}$
$Y = \beta_0 + \beta_1 X + \beta_2 D + \beta_3 DX + \varepsilon$	$\hat{Y} = \begin{cases} (b_0 + b_2) + (b_1 + b_3)X & \text{สำหรับ } D = 1 \\ b_0 + b_1 X & \text{สำหรับ } D = 0 \end{cases}$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

และสามารถเขียนรูปแบบการถดถอยได้ดังนี้

1. รูปแบบการถดถอยทั่งไป

$$Y = \beta_0 + \beta_1 X + \beta_2 D + \beta_3 DX + \varepsilon$$

จะแทนกรณีทั้ง 2 กลุ่ม มีเส้นการถดถอยที่ต่างกันทั้งจุดที่เส้นการถดถอยตัดแกน Y และค่าความลาดชัน สามารถทำการทดสอบสมมติฐานได้ดังนี้

$$H_0 : \beta_2 = \beta_3 = 0$$

$$H_1 : \beta_i \text{ อย่างน้อยหนึ่งค่าที่ไม่เป็น } 0 \ ; \ i = 2, 3$$

ตัวสถิติที่ใช้ในการทดสอบคือ

$$F = \frac{SSR(D, DX / X)/2}{SSE(X, D, DX)/(n-4)}$$

และจะปฏิเสธ H_0 เมื่อ $F \geq F_{\alpha,(2,n-4)}$

2. เมื่อ $\beta_2 = 0$ จะเขียนรูปแบบการถดถอยทั่วไปใหม่เป็น

$$Y = \beta_0 + \beta_1 X + \beta_3 DX + \varepsilon$$

รูปแบบใหม่จะแทนกรณีที่ทั้งสองกลุ่มมีเส้นการถดถอยซึ่งมีจุดที่เส้นการถดถอยตัดแกน Y เดียวกันแต่มีความลาดชันต่างกัน สามารถทำการทดสอบสมมติฐานได้ดังนี้

$$H_0 : \beta_2 = 0$$

$$H_1 : \beta_2 \neq 0$$

ตัวสถิติที่ใช้ในการทดสอบคือ

$$F = \frac{SSR(D / X, DX)/1}{SSE(X, D, DX)/(n-4)}$$

และจะปฏิเสธ H_0 เมื่อ $F \geq F_{\alpha,(1,n-4)}$

3. เมื่อ $\beta_3 = 0$ จะเขียนรูปแบบการถดถอยทั่วไปใหม่เป็น

$$Y = \beta_0 + \beta_1 X + \beta_2 D + \varepsilon$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปแบบใหม่จะแทนกรณีทั้งสองกลุ่มมีเส้นถดถอยที่ขนานกัน นั่นคือ มีจุดที่เส้นการถดถอยตัดแกน Y ต่างกันแต่มีความลาดชันเท่ากัน สามารถทำการทดสอบสมมติฐานได้ดังนี้

$$H_0 : \beta_3 = 0$$

$$H_1 : \beta_3 \neq 0$$

ตัวสถิติที่ใช้ในการทดสอบคือ

$$F = \frac{SSR(DX/X,D)/1}{SSE(X,D,DX)/(n-4)}$$

และจะปฏิเสธ H_0 เมื่อ $F \geq F_{\alpha,(1,n-4)}$

เมื่อทดสอบสมมติฐานพบว่า ความสัมพันธ์ระหว่างตัวแปรอิสระ X และตัวแปรตาม Y ของแต่ละกลุ่มเป็นอย่างใดแล้ว จะสร้างสมการถดถอยตามรูปแบบการถดถอยที่ทดสอบได้ และนำสมการถดถอยที่ได้ไปอธิบายความสัมพันธ์ระหว่างตัวแปรอิสระ X และตัวแปรตาม Y สำหรับข้อมูลสองกลุ่ม

กรณีที่ 2 ตัวแปรดัมมี่สำหรับข้อมูลมีมากกว่าสองกลุ่ม

เมื่อข้อมูลแยกออกได้เป็นหลายกลุ่ม โดยในแต่ละกลุ่มตัวแปรอิสระX และตัวแปรตาม Y มีความสัมพันธ์กันแบบเส้นตรง การตรวจสอบว่าเส้นการถดถอยในแต่ละกลุ่มเป็นเส้นเดียวกัน เป็นเส้นขนานกันหรือมีจุดที่เส้นการถดถอยตัดแกน Y ต่างกันแต่มีความลาดชันเดียวกัน เป็นเส้นที่จุดที่เส้นการถดถอยตัดแกน Y เดียวกันแต่ความลาดชันต่างกัน หรือเป็นเส้นที่มีจุดที่เส้นการถดถอยตัดแกน Y ต่างกันและค่าความลาดชันต่างกันหรือไม่ จะทำได้โดยการกำหนดตัวแปรดัมมี่แทนกลุ่ม จำนวนตัวแปรดัมมี่จะเท่ากับจำนวนกลุ่มหักออกด้วย 1 เช่น กรณีแบ่งข้อมูลออกเป็น 3 กลุ่ม จะกำหนดตัวแปรดัมมี่สองตัวแปรได้แก่ D_1 และ D_2 เมื่อ

$$D_1 = \begin{cases} 1 & \text{เมื่อค่าสังเกตอยู่ในกลุ่มที่ 1} \\ 0 & \text{เมื่อค่าสังเกตไม่อยู่ในกลุ่มที่ 1} \end{cases}$$

$$D_2 = \begin{cases} 1 & \text{เมื่อค่าสังเกตอยู่ในกลุ่มที่ 2} \\ 0 & \text{เมื่อค่าสังเกตไม่อยู่ในกลุ่มที่ 2} \end{cases}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

142087

และกำหนดตัวแปร D_1X , D_2X เป็นตัวแปรร่วมระหว่างตัวแปรดัมมี่ D และตัวแปรอิสระ X ดังในตารางที่ 2.4

ตารางที่ 2.4 แสดงรูปแบบการถดถอยและสมการถดถอยสำหรับข้อมูลสามกลุ่ม

รูปแบบการถดถอย	สมการถดถอย
1. $Y = \beta_0 + \beta_1 X + \varepsilon$	$\hat{Y} = b_0 + b_1 X$ สำหรับทั้งสามกลุ่ม
2. $Y = \beta_0 + \beta_1 X + \beta_2 D_1 + \beta_3 D_2 + \varepsilon$	$\hat{Y} = \begin{cases} (b_0 + b_2) + b_1 X & \text{สำหรับ } D_1 = 1 , D_2 = 0 \\ (b_0 + b_3) + b_1 X & \text{สำหรับ } D_1 = 0 , D_2 = 1 \\ b_0 + b_1 X & \text{สำหรับ } D_1 = 0 , D_2 = 0 \end{cases}$
3. $Y = \beta_0 + \beta_1 X + \beta_4 D_1 X + \beta_5 D_2 X + \varepsilon$	$\hat{Y} = \begin{cases} b_0 + (b_1 + b_4) X & \text{สำหรับ } D_1 = 1 , D_2 = 0 \\ b_0 + (b_1 + b_5) X & \text{สำหรับ } D_1 = 0 , D_2 = 1 \\ b_0 + b_1 X & \text{สำหรับ } D_1 = 0 , D_2 = 0 \end{cases}$
4. $Y = \beta_0 + \beta_1 X + \beta_2 D_1 + \beta_3 D_2 + \beta_4 D_1 X + \beta_5 D_2 X + \varepsilon$	$\hat{Y} = \begin{cases} (b_0 + b_2) + (b_1 + b_4) X & \text{สำหรับ } D_1 = 1 , D_2 = 0 \\ (b_0 + b_3) + (b_1 + b_5) X & \text{สำหรับ } D_1 = 0 , D_2 = 1 \\ b_0 + b_1 X & \text{สำหรับ } D_1 = 0 , D_2 = 0 \end{cases}$

และสามารถเขียนรูปแบบการถดถอยได้ดังนี้

1. รูปแบบการถดถอยทั่วไป

$$Y = \beta_0 + \beta_1 X + \beta_2 D_1 + \beta_3 D_2 + \beta_4 D_1 X + \beta_5 D_2 X + \varepsilon$$

จะแทนกรณีทั้ง3กลุ่ม มีเส้นการถดถอยที่ต่างกันทั้งจุดที่เส้นการถดถอยตัดแกน Y และค่าความลาดชัน สามารถทำการทดสอบสมมติฐานได้ดังนี้

$H_0 : \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$

$H_1 : \beta_i$ อย่างน้อยหนึ่งค่าที่ไม่เป็น 0 ; $i = 2 , 3 , 4 , 5$

ตัวสถิติที่ใช้ในการทดสอบคือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$F=\frac{SSR(D_1,D_2,D_1X,D_2X/X)/4}{SSE(X,D_1,D_2,D_1X,D_2X)/(n-6)}$$

และจะปฏิเสธ H_o เมื่อ $F \geq F_{\alpha,(4,n-6)}$

2. เมื่อ $\beta_4 = \beta_5 = 0$ จะเขียนรูปแบบการถดถอยทั่วไปใหม่เป็น

$$Y = \beta_0 + \beta_1 X + \beta_2 D_1 + \beta_3 D_2 + \varepsilon$$

รูปแบบใหม่จะแทนกรณีทั้งสามกลุ่มมีเส้นถดถอยที่ขนานกัน นั่นคือ มีจุดที่เส้นการถดถอยตัดแกน Y ต่างกันแต่มีความลาดชันเท่ากัน สามารถทำการทดสอบสมมติฐานได้ดังนี้

$H_o : \beta_4 = \beta_5 = 0$

$H_1 : \beta_i$ อย่างน้อยหนึ่งค่าที่ไม่เป็น 0 ; $i = 4, 5$

ตัวสถิติที่ใช้ในการทดสอบคือ

$$F=\frac{SSR(D_1X,D_2X/X,D_1,D_2)/2}{SSE(X,D_1,D_2,D_1X,D_2X)/(n-6)}$$

และจะปฏิเสธ H_o เมื่อ $F \geq F_{\alpha,(2,n-6)}$

3. เมื่อ $\beta_2 = \beta_3 = 0$ จะเขียนรูปแบบการถดถอยทั่วไปใหม่เป็น

$$Y = \beta_0 + \beta_1 X + \beta_4 D_1 X + \beta_5 D_2 X + \varepsilon$$

รูปแบบใหม่จะแทนกรณีที่ทั้งสามกลุ่มมีเส้นการถดถอยซึ่งมีจุดที่เส้นการถดถอยตัดแกน Y เดียวกันแต่มีความลาดชันต่างกัน สามารถทำการทดสอบสมมติฐานได้ดังนี้

$H_o : \beta_2 = \beta_3 = 0$

$H_1 : \beta_i$ อย่างน้อยหนึ่งค่าที่ไม่เป็น 0 ; $i = 2, 3$

ตัวสถิติที่ใช้ในการทดสอบคือ

$$F=\frac{SSR(D_1,D_2/X,D_1X,D_2X)/2}{SSE(X,D_1,D_2,D_1X,D_2X)/(n-6)}$$

และจะปฏิเสธ H_o เมื่อ $F \geq F_{\alpha,(2,n-6)}$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.3 การทดสอบสมมติฐานเกี่ยวกับค่าเฉลี่ยของประชากร 2 กลุ่มที่ไม่เป็นอิสระกัน

เป็นการศึกษาถึงสองประชากรที่ไม่อิสระกัน ซึ่งกลุ่มตัวอย่าง 2 กลุ่ม จะมีลักษณะดังนี้คือ

1. มีกลุ่มตัวอย่างเดียว แต่เก็บข้อมูล 2 ครั้ง เช่น ผู้วิจัยต้องการทราบผลการอบรมพนักงานทำให้ปริมาณเพิ่มขึ้นหรือไม่ จึงสุ่มพนักงานมากลุ่มหนึ่ง บันทึกปริมาณงานแต่ละวันของพนักงานกลุ่มนี้ และหลังจากการอบรมแล้ว บันทึกปริมาณงานของพนักงานกลุ่มนี้อีกครั้ง ตัวเลขคู่ลำดับของพนักงานแต่ละคนที่บันทึกไว้ก่อนและหลังการอบรม จะเป็นค่าของข้อมูลจากกลุ่มตัวอย่างที่ไม่เป็นอิสระกัน หรือเป็นข้อมูลที่จับคู่กัน (Matched pair).

2. แบ่งกลุ่มตัวอย่างออกเป็นคู่ๆ และในแต่ละคู่จะแยกเป็น 2 กลุ่มย่อย โดยถือว่าข้อมูลทั้ง 2 กลุ่มมีลักษณะคล้ายกัน เช่น มีคะแนนผลสัมฤทธิ์ทางการเรียนเท่ากันเป็นคู่ๆ กลุ่มตัวอย่างย่อย 2 กลุ่มนี้จึงมีความสัมพันธ์กันหรือไม่เป็นอิสระกัน คุณสมบัติหรือลักษณะที่ใช้เป็นหลักในการจับคู่นี้ นอกจากจะใช้คะแนนผลสัมฤทธิ์ทางการเรียน อาจเป็น อายุ น้ำหนัก หรือลักษณะอื่นใดก็ได้ที่เหมาะสมกับปัญหาที่สนใจในครั้งนี้

สมมติตัวอย่างขนาด n จะมีค่าสังเกตที่วัดได้เป็นคู่ลำดับ ดังนี้

$(X_{11}, X_{12}), (X_{21}, X_{22}), \ldots, (X_{n1}, X_{n2})$

ค่าความแตกต่างระหว่างแต่ละคู่จะเป็นดังนี้

$d_i = X_{i1} - X_{i2} \quad ; \quad i = 1, 2, \ldots, n$

ให้ : $\mu_d = \mu_1 - \mu_2$ เมื่อ μ_1, μ_2 คือค่าเฉลี่ยของประชากรแต่ละกลุ่มย่อย

โดยมีขั้นตอนในการทดสอบดังต่อไปนี้

ขั้นที่ 1 ตั้งสมมติฐานแบบใดแบบหนึ่งเกี่ยวกับผลต่างค่าเฉลี่ยของประชากรที่ไม่เป็นอิสระกัน

แบบ A	$H_0 : \mu_d = d_0$	;	$H_1 : \mu_d \neq d_0$
แบบ B	$H_0 : \mu_d = d_0$	;	$H_1 : \mu_d < d_0$
แบบ C	$H_0 : \mu_d = d_0$	;	$H_1 : \mu_d > d_0$

ขั้นที่ 2 คำนวณค่าสถิติที่ใช้ในการทดสอบสมมติฐานภายใต้ H_0 ดังนี้

กรณีที่ 1 ถ้าประชากรมีการแจกแจงแบบปกติ และทราบค่า σ_d^2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$Z = \frac{\bar{d} - d_0}{\sigma_d / \sqrt{n}} \sim N(0, 1)$$

กรณีที่ 2 ถ้าประชากรมีการแจกแจงแบบปกติ แต่ไม่ทราบค่า σ_d^2

$$t = \frac{\bar{d} - d_0}{S_d / \sqrt{n}} \sim t_{df=n-1}$$

กรณีที่ 3 ถ้าไม่ทราบการแจกแจงของประชากร แต่ขนาด n มีค่ามาก ($n \geq 30$)

$$Z = \frac{\bar{d} - d_0}{S_d / \sqrt{n}} \sim N(0, 1)$$

เมื่อ $\bar{d} = \frac{\sum_{i=1}^{n} d_i}{n}$; $S_d^2 = \frac{\sum_{i=1}^{n} (d_i - \bar{d})^2}{n-1}$

ขั้นที่ 3 กำหนดระดับนัยสำคัญเท่ากับ α และกำหนดช่วงวิกฤตสำหรับการทดสอบสมมติฐานได้ดังตารางที่ 2.5

ตารางที่ 2.5 แสดงสมมติฐาน ตัวทดสอบสถิติ และช่วงวิกฤติสำหรับการทดสอบสมมติฐานเกี่ยวกับค่าเฉลี่ยของประชากร 2 กลุ่มที่ไม่เป็นอิสระกัน

H_0	H_1	การแจกแจงของตัวทดสอบ	ช่วงวิกฤติ
$\mu_d = d_0$	$\mu_d \neq d_0$	$Z \sim N(0, 1)$ $t \sim t_{df=n-1}$	$\lvert Z \rvert \geq Z_{\frac{\alpha}{2}}$ $\lvert t \rvert \geq t_{\frac{\alpha}{2},(n-1)}$
$\mu_d = d_0$	$\mu_d < d_0$	$Z \sim N(0, 1)$ $t \sim t_{df=n-1}$	$Z \leq -Z_{\alpha}$ $t \leq -t_{\alpha,(n-1)}$
$\mu_d = d_0$	$\mu_d > d_0$	$Z \sim N(0, 1)$ $t \sim t_{df=n-1}$	$Z \geq Z_{\alpha}$ $t \geq t_{\alpha,(n-1)}$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.4 การหาคุณภาพแบบทดสอบ

คุณภาพของแบบทดสอบเกี่ยวข้องกับองค์ประกอบที่สำคัญ ดังนี้

1. ความเที่ยงตรง (Validity)

ความเที่ยงตรงถือเป็นคุณภาพของแบบทดสอบ หมายความว่า แบบทดสอบที่ผู้สอนได้สร้างไว้สามารถวัดได้ตรงตามวัตถุประสงค์ที่ต้องการจะวัด แบบทดสอบที่ดีจะต้องนำไปทดสอบเพื่อหาคุณภาพด้านความเที่ยงตรง จะถือได้ว่าเป็นแบบทดสอบที่มีคุณภาพตามวัตถุประสงค์ที่จะวัด และผลที่ได้จากการวัดจะถูกต้องตรงตามความต้องการ ความเที่ยงตรงของแบบทดสอบจำแนกเป็น 3 แบบ ดังนี้

1.1 ความเที่ยงตรงตามเนื้อหา

ความเที่ยงตรงตามเนื้อหา (content validity) หมายถึง การที่ผู้สอนออกแบบทดสอบได้ตรงตามเนื้อหาที่สอน ในการทดสอบความเที่ยงตรงตามเนื้อหาสามารถดำเนินการได้โดยใช้ผู้เชี่ยวชาญในด้านเนื้อหา พิจารณาถึงความสอดคล้องระหว่างวัตถุประสงค์กับแบบทดสอบ โดยพิจารณาเป็นรายข้อ วิธีการพิจารณาแบบนี้จะเรียกว่า การหาค่าสัมประสิทธิ์ความสอดคล้อง (Index of Item-Objective Congruence :IQC) โดยมีสูตรการคำนวณดังนี้

$$IQC = \frac{\sum R}{N}$$

เมื่อ IQC คือ ความสอดคล้องระหว่างวัตถุประสงค์กับแบบทดสอบ

$\sum R$ คือ ผลรวมของคะแนนจากผู้เชี่ยวชาญทั้งหมด

N คือ จำนวนผู้เชี่ยวชาญ

ในการตรวจสอบค่าความเที่ยงตรงด้านเนื้อหาสามารถทำได้ โดยนำแบบทดสอบให้ผู้เชี่ยวชาญพิจารณาว่า ข้อสอบแต่ละข้อมีความสอดคล้องกับวัตถุประสงค์เชิงพฤติกรรมหรือไม่อย่างไร ถ้ามีความสอดคล้องกันผู้เชี่ยวชาญจะให้ค่าเป็น “+1” แต่ถ้าผู้เชี่ยวชาญเห็นว่าข้อสอบข้อนั้นไม่มีความสอดคล้องกับวัตถุประสงค์จะให้ค่าเป็น “-1” และในกรณีที่ผู้เชี่ยวชาญไม่แน่ใจว่าข้อสอบข้อนั้นมีความสอดคล้องกับวัตถุประสงค์หรือไม่ ก็จะให้ค่าเป็น “0”

1.2 ความเที่ยงตรงตามเกณฑ์สัมพันธ์

ความเที่ยงตรงตามเกณฑ์สัมพันธ์ (criterion related validity) หมายถึง การวัดคุณภาพของแบบทดสอบ โดยเอาผลการวัดของแบบทดสอบไปหาความสัมพันธ์กับเกณฑ์ที่กำหนด เช่น ระดับผลการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เรียน เป็นต้น ถ้าผู้เรียนที่มีระดับผลการเรียนดี เมื่อทำข้อสอบชุดนั้นแล้วพบว่าได้คะแนนสูง แสดงว่าแบบทดสอบนั้นมีความเที่ยงตรงตามเกณฑ์สัมพันธ์ดี แต่ถ้ามีผลตรงกันข้าม แสดงว่า แบบทดสอบนั้นไม่มีความเที่ยงตรง การทดสอบความเที่ยงตรงตามเกณฑ์สัมพันธ์จัดแบ่งออกเป็น 2 ชนิด ดังนี้

(1) ความเที่ยงตรงเชิงสภาพ (concurrent validity) หมายถึง การนำเอาผลการวัดจากแบบทดสอบไปหาความสัมพันธ์กับผลการเรียนอื่นๆ ของผู้เรียนในปัจจุบัน เช่น การนำเอาผลการวัดจากแบบทดสอบเกี่ยวกับคอมพิวเตอร์เบื้องต้นที่สร้างขึ้น ไปหาความสัมพันธ์กับคะแนนการปฏิบัติการใช้โปรแกรมคอมพิวเตอร์ต่างๆ เป็นต้น ถ้าผลการหาความสัมพันธ์มีความสัมพันธ์กันสูง กล่าวคือ ผู้เรียนที่มีทักษะการปฏิบัติการงานด้านโปรแกรมคอมพิวเตอร์สูงจะทำแบบทดสอบนั้นได้ ทำนองเดียวกันคนที่ไม่มีทักษะการปฏิบัติการใช้โปรแกรมคอมพิวเตอร์ก็จะทำแบบทดสอบนั้นไม่ได้ ถ้าผลการหาความสัมพันธ์เป็นไปในทางเดียวกันคือ มีความสัมพันธ์กันสูง แสดงว่าแบบทดสอบนั้นมีความเที่ยงตรงเชิงสภาพสูง

การทดสอบความเที่ยงตรงเชิงสภาพ สามารถดำเนินการโดยการคำนวณหาค่าสัมประสิทธิสหสัมพันธ์ (correlation coefficient) ของเพียรสัน (Pearson) มีสูตรดังนี้

$$r_{XY} = \frac{N\sum XY - \sum X \sum Y}{\sqrt{\left[N\sum X^2 - (\sum X)^2\right]\left[N\sum Y^2 - (\sum Y)^2\right]}}$$

เมื่อ r_{XY} = สัมประสิทธิ์ความเชื่อมั่นของแบบทดสอบ

N = จำนวนผู้เรียนที่ทำแบบทดสอบ

$\sum X$ = ผลรวมคะแนนแบบทดสอบที่หาความเที่ยงตรงเชิงสภาพ

$\sum Y$ = ผลรวมคะแนนความรู้ของผู้เรียนที่เป็นเกณฑ์ หรืออาจใช้เกรดเฉลี่ย

(2) ความเที่ยงตรงเชิงพยากรณ์(predictive validity) เป็นการทดสอบความเที่ยงตรงที่ใช้หลักการเดียวกับการทดสอบความเที่ยงตรงเชิงสภาพ เพียงแต่ถ้าเป็นแบบเชิงพยากรณ์จะใช้ข้อมูลที่เป็นเกณฑ์ในอนาคต ไปหาความสัมพันธ์กับคะแนนจากการทำแบบทดสอบที่สร้างขึ้น เช่น ใช้ข้อมูลที่เป็นเกรดเฉลี่ยของผู้เรียนที่สำเร็จการศึกษาแล้ว หรือใช้คะแนนในรายวิชาใดๆ ที่ได้สอบผ่านไปแล้วมาเป็นเกณฑ์ เป็นต้น

ในการดำเนินการนั้น จะต้องทำการทดสอบผู้เรียนด้วยแบบทดสอบที่ต้องการหาความเที่ยงตรงเชิงพยากรณ์ก่อน เมื่อทดสอบแล้วจะต้องรอให้ผู้เรียนกลุ่มนี้ได้คะแนนที่จะใช้เป็นเกณฑ์ในการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หาความสัมพันธ์ จึงจะดำเนินการคำนวณหาความเที่ยงตรงเชิงพยากรณ์ได้ เช่น ถ้าต้องการทดสอบความเที่ยงตรงเชิงพยากรณ์ของแบบทดสอบวิชาคอมพิวเตอร์เบื้องต้น โดยใช้คะแนนจากผลการทดสอบวิชาคณิตศาสตร์เป็นเกณฑ์เทียบความสัมพันธ์ แต่เนื่องจากคะแนนจากผลการทดสอบวิชาคณิตศาสตร์ในขณะนั้นยังไม่มี เนื่องจากผู้เรียนกลุ่มที่จะใช้ทดลองนั้น ยังไม่ได้ผ่านการทดสอบวิชาคณิตศาสตร์ ดังนั้นจะต้องทดสอบผู้เรียนด้วยแบบทดสอบวิชาคอมพิวเตอร์เบื้องต้นก่อน จากนั้นให้รอจนกระทั่งผู้เรียนกลุ่มนี้ได้ผ่านการทดสอบวิชาคณิตศาสตร์ จึงสามารถนำคะแนนการทดสอบวิชาคณิตศาสตร์มาคำนวณร่วมกับคะแนนจากการทดสอบวิชาคอมพิวเตอร์เบื้องต้นที่ได้ทำการทดสอบก่อนหน้า เพื่อหาค่าความเที่ยงตรงเชิงพยากรณ์ ใช้สูตรเดียวกันกับการหาค่าความเที่ยงตรงเชิงสภาพ

1.3 ความเที่ยงตรงตามโครงสร้าง

ความเที่ยงตรงตามโครงสร้าง (construct validity) หมายถึง การวัดคุณภาพของแบบทดสอบว่าตรงตามลักษณะโครงสร้าง หรือวัดได้ครอบคลุมตามลักษณะโครงสร้างหรือไม่ โดยที่โครงสร้างหมายถึง โครงสร้างของแบบทดสอบมาตรฐาน โดยแบบทดสอบที่สร้างขึ้นจะมีมาตรฐานที่วัดลักษณะเดียวกันกับแบบทดสอบมาตรฐานหรือไม่ สามารถคำนวณหาความเที่ยงตรงตามโครงสร้างได้โดยใช้ค่าสัมประสิทธิ์สหสัมพันธ์ของเพียรสัน ซึ่งค่า X คือคะแนนที่ได้จากแบบทดสอบที่สร้างขึ้น และค่า Y คือ ค่าคะแนนที่ได้จากแบบทดสอบมาตรฐานที่วัดลักษณะเดียวกัน เมื่อคำนวณค่าได้แล้ว พบว่า ถ้าค่าที่คำนวณได้เข้าใกล้ 1 หมายถึง แบบทดสอบที่สร้างขึ้นนั้นมีความเที่ยงตรงเชิงโครงสร้างสูง ในทางกลับกัน ถ้าค่าที่คำนวณได้มีค่าเข้าใกล้ 0 แสดงว่าแบบทดสอบนั้นไม่มีความเที่ยงตรงเชิงโครงสร้าง

2. ความเชื่อมั่น (Reliability)

ความเชื่อมั่น หมายถึง ความคงเส้นคงวาของผลการวัดจากการที่นำแบบทดสอบชุดนั้นไปทดสอบกับผู้เรียนไม่ว่าจะทดสอบจำนวนกี่ครั้งคะแนนที่ได้จะไม่แตกต่างกัน ค่าความเชื่อมั่นนี้สามารถคำนวณเป็นตัวเลขได้หลายวิธี และแต่ละวิธีจะได้ค่าไม่เกิน 1 ถ้าค่าที่คำนวณได้มีค่าเข้าใกล้ 1 แสดงว่าแบบทดสอบนั้นมีค่าความเชื่อมั่นสูง วิธีการคำนวณหาค่าความเชื่อมั่นสามารถคำนวณหาค่าได้ หลายวิธีดังนี้

2.1 วิธีการสอบซ้ำ

วิธีการสอบซ้ำ(test-retest) เป็นวิธีการหาค่าความเชื่อมั่นของแบบสอบถามในความหมายของคำว่า ความคงที่(stability) โดยคะแนนที่ได้จากการสอบ 2 ครั้งจะต้องไม่มีความแตกต่างกัน ในการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

วัดผลจะวัดในเวลาที่ต่างกัน แล้วนำคะแนนที่ได้ทั้ง 2 ครั้งมาคำนวณหาค่าสหสัมพันธ์โดยใช้สูตรของเพียร์สัน

2.2 วิธีการใช้แบบทดสอบคู่ขนาน

วิธีการใช้แบบทดสอบคู่ขนาน(parallel form) หมายถึง การทดสอบความเชื่อมั่นโดยใช้แบบทดสอบ 2 ชุดที่มีเนื้อหาเดียวกัน ความยากง่ายระดับเดียวกัน มีโครงสร้างเดียวกัน จำนวนข้อเท่ากัน ไปทดสอบกับกลุ่มผู้เรียนทั้ง 2 ฉบับ นำคะแนนที่ได้ไปคำนวณหาค่าสหสัมพันธ์ โดยใช้สูตรของเพียร์สัน เหมือนกับวิธีการสอบซ้ำ

2.3 วิธีการแบ่งครึ่งแบบทดสอบ

วิธีการแบ่งครึ่งแบบทดสอบ (split-half) หมายถึง การนำเอาแบบทดสอบที่สร้างขึ้นโดยจัดแบ่งเป็น 2 ฉบับ จัดแบ่งตามข้อคู่และข้อคี่ ในการคำนวณหาค่าความเชื่อมั่นจะใช้หลักการเดียวกับวิธีการสอบคู่ขนาน แต่ค่าสหสัมพันธ์ที่คำนวณได้นี้จะเป็นค่าความเชื่อมั่นเพียงครึ่งฉบับเท่านั้น ดังนั้นจึงต้องนำค่าที่ได้ไปคำนวณหาความเชื่อมั่นทั้งฉบับ โดยอาศัยสูตรของสเพียร์แมน บราวน์(Spearman-Brown) ดังนี้

$$r_t = \frac{2r_{\frac{1}{2}}}{1+r_{\frac{1}{2}}}$$

เมื่อ r_t คือ สัมประสิทธิ์ความเชื่อมั่นแบบทดสอบทั้งฉบับ

$r_{\frac{1}{2}}$ คือ สัมประสิทธิ์ความเชื่อมั่นของแบบทดสอบครึ่งฉบับ

2.4 วิธีแบบคูเดอร์-ริชาร์ดสัน

การหาความเชื่อมั่นโดยวิธีของคูเดอร์-ริชาร์ดสัน(Kuder-Richardson:KR) โดยวิธีการนี้จะแตกต่างจากวิธีการหาความเชื่อมั่นแบบต่างๆ ที่กล่าวมา จะไม่ได้ใช้การหาค่าสหสัมพันธ์เพื่อทดสอบความเชื่อมั่น แต่จะใช้วิธีหาความสัมพันธ์ระหว่างองค์ประกอบภายใน ได้แก่ ความสัมพันธ์ระหว่างข้อสอบในฉบับเดียวกัน และการคำนวณหาค่าความสัมพันธ์คะแนนของข้อสอบแต่ละข้อ จะต้องแปลงให้เป็นคะแนน 2 ค่าเท่านั้น ได้แก่ ถ้าตอบถูกจะได้ค่า 1 และถ้าตอบผิดจะได้ค่า 0 สูตรในการหาความเชื่อมั่นแบบคูเดอร์-ริชาร์ดสัน จะจำแนกเป็น 2 สูตร ดังรายละเอียดต่อไปนี้

(1) KR-20 เป็นสูตรการหาค่าความเชื่อมั่นที่เหมาะสมสำหรับแบบทดสอบที่มีค่าความยากง่ายในลักษณะกระจาย สูตรที่ใช้ในการหามีรูปแบบดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$r_t = \frac{k}{k-1}\left[1-\frac{\sum pq}{S_t^2}\right]$$

เมื่อ r_t คือ สัมประสิทธิ์ของความเชื่อมั่นของแบบทดสอบทั้งฉบับ

k คือ จำนวนข้อของแบบทดสอบ

p คือ สัดส่วนของผู้เรียนที่ทำข้อสอบข้อนั้นถูกกับผู้เรียนทั้งหมด

q คือ สัดส่วนของผู้เรียนที่ทำข้อสอบข้อนั้นผิดกับผู้เรียนทั้งหมด

S_t^2 คือ ความแปรปรวนของคะแนนสอบทั้งฉบับ

(2) KR-21 เป็นสูตรในการหาค่าความเชื่อมั่นที่เหมาะสำหรับแบบทดสอบที่มีความยากง่ายของข้อสอบแต่ละข้อมีค่าใกล้เคียงกัน สูตรที่ใช้ในการคำนวณมีรูปแบบดังนี้

$$r_t = \frac{k}{k-1}\left[1-\frac{\bar{X}(n-\bar{X})}{kS_t^2}\right]$$

เมื่อ r_t คือ สัมประสิทธิ์ของความเชื่อมั่นของแบบทดสอบทั้งฉบับ

k คือ จำนวนข้อของแบบทดสอบ

$\bar{X}$ คือ ค่าเฉลี่ยของคะแนน

S_t^2 คือ ความแปรปรวนของคะแนนสอบทั้งฉบับ

2.5 วิธีการหาสัมประสิทธิ์แอลฟา

สัมประสิทธิ์แอลฟา(α – Coefficient) ของแบบทดสอบ เป็นค่าความเชื่อมั่นที่คำนวณหาได้จากสูตรครอนบราช (Cronbach) การหาค่าความเชื่อมั่นของแบบทดสอบนั้นจะหาค่าคะแนนที่ได้ของแบบทดสอบ อาจจะเป็นค่าอะไรก็ได้ที่มีค่ามากกว่า 1 สูตรที่ใช้ในการคำนวณมีดังนี้

$$r_t = \frac{k}{k-1}\left[1-\frac{\sum S_i^2}{S_t^2}\right]$$

เมื่อ α คือ ค่าสัมประสิทธิ์ความเชื่อมั่นของแบบทดสอบ

k คือ จำนวนข้อของแบบทดสอบ

S_i^2 คือ ความแปรปรวนของแบบทดสอบรายข้อ

S_t^2 คือ ความแปรปรวนของแบบทดสอบทั้งฉบับ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. อำนาจจำแนก (Discrimination)

อำนาจจำแนก หมายถึง การที่ข้อคำถามสามารถจัดแบ่งผู้เรียนออกเป็น 2 กลุ่มได้ โดยกลุ่มผู้เรียน 2 กลุ่มในที่นี้คือ ผู้เรียนกลุ่มเก่งและผู้เรียนกลุ่มอ่อน หรือกลุ่มที่ชอบหรือกลุ่มที่ไม่ชอบ ค่าอำนาจจำแนกที่คำนวณได้จะมีค่าอยู่ระหว่าง -1 ถึง 1 สามารถแปลความหมายค่าอำนาจจำแนกของข้อสอบได้ดังนี้

ค่าอำนาจจำแนกของข้อสอบ (r)	ความหมาย
0.60 – 1.00	อำนาจจำแนกดีมาก
0.40 – 0.59	อำนาจจำแนกดี
0.20 – 0.39	อำนาจจำแนกพอใช้
0.10 – 0.19	อำนาจจำแนกต่ำ (ควรปรับปรุงหรือตัดทิ้ง)
-1.00 – 0.09	อำนาจจำแนกต่ำมาก (ควรปรับปรุงหรือตัดทิ้ง)

การคำนวณหาค่าอำนาจจำแนก สามารถทำได้หลายวิธี ดังนี้

3.1 วิธีการตรวจให้คะแนน

วิธีการตรวจให้คะแนน เป็นวิธีการที่นำแบบทดสอบไปทดสอบกับกลุ่มผู้เรียนที่เป็นกลุ่มตัวอย่าง เมื่อทดสอบแล้วให้เรียงคะแนนที่ได้จากน้อยไปหามากหรือจากมากไปหาน้อยก็ได้ ผู้เรียนที่ได้คะแนนสูงถือว่าเป็นกลุ่มเก่ง และผู้เรียนที่ได้คะแนนต่ำถือว่าเป็นกลุ่มอ่อน เมื่อจัดเรียงลำดับคะแนนรวมของผู้เรียนทั้งหมดแล้ว หลังจากนั้นทำการคัดเลือกผู้เรียนที่ได้คะแนนสูงจำนวน 1/3 ของผู้เรียนทั้งหมด และผู้เรียนที่ได้คะแนนต่ำจำนวน 1/3 ของผู้เรียนทั้งหมดมาแทนค่า ในสูตร ดังนี้

$$D = \frac{R_U - R_L}{\frac{N}{2}}$$

เมื่อ D คือ ค่าอำนาจจำแนก

R_U คือ จำนวนผู้เรียนที่ตอบถูกในกลุ่มเก่ง

R_L คือ จำนวนผู้เรียนที่ตอบถูกในกลุ่มอ่อน

N คือ จำนวนผู้เรียนทั้งหมด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2 วิธีการใช้สัดส่วน

เป็นวิธีการที่ใช้หลักการเหมือนกับวิธีตรวจให้คะแนน เมื่อทดสอบผู้เรียนและทำการตรวจให้คะแนนแล้ว นำคะแนนมาจัดเรียง และหลังจากนั้นทำการคัดเลือกผู้เรียนที่ได้คะแนนสูงจำนวน 1/3 ของผู้เรียนทั้งหมด และผู้เรียนที่ได้คะแนนต่ำจำนวน 1/3 ของผู้เรียนทั้งหมด และทำการหาสัดส่วนระหว่างผู้เรียนกลุ่มเก่งและกลุ่มอ่อน โดยใช้สูตรต่อไปนี้

$$D = P_H - P_L$$

เมื่อ P_H คือ สัดส่วนของคะแนนของผู้เรียนกลุ่มเก่ง

P_L คือ สัดส่วนของคะแนนของผู้เรียนกลุ่มอ่อน

3.3 วิธีการใช้ค่าสหสัมพันธ์แบบพอยน์-ไบซีเรียล

สหสัมพันธ์แบบพอยน์-ไบซีเรียล(point biserial correlation) เป็นวิธีการที่จะต้องแปลงคะแนนของข้อคำถามที่ผู้เรียนทำคะแนนได้เป็นค่า 0 และ1 โดยให้ผู้เรียนทำถูกจะได้ 1 และถ้าทำผิดจะได้ 0 คะแนนที่ผู้เรียนทำได้จากข้อคำถาม ในการคำนวณหาค่าอำนาจจำแนกโดยวิธีนี้ จะต้องดำเนินการไปทีละข้อคำถาม โดยใช้สูตร ดังนี้

$$r_p = \frac{\bar{X}_p - \bar{X}_f}{S_t} \cdot \sqrt{pq}$$

เมื่อ r_p คือ ค่าอำนาจจำแนกแบบพอยน์-ไบซีเรียล

$\bar{X}_p$ คือ คะแนนเฉลี่ยของกลุ่มที่ทำข้อสอบข้อนั้นได้

$\bar{X}_f$ คือ คะแนนเฉลี่ยของกลุ่มที่ทำข้อสอบข้อนั้นไม่ได้

p คือ สัดส่วนผู้เรียนที่ทำข้อสอบข้อนั้นได้

q คือ สัดส่วนผู้เรียนที่ทำข้อสอบข้อนั้นไม่ได้

S_t คือ ส่วนเบี่ยงเบนมาตรฐานของแบบทดสอบทั้งฉบับ

4. ความยากง่าย (Difficulty)

ความยากง่าย หมายถึง ความยากหรือความง่ายของข้อสอบ โดยทั่วไปข้อสอบแต่ละข้อ ควรจะมีความยากหรือความง่ายพอเหมาะ คือมีสัดส่วนความยาก 50% และสัดส่วนความง่าย50% แต่การที่จะจัดทำข้อสอบให้มีความยากง่ายในอัตราส่วน 50/50 นั้นถือเป็นเรื่องที่ยาก เพราะข้อสอบนั้นต้องนำไปทดสอบหลายๆ ครั้ง และปรับปรุงจนได้ค่าความยากง่ายใกล้เคียงกับ 50%

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยทั่วไปแบบทดสอบที่จะนำมาหาความยากง่ายนั้น จะเป็นแบบทดสอบที่ใช้วัดผลสัมฤทธิ์ทางการเรียนหรือแบบทดสอบความถนัดที่มุ่งวัดสติปัญญาผู้เรียน ความยากง่ายของข้อสอบมีค่าไม่เกิน 1 แต่ค่าที่ยอมรับได้จะอยู่ระหว่าง 0.2 ถึง 0.8 ถ้าข้อสอบมีค่าเกิน 0.80 แสดงว่า ข้อสอบนั้นมีความง่ายมากเกินไปต้องตัดออกหรือปรับปรุงใหม่ แต่ถ้าข้อสอบมีค่าต่ำกว่า 0.20 ถือว่า ข้อสอบนั้นมีความยากเกินไปต้องตัดออกหรือปรับปรุงเช่นเดียวกัน ดังนั้น สามารถแปลความหมายค่าความยากง่ายของข้อสอบได้ดังนี้

ความยากง่ายของข้อสอบ (P)	ความหมาย
0.81 – 1.00	ง่ายมาก (ควรปรับปรุงหรือตัดทิ้ง)
0.60 – 0.80	ค่อนข้างง่าย (ดี)
0.40 – 0.59	ยากพอเหมาะ (ดีมาก)
0.20 – 0.39	ค่อนข้างยาก (ดี)
0.00 – 0.19	ยากมาก (ควรปรับปรุงหรือตัดทิ้ง)

สูตรในการคำนวณหาความยากง่ายมีดังนี้

$$P = \frac{R}{N}$$

เมื่อ P คือ ค่าความยากง่าย

R คือ จำนวนผู้เรียนที่ทำข้อสอบข้อนั้นถูก

N คือ จำนวนผู้เรียนทั้งหมด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3

การดำเนินงานวิจัย

ในการดำเนินการวิจัยครั้งนี้ เพื่อให้เป็นไปตามวัตถุประสงค์ที่ได้กำหนดไว้ ผู้วิจัยจึงได้ดำเนินการตามขั้นตอนในการวิจัยดังต่อไปนี้

3.1 ข้อมูลที่ใช้ในงานวิจัย

ข้อมูลที่ใช้ในงานวิจัยครั้งนี้ ประกอบด้วย เพศ เกรดเฉลี่ยสะสม คะแนนสอบกลางภาค และคะแนนสอบปลายภาค โดยรวบรวมจากผู้ที่เรียนในวิชาสถิติเบื้องต้น ซึ่งมีรายละเอียดดังนี้

- นักศึกษาสาขาวิชาสถิติประยุกต์ปีการศึกษา 2555 มีจำนวน 134 คน
- นักศึกษาสาขาวิชาสถิติประยุกต์ปีการศึกษา 2556 มีจำนวน 136 คน
- นักศึกษาสาขาวิชาชีวสถิติปีการศึกษา 2557 มีจำนวน 120 คน
- นักศึกษาสาขาวิชาสัตวศาสตร์ปีการศึกษา 2555 มีจำนวน 100 คน

3.2 สถิติที่ใช้ในการวิเคราะห์

สถิติที่ใช้ในการทำวิจัยครั้งนี้แบ่งออกเป็น 2 ส่วน คือ

ส่วนที่ 1 เป็นการวิเคราะห์คุณภาพของข้อสอบ ประกอบด้วย

1. ค่าความยากของข้อสอบ
2. ค่าอำนาจจำแนกข้อสอบ มี 2 ค่า คือ ค่าอำนาจจำแนกรายข้อแบบสัดส่วน และค่าอำนาจจำแนกรายข้อแบบพอยน์ไบซีเรียล จะทำการทดสอบค่าทั้งสองว่ามีความแตกต่างกันหรือไม่ โดยใช้ตัวสถิติ t สำหรับค่าเฉลี่ยของประชากร 2 กลุ่มที่ไม่เป็นอิสระกัน
3. ค่าความเชื่อมั่นของข้อสอบ โดยวิธีแบบคูเดอร์-ริชาร์ดสัน

ส่วนที่ 2 เป็นการวิเคราะห์การถดถอยกับตัวแปรดัมมี่ แบ่งออกเป็น

- การถดถอยกับตัวแปรดัมมี่สำหรับข้อมูลสองกลุ่ม จากรูปแบบเต็ม

$$Y = \beta_0 + \beta_1 X + \beta_2 D + \beta_3 DX + \varepsilon$$

- การถดถอยกับตัวแปรดัมมี่สำหรับข้อมูลสามกลุ่ม จากรูปแบบเต็ม

$$Y = \beta_0 + \beta_1 X + \beta_2 D_1 + \beta_3 D_2 + \beta_4 D_1 X + \beta_5 D_2 X + \varepsilon$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.3 ขั้นตอนในการดำเนินงาน

1. ทำการเก็บรวบรวมข้อมูล เพศ เกรดเฉลี่ยสะสม คะแนนสอบกลางภาค และคะแนนสอบปลายภาค ของนักศึกษาแต่ละคนที่เรียนในวิชาสถิติเบื้องต้นสาขาวิชาสถิติประยุกต์ปีการศึกษา 2555 – 2556 สาขาวิชาจุลชีววิทยา ปีการศึกษา 2557 และสาขาวิชาสัตวศาสตร์ ปีการศึกษา 2555
2. นำข้อสอบกลางภาคและปลายภาคในวิชาสถิติเบื้องต้นมาทำการหาคุณภาพข้อสอบ ด้วยค่าความยากของข้อสอบ ค่าอำนาจจำแนกข้อสอบ และค่าความเชื่อมั่นของข้อสอบ โดยค่าค่าอำนาจจำแนกข้อสอบ มี 2 ค่า คือ ค่าอำนาจจำแนกรายข้อแบบสัดส่วน และค่าอำนาจจำแนกรายข้อแบบพอยน์ไบซีเรียล จึงได้ทำการทดสอบค่าทั้งสองว่ามีความแตกต่างกันหรือไม่ โดยใช้ตัวสถิติ t สำหรับค่าเฉลี่ยของประชากร 2 กลุ่มที่ไม่เป็นอิสระกัน
3. นำตัวแปร เพศ สาขาวิชา เกรดเฉลี่ยสะสม คะแนนสอบกลางภาค และคะแนนสอบปลายภาค ของนักศึกษาแต่ละคนที่รวบรวมมาได้ในวิชาสถิติเบื้องต้น มาทำการวิเคราะห์การถดถอยกับตัวแปรดัมมี่สำหรับข้อมูลสองกลุ่มและสามกลุ่มตามลำดับ
4. นำผลการวิเคราะห์ที่ได้มาทำการสรุปผล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4

ผลการวิเคราะห์ข้อมูล

ในการวิจัยครั้งนี้ ได้ทำการเก็บข้อมูลที่ใช้ในการวิเคราะห์ โดยสามารถแบ่งออกเป็น 2 ส่วน ดังนี้

ส่วนที่ 1 เป็นข้อสอบกลางภาคและปลายภาควิชาสถิติเบื้องต้น

ส่วนที่ 2 เป็นคะแนนสอบกลางภาคและปลายภาควิชาสถิติเบื้องต้น

ซึ่งทั้ง 2 ส่วนนี้จะเก็บจากนักศึกษาที่เรียนในวิชาสถิติเบื้องต้น ประกอบด้วย

- นักศึกษาสาขาวิชาสถิติประยุกต์ปีการศึกษา 2555 มีจำนวน 134 คน
- นักศึกษาสาขาวิชาสถิติประยุกต์ปีการศึกษา 2556 มีจำนวน 136 คน
- นักศึกษาสาขาวิชาจุลชีววิทยาปีการศึกษา 2557 มีจำนวน 120 คน
- นักศึกษาสาขาวิชาสัตวศาสตร์ปีการศึกษา 2555 มีจำนวน 100 คน

โดยนำข้อมูลมาทำการวิเคราะห์ อาศัยโปรแกรมสำเร็จรูป SPSS และโปรแกรม Microsoft Excel มาช่วยในการประมวลผล และจะนำเสนอผลการวิเคราะห์ออกเป็น 3 ส่วน คือ

1. ผลการวิเคราะห์คุณภาพข้อสอบ
2. ผลการวิเคราะห์การถดถอยกับตัวแปรดัมมี่สำหรับข้อมูลสองกลุ่ม
3. ผลการวิเคราะห์การถดถอยกับตัวแปรดัมมี่สำหรับข้อมูลสามกลุ่ม

4.1 ผลการวิเคราะห์คุณภาพข้อสอบ

ข้อสอบที่ใช้ในการวิเคราะห์ครั้งนี้เป็นข้อสอบกลางภาคและปลายภาควิชาสถิติเบื้องต้น ในสาขาวิชาสถิติประยุกต์ ปีการศึกษา 2555 – 2556 สาขาวิชาจุลชีววิทยา ปีการศึกษา 2557 และสาขาวิชาสัตวศาสตร์ ปีการศึกษา 2555 ซึ่งในการวิเคราะห์คุณภาพข้อสอบนั้น จะพิจารณาจากค่าความยากของข้อสอบ ค่าอำนาจจำแนก และค่าความเชื่อมั่นของข้อสอบทั้งฉบับ โดยค่าอำนาจจำแนกมีอยู่ 2 ค่า คือ

- ค่าอำนาจจำแนกรายข้อแบบสัดส่วน
- ค่าอำนาจจำแนกรายข้อแบบพอยน์ไบซีเรียล

จึงได้ทำการเปรียบเทียบความแตกต่างของค่าอำนาจจำแนกรายข้อแบบสัดส่วนและแบบพอยไบซีเรียล โดยใช้ตัวสถิติ t สำหรับค่าเฉลี่ยของประชากร 2 กลุ่มที่ไม่เป็นอิสระกันได้ผลการวิเคราะห์ดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.1 การเปรียบเทียบความแตกต่างของค่าอำนาจจำแนกรายข้อแบบสัดส่วนและแบบพอยน์ไบซีเรียลโดยใช้ตัวสถิติ t สำหรับค่าเฉลี่ยของประชากร 2 กลุ่มที่ไม่เป็นอิสระกัน

สาขาวิชา	ประเภทข้อสอบ	Paired Differences			t	df	Sig.(2-tailed)
		Mean	Std. Deviation	Std. Error Mean			
สถิติประยุกต์ ปีการศึกษา 2555	กลางภาค	0.0110	0.0547	0.0109	1.006	24	0.325
	ปลายภาค	0.0034	0.0567	0.0113	0.300	24	0.767
สถิติประยุกต์ ปีการศึกษา 2556	กลางภาค	0.0047	0.0544	0.0109	0.430	24	0.671
	ปลายภาค	0.0251	0.0676	0.0135	1.854	24	0.076
จุลชีววิทยา ปีการศึกษา 2557	กลางภาค	0.0201	0.0489	0.0109	1.836	19	0.082
	ปลายภาค	0.0265	0.0634	0.0142	1.865	19	0.078
สัตวศาสตร์ ปีการศึกษา 2555	กลางภาค	0.0052	0.0536	0.0107	0.489	24	0.629
	ปลายภาค	0.0017	0.0732	0.0164	0.104	19	0.918

ในการทดสอบสมมติฐานเกี่ยวกับความแตกต่างระหว่างค่าอำนาจจำแนกรายข้อแบบสัดส่วนและแบบพอยน์ไบซีเรียล สามารถตั้งสมมติฐานได้ดังนี้

H_0 : $\mu_d = 0$ (ค่าอำนาจจำแนกรายข้อแบบสัดส่วนและแบบพอยน์ไบซีเรียลไม่แตกต่างกัน)

H_1 : $\mu_d \neq 0$ (ค่าอำนาจจำแนกรายข้อแบบสัดส่วนและแบบพอยน์ไบซีเรียลแตกต่างกัน)

จากตารางที่ 4.1 เมื่อกำหนดระดับนัยสำคัญที่ 0.05 ($\alpha = 0.05$) จะเห็นว่า ค่า Sig.(2-tailed) > 0.05 ทุกค่า จึงทำการยอมรับ H_0 แสดงว่า ค่าอำนาจจำแนกรายข้อแบบสัดส่วนและแบบพอยน์ไบซีเรียลไม่แตกต่างกัน ดังนั้น ในงานวิจัยครั้งนี้ จะนำเสนอผลการวิเคราะห์ค่าอำนาจจำแนกแบบสัดส่วนเท่านั้น

4.1.1 คุณภาพข้อสอบวิชาสถิติเบื้องต้นสาขาวิชาสถิติประยุกต์ ปีการศึกษา 2555

4.1.1.1 ข้อสอบกลางภาควิชาสถิติเบื้องต้นสาขาวิชาสถิติประยุกต์ ปีการศึกษา 2555

จากการวิเคราะห์พบว่าข้อสอบกลางภาควิชาสถิติเบื้องต้นสาขาวิชาสถิติประยุกต์ ปีการศึกษา 2555 มีค่าความยากเฉลี่ยทั้งฉบับ $(\overline{P})$ เท่ากับ 0.450 และค่าความเชื่อมั่นทั้งฉบับ(KR-20) มีค่าเท่ากับ 0.517 ส่วนค่าความยากรายข้อ (P) และค่าอำนาจจำแนกรายข้อ (r) แสดงในตารางต่อไปนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.2 คุณภาพข้อสอบกลางภาคแบบรายข้อวิชาสถิติเบื้องต้นสาขาวิชาสถิติประยุกต์ ปีการศึกษา 2555

ข้อที่	ค่าความยาก (P)	ค่าอำนาจจำแนก (r)	ความหมาย	
1	0.648	0.345	ข้อสอบค่อนข้างง่าย	อำนาจจำแนกพอใช้
2	0.545	0.509	ข้อสอบยากพอเหมาะ	อำนาจจำแนกดี
3	0.291	0.255	ข้อสอบค่อนข้างยาก	อำนาจจำแนกพอใช้
4	0.236	0.000	ข้อสอบค่อนข้างยาก	อำนาจจำแนกต่ำมาก
5	0.564	0.527	ข้อสอบยากพอเหมาะ	อำนาจจำแนกดี
6	0.333	0.182	ข้อสอบค่อนข้างยาก	อำนาจจำแนกต่ำ
7	0.691	0.255	ข้อสอบค่อนข้างง่าย	อำนาจจำแนกพอใช้
8	0.200	0.200	ข้อสอบค่อนข้างยาก	อำนาจจำแนกพอใช้
9	0.376	0.455	ข้อสอบค่อนข้างยาก	อำนาจจำแนกดี
10	0.364	0.418	ข้อสอบค่อนข้างยาก	อำนาจจำแนกดี
11	0.600	0.418	ข้อสอบค่อนข้างง่าย	อำนาจจำแนกดี
12	0.521	0.473	ข้อสอบยากพอเหมาะ	อำนาจจำแนกดี
13	0.406	0.418	ข้อสอบยากพอเหมาะ	อำนาจจำแนกดี
14	0.552	0.073	ข้อสอบยากพอเหมาะ	อำนาจจำแนกต่ำมาก
15	0.212	0.182	ข้อสอบค่อนข้างยาก	อำนาจจำแนกต่ำ
16	0.667	0.218	ข้อสอบค่อนข้างง่าย	อำนาจจำแนกพอใช้
17	0.448	0.327	ข้อสอบยากพอเหมาะ	อำนาจจำแนกพอใช้
18	0.291	0.127	ข้อสอบค่อนข้างยาก	อำนาจจำแนกต่ำ
19	0.800	0.182	ข้อสอบค่อนข้างง่าย	อำนาจจำแนกต่ำ
20	0.673	0.291	ข้อสอบค่อนข้างง่าย	อำนาจจำแนกพอใช้
21	0.406	0.273	ข้อสอบยากพอเหมาะ	อำนาจจำแนกพอใช้
22	0.327	0.182	ข้อสอบค่อนข้างยาก	อำนาจจำแนกต่ำ
23	0.279	0.345	ข้อสอบค่อนข้างยาก	อำนาจจำแนกพอใช้
24	0.224	0.091	ข้อสอบค่อนข้างยาก	อำนาจจำแนกต่ำมาก
25	0.606	0.455	ข้อสอบค่อนข้างง่าย	อำนาจจำแนกดี

จากตารางที่ 4.2 พบว่า ข้อสอบกลางภาควิชาสถิติเบื้องต้นสาขาวิชาสถิติประยุกต์ ปีการศึกษา 2555 มีทั้งหมด 25 ข้อ และมีจำนวนข้อที่ถือว่าเป็นข้อสอบที่ใช้ได้อยู่ 17 ข้อ คิดเป็นร้อยละ 68

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.1.2 ข้อสอบปลายภาควิชาสถิติเบื้องต้นสาขาวิชาสถิติประยุกต์ ปีการศึกษา 2555

จากการวิเคราะห์พบว่า ข้อสอบปลายภาควิชาสถิติเบื้องต้นสาขาวิชาสถิติประยุกต์ ปีการศึกษา 2555 มีค่าความยากเฉลี่ยทั้งฉบับ $(\overline{P})$ เท่ากับ 0.494 และค่าความเชื่อมั่นทั้งฉบับ(KR-20) มีค่าเท่ากับ 0.691 ส่วนค่าความยากรายข้อ (P) และค่าอำนาจจำแนกรายข้อ (r) แสดงในตารางต่อไปนี้

ตารางที่ 4.3 คุณภาพข้อสอบปลายภาคแบบรายข้อ วิชาสถิติเบื้องต้นสาขาวิชาสถิติประยุกต์ ปีการศึกษา 2555

ข้อที่	ค่าความยาก (P)	ค่าอำนาจจำแนก (r)	ความหมาย	
1	0.648	0.418	ข้อสอบค่อนข้างง่าย	อำนาจจำแนกดี
2	0.648	0.455	ข้อสอบค่อนข้างง่าย	อำนาจจำแนกดี
3	0.758	0.200	ข้อสอบค่อนข้างง่าย	อำนาจจำแนกพอใช้
4	0.242	0.091	ข้อสอบค่อนข้างยาก	อำนาจจำแนกต่ำมาก
5	0.176	0.164	ข้อสอบยากมาก	อำนาจจำแนกต่ำ
6	0.448	0.509	ข้อสอบยากพอเหมาะ	อำนาจจำแนกดี
7	0.545	0.400	ข้อสอบยากพอเหมาะ	อำนาจจำแนกดี
8	0.612	0.473	ข้อสอบค่อนข้างง่าย	อำนาจจำแนกดี
9	0.394	0.509	ข้อสอบค่อนข้างยาก	อำนาจจำแนกดี
10	0.364	0.327	ข้อสอบค่อนข้างยาก	อำนาจจำแนกพอใช้
11	0.503	0.345	ข้อสอบยากพอเหมาะ	อำนาจจำแนกพอใช้
12	0.255	0.164	ข้อสอบค่อนข้างยาก	อำนาจจำแนกต่ำ
13	0.370	0.127	ข้อสอบค่อนข้างยาก	อำนาจจำแนกต่ำ
14	0.479	0.382	ข้อสอบยากพอเหมาะ	อำนาจจำแนกพอใช้
15	0.612	0.345	ข้อสอบค่อนข้างง่าย	อำนาจจำแนกพอใช้
16	0.842	0.255	ข้อสอบง่ายมาก	อำนาจจำแนกพอใช้
17	0.673	0.509	ข้อสอบค่อนข้างง่าย	อำนาจจำแนกดี
18	0.261	0.291	ข้อสอบค่อนข้างยาก	อำนาจจำแนกพอใช้
19	0.855	0.309	ข้อสอบง่ายมาก	อำนาจจำแนกพอใช้
20	0.255	0.291	ข้อสอบค่อนข้างยาก	อำนาจจำแนกพอใช้
21	0.412	0.473	ข้อสอบยากพอเหมาะ	อำนาจจำแนกดี
22	0.558	0.418	ข้อสอบยากพอเหมาะ	อำนาจจำแนกดี
23	0.442	0.491	ข้อสอบยากพอเหมาะ	อำนาจจำแนกดี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.3 (ต่อ)

ข้อที่	ค่าความยาก (P)	ค่าอำนาจจำแนก (r)	ความหมาย	
24	0.782	0.200	ข้อสอบค่อนข้างง่าย	อำนาจจำแนกพอใช้
25	0.212	0.273	ข้อสอบค่อนข้างยาก	อำนาจจำแนกพอใช้

จากตารางที่ 4.3 พบว่า ข้อสอบปลายภาควิชาสถิติเบื้องต้นสาขาวิชาสถิติประยุกต์ ปีการศึกษา 2555 มีทั้งหมด 25 ข้อ และมีจำนวนข้อที่ถือว่าเป็นข้อสอบที่ใช้ได้อยู่ 19 ข้อ คิดเป็นร้อยละ 76

4.1.2 คุณภาพข้อสอบวิชาสถิติเบื้องต้นสาขาวิชาสถิติประยุกต์ ปีการศึกษา 2556

4.1.2.1 ข้อสอบกลางภาควิชาสถิติเบื้องต้นสาขาวิชาสถิติประยุกต์ ปีการศึกษา 2556

จากการวิเคราะห์พบว่า ข้อสอบกลางภาควิชาสถิติเบื้องต้นสาขาวิชาสถิติประยุกต์ ปีการศึกษา 2556 มีค่าความยากเฉลี่ยทั้งฉบับ $(\overline{P})$ เท่ากับ 0.533 และค่าความเชื่อมั่นทั้งฉบับ(KR-20) มีค่าเท่ากับ 0.688 ส่วนค่าความยากรายข้อ (P) และค่าอำนาจจำแนกรายข้อ (r) แสดงในตารางต่อไปนี้

ตารางที่ 4.4 คุณภาพข้อสอบกลางภาคแบบรายข้อ วิชาสถิติเบื้องต้นสาขาวิชาสถิติประยุกต์ ปีการศึกษา 2556

ข้อที่	ค่าความยาก (P)	ค่าอำนาจจำแนก (r)	ความหมาย	
1	0.852	0.327	ข้อสอบง่ายมาก	อำนาจจำแนกพอใช้
2	0.419	0.385	ข้อสอบยากพอเหมาะ	อำนาจจำแนกพอใช้
3	0.368	0.231	ข้อสอบค่อนข้างยาก	อำนาจจำแนกพอใช้
4	0.710	0.635	ข้อสอบค่อนข้างง่าย	อำนาจจำแนกดีมาก
5	0.168	0.038	ข้อสอบยากมาก	อำนาจจำแนกต่ำมาก
6	0.452	0.385	ข้อสอบยากพอเหมาะ	อำนาจจำแนกพอใช้
7	0.619	0.269	ข้อสอบค่อนข้างง่าย	อำนาจจำแนกพอใช้
8	0.845	0.250	ข้อสอบง่ายมาก	อำนาจจำแนกพอใช้
9	0.710	0.288	ข้อสอบค่อนข้างง่าย	อำนาจจำแนกพอใช้
10	0.439	0.558	ข้อสอบยากพอเหมาะ	อำนาจจำแนกดี
11	0.677	0.519	ข้อสอบค่อนข้างง่าย	อำนาจจำแนกดี
12	0.555	0.442	ข้อสอบยากพอเหมาะ	อำนาจจำแนกดี
13	0.342	0.269	ข้อสอบค่อนข้างยาก	อำนาจจำแนกพอใช้
14	0.574	0.135	ข้อสอบยากพอเหมาะ	อำนาจจำแนกต่ำ
15	0.245	0.288	ข้อสอบค่อนข้างยาก	อำนาจจำแนกพอใช้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ในการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของผลงานทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.4 (ต่อ)

ข้อที่	ค่าความยาก (P)	ค่าอำนาจจำแนก (r)	ความหมาย	
16	0.665	0.346	ข้อสอบค่อนข้างง่าย	อำนาจจำแนกพอใช้
17	0.465	0.308	ข้อสอบยากพอเหมาะ	อำนาจจำแนกพอใช้
18	0.606	0.519	ข้อสอบค่อนข้างง่าย	อำนาจจำแนกดี
19	0.819	0.192	ข้อสอบง่ายมาก	อำนาจจำแนกต่ำ
20	0.703	0.365	ข้อสอบค่อนข้างง่าย	อำนาจจำแนกพอใช้
21	0.355	0.500	ข้อสอบค่อนข้างยาก	อำนาจจำแนกดี
22	0.323	0.308	ข้อสอบค่อนข้างยาก	อำนาจจำแนกพอใช้
23	0.439	0.404	ข้อสอบยากพอเหมาะ	อำนาจจำแนกดี
24	0.271	0.288	ข้อสอบค่อนข้างยาก	อำนาจจำแนกพอใช้
25	0.703	0.365	ข้อสอบค่อนข้างง่าย	อำนาจจำแนกพอใช้

จากตารางที่ 4.4 พบว่า ข้อสอบกลางภาคของวิชาสถิติเบื้องต้นสาขาวิชาสถิติประยุกต์ ปีการศึกษา 2556 มีทั้งหมด 25 ข้อ และมีจำนวนข้อที่ถือว่าเป็นข้อสอบที่ใช้ได้อยู่ 20 ข้อ คิดเป็นร้อยละ 80

4.1.2.2 ข้อสอบปลายภาควิชาสถิติเบื้องต้นสาขาวิชาสถิติประยุกต์ ปีการศึกษา 2556

จากการวิเคราะห์พบว่า ข้อสอบปลายภาควิชาสถิติเบื้องต้นสาขาวิชาสถิติประยุกต์ ปีการศึกษา 2556 มีค่าความยากเฉลี่ยทั้งฉบับ $(\overline{P})$ เท่ากับ 0.529 และค่าความเชื่อมั่นทั้งฉบับ (KR-20) มีค่าเท่ากับ 0.777 ส่วนค่าความยากรายข้อ (P) และค่าอำนาจจำแนกรายข้อ (r) แสดงในตารางต่อไปนี้

ตารางที่ 4.5 คุณภาพข้อสอบปลายภาคแบบรายข้อ วิชาสถิติเบื้องต้นสาขาวิชาสถิติประยุกต์ ปีการศึกษา 2556

ข้อที่	ค่าความยาก (P)	ค่าอำนาจจำแนก (r)	ความหมาย	
1	0.916	0.154	ข้อสอบง่ายมาก	อำนาจจำแนกต่ำ
2	0.806	0.115	ข้อสอบค่อนข้างง่าย	อำนาจจำแนกต่ำ
3	0.452	0.558	ข้อสอบยากพอเหมาะ	อำนาจจำแนกดี
4	0.510	0.750	ข้อสอบยากพอเหมาะ	อำนาจจำแนกดีมาก
5	0.174	0.327	ข้อสอบยากมาก	อำนาจจำแนกพอใช้
6	0.723	0.385	ข้อสอบค่อนข้างง่าย	อำนาจจำแนกพอใช้
7	0.271	0.115	ข้อสอบค่อนข้างยาก	อำนาจจำแนกต่ำ
8	0.710	0.481	ข้อสอบค่อนข้างง่าย	อำนาจจำแนกดี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่อนุญาตให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของผลงานทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.5 (ต่อ)

ข้อที่	ค่าความยาก (P)	ค่าอำนาจจำแนก (r)	ความหมาย	
9	0.465	0.385	ข้อสอบยากพอเหมาะ	อำนาจจำแนกพอใช้
10	0.542	0.538	ข้อสอบยากพอเหมาะ	อำนาจจำแนกดี
11	0.568	0.385	ข้อสอบยากพอเหมาะ	อำนาจจำแนกพอใช้
12	0.606	0.519	ข้อสอบค่อนข้างง่าย	อำนาจจำแนกดี
13	0.561	0.731	ข้อสอบยากพอเหมาะ	อำนาจจำแนกดีมาก
14	0.626	0.692	ข้อสอบค่อนข้างง่าย	อำนาจจำแนกดีมาก
15	0.232	0.038	ข้อสอบค่อนข้างยาก	อำนาจจำแนกต่ำมาก
16	0.452	0.596	ข้อสอบยากพอเหมาะ	อำนาจจำแนกดี
17	0.716	0.481	ข้อสอบค่อนข้างง่าย	อำนาจจำแนกดี
18	0.561	0.673	ข้อสอบยากพอเหมาะ	อำนาจจำแนกดีมาก
19	0.110	0.077	ข้อสอบยากมาก	อำนาจจำแนกต่ำมาก
20	0.587	0.538	ข้อสอบยากพอเหมาะ	อำนาจจำแนกดี
21	0.535	0.673	ข้อสอบยากพอเหมาะ	อำนาจจำแนกดีมาก
22	0.806	0.365	ข้อสอบค่อนข้างง่าย	อำนาจจำแนกพอใช้
23	0.310	0.192	ข้อสอบค่อนข้างยาก	อำนาจจำแนกต่ำ
24	0.310	0.308	ข้อสอบค่อนข้างยาก	อำนาจจำแนกพอใช้
25	0.665	0.327	ข้อสอบค่อนข้างง่าย	อำนาจจำแนกพอใช้

จากตารางที่ 4.5 พบว่า ข้อสอบปลายภาคของวิชาสถิติเบื้องต้นสาขาวิชาสถิติประยุกต์ ปีการศึกษา 2556 มีทั้งหมด 25 ข้อ และมีจำนวนข้อที่ถือว่าเป็นข้อสอบที่ใช้ได้อยู่ 18 ข้อ คิดเป็นร้อยละ 72

4.1.3 คุณภาพข้อสอบวิชาสถิติเบื้องต้นสาขาวิชาจุลชีววิทยา ปีการศึกษา 2557

4.1.3.1 ข้อสอบกลางภาควิชาสถิติเบื้องต้นสาขาวิชาจุลชีววิทยา ปีการศึกษา 2557

จากการวิเคราะห์พบว่า ข้อสอบกลางภาควิชาสถิติเบื้องต้นสาขาวิชาจุลชีววิทยา ปีการศึกษา 2557 มีค่าความยากเฉลี่ยทั้งฉบับ $(\overline{P})$ เท่ากับ 0.503 และค่าความเชื่อมั่นทั้งฉบับ(KR-20) มีค่าเท่ากับ 0.679 ส่วนค่าความยากรายข้อ (P) และค่าอำนาจจำแนกรายข้อ (r) แสดงในตารางต่อไปนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.6 คุณภาพข้อสอบกลางภาคแบบรายข้อ วิชาสถิติเบื้องต้นสาขาวิชาจุลชีววิทยา ปีการศึกษา 2557

ข้อที่	ค่าความยาก (P)	ค่าอำนาจจำแนก (r)	ความหมาย	
1	0.701	0.310	ข้อสอบค่อนข้างง่าย	อำนาจจำแนกพอใช้
2	0.535	0.500	ข้อสอบยากพอเหมาะ	อำนาจจำแนกดี
3	0.362	0.238	ข้อสอบค่อนข้างยาก	อำนาจจำแนกพอใช้
4	0.701	0.429	ข้อสอบค่อนข้างง่าย	อำนาจจำแนกดี
5	0.346	0.524	ข้อสอบค่อนข้างยาก	อำนาจจำแนกดี
6	0.630	0.500	ข้อสอบค่อนข้างง่าย	อำนาจจำแนกดี
7	0.425	0.357	ข้อสอบยากพอเหมาะ	อำนาจจำแนกพอใช้
8	0.606	0.524	ข้อสอบค่อนข้างง่าย	อำนาจจำแนกดี
9	0.787	0.310	ข้อสอบค่อนข้างง่าย	อำนาจจำแนกพอใช้
10	0.472	0.429	ข้อสอบยากพอเหมาะ	อำนาจจำแนกดี
11	0.315	0.357	ข้อสอบค่อนข้างยาก	อำนาจจำแนกพอใช้
12	0.425	0.524	ข้อสอบยากพอเหมาะ	อำนาจจำแนกดี
13	0.394	0.238	ข้อสอบค่อนข้างยาก	อำนาจจำแนกพอใช้
14	0.354	0.262	ข้อสอบค่อนข้างยาก	อำนาจจำแนกพอใช้
15	0.252	0.286	ข้อสอบค่อนข้างยาก	อำนาจจำแนกพอใช้
16	0.827	0.310	ข้อสอบง่ายมาก	อำนาจจำแนกพอใช้
17	0.417	0.452	ข้อสอบยากพอเหมาะ	อำนาจจำแนกดี
18	0.449	0.405	ข้อสอบยากพอเหมาะ	อำนาจจำแนกดี
19	0.677	0.571	ข้อสอบค่อนข้างง่าย	อำนาจจำแนกดี
20	0.394	0.333	ข้อสอบค่อนข้างยาก	อำนาจจำแนกพอใช้

จากตารางที่ 4.6 พบว่า ข้อสอบกลางภาคของวิชาสถิติเบื้องต้นสาขาวิชาจุลชีววิทยา ปีการศึกษา 2557 มีทั้งหมด 20 ข้อ และมีจำนวนข้อที่ถือว่าเป็นข้อสอบที่ใช้ได้อยู่ 19 ข้อ คิดเป็นร้อยละ 95

4.1.3.2 ข้อสอบปลายภาควิชาสถิติเบื้องต้นสาขาวิชาจุลชีววิทยา ปีการศึกษา 2557

จากการวิเคราะห์พบว่า ข้อสอบปลายภาควิชาสถิติเบื้องต้นสาขาวิชาจุลชีววิทยา ปีการศึกษา 2557 มีค่าความยากเฉลี่ยทั้งฉบับ $(\overline{P})$ เท่ากับ 0.594 และค่าความเชื่อมั่นทั้งฉบับ(KR-20) มีค่าเท่ากับ 0.832 ส่วนค่าความยากรายข้อ (P) และค่าอำนาจจำแนกรายข้อ (r) แสดงในตารางต่อไปนี้

เอกสารนี้มีลิขสิทธิ์ ห้ามนำไปใช้ในเชิงพาณิชย์ ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.7 คุณภาพข้อสอบปลายภาคแบบรายข้อ วิชาสถิติเบื้องต้นสาขาวิชาจุลชีววิทยา ปีการศึกษา 2557

ข้อที่	ค่าความยาก (P)	ค่าอำนาจจำแนก (r)	ความหมาย	
1	0.850	0.200	ข้อสอบง่ายมาก	อำนาจจำแนกพอใช้
2	0.625	0.500	ข้อสอบค่อนข้างง่าย	อำนาจจำแนกดี
3	0.208	0.300	ข้อสอบค่อนข้างยาก	อำนาจจำแนกพอใช้
4	0.767	0.400	ข้อสอบค่อนข้างง่าย	อำนาจจำแนกดี
5	0.650	0.550	ข้อสอบค่อนข้างง่าย	อำนาจจำแนกดี
6	0.742	0.200	ข้อสอบค่อนข้างง่าย	อำนาจจำแนกพอใช้
7	0.700	0.500	ข้อสอบค่อนข้างง่าย	อำนาจจำแนกดี
8	0.517	0.625	ข้อสอบยากพอเหมาะ	อำนาจจำแนกดีมาก
9	0.683	0.375	ข้อสอบค่อนข้างง่าย	อำนาจจำแนกพอใช้
10	0.233	0.100	ข้อสอบค่อนข้างยาก	อำนาจจำแนกต่ำ
11	0.825	0.375	ข้อสอบง่ายมาก	อำนาจจำแนกพอใช้
12	0.708	0.550	ข้อสอบค่อนข้างง่าย	อำนาจจำแนกดี
13	0.550	0.725	ข้อสอบยากพอเหมาะ	อำนาจจำแนกดีมาก
14	0.542	0.525	ข้อสอบยากพอเหมาะ	อำนาจจำแนกดี
15	0.650	0.400	ข้อสอบค่อนข้างง่าย	อำนาจจำแนกดี
16	0.500	0.600	ข้อสอบยากพอเหมาะ	อำนาจจำแนกดีมาก
17	0.417	0.600	ข้อสอบยากพอเหมาะ	อำนาจจำแนกดีมาก
18	0.533	0.600	ข้อสอบยากพอเหมาะ	อำนาจจำแนกดีมาก
19	0.575	0.575	ข้อสอบยากพอเหมาะ	อำนาจจำแนกดี
20	0.750	0.550	ข้อสอบค่อนข้างง่าย	อำนาจจำแนกดี

จากตารางที่ 4.7 พบว่า ข้อสอบปลายภาคของวิชาสถิติเบื้องต้นสาขาวิชาจุลชีววิทยา ปีการศึกษา 2557 มีทั้งหมด 20 ข้อ และมีจำนวนข้อที่ถือว่าเป็นข้อสอบที่ใช้ได้อยู่ 17 ข้อ คิดเป็นร้อยละ 85

4.1.4 คุณภาพข้อสอบวิชาสถิติเบื้องต้นสาขาวิชาสัตวศาสตร์ ปีการศึกษา 2555

4.1.4.1 ข้อสอบกลางภาควิชาสถิติเบื้องต้นสาขาวิชาสัตวศาสตร์ ปีการศึกษา 2555

จากการวิเคราะห์พบว่าข้อสอบกลางภาควิชาสถิติเบื้องต้นสาขาวิชาสัตวศาสตร์ ปีการศึกษา 2555 มีค่าความยากเฉลี่ยทั้งฉบับ $(\bar{P})$ เท่ากับ 0.359 และค่าความเชื่อมั่นทั้งฉบับ(KR-20) มีค่าเท่ากับ 0.260 ส่วนค่าความยากรายข้อ (P) และค่าอำนาจจำแนกรายข้อ (r) แสดงในตารางต่อไปนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.8 คุณภาพข้อสอบกลางภาคแบบรายข้อ วิชาสถิติเบื้องต้นสาขาวิชาสัตวศาสตร์ ปีการศึกษา 2555

ข้อที่	ค่าความยาก (P)	ค่าอำนาจจำแนก (r)	ความหมาย	
1	0.592	0.371	ข้อสอบยากพอเหมาะ	อำนาจจำแนกพอใช้
2	0.427	0.571	ข้อสอบยากพอเหมาะ	อำนาจจำแนกดี
3	0.311	0.229	ข้อสอบค่อนข้างยาก	อำนาจจำแนกพอใช้
4	0.320	0.143	ข้อสอบค่อนข้างยาก	อำนาจจำแนกต่ำ
5	0.408	0.286	ข้อสอบยากพอเหมาะ	อำนาจจำแนกพอใช้
6	0.126	0.114	ข้อสอบยากมาก	อำนาจจำแนกต่ำ
7	0.408	0.114	ข้อสอบยากพอเหมาะ	อำนาจจำแนกต่ำ
8	0.155	0.057	ข้อสอบยากมาก	อำนาจจำแนกต่ำมาก
9	0.262	0.000	ข้อสอบค่อนข้างยาก	อำนาจจำแนกต่ำมาก
10	0.126	0.029	ข้อสอบยากมาก	อำนาจจำแนกต่ำมาก
11	0.583	0.629	ข้อสอบยากพอเหมาะ	อำนาจจำแนกดีมาก
12	0.398	0.400	ข้อสอบค่อนข้างยาก	อำนาจจำแนกดี
13	0.252	0.257	ข้อสอบค่อนข้างยาก	อำนาจจำแนกพอใช้
14	0.515	0.171	ข้อสอบยากพอเหมาะ	อำนาจจำแนกต่ำ
15	0.146	0.200	ข้อสอบยากมาก	อำนาจจำแนกพอใช้
16	0.631	0.229	ข้อสอบค่อนข้างง่าย	อำนาจจำแนกพอใช้
17	0.291	0.114	ข้อสอบค่อนข้างยาก	อำนาจจำแนกต่ำ
18	0.282	0.229	ข้อสอบค่อนข้างยาก	อำนาจจำแนกพอใช้
19	0.602	0.200	ข้อสอบค่อนข้างง่าย	อำนาจจำแนกพอใช้
20	0.660	0.171	ข้อสอบค่อนข้างง่าย	อำนาจจำแนกต่ำ
21	0.282	0.314	ข้อสอบค่อนข้างยาก	อำนาจจำแนกพอใช้
22	0.214	0.200	ข้อสอบค่อนข้างยาก	อำนาจจำแนกพอใช้
23	0.136	0.057	ข้อสอบยากมาก	อำนาจจำแนกต่ำมาก
24	0.204	0.086	ข้อสอบค่อนข้างยาก	อำนาจจำแนกต่ำมาก
25	0.650	0.257	ข้อสอบค่อนข้างง่าย	อำนาจจำแนกพอใช้

จากตารางที่ 4.8 พบว่า ข้อสอบกลางภาคของวิชาสถิติเบื้องต้นสาขาวิชาสัตวศาสตร์ ปีการศึกษา 2555 มีทั้งหมด 25 ข้อ และมีจำนวนข้อที่ถือว่าเป็นข้อสอบที่ใช้ได้อยู่ 13 ข้อ คิดเป็นร้อยละ 52

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.4.2 ข้อสอบปลายภาควิชาสถิติเบื้องต้นสาขาวิชาสัตวศาสตร์ ปีการศึกษา 2555

จากการวิเคราะห์พบว่า ข้อสอบปลายภาควิชาสถิติเบื้องต้นสาขาวิชาสัตวศาสตร์ ปีการศึกษา 2555 มีค่าความยากเฉลี่ยทั้งฉบับ $(\overline{P})$ เท่ากับ 0.403 และค่าความเชื่อมั่นทั้งฉบับ(KR-20) มีค่าเท่ากับ 0.555 ส่วนค่าความยากรายข้อ (P) และค่าอำนาจจำแนกรายข้อ (r) แสดงในตารางต่อไปนี้

ตารางที่ 4.9 คุณภาพข้อสอบปลายภาคแบบรายข้อ วิชาสถิติเบื้องต้นสาขาวิชาสัตวศาสตร์ ปีการศึกษา 2555

ข้อที่	ค่าความยาก (P)	ค่าอำนาจจำแนก (r)	ความหมาย	
1	0.294	0.412	ข้อสอบค่อนข้างยาก	อำนาจจำแนกดี
2	0.480	0.441	ข้อสอบยากพอเหมาะ	อำนาจจำแนกดี
3	0.157	0.029	ข้อสอบยากมาก	อำนาจจำแนกต่ำมาก
4	0.235	0.000	ข้อสอบค่อนข้างยาก	อำนาจจำแนกต่ำมาก
5	0.294	0.029	ข้อสอบค่อนข้างยาก	อำนาจจำแนกต่ำมาก
6	0.343	0.500	ข้อสอบค่อนข้างยาก	อำนาจจำแนกดี
7	0.588	0.294	ข้อสอบยากพอเหมาะ	อำนาจจำแนกพอใช้
8	0.676	0.500	ข้อสอบค่อนข้างง่าย	อำนาจจำแนกดี
9	0.216	0.206	ข้อสอบค่อนข้างยาก	อำนาจจำแนกพอใช้
10	0.206	0.206	ข้อสอบค่อนข้างยาก	อำนาจจำแนกพอใช้
11	0.235	0.235	ข้อสอบค่อนข้างยาก	อำนาจจำแนกพอใช้
12	0.284	0.412	ข้อสอบค่อนข้างยาก	อำนาจจำแนกดี
13	0.500	0.265	ข้อสอบยากพอเหมาะ	อำนาจจำแนกพอใช้
14	0.422	0.471	ข้อสอบยากพอเหมาะ	อำนาจจำแนกดี
15	0.559	0.147	ข้อสอบยากพอเหมาะ	อำนาจจำแนกต่ำ
16	0.745	0.265	ข้อสอบค่อนข้างง่าย	อำนาจจำแนกพอใช้
17	0.363	0.324	ข้อสอบค่อนข้างยาก	อำนาจจำแนกพอใช้
18	0.490	0.735	ข้อสอบยากพอเหมาะ	อำนาจจำแนกดีมาก
19	0.647	0.471	ข้อสอบค่อนข้างง่าย	อำนาจจำแนกดี
20	0.324	0.441	ข้อสอบค่อนข้างยาก	อำนาจจำแนกดี

จากตารางที่ 4.9 พบว่า ข้อสอบปลายภาคของวิชาสถิติเบื้องต้นสาขาวิชาสัตวศาสตร์ ปีการศึกษา 2555 มีทั้งหมด 20 ข้อ และมีจำนวนข้อที่ถือว่าเป็นข้อสอบที่ใช้ได้อยู่ 16 ข้อ คิดเป็นร้อยละ 80

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากผลการวิเคราะห์คุณภาพข้อสอบกลางภาคและปลายภาคในวิชาสถิติเบื้องต้นข้างต้น สามารถนำมาสรุปได้ดังตารางต่อไปนี้

ตารางที่ 4.10 คุณภาพข้อสอบกลางภาคและปลายภาควิชาสถิติเบื้องต้นจำแนกตามสาขาวิชา

สาขาวิชา	ประเภทข้อสอบ	จำนวนข้อ	ค่าความยากเฉลี่ยทั้งฉบับ $(\bar{P})$	ค่าความเชื่อมั่นทั้งฉบับ KR-20	จำนวนข้อสอบที่ใช้ได้ (%)
สถิติประยุกต์ ปีการศึกษา 2555	กลางภาค	25	0.450	0.517	68
	ปลายภาค	25	0.494	0.691	76
สถิติประยุกต์ ปีการศึกษา 2556	กลางภาค	25	0.533	0.688	80
	ปลายภาค	25	0.529	0.777	72
จุลชีววิทยา ปีการศึกษา 2557	กลางภาค	20	0.503	0.679	95
	ปลายภาค	20	0.594	0.832	85
สัตวศาสตร์ ปีการศึกษา 2555	กลางภาค	25	0.359	0.260	52
	ปลายภาค	20	0.403	0.555	80

จากตารางที่ 4.10 พบว่า คุณภาพข้อสอบกลางภาคและปลายภาคในวิชาสถิติเบื้องต้นทั้ง 8 ฉบับ มีอยู่ 7 ฉบับที่มีค่าความยากเฉลี่ยทั้งฉบับ $(\bar{P})$ อยู่ในช่วง 0.40 – 0.50 ค่าความเชื่อมั่นทั้งฉบับ (KR-20) อยู่ในช่วง 0.50 – 0.80 และมีจำนวนข้อสอบที่ใช้ได้มากกว่าร้อยละ 65 แต่มีข้อสอบเพียงฉบับเดียวคือข้อสอบกลางภาคสาขาวิชาสัตวศาสตร์ ที่มีค่าความยากเฉลี่ยทั้งฉบับ $(\bar{P})$ เท่ากับ 0.359 ค่าความเชื่อมั่นทั้งฉบับ (KR-20) เท่ากับ 0.260 และมีจำนวนข้อสอบที่ใช้ได้ร้อยละ 52

จากผลการวิเคราะห์ข้างต้นจึงถือได้ว่า ข้อสอบที่นำมาวิเคราะห์ทั้ง 8 ฉบับนี้มีคุณภาพข้อสอบที่ใกล้เคียงกัน ดังนั้น การนำเอาคะแนนสอบกลางภาคและปลายภาควิชาสถิติเบื้องต้นมาใช้ในการวิเคราะห์ถือเป็นการกำจัดความผันแปรที่อาจเกิดจากคุณภาพของข้อสอบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2 ผลการวิเคราะห์การถดถอยกับตัวแปรดัมมี่สำหรับข้อมูลสองกลุ่ม

ข้อมูลที่ใช้ในวิเคราะห์ประกอบด้วย

- ตัวแปรตามคือ คะแนนสอบปลายภาควิชาสถิติเบื้องต้นของนักศึกษา
- ส่วนตัวแปรอิสระ คือ คะแนนสอบกลางภาควิชาสถิติเบื้องต้น เพศ และเกรดเฉลี่ยสะสมของนักศึกษา

ซึ่งสามารถสร้างรูปแบบการถดถอยที่ใช้ในการวิเคราะห์ได้ดังนี้

รูปแบบที่ 1 $Y = \beta_0 + \beta_1 X_1 + \varepsilon$

รูปแบบที่ 2 $Y = \beta_0 + \beta_1 X_1 + \beta_2 D + \varepsilon$

รูปแบบที่ 3 $Y = \beta_0 + \beta_1 X_1 + \beta_3 D X_1 + \varepsilon$

รูปแบบที่ 4 $Y = \beta_0 + \beta_1 X_1 + \beta_2 D + \beta_3 D X_1 + \varepsilon$

รูปแบบที่ 5 $Y = \beta_0 + \beta_1 X_1 + \beta_4 X_2 + \varepsilon$

เมื่อ Y คือ ตัวแปรตาม ในที่นี้คือ คะแนนสอบปลายภาค

X_1 คือ ตัวแปรอิสระตัวที่ 1 ในที่นี้ คือ คะแนนสอบกลางภาค

X_2 คือ ตัวแปรอิสระตัวที่ 2 ในที่นี้ คือ เกรดเฉลี่ยสะสม

D คือ ตัวแปรดัมมี่ ในที่นี้

$D = 0$ สำหรับนักศึกษาเพศหญิง และ $D = 1$ สำหรับนักศึกษาเพศชาย

$\beta_0, \beta_1, \beta_2, \beta_3, \beta_4$ คือ สัมประสิทธิ์การถดถอย

4.2.1 ผลการวิเคราะห์วิชาสถิติเบื้องต้นของนักศึกษาสาขาวิชาสถิติประยุกต์ ปีการศึกษา 2555

ก). จากรูปแบบ $Y = \beta_0 + \beta_1 X_1 + \varepsilon$ ทำการทดสอบสมมติฐาน

$H_0 : \beta_1 = 0$ $H_1 : \beta_1 \neq 0$

ตารางการวิเคราะห์ความแปรปรวน

Source	df	SS	MS	F	P-value
Regression on X_1	1	11193.163	11193.163	54.033	< 0.001
Error	132	27344.397	207.155		
Total	133	38537.560			

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารนี้ทุกครั้งที่มีการนำไปใช้

เมื่อกำหนด $\alpha = 0.05$ ค่า P-value มีค่า < 0.001 ซึ่งน้อยกว่า α จึงทำการปฏิเสธ H_0 แสดงว่ารูปแบบการถดถอยที่เหมาะสม คือ $Y = \beta_0 + \beta_1 X_1 + \varepsilon$

ข). จากรูปแบบ $Y = \beta_0 + \beta_1 X_1 + \beta_2 D + \varepsilon$ ทำการทดสอบสมมติฐาน

$H_0 : \beta_2 = 0$ $H_1 : \beta_2 \neq 0$

ตารางการวิเคราะห์ความแปรปรวน

Source	df	SS	MS	F	P-value
Regression on X_1 , D	2	12700.357	6350.178	32.197	< 0.001
due to X_1	1	11193.163	-	-	
due to D / X_1	1	1507.194	1507.194	7.642	0.007
Error	131	25837.203	197.231		
Total	133	38537.560			

เมื่อกำหนด $\alpha = 0.05$ ค่า P-value มีค่าเท่ากับ 0.007 ซึ่งน้อยกว่า α จึงทำการปฏิเสธ H_0 แสดงว่ารูปแบบการถดถอยที่เหมาะสมคือ $Y = \beta_0 + \beta_1 X_1 + \beta_2 D + \varepsilon$

ค). จากรูปแบบ $Y = \beta_0 + \beta_1 X_1 + \beta_3 DX_1 + \varepsilon$ ทำการทดสอบสมมติฐาน

$H_0 : \beta_3 = 0$ $H_1 : \beta_3 \neq 0$

ตารางการวิเคราะห์ความแปรปรวน

Source	df	SS	MS	F	P-value
Regression on X_1 , DX_1	2	12053.549	6026.775	29.811	< 0.001
due to X_1	1	11193.163			
due to DX_1 / X_1	1	860.386	860.386	4.256	0.041
Error	131	26484.010	202.168		
Total	133	38537.560			

เมื่อกำหนด $\alpha = 0.05$ ค่า P-value มีค่าเท่ากับ 0.041 ซึ่งน้อยกว่า α จึงทำการปฏิเสธ H_0 แสดงว่ารูปแบบการถดถอยที่เหมาะสมคือ $Y = \beta_0 + \beta_1 X_1 + \beta_3 DX_1 + \varepsilon$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ง). จากรูปแบบ $Y = \beta_0 + \beta_1 X_1 + \beta_2 D + \beta_3 DX_1 + \varepsilon$ ทำการทดสอบสมมติฐาน

$H_0 : \beta_2 = 0$ $H_1 : \beta_2 \neq 0$

ตารางการวิเคราะห์ความแปรปรวน

Source	df	SS	MS	F	P-value
Regression on X_1 , D , DX_1	3	13574.165	4524.722	23.563	< 0.001
due to X_1 , DX_1	2	12053.549			
due to D / X_1 , DX_1	1	1520.616	1520.616	7.919	0.006
Error	130	24963.394	192.026		
Total	133	38537.560			

เมื่อกำหนด $\alpha = 0.05$ ค่า P-value มีค่าเท่ากับ 0.006 ซึ่งน้อยกว่า α จึงทำการปฏิเสธ H_0 แสดงว่ารูปแบบการถดถอยที่เหมาะสมคือ $Y = \beta_0 + \beta_1 X_1 + \beta_2 D + \beta_3 DX_1 + \varepsilon$

จ). จากรูปแบบ $Y = \beta_0 + \beta_1 X_1 + \beta_2 D + \beta_3 DX_1 + \varepsilon$ ทำการทดสอบสมมติฐาน

$H_0 : \beta_3 = 0$ $H_1 : \beta_3 \neq 0$

ตารางการวิเคราะห์ความแปรปรวน

Source	df	SS	MS	F	P-value
Regression on X_1 , D , DX_1	3	13574.165	4524.722	23.563	< 0.001
due to X_1 , D	2	12700.357			
due to DX_1 / X_1 , D	1	873.808	873.808	4.550	0.035
Error	130	24963.394	192.026		
Total	133	38537.560			

เมื่อกำหนด $\alpha = 0.05$ ค่า P-value มีค่าเท่ากับ 0.035 ซึ่งน้อยกว่า α จึงทำการปฏิเสธ H_0 แสดงว่ารูปแบบการถดถอยที่เหมาะสมคือ $Y = \beta_0 + \beta_1 X_1 + \beta_2 D + \beta_3 DX_1 + \varepsilon$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ฉ). จากรูปแบบ $Y = \beta_0 + \beta_1 X_1 + \beta_4 X_2 + \varepsilon$ ทำการทดสอบสมมติฐาน

$H_o : \beta_4 = 0$ $\quad$ $H_1 : \beta_4 \neq 0$

ตารางการวิเคราะห์ความแปรปรวน

Source	df	SS	MS	F	P-value
Regression on X_1 , X_2	2	17266.174	8633.087	53.167	< 0.001
due to X_1	1	11193.163			
due to X_2 / X_1	1	6073.011	6073.011	37.401	< 0.001
Error	131	21271.386	162.377		
Total	133	38537.560			

เมื่อกำหนด $\alpha = 0.05$ ค่า P-value มีค่า < 0.001 ซึ่งน้อยกว่า α จึงทำการปฏิเสธ H_o แสดงว่ารูปแบบการถดถอยที่เหมาะสมคือ $Y = \beta_0 + \beta_1 X_1 + \beta_4 X_2 + \varepsilon$

จากผลการวิเคราะห์รูปแบบในข้อ ก. ถึงข้อ ฉ. สามารถสรุปผลการวิเคราะห์ได้ดังตารางต่อไปนี้

ตารางที่ 4.11 รูปแบบการถดถอยกับตัวแปรดัมมี่สำหรับข้อมูลสองกลุ่มของนักศึกษาสาขาวิชาสถิติประยุกต์ ปีการศึกษา 2555

รูปแบบการถดถอยในข้อ	ทดสอบ	รูปแบบการถดถอยที่เหมาะสม	ค่า R^2_a
ก). $Y = \beta_0 + \beta_1 X_1 + \varepsilon$	$\beta_1 = 0$	$Y = \beta_0 + \beta_1 X_1 + \varepsilon$	0.285
ข). $Y = \beta_0 + \beta_1 X_1 + \beta_2 D + \varepsilon$	$\beta_2 = 0$	$Y = \beta_0 + \beta_1 X_1 + \beta_2 D + \varepsilon$	0.319
ค). $Y = \beta_0 + \beta_1 X_1 + \beta_3 DX_1 + \varepsilon$	$\beta_3 = 0$	$Y = \beta_0 + \beta_1 X_1 + \beta_3 DX_1 + \varepsilon$	0.302
ง). $Y = \beta_0 + \beta_1 X_1 + \beta_2 D + \beta_3 DX_1 + \varepsilon$	$\beta_2 = 0$	$Y = \beta_0 + \beta_1 X_1 + \beta_2 D + \beta_3 DX_1 + \varepsilon$	0.337
จ). $Y = \beta_0 + \beta_1 X_1 + \beta_2 D + \beta_3 DX_1 + \varepsilon$	$\beta_3 = 0$	$Y = \beta_0 + \beta_1 X_1 + \beta_2 D + \beta_3 DX_1 + \varepsilon$	0.337
ฉ). $Y = \beta_0 + \beta_1 X_1 + \beta_4 X_2 + \varepsilon$	$\beta_4 = 0$	$Y = \beta_0 + \beta_1 X_1 + \beta_4 X_2 + \varepsilon$	0.440

จากตารางที่ 4.11 พบว่า เมื่อพิจารณาตัวแปรอิสระ X_1 เพียงตัวเดียวในสมการถดถอย จะส่งผลให้ค่า R^2_a มีค่าเท่ากับ 0.285 และเมื่อพิจารณาตัวแปรอิสระ X_1 และตัวแปรดัมมี่ D และ DX_1 เข้าสู่สมการถดถอย ค่า R^2_a มีค่าเป็น 0.337 แต่ถ้าพิจารณาตัวแปรอิสระ X_1 และ X_2 เข้าสู่สมการถดถอย จะทำให้ค่า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของผลงานทุกครั้งที่มีการนำไปใช้

R^2_a มีค่าเป็น 0.440 ดังนั้น กรณีที่ใช้ตัวแปรดัมมี่จะได้รูปแบบสมการถดถอยที่เหมาะสมสำหรับข้อมูลชุดนี้คือ $Y = \beta_0 + \beta_1 X_1 + \beta_2 D + \beta_3 D X_1 + \varepsilon$ ค่า R^2_a มีค่าเป็น 0.337 สำหรับกรณีเพิ่มตัวแปรอิสระ X_2 จะได้รูปแบบสมการถดถอยที่เหมาะสมสำหรับข้อมูลชุดนี้คือ $Y = \beta_0 + \beta_1 X_1 + \beta_4 X_2 + \varepsilon$ ค่า R^2_a มีค่าเป็น 0.440 จะเห็นได้ว่า ในการเพิ่มตัวแปรอิสระ X_2 จะเหมาะสมกว่าในการใช้ตัวแปรดัมมี่สำหรับข้อมูลชุดนี้

4.2.2 ผลการวิเคราะห์วิชาสถิติเบื้องต้นของนักศึกษาสาขาสถิติประยุกต์ ปีการศึกษา 2556

ก). จากรูปแบบ $Y = \beta_0 + \beta_1 X_1 + \varepsilon$ ทำการทดสอบสมมติฐาน

$H_0 : \beta_1 = 0$ $H_1 : \beta_1 \neq 0$

ตารางการวิเคราะห์ความแปรปรวน

Source	df	SS	MS	F	P-value
Regression on X_1	1	15350.389	15350.389	58.249	< 0.001
Error	134	35313.015	263.530		
Total	135	50663.404			

เมื่อกำหนด $\alpha = 0.05$ ค่า P-value มีค่า < 0.001 ซึ่งน้อยกว่า α จึงทำการปฏิเสธ H_0 แสดงว่ารูปแบบการถดถอยที่เหมาะสมคือ $Y = \beta_0 + \beta_1 X_1 + \varepsilon$

ข). จากรูปแบบ $Y = \beta_0 + \beta_1 X_1 + \beta_2 D + \varepsilon$ ทำการทดสอบสมมติฐาน

$H_0 : \beta_2 = 0$ $H_1 : \beta_2 \neq 0$

ตารางการวิเคราะห์ความแปรปรวน

Source	df	SS	MS	F	P-value
Regression on X_1 , D	2	15767.131	7883.566	30.047	< 0.001
due to X_1	1	15350.389			
due to D / X_1	1	416.742	416.742	1.588	0.210
Error	133	34896.273	262.376		
Total	135	50663.404			

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมื่อกำหนด $\alpha = 0.05$ ค่า P-value มีค่าเท่ากับ 0.210 ซึ่งมากกว่า α จึงทำการยอมรับ H_o แสดงว่ารูปแบบการถดถอยที่เหมาะสมคือ $Y = \beta_0 + \beta_1 X_1 + \varepsilon$

ค). จากรูปแบบ $Y = \beta_0 + \beta_1 X_1 + \beta_3 DX_1 + \varepsilon$ ทำการทดสอบสมมติฐาน

$H_o : \beta_3 = 0$ $H_1 : \beta_3 \neq 0$

ตารางการวิเคราะห์ความแปรปรวน

Source	df	SS	MS	F	P-value
Regression on X_1 , DX_1	2	15749.036	7874.518	29.997	< 0.001
due to X_1	1	15350.389			
due to DX_1 / X_1	1	398.647	398.647	1.518	0.220
Error	133	34914.368	262.514		
Total	135	50663.404			

เมื่อกำหนด $\alpha = 0.05$ ค่า P-value มีค่าเท่ากับ 0.220 ซึ่งมากกว่า α จึงทำการยอมรับ H_o แสดงว่ารูปแบบการถดถอยที่เหมาะสมคือ $Y = \beta_0 + \beta_1 X_1 + \varepsilon$

ง). จากรูปแบบ $Y = \beta_0 + \beta_1 X_1 + \beta_2 D + \beta_3 DX_1 + \varepsilon$ ทำการทดสอบสมมติฐาน

$H_o : \beta_2 = 0$ $H_1 : \beta_2 \neq 0$

ตารางการวิเคราะห์ความแปรปรวน

Source	df	SS	MS	F	P-value
Regression on X_1 , D , DX_1	3	15768.300	5256.100	19.883	< 0.001
due to X_1 , DX_1	2	15749.036			
due to D / X_1 , DX_1	1	19.264	19.264	0.073	0.787
Error	132	34895.104	264.357		
Total	135	50663.404			

เมื่อกำหนด $\alpha = 0.05$ ค่า P-value มีค่าเท่ากับ 0.787 ซึ่งมากกว่า α จึงทำการยอมรับ H_o แสดงว่ารูปแบบการถดถอยที่เหมาะสมคือ $Y = \beta_0 + \beta_1 X_1 + \beta_3 DX_1 + \varepsilon$

เอกสารนี้ใช้เพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่อนุญาตให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ข). จากรูปแบบ $Y = \beta_0 + \beta_1 X_1 + \beta_2 D + \beta_3 DX_1 + \varepsilon$ ทำการทดสอบสมมติฐาน

$H_0 : \beta_3 = 0$ $H_1 : \beta_3 \neq 0$

ตารางการวิเคราะห์ความแปรปรวน

Source	df	SS	MS	F	P-value
Regression on X_1 , D , DX_1	3	15768.300	5256.100	19.883	< 0.001
due to X_1 , D	2	15767.131			
due to DX_1 / X_1 , D	1	1.169	1.169	0.004	0.950
Error	132	34895.104	264.357		
Total	135	50663.404			

เมื่อกำหนด $\alpha = 0.05$ ค่า P-value มีค่าเท่ากับ 0.950 ซึ่งมากกว่า α จึงทำการยอมรับ H_0 แสดงว่ารูปแบบการถดถอยที่เหมาะสมคือ $Y = \beta_0 + \beta_1 X_1 + \beta_2 D + \varepsilon$

ฉ). จากรูปแบบ $Y = \beta_0 + \beta_1 X_1 + \beta_4 X_2 + \varepsilon$ ทำการทดสอบสมมติฐาน

$H_0 : \beta_4 = 0$ $H_1 : \beta_4 \neq 0$

ตารางการวิเคราะห์ความแปรปรวน

Source	df	SS	MS	F	P-value
Regression on X_1 , X_2	2	23581.721	11790.861	57.906	< 0.001
due to X_1	1	15350.389			
due to X_2 / X_1	1	8231.332	8231.332	40.425	< 0.001
Error	133	27081.683	203.622		
Total	135	50663.404			

เมื่อกำหนด $\alpha = 0.05$ ค่า P-value มีค่าเท่ากับ 0.00 ซึ่งน้อยกว่า α จึงทำการปฏิเสธ H_0 แสดงว่ารูปแบบการถดถอยที่เหมาะสมคือ $Y = \beta_0 + \beta_1 X_1 + \beta_4 X_2 + \varepsilon$

จากผลการวิเคราะห์รูปแบบในข้อ ก. ถึงข้อ ฉ. สามารถสรุปผลการวิเคราะห์ได้ดังตารางต่อไปนี้

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.12 รูปแบบการถดถอยกับตัวแปรดัมมี่สำหรับข้อมูลสองกลุ่มของนักศึกษาสาขาวิชาสถิติประยุกต์ ปีการศึกษา 2556

รูปแบบการถดถอยในข้อ	ทดสอบ	รูปแบบการถดถอยที่เหมาะสม	ค่า R^2_a
ก). $Y = \beta_0 + \beta_1 X_1 + \varepsilon$	$\beta_1 = 0$	$Y = \beta_0 + \beta_1 X_1 + \varepsilon$	0.298
ข). $Y = \beta_0 + \beta_1 X_1 + \beta_2 D + \varepsilon$	$\beta_2 = 0$	$Y = \beta_0 + \beta_1 X_1 + \varepsilon$	0.298
ค). $Y = \beta_0 + \beta_1 X_1 + \beta_3 DX_1 + \varepsilon$	$\beta_3 = 0$	$Y = \beta_0 + \beta_1 X_1 + \varepsilon$	0.298
ง). $Y = \beta_0 + \beta_1 X_1 + \beta_2 D + \beta_3 DX_1 + \varepsilon$	$\beta_2 = 0$	$Y = \beta_0 + \beta_1 X_1 + \beta_3 DX_1 + \varepsilon$	0.300
จ). $Y = \beta_0 + \beta_1 X_1 + \beta_2 D + \beta_3 DX_1 + \varepsilon$	$\beta_3 = 0$	$Y = \beta_0 + \beta_1 X_1 + \beta_2 D + \varepsilon$	0.301
ฉ). $Y = \beta_0 + \beta_1 X_1 + \beta_4 X_2 + \varepsilon$	$\beta_4 = 0$	$Y = \beta_0 + \beta_1 X_1 + \beta_4 X_2 + \varepsilon$	0.457

จากตารางที่ 4.12 พบว่า เมื่อพิจารณาตัวแปรอิสระ X_1 เพียงตัวเดียวในสมการถดถอย จะส่งผลให้ค่า R^2_a มีค่าเท่ากับ 0.298 และเมื่อพิจารณาตัวแปรอิสระ X_1 และตัวแปรดัมมี่ D และ/หรือ DX_1 เข้าสู่สมการถดถอย จะเห็นว่า ตัวแปรดัมมี่ไม่มีอิทธิพลต่อตัวแปรตาม Y แต่ถ้าพิจารณาตัวแปรอิสระ X_1 และ X_2 เข้าสู่สมการถดถอย จะทำให้ค่า R^2_a มีค่าเป็น 0.457 ดังนั้น รูปแบบสมการถดถอยที่เหมาะสมสำหรับข้อมูลชุดนี้คือ $Y = \beta_0 + \beta_1 X_1 + \beta_4 X_2 + \varepsilon$ ค่า R^2_a มีค่าเป็น 0.457 จะเห็นได้ว่า การใช้ตัวแปรดัมมี่สำหรับข้อมูลชุดนี้ไม่เหมาะสม

4.2.3 ผลการวิเคราะห์วิชาสถิติเบื้องต้นของนักศึกษาสาขาวิชาจุลชีววิทยาปีการศึกษา 2557

ก). จากรูปแบบ $Y = \beta_0 + \beta_1 X_1 + \varepsilon$ ทำการทดสอบสมมติฐาน

$H_0 : \beta_1 = 0$ $\quad$ $H_1 : \beta_1 \neq 0$

ตารางการวิเคราะห์ความแปรปรวน

Source	df	SS	MS	F	P-value
Regression on X_1	1	16148.481	16148.481	64.231	< 0.001
Error	118	29666.819	251.414		
Total	119	45815.300			

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมื่อกำหนด $\alpha = 0.05$ ค่า P-value มีค่า < 0.001 ซึ่งน้อยกว่า α จึงทำการปฏิเสธ H_0 แสดงว่า รูปแบบการถดถอยที่เหมาะสมคือ $Y = \beta_0 + \beta_1 X_1 + \varepsilon$

ข). จากรูปแบบ $Y = \beta_0 + \beta_1 X_1 + \beta_2 D + \varepsilon$ ทำการทดสอบสมมติฐาน

$H_0 : \beta_2 = 0$ $H_1 : \beta_2 \neq 0$

ตารางการวิเคราะห์ความแปรปรวน

Source	df	SS	MS	F	P-value
Regression on X_1 , D	2	16361.237	8180.619	32.496	< 0.001
due to X_1	1	16148.481			
due to D/X_1	1	212.756	212.756	0.845	0.360
Error	117	29454.063	251.744		
Total	119	45815.300			

เมื่อกำหนด $\alpha = 0.05$ ค่า P-value มีค่าเท่ากับ 0.360 ซึ่งมากกว่า α จึงทำการยอมรับ H_0 แสดงว่ารูปแบบการถดถอยที่เหมาะสมคือ $Y = \beta_0 + \beta_1 X_1 + \varepsilon$

ค). จากรูปแบบ $Y = \beta_0 + \beta_1 X_1 + \beta_3 DX_1 + \varepsilon$ ทำการทดสอบสมมติฐาน

$H_0 : \beta_3 = 0$ $H_1 : \beta_3 \neq 0$

ตารางการวิเคราะห์ความแปรปรวน

Source	df	SS	MS	F	P-value
Regression on X_1 , DX_1	2	16160.871	8080.435	31.881	< 0.001
due to X_1	1	16148.481			
due to DX_1 / X_1	1	12.390	12.390	0.049	0.825
Error	117	29654.429	253.457		
Total	119	45815.300			

เมื่อกำหนด $\alpha = 0.05$ ค่า P-value มีค่าเท่ากับ 0.825 ซึ่งมากกว่า α จึงทำการยอมรับ H_0 แสดงว่ารูปแบบการถดถอยที่เหมาะสมคือ $Y = \beta_0 + \beta_1 X_1 + \varepsilon$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ง). จากรูปแบบ $Y = \beta_0 + \beta_1 X_1 + \beta_2 D + \beta_3 DX_1 + \varepsilon$ ทำการทดสอบสมมติฐาน

$H_0 : \beta_2 = 0$ $H_1 : \beta_2 \neq 0$

ตารางการวิเคราะห์ความแปรปรวน

Source	df	SS	MS	F	P-value
Regression on X_1 , D, DX_1	3	17872.777	5957.592	24.732	< 0.001
due to X_1 , DX_1	2	16160.871			
due to D / X_1 , DX_1	1	1711.906	1711.906	7.107	0.009
Error	116	27942.523	240.884		
Total	119	45815.300			

เมื่อกำหนด $\alpha = 0.05$ ค่า P-value มีค่าเท่ากับ 0.009 ซึ่งน้อยกว่า α จึงทำการปฏิเสธ H_0 แสดงว่ารูปแบบการถดถอยที่เหมาะสมคือ $Y = \beta_0 + \beta_1 X_1 + \beta_2 D + \beta_3 DX_1 + \varepsilon$

จ). จากรูปแบบ $Y = \beta_0 + \beta_1 X_1 + \beta_2 D + \beta_3 DX_1 + \varepsilon$ ทำการทดสอบสมมติฐาน

$H_0 : \beta_3 = 0$ $H_1 : \beta_3 \neq 0$

ตารางการวิเคราะห์ความแปรปรวน

Source	df	SS	MS	F	P-value
Regression on X_1 , D , DX_1	3	17872.777	5957.592	24.732	< 0.001
due to X_1 , D	2	16361.237			
due to DX_1 / X_1 , D	1	1511.54	1511.54	6.275	0.014
Error	116	27942.523	240.884		
Total	119	45815.300			

เมื่อกำหนด $\alpha = 0.05$ ค่า P-value มีค่าเท่ากับ 0.014 ซึ่งน้อยกว่า α จึงทำการปฏิเสธ H_0 แสดงว่ารูปแบบการถดถอยที่เหมาะสมคือ $Y = \beta_0 + \beta_1 X_1 + \beta_2 D + \beta_3 DX_1 + \varepsilon$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ฉ). จากรูปแบบ $Y = \beta_0 + \beta_1 X_1 + \beta_4 X_2 + \varepsilon$ ทำการทดสอบสมมติฐาน

$H_0 : \beta_4 = 0$ $\quad$ $H_1 : \beta_4 \neq 0$

ตารางการวิเคราะห์ความแปรปรวน

Source	df	SS	MS	F	P-value
Regression on X_1 , X_2	2	24970.741	12485.371	70.080	< 0.001
due to X_1	1	16148.481			
due to X_2 / X_1	1	8822.260	8822.260	49.519	< 0.001
Error	117	20844.559	178.159		
Total	119	45815.300			

เมื่อกำหนด $\alpha = 0.05$ ค่า P-value มีค่า < 0.001 ซึ่งน้อยกว่า α จึงทำการปฏิเสธ H_0 แสดงว่ารูปแบบการถดถอยที่เหมาะสมคือ $Y = \beta_0 + \beta_1 X_1 + \beta_4 X_2 + \varepsilon$

จากผลการวิเคราะห์รูปแบบในข้อ ก. ถึงข้อ ฉ. สามารถสรุปผลการวิเคราะห์ได้ดังตารางต่อไปนี้

ตารางที่4.13 รูปแบบการถดถอยกับตัวแปรดัมมี่สำหรับข้อมูลสองกลุ่มของนักศึกษาสาขาวิชาจุลชีววิทยา ปีการศึกษา 2557

รูปแบบการถดถอยในข้อ	ทดสอบ	รูปแบบการถดถอยที่เหมาะสม	ค่า R^2_a
ก). $Y = \beta_0 + \beta_1 X_1 + \varepsilon$	$\beta_1 = 0$	$Y = \beta_0 + \beta_1 X_1 + \varepsilon$	0.347
ข). $Y = \beta_0 + \beta_1 X_1 + \beta_2 D + \varepsilon$	$\beta_2 = 0$	$Y = \beta_0 + \beta_1 X_1 + \varepsilon$	0.347
ค). $Y = \beta_0 + \beta_1 X_1 + \beta_3 DX_1 + \varepsilon$	$\beta_3 = 0$	$Y = \beta_0 + \beta_1 X_1 + \varepsilon$	0.347
ง). $Y = \beta_0 + \beta_1 X_1 + \beta_2 D + \beta_3 DX_1 + \varepsilon$	$\beta_2 = 0$	$Y = \beta_0 + \beta_1 X_1 + \beta_2 D + \beta_3 DX_1 + \varepsilon$	0.374
จ). $Y = \beta_0 + \beta_1 X_1 + \beta_2 D + \beta_3 DX_1 + \varepsilon$	$\beta_3 = 0$	$Y = \beta_0 + \beta_1 X_1 + \beta_2 D + \beta_3 DX_1 + \varepsilon$	0.374
ฉ). $Y = \beta_0 + \beta_1 X_1 + \beta_4 X_2 + \varepsilon$	$\beta_4 = 0$	$Y = \beta_0 + \beta_1 X_1 + \beta_4 X_2 + \varepsilon$	0.537

จากตารางที่ 4.13 พบว่า เมื่อพิจารณาตัวแปรอิสระ X_1 เพียงตัวเดียวในสมการถดถอย จะส่งผลให้ค่า R^2_a มีค่าเท่ากับ 0.347 และเมื่อพิจารณาตัวแปรอิสระ X_1 และตัวแปรดัมมี่ D และ DX_1 เข้าสู่สมการถดถอย ค่า R^2_a มีค่าเป็น 0.374 แต่ถ้าพิจารณาตัวแปรอิสระ X_1 และ X_2 เข้าสู่สมการถดถอย จะทำให้ค่า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

R^2_a มีค่าเป็น 0.537 ดังนั้น กรณีที่ใช้ตัวแปรดัมมี่จะได้รูปแบบสมการถดถอยที่เหมาะสมสำหรับข้อมูลชุดนี้คือ $Y = \beta_0 + \beta_1 X_1 + \beta_2 D + \beta_3 DX_1 + \varepsilon$ ค่า R^2_a มีค่าเป็น 0.374 สำหรับกรณีเพิ่มตัวแปรอิสระ X_2 จะได้รูปแบบสมการถดถอยที่เหมาะสมสำหรับข้อมูลชุดนี้คือ $Y = \beta_0 + \beta_1 X_1 + \beta_4 X_2 + \varepsilon$ ค่า R^2_a มีค่าเป็น 0.537 จะเห็นได้ว่า ในการเพิ่มตัวแปรอิสระ X_2 จะเหมาะสมกว่าในการใช้ตัวแปรดัมมี่สำหรับข้อมูลชุดนี้

4.2.4 ผลการวิเคราะห์วิชาสถิติเบื้องต้นของนักศึกษาสาขาวิชาสัตวศาสตร์ ปีการศึกษา 2555

ก). จากรูปแบบ $Y = \beta_0 + \beta_1 X_1 + \varepsilon$ ทำการทดสอบสมมติฐาน

$H_0 : \beta_1 = 0$ $H_1 : \beta_1 \neq 0$

ตารางการวิเคราะห์ความแปรปรวน

Source	df	SS	MS	F	P-value
Regression on X_1	1	8822.663	8822.663	49.631	< 0.001
Error	98	17420.897	177.764		
Total	99	26243.560			

เมื่อกำหนด $\alpha = 0.05$ ค่า P-value มีค่า < 0.001 ซึ่งน้อยกว่า α จึงทำการปฏิเสธ H_0 แสดงว่ารูปแบบการถดถอยที่เหมาะสมคือ $Y = \beta_0 + \beta_1 X_1 + \varepsilon$

ข). จากรูปแบบ $Y = \beta_0 + \beta_1 X_1 + \beta_2 D + \varepsilon$ ทำการทดสอบสมมติฐาน

$H_0 : \beta_2 = 0$ $H_1 : \beta_2 \neq 0$

ตารางการวิเคราะห์ความแปรปรวน

Source	df	SS	MS	F	P-value
Regression on X_1 , D	2	9341.106	4670.553	26.803	< 0.001
due to X_1	1	8822.663			
due to D / X_1	1	518.443	518.443	2.975	0.088
Error	97	16902.454	174.252		
Total	99	26243.560			

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมื่อกำหนด $\alpha = 0.05$ ค่า P-value มีค่าเท่ากับ 0.088 ซึ่งมากกว่า α จึงทำการยอมรับ H_o แสดงว่ารูปแบบการถดถอยที่เหมาะสมคือ $Y = \beta_0 + \beta_1 X_1 + \varepsilon$

ค). จากรูปแบบ $Y = \beta_0 + \beta_1 X_1 + \beta_3 DX_1 + \varepsilon$ ทำการทดสอบสมมติฐาน

$H_o : \beta_3 = 0$ $H_1 : \beta_3 \neq 0$

ตารางการวิเคราะห์ความแปรปรวน

Source	df	SS	MS	F	P-value
Regression on X_1 , DX_1	2	9450.651	4725.325	27.295	< 0.001
due to X_1	1	8822.663			
due to DX_1 / X_1	1	627.988	627.988	3.627	0.060
Error	97	16792.909	173.123		
Total	99	26243.560			

เมื่อกำหนด $\alpha = 0.05$ ค่า P-value มีค่าเท่ากับ 0.060 ซึ่งมากกว่า α จึงทำการยอมรับ H_o แสดงว่ารูปแบบการถดถอยที่เหมาะสมคือ $Y = \beta_0 + \beta_1 X + \varepsilon$

ง). จากรูปแบบ $Y = \beta_0 + \beta_1 X_1 + \beta_2 D + \beta_3 DX_1 + \varepsilon$ ทำการทดสอบสมมติฐาน

$H_o : \beta_2 = 0$ $H_1 : \beta_2 \neq 0$

ตารางการวิเคราะห์ความแปรปรวน

Source	df	SS	MS	F	P-value
Regression on X_1 , D , DX_1	3	9471.054	3157.018	18.070	< 0.001
due to X_1 , DX_1	2	9450.651			
due to D / X_1 , DX_1	1	20.403	20.403	0.117	0.733
Error	96	16772.506	174.714		
Total	99	26243.560			

เมื่อกำหนด $\alpha = 0.05$ ค่า P-value มีค่าเท่ากับ 0.733 ซึ่งมากกว่า α จึงทำการยอมรับ H_o แสดงว่ารูปแบบการถดถอยที่เหมาะสมคือ $Y = \beta_0 + \beta_1 X_1 + \beta_3 DX_1 + \varepsilon$

เอกสารนี้ใช้เพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่อนุญาตให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ข). จากรูปแบบ $Y = \beta_0 + \beta_1 X_1 + \beta_2 D + \beta_3 DX_1 + \varepsilon$ ทำการทดสอบสมมติฐาน

$H_0 : \beta_3 = 0$ $H_1 : \beta_3 \neq 0$

ตารางการวิเคราะห์ความแปรปรวน

Source	df	SS	MS	F	P-value
Regression on X_1 , D , DX_1	3	9471.054	3157.018	18.070	< 0.001
due to X_1 , D	2	9341.106			
due to DX_1 / X_1 , D	1	129.948	129.948	0.744	0.391
Error	96	16772.506	174.714		
Total	99	26243.560			

เมื่อกำหนด $\alpha = 0.05$ ค่า P-value มีค่าเท่ากับ 0.391 ซึ่งมากกว่า α จึงทำการยอมรับ H_0 แสดงว่ารูปแบบการถดถอยที่เหมาะสมคือ $Y = \beta_0 + \beta_1 X_1 + \beta_2 D + \varepsilon$

ฉ). จากรูปแบบ $Y = \beta_0 + \beta_1 X_1 + \beta_4 X_2 + \varepsilon$ ทำการทดสอบสมมติฐาน

$H_0 : \beta_4 = 0$ $H_1 : \beta_4 \neq 0$

ตารางการวิเคราะห์ความแปรปรวน

Source	df	SS	MS	F	P-value
Regression on X_1 , X_2	2	13951.325	6975.662	55.046	< 0.001
due to X_1	1	8822.663			
due to X_2 / X_1	1	5128.662	5128.662	40.471	< 0.001
Error	97	12292.235	126.724		
Total	99	26243.560			

เมื่อกำหนด $\alpha = 0.05$ ค่า P-value มีค่า < 0.001 ซึ่งน้อยกว่า α จึงทำการปฏิเสธ H_0 แสดงว่ารูปแบบการถดถอยที่เหมาะสมคือ $Y = \beta_0 + \beta_1 X_1 + \beta_4 X_2 + \varepsilon$

จากผลการวิเคราะห์รูปแบบในข้อ ก. ถึงข้อ ฉ. สามารถสรุปผลการวิเคราะห์ได้ดังตารางต่อไปนี้

เอกสารนี้เป็นการเผยแพร่เพื่อการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่4.14 รูปแบบการถดถอยกับตัวแปรดัมมี่สำหรับข้อมูลสองกลุ่มของนักศึกษาสาขาวิชาสัตวศาสตร์ ปีการศึกษา 2555

รูปแบบการถดถอยในข้อ	ทดสอบ	รูปแบบการถดถอยที่เหมาะสม	ค่า R^2_a
ก). $Y = \beta_0 + \beta_1 X_1 + \varepsilon$	$\beta_1 = 0$	$Y = \beta_0 + \beta_1 X_1 + \varepsilon$	0.329
ข). $Y = \beta_0 + \beta_1 X_1 + \beta_2 D + \varepsilon$	$\beta_2 = 0$	$Y = \beta_0 + \beta_1 X_1 + \varepsilon$	0.329
ค). $Y = \beta_0 + \beta_1 X_1 + \beta_3 DX_1 + \varepsilon$	$\beta_3 = 0$	$Y = \beta_0 + \beta_1 X_1 + \varepsilon$	0.329
ง). $Y = \beta_0 + \beta_1 X_1 + \beta_2 D + \beta_3 DX_1 + \varepsilon$	$\beta_2 = 0$	$Y = \beta_0 + \beta_1 X_1 + \beta_3 DX_1 + \varepsilon$	0.347
จ). $Y = \beta_0 + \beta_1 X_1 + \beta_2 D + \beta_3 DX_1 + \varepsilon$	$\beta_3 = 0$	$Y = \beta_0 + \beta_1 X_1 + \beta_2 D + \varepsilon$	0.343
ฉ). $Y = \beta_0 + \beta_1 X_1 + \beta_4 X_2 + \varepsilon$	$\beta_4 = 0$	$Y = \beta_0 + \beta_1 X_1 + \beta_4 X_2 + \varepsilon$	0.522

จากตารางที่ 4.14 พบว่า เมื่อพิจารณาตัวแปรอิสระ X_1 เพียงตัวเดียวในสมการถดถอย จะส่งผลให้ค่า R^2_a มีค่าเท่ากับ 0.329 และเมื่อพิจารณาตัวแปรอิสระ X_1 และตัวแปรดัมมี่ D และ/หรือ DX_1 เข้าสู่สมการถดถอย จะเห็นว่า ตัวแปรดัมมี่ไม่มีอิทธิพลต่อตัวแปรตาม Y แต่ถ้าพิจารณาตัวแปรอิสระ X_1 และ X_2 เข้าสู่สมการถดถอย จะทำให้ค่า R^2_a มีค่าเป็น 0.522 ดังนั้น รูปแบบสมการถดถอยที่เหมาะสมสำหรับข้อมูลชุดนี้คือ $Y = \beta_0 + \beta_1 X_1 + \beta_4 X_2 + \varepsilon$ ค่า R^2_a มีค่าเป็น 0.522 จะเห็นได้ว่า การใช้ตัวแปรดัมมี่สำหรับข้อมูลชุดนี้ไม่เหมาะสม

4.3 ผลการวิเคราะห์การถดถอยกับตัวแปรดัมมี่สำหรับข้อมูลสามกลุ่ม

ข้อมูลที่ใช้ในวิเคราะห์ประกอบด้วย

- ตัวแปรตามคือ คะแนนสอบปลายภาควิชาสถิติเบื้องต้นของนักศึกษา
- ส่วนตัวแปรอิสระ คือ คะแนนสอบกลางภาควิชาสถิติเบื้องต้น สาขาวิชา และเกรดเฉลี่ยสะสมของนักศึกษา

ซึ่งสามารถสร้างรูปแบบการถดถอยที่ใช้ในการวิเคราะห์ได้ดังนี้

รูปแบบที่ 1 $Y = \beta_0 + \beta_1 X_1 + \varepsilon$

รูปแบบที่ 2 $Y = \beta_0 + \beta_1 X_1 + \beta_2 D_1 + \beta_3 D_2 + \varepsilon$

รูปแบบที่ 3 $Y = \beta_0 + \beta_1 X_1 + \beta_4 D_1 X_1 + \beta_5 D_2 X_1 + \varepsilon$

รูปแบบที่ 4 $Y = \beta_0 + \beta_1 X_1 + \beta_2 D_1 + \beta_3 D_2 + \beta_4 D_1 X_1 + \beta_5 D_2 X_1 + \varepsilon$

รูปแบบที่ 5 $Y = \beta_0 + \beta_1 X_1 + \beta_6 X_2 + \varepsilon$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมื่อ Y คือ ตัวแปรตาม ในที่นี้คือ คะแนนสอบปลายภาค

X_1 คือ ตัวแปรอิสระตัวที่ 1 ในที่นี้ คือ คะแนนสอบกลางภาค

X_2 คือ ตัวแปรอิสระตัวที่ 2 ในที่นี้ คือ เกรดเฉลี่ยสะสม

D_1 คือ ตัวแปรดัมมี่ตัวที่ 1 ในที่นี้หมายถึง

$D_1 = 0$ สำหรับนักศึกษาที่ไม่ใช่สาขาวิชาสถิติประยุกต์

$D_1 = 1$ สำหรับนักศึกษาสาขาวิชาสถิติประยุกต์

D_2 คือ ตัวแปรดัมมี่ตัวที่ 2 ในที่นี้หมายถึง

$D_2 = 0$ สำหรับนักศึกษาที่ไม่ใช่สาขาวิชาชีววิทยา

$D_2 = 1$ สำหรับนักศึกษาสาขาวิชาชีววิทยา

$\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6$ คือ สัมประสิทธิ์การถดถอย

ก). จากรูปแบบ $Y = \beta_0 + \beta_1 X_1 + \varepsilon$ ทำการทดสอบสมมติฐาน

$H_0 : \beta_1 = 0$ $\quad H_1 : \beta_1 \neq 0$

ตารางการวิเคราะห์ความแปรปรวน

Source	df	SS	MS	F	P-value
Regression on X_1	1	62029.623	62029.623	268.582	< 0.001
Error	488	112704.777	230.952		
Total	489	174734.400			

เมื่อกำหนด $\alpha = 0.05$ ค่า P-value มีค่า < 0.001 ซึ่งน้อยกว่า α จึงทำการปฏิเสธ H_0 แสดงว่ารูปแบบการถดถอยที่เหมาะสมคือ $Y = \beta_0 + \beta_1 X_1 + \varepsilon$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ข). จากรูปแบบ $Y = \beta_0 + \beta_1 X_1 + \beta_2 D_1 + \beta_3 D_2 + \varepsilon$ ทำการทดสอบสมมติฐาน

$H_0 : \beta_2 = \beta_3 = 0$ $H_1 : \beta_2$ และ β_3 ไม่เป็น 0 พร้อมกัน

ตารางการวิเคราะห์ความแปรปรวน

Source	df	SS	MS	F	P-value
Regression on X_1 , D_1 , D_2	3	62085.870	20695.290	89.286	< 0.001
due to X_1	1	62029.623			
due to D_1 , D_2 / X_1	2	56.247	28.124	0.121	0.886
Error	486	112648.530	231.787		
Total	489	174734.400			

เมื่อกำหนด $\alpha = 0.05$ ค่า P-value มีค่าเท่ากับ 0.886 ซึ่งมากกว่า α จึงทำการยอมรับ H_0 แสดงว่ารูปแบบการถดถอยที่เหมาะสมคือ $Y = \beta_0 + \beta_1 X_1 + \varepsilon$

ค). จากรูปแบบ $Y = \beta_0 + \beta_1 X_1 + \beta_4 D_1 X_1 + \beta_5 D_2 X_1 + \varepsilon$ ทำการทดสอบสมมติฐาน

$H_0 : \beta_4 = \beta_5 = 0$ $H_1 : \beta_4$ และ β_5 ไม่เป็น 0 พร้อมกัน

ตารางการวิเคราะห์ความแปรปรวน

Source	df	SS	MS	F	P-value
Regression on X_1 , D_1X_1 , D_2X_1	3	62276.412	20758.804	89.712	< 0.001
due to X_1	1	62029.623			
due to D_1X_1 , D_2X_1 / X_1	2	246.789	123.395	0.533	0.587
Error	486	112457.988	231.395		
Total	489	174734.400			

เมื่อกำหนด $\alpha = 0.05$ ค่า P-value มีค่าเท่ากับ 0.587 ซึ่งมากกว่า α จึงทำการยอมรับ H_0 แสดงว่ารูปแบบการถดถอยที่เหมาะสมคือ $Y = \beta_0 + \beta_1 X_1 + \varepsilon$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ง). จากรูปแบบ $Y = \beta_0 + \beta_1 X_1 + \beta_2 D_1 + \beta_3 D_2 + \beta_4 D_1 X_1 + \beta_5 D_2 X_1 + \varepsilon$ ทำการทดสอบสมมติฐาน

$H_0 : \beta_2 = \beta_3 = 0$ $H_1 : \beta_2$ และ β_3 ไม่เป็น 0 พร้อมกัน

ตารางการวิเคราะห์ความแปรปรวน

Source	df	SS	MS	F	P-value
Regression on $X_1, D_1, D_2, D_1X_1, D_2X_1$	5	62940.693	12588.139	54.499	< 0.001
due to X_1, D_1X_1, D_2X_1	3	62276.412			
due to $D_1, D_2 / X_1, D_1X_1, D_2X_1$	2	664.281	332.141	1.438	0.238
Error	484	111793.707	230.979		
Total	489	174734.400			

เมื่อกำหนด $\alpha = 0.05$ ค่า P-value มีค่าเท่ากับ 0.238 ซึ่งมากกว่า α จึงทำการยอมรับ H_0 แสดงว่ารูปแบบการถดถอยที่เหมาะสมคือ $Y = \beta_0 + \beta_1 X_1 + \beta_4 D_1 X_1 + \beta_5 D_2 X_1 + \varepsilon$

จ). จากรูปแบบ $Y = \beta_0 + \beta_1 X_1 + \beta_2 D_1 + \beta_3 D_2 + \beta_4 D_1 X_1 + \beta_5 D_2 X_1 + \varepsilon$ ทำการทดสอบสมมติฐาน

$H_0 : \beta_4 = \beta_5 = 0$ $H_1 : \beta_4$ และ β_5 ไม่เป็น 0 พร้อมกัน

ตารางการวิเคราะห์ความแปรปรวน

Source	df	SS	MS	F	P-value
Regression on $X_1, D_1, D_2, D_1X_1, D_2X_1$	5	62940.693	12588.139	54.499	< 0.001
due to X_1, D_1, D_2	3	62085.870			
due to $D_1X_1, D_2X_1 / X_1, D_1, D_2$	2	854.823	427.412	1.850	0.158
Error	484	111793.707	230.979		
Total	489	174734.400			

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมื่อกำหนด $\alpha = 0.05$ ค่า P-value มีค่าเท่ากับ 0.158 ซึ่งมากกว่า α จึงทำการยอมรับ H_0 แสดงว่ารูปแบบการถดถอยที่เหมาะสมคือ $Y = \beta_0 + \beta_1 X_1 + \beta_2 D_1 + \beta_3 D_2 + \varepsilon$

ฉ). จากรูปแบบ $Y = \beta_0 + \beta_1 X_1 + \beta_6 X_2 + \varepsilon$ ทำการทดสอบสมมติฐาน

$H_0 : \beta_6 = 0$ $H_1 : \beta_6 \neq 0$

ตารางการวิเคราะห์ความแปรปรวน

Source	df	SS	MS	F	P-value
Regression on X_1 , X_2	2	88432.134	44216.067	249.509	< 0.001
due to X_1	1	62029.623			
due to X_2 / X_1	1	26402.511	26402.511	148.988	< 0.001
Error	487	86302.266	177.212		
Total	489	174734.400			

เมื่อกำหนด $\alpha = 0.05$ ค่า P-value มีค่า < 0.001 ซึ่งน้อยกว่า α จึงทำการปฏิเสธ H_0 แสดงว่ารูปแบบการถดถอยที่เหมาะสมคือ $Y = \beta_0 + \beta_1 X_1 + \beta_6 X_2 + \varepsilon$

จากผลการวิเคราะห์รูปแบบในข้อ ก. ถึงข้อ ฉ. สามารถสรุปผลการวิเคราะห์ได้ดังตารางต่อไปนี้

ตารางที่4.15 รูปแบบการถดถอยกับตัวแปรดัมมี่สำหรับข้อมูลสามกลุ่มของนักศึกษาสาขาวิชาสถิติประยุกต์ สาขาวิชาชีวสถิติและสาขาวิชาสัตวศาสตร์

รูปแบบการถดถอยในข้อ	ทดสอบ	รูปแบบการถดถอยที่เหมาะสม	ค่า R^2_a
ก). $Y = \beta_0 + \beta_1 X_1 + \varepsilon$	$\beta_1 = 0$	$Y = \beta_0 + \beta_1 X_1 + \varepsilon$	0.354
ข). $Y = \beta_0 + \beta_1 X_1 + \beta_2 D_1 + \beta_3 D_2 + \varepsilon$	$\beta_2 = \beta_3 = 0$	$Y = \beta_0 + \beta_1 X_1 + \varepsilon$	0.354
ค). $Y = \beta_0 + \beta_1 X_1 + \beta_4 D_1 X_1 + \beta_5 D_2 X_1 + \varepsilon$	$\beta_4 = \beta_5 = 0$	$Y = \beta_0 + \beta_1 X_1 + \varepsilon$	0.354
ง). $Y = \beta_0 + \beta_1 X_1 + \beta_2 D_1 + \beta_3 D_2 + \beta_4 D_1 X_1 + \beta_5 D_2 X_1 + \varepsilon$	$\beta_2 = \beta_3 = 0$	$Y = \beta_0 + \beta_1 X_1 + \beta_4 D_1 X_1 + \beta_5 D_2 X_1 + \varepsilon$	0.352
จ). $Y = \beta_0 + \beta_1 X_1 + \beta_2 D_1 + \beta_3 D_2 + \beta_4 D_1 X_1 + \beta_5 D_2 X_1 + \varepsilon$	$\beta_4 = \beta_5 = 0$	$Y = \beta_0 + \beta_1 X_1 + \beta_2 D_1 + \beta_3 D_2 + \varepsilon$	0.351
ฉ). $Y = \beta_0 + \beta_1 X_1 + \beta_6 X_2 + \varepsilon$	$\beta_6 = 0$	$Y = \beta_0 + \beta_1 X_1 + \beta_6 X_2 + \varepsilon$	0.504

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางที่ 4.15 พบว่า เมื่อพิจารณาตัวแปรอิสระ X_1 เพียงตัวเดียวในสมการถดถอย จะส่งผลให้ค่า R^2_a มีค่าเท่ากับ 0.354 และเมื่อพิจารณาตัวแปรอิสระ X_1 และตัวแปรดัมมี่ D_1, D_2 และ/หรือ D_1X_1, D_2X_1 เข้าสู่สมการถดถอย จะเห็นว่า ตัวแปรดัมมี่ไม่มีอิทธิพลต่อตัวแปรตาม Y แต่ถ้าพิจารณาตัวแปรอิสระ X_1 และ X_2 เข้าสู่สมการถดถอย จะทำให้ค่า R^2_a มีค่าเป็น 0.504 ดังนั้น รูปแบบสมการถดถอยที่เหมาะสมสำหรับข้อมูลชุดนี้คือ $Y = \beta_0 + \beta_1 X_1 + \beta_6 X_2 + \varepsilon$ ค่า R^2_a มีค่าเป็น 0.504 จะเห็นได้ว่า การใช้ตัวแปรดัมมี่สำหรับข้อมูลชุดนี้ไม่เหมาะสม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5

สรุปผลการวิเคราะห์และข้อเสนอแนะ

ข้อมูลที่ใช้ในการวิเคราะห์ครั้งนี้ ประกอบด้วยข้อสอบกลางภาคและปลายภาควิชาสถิติเบื้องต้นของนักศึกษาสาขาวิชาสถิติประยุกต์ปีการศึกษา 2555-2556 สาขาวิชาจุลชีววิทยาปีการศึกษา 2557 และสาขาวิชาสัตวศาสตร์ปีการศึกษา 2555 รวมทั้งเพศและเกรดเฉลี่ยสะสมของนักศึกษา โดยใช้โปรแกรมสำเร็จรูป SPSS และโปรแกรม Microsoft Excel ในการประมวล สามารถสรุปการวิเคราะห์ได้ดังต่อไปนี้

5.1 คุณภาพข้อสอบ

5.1.1 ข้อสอบวิชาสถิติเบื้องต้นสาขาวิชาสถิติประยุกต์ ปีการศึกษา 2555

ข้อสอบวิชาสถิติเบื้องต้นสาขาวิชาสถิติประยุกต์ ปีการศึกษา 2555 ที่ใช้ในการวิเคราะห์ครั้งนี้มีทั้งหมด 134 ชุด ในส่วนของข้อสอบกลางภาคมีค่าความยากเฉลี่ยทั้งฉบับ $(\overline{P})$ เท่ากับ 0.450 ค่าความเชื่อมั่นทั้งฉบับ (KR-20) มีค่าเท่ากับ 0.517 และมีจำนวนข้อที่ถือว่าเป็นข้อสอบที่ใช้ได้คิดเป็นร้อยละ 68 ส่วนข้อสอบปลายภาคมีค่าความยากเฉลี่ยทั้งฉบับ $(\overline{P})$ เท่ากับ 0.494 ค่าความเชื่อมั่นทั้งฉบับ (KR-20) มีค่าเท่ากับ 0.691 และมีจำนวนข้อที่ถือว่าเป็นข้อสอบที่ใช้ได้คิดเป็นร้อยละ 76

5.1.2 ข้อสอบวิชาสถิติเบื้องต้นสาขาวิชาสถิติประยุกต์ ปีการศึกษา 2556

ข้อสอบวิชาสถิติเบื้องต้นสาขาวิชาสถิติประยุกต์ ปีการศึกษา 2556 ที่ใช้ในการวิเคราะห์ครั้งนี้มีทั้งหมด 136 ชุด ในส่วนของข้อสอบกลางภาคมีค่าความยากเฉลี่ยทั้งฉบับ $(\overline{P})$ เท่ากับ 0.533 ค่าความเชื่อมั่นทั้งฉบับ (KR-20) มีค่าเท่ากับ 0.688 และมีจำนวนข้อที่ถือว่าเป็นข้อสอบที่ใช้ได้คิดเป็นร้อยละ80 ส่วนข้อสอบปลายภาคมีค่าความยากเฉลี่ยทั้งฉบับ $(\overline{P})$ เท่ากับ 0.529 ค่าความเชื่อมั่นทั้งฉบับ (KR-20)มีค่าเท่ากับ 0.777 และมีจำนวนข้อที่ถือว่าเป็นข้อสอบที่ใช้ได้คิดเป็นร้อยละ 72

5.1.3 ข้อสอบวิชาสถิติเบื้องต้นสาขาวิชาจุลชีววิทยา ปีการศึกษา 2557

ข้อสอบวิชาสถิติเบื้องต้นสาขาวิชาจุลชีววิทยา ปีการศึกษา 2557 ที่ใช้ในการวิเคราะห์มีทั้งหมด 120 ชุด โดยข้อสอบกลางภาคมีค่าความยากเฉลี่ยทั้งฉบับ $(\overline{P})$ เท่ากับ 0.503 ค่าความเชื่อมั่นทั้งฉบับ (KR-20) มีค่าเท่ากับ 0.679 และมีจำนวนข้อที่ถือว่าเป็นข้อสอบที่ใช้ได้คิดเป็นร้อยละ 95 ส่วนข้อสอบปลายภาคมีค่าความยากเฉลี่ยทั้งฉบับ $(\overline{P})$ เท่ากับ 0.594 ค่าความเชื่อมั่นทั้งฉบับ (KR-20) มีค่าเท่ากับ 0.832 และมีจำนวนข้อที่ถือว่าเป็นข้อสอบที่ใช้ได้คิดเป็นร้อยละ 85

เอกสารนี้เป็นลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย ห้ามมิให้ทำซ้ำ ดัดแปลง หรือจำหน่ายเพื่อการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น มีทั้งห้ามมิให้ลบข้อความนี้ออก และห้ามมิให้ลบรอยน้ำของเอกสารทุกครั้งที่มีการนำไปใช้

5.1.4 ข้อสอบวิชาสถิติเบื้องต้นสาขาวิชาสัตวศาสตร์ ปีการศึกษา 2555

ข้อสอบวิชาสถิติเบื้องต้นสาขาวิชาสัตวศาสตร์ ปีการศึกษา 2555 ที่ใช้ในการวิเคราะห์มีทั้งหมด 100 ชุด โดยข้อสอบกลางภาคมีค่าความยากเฉลี่ยทั้งฉบับ $(\overline{P})$ เท่ากับ 0.359 ค่าความเชื่อมั่นทั้งฉบับ (KR-20) มีค่าเท่ากับ 0.260 และมีจำนวนข้อที่ถือว่าเป็นข้อสอบที่ใช้ได้คิดเป็นร้อยละ 52 ส่วนข้อสอบปลายภาคมีค่าความยากเฉลี่ยทั้งฉบับ $(\overline{P})$ เท่ากับ 0.403 ค่าความเชื่อมั่นทั้งฉบับ (KR-20) มีค่าเท่ากับ 0.555 และมีจำนวนข้อที่ถือว่าเป็นข้อสอบที่ใช้ได้คิดเป็นร้อยละ 80

ดังนั้น คุณภาพข้อสอบกลางภาคและปลายภาคในวิชาสถิติเบื้องต้นที่ใช้ในการวิเคราะห์ครั้งนี้ มีทั้งหมด 8 ฉบับ จะมีอยู่ 7 ฉบับที่มีค่าความยากเฉลี่ยทั้งฉบับ $(\overline{P})$ อยู่ที่ 0.40 – 0.50 ค่าความเชื่อมั่นทั้งฉบับ(KR-20) อยู่ที่ 0.50 – 0.80 และมีจำนวนข้อสอบที่ใช้ได้มากกว่าร้อยละ 65 แต่จะมีเพียงข้อสอบฉบับเดียวคือ ข้อสอบกลางภาคสาขาวิชาสัตวศาสตร์ ที่มีค่าความยากเฉลี่ยทั้งฉบับ $(\overline{P})$ เท่ากับ 0.359 ค่าความเชื่อมั่นทั้งฉบับ (KR-20) เท่ากับ 0.260 และมีจำนวนข้อสอบที่ใช้ได้ร้อยละ 52

เมื่อคุณภาพของข้อสอบกลางภาคและปลายภาคของวิชาสถิติเบื้องต้น ที่ใช้ในการวิเคราะห์ทั้ง 8 ฉบับไม่มีความแตกต่างกันมากนัก ส่งผลให้การนำคะแนนกลางภาคและปลายภาควิชาสถิติเบื้องต้นมาใช้ในการวิเคราะห์มีความน่าเชื่อถือของข้อมูลมากขึ้นไปด้วย ซึ่งถือเป็นการกำจัดความผันแปรที่อาจเกิดจากคุณภาพของข้อสอบ

5.2 การถดถอยกับตัวแปรดัมมี่สำหรับข้อมูลสองกลุ่ม

ข้อมูลที่ใช้ในวิเคราะห์ประกอบด้วย ตัวแปรตามคือ คะแนนสอบปลายภาควิชาสถิติเบื้องต้นของนักศึกษา(Y) ส่วนตัวแปรอิสระ คือ คะแนนสอบกลางภาควิชาสถิติเบื้องต้น(X_1) และเกรดเฉลี่ยสะสมของนักศึกษา (X_2) และตัวแปรดัมมี่ คือ เพศของนักศึกษา (D) สามารถสรุปผลได้ดังนี้

5.2.1 วิชาสถิติเบื้องต้นสาขาวิชาสถิติประยุกต์ ปีการศึกษา 2555

เมื่อพิจารณาตัวแปรอิสระ X_1 เพียงตัวเดียวในสมการถดถอย จะส่งผลให้ค่า R^2_a มีค่าเท่ากับ 0.285 และเมื่อพิจารณาตัวแปรอิสระ X_1 และตัวแปรดัมมี่ D และ DX_1 เข้าสู่สมการถดถอย ค่า R^2_a มีค่าเป็น 0.337 แต่ถ้าพิจารณาตัวแปรอิสระ X_1 และ X_2 เข้าสู่สมการถดถอย จะทำให้ค่า R^2_a มีค่าเป็น 0.440 ดังนั้น กรณีที่ใช้ตัวแปรดัมมี่จะได้รูปแบบสมการถดถอยที่เหมาะสมสำหรับข้อมูลชุดนี้คือ $Y = \beta_0 + \beta_1 X_1 + \beta_2 D + \beta_3 DX_1 + \varepsilon$ ค่า R^2_a มีค่าเป็น 0.337 สำหรับกรณีเพิ่มตัวแปรอิสระ X_2 จะได้รูปแบบสมการถดถอยที่เหมาะสมสำหรับข้อมูลชุดนี้คือ $Y = \beta_0 + \beta_1 X_1 + \beta_4 X_2 + \varepsilon$ ค่า R^2_a มีค่า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของผลงานทุกครั้งที่มีการนำไปใช้

เป็น 0.440 จะเห็นได้ว่า ในการเพิ่มตัวแปรอิสระ X_2 จะเหมาะสมกว่าในการใช้ตัวแปรดัมมี่สำหรับข้อมูลชุดนี้ ดังนั้น ในการประมาณคะแนนสอบปลายภาควิชาสถิติเบื้องต้นของนักศึกษาสาขาวิชาสถิติประยุกต์ ปีการศึกษา 2555 สามารถใช้เพศของนักศึกษามาเป็นตัวแปรดัมมี่ได้ แต่การใช้เกรดเฉลี่ยสะสมของนักศึกษาเป็นตัวแปรอิสระอีกตัวหนึ่ง จะส่งผลให้การประมาณคะแนนสอบปลายภาคของนักศึกษาได้ผลที่ดีกว่า

5.2.2 วิชาสถิติเบื้องต้นสาขาวิชาสถิติประยุกต์ ปีการศึกษา 2556

เมื่อพิจารณาตัวแปรอิสระ X_1 เพียงตัวเดียวในสมการถดถอย จะส่งผลให้ค่า R^2_a มีค่าเท่ากับ 0.298 และเมื่อพิจารณาตัวแปรอิสระ X_1 และตัวแปรดัมมี่ D และ/หรือ DX_1 เข้าสู่สมการถดถอย จะเห็นว่า ตัวแปรดัมมี่ไม่มีอิทธิพลต่อตัวแปรตาม Y แต่ถ้าพิจารณาตัวแปรอิสระ X_1 และ X_2 เข้าสู่สมการถดถอย จะทำให้ค่า R^2_a มีค่าเป็น 0.457 ดังนั้น รูปแบบสมการถดถอยที่เหมาะสมสำหรับข้อมูลชุดนี้คือ $Y = \beta_0 + \beta_1 X_1 + \beta_4 X_2 + \varepsilon$ ค่า R^2_a มีค่าเป็น 0.457 จะเห็นได้ว่า การใช้ตัวแปรดัมมี่สำหรับข้อมูลชุดนี้ไม่เหมาะสม ดังนั้น ในการประมาณคะแนนสอบปลายภาควิชาสถิติเบื้องต้นของนักศึกษาสาขาวิชาสถิติประยุกต์ ปีการศึกษา 2556 ไม่สามารถใช้เพศของนักศึกษามาเป็นตัวแปรดัมมี่ได้ การใช้เกรดเฉลี่ยสะสมของนักศึกษาเป็นตัวแปรอิสระอีกตัวหนึ่ง จะส่งผลให้การประมาณคะแนนสอบปลายภาคของนักศึกษาได้ผลที่ดีกว่า

5.2.3 วิชาสถิติเบื้องต้นสาขาวิชาจุลชีววิทยา ปีการศึกษา 2557

เมื่อพิจารณาตัวแปรอิสระ X_1 เพียงตัวเดียวในสมการถดถอย จะส่งผลให้ค่า R^2_a มีค่าเท่ากับ 0.347 และเมื่อพิจารณาตัวแปรอิสระ X_1 และตัวแปรดัมมี่ D และ DX_1 เข้าสู่สมการถดถอย ค่า R^2_a มีค่าเป็น 0.374 แต่ถ้าพิจารณาตัวแปรอิสระ X_1 และ X_2 เข้าสู่สมการถดถอย จะทำให้ค่า R^2_a มีค่าเป็น 0.537 ดังนั้น กรณีที่ใช้ตัวแปรดัมมี่จะได้รูปแบบสมการถดถอยที่เหมาะสมสำหรับข้อมูลชุดนี้คือ $Y = \beta_0 + \beta_1 X_1 + \beta_2 D + \beta_3 DX_1 + \varepsilon$ ค่า R^2_a มีค่าเป็น 0.374 สำหรับกรณีเพิ่มตัวแปรอิสระ X_2 จะได้รูปแบบสมการถดถอยที่เหมาะสมสำหรับข้อมูลชุดนี้คือ $Y = \beta_0 + \beta_1 X_1 + \beta_4 X_2 + \varepsilon$ ค่า R^2_a มีค่าเป็น 0.537 จะเห็นได้ว่า ในการเพิ่มตัวแปรอิสระ X_2 จะเหมาะสมกว่าในการใช้ตัวแปรดัมมี่สำหรับข้อมูลชุดนี้ ดังนั้น ในการประมาณคะแนนสอบปลายภาควิชาสถิติเบื้องต้นของนักศึกษาสาขาวิชาจุลชีววิทยา ปีการศึกษา 2557 สามารถใช้เพศของนักศึกษามาเป็นตัวแปรดัมมี่ได้ แต่การใช้เกรดเฉลี่ยสะสมของนักศึกษาเป็นตัวแปรอิสระอีกตัวหนึ่ง จะส่งผลให้การประมาณคะแนนสอบปลายภาคของนักศึกษาได้ผลที่ดีกว่า

เอกสารนี้สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.2.4 วิชาสถิติเบื้องต้นสาขาวิชาสัตวศาสตร์ ปีการศึกษา 2555

เมื่อพิจารณาตัวแปรอิสระ X_1 เพียงตัวเดียวในสมการถดถอย จะส่งผลให้ค่า R^2_a มีค่าเท่ากับ 0.329 และเมื่อพิจารณาตัวแปรอิสระ X_1 และตัวแปรดัมมี่ D และ/หรือ DX_1 เข้าสู่สมการถดถอย จะเห็นว่า ตัวแปรดัมมี่ไม่มีอิทธิพลต่อตัวแปรตาม Y แต่ถ้าพิจารณาตัวแปรอิสระ X_1 และ X_2 เข้าสู่สมการถดถอย จะทำให้ค่า R^2_a มีค่าเป็น 0.522 ดังนั้น รูปแบบสมการถดถอยที่เหมาะสมสำหรับข้อมูลชุดนี้คือ $Y = \beta_0 + \beta_1 X_1 + \beta_4 X_2 + \varepsilon$ ค่า R^2_a มีค่าเป็น 0.522 จะเห็นได้ว่า การใช้ตัวแปรดัมมี่สำหรับข้อมูลชุดนี้ไม่เหมาะสม ดังนั้น ในการประมาณคะแนนสอบปลายภาควิชาสถิติเบื้องต้นของนักศึกษาสาขาวิชาสัตวศาสตร์ ปีการศึกษา 2555 ไม่สามารถใช้เพศของนักศึกษามาเป็นตัวแปรดัมมี่ได้ การใช้เกรดเฉลี่ยสะสมของนักศึกษาเป็นตัวแปรอิสระอีกตัวหนึ่ง จะส่งผลให้การประมาณคะแนนสอบปลายภาคของนักศึกษาได้ผลที่ดีกว่า

จากการวิเคราะห์ข้อมูลคะแนนสอบกลางภาคและปลายภาคในวิชาสถิติเบื้อต้นของนักศึกษาสาขาวิชาสถิติประยุกต์ ปีการศึกษา 2555 – 2556 นักศึกษาสาขาวิชาจุลชีววิทยา ปีการศึกษา 2557 และนักศึกษาสาขาวิชาสัตวศาสตร์ ปีการศึกษา 2555 จะสรุปได้ว่า การใช้ตัวแปรอิสระเป็นคะแนนสอบกลางภาค และเกรดเฉลี่ยสะสมของนักศึกษา จะมีส่วนในการอธิบายความผันแปรของคะแนนสอบปลายภาคของนักศึกษาในวิชาสถิติเบื้องต้นได้ดีกว่าการใช้ตัวแปรอิสระเป็นคะแนนสอบกลางภาค และเพศเป็นตัวแปรดัมมี่ในการแบ่งกลุ่มข้อมูล

5.3 การถดถอยกับตัวแปรดัมมี่สำหรับข้อมูลสามกลุ่ม

ข้อมูลที่ใช้ในวิเคราะห์ประกอบด้วย ตัวแปรตามคือ คะแนนสอบปลายภาควิชาสถิติเบื้องต้นของนักศึกษา(Y) ส่วนตัวแปรอิสระ คือ คะแนนสอบกลางภาควิชาสถิติเบื้องต้น(X_1) และเกรดเฉลี่ยสะสมของนักศึกษา (X_2) และตัวแปรดัมมี่ คือ สาขาวิชาของนักศึกษา (D_1, D_2) สามารถสรุปผลได้ดังนี้

เมื่อพิจารณาตัวแปรอิสระ X_1 เพียงตัวเดียวในสมการถดถอย จะส่งผลให้ค่า R^2_a มีค่าเท่ากับ 0.354 และเมื่อพิจารณาตัวแปรอิสระ X_1 และตัวแปรดัมมี่ D_1, D_2 และ/หรือ D_1X_1 , D_2X_1 เข้าสู่สมการถดถอย จะเห็นว่า ตัวแปรดัมมี่ไม่มีอิทธิพลต่อตัวแปรตาม Y แต่ถ้าพิจารณาตัวแปรอิสระ X_1 และ X_2 เข้าสู่สมการถดถอย จะทำให้ค่า R^2_a มีค่าเป็น 0.504 ดังนั้น รูปแบบสมการถดถอยที่เหมาะสมสำหรับข้อมูลชุดนี้คือ $Y = \beta_0 + \beta_1 X_1 + \beta_6 X_2 + \varepsilon$ ค่า R^2_a มีค่าเป็น 0.504 จะเห็นได้ว่า การใช้ตัวแปรดัมมี่สำหรับข้อมูลชุดนี้ไม่เหมาะสม ดังนั้น การใช้ตัวแปรอิสระเป็นคะแนนสอบกลางภาค และเกรดเฉลี่ยสะสมของนักศึกษา จะ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

มีส่วนในการอธิบายความผันแปรของคะแนนสอบปลายภาคของนักศึกษาในวิชาสถิติเบื้องต้นได้ดีกว่าการใช้ตัวแปรอิสระเป็นคะแนนสอบกลางภาคและสาขาวิชาเป็นตัวแปรดัมมี่ในการแบ่งกลุ่มข้อมูล

5.4 ข้อเสนอแนะ

ตัวแปรดัมมี่เป็นตัวแปรที่ถูกสร้างขึ้นในการวิเคราะห์การถดถอย เพื่อใช้ในการแบ่งกลุ่มของข้อมูล ถ้าหากข้อมูลที่นำมาวิเคราะห์ไม่สามารถแบ่งกลุ่มข้อมูลอย่างเห็นได้ชัด ก็จะทำให้ตัวแปรดัมมี่นั้นไม่มีอิทธิพลต่อตัวแปรตามในรูปแบบการถดถอย ดังนั้น หากจะใช้ตัวแปรดัมมี่ให้เกิดประโยชน์ ก็ควรจะศึกษาข้อมูลที่นำมาใช้ในการวิเคราะห์ให้มั่นใจว่า ข้อมูลชุดนั้นๆ สามารถแบ่งข้อมูลออกเป็นกลุ่มได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บรรณานุกรม

กัลยา วานิชย์บัญชา. (2551). **การใช้ SPSS for Windows ในการวิเคราะห์ข้อมูล.** พิมพ์ครั้งที่ 11. กรุงเทพฯ: ธรรมสาร.

ชวลิต ทับสีรักษ์. (2555). **ตัวแปรเชิงคุณภาพกับการวิเคราะห์ถดถอย.** วารสารการวัดผลการศึกษา. 17(1): 31 – 42.

ทรงศิริ แต้สมบัติ. (2542). **การวิเคราะห์การถดถอย.** กรุงเทพฯ: สำนักพิมพ์มหาวิทยาลัยเกษตรศาสตร์.

พิชิต ฤทธิ์จรูญ. (2552). **หลักการวัดและประเมินผลการศึกษา.** พิมพ์ครั้งที่ 5. กรุงเทพฯ: เฮาส์ ออฟ เคอร์มิสจำกัด.

เยาวดี รางชัยกุล วิบูลย์ศรี. (2552). **การวัดผลและการสร้างแบบสอบผลสัมฤทธิ์.** พิมพ์ครั้งที่ 8. กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

ล้วน สายยศ และอังคณา สายยศ.(2543). **เทคนิคการวัดผลการเรียนรู้.** พิมพ์ครั้งที่ 2. กรุงเทพฯ: สุวีริยาสาส์น.

ศรีเพ็ญ ทรัพย์มนชัย. (2545). **การวิเคราะห์การถดถอยเชิงธุรกิจ.** กรุงเทพฯ: ปิ่นเกล้าการพิมพ์.

สุพล ดุรงค์วัฒนา. (2537). **การพยากรณ์ทางธุรกิจ.** กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

Draper, N. R. and Smith, H. (1998). **Applied Regression Analysis.** (3rd ed). New York: John Wiley & Sons, Inc.

Montgomery, D.C. & Peck, E.A. (1992). **Introduction to linear regression analysis.** (2nd ed). New York: John Wiley & Sons, Inc.

Seber, A. F. & Lee, A. J. (2003). **Linear regression analysis.** (2nd ed). New Jersey: John Wiley & Sons, Inc.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้