

ชื่อเรื่องการค้นคว้าแบบอิสระเชิงวิทยานิพนธ์ การศึกษาการวิเคราะห์ข้อมูลเชิงกลุ่ม

ชื่อผู้เขียน

นายสุทธิเดช ตันท์พาณิชย์

วิทยาสาสตรมหาบัณทิต

สาขาวิชาสถิติประยุกต์

คณะกรรมการสอบการค้นคว้าแบบอิสระเชิงวิทยานิพนธ์

อาจารย์	พุดิพงษ์ พุกกะมาน	ประธานกรรมการ
อาจารย์ ดร.	ชูเพ็ญศรี วงศ์พุทธา	กรรมการ
ผู้ช่วยศาสตราจารย์	อังกาบ บุญย้อย	กรรมการ

บทคัดย่อ

งานการค้นคว้าแบบอิสระเชิงวิทยานิพนธ์นี้มีวัตถุประสงค์เพื่อ ศึกษาเทคนิคการวิเคราะห์ข้อมูลเชิงกลุ่ม โดยการสร้างตัวแบบเชิงสถิติภายใต้วัตถุประสงค์เพื่อทำการอนุมานเกี่ยวกับเซตของพารามิเตอร์ที่ใช้ในการอธิบาย และตรวจสอบโครงสร้างความสัมพันธ์ของตัวแปรต่างๆ โดยตัวแปรจะมีลักษณะเชิงกลุ่ม ข้อมูลจึงเกี่ยวข้องกับเซตของจำนวนนับที่สังเกตได้เป็นสำคัญ เช่น เพศ ระดับการศึกษา ระดับความคิดเห็น และ อาชีพ เป็นต้น หลักการสำคัญในการสร้างตัวแบบเชิงสถิติที่ใช้ในการวิเคราะห์ข้อมูลเชิงกลุ่มต้องคำนึงถึงลักษณะของข้อมูล หรือตัวแปรที่สนใจ โดยมีตัวแบบต่าง ๆ ดังต่อไปนี้ คือ

1. ตัวแบบโลจิท ตัวแบบโพรบิท และตัวแบบคอมพลีเมนต์ลอจิท-ลอจิท จะเป็นตัวแบบที่มีลักษณะข้อมูลโดยทั่วไปเหมือนกัน คือ มีตัวแปรตามแยกเป็น 2 กลุ่ม ส่วนตัวแปรอิสระจะเป็นตัวแปรต่อเนื่อง หรือเชิงกลุ่มก็ได้ แต่มีข้อแตกต่างกัน คือ ตัวแบบโพรบิท และตัวแบบคอมพลีเมนต์ลอจิท-ลอจิทเป็นตัวแบบที่มีเทอมหลักคล้ายกับตัวแบบโลจิทแต่ตัวแบบคอมพลีเมนต์ลอจิท-ลอจิทเหมาะสมกับข้อมูลที่มีลักษณะฟังก์ชันความหนาแน่นสะสมที่มีการถู่เข้า และถู่ออกไม่สม่ำเสมอและไม่สมมาตรที่ค่าความน่าจะเป็น 0.5

2. ตัวแบบการถดถอยโลจิสติก เป็นตัวแบบที่ประยุกต์มาจากตัวแบบการวิเคราะห์การถดถอยโดยมีข้อแตกต่างกัน คือ ตัวแปรตามของตัวแบบการถดถอยโลจิสติก เป็นตัวแปรซึ่งจัดเป็น 2 กลุ่ม ส่วนตัวแปรตามของการวิเคราะห์การถดถอยจะเป็นตัวแปรต่อเนื่อง สำหรับตัวแปรอิสระของทั้งสองตัวแบบข้างต้นจะเป็นตัวแปรแบบต่อเนื่อง หรือตัวแปรเชิงกลุ่มก็ได้ โดยในการวิเคราะห์การถดถอยที่มีตัวแปรตามแบบต่อเนื่องจะเรียกว่า ตัวแปรหุ่น (Dummy Variable)

3. ตัวแบบลอกลินีเยอร์ เป็นตัวแบบที่มีตัวแปรอิสระเป็นตัวแปรเชิงกลุ่ม และตัวแปรตามเป็นจำนวนนับในแต่ละช่องในตารางหลายทาง ตัวแบบดังกล่าวสามารถวิเคราะห์ความสัมพันธ์ของตัวแปรครั้งละหลายตัวแปร และสามารถอธิบายลักษณะความสัมพันธ์ระหว่างตัวแปรเชิงกลุ่มได้

4. ตัวแบบพหุนาม เป็นตัวแบบที่ให้ความสำคัญกับตารางการณัศจรรย์หลายทาง โดยกำหนดให้ตัวแปรตัวใดตัวหนึ่งเป็นตัวแปรตาม ส่วนตัวแปรอื่นๆที่เหลือเป็นตัวแปรอิสระ โดยที่ตัวแปรอิสระอาจเป็นตัวแปรเชิงกลุ่ม หรือแบบผสมก็ได้ และกำหนดให้ความถี่ของกลุ่มในระดับต่าง ๆ ที่เป็นตัวแปรอิสระนั้นมีค่าคงที่

5. ตัวแบบความสัมพันธ์แบบ Linear-by-Linear ตัวแบบ Row Effect และตัวแบบ Column Effect เป็นตัวแบบที่ใช้ในการวิเคราะห์ตัวแปรเชิงกลุ่มแบบมีลำดับ ซึ่งเป็นการศึกษาในกรณีที่ผู้ทำการวิจัยให้ความสนใจกับลำดับของตัวแปรเชิงกลุ่มในตารางการณัศจรรย์ โดยระดับของตัวแปรนั้นถูกจัดเป็นกลุ่มแบบมีลำดับและมีการจัดเรียงกันตามธรรมชาติ ใช้การวิเคราะห์โดยให้คะแนน (Score) แก่กลุ่มต่าง ๆ โดยคำนึงถึงลำดับก่อนหลัง แล้วอาจใช้การวิเคราะห์การถดถอยมาวิเคราะห์ข้อมูลในตารางการณัศจรรย์ที่มีตัวแปรเชิงกลุ่มเป็นแบบมีลำดับ

Research Title Study of Categorical Data Analysis

Author Mr. Suttiedeck Tanparnit

M.S. Applied Statistics

Examining Committee

Lecturer Putipong	Bookkamana	Chairman
Lecturer Dr. Chupensri	Wongbuddha	Member
Asst. Prof. Angkarb	Boonyoi	Member

Abstract

This independent study is centered on the analysis of categorical data and model building to infer the set of the studied parameters. The categorical data analysis has been used to investigate the relationship of the variables which mostly were category or countable variable such as sex, educational background, attitude and occupation. The most important principal of categorical data analysis are to build the statistical models and select the appropriate model to satisfy the data or the variable. Several statistical models which were demonstrated in this study including:

1. Logit Models, Probit Model and Complementary Log-Log Model. These models have the same general data that is the dependent data will provide to be 2 groups and the independent variable can be continuous variable or categorical variable. Although these models will have the same general data, it also has some different as well. The Probit Model and Complementary Log-Log Model have a model which has a main term like the model of Logit Model but Complementary Log-Log Model is suitable for the data which has cumulative density function which convergent and divergent are irregular and asymmetric about the point where probability equal 0.5.

2. The Logistic Regression Model is adapted from Regression Analysis Model but it has some different points between each other. The dependent variable of the Logistic Regression Model is set for 2 groups but the dependent variable of Regression Analysis Model is continuous variable. However, the independent variable all of this can be continuous variable or categorical variable. The continuous dependent variable of Regression Analysis is called Dummy Variable.

3. Loglinear Models are the models that is independent variable categorical variable and the dependent variables are the countable variables in each cell in the multiple contingency table. The models can analyze the relationship for each model in the same time and explain the relationship between the categorical variables.

4. Multinomial Models gave the attention to the multiple contingency table. It will set some variable to be dependent variable and another variables will be the independent variable. The independent variable will be categorical variable or mixed variable and sets the frequency of the group in each level as the constant.

5. Linear-by-Linear Model, Row Effect Model and Column Effect Model. These Models are analyzed the categorical ordinal variable which is studied in case the author (researcher) is interested the order of categorical variable in contingency table. The level of variable is set to be a group of order and naturally. It is analyzed by using the score of each group orderly and will use Regression analysis to analyze the data in the contingency table which the categorical variables are ordinal variables.