บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในปัจจุบันการนำความรู้ทางด้านสถิติไปประยุกต์ใช้ในสาขาต่าง ๆ มีมากขึ้นอันเนื่องมาจากวิธีการทางสถิติมีวิธีดำเนินการอย่างมีระบบภายใต้เหตุและผลทั้งนี้การประยุกต์ใช้จำเป็นต้องอาศัยวิธีการทางสถิติที่เหมาะสมกับสาขานั้น ๆ เพื่อให้การวิจัยมีความถูกต้องสมบูรณ์มากที่สุด แต่สิ่งที่สำคัญที่สุดอีกประการหนึ่งที่จะทำให้การวิเคราะห์ทางสถิติบรรลุตามวัตถุประสงค์คือลักษณะของข้อมูลที่นำมาใช้ในการวิเคราะห์ ซึ่งอาจจะเป็นข้อมูลเชิงปริมาณ (Quantitative data) หรือข้อมูลเชิงคุณภาพ(Qualitative data) ก็ได้ โดยทั่วไปการวิเคราะห์เชิงสถิติได้รับการพัฒนา เพื่อให้สอดคล้องกับข้อมูลเชิงปริมาณ แต่มีบ่อยครั้งที่ในทางปฏิบัติการเก็บรวบรวมข้อมูลมักจะมีข้อมูลเชิงคุณภาพเข้ามาเกี่ยวข้องด้วย เช่น เพศ (ชาย, หญิง) การตัดสินใจในการซื้อสินค้า (ซื้อ, ไม่ซื้อ) อาการเกิดมะเร็งที่ปอด (เกิด, ไม่เกิด) เป็นต้น ข้อมูลเชิงคุณภาพในลักษณะเช่นนี้จะมีค่าที่สามารถวัดได้เพียง 2 ค่า หรือสนใจศึกษาเพียง 2 ค่าเท่านั้น ข้อมูลเหล่านี้พบได้มากในงานวิจัยทางด้านสังคมศาสตร์ เศรษฐศาสตร์และการแพทย์ เป็นต้น ถ้าข้อมูลที่จะสนใจจะศึกษา(ตัวแปรตาม) มีลักษณะดังกล่าว จะเรียกตัวแปรตามชนิดนี้ว่า " Dichotomous dependent Variable " ซึ่งเป็นตัวแปรที่ต้องการพยากรณ์ โดยมีตัวแปรอิสระที่มีความสัมพันธ์กับตัวแปรตามมาช่วยในการพยากรณ์

การวิเคราะห์การถดถอยโดยวิธีกำลังสองน้อยที่สุด เป็นวิธีการหนึ่งที่ผู้วิจัยนิยมใช้เพื่อเป็นเครื่องมือในการพยากรณ์ หรือ คาดคะเนเหตุการณ์ล่วงหน้า เนื่องจากเป็นวิธีที่สะดวกและง่ายต่อการแปลความหมาย แต่การนำมาประยุกต์ใช้กับตัวแปรตามสองด้านนี้ทำให้เกิดข้อขัดแย้งกับข้อตกลงเบื้องต้นของการวิเคราะห์การถดถอยและเกิดปัญหาทางด้านความเหมาะสมในการพยากรณ์ดังต่อไปนี้

1.) ปัญหา Heteroscedasticity เนื่องจาก $V(e_i)$ ไม่คงที่โดยผันแปรค่าไปตามค่าของตัวแปรอิสระ นั่นคือ $V(e_i) = \sigma^2 V \neq \sigma^2 I$

2.) ปัญหาในการประมาณค่าพารามิเตอร์โดยวิธีกำลังสองน้อยที่สุดแบบธรรมดา (OLS) ทำให้ β_j ที่ได้ขาดประสิทธิภาพ และ $V(\beta_j)$ จะสูงเกินไปขาดคุณสมบัติของ

BLUE อันจะมีผลทำให้ช่วงความเชื่อมั่นของ β_j และช่วงการพยากรณ์ $E(y_i)$ กว้างเกินไป รวมทั้งมีผลทำให้ F-test, T-test ที่ได้มีค่าต่ำกว่าที่ควรจะเป็น ดังนั้น การทดสอบสมมติฐาน $Ho : \beta_1 = \beta_2 = \ldots = \beta_p = 0$ หรือ $Ho : \beta_j = 0$ มีโอกาสที่จะยอมรับมีมากขึ้น ทำให้ตัวแปรอิสระบางตัวถูกกัดทิ้งไปทั้งที่ตัวแปรอิสระตัวนั้น อาจมีความสำคัญเป็นอย่างยิ่งก็ได้

3.) ปัญหาในการพยากรณ์ค่าตัวแปรตาม(y) คือ การพยากรณ์ไม่สามารถระบุได้ว่าค่าที่แท้จริงของ y คือ 0 หรือ 1 แบ่งเป็น 2 กรณี ดังนี้

3.1 ถ้าค่าพยากรณ์อยู่นอกขอบเขตที่แท้จริงของ $y \in [0,1]$ ทำให้ค่าที่ได้ผิดจากความเป็นจริง และระบุค่าที่แท้จริงไม่ได้

3.2 ถ้าค่าพยากรณ์อยู่ในช่วง [0,1] การระบุค่าที่แท้จริงของ y จะกระทำไม่ได้และถ้าแปลผลในรูปของร้อยละจะทำให้ขัดแย้งกับข้อ 3.1 ที่มีค่าติดลบหรือมากกว่า 100% ซึ่งไม่สอดคล้องกันในขณะที่ y มาจากวิธีการเดียวกัน

4.) ปัญหาการแจกแจงของตัวแปรตาม และการแจกแจงของความผิดพลาดที่มีการแจกแจงแบบเบอร์นูลลีและแบบทวินามนั้น จะประมาณด้วยการแจกแจงแบบปกติเมื่อเพิ่มขนาดตัวอย่างให้มากขึ้นทำให้ลดปัญหาในข้อนี้ได้ และจะทำให้ได้ข้อสมมติของความผิดพลาด (ERROR) ตามข้อตกลงเดิม คือ มีการแจกแจงแบบปกติ

เพื่อแก้ไขปัญหาดังกล่าวข้างต้น การวิเคราะห์การถดถอยจึงได้รับการพัฒนาขึ้นมาอีกชนิดหนึ่งซึ่งเรียกว่า"การวิเคราะห์การถดถอยทวิ(Binary Regression Analysis)"

เนื่องจากการวิเคราะห์ถดถอยทวินี้ ตัวแปรตามที่ใช้ในการวิเคราะห์มีเพียง 2 ค่า ด้วยเหตุนี้จึงสามารถจำแนกกลุ่มของตัวแปรตามออกได้เป็น 2 กลุ่ม คือ กลุ่มที่ตัวแปรตามมีค่าเป็น 1 และ กลุ่มที่ตัวแปรตามมีค่าเป็น 0 จากแนวคิดในการจำแนกกลุ่มนี้ ทำให้นักวิจัยได้นำเอาไปประยุกต์ใช้ เพื่อเปรียบเทียบกับการวิเคราะห์จำแนกประเภทในรูปแบบต่าง ๆ กัน จะเห็นได้จากงานวิจัยที่เกี่ยวข้อง ที่นำมาอ้างอิงในสาขาต่าง ๆ เช่น สาขาการแพทย์ (Cupples L.D et.al., 1984 : EFRON, B. 1975)สาขาเศรษฐศาสตร์ (Zellner, A. and Lee, T.H. 1965) เป็นต้น

จากกรณีศึกษาที่อ้างถึงข้างต้น ผู้วิจัยต่างก็มีข้อสมมติเกี่ยวกับประชากรของตัวแปรอิสระว่ามีการแจกแจงแบบปกติ เพื่อให้ข้อมูลมีลักษณะเป็นธรรมชาติสะดวกต่อการอ้างอิงในทางทฤษฎีของการแจกแจงนี้ แต่ในทางปฏิบัติเราไม่อาจควบคุมค่าของตัวแปรอิสระให้มีการแจกแจงแบบปกติตามข้อสมมติ เช่น ข้อมูลทางด้านการแพทย์ ทางด้านเศรษฐศาสตร์ ทางด้านวิศวกรรมศาสตร์ เป็นต้น ในทางทฤษฎีสามารถพิสูจน์ได้ว่าข้อมูลดังกล่าวเหล่านี้มี

การแจกแจงแบบเบ้ โดยมีนักสถิติหลายท่านให้การพิสูจน์เพื่อสนับสนุนคำกล่าวข้างต้น ดังเช่น Elandt-Johnson R., EFRON B., Weibull เป็นต้น ดังนั้น จึงเป็นสิ่งที่น่าสนใจศึกษาต่อไปว่า การวิเคราะห์ทางสถิติโดยใช้การวิเคราะห์การถดถอยทวิ (Binary Regression) และ การวิเคราะห์จำแนกประเภท (Discriminant Analysis) ในการจำแนกกลุ่มโดยตัวแปรอิสระมีการแจกแจงแบบเบ้ เราควรจะเลือกใช้วิธีการใดจึงจะเหมาะสมมากที่สุดกับลักษณะของข้อมูล ข้อขัดแย้งและปัญหาที่เราประสบอยู่

1.2 วัตถุประสงค์ของการวิจัย

1.) เพื่อศึกษาเปรียบเทียบประสิทธิภาพในการจำแนกกลุ่ม กรณีมี 2 กลุ่มประชากรโดยศึกษาเปรียบเทียบระหว่าง วิธีการต่อไปนี้

1.1 การวิเคราะห์การถดถอย

1.2 การวิเคราะห์จำแนกประเภท

2.) เพื่อศึกษาเปรียบเทียบประสิทธิภาพในการประมาณค่าสำหรับการวิเคราะห์โดยใช้การถดถอยโดยใช้ตัวประมาณค่าต่อไปนี้

2.1 Ordinary Least Square estimator

2.2 Estimated Gerneralized Least Square estimator

(กรณีที่มีการแปลงค่าพยากรณ์ให้อยู่ในช่วง[0,1] ด้วยเส้นโค้งของการแจกแจงต่าง ๆ กัน)

1.3 สมมติฐานของการวิจัย

1.) ประสิทธิภาพในการจำแนกกลุ่ม โดยการวิเคราะห์จำแนกประเภทจะดีกว่าการวิเคราะห์การถดถอยทั้ง 2 วิธี

2.) ประสิทธิภาพในการประมาณค่าสัมประสิทธิ์การถดถอยพหุโดยการวิเคราะห์การถดถอยทวิจะดีกว่าการวิเคราะห์การถดถอยโดยวิธีกำลังสองน้อยที่สุดแบบธรรมดา

1.4 ขอบเขตของการวิจัย

1.) ลักษณะข้อมูล

1.1 ตัวแปรตาม (y) เป็นข้อมูลเชิงคุณภาพที่มี 2 ค่า คือ 1 และ 0 โดยในแต่ละขนาดตัวอย่างจะกำหนดสัดส่วนระหว่าง 1 และ 0 จำนวน 10 สัดส่วน ดังนี้ 0.50:0.50, 0.55:0.45, ... , 0.95:0.05

1.2 ตัวแปรอิสระ (x) เป็นข้อมูลเชิงปริมาณที่มีการแจกแจงแบบเบ้

ซึ่งเป็นการแจกแจงที่นักวิจัยนำไปประยุกต์ใช้ในสาขาวิชาต่าง ๆ และพบได้ในทางปฏิบัติเป็นส่วนมาก นอกจากนี้ยังได้นำการแจกแจงแบบปกติมาตรฐาน มาเพื่อเปรียบเทียบให้เห็นความชัดเจนของวิธีการที่ใช้จำแนกกลุ่มมากยิ่งขึ้น สำหรับการแจกแจงที่ใช้ศึกษามีดังต่อไปนี้

ก. <u>การแจกแจงแบบไวบูลล์</u> (Weibull Distribution)

ฟังก์ชันความหนาแน่นอยู่ในรูปของ

$$f(x) = \begin{cases} \dfrac{\alpha X^{\alpha} \exp(X/\beta)^{\alpha}}{\beta^{\alpha}} & , X>0,\ \alpha>0,\ \beta>0 \\ 0 & \text{อื่น ๆ} \end{cases}$$

เมื่อ β เป็น scale parameter

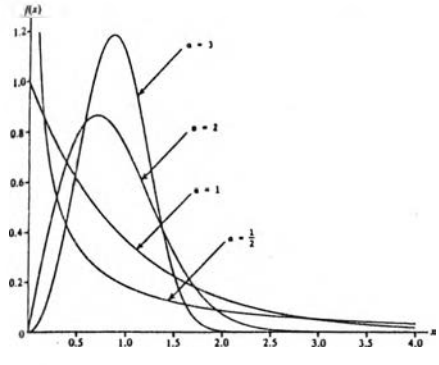
α เป็น shape parameter

$$E(x) = \beta\, \Gamma[1+1/\alpha]$$

$$V(x) = \beta^{2}[\Gamma(1+2/\alpha) - \Gamma^{2}(1+1/\alpha)]$$

$$CV(x) = \left[\frac{\Gamma(1+2/\alpha)}{\Gamma^{2}(1+1/\alpha)} - 1 \right]^{1/2}$$

ในการวิจัยครั้งนี้ จะศึกษาที่ $\beta = 1$, $\alpha = 1$



Weibull(α,1) density functions.

ข. <u>การแจกแจงแบบแกมมา</u> (Gamma Distribution)

ฟังก์ชันความหนาแน่นอยู่ในรูปของ

$$f(x) = \begin{cases} \dfrac{X^{\alpha-1}\exp(-X/\beta)}{\beta^{\alpha}\,\Gamma(\alpha)} & ,\quad X>0,\ \alpha>0,\ \beta>0 \\ 0 & \text{อื่น ๆ} \end{cases}$$

เมื่อ β เป็น scale parameter

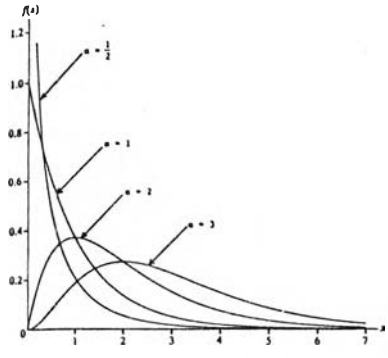
α เป็น shape parameter

$E(x) = \beta\alpha$

$V(x) = \beta^2\alpha$

$CV(x) = 1/\sqrt{\alpha}$

ในการวิจัยครั้งนี้ จะศึกษา $\beta = 1$, $\alpha = 1$



gamma(α,1) density functions.

ค. <u>การแจกแจงแบบลอกนอร์มอล</u> (Lognormal Distribution)
ฟังก์ชันความหนาแน่นอยู่ในรูป

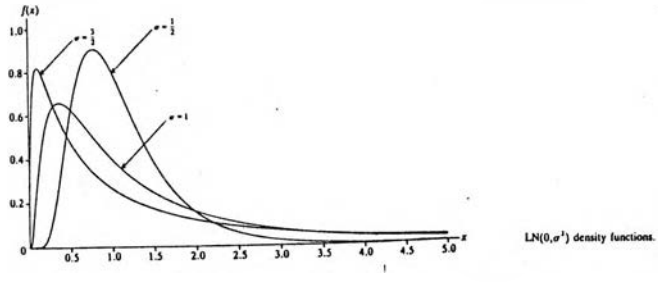
$$f(x) = \begin{cases} \dfrac{1}{x\delta\sqrt{2\pi}} \exp\left[-\dfrac{1}{2}(\log_e x-\mu)/\delta^2\right] & ., X>0, -\alpha<\mu<\alpha \\ 0 & \text{อื่น ๆ} \end{cases}$$

$E(x) = \exp(\mu+\delta^2/2)$

$V(x) = \exp(2\mu\rho+\delta^2)\ (\exp(\delta^2) - 1)$

$CV(x) = \sqrt{\exp(\delta^2)-1}$

เมื่อ μ และ δ^2 เป็นค่าเฉลี่ยและความแปรปรวนของ Y
โดยที่ $Y = \log_e(x)$ และ Y มีการแจกแจงปกติ



ในการวิจัยครั้งนี้ จะศึกษากรณี $\mu = 0, \sigma^2 = 1$

2.) จำนวนตัวแปรอิสระและขนาดตัวอย่าง

2.1 จำนวนตัวแปรอิสระ 1 ตัวแปร และ 2 ตัวแปร โดยถ้าใช้ตัวแปรอิสระ 2 ตัวแปร จะถือว่าตัวแปรอิสระเหล่านี้ผ่านการคัดเลือก และทดสอบความเป็นอิสระซึ่งกันและกันแล้ว พร้อมที่จะนำมาช่วยในการจำแนกกลุ่ม

2.2 จำนวนตัวอย่าง (n) เท่ากับ 10, 30 และ 60

1.5 ประโยชน์ที่คาดว่าจะได้รับ

จากขอบเขตการวิจัยที่กำหนดสามารถใช้ ผลการศึกษาเปรียบเทียบที่ได้ เพื่อใช้เป็นแนวทางในการศึกษาให้สอดคล้องกับข้อมูลที่มีอยู่ ดังนี้

1.) เป็นแนวทางในการตัดสินใจ ว่าควรจะใช้วิธีการใดในการจำแนกกลุ่ม จึงจะทำให้ประสิทธิภาพของการจำแนกกลุ่มมีค่ามากที่สุด

2.) เป็นแนวทางในการตัดสินใจ ว่าควรจะใช้วิธีการใดในการประมาณค่าสัมประสิทธิ์การถดถอยพหุ จึงจะทำให้ผลการประมาณค่ามีความผิดพลาดน้อยที่สุด

3.) เป็นแนวทางในการศึกษาเพื่อเปรียบเทียบกับวิธีการทางสถิติอื่น ๆ ที่เกี่ยวข้องต่อไป