หอสมุดกลาง สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ปัจจุบันนี้การวิจัยในสาขาต่าง ๆ โดยเฉพาะวิทยาศาสตร์และสังคมศาสตร์ จำเป็นต้องอาศัยความรู้ทางด้านสถิติซึ่งเป็นวิทยาศาสตร์ประยุกต์และมีวิธีดำเนินการที่เป็นระบบ การวิจัยในสาขาวิชาต่าง ๆ นั้น ส่วนใหญ่แล้วต้องใช้วิธีการทางสถิติช่วยในการวิเคราะห์เพื่อหาคำตอบสำหรับการวิจัยนั้น ๆ โดยเฉพาะอย่างยิ่งการหาคำตอบเพื่อคาดคะเนเหตุการณ์ล่วงหน้าหรือการพยากรณ์ ซึ่งผู้วิจัยส่วนใหญ่จะใช้วิธีการวิเคราะห์ความถดถอย(Regression Analysis)

วิธีการวิเคราะห์ความถดถอยที่นักวิจัยส่วนใหญ่นิยมใช้กันอย่างแพร่หลายคือ การวิเคราะห์ความถดถอยเชิงเส้น (Linear Regression Analysis) ซึ่งมีรูปแบบดังนี้คือ

$$\underset{\sim}{Y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$$

เมื่อ $\underset{\sim}{Y}$ เป็นเวคเตอร์ของตัวแปรตามขนาด (nxl)

X เป็นเมตริกซ์ของตัวแปรอิสระขนาด (nxp)

$\underset{\sim}{\beta}$ เป็นเวคเตอร์ของค่าสัมประสิทธิ์ถดถอยเชิงเส้นขนาด (pxl)

$\underset{\sim}{\varepsilon}$ เป็นเวคเตอร์ของความคลาดเคลื่อนขนาด (nxl)

n เป็นจำนวนข้อมูล

p เป็นจำนวนสัมประสิทธิ์ถดถอยเชิงเส้น

ตามปกติเมื่อผู้วิจัยนำข้อมูลมาวิเคราะห์ความถดถอยเชิงเส้น ภายหลังจากที่ได้ตัวแบบถดถอยเชิงเส้น(linear regression model) เพื่อนำไปใช้ในการพยากรณ์ ผู้วิจัยส่วนใหญ่มิได้ทำการตรวจสอบความเหมาะสมของตัวแบบถดถอยเชิงเส้นก่อนที่จะนำไปใช้ในการพยากรณ์จริง ๆ ในทางปฏิบัติบางครั้งผู้วิจัยอาจมองข้ามตัวแปรที่สำคัญ ๆ บางตัว หรือข้อมูลที่ถูกนำมาวิเคราะห์ไม่ดีเท่าที่ควร สาเหตุเหล่านี้อาจก่อให้เกิดข้อผิดพลาดในการนำตัวแบบถดถอยเชิงเส้น

ที่ได้ไปใช้ในการพยากรณ์ ซึ่งเป็นข้อบกพร่องของผู้วิจัยส่วนใหญ่ ดังนั้นก่อนที่จะนำตัวแบบถดถอยเชิงเส้นไปใช้ในการพยากรณ์ ผู้วิจัยควรทำการตรวจสอบความเหมาะสมเสียก่อน จากการศึกษาของเบิร์ค (1984: 331-338) พบว่า วิธีการตรวจสอบความเหมาะสมของตัวแบบที่ดีวิธีหนึ่งก็คือ การตรวจสอบความเหมาะสมของตัวแบบถดถอยด้วยข้อมูลใหม่ (Validating Regression Model with New Data) วิธีนี้เป็นวิธีการสร้างตัวแบบถดถอยด้วยข้อมูลที่มีอยู่ในอดีต จากตัวแบบถดถอยจะทำให้ได้ค่าประมาณ ($\hat{Y}$) ซึ่งจะถูกนำมาเปรียบเทียบกับข้อมูลที่เก็บขึ้นใหม่ (new data) เพื่อพิจารณาความแตกต่างที่เกิดขึ้นว่ามีมากน้อยเพียงใด แต่วิธีนี้มีข้อเสียที่ว่าคุณภาพในการตรวจสอบความเหมาะสมของตัวแบบขึ้นอยู่กับความสัมพันธ์ของข้อมูลเก่าและข้อมูลใหม่ (Kenneth 1984:331-338) นอกจากนี้ในทางปฏิบัติ ผู้วิจัยอาจไม่มีข้อมูลใหม่มาใช้ในการวิเคราะห์ ดังนั้นในสถานการณ์เช่นนี้ผู้วิจัยสามารถแก้ปัญหาได้โดยทำการแบ่งข้อมูล (splitting data) เป็น 2 ชุดคือ

1. ข้อมูลที่ใช้ในการประมาณค่า (estimation data) ข้อมูลชุดนี้จะถูกใช้ในการประมาณค่าสัมประสิทธิ์ถดถอยเชิงเส้น

2. ข้อมูลที่ใช้ในการพยากรณ์ (prediction data) จะใช้ข้อมูลชุดนี้ในการเปรียบเทียบกับค่าประมาณที่เกิดจากข้อมูลที่ใช้ในการประมาณค่า

วิธีการแบ่งข้อมูลที่นิยมใช้กันอย่างแพร่หลายก็คือ การแบ่งข้อมูลแบบครึ่งหนึ่ง (Ronald 1977 : 415-428) (half-and-half splitting data) การแบ่งข้อมูลในลักษณะนี้เป็นวิธีที่ง่ายและสะดวก แต่ก็มีข้อเสียที่ว่า ไม่มีหลักเกณฑ์ที่แน่นอนในการแบ่ง ซึ่งอาจเป็นเหตุที่ทำให้เกิดข้อผิดพลาดในการตรวจสอบความเหมาะสมของตัวแบบ

การแบ่งข้อมูลอีกวิธีหนึ่งที่นิยมใช้กันมากก็คือ ใช้ระยะเวลาเป็นเกณฑ์ในการแบ่งข้อมูล ผู้ที่ทำการศึกษาเรื่องนี้คือ เคดี้และแอลเลน (Cady and Allen 1972 : 211-214) และมาร์ควอดต์และสนี (Marquardt and Snee 1975 : 3-19) โดยใช้ข้อมูลที่เก็บในระยะแรก ๆ เป็นข้อมูลที่ใช้ในการประมาณค่า และใช้ข้อมูลที่เก็บในระยะต่อมาเป็นข้อมูลที่ใช้ในการพยากรณ์ ตามปกติผู้วิจัยจะตัดสินใจเองว่าจะใช้ระยะเวลาที่จุดใดแบ่งข้อมูล ซึ่งส่วนใหญ่จะแบ่งให้จำนวนข้อมูลที่ใช้ในการประมาณค่ามีมากกว่าจำนวนข้อมูลที่ใช้ในการพยากรณ์ แต่ก็มีผู้ไม่เห็นด้วยกับการใช้ระยะเวลาเป็นเกณฑ์ในการแบ่งข้อมูลคือ แดรปเปอร์และสมิท (Draper and Smith) ได้

แย้งในเรื่องนี้ว่า ตัวแบบที่ได้จากข้อมูลที่เก็บในระยะเวลาหนึ่ง อาจจะนำไปใช้ในการพยากรณ์ในอีกระยะเวลาหนึ่งได้ ไม่ดีเท่าที่ควร (Draper 1966) เนื่องจากข้อมูลที่เก็บในระยะเวลาที่ต่างกันอาจมีลักษณะหรือโครงสร้างของความสัมพันธ์ (coorrelation structure) ที่แตกต่างกัน ดังนั้นการใช้ระยะเวลาเป็นเกณฑ์ในการแบ่งข้อมูลผู้วิจัยควรต้องคำนึงถึงปัญหาที่อาจเกิดขึ้น

จากปัญหาที่กล่าวมาแล้วข้างต้นทำให้เกิดแนวความคิดที่ว่าวิธีการแบ่งข้อมูลควรต้องมีหลักเกณฑ์ที่ดี ผู้ที่ศึกษาเรื่องนี้คือ สนี (Snee 1977 : 415-428) ได้อธิบายถึงวิธีดูเพล็กซ์ (DUPLEX Algorithm) ในการแบ่งข้อมูล ซึ่งวิธีนี้ได้พัฒนามาจากวิธีคาเด็กซ์ (CADEX Algorithm) ของ เคนนาร์ดและสโตน (Kennard and Stone 1969 : 137-148) จุดประสงค์ของการแบ่งข้อมูลด้วยวิธีดูเพล็กซ์คือ เพื่อทำให้ตัวแปรอิสระของข้อมูลที่ใช้ในการประมาณค่าและของข้อมูลที่ใช้ในการพยากรณ์มีลักษณะคล้ายกับข้อมูลเดิมก่อนที่จะทำการแบ่งให้มากที่สุด ซึ่งหลักเกณฑ์ในการพิจารณาตัวแปรอิสระของข้อมูลทั้ง 2 ชุด กำหนดจาก

$$(|X'X|\ est/|\ X'X\ |pre)^{1/p}$$

เมื่อ $|X'X|$ est เป็นดีเทอร์มิแนนท์ของเมตริกซ์ $X'X$ ของข้อมูลที่ใช้ในการประมาณค่า

$|X'X|$ pre เป็นดีเทอร์มิแนนท์ของเมตริกซ์ $X'X$ ของข้อมูลที่ใช้ในการพยากรณ์

ผลลัพธ์ของดีเทอร์มิแนนท์จากสูตรข้างต้นจะเข้าใกล้ 1 เมื่อตัวแปรอิสระจากข้อมูลทั้ง 2 ชุด มีลักษณะใกล้เคียงกัน แต่ถ้าหากค่าผลลัพธ์ของดีเทอร์มิแนนท์ที่ได้มีค่าต่างไปจาก 1 มาก ก็แสดงว่าตัวแปรอิสระจากข้อมูลทั้ง 2 ชุดมีลักษณะที่แตกต่างกัน

จากคุณสมบัติทางสถิตินี้ทำให้เกิดแนวความคิดเกี่ยวกับการตรวจสอบความเหมาะสมของตัวแบบถดถอยเชิงเส้นซึ่งพิจารณาที่ความเหมาะสมของข้อมูลคือ เมื่อนำข้อมูลที่ใช้ในการประมาณค่าและข้อมูลที่ใช้ในการพยากรณ์มาหาค่าสัมประสิทธิ์ถดถอยเชิงเส้น ๆ ที่ได้จากทั้ง 2 สมการน่าจะมีค่าที่ใกล้เคียงกัน กล่าวคือมีความคงที่ในค่าสัมประสิทธิ์ถดถอยเชิงเส้น แต่ในทางปฏิบัติอาจเป็นไปได้ที่ค่าสัมประสิทธิ์ถดถอยเชิงเส้นจากทั้ง 2 สมการมีความแตกต่างกัน ซึ่งอาจมีสาเหตุมาจากข้อบกพร่องในการเก็บข้อมูลที่มีผลทำให้ข้อมูลมีความแตกต่างกัน เช่น การเก็บข้อมูลในระยะเวลาที่ต่างกัน การเก็บข้อมูลจากนอกขอบเขตประชากรที่ต้องการ เป็นต้น ถ้าภายหลัง

จากการตรวจสอบพบว่า สัมประสิทธิ์ถดถอยเชิงเส้นจากทั้ง 2 สมการมีความแตกต่างกัน ผู้วิจัยก็ไม่ควรนำข้อมูลทั้งหมดมาประมาณค่าสัมประสิทธิ์ถดถอยเชิงเส้นเพื่อนำไปใช้พยากรณ์ด้วยความเสี่ยงสูง ดังนั้นการตรวจสอบความเหมาะสมของตัวแบบถดถอยเชิงเส้นในเรื่องของข้อมูลจึงเป็นสิ่งที่สมควรกระทำ โดยความเหมาะสมในที่นี้จะพิจารณาจากความคงที่ของสัมประสิทธิ์ถดถอยเชิงเส้น เบิร์คได้แนะนำถึงการทดสอบความคงที่ของสัมประสิทธิ์ถดถอยเชิงเส้นโดยใช้การทดสอบเอฟ (F-test) ซึ่งคิดขึ้นโดย เกรกอรี่ ซี. เช่า (Gregory C. Chow) หรือเรียกอีกอย่างหนึ่งก็คือ การทดสอบเช่า (Chow test) ตามปกติการทดสอบสมมติฐานเกี่ยวกับพารามิเตอร์โดยใช้การทดสอบเช่านั้นจะต้องมีข้อตกลงเบื้องต้นเกี่ยวกับความคลาดเคลื่อนดังนี้คือ ความคลาดเคลื่อนจะต้องมีการแจกแจงแบบปกติ ที่มีค่าเฉลี่ยเป็น 0 ค่าความแปรปรวนเป็น σ^2 และเป็นอิสระซึ่งกันและกันหรือ $\varepsilon \sim NID(0, \sigma^2 I)$ อย่างไรก็ตามในทางปฏิบัติบ่อยครั้งที่เราพบว่า ข้อมูลที่ถูกนำมาใช้วิเคราะห์นั้นมีอยู่ไม่น้อยที่ไม่เป็นไปตามข้อตกลงเบื้องต้นดังกล่าว คือลักษณะการแจกแจงของความคลาดเคลื่อนมักจะไม่เป็นแบบปกติ เช่น อาจเป็นการแจกแจงที่มีหางยาว (long tails) หรือมีการกระจายไปทางหางมาก (heavy tails) ซึ่งการแจกแจงลักษณะเช่นนี้จะทำให้เกิดค่าความคลาดเคลื่อนบางค่ามีค่าสูงมาก ๆ หรือต่ำมาก ๆ ซึ่งจะไปมีผลทำให้ค่าสังเกต (y) บางค่าเป็นค่าผิดปกติ * (outliers) ดังนั้นสิ่งที่น่าสนใจศึกษาอีกเรื่องหนึ่งของการวิจัยในครั้งนี้คือ ความแกร่ง (robustness) ของการทดสอบเช่า โดยจะพิจารณาในลักษณะที่ว่าสถิติทดสอบนั้นจะต้องมีความไว (sensitive) ต่อการเปลี่ยนแปลงของปัจจัยที่ต้องการทดสอบและจะต้องไม่มีความไว (insensitive) ต่อการเปลี่ยนแปลงของสิ่งอื่นที่ไม่ใช่ปัจจัยที่ต้องการทดสอบ ทั้งนี้จะพิจารณาตามคำกล่าวของเนย์แมน (Neyman 1950 : 265 อ้างโดย Direk Srisukho 1974 : 38) ซึ่งกล่าวว่า "เมื่อต้องการที่จะเลือกใช้สถิติทดสอบ เราต้องเริ่มพิจารณาถึงความสามารถในการควบคุมความผิดพลาดประเภทที่ 1 (Type I Error) ก่อน โดยยอมให้ความน่าจะเป็นที่เกิดความผิดพลาดประเภทที่ 1 ไม่เกินขอบเขตของเกณฑ์ที่กำหนดไว้"

---

*ค่าผิดปกติ หมายถึง ค่าสังเกตที่มีค่ามากหรือน้อยกว่าค่าสังเกตอื่น ๆ อย่างผิดปกติ

ในการศึกษาครั้งนี้ผู้วิจัยได้ทำการศึกษาทั้ง การแบ่งข้อมูลด้วยวิธีดูเพล็กซ์ วิธีการประมาณค่า การทดสอบสมมติฐานเกี่ยวกับพารามิเตอร์ และการคำนวณความผิดพลาดประเภทที่ 1 เมื่อสภาพการแจกแจงไม่เป็นไปตามข้อตกลงเบื้องต้นดังกล่าว ซึ่งกระทำได้ยากมาก ดังนั้นการวิจัยในครั้งนี้จึงศึกษาโดยใช้วิธีมอนติคาร์โล (Monte Carlo method) ซึ่งสามารถกำหนดขนาดของข้อมูล ตัวแปร ลักษณะการแจกแจงของความคลาดเคลื่อน ค่าเฉลี่ย และความแปรปรวนได้

1.2 <u>วัตถุประสงค์ของการวิจัย</u>

1.2.1 เพื่อเสนอวิธีการตรวจสอบความเหมาะสมของตัวแบบถดถอยเชิงเส้นโดยการแบ่งข้อมูลด้วยวิธีดูเพล็กซ์ (Validating Linear Regression Model by Splitting Data with DUPLEX Algorithm)

1.2.2 เพื่อศึกษาความแกร่งของการทดสอบเช่า เมื่อสภาพการแจกแจงของความคลาดเคลื่อนไม่เป็นไปตามข้อตกลงเบื้องต้น

1.3 <u>สมมติฐานของการวิจัย</u>

ภายใต้ลักษณะการแจกแจงของความคลาดเคลื่อนที่มีรูปแบบต่าง ๆ กัน จะมีผลให้ความแกร่งของการทดสอบเช่าแตกต่างกัน

1.4 <u>ข้อตกลงเบื้องต้น</u>

1.4.1 ตัวแปรอิสระแต่ละตัวเป็นค่าคงที่ ที่มีการแจกแจงแบบเดียวกัน และเป็นอิสระซึ่งกันและกัน

1.4.2 ความคลาดเคลื่อนเป็นตัวแปรสุ่มแบบต่อเนื่อง ที่มีการแจกแจงแบบเดียวกัน และเป็นอิสระซึ่งกันและกัน

1.4.3 ใช้วิธีกำลังสองน้อยที่สุด (Least Squares Method) ในการประมาณค่าสัมประสิทธิ์ถดถอยเชิงเส้น

1.4.4 ในการแบ่งข้อมูลด้วยวิธีดูเพล็กซ์ จำนวนข้อมูลต้องมีไม่น้อยกว่า 2P + 20 ถึง 2P + 30 โดย P เป็นจำนวนสัมประสิทธิ์ถดถอยเชิงเส้น ทั้งนี้เพื่อให้องค์อิสระของความคลาดเคลื่อนมีเพียงพอสำหรับการทดล่อบสมมติฐาน

1.4.5 การวิจัยครั้งนี้ถือว่า ความคงที่ของสัมประสิทธิ์ถดถอยเชิงเส้นเป็นเกณฑ์ที่สำคัญในการพิจารณาความเหมาะสมของตัวแบบถดถอยเชิงเส้น

1.4.6 ใช้ความผิดพลาดประเภทที่ 1 เป็นเกณฑ์ในการพิจารณาความแกร่งของการทดล่อบเชา

1.5 <u>ขอบเขตของการวิจัย</u>

1.5.1 ศึกษาความน่าจะเป็นของความผิดพลาดประเภทที่ 1 ของการทดล่อบเชาในการทดล่อบเกี่ยวกับพารามิเตอร์ เมื่อความคลาดเคลื่อนมีการแจกแจงแบบโลจิสติค ดับเบิ้ล-เอ็กซ์โปเนนเชียล และปกติปลอมปน

1.5.2 การวิจัยครั้งนี้กำหนดให้ $\beta' = [50\ 1\ 1\ \ldots\ 1]_{(1xp)}$ ในประชากรทุกรูปแบบที่ศึกษา*

1.5.3 จำนวนข้อมูลที่ศึกษามีขนาดเป็น 30 40 50 60 70 80 90 และ 100 ส่วนจำนวนตัวแปรอิสระที่ศึกษาคือ 1 2 4 6 8 และ 10

1.5.4 ในการทดล่อบความคงที่ของสัมประสิทธิ์ถดถอยเชิงเส้นจะทดล่อบโดยใช้การทดล่อบเชา ณ ระดับนัยสำคัญ .01 .05 และ .10

---

*ในการวิจัยครั้งนี้ กำหนดค่าพารามิเตอร์เป็นค่าดังกล่าวเนื่องจาก การทดลองกระทำที่ขนาดตัวอย่าง 30 และ 100 พบว่า ค่าพารามิเตอร์ไม่มีผลต่อค่าความน่าจะเป็นของความผิดพลาดประเภทที่ 1

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

1.5.5 ในการวิจัยครั้งนี้จะศึกษาเมื่อความคลาดเคลื่อนมีค่าเฉลี่ยเป็น 0 และค่าความแปรปรวนเป็น 100*

1.5.6 ในกรณีการแจกแจงแบบปกติปลอมปน จะศึกษาที่เปอร์เซ็นต์การปลอมปนเป็น 1 5 10 และ 25 สำหรับสเกลแฟคเตอร์** (scale factor) มี 2 ระดับคือ 3 และ 10

1.5.7 การวิจัยครั้งนี้จำลองการทดลองขึ้นโดยใช้เทคนิคมอนติคาร์โล (Monte Carlo Technique) จากเครื่องคอมพิวเตอร์ IBM 370/3031 ซึ่งจะศึกษาเมื่อความคลาดเคลื่อน มีการแจกแจงแบบโลจิสติค ดับเบิ้ลเอ็กซ์โปเนนเชียล และปกติปลอมปน โดยมีฟังก์ชันความน่าจะเป็น ค่าคาดหวัง และความแปรปรวน ของการแจกแจงแต่ละรูปแบบเป็นดังนี้

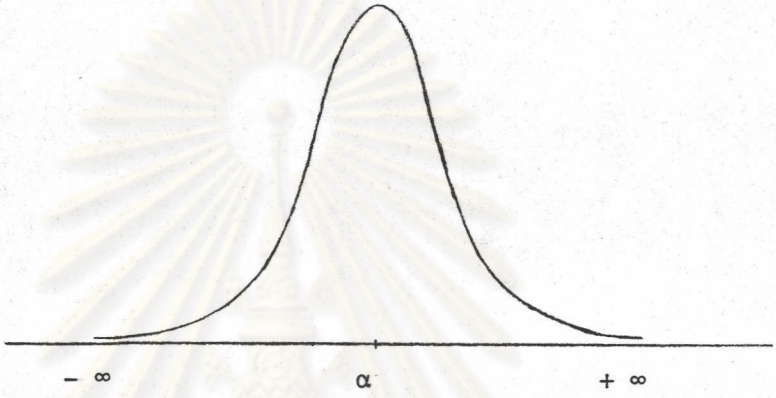
---

*การวิจัยในครั้งนี้กำหนดค่าความแปรปรวนเป็นค่าดังกล่าวเนื่องจาก การทดลองกระทำที่ขนาดตัวอย่าง 30 และ 100 พบว่า ค่าความแปรปรวนไม่มีผลต่อค่าความน่าจะเป็นของความผิดพลาดประเภทที่ 1 และในกรณีที่ความคลาดเคลื่อนมีการแจกแจงแบบปกติปลอมปน ค่าความแปรปรวนจะต่างจากค่าดังกล่าวเนื่องจากค่าสเกลแฟคเตอร์และเปอร์เซ็นต์การปลอมปนมีขนาดต่าง ๆ กัน

**สเกลแฟคเตอร์ หมายถึง ค่าที่ปรับให้ข้อมูลมีการกระจายมากขึ้น โดยสเกลแฟคเตอร์ที่มีค่าสูงจะทำให้เกิดค่าผิดปกติในข้อมูล และจากการศึกษาที่ผ่านมาพบว่า สเกลแฟคเตอร์ที่มีค่าน้อยกว่า 3 จะมีโอกาสของการเกิดค่าผิดปกติน้อย ส่วนสเกลแฟคเตอร์ที่มีค่ามากกว่า 10 จะมีโอกาสของการเกิดค่าผิดปกติมาก จึงทำการศึกษาเพียง 2 ระดับดังกล่าว

1.5.7.1 การแจกแจงแบบโลจิสติค (Logistic Distribution)

$$f(x) = \frac{1}{\beta} \cdot \frac{e^{\frac{-(x-\alpha)}{\beta}}}{\left[1 + e^{\frac{-(x-\alpha)}{\beta}}\right]^2}, \quad -\infty < x < \infty$$



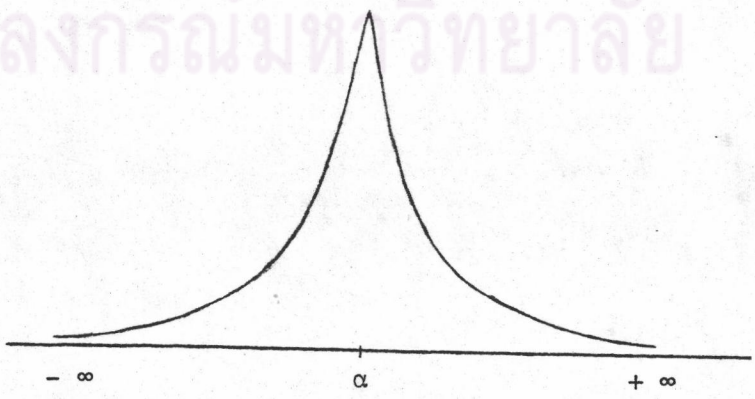
$$E(x) = \alpha$$

$$V(x) = \frac{1}{3}\Pi^2\beta^2$$

1.5.7.2 การแจกแจงแบบดับเบิ้ลเอ็กซ์โปเนนเชียล (Double Exponential Distribution)

$$f(x) = \frac{1}{2\beta} \cdot e^{\frac{-|x-\alpha|}{\beta}}, \quad -\infty < x < \infty$$

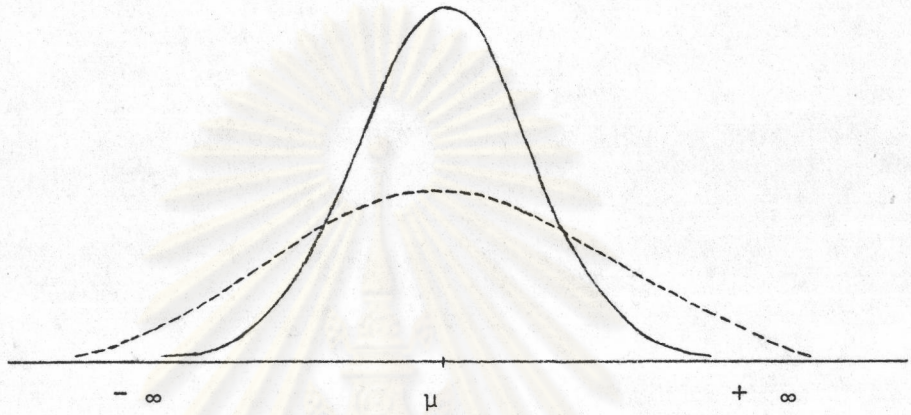
ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย



$$E(x) = \alpha$$

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

1.5.7.3 การแจกแจงแบบปกติปลอมปน (Scale Contaminated Normal Distribution)



ลักษณะการแจกแจงแบบปกติปลอมปนที่พิจารณาในที่นี้เป็นการแจกแจงที่แปลงมาจากการแจกแจงแบบปกติ ซึ่งมีฟังก์ชันการแปลงเป็นดังนี้

$$F = (1 - P)\, N(\mu , \sigma^2) + PN (\mu , c^2\sigma^2) , c > 0$$

หมายความว่า x มาจากการแจกแจง $N(\mu , \sigma^2)$ ด้วยความน่าจะเป็น $(1 - P)$ และมาจากการแจกแจง $N(\mu , c^2\sigma^2)$ ด้วยความน่าจะเป็น P

ในที่นี้ μ และ σ^2 เป็นค่าเฉลี่ยและความแปรปรวนของความคลาดเคลื่อน

$(\mu = 0 , \sigma^2 = 100)$

P และ c เป็นค่าที่กำหนดสัดส่วนการปลอมปนและสเกลแฟคเตอร์ ซึ่งจะทำให้เกิดค่าผิดปกติ

1.5.8 การจำลองการทดลองจะกระทำซ้ำกันไม่ต่ำกว่า 100 ครั้ง ในแต่ละสถานการณ์ของการทดลอง

1.6 คำจำกัดความ

1.6.1 ความผิดพลาดประเภทที่ 1 (Type I error) เป็นความผิดพลาดที่เกิดจากการปฏิเสธสมมติฐาน H_o เมื่อสมมติฐาน H_o ถูก

1.6.2 ความแกร่ง (Robustness) ของการทดสอบ หมายถึงคุณสมบัติของการทดสอบที่ไม่ไวต่อการเปลี่ยนแปลงของปัจจัยอื่นที่ไม่ใช่ปัจจัยที่ต้องการทดสอบ เช่น การฝ่าฝืนข้อตกลงเบื้องต้นของการทดสอบนั้น ซึ่งสิ่งที่ใช้พิจารณาความแกร่งของการทดสอบคือ ความน่าจะเป็นของความผิดพลาดประเภทที่ 1

1.7 ประโยชน์ของการวิจัย

เพื่อเป็นการเสนอวิธีการตรวจสอบความเหมาะสมของตัวแบบถดถอยเชิงเส้นโดยการแบ่งข้อมูลด้วยวิธีดูเพล็กซ์ซึ่งเป็นวิธีที่ยังไม่แพร่หลายนักสำหรับผู้วิจัยทั้งในสาขาสถิติและสาขาอื่น ๆ นอกจากนี้ยังช่วยให้มีข้อสรุปที่เป็นประโยชน์ในการทดสอบเกี่ยวกับพารามิเตอร์ได้อย่างมีประสิทธิภาพและเหมาะสม เมื่อความคลาดเคลื่อนไม่เป็นไปตามข้อตกลงเบื้องต้น

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย