การแปลงข้อมูลผลการวิจัยโดยวิธีทางสถิติ

*ทรงพล รดิศพงศ์**

ในปัจจุบันการวิจัยทางด้านต่าง ๆ มีบทบาทที่สำคัญต่อการดำเนินชีวิตของคนเราทั้งทางตรงและทางอ้อมเป็นอย่างมาก เนื่องจากการวิจัยก่อให้เกิดความรู้ใหม่ที่เป็นคำตอบของปัญหาต่าง ๆ ได้ การวิจัยเกี่ยวกับปัญหาด้านต่าง ๆ มีวิธีการดำเนินการได้หลายวิธีด้วยกัน แต่วิธีหนึ่งในจำนวนหลายวิธีที่มีการยอมรับและเชื่อถือกันมาก คือ การวิจัยที่อาศัยข้อมูลและระเบียบวิธีทางสถิติเข้ามาช่วย ดังนั้นวิธีการดังกล่าวจึงถูกนำมาใช้กันอย่างกว้างขวางกับงานทางสังคมศาสตร์และวิทยาศาสตร์ทุกสาขา ระเบียบวิธีการทางสถิติเป็นวิธีการทางวิทยาศาสตร์ (Scientific Method) ที่ใช้เป็นเครื่องมือวินิจฉัยเกี่ยวกับความถูกต้องของสารสนเทศ เพื่อหาผลสรุปที่ถูกต้องของการศึกษาค้นคว้าสำหรับช่วยในการตัดสินใจอย่างมีประสิทธิภาพ ระเบียบวิธีทางสถิติที่ใช้ในงานวิจัยต่าง ๆ มีขั้นตอนสำคัญ 4 ขั้นตอนคือ การเก็บรวบรวมข้อมูล (Data Collection) การประมวลผลและการนำเสนอข้อมูล (Data Processing and Presentation) การวิเคราะห์ข้อมูล (Data Analysis) และการตีความหมายข้อมูล (Data Interpretation)

จากขั้นตอนของระเบียบวิธีทางสถิติดังกล่าวจะเห็นได้ว่าการวิเคราะห์ข้อมูลทางสถิติเป็นขั้นตอนที่สำคัญของการวิจัย เพราะจะให้ได้มาซึ่งข้อสรุปของการวิจัย การวิจัยเชิงทดลองอยู่ในการวิจัยประเภทหนึ่ง โดยทั่วไปมุ่งศึกษาเปรียบเทียบความแตกต่างระหว่างวิธีการทดลองหรือทรีทเมนต์ (Treatment) ที่ให้กับหน่วยทดลอง (Experimental Unit) ว่ามีอิทธิพลแตกต่างกันหรือไม่ ซึ่งผู้วิจัยจะต้องมีความเข้าใจในเรื่องที่จะศึกษาและออกแบบการทดลองได้เป็นอย่างดี ในการศึกษาเพื่อทดสอบค่าเฉลี่ยประชากรของแต่ละทรีทเมนต์มีค่าเท่ากันหรือไม่ หรือทดสอบอิทธิพลของทรีทเมนต์มีนัยสำคัญหรือไม่นั้น มีข้อตกลงเบื้องต้น (Assumption) คือ ความคลาดเคลื่อนสุ่มมาจากประชากรเดียวกันที่มีการแจกแจงแบบปกติ โดยมีค่าเฉลี่ยศูนย์และความแปรปรวนเท่ากันในทุกทรีทเมนต์และแต่ละตัวก็เป็นอิสระซึ่งกันและกัน หากพบว่าข้อมูลไม่เป็นไปตามข้อตกลงเบื้องต้นและต้องการใช้การทดสอบด้วยวิธีการทางสถิติแบบพาราเมตริก (Parametric test) จะทำให้ประสิทธิภาพของการทดสอบลดลง นำไปสู่การสรุปผลการวิเคราะห์ที่ผิดพลาดและขาดความน่าเชื่อถือ ซึ่งวิธีการแก้ปัญหาในกรณีที่ข้อมูลมีการแจกแจงแบบไม่ปกติและความแปรปรวนมีค่าไม่เท่ากัน คือ การแปลงข้อมูล (Data Transformation) โดยใช้วิธีการทางสถิติ (Statistical Methods)

การแปลงข้อมูล หมายถึง การเปลี่ยนสภาพของข้อมูลที่เราศึกษาให้มีการแจกแจงแบบปกติหรือทำให้ความแปรปรวนมีค่าเท่ากัน เนื่องจากข้อตกลงเบื้องต้นของการทดสอบสถิติบางตัวได้กำหนดไว้ เช่น การทดสอบค่าเฉลี่ย การทดสอบความแปรปรวน (Analysis of Variance) การวิเคราะห์การถดถอย (Regression Analysis) เป็นต้น การแปลงข้อมูลทำได้โดยใช้วิธีการทางสถิติ มีวัตถุประสงค์เพื่อแปลงข้อมูลเป็นมาตราใหม่แล้วข้อมูลมีการแจกแจงแบบปกติหรือใกล้เคียงกับแบบปกติ ค่าเฉลี่ยและความแปรปรวนของข้อมูลที่แปลงแล้วเป็นอิสระต่อกันจะทำให้ความแปรปรวนมีค่าเท่ากัน การแปลงข้อมูลกระทำได้โดยการยกกำลังข้อมูลเดิม โดยประมาณตัวยกกำลังตามเกณฑ์ที่กำหนดจากความสัมพันธ์ระหว่างความแปรปรวนที่เป็นสัดส่วนกับค่าเฉลี่ยของประชากรแต่ละทรีทเมนต์ ดังนี้

1) การแปลงโดยใช้รากที่สอง (Square root transformation)

ถ้าข้อมูลมีความแปรปรวน (σ_i^2) และค่าเฉลี่ย (μ_i) เป็นสัดส่วนกัน นั่นคือ $\sigma_i^2 = k\,\mu_i$ จะทำให้ความแปรปรวนไม่เป็นอิสระจากค่าเฉลี่ย การแปลงข้อมูลโดยใช้

$y_{ij}^* = \sqrt{y_{ij}}$ เมื่อ y_{ij} คือ ค่าสังเกตเดิม

k คือ ค่าคงที่

2) การแปลงโดยใช้ลอการิทึม (Logarithmic transformation)

ถ้าข้อมูลมีค่าส่วนเบี่ยงเบนมาตรฐาน (σ_i) และค่าเฉลี่ยเป็นสัดส่วนกัน นั่นคือ $\sigma_i = k\,\mu_i$ การแปลงข้อมูล โดยใช้

$y_{ij}^* = \log(y_{ij})$

* นักวิทยาศาสตร์ชำนาญการ โครงการฟิสิกส์และวิศวกรรม

3) การแปลงโดยใช้รากที่สองของการกลับเศษส่วน (Reciprocal square root transformation)

ถ้าข้อมูลมีค่าส่วนเบี่ยงเบนมาตรฐานและค่าเฉลี่ยยกกำลังสามส่วนสองเป็นสัดส่วนกัน นั่นคือ $\sigma_i = k\ \mu_i^{3/2}$ การ แปลงข้อมูลโดยใช้

$$y_{ij}^{*} = \frac{1}{\sqrt{y_{ij}}}$$

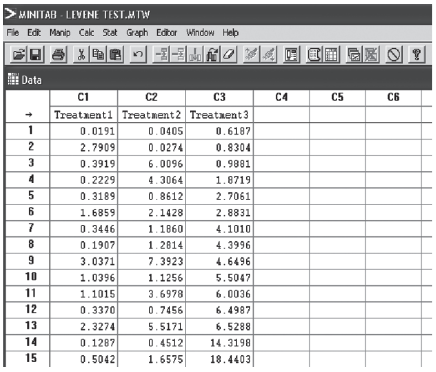
4) การแปลงโดยใช้เศษส่วน (Reciprocal transformation)

ถ้าข้อมูลมีค่าส่วนเบี่ยงเบนมาตรฐานและค่าเฉลี่ยยกกำลังสองเป็นสัดส่วนกัน นั่นคือ $\sigma_i = k\ \mu_i^{2}$ การแปลงข้อมูลโดยใช้

$$y_{ij}^{*} = \frac{1}{\sqrt{y_{ij}}}$$

ตัวอย่างการแปลงข้อมูล

1. ข้อมูลที่มีการแจกแจงแบบไม่ปกติและความแปรปรวนมีค่าไม่เท่ากัน



MINITAB - LEVENE TEST.MTW

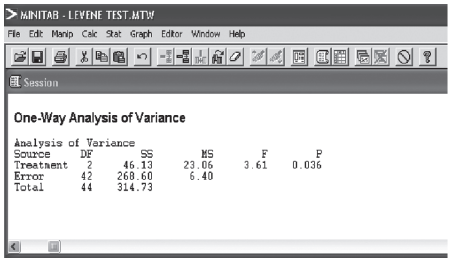
File Edit Manip Calc Stat Graph Editor Window Help

Data

	C1	C2	C3	C4	C5	C6
→	Treatment1	Treatment2	Treatment3			
1	0.0191	0.0405	0.6187			
2	2.7909	0.0274	0.8304			
3	0.3919	6.0096	0.9881			
4	0.2229	4.3064	1.8719			
5	0.3189	0.8612	2.7061			
6	1.6859	2.1428	2.8831			
7	0.3446	1.1860	4.1010			
8	0.1907	1.2814	4.3996			
9	3.0371	7.3923	4.6496			
10	1.0396	1.1256	5.5047			
11	1.1015	3.6978	6.0036			
12	0.3370	0.7456	6.4987			
13	2.3274	5.5171	6.5288			
14	0.1287	0.4512	14.3198			
15	0.5042	1.6575	18.4403			

ภาพที่ 1 ข้อมูลตัวอย่าง

1.1 ทดสอบความเท่ากันของความแปรปรวนของประชากรด้วยสถิติของเลวีน (Levene Test) ที่ระดับนัยสำคัญ (α) 0.05



MINITAB - LEVENE TEST.MTW

File Edit Manip Calc Stat Graph Editor Window Help

Session

One-Way Analysis of Variance

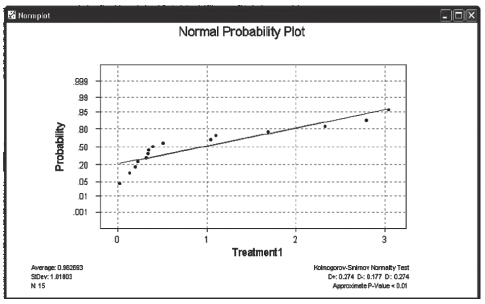
Analysis of Variance

Source	DF	SS	MS	F	P
Treatment	2	46.13	23.06	3.61	0.036
Error	42	268.60	6.40		
Total	44	314.73			

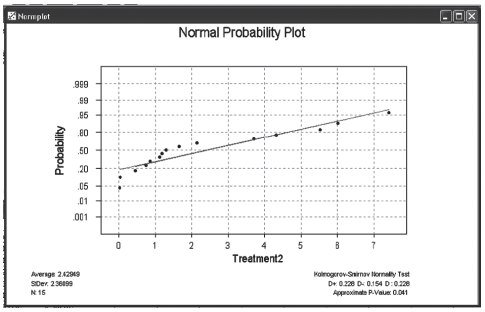
ภาพที่ 2 ผลการวิเคราะห์ความแปรปรวน

สามารถแปลผลได้ว่าความแปรปรวนของประชากรมีค่าไม่เท่ากัน

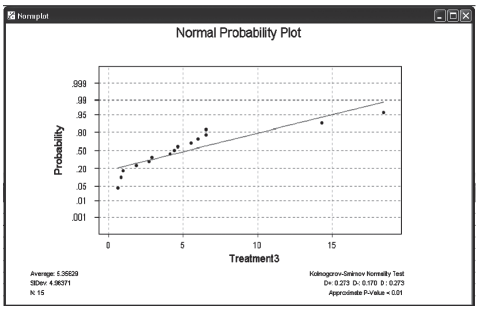
1.2 ทดสอบการแจกแจงแบบปกติของประชากรด้วยสถิติโคลโมโกรอฟ-สเมอร์นอฟ (Kolmogorov –Smirnov) ที่ระดับนัยสำคัญ 0.05 และกราฟความน่าจะเป็นของส่วนตกค้าง (Normal Probability Plot of the Residuals)



ภาพที่ 3 ทดสอบการแจกแจงแบบปกติที่ทรีทเมนต์ 1



ภาพที่ 4 ทดสอบการแจกแจงแบบปกติที่ทรีทเมนต์ 2



ภาพที่ 5 ทดสอบการแจกแจงแบบปกติที่ทรีทเมนต์ 3

สามารถแปลผลได้ว่าประชากรมีการแจกแจงแบบไม่ปกติทั้ง 3 ทรีทเมนต์

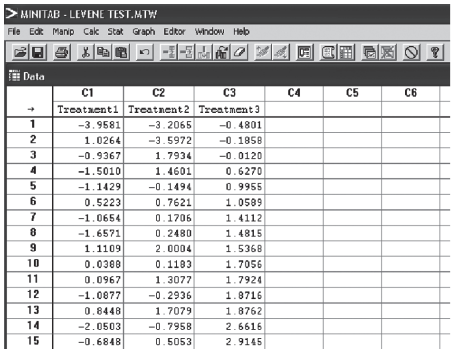
2. การประมาณตัวยกกำลังตามเกณฑ์ที่กำหนดจากความสัมพันธ์ระหว่างความแปรปรวนที่เป็นสัดส่วนกับค่าเฉลี่ยของประชากรแต่ละทรีทเมนต์

ตารางที่ 1 ความสัมพันธ์ระหว่าง σ_i^2 กับ μ_i

ความสัมพันธ์	ทรีทเมนต์ 1	ทรีทเมนต์ 2	ทรีทเมนต์ 3
$\sigma_i^2 = k\,\mu_i$	1.0765	2.9532	4.5999
$\sigma_i = k\,\mu_i$	1.0575	0.9726	0.9267
$\sigma_i = k\,\mu_i^{3/2}$	1.0778	0.5505	0.4004
$\sigma_i = k\,\mu_i^2$	1.0985	0.3116	0.1730

พบว่าความสัมพันธ์ $\sigma_i = k\,\mu_i$ คงที่มากที่สุด จึงใช้รูปแบบการแปลงแบบลอการิทึม

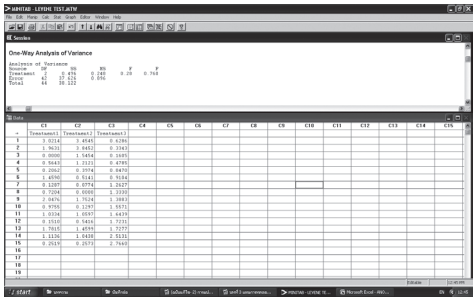
3. ทำการแปลงข้อมูลโดยใช้ลอการิทึม



	C1	C2	C3	C4	C5	C6
→	Treatment1	Treatment2	Treatment3			
1	-3.9581	-3.2065	-0.4801			
2	1.0264	-3.5972	-0.1858			
3	-0.9367	1.7934	-0.0120			
4	-1.5010	1.4601	0.6270			
5	-1.1429	-0.1494	0.9955			
6	0.5223	0.7621	1.0589			
7	-1.0654	0.1706	1.4112			
8	-1.6571	0.2480	1.4815			
9	1.1109	2.0004	1.5368			
10	0.0388	0.1193	1.7056			
11	0.0967	1.3077	1.7924			
12	-1.0877	-0.2936	1.8716			
13	0.8448	1.7079	1.8762			
14	-2.0503	-0.7958	2.6616			
15	-0.6848	0.5053	2.9145			

ภาพที่ 6 ข้อมูลตัวอย่างที่ได้จากการแปลง

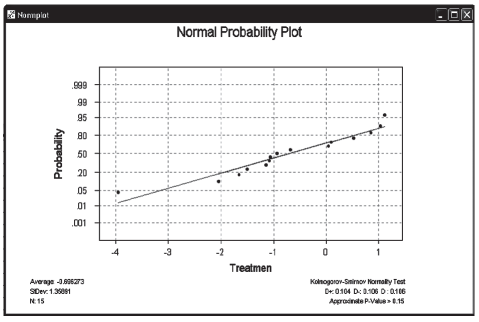
3.1 ทดสอบความเท่ากันของความแปรปรวนของประชากรจากข้อมูลที่ได้จากการแปลงด้วยสถิติของเลวีน ที่ระดับนัยสำคัญ 0.05



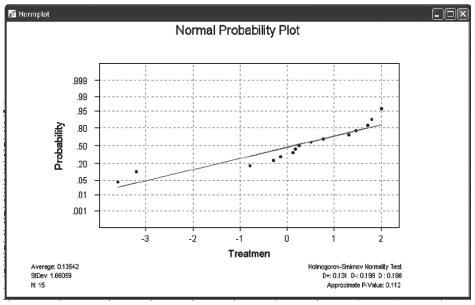
ภาพที่ 7 ผลการวิเคราะห์ความแปรปรวนของข้อมูลที่ได้จากการแปลง

สามารถแปลผลได้ว่าความแปรปรวนของประชากรมีค่าเท่ากัน

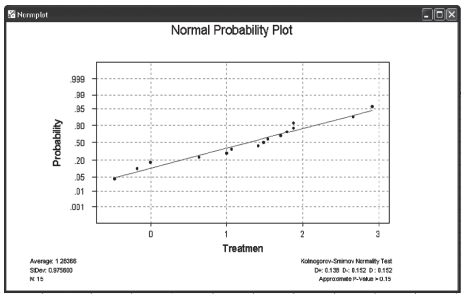
3.2 ทดสอบการแจกแจงแบบปกติของประชากรจากข้อมูลที่ได้จากการแปลงด้วยสถิติโคลโมโกรอฟ-สเมอร์นอฟ ที่ระดับนัยสำคัญ 0.05 และกราฟความน่าจะเป็นของส่วนตกค้าง



ภาพที่ 8 ทดสอบการแจกแจงแบบปกติของข้อมูลที่ได้จากการแปลงที่ทรีทเมนต์ 1



ภาพที่ 9 ทดสอบการแจกแจงแบบปกติของข้อมูลที่ได้จากการแปลงที่ทรีทเมนต์ 2



ภาพที่ 10 ทดสอบการแจกแจงแบบปกติของข้อมูลที่ได้จากการแปลงที่ทรีทเมนต์ 3

สามารถแปลผลได้ว่าประชากรมีการแจกแจงแบบปกติทั้ง 3 ทรีทเมนต์

ดังนั้นในการวิจัยเพื่อศึกษาอิทธิพลของทรีทเมนต์ สิ่งที่นักวิจัยพึงที่จะต้องตรวจสอบเบื้องต้นก่อนทำการวิเคราะห์ข้อมูล คือ ข้อมูลมาจากประชากรที่มีการแจกแจงแบบปกติและความแปรปรวนมีค่าเท่ากัน หากพบว่าข้อมูลไม่เป็นไปตามข้อตกลงเบื้องต้นเกี่ยวกับประชากรมีการแจกแจงแบบปกติและความแปรปรวนมีค่าเท่ากัน ควรทำการแปลงข้อมูลโดยวิธีการทางสถิติ เพื่อทำให้ข้อมูลผลของการทดสอบดีขึ้น นำไปสู่การสรุปผลการวิเคราะห์ที่ถูกต้อง มีความน่าเชื่อถือ

เอกสารอ้างอิง

Kuehl, R. O. **Design of experiments: statistical principles of research design and analysis.** 2nd ed., Duxbury: Pacific Grove. 2000. 688 p.

Kutner, M. H., et al. **Applied linear statistical models.** 5th ed., [n.p]: McGraw Hill, 2005.

Montgomery, D. C. **Design and analysis of experiments.** 6th ed., New York: Wiley, 2005.

กัลยา วานิชย์บัญชา. **การวิเคราะห์ข้อมูลหลายตัวแปร.** กรุงเทพมหานคร: จุฬาลงกรณ์มหาวิทยาลัย. 2548.

สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ. คณะวิทยาศาสตร์ ภาควิชาสถิติประยุกต์. **การวางแผนการทดลอง เล่ม 1** โดย วิจิตรา พลเยี่ยม. กรุงเทพมหานคร: สถาบันฯ, 2549.