บทที่ 2

วรรณคดีที่เกี่ยวข้อง

ประวัติและความเป็นมาของการวิเคราะห์ความแปรปรวน

การทดสอบเอฟ หรือ การวิเคราะห์ความแปรปรวน (ANOVA) เป็นวิธีการหนึ่งที่ Sir Ronald A. Fisher คิดค้นขึ้น (Krukal and Tanur 1978 352 - 357) ในตอนแรกงานที่ใช้เทคนิคนี้ในการวิเคราะห์ คือ งานในด้านเกษตรกรรมและชีววิทยา แต่ปัจจุบันนิยมใช้กันมาในทุก ๆ สาขาวิชา โดยเฉพาะอย่างยิ่งในสาขาสังคมศาสตร์ ค่าสถิติ F นี้ Snedecor ได้เป็นผู้เสนอชื่อขึ้นเพื่อให้เป็นเกียรติแก่ R.A. Fisher

การทดสอบเอฟ (F - Test) เป็นกระบวนการหนึ่งทางคณิตศาสตร์ เพื่อใช้ในการทดสอบสมมติฐานความเท่ากันของมัชฌิมเลขคณิตจากผลการทดลองของแต่ละทรีทเมนต์ (Treatment Population Mean) โดยทางทฤษฎีค่า F เป็นการแจกแจงที่ได้มาจากสัดส่วนของ χ^2 2 ค่า คือ

$$F = \frac{\chi_1^2 / v_1}{\chi_2^2 / v_2} = \frac{S_1^2 / \sigma_1^2}{S_2^2 / \sigma_2^2} = \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2}$$

ซึ่งค่า χ_1^2 เป็นการแจกแจงของ Chi - Square มีชั้นแห่งความเป็นอิสระ $v_1 = n_1 - 1$

χ_2^2 เป็นการแจกแจงของ Chi - Square มีชั้นแห่งความเป็นอิสระ $v_2 = n_2 - 1$

แต่ค่า χ^2 นั้นได้มาจากค่า Z^2 และค่า Z มีข้อตกลงเบื้องต้นข้อหนึ่งว่าลักษณะการแจกแจงของประชากรต้องเป็นโค้งปกติ (Normal Distribution) ดังนั้นค่า F จึงต้องระบุข้อตกลงเบื้องต้นอย่างเข้มงวดด้วย

ค่าสถิติ F จะมีการแจกแจงแบบ F (F - Distribution) โดยมีลักษณะการแจกแจงเป็นโค้งเบ้ทางขวา (positive Skewed Distribution) ค่าพิสัยอยู่ระหว่าง 0 กับ ∞ ค่ามัธยฐานเท่ากับหรือน้อยกว่า 1 ค่ามัชฌิมเลขคณิตเท่ากับ $v_2/(v_2-2)$ โดยที่ v_2 มากกว่า 2 ค่าความแปรปรวนเท่ากับ $\dfrac{2\,v_2^2(v_1+v_2+2)}{v_2(v_2-2)(v_2-4)}$ โดยที่ v_2 มากกว่า 4 ค่าฐานนิยมเท่ากับ $\dfrac{v_2(v_1-2)}{v_1(v_2+2)}$ โดยที่ v_1 มากกว่า 2 ซึ่งค่า

v_1, v_2 เป็นค่าชั้นแห่งความเป็นอิสระ (degree of freedum) ของผลบวกกำลังสองระหว่างกลุ่ม (Sum of Square between Groups) และผลบวกกำลังสองภายในกลุ่ม (Sum of Square Within Groups) (Kruskal and Tanur 1978 : 164)

ตารางค่า F สร้างขึ้นโดย P.C. Tang (1938)(Feldt 1958 : 871) และต่อมาในปี ค.ศ. 1967 M.L. Tiku ได้พัฒนารายละเอียดของตาราง F ให้ดีขึ้น โดยตาราง F แสดงพื้นที่ใต้โค้งที่รวมกันเท่ากับ 1 การใช้ตาราง F ต้องทราบค่าชั้นแห่งความเป็นอิสระของเศษและส่วน พื้นที่ที่เกี่ยวข้องกับ F นิยมใช้เฉพาะที่อยู่ปลายทางขวามือของโค้ง

โมเดลสำหรับการวิเคราะห์ความแปรปรวน

การวิเคราะห์ความแปรปรวนมีหลายแบบด้วยกันขึ้นอยู่กับการออกแบบการทดลองดังนี้ คือ

1. ตัวแปรอิสระ 1 ตัว หลายระดับการทดลอง และตัวแปรตาม 1 ตัว มีดังนี้ คือ

1.1 การวิเคราะห์ความแปรปรวนแบบสุ่มสมบูรณ์ (Completely Randomized Design)

1.2 การวิเคราะห์ความแปรปรวนแบบกลุ่มสุ่ม (Randomized Block Design)

1.3 การวิเคราะห์ความแปรปรวนแบบจตุรัส ลาติน(Latin Square Design)

2. ตัวแปรอิสระหลายตัว แต่ละตัวมีหลายระดับการทดลอง ตัวแปรตาม 1 ตัว มีดังนี้ คือ

2.1 การวิเคราะห์ความแปรปรวนแบบเฟคทอเรียล (Factorial Design)

2.2 การวิเคราะห์ความแปรปรวนแบบแยกส่วน (Spilt - Plot Design)

การวิเคราะห์ความแปรปรวนตามโมเดลดังกล่าวแล้ว สามารถจะจำแนกเป็นโมเดลเชิงเส้นตรงของผลทดลองตามกำหนด(Fixed Effects Linear Model) และโมเดลเชิงเส้นตรงของผลทดลองตามการสุ่ม (Random Effects Linear Model) สำหรับงานวิจัยนี้จะได้ยกตัวอย่างเฉพาะโมเดลที่ใช้กันมาก คือ โมเดลเชิงเส้นตรงของผลทดลองตามกำหนด (Fixed Effect Linear Model) มีรูปแบบดังนี้

$$X_{ij} = \mu + \alpha_j + \varepsilon_{ij}$$

เมื่อค่า X_{ij} คือ ผลการวัดตัวแปรตามซึ่งได้จากตัวอย่างที่สุ่มมาจากประชากร

μ คือ มัชฌิมเลขคณิตของประชากร

α_j คือ ผลจากการทดลอง (Treatment Effects) ซึ่งจะมีค่าคงที่สำหรับตัวอย่างทั้งหมดภายในประชากร

ε_{ij} คือ ความคลาดเคลื่อนทางการทดลอง (Error Effect) ซึ่งเป็นอิสระสำหรับค่าความคลาดเคลื่อนอื่นๆ

และตารางวิเคราะห์ความแปรปรวน มีรูปแบบ ดังนี้ คือ

แหล่งความแปรปรวน	SS	df	MS	F
ระหว่างกลุ่ม (B)	$n\sum_{j=1}^{k}(\overline{X}_{.j}-\overline{X}_{..})^2$	$k-1$	SS_B/df	MS_B/MS_W
ภายในกลุ่ม (W)	$\sum_{j=1}^{k}\sum_{i=1}^{n}(X_{ij}-\overline{X}_{.j})^2$	$N-k$	SS_W/df	
ทั้งหมด (T)	$SS_B + SS_W$	$N-1$		

โดยที่ N คือ จำนวนข้อมูลทั้งหมด

n คือ จำนวนข้อมูลในแต่ละระดับการทดลอง

k คือ จำนวนระดับการทดลอง

X_{ij} คือ คะแนนแต่ละตัวซึ่งจะเป็นที่ i ของกลุ่มที่ j

$\overline{X}_{.j}$ คือ มัชฌิมเลขคณิตของข้อมูลในแต่ละระดับการทดลอง

$\overline{X}_{..}$ คือ มัชฌิมเลขคณิตของข้อมูลทั้งหมด

ข้อตกลงเบื้องต้นของ ANOVA

1. ข้อตกลงเบื้องต้นเกี่ยวกับประชากรมีการแจกแจงแบบปกติ (Assumption od Normally Distributed Population) หมายความว่า เศษและส่วนของอัตราส่วน F (F - Ratio) ต้องเป็นอิสระต่อกัน โดยข้อมูลต้องได้รับเลือกอย่างสุ่มจากประชากรที่มีการแจกแจงปกติ นอกจากนี้ความคลาดเคลื่อนทางการทดลองในโมเดล

ผลการทดลองตามกำหนดและตามการสุ่ม (Fixed - Effects and Random - Effects Modols) ต้องมีการแจกแจงแบบปกติในแต่ละประชากรทดลอง

2. ข้อตกลงเบื้องต้นของความเป็นเอกพันธุ์ของความแปรปรวนคลาดเคลื่อนของประชากร (Assumption of Homogeneity of Population Error Variance) หมายความว่า ความแปรปรวนที่เนื่องจากความคลาดเคลื่อนภายในประชากรทดลองแต่ละกลุ่มเท่ากัน (Kirk 1969 : 61)

3. ข้อตกลงเบื้องต้นของการรวมกันได้ของผลการทดลอง (Assumption Addivity of Effects) หมายความว่า แรงปะทะ (Interaction) ระหว่างระดับการทดลองและกลุ่มมีค่าเท่ากับศูนย์ หรือหมายถึงความแปรปรวนร่วมระหว่างระดับการทดลองแต่ละคู่เท่ากัน (Scheffé 1970 : 94)

4. ข้อตกลงเบื้องต้นเกี่ยวกับความเป็นอิสระในการสุ่มตัวอย่าง (Assumption of Indenpendence Within Samples) และความเป็นอิสระระหว่างกลุ่มตัวอย่าง (Assumption of Independence Between Samples) ในกรณีของการสุ่มแบบสมบูรณ์ (Marascuilo 1971 : 344)

<u>สถิติที่ใช้ทดสอบข้อตกลงเบื้องต้นของ ANOVA</u>

สถิติที่ใช้ทดสอบข้อตกลงเบื้องต้นเกี่ยวกับความเป็นเอกพันธุ์ของความแปรปรวนความคลาดเคลื่อนของประชากร โดยตั้งสมมติฐานว่า

$$H_0 : \sigma_1^2 = \sigma_2^2 = \ldots\ldots\ldots = \sigma_k^2$$

ได้แก่ Bartlett's Test , Hartley's Test and Cochran's Test (Kirk 1969 : 61 - 62)

1. <u>Bartlett's Test</u> (1937) สถิติที่ใช้ คือ

$$B = \frac{2.30259}{C}\left[V \log_{10} MS_e - \sum_{j=1}^{k}(V_j \log_{10}\hat{\sigma}_j^2)\right]$$

โดยที่

$$C = 1 + \frac{\sum_{j=1}^{k} 1/V_j - 1/V}{3(k-1)}$$

V_j = ชั้นของความเป็นอิสระ $\hat{\sigma}_j^2$

V = ชั้นของความเป็นอิสระ MS_e

$\hat{\sigma}_j^2$ = ตัวประมาณค่าที่ไม่ลำเอียงของความแปรปรวนของประชากร

คำนวณจาก

$$\hat{\sigma}_j^2 = \left[\sum_1^n X^2 - (\sum_1^n X)^2/n \right] / (n-1)$$

$$MS_e = \sum_{j=1}^{k} \hat{\sigma}_j^2 / V$$

k = จำนวนความแปรปรวน

ถ้า V_j มากกว่าหรือเท่ากับ 5 ค่า B จะมีการแจกแจงเป็น X^2 ที่ชั้นแห่งความเป็นอิสระเท่ากับ k - 1

ถ้า V_j น้อยกว่า 5 ให้ใช้ตารางที่คำนวณโดย Merrington และ Thompson (1946)

2. Hartley ' s F - Max Test (1940 , 1950)

สถิติที่ใช้คือ

$$F_{max} = \frac{\hat{\sigma}^2_{max}}{\hat{\sigma}^2_{min}}$$

โดยชั้นแห่งความเป็นอิสระเท่ากับ k และ n - 1

เมื่อ k = จำนวนความแปรปรวน

n = จำนวนการสังเกตในแต่ละระดับของการทดลอง

สมมติฐานที่ว่าด้วยความเท่ากันของความแปรปรวนจะได้รับการปฏิเสธ ถ้า F_{max} มากกว่า F จากตาราง ถ้าจำนวน n ในแต่ละระดับการทดลองแตกต่างกันเพียงเล็กน้อย ให้ใช้ค่า n ค่าที่มากที่สุด เพื่อกำหนดชั้นแห่งความมีนัยสำคัญสำหรับการทดลองนั้น ซึ่งอาจทำให้ได้ผลเปลี่ยนแปลงไปเพียงเล็กน้อย คือทำให้จำนวนครั้งในการปฏิเสธสมมติฐานมากกว่าที่ควรจะเป็น

3. Cochran ' s Test (1941)

สถิติที่ใช้คือ

$$C = \frac{\hat{\sigma}_j^2 \ max}{\sum_{j=1}^{k} \hat{\sigma}_j^2}$$

โดยที่ $\hat{\sigma}_j^2$ มากที่สุด คือ ค่าความแปรปรวนที่ใหญ่ที่สุดในจำนวนความแปรปรวนทั้งหลาย

$\sum_{j=1}^{k} \hat{\sigma}_j^2$ คือ ผลบวกของความแปรปรวนทั้งหมด

C มีการแจกแจงเป็น F โดยสมมติฐานที่ว่าด้วยความเท่ากันของความแปรปรวนจะได้รับการปฏิเสธ ถ้าค่า C มากกว่า F จากตาราง และชั้นแห่งความเป็น

อิสระสำหรับการทดสอบนี้ เท่ากับ k และ n - 1 เหมือนกับค่าของ F_{max}

ข้อคิดเห็นเกี่ยวกับการฝ่าฝืนข้อตกลงเบื้องต้นในการวิเคราะห์ความแปรปรวน

(Effects of Failure To Meet Assumptions in Analysis of Variance)

Cochran (1947) (Kirk 1969:60) ชี้ให้เห็นว่ามันเป็นไปไม่ได้ที่จะแน่ใจว่า ข้อมูลที่ได้ตรงตามข้อตกลงเบื้องต้นที่ต้องการ การวิเคราะห์ความแปรปรวนจึงได้ค่าประมาณมากกว่าจะได้ค่าที่แท้จริง

Eisenhort , Cochran และ Bartlett ให้ผลสรุปในแนวเดียวกันว่า การฝ่าฝืนข้อตกลงเบื้องต้นมีผลต่อระดับความมีนัยสำคัญและความไว (Sensitivity) ของการทดสอบเอฟ (Cochran and Cox 1957 : 91 - 92) กล่าวคือ ถ้าผู้วิจัยทดสอบที่ระดับนัยสำคัญ .05 ซึ่งแท้จริงอาจจะอยู่ที่ระดับ.04 หรือ .07 ก็ได้ การขาดความไวก็จะทำให้มีผลต่ออำนาจในการทดสอบเอฟลดลง และการแจกแจง F นั้นถ้าขาดข้อตกลงเบื้องต้นบางส่วน ก็จะทำให้เกิดผลลำเอียงในการทดสอบด้วย

Bartlett (1937) , Cochran (1947) และ Edwards (1960 , 1968) ให้ผลสรุปในแนวเดียวกันว่า การวิเคราะห์ความแปรปรวนมีความแกร่ง (Robust) ถึงแม้ว่าจะฝ่าฝืน (Violate) ข้อตกลงเบื้องต้นเกี่ยวกับความเป็นเอกพันธุ์ในการแปรปรวนของความคลาดเคลื่อนของประชากร หรือฝ่าฝืนข้อตกลงเบื้องต้นเกี่ยวกับประชากรมีการแจกแจงปกติเพียงเล็กน้อย

ANOVA สำหรับผลการทดลองตามกำหนด (Fixed - Effects ANOVA) มีความแกร่งสำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 ถึงแม้ว่าจะฝ่าฝืนข้อตกลงเบื้องต้นของการทดสอบเอฟ โดยเฉพาะข้อตกลงเบื้องต้นเกี่ยวกับความเป็นเอกพันธุ์ของความแปรปรวนของประชากร ไม่มีผลกระทบต่อการทดสอบเอฟเลย(Kruskal and Tanur 1978:550)

จากการศึกษาของ Wayne Lee (1975 : 282 - 283) พบว่า ข้อตกลงเบื้องต้นทั้ง 3 ข้อ ถ้าล้มเหลวเพียงข้อเดียวหรือหลายข้อ อัตราส่วน F (F - Ratio) ก็ยังคงมีการแจกแจงความน่าจะเป็นภายใต้การศึกษาซ้ำ (Replication) แต่การแจกแจงนี้จะแตกต่างมากบ้างน้อยบ้างจากการแจกแจง F (F - Distribution) อย่างจริงๆ ขึ้นอยู่กับว่าข้อมูลนั้นแตกต่างจากข้อตกลงเบื้องต้นมากน้อยแค่ไหน เราเรียก การแจกแจงของอัตราส่วน F ในสถานการณ์การทดลองจริง ๆ ว่า การแจกแจงประสิทธิภาพของอัตรา

ส่วน F (Effective Distribution of the F - Ratio) ซึ่งจะแตกต่างไปจากการแจกแจงF ในทางทฤษฎี(Theoretical F - Distribution) เมื่อข้อตกลงเบื้องต้นล้มเหลว ภายใต้สมมติฐานสูญ (Null Hypothesis) อัตราส่วน F จะใหญ่กว่า $F_{100(1-\alpha)}$ (df1 , df 2) ที่ระดับนัยสำคัญ α ความน่าจะเป็นที่แท้จริงก็คืออัตราส่วนF จะใหญ่กว่า $F_{100(1-\alpha)}$ (df1 , df2) เท่ากับสัดส่วนของการแจกแจงประสิทธิภาพที่ใหญ่กว่า $F_{100(1-\alpha)}$,(df1 , df2) สัดส่วนนี้อาจใหญ่กว่าหรือเล็กกว่า α ได้ เมื่อสัดส่วนนี้ใหญ่กว่า α เรากล่าวว่า การทดสอบ F มีความเอนเอียงทางบวก (Positively Biased) ส่วนสัดส่วนที่เล็กกว่า α เรากล่าวว่าการทดสอบ F มีความเอนเอียงทางลบ (Negatively Biased) ปัญหาสำหรับการทดสอบความเอนเอียงทางบวก ก็คือ สมมติฐานสูญถูกปฏิเสธบ่อยครั้ง นั่นคืออัตราส่วน F ที่มีขนาดใหญ่เกิดขึ้นด้วยความน่าจะเป็นที่มากกว่าในทางทฤษฎี ดังนั้นผู้วิจัยจึงมักเขียนรายงานผิด คือ เกิดความคลาดเคลื่อนประเภทที่ 1 (Type I Error) ขึ้น ในทางตรงกันข้ามสำหรับปัญหาการทดสอบที่เอนเอียงทางลบ สมมติฐานสูญที่ผิดถูกยอมรับอย่างบ่อยครั้ง นั่นคืออัตราส่วน F ที่เล็กเกิดขึ้นด้วยความน่าจะเป็นที่มากกว่าในทางทฤษฎี ผู้วิจัยจึงมักเขียนรายงานผิด เกิดความคลาดเคลื่อนประเภทที่ 2 (Type II Error)

Norton (Lee 1975 :284) พบว่า การแจกแจงของอัตราส่วน F สำหรับการแจกแจงของคะแนนที่ไม่เป็นปกติ (Non - Normal Score Distribution) ไม่แตกต่างกับการแจกแจงของ F ในทางทฤษฎี ดังนั้นประสิทธิภาพของอัตราความคลาดเคลื่อนประเภทที่ 1 จากผลการทดลอง (Actual Type I Error) ในกรณีนี้จึงไม่ผิดไปจากอัตราความคลาดเคลื่อนที่ระบุ (Nominated α) และนอร์ตันได้ศึกษาผลงานของความเบ้ (Skewness) และความโด่ง (Kurtosis)ที่มีประสิทธิภาพของ อัตราความคลาดเคลื่อนประเภทที่ 1 จากผลการทดลอง พบว่า เมื่อข้อมูลมีการแจกแจงแบบโลจิสติก (Logistic) อัตราความคลาดเคลื่อนประเภทที่ 1 จากผลการทดลองมีค่าประมาณ .08 และ .03 เมื่อเปรียบเทียบกับค่าอัตราความคลาดเคลื่อนที่ระบุเท่ากับ .05 และ .01 ตามลำดับ ส่วนการแจกแจงแบบ highly Platykurtic ค่าอัตราความคลาดเคลื่อนประเภทที่ 1 จากผลการทดลอง มีค่าใกล้เคียงกับค่าอัตราความคลาดเคลื่อนที่ระบุ นอกจานี้ นอร์ตัน พบว่า การทดสอบ F มีความแกร่งสำหรับ

ข้อมูลที่ฝ่าฝืนข้อตกลงเบื้องต้น ที่เกี่ยวกับประชากรที่มีการแจกแจงแบบปกติ หรือที่เกี่ยวกับความเป็นเอกพันธุ์ของความแปรปรวนคลาดเคลื่อนของประชากร หรือทั้ง 2 อย่าง

Pearson (1931) ในหนังสือ Lindquist (Kirk 1969 : 60 - 61) เสนอแนะว่าประชากรอาจแตกต่างไปจากการแจกแจงแบบปกติในลักษณะที่มีความเบ้ (Skewness) และความโด่ง (Kurtosis) หรือทั้ง 2 ลักษณะได้และเมื่อนำมาวิเคราะห์ความแปรปรวน พบว่า การแจกแจงของ F จะไม่ได้รับความกระทบกระเทือนมากนักเมื่อประชากรไม่สมมาตร และถึงแม้จะมีความเบ้หรือความโด่งก็ตาม ยกเว้นในกรณีที่แตกต่างกันออกไปมาก เช่น ประชากรมีการแจกแจงโด่งมากหรือแบนมากผิดปกติในโมเดลผลการทดลองตามกำหนด (Fixed - Effect Model) ถึงแม้ว่าประชากรจะมีการแจกแจงแตกต่างจากปกติมากพอประมาณก็ตามแต่ถ้าประชากรมีการแจกแจงรูปร่างคล้ายกันมาก เช่น เบ้ทางบวกทุกกลุ่ม การวิเคราะห์ความแปรปรวนก็ยังคงมีความแกร่ง (Robust)

การแปลงข้อมูล (Transformation of Data)

เป็นการเปลี่ยนแปลงข้อมูลอย่างมีระบบ เพื่อมุ่งหวังที่จะช่วยให้อำนาจการทดสอบมีค่าสูงสุด (Kirk 1969 : 556) จุดประสงค์หลักของการแปลงข้อมูลก็เพื่อทำให้ข้อมูลที่ได้จากการวิจัยเป็นไปตามข้อตกลงเบื้องต้นของการทดสอบที่ต้องการใช้ เช่น แบบทดสอบเอฟ

Wayne Lee (1975 : 288) กล่าวว่าจุดประสงค์ของการแปลงข้อมูลมี 2 ประการ คือ

1. เพื่อทำให้ข้อมูลมีการแจกแจงปกติ และค่าความแปรปรวนมีความเป็นเอกพันธุ์
2. กำจัดเทอมของแรงปะทะ (Interaction) เพื่อจะได้ข้อมูลเป็นโมเดลที่มีความเป็นบวก (Additive Model)

Kirk (1969 : 63) ระบุว่าจุดประสงค์ของการแปลงข้อมูลมี 3 ประการ คือ

1. เพื่อให้ได้ข้อมูลมีความแปรปรวนความคลาดเคลื่อนเป็นเอกพันธุ์
2. เพื่อให้ได้การแจกแจงข้อมูลในระดับของการทดลองหรือการแจกแจงภายในเซลเป็นปกติ (Normality of Treatment - Level Distribution or Within - Cell Distribution)

3. เพื่อให้ได้ผลของการทดลองมีความเป็นบวก(Additivity of Treatment Effect)

Schlesselman (1973 : 369) กล่าวว่าจุดประสงค์ของการแปลงข้อมูลมี 2 ประการ คือ

1. เพื่อให้ได้ข้อมูลที่ถูกต้องโดยประมาณตามทฤษฎี (Theoretical Approximation)

2. เพื่อจัดการกับข้อมูลเพื่อให้เป็นไปตามข้อตกลงเบื้องต้นภายใต้การวิเคราะห์ข้อมูลตามระเบียบแบบแผน (Conventional Analysis)

ประโยชน์ของการแปลงข้อมูล

การแปลงข้อมูลในรูปแบบต่าง ๆ มีประโยชน์ดังนี้ คือ (Krukal and Tanur 1978 : 1044 – 1048)

1. ทำให้ความสัมพันธ์ของตัวแปร 2 ตัว เป็นเส้นโค้งที่แก้ไขได้ง่าย คือ อยู่ในลักษณะของเส้นตรง หรือตระกูลของเส้นโค้ง

2. ทำให้ค่าความแปรปรวนของข้อมูลมีค่าคงที่ เพื่อสะดวกในการวิเคราะห์ความแปรปราน

3. เทคนิกของการวิเคราะห์ความแปรปรวนจะถูกต้องถ้าข้อมูลมีการแจกแจงปกติ ดังนั้นการแปลงข้อมูลจะช่วยปรับปรุงลักษณะการแจกแจงของประชากรให้ดีขึ้น ตัวอย่างเช่นถ้าข้อมูลมีการแจกแจงเบ้ขวา การแปลงข้อมูลโดยใช้ล็อกฐาน 10 จะช่วยให้การแจกแจงเป็นปกติ

4. การแปลงข้อมูลในรูปแบบต่าง ๆ ช่วยในด้านการตีความของข้อมูล (Interpretation) เพื่อให้การวิเคราะห์ภายหลังได้ข้อมูลที่ถูกต้อง และมีความหมายชัดเจนยิ่งขึ้น

5. ถ้าข้อมูลอยู่ในรูปแบบต่าง ๆ เช่น เศษส่วนหรือเปอร์เซนต์ การแปลงข้อมูลในรูปแบบต่าง ๆจะช่วยทำให้สามารถวิเคราะห์ความแตกต่างได้ถูกต้อง ตัวอย่างเช่น ถ้าข้อมูลมีการแจกแจงแบบพัวซอง (Poisson Distribution) ก็ใช้การแปลงข้อมูลโดยใช้รูทสองก่อนการวิเคราะห์ ซึ่งประโยชน์ที่จะได้รับ ก็คือทำให้ข้อมูลเป็นไปตามข้อตกลงเบื้องต้นของการวิเคราะห์ความแปรปรวน

6. ปรับปรุงข้อตกลงเบื้องต้นของการวิเคราะห์ความแปรปรวนที่ว่าผลการทดลองมีความเป็นบวก หมายความว่าช่วยทำให้ความแปรปรวนระหว่างระดับการทดลองแต่ละคู่มีค่าเท่ากัน

ลักษณะการแปลงข้อมูลในรูปแบบต่าง ๆ

การแปลงข้อมูลเชิงเส้นตรง (Linear Transformation)

การแปลงข้อมูลเชิงเส้นตรง เรียกอีกชื่อว่า Coding มีรูปแบบคือ $Y = c_1x + c_2$, $c_1 \neq 0$ หรือ $Y = c_1x$ โดย Y เป็นการแปลงรูปข้อมูลเชิงเส้นตรงของ x และ c_1 , c_2 เป็นค่าคงที่ เมื่อข้อมูลถูกแปลงรูปเชิงเส้นตรง ค่ามัชฌิมเลขคณิตและค่าความแปรปรวนของข้อมูลจะถูกเปลี่ยนไปด้วย ดังนี้คือ

$$\mu' = c_1\mu + c_2$$

และ $$\sigma^2 = c_1^2\sigma^2$$

โดยที่ μ และ σ^2 เป็นค่ามัชฌิมเลขคณิตและความแปรปรวนของข้อมูลเดิม (Kreyszig 1970 : 89 - 90)

เมื่อข้อมูลถูกแปลงรูปเชิงเส้นตรง จะไม่มีผลกระทบต่อการวิเคราะห์ความแปรปรวน เพราะทำให้ค่า Mean Square ของเศษและส่วนเปลี่ยนไปในอัตราที่เท่ากัน นอกจากนั้นการแปลงข้อมูลเชิงเส้นตรงยังไม่สามารถทำให้ความล้มเหลวของข้อตกลงเบื้องต้นของการทดสอบเอฟดีขึ้นได้

Non - Linear and Monotonic Transformation

การแปลงรูปแบบข้อมูล 2 ชนิด ดังกล่าวตามหัวข้อนี้ มีลักษณะที่เหมือนกัน แต่นักสถิติเรียกชื่อต่างกัน ตัวอย่างเช่น Kruskal and Tanur เรียกการแปลงข้อมูลแบบนี้ว่า Non - Linear Transformation แต่ Schlesselman เรียกว่า Monotonic Transformation การแปลงรูปทั้ง 2 แบบนี้ เป็นการแปลงข้อมูลที่เมื่อค่า x มีการเปลี่ยนแปลง ค่า $\hat{y}$ ก็มีการเปลี่ยนแปลงด้วย ซึ่งอาจอยู่ในลักษณะที่เปลี่ยนแปลงไปในทางเดียวกัน หรือในทางตรงกันข้ามก็ได้

<u>Monotonic Transformation</u> (Kruskal and Tanur 1978 : 1050) แบ่งออกเป็น 2 ลักษณะ คือ

1. Increasing Monotonic Transformation คือ การแปลงข้อมูลที่เมื่อค่า y มีค่าเพิ่มขึ้น เมื่อ x มีค่าเพิ่ม เมื่อนำไปเขียนกราฟ เส้นโค้งจะขึ้นเรื่อย ๆ ไปทางขวา (Go To Right)

ตัวอย่างเช่น $Y = \log x$

$Y = \ln x$

$Y = \sqrt{x}$

2. Decreasing Monotonic Transformation คือ การแปลงข้อมูลที่ค่า y จะลดลง เมื่อ x มีค่าเพิ่มขึ้น ตัวอย่างเช่น $Y = 1/x$

<u>Non - Linear Transformation</u>

การแปลงข้อมูลแบบ Non - Linear ที่น่าใจ มีดังนี้ คือ

1 การแปลงข้อมูลโดยใช้รูทสอง (Square Root Transformation) รูปแบบคือ $Y' = \sqrt{y}$ เมื่อ Y' เป็นข้อมูลที่ถูกแปลงขึ้นโดยใช้รูทสอง <u>มักใช้กับข้อมูลที่มีการแจกแจงแบบพัวซอง</u> (Poisson Distribution) โดยค่ามัชฌิมเลขคณิตและความแปรปรวนของทรีทเมนต์เป็นสัดส่วนกันคือ $\sigma^2 y = k\bar{Y}$ การแจกแจงแบบนี้เป็นผลเนื่องจากตัวแปรตาม (Dependent Variable) เป็นความถี่ของเหตุการณ์ที่มีโอกาสเกิดขึ้นได้น้อยมาก เช่น การนับจำนวนเหตุการณ์ผิดปกติหรืออุบัติเหตุต่าง ๆ เป็นต้น (Kirk 1969 : 64 – 65) การแปลงข้อมูลโดยวิธีนี้จะมีประสิทธิภาพ <u>ถ้าความแปรปรวนของข้อมูลมีค่าเท่ากับ 0.25 ถ้าข้อมูลมีขนาดเล็กมาก</u> Freeman และ Turkey (1950) ได้เสนอให้ใช้รูปแบบการแปลงดังนี้คือ $Y' = \sqrt{y + 1}$ หรือ $Y' = \sqrt{y} + \sqrt{y+1}$ ซึ่งจะให้ผลดีกว่า และต่อมาในปี ค.ศ. 1954 Mosteller และ Bush ได้สร้างตาราง $\sqrt{y} + \sqrt{y + 1}$ ขึ้น เพื่อสะดวกในการคิดคำนวณ

การแปลงข้อมูลโดยใช้รูทสอง มีผลต่อรูปร่างของการแจกแจงความถี่ของความคลาดเคลื่อน (Shape of the frequency Distribution of Error) และการรวมกันได้ของผลการทดลอง (Additivity of Effects) ถ้าผลการทดลอง

ระหว่างทรีทเมนต์และกลุ่มมีค่าความเป็นบวกในข้อมูลเดิมเมื่อแปลงข้อมูลโดยใช้รูทสองแล้ว จะไม่พบค่าการรวมกันไ ..นั้นอีก (Snedecor and Cochran 1976 : 325 – 327) นอกจากนั้นยังช่วยให้ค่าเฉลี่ยและค่าความแปรปรวนเป็นอิสระจากกัน

2. การแปลงข้อมูลโดยใช้ล็อกฐาน 10(Logarithemic Transformation) รูปแบบคือ Y = log y หรือ $Y' = \log (y + 1)$ เมื่อ Y' เป็นข้อมูลที่ถูกแปลงโดยใช้ล็อกฐาน 10 โดยรูปแบบหลังใช้เมื่อข้อมูลบางตัวมีค่าน้อยมาก หรือเท่ากับ 0 (Kirk 1969 : 64) การแปลงข้อมูลโดยใช้ล็อกฐาน 10 ใช้ในกรณีที่ความแปรปรวนมีค่าเป็นสัดส่วนกำลัง 2 ของค่าเฉลี่ยของการทดลองหรืออีกนัยหนึ่ง สัมประสิทธิ์ของความแปรปรวนของค่าเฉลี่ยของทรีทเมนต์เป็นค่าคงที่ ซึ่งจะช่วยให้ค่าความแปรปรวนของข้อมูลมีค่าเท่ากัน นอกจากนี้จะทำให้อิทธิ..ลต่าง ๆ ที่มีผลแบบคูณ (Multiplicative) ในข้อมูลเดิมเปลี่ยนเป็นข้อมูลใหม่ที่มีผลความเป็นบวก (Additivity) (Steel and Torrie 1960 : 235)

การแปลงข้อมูลแบบนี้มีประสิทธิภาพสำหรับการแจกแจงปกติที่มีความเบ้เป็นบวก (Positive Skewness) ซึ่งการแจกแจงแบบนี้มักเกิดขึ้นกับงานวิจัยทางจิตวิทยา เมื่อเกณฑ์ที่วัดเป็นสเกลของเวลา เช่น จำนวนเวลาที่เป็นวินาทีที่ทำงานเสร็จ(Winer 1971:400)

3. การแปลงข้อมูลโดยใช้ล็อกฐานอี (Natural Logarithemic Transformation)

รูปแบบคือ $Y' = \ln y$ เมื่อ Y' เป็นข้อมูลที่ถูกแปลงโดยใช้ล็อกฐานอี หลักการและประโยชน์คล้ายกับการแปลงข้อมูลโดยใช้ล็อกฐาน 10

4. การแปลงข้อมูลโดยใช้วิธีกลับเศษส่วน (Reciprocal Transformation)

รูปแบบ คือ $Y' = 1/y$ เมื่อ Y' เป็นข้อมูลที่ถูกแปลงโดยวิธีกลับเศษส่วน ใช้เมื่อค่ากำลังสองของมัชฌิมเลขคณิตของทรีทเมนต์ และค่าส่วนเบี่ยงเบนมาตรฐานเป็นสัดส่วนกัน ถ้าข้อมูลมีค่าน้อยมาก หรือเท่ากับ 0 จะใช้รูปแบบดังนี้คือ $Y' = 1/(y+1)$ ซึ่งจะให้ผลที่ดีกว่า (Kirk 1969 : 66) การแปลงข้อมูลแบบนี้จะมีประสิทธิภาพ ถ้าตัวแปรเป็นเวลาที่ใช้ตอบโต้กลับ (Reaction)

5. การแปลงข้อมูลโดยใช้อาร์ค ..าย (Arcsine of Angular Transformation)

รูปแบบคือ $Y' = 2 \text{ arcsine } y$ เมื่อ Y' เป็นข้อมูลที่แปลงรูปโดยใช้อาร์คซาย และ y เป็นค่าสัดส่วน (Proportion) ผู้วิจัยไม่จำเป็นต้องคำนวณค่า Y' จากสูตร เพราะมีตารางสำเร็จสำหรับคำนวณค่า Y' เมื่อค่า Y มีค่าตั้งแต่ .001 ถึง .999 และค่า Y' มีหน่วยเป็นเรเดียน ในปี ค.ศ.1947 Bartlett เสนอแนะว่าสำหรับข้อมูลที่มีค่าใกล้ 0 หรือ 1 ให้ใช้รูปแบบการแปลงดังนี้คือ

$$Y' = 2 \text{ arcsine} \sqrt{y \pm 1/2n}$$

โดยที่ n เป็นจำนวนข้อมูล เครื่องหมาย + ใช้สำหรับค่า y ที่มีค่าใกล้ 0 และเครื่องหมาย – ใช้สำหรับค่า y ที่มีค่าใกล้ 1 (Winer 1971 : 399 – 400)

การแปลงข้อมูลแบบนี้จะมีประสิทธิภาพเมื่อค่ามัชฌิมเลขคณิตและความแปรปรวนเป็นสัดส่วนกัน ดังนี้ $\sigma^2_{ij} = \mu_{ij} (1 - \mu_{ij})$ และการแจกแจงเป็นแบบไบโนเมียล กรณีเช่นนี้ใช้เมื่อจำนวนของการทดลองคงที่ และ y เป็นความน่าจะเป็นในการสนองตอบซึ่งแปรเปลี่ยนไปตามระดับการทดลอง (Kirk 1969 : 66)

นอกจากรูปแบบการแปลงข้อมูลข้างต้นแล้ว (Kruskal and Tanur ได้แบ่งออกเป็น 2 ลักษณะ ดังนี้คือ

1. One – Bend Transformation เป็นการแปลงข้อมูลที่เมื่อนำมาเขียนกราฟ จะได้รูปกราฟมีลักษณะเป็นเส้นโค้ง มีเพียงโค้งเดียวเท่านั้น (Kruskal Tanur 1978 : 1051)

ตัวอย่างเช่น	$Y = \log (x+k)$	เรียกว่า Logarithemic Family
	$Y = \sqrt{x + k}$	เรียกว่า Square Root Family
	✓ $Y = x^c$	เรียกว่า Power Family
	✓ $Y = e^{cx}$	เรียกว่า Exponential Family
	$Y = (x + k)^c$	เรียกว่า Simple Family

โดยที่ k , c เป็นค่าคงที่ และ e = 2.718

2. Two – Bend Transformation เป็นการแปลงข้อมูลที่เมื่อนำมาเขียนกราฟ จะได้รูปกราฟมีลักษณะเป็นเส้นโค้ง มี 2 โค้ง (bend) (Kruskal and Tanur 1978 : 1051)

ตัวอย่างเช่น $Y = \arcsin x$ เรียกว่า Angular or Arcsine Transformation

$Y = 1/2 \log x/1 - x$ เรียกว่า Logit Transformation

$Y = 1/2 \log 1+r/1-r$ เรียกว่า Fisher ' s Z Transformation

$Y = \text{arc tanh } r$ เรียกว่า Hyperbolic Arctangent Transformation

การวิจัยที่เกี่ยวข้อง

Freeman และ Tukey (1939 : 607 – 611) ได้ทดสอบความสามารถในการแปลงข้อมูลโดยใช้รูทสอง (Square Root Transformation) ในลักษณะต่าง ๆ 5 แบบด้วยกัน คือ 1. $Y = \sqrt{x}$ 2. $Y = \sqrt{x+1}$ 3. $Y = \sqrt{x} + \sqrt{x+1}$ 4. $Y = \sqrt{x + 3/8}$ 5. $Y = \sqrt{x + 1/2}$ เมื่อ $\bar{x}$ เป็นข้อมูลที่มีการแจกแจงแบบพัวซอง (Poisson Distribution) เพื่อเปรียบเทียบดูว่าการแปลงข้อมูลโดยใช้รูทสองในลักษณะแบบใดที่ทำให้ค่าความแปรปรวนความคลาดเคลื่อนมีค่าคงที่ โดยใช้กลุ่มตัวอย่างขนาดต่าง ๆ กัน คือ 5, 10, 20 ผลปรากฏว่า การแปลงข้อมูลโดยใช้รูทสองแบบ $Y = \sqrt{x} + \sqrt{x+1}$ ให้ผลการวิเคราะห์ที่ดีที่สุด

Bartlett (Kempthrone 1952 : 156) เสนอแนะว่า การที่ผู้วิจัย จะวิเคราะห์ความแปรปรวนเพื่อสรุปผลงานการวิจัย จะต้องคำนึงถึงข้อมูลว่าข้อมูลที่ได้มานั้นเป็นไปตามข้อตกลงเบื้องต้นหรือไม่ โดยเฉพาะข้อตกลงเบื้องต้นที่เกี่ยวกับความเป็นเอกพันธุ์ของความแปรปรวน ถ้าข้อมูลไม่เป็นไปตามข้อตกลงเบื้องต้น เขาให้เปลี่ยนข้อมูลในรูปแบบต่าง ๆ เสียก่อนที่จะวิเคราะห์ โดยพิจารณาความสัมพันธ์ระหว่างค่าความแปรปรวนและมัชฌิมเลขคณิตเป็นหลักในการแปลงข้อมูล และได้เสนอตารางการแปลงข้อมูล ดังนี้

ค่าความแปรปรวนในเทอมมัชฌิมเลขคณิต (m)	วิธีการแปลงข้อมูล	ความแปรปรวนของข้อมูลที่แปลง
m , λm	$\sqrt{x}$ or $\sqrt{x + 1/2}$	0.25
$\lambda^2 m^2$	$\log_e x$, $\log_{10} x$, $\log_{10}(x + 1)$	λ^2 or $0.189 \lambda^2$
$\frac{2 m^2}{n - 1}$	$\log_e x$	$2/(n - 1)$
$\frac{m(1 - m)}{n}$	$\sin^{-1} x$(degree)	$821/n$
	$\sin^{-1} x$(radian)	$0.25/n$
$\lambda^2 m^2(1 - m^2)$	$\log_e \frac{x}{1 - x}$	λ^2
$\frac{(1 - m^2)^2}{n - 1}$	$1/2 \log_e \frac{(1 + x)}{(1 - x)}$	$1/(n - 3)$

Steel และ Torrie (1960 : 236 – 237) เสนอแนะว่า การแปลงข้อมูลมีประโยชน์มากสำหรับข้อมูลที่มีขนาดเล็ก เพื่อค้นหาว่ามีค่าสหสัมพันธ์ระหว่างค่ามัชฌิมเลขคณิตของกลุ่มทดลองและค่าความแปรปรวนภายในกลุ่มหรือไม่ (Correlation between the treatment Mean and their Within Treatment Variance) แต่ถ้าพบว่าค่าความแปรปรวนของข้อมูลมีค่าแตกต่างกันเพียงเล็กน้อย ค่าของการแปลงข้อมูลด้วยวิธีต่าง ๆ ก็อาจไร้ผล คืออาจจะไม่ให้ผลลัพธ์ที่ดีที่สุดเสมอไป ทำให้ผู้อ่านไม่กระจ่างในรายงานผลการวิจัย ดังนั้นจึงจำเป็นต้องเปลี่ยนข้อมูลที่ถูกแปลงนั้นกลับคืนสู่ข้อมูลเดิม

Ekeblad (1962 : 200 – 202) ได้ทดสอบความสามารถในการแปลงการแจกแจงของข้อมูลที่เบ้ ให้เป็นข้อมูลที่มีการแจกแจงอย่างสมมาตร โดยใช้วิธีการแปลงข้อมูลโดยใช้ล็อกฐาน 10 พบว่า ด้วยกลุ่มตัวอย่างขนาดเล็ก และมีการแจกแจงที่เบ้ มัชฌิมเลขคณิตมีค่ามากกว่ามัธยฐานหลายเท่า สามารถทำให้ข้อมูลมีการแจกแจงปกติได้ โดยวิธีการทางคณิตศาสตร์ (Standard Mathematical Procedure) คือใช้วิธีการแปลงข้อมูลโดยใช้ล็อกฐาน 10 และได้เสนอแนะว่าวิธีการแปลงข้อมูลโดยใช้รูทสองและโดยวิธีกลับเศษส่วน ก็สามารถทำให้ข้อมูลมีการแจกแจงปกติได้

Budescu และ Appelbaum (1981 : 55 – 74) ใช้เทคนิคของมอนติคาร์โล ซิมมูเลชั่น เพื่อศึกษาผลของการแปลงข้อมูลเพื่อให้ได้ค่าความแปรปรวนมีเสถียรภาพ (Variance Stabilizing Transformation) ที่มีต่ออำนาจการทดสอบเอฟ โดยจำแนกการแจกแจงของประชากรเป็น 5 สถานการณ์ คือ แบบไบโนเมียล แบบปกติที่มีค่ามัชฌิมเลขคณิตและค่าความแปรปรวนเท่ากับการแจกแจงแบบไบโนเมียล แบบพัวซอง แบบปกติที่มีค่ามัชฌิมเลขคณิตและค่าความแปรปรวนเท่ากับการแจกแจงแบบพัวซอง และการแจกแจงปกติ ใช้กลุ่มตัวอย่าง 4 กลุ่มมีขนาดไม่เท่ากัน ด้วยแผนการทดลองแบบ CR – k ทำการจำลองการทดลองซ้ำ 1500 ครั้ง พบว่า การแปลงข้อมูลไม่มีผลต่ออำนาจการทดสอบเอฟที่ระดับนัยสำคัญ .01 และ .05 นอกจากนี้ยังทำให้อำนาจการทดสอบเอฟมีค่าลดลง โดยเฉพาะในกรณีกลุ่มตัวอย่างขนาดเล็กมาจากประชากรที่มีค่าความแปรปรวนสูง และกลุ่มตัวอย่างขนาดใหญ่มาจากประชากรที่มีค่าความแปรปรวนต่ำ

Srisukho (1974 : 1 – 3) ใช้เทคนิคของมอนติคาร์โล ซิมมูเลชั่น เพื่อศึกษาอำนาจการทดสอบเอฟจากข้อมูลที่มีการแจกแจงของประชากร 6 รูปแบบ คือ แบบปกติ, แบบยูนิฟอร์ม, แบบ Double Exponential แบบ Moderate Negatively Skewed แบบ Extreme Negatively Skewed และแบบ Bimodal ขนาดกลุ่มตัวอย่างเท่ากับ 5, 10, 15 และ 20 ทำการจำลองการทดลองซ้ำ 1000 ครั้ง และตั้งระดับนัยสำคัญที่ α เท่ากับ .01 และ .05 พบว่า อำนาจการทดสอบเอฟมีความแกร่ง (Robust) สำหรับข้อมูลที่มีการฝ่าฝืนข้อตกลงเบื้องต้นเกี่ยวกับประชากรมีการแจกแจงแบบปกติ นอกจากนั้นยังพบว่าอัตราความคลาดเคลื่อนประเภทที่ 1 จากผลการทดลองมีค่ามากกว่าอัตราความคลาดเคลื่อนที่ระบุในระดับนัยสำคัญเท่ากับ .01 เมื่อกลุ่มตัวอย่างมีขนาดเท่ากับ 5 และ 10 ซึ่งมีการแจกแจงแบบยูนิฟอร์ม และกลุ่มตัวอย่างขนาดเท่ากับ 15 ที่มีการแจกแจงแบบ Bimodal และแบบเบ้ (Skewed Distribution)