บทที่ 2
เอกสารงานวิจัยที่เกี่ยวข้อง

ในการวิจัยครั้งนี้ทำการศึกษาอำนาจของการทดสอบ ของการวิเคราะห์ความแปรปรวนแบบจำแนกทางเดียว เมื่อฝ่าฝืนข้อกำหนดเบื้องต้นเกี่ยวกับการเท่ากันของความแปรปรวนของความคลาดเคลื่อน ซึ่งในบทนี้จะกล่าวถึงวรรณกรรมต่างๆที่เกี่ยวข้องกับงานวิจัย รวมทั้งนำเสนอผลงานวิจัยที่เกี่ยวข้อง ดังรายละเอียดต่อไปนี้

2.1 ทฤษฎีที่เกี่ยวข้อง

2.1.1 การวิเคราะห์ความแปรปรวนแบบจำแนกทางเดียว

การวิเคราะห์ความแปรปรวนแบบจำแนกทางเดียว (One-Way Analysis of Variance) เป็นการจำแนกข้อมูลด้วยตัวแปรเดียว นั่นคือวิเคราะห์ความแตกต่างของข้อมูลโดยพิจารณาจากปัจจัยที่มีผลต่อข้อมูลเพียงปัจจัยเดียว หรือเป็นการวิเคราะห์ความแตกต่างของระดับต่างๆของปัจจัยโดยมีวัตถุประสงค์เพื่อทดสอบความแตกต่างระหว่างค่าเฉลี่ยของประชากรที่ได้รับปัจจัยที่ระดับต่างๆ โดยถือว่าหน่วยที่ได้รับปัจจัยระดับเดียวกันเป็นประชากรหนึ่งๆ และหน่วยที่ได้รับปัจจัยคนละระดับเป็นคนละประชากร (กัลยา วานิชย์บัญชา, 2545) ดังโครงร่างข้อมูลในตารางที่ 2.1

ตารางที่ 2.1 โครงร่างข้อมูลของแต่ละประชากร ในการวิเคราะห์ความแปรปรวนแบบจำแนกทางเดียว

ค่าสังเกต	ประชากรที่					
	1	2	…	i	…	t
1	Y_{11}	Y_{21}		Y_{i1}		Y_{t1}
2	Y_{12}	Y_{22}		Y_{i2}		Y_{t2}
.	.	.		.		.
.	.	.		.		.
.	.	.	…	.	…	.
j	Y_{1j}	Y_{2j}		Y_{ij}		Y_{tj}
.	.	.		.		.
.	.	.		.		.
.	.	.		.		.
n_j	Y_{2n_1}	Y_{2n_2}		Y_{in_i}		Y_{tn_t}
ขนาดตัวอย่าง	n_1	n_2	…	n_i	…	n_t
ค่าเฉลี่ยตัวอย่าง	$\overline{Y}_{1.}$	$\overline{Y}_{2.}$	…	$\overline{Y}_{i.}$	…	$\overline{Y}_{t.}$

จากตารางตัวอย่างขนาด n_i ในแต่ละประชากรถูกสุ่มมาโดยอิสระจาก t ประชากร ที่มีค่าเฉลี่ย $\mu_1, \mu_2, ... \mu_t$ ตามลำดับ โดยค่าสังเกตที่ j จากประชากรที่ i หรือ Y_{ij} มีค่าเท่ากับค่าเฉลี่ยของประชากรกลุ่มที่ i หรือ μ_i บวกกับค่าเบี่ยงเบนไปจากค่าเฉลี่ย μ_i ซึ่งเรียกว่า ความคลาดเคลื่อน (error) ซึ่งเขียนแทนด้วย ε_{ij} โดย ε_{ij} เป็นตัวแปรสุ่ม นั่นคือ

$$Y_{ij} = \mu_i + \varepsilon_{ij}$$

หรือ
$$Y_{ij} = \mu + (\mu_i - \mu) + \varepsilon_{ij}$$

เมื่อ
$$\mu = \frac{\sum_{i=1}^{t} \mu_i}{t}$$ คือ ค่าเฉลี่ยรวม

$$\mu_i - \mu = \tau_i$$ คือ ผลกระทบของสิ่งทดลองที่ i

ดังนั้นตัวแบบเชิงเส้นทางสถิติ (Linear Statistical Model) มีรูปแบบดังนี้

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij} \quad , \quad i = 1,2,\ldots,t$$
$$j = 1,2,\ldots,n_i$$

โดยที่ Y_{ij} คือ ค่าสังเกตที่ได้จากสิ่งทดลองที่ i ซ้ำที่ j

μ คือ ค่าเฉลี่ยรวม

τ_i คือ อิทธิพลของสิ่งทดลอง (treatment effect) ที่ i โดยที่

$$\tau_i = \mu_i - \mu$$

ε_{ij} คือ ความคลาดเคลื่อนของการทดลอง (experimental error) จากสิ่งทดลองที่ i ซ้ำที่ j โดย $\varepsilon_{ij} \sim NID(0, \sigma^2)$

t คือ จำนวนของสิ่งทดลองหรือจำนวนกลุ่มประชากร

n_i คือ จำนวนซ้ำหรือขนาดตัวอย่างของกลุ่มที่ i

ตัวแบบการวิเคราะห์ความแปรปรวนแบบจำแนกทางเดียวมีข้อสมมติ (assumption) ของตัวแปรสุ่ม ε_{ij} ว่าในแต่ละสิ่งทดลอง ε_{ij} มีการแจกแจงแบบปกติและเป็นอิสระกัน (normally and independently distributed) โดยมีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ σ^2 ซึ่งมีค่าเท่ากันแต่ไม่ทราบค่า เขียนย่อได้เป็น $\varepsilon_{ij} \sim NID(0,\sigma^2)$ นอกจากข้อสมมติเกี่ยวกับตัวแปรสุ่ม ε_{ij} แล้วยังถือว่า τ_i ในตัวแบบเป็นผลกระทบชนิดคงที่ (fixed effects) ซึ่งหมายความว่าประชากรที่ 1,2,…,t ได้ถูกกำหนดไว้แล้ว ดังนั้นผลของการทดสอบสมมติฐานจึงครอบคลุมเฉพาะประชากรต่างๆที่ได้กำหนดไว้แต่แรกเท่านั้น มิได้ขยายความไปถึงประชากรอื่นๆที่ไม่ได้ทำการ

ทดลอง จากการที่ τ_i เป็นผลกระทบชนิดคงที่ ทำให้ $\sum_{i=1}^{t}\tau_i=\sum_{i=1}^{t}(\mu_i-\mu)=0$ และเรียกตัวแบบข้างต้นว่า ตัวแบบผลกระทบชนิดคงที่ (fixed effects model) (ชัชวาลย์ เรืองประพันธ์, 2543)

จากตัวแบบผลกระทบชนิดคงที่และข้อสมมติเกี่ยวกับตัวแปรสุ่ม ε_{ij} จะได้ว่าในแต่ละสิ่งทดลอง Y_{ij} เป็นตัวแปรสุ่มที่มีการแจกแจงแบบปกติและเป็นอิสระกันที่มี

ค่าเฉลี่ย $= E[Y_{ij}]$

$= E[\mu_i+\varepsilon_{ij}]$

$= \mu_i$

$= \mu+\tau_i$

ความแปรปรวน $= \mathrm{var}[Y_{ij}]$

$= \mathrm{var}[\mu+\tau_i+\varepsilon_{ij}]$

$= \mathrm{var}[\varepsilon_{ij}]$

$= \sigma^2$ ซึ่งไม่ทราบค่าและมีค่าเท่ากันทุกประชากร

นั่นคือความแปรปรวน Y_{ij} และความแปรปรวน ε_{ij} มีค่าเท่ากัน และสามารถเขียนได้ว่า $Y_{ij} \sim NID(\mu_i,\sigma^2)$ หรือ $Y_{ij} \sim NID(\mu+\tau_i,\sigma^2)$ โดยสมมติฐานในการทดสอบค่าเฉลี่ยของ t ประชากร หรือสิ่งทดลอง คือ

$H_0: \mu_1=\mu_2=...=\mu_t$ เทียบกับ

H_1 : มีค่าเฉลี่ยอย่างน้อย 2 กลุ่มที่แตกต่างกัน

โดยที่ μ_i แทนค่าเฉลี่ยของประชากรกลุ่มที่ i ; i = 1,2,..., t

ซึ่งมีความหมายเช่นเดียวกับ สมมติฐาน

$H_0: \tau_1=\tau_2=...=\tau_t=0$ เทียบกับ

H_1 : τ_i อย่างน้อยหนึ่งค่าไม่เท่ากับศูนย์

สำหรับวิธีการทดสอบสมมติฐานในการวิเคราะห์ความแปรปรวนซึ่งเป็นวิธีการทดสอบความแตกต่างระหว่างค่าเฉลี่ยของประชากร จะใช้สถิติทดสอบเอฟ สำหรับวิธีการในการวิเคราะห์จะแบ่งความผันแปร (Variation หรือ Sum of Squares) ทั้งหมดของข้อมูลออกเป็นส่วนๆ ตามแหล่งความผันแปรที่แตกต่างกัน จากนั้นจะหาค่าประมาณของ σ^2 จากแหล่งความผันแปรต่างๆ ในการดำเนินการทดสอบ เราจะนำค่าประมาณของ σ^2 มาเปรียบเทียบกัน สำหรับการแจกแจงเอฟ (F – distribution) ได้มีการเสนอโดย Sir R.A. Fisher

โดยมี สูตร $$Z = \frac{1}{2} \ln \frac{S_1^2}{S_2^2}$$

ซึ่งค่าลอการิทึม (logarithm) ทำให้ยุ่งยากในการคำนวณ W. Snedecor จึงปรับปรุงใหม่

โดยใช้สูตร $$F = \frac{S_1^2}{S_2^2}$$

และค่า S_1^2 และ S_2^2 คือค่าความแปรปรวนของแหล่งความผันแปรระหว่างกลุ่ม และแหล่งความผันแปรภายในกลุ่ม ตามลำดับ ดังนั้น Snedecor ได้ตั้งชื่อการแจกแจงใหม่นี้ว่า การแจกแจงเอฟ เพื่อเป็นเกียรติแก่ R. A. Fisher

2.1.2 ความคลาดเคลื่อนของการทดสอบสมมติฐานเชิงสถิติ

จากการทดสอบสมมติฐานจะมีการตัดสินใจอยู่ 2 ทาง คือ จะยอมรับสมมติฐานหลัก หรือ จะปฏิเสธสมมติฐานหลัก ถ้าทางหนึ่งเกิดขึ้นแล้วอีกทางหนึ่งจะไม่เกิดขึ้น กล่าวคือ การตัดสินใจจะเกิดได้อย่างเดียวไม่สามารถตัดสินใจพร้อมๆกันทั้งสองทางได้ ซึ่งจะเป็นทางใดนั้นกระทำได้โดยการทดสอบ (test) สำหรับการทดสอบสมมติฐานเชิงสถิติ การที่จะยอมรับสมมติฐานหลัก หมายความว่า ค่าสถิติที่คำนวณได้จากตัวอย่างสนับสนุนให้ยอมรับสมมติฐานหลัก คือทำให้โอกาสหรือความน่าจะเป็นที่สมมติฐานหลักจะเป็นจริงได้สูงพอเท่านั้น แต่มิได้หมายความว่าการยอมรับสมมติฐานหลักจะถูกต้องตามหลักการวิทยาศาสตร์เสมอไปทุกครั้ง ในทำนองเดียวกันการปฏิเสธสมมติฐานหลัก ก็มิได้หมายความว่า สมมติฐานหลักไม่เป็นจริงเสมอไป ดังนั้นในทางสถิติการยอมรับหรือปฏิเสธสมมติฐานหลักจึงต้องมีความคลาดเคลื่อนที่จะตัดสินใจผิดอยู่บ้าง แต่เราก็พยายามที่จะทำให้ความคลาดเคลื่อนในการตัดสินใจผิดอยู่ในระดับที่ไม่สูงเกินไปนัก สำหรับความคลาดเคลื่อนจากการตัดสินใจในการทดสอบสมมติฐาน แบ่งออกเป็น 2 ประเภท คือ

2.1.2.1 ความคลาดเคลื่อนประเภทที่ 1 (Type I Error : α) เป็นความคลาดเคลื่อนที่เกิดขึ้นจากกรณีสมมติฐานหลักที่ตั้งไว้เป็นสมมติฐานที่ถูกต้อง ซึ่งควรยอมรับว่าถูกต้อง แต่หลักฐานที่ได้จากทดสอบสมมติฐานโดยอาศัยข้อมูลที่สังเกตได้จากตัวอย่างทำให้ปฏิเสธหรือไม่ยอมรับสมมติฐานหลักที่ตั้งไว้ (ไปยอมรับสมมติฐานรองแทน) และถ้าให้โอกาสที่เกิดความคลาดเคลื่อนประเภทที่ 1 มีค่าเท่ากับ α สามารถเขียนอยู่ในรูปฟังก์ชันความน่าจะเป็นได้ดังนี้

$$\alpha = P(\text{ปฏิเสธ } H_0 / H_0 \text{ เป็นจริง})$$

2.1.2.2 ความคลาดเคลื่อนประเภทที่ 2 (Type II Error : β) เป็นความคลาดเคลื่อนที่เกิดขึ้นจากกรณีสมมติฐานหลักที่ตั้งไว้แท้จริงแล้วเป็นสมมติฐานที่ผิดซึ่งควรปฏิเสธ แต่หลักฐานที่ได้จากข้อมูลที่สังเกตได้จากตัวอย่างทำให้ยอมรับสมมติฐานหลักที่ตั้งไว้ (ปฏิเสธสมมติ

ฐานรอง) และถ้าให้โอกาสที่เกิดความคลาดเคลื่อนประเภทที่ 2 มีค่าเท่ากับ β สามารถเขียนอยู่ในรูปฟังก์ชันความน่าจะเป็นได้ดังนี้

$$\beta = P(\text{ปฏิเสธ } H_1 / H_1 \text{ เป็นจริง})$$

ทั้งนี้ความคลาดเคลื่อนทั้ง 2 แบบ แม้จะมีโอกาสเกิดขึ้นแต่ไม่มีโอกาสที่จะเกิดขึ้นพร้อมๆกันได้ กล่าวคือ กรณีที่ตัดสินใจปฏิเสธสมมติฐานหลักเท่านั้นที่จะมีโอกาสเกิดความคลาดเคลื่อนประเภทที่ 1 และกรณีที่ตัดสินใจยอมรับสมมติฐานหลักเท่านั้นที่เสี่ยงต่อการเกิดความคลาดเคลื่อนประเภทที่ 2 อย่างไรก็ตามความคลาดเคลื่อนทั้งสองประเภท เมื่อเกิดขึ้นแล้วมีผลเสียต่องานทัดเทียมกันและเราพยายามให้เกิดความคลาดเคลื่อนน้อยที่สุดเท่าที่จะเป็นไปได้ ซึ่งเราสามารถควบคุมความคลาดเคลื่อนประเภทที่ 2 (β) ได้ โดยเลือกตัวแบบทดสอบที่ทำให้เกิดความคลาดเคลื่อนต่ำสุด เพราะฉะนั้นหน้าที่ของผู้ใช้จึงเหลือเพียงการควบคุมมิให้เกิดความคลาดเคลื่อนประเภทที่ 1(α) มีค่าสูงเกินไป แต่ไม่ควรควบคุมให้ค่า α ต่ำมากนัก เพราะมีผลทำให้ β มีค่าสูงขึ้นตาม ทั้งนี้เนื่องจากค่า α และ β มีความสัมพันธ์ในลักษณะที่เป็นสัดส่วนผกผัน (inverse relationship) (อโนทัย ตรีวานิช, 2539) เมื่อพิจารณาความคลาดเคลื่อนประเภทที่ 1 ก็อาจจะเกิดขึ้นเนื่องจากข้อมูลที่นำมาวิเคราะห์ ไม่เป็นไปตามข้อกำหนดเบื้องต้นหรือข้อสมมติ (assumption) ซึ่งจะทำให้เกิดความคลาดเคลื่อนบางค่ามีค่าสูงมากๆ หรือต่ำมากๆ จนค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 เกินกว่าระดับนัยสำคัญ (α) ที่กำหนดไว้

2.1.3 การจำลองการทดลองโดยอาศัยเทคนิคมอนติคาร์โล

การศึกษาเกี่ยวกับการควบคุมความคลาดเคลื่อนประเภทที่ 1 นั้น ได้ทำการศึกษาโดยการทำการทดลอง โดยใช้เทคนิคการจำลองที่เรียกว่า เทคนิคมอนติคาร์โล (Monte Carlo Technique) ซึ่งเป็นเทคนิคที่ใช้แก้ปัญหาในการคำนวณทางคณิตศาสตร์และเป็นเทคนิคที่นิยมใช้กันอย่างแพร่หลายในปัจจุบันโดยจะใช้ตัวเลขสุ่ม (random number) มาช่วยในการหาคำตอบของปัญหาที่ต้องการศึกษา ซึ่งจะทำให้ได้ผลสรุปจากสภาพการณ์ที่สร้างขึ้นในการทดลอง โดยสามารถกำหนดขนาดตัวอย่าง ลักษณะการแจกแจงของความคลาดเคลื่อน ค่าเฉลี่ยและความแปรปรวนได้ตามความต้องการของผู้ศึกษา ซึ่งขั้นตอนของเทคนิคมอนติคาร์โล นิยมดำเนินการเป็นขั้นตอนดังนี้ (สุพรรณี อึ้งปัญสัตวงศ์, 2541)

2.1.3.1 การสร้างตัวเลขสุ่ม เนื่องจากการใช้ตัวเลขสุ่มเป็นสิ่งที่สำคัญมากในเทคนิคมอนติคาร์โล ทั้งนี้เพราะ หลักการของเทคนิคมอนติคาร์โลนั้นจะใช้ตัวเลขสุ่มมาช่วยในการหาคำตอบให้กับปัญหาลักษณะของตัวเลขสุ่มจะมีการแจกแจงแบบยูนิฟอร์มในช่วง 0 ถึง 1 (Uniform(0,1))

π = 3.14159

e = 2.71828

x = ค่าของข้อมูลจากกลุ่มตัวอย่าง

ค่า μ และ σ เป็นพารามิเตอร์ ที่บอกถึงลักษณะของประชากรว่า ประชากรนั้นมีค่าเฉลี่ยอยู่ที่ใด และมีการกระจายมากน้อยเพียงใด ซึ่งลักษณะและคุณสมบัติของการแจกแจงแบบปกติ มีดังนี้

- ลักษณะของเส้นโค้งเป็นรูประฆังคว่ำ (Bell Shaped)
- เส้นแบ่งครึ่งโค้งจะอยู่ที่ค่าเฉลี่ยของข้อมูล และเส้นนี้ทำให้เส้นโค้งที่อยู่สองข้างมีลักษณะสมมาตร (Symmetry)
- จุดที่เป็นค่าเฉลี่ย มัธยฐาน และฐานนิยม เป็นจุดเดียวกันหรือมีค่าเท่ากัน
- มีความโด่ง (kurtosis) ของเส้นโค้งเท่ากับ 3 ซึ่งเรียกว่า เมโซเคอร์ติค (Mesokurtic) และจุดเปลี่ยนโค้งทั้งสองข้างจะอยู่ ณ ระยะ 1 เท่าของส่วนเบี่ยงเบนมาตรฐาน
- ค่าความเบ้ (Skewness) เท่ากับศูนย์
- ปลายทั้งสองข้างของเส้นโค้งจะค่อย ๆ ลดต่ำลง แต่ไม่จรดกับฐานของเส้นโค้งหรือแกนนอน
- ถ้าลากเส้นตั้งฉากจากแกน x (ซึ่งเป็นแกนนอน) ไปยังเส้นโค้งโดยที่เส้นดังกล่าวห่างจากจุดเฉลี่ย (μ) ทั้งด้านซ้ายและด้านขวาด้วยระยะหนึ่งเท่า, สองเท่า และสามเท่าของค่าเบี่ยงเบนมาตรฐาน (σ) พื้นที่ที่ปิดกั้นเส้นตั้งฉากกับเส้นโค้งจะเท่ากับ 68% 95% และ 99.7% ของพื้นที่ทั้งหมดตามลำดับ
- พารามิเตอร์ μ และ σ จะเป็นตัวกำหนดตำแหน่งที่ตั้งของเส้นโค้งและความโค้งหรือแบบของเส้นโค้งตามลำดับ

ซึ่งการผลิตตัวแปรสุ่มที่มีการแจกแจงแบบปกติที่นิยมใช้มี 3 วิธี (ศิริจันทร์ ทองประเสริฐ, 2540) คือ

(1) วิธีการของ Box และ Muller ซึ่งใช้สร้างค่าตัวแปรสุ่มที่มีการแจกแจงแบบปกติมาตรฐานพร้อม ๆ กัน 2 ค่า เป็นอิสระกันโดยใช้ตัวผลิต (generator) Z_1 และ Z_2

$$Z_1 = \left(\sqrt{-2\ln R_1}\right)\cos\left(2\pi R_2\right)$$
$$Z_2 = \left(\sqrt{-2\ln R_1}\right)\sin\left(2\pi R_2\right)$$

R_1 และ R_2 เป็นตัวเลขสุ่ม เมื่อได้ตัวแปรสุ่มที่มีการแจกแจงแบบปกติมาตรฐานแล้ว ทำการแปลงตัวแปรสุ่มดังกล่าวโดยอาศัยฟังก์ชัน

$$Z_1' = \mu + \sigma Z_1$$

และ

$$Z_2' = \mu + \sigma Z_2$$

ซึ่งจะได้ว่า Z_1' และ Z_2' มีการแจกแจงเป็นแบบปกติที่มีค่าเฉลี่ย $E(x) = \mu$ และความแปรปรวน $V(x) = \sigma^2$ นั่นคือ $Z_i' \sim N(\mu, \sigma^2); i = 1,2$

(2) วิธีการของ Marsaglia-Bray ซึ่งสร้างตัวแปรสุ่มที่มีการแจกแจงแบบปกติพร้อม ๆ กัน 2 ค่า โดยดัดแปลงจากวิธีการของ Box และ Muller โดยใช้สูตร

$$Z_1' = \mu + \left(\frac{V_1}{S}\sqrt{-2\ln S}\right)\sigma$$

$$Z_2' = \mu + \left(\frac{V_2}{S}\sqrt{-2\ln S}\right)\sigma$$

เมื่อ

$$V_1 = -1 + 2R_1$$

$$V_2 = -1 + 2R_2$$

$$S = V_1^2 + V_2^2$$

ถ้า $S > 1$ ให้หาตัวเลขสุ่มใหม่ เพื่อหาค่า V_1 และ V_2 ใหม่จนกว่าจะได้ $S < 1$

(3) ทฤษฎีเข้าสู่ศูนย์กลาง (Central Limit Theorem) สร้างตัวแปรที่มีการ แจกแจงแบบปกติ โดยใช้สูตร

$$Z' = \mu + (\sum_{i=1}^{12} R_i - 6)\sigma$$

ในกรณีที่ $(\sum_{i=1}^{12} R_i - 6)$ มีค่าเกินกว่า 2 ให้ปรับค่าเป็น

$$Z' = \mu + R\sigma$$

เมื่อ $R = ((((C_1N^2 + C_2)N^2 + C_3)N^2 + C_4)N^2 + C_5)N$

โดย $N = (\sum_{i=1}^{12} R_i - 6)/4$

และค่าคงที่ $C_1 = 0.029899776$ $C_2 = 0.008355968$

$C_3 = 0.076542912$ $C_4 = 0.252408784$

$C_5 = 3.949846138$

สำหรับการวิจัยครั้งนี้ได้มีการผลิตตัวแปรสุ่มที่มีการแจกแจงแบบปกติ โดยใช้วิธีทฤษฎีเข้าสู่ศูนย์กลาง (Central Limit Theorem) เนื่องจากต้องการผลิตตัวแปรที่มีการแจกแจงแบบปกติมาตรฐานครั้งละ 1 ค่า

2.1.3.3 การทดลองกระทำ โดยใช้กระบวนการของการสุ่ม (random process) มากระทำในลักษณะซ้ำๆกัน ซึ่งในการวิจัยครั้งนี้ในแต่ละแผนแบบการจำลองได้กระทำซ้ำ 1,000 รอบ เพื่อหาคำตอบของปัญหาที่ต้องการศึกษา และจากการศึกษา 1,000 รอบ จะได้จำนวนครั้งของการปฏิเสธสมมติฐาน เพื่อที่จะนำจำนวนครั้งของการปฏิเสธสมมติฐานไปคำนวณหาค่าอำนาจของการทดสอบ

2.2 ผลงานวิจัยที่เกี่ยวข้อง

การวิเคราะห์ความแปรปรวนเมื่อมีการฝ่าฝืนข้อกำหนดเบื้องต้น มีผู้ศึกษาผลกระทบต่างๆ ดังนี้

Norton (1946 อ้างถึงใน Lindquist, 1953) ได้กล่าวว่า

- การแจกแจงเอฟ เมื่อประชากรแต่ละกลุ่มมีการแจกแจงแบบสมมาตร จะไม่มีผลกระทบเกี่ยวกับการขาดข้อตกลง นั่นก็แสดงว่าจะไม่มีผลกระทบเมื่อประชากรมีความโด่งผิดปกติ
- การแจกแจงเอฟมีความแกร่งต่อการฝ่าฝืนข้อตกลงที่ว่าความแปรปรวนแต่ละกลุ่มต้องเท่ากัน เมื่อขนาดตัวอย่างของแต่ละกลุ่มมีขนาดเท่ากัน แต่ถ้าขนาดตัวอย่างแต่ละกลุ่มไม่เท่ากันจะส่งผลกระทบต่อระดับนัยสำคัญ

Guildford (1965) ได้ทำการศึกษาเกี่ยวกับข้อตกลงเบื้องต้นของ ANOVA พบว่า ค่า F ที่ได้ค่อนข้างจะไม่แตกต่างมากนักเมื่อการแจกแจงของประชากรไม่เป็นไปตามข้อกำหนด แต่จากการศึกษาเกี่ยวกับการไม่เท่ากันของความแปรปรวน พบว่ามีผลกระทบต่อค่าสถิติทดสอบ F ดังนั้นในกรณีค่าความแปรปรวนแตกต่างกันนั้น ควรคำนึงถึงค่าระดับนัยสำคัญ คือ ถ้าการทดสอบมีระดับนัยสำคัญที่ 0.05 ผลดังกล่าวอาจแสดงความมีนัยสำคัญที่แท้จริงอยู่ระหว่าง 0.04-0.07 และถ้าการทดสอบมีระดับนัยสำคัญที่ 0.01 ความมีนัยสำคัญที่แท้จริงอาจอยู่ระหว่าง 0.005-0.02 ความมีนัยสำคัญที่ได้จากข้อมูลที่มีลักษณะเช่นนี้มักต่ำกว่าระดับที่แท้จริง (อาจมีบ้างที่สูงกว่า)

Glass, J. C. (1970) ได้ให้ข้อสังเกตที่ผู้วิจัยจะนำไปพิจารณาออกแบบการวิจัยและการวิเคราะห์ข้อมูล ได้ดังนี้

- กรณีที่มีการฝ่าฝืนข้อตกลงเบื้องต้นของการเท่ากันของความแปรปรวน ถ้าผู้วิจัยกำหนดขนาดตัวอย่างให้มีจำนวนเท่ากันทุกกลุ่ม แล้วระดับของการยอมรับหรือปฏิเสธสมมติฐาน เกือบจะไม่แตกต่างกันกับการไม่ฝ่าฝืนข้อกำหนดเบื้องต้น

- กรณีที่ขนาดของตัวอย่างมีจำนวนไม่เท่ากัน และความแปรปรวนของประชากรแต่ละกลุ่มไม่เท่ากัน ถ้าขนาดตัวอย่างมีขนาดเล็กแต่ความแปรปรวนของประชากรมีขนาดใหญ่ด้วย โอกาสที่ผู้วิจัยจะสรุปผิดพลาด ทำให้เกิดความคลาดเคลื่อนประเภทที่ 1 จะมีมากกว่าระดับนัยสำคัญที่ผู้วิจัยกำหนด

- กรณีที่ขนาดของตัวอย่างมีจำนวนไม่เท่ากัน และความแปรปรวนของประชากรแต่ละกลุ่มไม่เท่ากัน ถ้าขนาดตัวอย่างมีขนาดใหญ่และความแปรปรวนของประชากรก็มีขนาดใหญ่ด้วย โอกาสที่ผู้วิจัยจะสรุปผิดพลาดทำให้เกิดความคลาดเคลื่อน ประเภทที่1 จะมีน้อยกว่าระดับนัยสำคัญที่ผู้วิจัยกำหนด

- กรณีที่มีการฝ่าฝืนข้อตกลงของการแจกแจงแบบปกติ พบว่าการฝ่าฝืนข้อตกลงดังกล่าวส่งผลต่อการที่ผู้วิจัยจะสรุปผิดพลาดมีโอกาสน้อยมาก กล่าวคือจะคลาดเคลื่อนไม่เกินระดับความคลาดเคลื่อนที่กำหนด

- การฝ่าฝืนข้อตกลงการเป็นอิสระของความคลาดเคลื่อน มีผลต่อความผิดพลาดในการสรุปผลการวิจัยอย่างยิ่ง ควรเลือกใช้สถิติที่ใช้ทดสอบสมมติฐานที่เหมาะสมแบบอื่น ไม่เหมาะที่จะใช้การวิเคราะห์ความแปร ปรวนตามแบบแผนที่กำหนดไว้

Myers, J. L. (1972) ได้กล่าวถึงการฝ่าฝืนข้อกำหนดไว้ว่า จะทำให้ค่าความน่าจะเป็นที่แท้จริงของพื้นที่วิกฤต (ξ) ไม่เท่ากับค่าระดับนัยสำคัญของการทดสอบที่เราตั้ง (α) โดยที่ค่าที่แท้จริงของพื้นที่วิกฤตอาจมากกว่าหรือน้อยกว่าค่า α ซึ่งก็ขึ้นอยู่กับรูปแบบของการถูกรบกวน

- กรณีที่ข้อมูลไม่เป็นปกติ ไม่มีผลกระทบต่อความคลาดเคลื่อนประเภทที่ 1 หรือประเภทที่ 2 ในอัตราส่วนที่มาก ถ้าเพิ่มค่าองศาอิสระ (degree of freedom) ของค่าความคลาดเคลื่อน ซึ่งเพิ่มโดยการเพิ่มจำนวนขนาดตัวอย่าง

- ถ้าความแปรปรวนของข้อมูลแต่ละกลุ่มแตกต่างกัน จะส่งผลกระทบต่อการวิเคราะห์ความแปรปรวน ซึ่งการฝ่าฝืนข้อกำหนดข้อนี้มีความไวมากกว่าการฝ่าฝืนการเป็นปกติ สำหรับการป้องกันการเกิดปัญหาการเกิดความคลาดเคลื่อนประเภทที่ 1 ทำโดยการให้จำนวนขนาดตัวอย่างของแต่ละกลุ่มเท่ากัน ถ้าในแต่ละกลุ่มของการทดลองมีขนาดตัวอย่างไม่ท่ากัน หรือถ้าความแปรปรวนของความคลาดเคลื่อนของบางกลุ่มมีค่าต่างจากพวกมาก จะทำให้ค่าความน่าจะเป็นที่แท้จริงของพื้นที่วิกฤต (ξ)ไม่เท่ากับค่าที่เราตั้ง (α) สำหรับกรณีที่เกิดการฝ่าฝืนข้อกำหนดเบื้องต้นทั้งสองข้อพร้อมกันคือความแปรปรวนแต่ละกลุ่มไม่เท่ากันและข้อมูลไม่เป็นปกติไม่ได้ส่งผลกระทบต่อความคลาดเคลื่อนประเภทที่ 1 มากกว่าการฝ่าฝืนข้อกำหนดเบื้องต้นของการเท่ากันของความแปรปรวนของความคลาดเคลื่อนเพียงข้อเดียว

- ถ้าข้อมูลไม่เป็นไปตามข้อกำหนดเกี่ยวกับการเป็นอิสระ จะมีผลกระทบต่อการแจกแจงของอัตราส่วนของ mean squares กรณีที่ mean squares มีความสัมพันธ์ทางบวกจะส่งผลกระทบต่อความคลาดเคลื่อน ประเภทที่ 1 มากขึ้น และถ้ามีความสัมพันธ์ทางลบจะเกิดผลกระทบลดลง

Miller, R. G. (1986) ได้กล่าวถึงข้อควรระวังในการวิเคราะห์ความแปรปรวน เพื่อให้ผลไม่เอนเอียง ในกรณีกำหนดปัจจัย ไว้ดังนี้

- การขาดข้อกำหนดที่ว่า การแจกแจงเป็นปกติ มีผลกระทบต่อระดับนัยสำคัญของการทดสอบเอฟเพียงเล็กน้อย โดยส่งผลกระทบน้อยกว่ากรณีกลุ่มตัวอย่าง 2 กลุ่ม ในการทดลองเมื่อเราเพิ่มขนาดตัวอย่างของแต่ละกลุ่มให้ใหญ่ จะประมาณ central limit theorem โดยจะทำให้ค่าเฉลี่ยแต่ละกลุ่มประมาณเข้าสู่ปกติ

- เมื่อความแปรปรวนแต่ละกลุ่มแตกต่างกันแล้ว sum of squares ของการทดสอบจะให้น้ำหนักไม่เท่ากัน แล้วทำให้การแจกแจงไม่เป็น χ^2 เมื่อขนาดตัวอย่างแต่ละกลุ่มไม่เท่ากันจะส่งผลกระทบยิ่งขึ้น ถ้าความแปร ปรวนที่ใหญ่ที่สุดอยู่ในกลุ่มที่มีขนาดตัวอย่างต่ำสุดเป็นเรื่องที่น่ากังวลมาก เนื่องจากว่าระดับนัยสำคัญจะผิดไป ถ้าจะลดผลกระทบให้น้อยที่สุด ถ้าเป็นไปได้ให้ทำการทดลองที่ขนาดตัวอย่างแต่ละกลุ่มเท่ากัน

Krutchkoff, R. G. (1988) ได้ทำการศึกษากรณีความแปรปรวนแต่ละกลุ่มไม่เท่ากันของการทดสอบเอฟ ได้ผลว่า เมื่อขนาดกลุ่มเล็ก (3) และจำนวนซ้ำหรือขนาดตัวอย่างแต่ละกลุ่มใหญ่ (20) ยังพบว่าไม่มีผลกระทบ ถ้าจำนวนกลุ่มในการทดสอบมีจำนวนมากและขนาดตัวอย่างแต่ละกลุ่มมีขนาดเล็ก จะทำให้พลังในการทดสอบต่ำกว่าการทดสอบเอฟที่ความแปรปรวนของแต่ละกลุ่มเท่ากัน แต่แนะนำว่ากรณีความแปรปรวนแต่ละกลุ่มไม่เท่ากันยังสามารถใช้การทดสอบเอฟได้เมื่อจำนวนกลุ่มมากกว่า 10 กลุ่ม ค่าสังเกตไม่น้อยกว่า 3 และกรณีที่เกิดปัญหา ความแปรปรวนแต่ละกลุ่มไม่เท่ากัน ถ้าเราให้ขนาดตัวอย่างของแต่ละกลุ่มเท่ากันก็จะลดปัญหาได้ เช่น ถ้าให้ α = 0.05 ค่าระดับนัยสำคัญที่แท้จริงที่ได้ก็จะใกล้เคียง 0.05 ซึ่งยอมรับได้

เมื่อพบว่าข้อมูลที่ได้จากการศึกษาไม่เป็นไปตามข้อกำหนดเบื้องต้นเกี่ยวกับความเป็นเอกภาพของความแปรปรวน โดยปกติทั่วไปจะปฏิบัติอย่างใดอย่างหนึ่ง ดังนี้

- ยังคงใช้สถิติเอฟ ทดสอบข้อมูล
- ทำการแปลง (transform) ข้อมูล
- ใช้การทดสอบนอนพาราเมตริก เช่น Kruskal-Wallis Test

อย่างไรก็ตามก็ไม่สามารถยืนยันได้ว่า การตัดสินใจใดผิดหรือเหมาะสมที่สุด

Montgomery, D. C. (1997) ได้กล่าวถึงการฝ่าฝืนข้อกำหนดและผลกระทบ ดังนี้

- การฝ่าฝืนข้อกำหนดเกี่ยวกับการแจกแจงแบบปกติ ในกรณีกำหนดปัจจัย (fixed effect) จะไม่น่ากังวลนัก หรือกล่าวได้ว่าการวิเคราะห์ความแปรปรวนมีความแกร่งต่อข้อกำหนดความเป็นปกติ

- การฝ่าฝืนข้อกำหนดเกี่ยวกับการเท่ากันของความแปรปรวน เป็นปัญหาที่ที่ค่อนข้างน่าหนักใจในกรณีขนาดตัวอย่างแต่ละกลุ่มไม่เท่ากัน ในกรณีที่ขนาดตัวอย่างแต่ละกลุ่มเท่ากันจะมีผลกระทบเพียงเล็กน้อยต่อการทดสอบเอฟ

- การฝ่าฝืนข้อกำหนดเกี่ยวกับการเป็นอิสระ อาจจะเกิดขึ้นได้จากการเก็บข้อมูลที่ยุ่งยาก ซึ่งเราสามารถป้องกันปัญหาของข้อมูล โดยทำการทดลองแบบสุ่ม ซึ่งก็จะทำให้ข้อมูลที่เก็บมาได้เป็นอิสระ

จินตนา ศรีสันสนีย (2541) ได้ทำการศึกษาผลกระทบต่อการฝ่าฝืนข้อกำหนดเบื้องต้นเกี่ยวกับการเป็นเอกภาพของความแปรปรวนและการแจกแจงแบบปกติ ในการควบคุมความคลาดเคลื่อนประเภทที่ 1 ในการวิเคราะห์ความแปรปรวนแบบจำแนกทางเดียว จากการวิจัยพบว่า

- เมื่อข้อมูลมีการแจกแจงแบบปกติความแปรปรวนเท่ากัน จำนวนกลุ่มตัวอย่างและจำนวนซ้ำไม่มีผลต่ออัตราส่วนความคลาดเคลื่อนประเภทที่ 1 โดยส่วนใหญ่สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้

- เมื่อข้อมูลมีการแจกแจงแบบปกติ ความแปรปรวนไม่เท่ากัน และจำนวนซ้ำไม่เท่ากัน ส่วนใหญ่อัตราส่วนความคลาดเคลื่อนประเภทที่ 1 จะมากกว่าอัตราส่วนความคลาดเคลื่อนที่ระบุแต่มากกว่าเพียงเล็กน้อยเท่านั้น ดังนั้นกรณีนี้ยังคงเหมาะสมที่จะใช้ได้

- เมื่อข้อมูลมีการแจกแจงแบบปกติ ความแปรปรวนไม่เท่ากัน และจำนวนซ้ำเท่ากัน ส่วนใหญ่อัตราส่วนความคลาดเคลื่อนประเภทที่ 1 จะมากกว่าความคลาดเคลื่อนที่ระบุ

- เมื่อข้อมูลมีการแจกแจงแบบเบ้ ความแปรปรวนเท่ากัน เมื่อจำนวนกลุ่มตัวอย่าง 3 และ 4 ที่จำนวนซ้ำเท่ากันและไม่เท่ากัน ยังคงสามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ แต่เมื่อจำนวนกลุ่มตัวอย่างมากกว่า 4 กลุ่มขึ้นไป อัตราส่วนความคลาดเคลื่อนประเภทที่ 1 จะต่างกับที่ระบุไว้มาก และยิ่งเบ้มาก อัตราส่วนความคลาดเคลื่อนยิ่งต่างจากที่ระบุไว้ยิ่งขึ้น