บทที่ 2

เอกสาร และผลการวิจัยที่เกี่ยวข้อง

การเสนอมโนทัศน์ และผลการวิจัยที่เกี่ยวข้อง สำหรับการวิจัยครั้งนี้ จำแนกเสนอเป็น 4 ตอนดังนี้

1. หลักการวิเคราะห์ความแปรปรวน
2. การแจกแจงที่เกี่ยวข้องกับสถิติทดสอบที่ใช้ในการวิจัย
3. สถิติทดสอบในงานวิจัยนี้
4. งานวิจัยที่เกี่ยวข้อง

1. <u>หลักการวิเคราะห์ความแปรปรวน</u>

การทดสอบสมมติฐานเกี่ยวกับความแตกต่างของค่าเฉลี่ยประชากรมากกว่าสองกลุ่มขึ้นไป ถ้าจะสรุปผลการทดสอบ โดยพิจารณาจากผลการทดสอบความแตกต่างระหว่างค่าเฉลี่ยของประชากรทีละสองประชากร โดยใช้ z-test หรือ t-test จะต้องทำการทดสอบทีละคู่หลายครั้ง ซึ่งวิธีดังกล่าวนี้จะทำให้ผลการทดสอบมีโอกาสผิดพลาดมากขึ้น คือโอกาสที่ผู้วิจัยจะปฏิเสธสมมติฐานที่ถูก (Type I error) มีค่าสูงขึ้น เมื่อจำนวนกลุ่มตัวอย่างมีอยู่ k กลุ่ม โอกาสที่จะปฏิเสธสมมติฐานผิดพลาดอย่างน้อย 1 ครั้ง ของการทดสอบ C ครั้ง (C คือ จำนวนคู่ที่จะต้องทดสอบทั้งหมด โดยคิดจาก Combination ได้ kC_2 ครั้ง) จะเป็น $1-(1-\alpha)^C$ (Hopkins and Glass 1978: 333)

ตารางที่ 1 เปรียบเทียบโอกาสสูงสุดที่จะปฏิเสธสมมติฐานที่ถูกอย่างน้อย 1 ครั้ง ในการทดสอบค่าเฉลี่ยทีละคู่ โดยใช้การทดสอบที (t-test) เมื่อกำหนด $\alpha = .05$

จำนวนกลุ่ม (k)	จำนวนคู่ของการเปรียบเทียบ (C)	$1 - (1 - \alpha)^C$
2	1	0.05
3	3	0.14
5	10	0.40
10	45	0.90

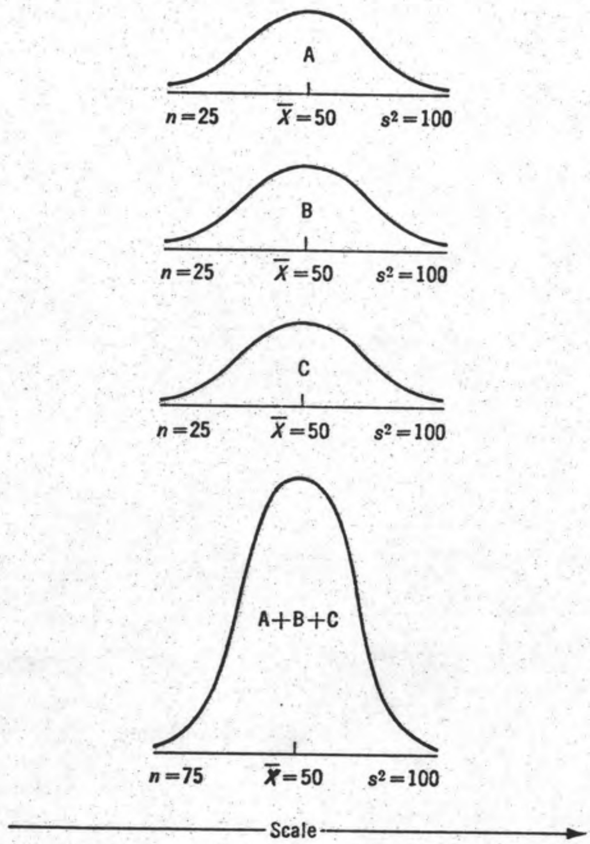
ดังนั้น ในปี ค.ศ. 1923 เซอร์ไรนัล ฟิเชอร์ นักสถิติชาวอังกฤษ จึงคิดวิธีการทดสอบความแตกต่างระหว่างค่าเฉลี่ยของประชากรมากกว่าสองกลุ่มขึ้น เรียกว่า การวิเคราะห์ความแปรปรวน (Analysis of Variance: ANOVA) โดยมีสมมติฐานการทดสอบว่า "ค่าเฉลี่ยของกลุ่มที่นำมาทดสอบทุกกลุ่มไม่แตกต่างกัน"

$H_0: \mu_1 = \mu_2 = \ldots = \mu_k$

H_A: มีค่าเฉลี่ยอย่างน้อย 2 ค่า ที่แตกต่างกัน

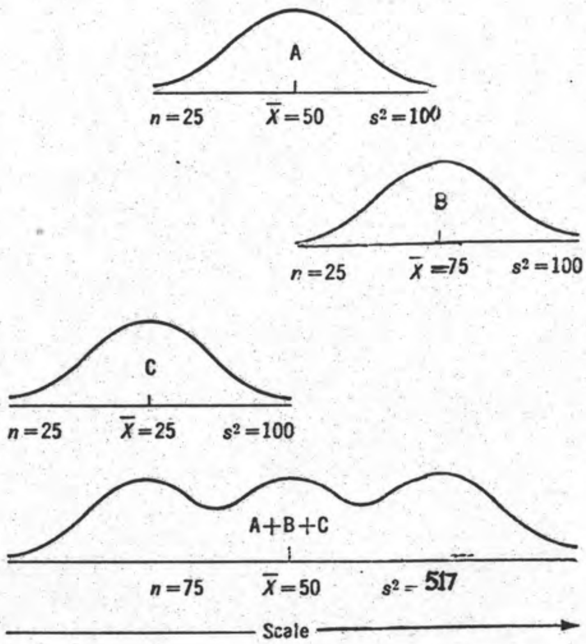
สถิติที่นำมาทดสอบในการวิเคราะห์ความแปรปรวน คือ อัตราส่วนเอฟ (F - ratio) หรือ สถิติทดสอบเอฟ (F - test)

เป้าหมายในการวิเคราะห์ความแปรปรวนมุ่งทดสอบความแตกต่างระหว่างค่าเฉลี่ยของกลุ่มประชากรมากกว่าสองกลุ่ม แต่การทดสอบยึดทดสอบความแปรปรวนของกลุ่ม ซึ่งดูเหมือนขัดแย้งกัน ทั้งนี้เพราะความแปรปรวนของค่าการวัดในคะแนนชุดใด เป็นการวัดการกระจาย (measure of variability) ของคะแนนในกลุ่ม ส่วนค่าเฉลี่ยเป็นการวัดตัวกลาง (measure of Central tendency) จากข้อขัดแย้งนี้ สามารถอธิบายความสัมพันธ์ระหว่างความแปรปรวนและค่าเฉลี่ยของคะแนนจากภาพดังนี้ (Popham and Sirotnik 1973: 168-169)



แผนภาพที่ 1 ค่าเฉลี่ยของทั้งสามกลุ่มมีค่าเท่ากัน

จากภาพที่ 1 แสดงค่าเฉลี่ยของกลุ่ม A, B และ C มีค่าเท่ากันคือ 50 ความแปรปรวนของแต่ละกลุ่มเป็น 100 เท่ากัน ภาพล่างสุดเป็นการกระจายของกลุ่มรวม ซึ่งจะมีโค้งการแจกแจงสูงขึ้นจากแต่ละกลุ่มย่อย เพราะจำนวนคะแนนเพิ่มเป็น 75 แต่ค่าเฉลี่ยและความแปรปรวนยังคงมีค่าเท่ากันในแต่ละกลุ่มย่อย ภาพนี้ใช้ให้เห็นว่าในกรณีที่ทั้งสามกลุ่มมีค่าเฉลี่ยเท่ากัน ความแปรปรวนเฉลี่ยของสามกลุ่มจะมีค่าเท่าหรือใกล้เคียงกับความแปรปรวนของกลุ่มรวม



แผนภาพที่ 2 ค่าเฉลี่ยของทั้งสามกลุ่มแตกต่างกัน

จากภาพที่ 2 แสดงความแตกต่างของค่าเฉลี่ยทั้งสามกลุ่ม โดยมีจำนวนคะแนน และความแปรปรวนในแต่ละกลุ่มเท่ากัน ภาพล่างสุดเป็นการกระจายของคะแนนรวมทั้งหมด ซึ่งจะมีช่วงกว้างขึ้นจากแต่ละกลุ่มย่อย ความแตกต่างดังกล่าวเกิดขึ้น เพราะความแปรปรวนของกลุ่มรวมมีค่ามาก

องค์ประกอบที่บ่งชี้ว่าความแปรปรวนของกลุ่มรวมมีค่ามากน้อยมี 2 ประการ ได้แก่

1. ความแปรปรวนระหว่างกลุ่ม (between-group Variance) เป็นความแปรปรวนที่เกิดจากค่าเฉลี่ยของแต่ละกลุ่มเบี่ยงเบนจากค่าเฉลี่ยของกลุ่มรวม (grand mean) กล่าวคือ ถ้าค่าเฉลี่ยของกลุ่มย่อยแตกต่างกันน้อย ค่าเฉลี่ยเหล่านี้ย่อมเบี่ยงเบนจากกลุ่มรวมไม่มาก และการกระจายของคะแนนในกลุ่มรวมจะแคบลงด้วย ในทางตรงข้าม ถ้าค่าเฉลี่ยของกลุ่มย่อยแตกต่างกันมาก การกระจายของคะแนนในกลุ่มรวมจะกว้างขึ้น

2. ความแปรปรวนภายในแต่ละกลุ่ม (within group Variance) เป็นความแปรปรวนที่เกิดจากคะแนนในแต่ละกลุ่มเบี่ยงเบนจากค่าเฉลี่ยของแต่ละกลุ่ม กล่าวคือ ถ้าในแต่ละกลุ่มย่อยมีคะแนนแคบ (ความแปรปรวนต่ำ) เมื่อรวมการกระจายของคะแนนจากทุกกลุ่ม

มาเป็นกลุ่มรวม ย่อมมีพิสัยของคะแนนแคบด้วย และค่าความแปรปรวนย่อมต่ำ ในทำนองเดียวกัน การกระจายของคะแนนในแต่ละกลุ่มยิ่งกว้างเท่าไร ผลรวมการกระจายของคะแนนในกลุ่มรวม ย่อมมีพิสัยของคะแนนกว้างและค่าความแปรปรวนย่อมมากด้วย

พิจารณาจากองค์ประกอบ 2 ประการ จะเห็นว่าความแปรปรวนระหว่างกลุ่ม เป็นผลมาจากความแตกต่างของค่าเฉลี่ย หรือเป็นผลของตัวจัดกระทำ (treatment) ส่วนความแปรปรวนภายในกลุ่มเป็นผลสืบเนื่องจากสภาพความแตกต่างของคะแนนของบุคคล (subject) ซึ่งมีมูลเหตุจากแหล่งต่าง ๆ ที่ทำให้เกิดสภาพความแตกต่างหรือเรียกว่าเป็นความคลาดเคลื่อน (error)

แนวคิดดังกล่าวข้างต้นเป็นหลักสำคัญในการทดสอบสมมติฐานของการวิเคราะห์ความแปรปรวน โดยใช้อัตราส่วนของความแปรปรวนระหว่างกลุ่มกับความแปรปรวนภายในกลุ่ม เป็นหลักในการลงสรุปหรือตัดสินใจ และเรียกอัตราส่วนนี้ว่าอัตราส่วนเอฟ (F – ratio)

<u>โมเดลสำหรับการวิเคราะห์ความแปรปรวน</u>

การวิเคราะห์ความแปรปรวน สามารถประยุกต์ใช้ได้หลายแบบด้วยกัน ขึ้นอยู่กับการออกแบบทดลอง แต่พอจะสรุปลักษณะการใช้ที่สำคัญ ๆ ได้ดังนี้ 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 การวิเคราะห์ความแปรปรวนแบบแฟคทอเรียล (Factorial Design)

2.2 การวิเคราะห์ความแปรปรวนแบบแยกส่วน (Spilt – Plot Design)

การวิเคราะห์ความแปรปรวนตามโมเดลดังกล่าว สามารถจะจำแนกเป็นโมเดลเชิงเส้นตรงของผลทดลองตามกำหนด (Fixed Effects Linear Model) และโมเดลเชิงเส้นตรงของผลทดลองตามการสุ่ม (Random Effects Linear Model) สำหรับงานวิจัยนี้จะยกตัวอย่างเฉพาะโมเดลที่ใช้กันมาก คือ โมเดลเชิงเส้นตรงของผลทดลองตามกำหนด (Fixed Effect Linear Model) ซึ่งมีรูปแบบดังนี้

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

เมื่อ X_{ij} คือ ผลการวัดตัวแปรตามของตัวอย่างที่สุ่มมาจากประชากร

μ คือ มัชฌิมเลขคณิตของประชากร

α_j คือ ผลจากการทดลอง (Treatment Effect) ซึ่งจะมีค่าคงที่สำหรับตัวอย่างทั้งหมดภายในแต่ละประชากร

ε_{ij} คือ ความคลาดเคลื่อนทางการทดลอง ซึ่งเป็นอิสระจากค่าความคลาดเคลื่อนอื่น ๆ (Error Effect)

และตารางวิเคราะห์ความแปรปรวน มีรูปแบบดังนี้ คือ

แหล่งความแปรปรวน	SS	df	MS	F
ระหว่างกลุ่ม (B)	$\sum_{j=1}^{k} n_j(\bar{X}_{.j} - \bar{X}_{..})^2$	$k - 1$	SS_B/df	$\frac{MS_B}{MS_w}$
ภายในกลุ่ม (w)	$\sum_{j=1}^{k} \sum_{i=1}^{n} (X_{ij} - \bar{X}_{.j})^2$	$N - k$	SS_w/df	
ทั้งหมด (T)	$SS_B + SS_w$	$N - 1$		

เมื่อ N คือ จำนวนข้อมูลทั้งหมด

n_j คือ จำนวนข้อมูลในแต่ละระดับการทดลอง

k คือ จำนวนระดับการทดลอง

X_{ij} คือ คะแนนแต่ละตัวซึ่งจะเป็นที่ i ของกลุ่มที่ j

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

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

<u>ข้อตกลงเบื้องต้นของ ANOVA</u>

1. ประชากรมีการแจกแจงแบบปกติ (Normally Distributed Population) เศษและส่วนของอัตราส่วน F(F - Ratio) ต้องเป็นอิสระต่อกัน โดยข้อมูลได้รับการสุ่มจากประชากรที่มีการแจกแจงปกติ

2. ความเป็นเอกพันธุ์ของความแปรปรวนของประชากร (Homogeneity of Population Error Variances) หมายความว่า ความแปรปรวนที่เนื่องจากความคลาดเคลื่อนภายในประชากรทดลองแต่ละกลุ่มเท่ากัน

3. การรวมกันได้ของผลการทดลอง (Addivity of Effects) หมายความว่า แรงปะทะ (Interaction) ระหว่างระดับการทดลอง และแต่ละกลุ่มมีค่าเท่ากับศูนย์ หรือหมายถึงความแปรปรวนร่วมระหว่างระดับการทดลองแต่ละคู่เท่ากัน

4. ความเป็นอิสระในการสุ่มตัวอย่าง (Independence Within Samples) และความเป็นอิสระระหว่างกลุ่มตัวอย่างในกรณีของการสุ่มแบบสมบูรณ์

2. <u>การแจกแจงที่เกี่ยวข้องกับสถิติทดสอบในงานวิจัยนี้</u>

เนื่องจากสถิติทดสอบเอฟ มีการแจกแจงแบบเอฟ และสถิติทดสอบเอฟสตาร์ ประมาณด้วยการแจกแจงเอฟ ส่วนสถิติทดสอบยู ประมาณด้วยการแจกแจงไคสแควร์ ในส่วนนี้จึงเสนอการแจกแจงเอฟและการแจกแจงไคสแควร์ ตามลำดับดังนี้

2.1 <u>การแจกแจงเอฟ</u> (F - Distribution)

Sir Ronald A Fisher เป็นบุคคลแรกที่เสนอการแจกแจงเอฟ เดิมใช้ชื่อว่าการแจกแจงซี (Z - distribution) โดยมีสูตร $Z = \frac{1}{2} \ln \frac{S_1^2}{S_2^2}$ ซึ่งค่าลอการิทึม (logarithm) ทำให้ยุ่งยากในการคำนวณ W. Snedecor จึงปรับปรุงใหม่ โดยใช้สูตร $F = \frac{S_1^2}{S_2^2}$ ค่า S_1^2

และ S_2^2 คือค่าความแปรปรวนของกลุ่มตัวอย่างจากประชากรที่ 1 และกลุ่มตัวอย่างจากประชากรที่ 2 ตามลำดับ Snedecor ได้ตั้งชื่อการแจกแจงใหม่นี้ว่า การแจกแจงเอฟ (F - distribution) เพื่อเป็นเกียรติแก่ Ronald A Fisher

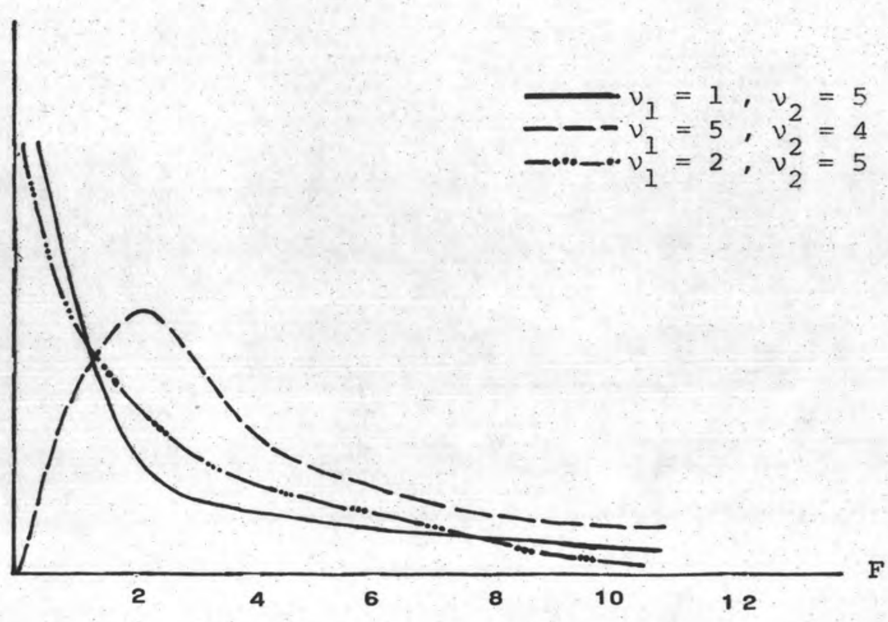
2.1.1 ฟังก์ชั่นการแจกแจงของเอฟ

ตัวแปรสุ่มที่มีการแจกแจงแบบเอฟได้จากอัตราส่วนของตัวแปรสุ่มแบบไคสแควร์ 2 ตัว ที่เป็นอิสระต่อกัน ซึ่งแต่ละตัวหารด้วยจำนวนชั้นความเป็นอิสระ (degree of freedom) ของตัวมันเอง และมีฟังก์ชั่นความน่าจะเป็นดังนี้

$$g(f) = \frac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)\left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}}}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} . f^{\frac{\nu_1}{2}-1} . \left(1+\frac{\nu_1}{\nu_2}\right) . f^{-\left(\frac{\nu_1+\nu_2}{2}\right)} \quad ; f > 0$$

เมื่อ Γ คือ ฟังก์ชั่นแกมม่า

ν_1 และ ν_2 เป็นชั้นความเป็นอิสระแบบไคสแควร์ที่เป็นเศษและเป็นส่วนตามลำดับ



แผนภาพที่ 3 เปรียบเทียบการแจกแจงเอฟที่ชั้นความเป็นอิสระเท่ากับ ($\nu_1 = 2, \nu_2 = 5$), ($\nu_1 = 1$, $\nu_2 = 5$) และ ($\nu_1 = 5$, $\nu_2 = 4$)

2.1.2 <u>คุณสมบัติของการแจกแจงเอฟ</u>

1. รูปการแจกแจงขึ้นอยู่กับ ν_1 และ ν_2 และพิสัยของค่า F อยู่ระหว่าง $0 \leq F \leq \infty$

2. การแจกแจงมีลักษณะเบ้ขวา หรือเบ้ทางบวก (positively Skewed) และมีจุดที่สูงสุดของโค้งเพียงจุดเดียว (unimodal)

3. การแจกแจง F มีค่าเฉลี่ยเลขคณิตเท่ากับ $\frac{\nu_2}{\nu_2 - 2}$ เมื่อ $\nu_2 > 2$ และความแปรปรวนเท่ากับ $\frac{2\nu_2^2(\nu_1 + \nu_2 - 2)}{\nu_1(\nu_2 - 2)^2(\nu_2 - 4)}$ เมื่อ $\nu_2 > 4$

4. รูปของการแจกแจง $F(\nu_1, \nu_2)$ ไม่เหมือนกับการแจกแจงของ $F(\nu_2, \nu_1)$ เพราะกำลังของ F ในเศษเปลี่ยนไป และ $_{1-P}F_{\nu_1 - \nu_2} = \frac{1}{_{P}F_{\nu_2, \nu_1}}$

เมื่อ p คือระดับของความมีนัยสำคัญ

2.1.3 <u>ความสำคัญของการแจกแจงเอฟ</u>

การแจกแจงเอฟสามารถนำไปใช้ประโยชน์ในการทดสอบทางสถิติได้หลายประการ ที่ใช้กันมากสรุปได้ดังนี้

1. ทดสอบความแตกต่างระหว่างความแปรปรวนของประชากร 2 ชุด

2. ทดสอบเกี่ยวกับค่าเฉลี่ยเลขคณิตของประชากร ที่มีการแจกแจงปกติ แต่ไม่ทราบความแปรปรวน

3. ทดสอบเกี่ยวกับค่าสัดส่วน

4. เปรียบเทียบ ค่าเฉลี่ยเลขคณิตของประชากรสองชุด ซึ่งมีการแจกแจงปกติ และไม่ทราบค่าความแปรปรวน แต่ตั้งข้อสมมติฐานไว้ว่ามีค่าเท่ากัน

5. การวิเคราะห์ความแปรปรวน

2.2 <u>การแจกแจงไคสแควร์</u> (Chi - Square Distribution)

ER. Helmert นักฟิสิกส์ชาวเยอรมันเป็นผู้พบคนแรกในปี ค.ศ. 1876 ต่อมา Karl Pearson นักสถิติชาวอังกฤษได้ค้นพบใหม่ใน ค.ศ. 1900 โดยการเสนอการทดสอบภาวะสารูปสนิทดีที่ข้อมูลแยกออกจากกัน เมื่อทราบค่าความถี่คาดหวัง (m_i) ล่วงหน้า และ x_i ใดๆ ได้จากการสังเกตจะต้องมีการแจกแจงแบบมัลติโนเมียล (Multinomial Distribution) X_i นั้น อาจมีการแจกแจงเป็นโค้งปกติเมื่อค่าความถี่ที่คาดหวังมีขนาดใหญ่พอทุก ๆ ค่า ซึ่งเป็นที่มาของกลุ่มตัวอย่างขนาดใหญ่ค่า χ^2 จะมีชั้นความเป็นอิสระเท่ากับ $k - 1$

ต่อมาฟิชเชอร์ (R.A. Fisher) ได้เสนอว่าถ้าค่าสังเกต X_i มีการแจกแจงแบบพัวซอง (Poisson Distribution) มี m_i เป็นค่าความถี่คาดหวังหรือมัชฌิมและขนาดของกลุ่มตัวอย่าง n ใหญ่มากแล้ว $Y_i = \frac{X_i - m_i}{\sqrt{m_i}}$ จะมีการแจกแจงแบบปกติ เมื่อมัชฌิมเลขคณิตคือ m_i และส่วนเบี่ยงเบนมาตรฐานเท่ากับ $\sqrt{m_i}$ ดังนั้นการแจกแจงของ χ^2 คือ $Y_1^2 + Y_2^2 + \ldots + Y_k^2$ เมื่อ Y_i มีการแจกแจงอย่างอิสระ และ

$$\sum_{i=1}^{k} Y_i \sqrt{m_i} = \sum_{i=1}^{k} (X_i - m_i) = 0$$

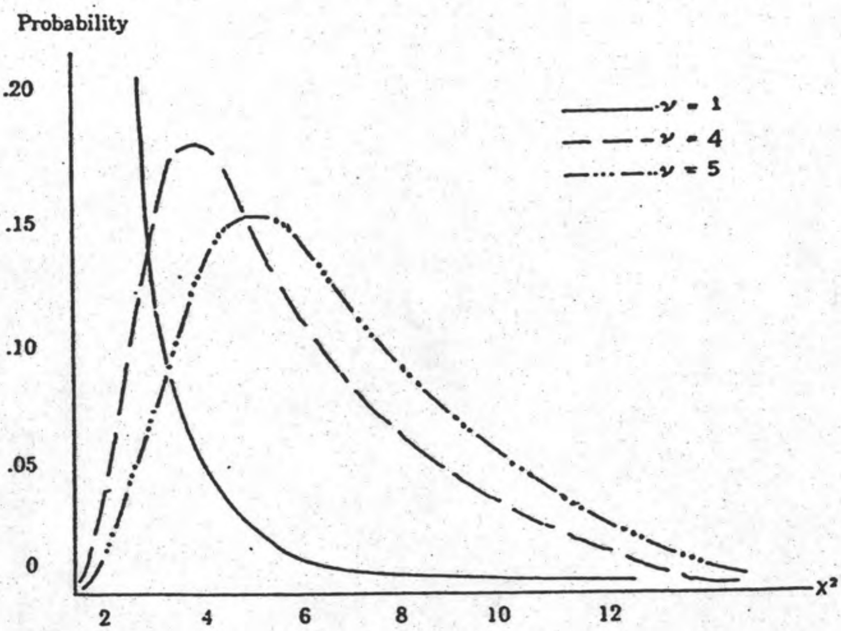
2.2.1 ฟังก์ชั่นการแจกแจงของไคสแควร์

การแจกแจงแบบไคสแควร์ เป็นรูปแบบหนึ่งของการแจกแจงแบบแกมม่า (Gamma Distribution) เมื่อพารามิเตอร์ $\alpha = \frac{\nu}{2}$ และ $\beta = 2$ โดยที่ ν เป็นจำนวนเต็มบวกตัวแปรสุ่มที่มีการแจกแจงแบบไคส์แควร์ ด้วยชั้นความเป็นอิสระ = ν จะมีฟังก์ชั่นความน่าจะเป็นดังนี้

$$f(x) = \frac{1}{2^{\nu/2} \cdot \Gamma\left(\frac{\nu}{2}\right)} \cdot e^{-\frac{x}{2}} X^{\frac{\nu}{2}-1} \qquad , X > 0$$

เมื่อ Γ คือ ฟังก์ชั่นแกมม่า

ตัวแปรสุ่ม X มีการแจกแจงแบบไคสแควร์ที่ชั้นความเป็นอิสระเท่ากับ ν จะแทนด้วยสัญญลักษณ์ χ^2_ν และค่าของฟังกชั่นขึ้นอยู่กับขนาดของชั้นความเป็นอิสระ ν ดังนั้น การแจกแจงไคส์แควร์จะแตกต่างกัน เมื่อชั้นความเป็นอิสระแตกต่างกัน



แผนภาพที่ 4 เปรียบเทียบการแจกแจงไคสแควร์ที่ชั้นความเป็นอิสระเท่ากับ 1, 4 และ 5

2.2.2 <u>คุณสมบัติของไคสแควร์</u>

1. การแจกแจงของไคสแควร์ตามทฤษฎีเป็นการแจกแจงที่ต่อเนื่อง
2. ลักษณะของส่วนโค้งเบ้ขวา และความโค้งจะเริ่มจาก 0 ถึง ∞
3. ค่าออร์ดิเนตสูงสุด อยู่ที่ $\chi^2 = \nu - 1$ เมื่อ ν คือชั้นความเป็นอิสระ
4. มีลักษณะเป็นส่วนโค้งของเปียร์สันชนิดที่ 3 (Pearson Type III curve)
5. การกระจายเป็นไปตามกฎของการแจกแจงแบบมัลติโนเมียล (Multinomial Distribution)
6. สำหรับการแจกแจงของไคสแควร์ เมื่อชั้นความเป็นอิสระเท่ากับ ν จะมีลักษณะดังนี้

6.1 ค่าเฉลี่ยเท่ากับ ν และความแปรปรวนเท่ากับ 2ν

6.2 ฐานนิยม (mode) เท่ากับ $\nu - 2$ เมื่อ $\nu > 2$

6.3 ค่าความเบ้ (Skewness) เท่ากับ $\sqrt{8/\nu}$

6.4 $\chi^2_{\nu_1} + \chi^2_{\nu_2}$ มีการแจกแจงเป็น $\chi^2_{\nu_1 + \nu_2}$ เมื่อ ν_1 และ ν_2 เป็นชั้นความเป็นอิสระ

6.5 เมื่อ ν มีค่ามากขึ้นการแจกแจงไคสแควร์จะเข้าใกล้การแจกแจงแบบปกติที่มีค่าเฉลี่ยเท่ากับ ν ความแปรปรวนเท่ากับ 2ν

3. สถิติทดสอบในงานวิจัยนี้

3.1 สถิติทดสอบเอฟ (F – test)

เนื่องจากสถิติทดสอบเอฟสามารถใช้ทดสอบได้หลายรูปแบบ แต่ที่ใช้ทดสอบในงานวิจัยนี้เป็นการทดสอบสมมติฐานที่ระบุถึงความเท่ากันของค่าเฉลี่ยของประชากรกลุ่มตัวอย่างที่สุ่มมาจากประชากรมากกว่า 2 กลุ่มขึ้นไป ที่เป็นอิสระต่อกัน เมื่อลักษณะข้อมูลอยู่ในมาตราอันตรภาค (Interval Scale) เป็นอย่างน้อย

3.1.1 ข้อตกลงเบื้องต้น (Kirk 1982: 56)

1. กลุ่มตัวอย่างถูกสุ่มมาจากประชากรที่แจกแจงแบบปกติ
2. กลุ่มตัวอย่างเป็นอิสระต่อกัน
3. อัตราส่วนเอฟเป็นค่าประมาณจากประชากรที่มีความแปรปรวนเท่ากัน
4. เศษและส่วนของอัตราส่วนเอฟเป็นอิสระต่อกัน

3.1.2 การคำนวณค่าทดสอบ

ถ้า $X_{11}, X_{21}, \ldots, X_{N1}$

$X_{12}, X_{22}, \ldots, X_{N2}$

$X_{13}, X_{23}, \ldots, X_{N3}$

$\vdots \quad \vdots \quad \vdots$

$X_{1k}, X_{2k}, \ldots, X_{Nk}$ เป็นค่าของตัวแปรจากกลุ่มตัวอย่างที่สุ่มมาจากประชากร K กลุ่ม จะคำนวณค่าทดสอบเอฟได้จากสูตร

$$F = \frac{1/(K-1) \sum_{k=1}^{K} N_k (\bar{X}_k - \bar{X})^2}{1/(N-K) \sum_{k=1}^{K} (N_k - 1) S_k^2}$$

เมื่อ K เป็นจำนวนกลุ่มตัวอย่าง

N_k เท่ากับจำนวนค่าสังเกตของกลุ่มที่ k

$\bar{X}_k$ เท่ากับค่าเฉลี่ยของค่าสังเกตกลุ่มที่ k

S_k^2 เท่ากับความแปรปรวนของค่าสังเกตกลุ่มที่ k

$\bar{X}$ เท่ากับค่าเฉลี่ยของค่าสังเกตทั้งหมด

N เท่ากับจำนวนค่าสังเกตทั้งหมด

ค่าสถิติเอฟ จะมีการแจกแจงด้วยชั้นความเป็นอิสระเท่ากับ $\nu_1 = K - 1$ และ $\nu_2 = N - K$

3.2 สถิติทดสอบเอฟสตาร์ (F^* - test)

สถิติทดสอบเอฟสตาร์ หรืออาจเรียกว่า F - like ซึ่งหมายถึงสถิติทดสอบที่สร้างขึ้นมาใช้แทนการทดสอบเอฟ ในกรณีที่ข้อตกลงเบื้องต้นด้านความแปรปรวนไม่เป็นจริง โดยมีนักสถิติหลายท่านเสนอวิธีการทดสอบ และประมาณค่าชั้นแห่งความอิสระในรูปแบบต่าง ๆ ได้แก่ Satterthwaite (1941), Welch (1951) และ Marascuilo (1971) เป็นต้น และจะประมาณการแจกแจงของการทดสอบด้วยการแจกแจงแบบเอฟทั้งสิ้น

ในการวิจัยครั้งนี้ได้เลือกใช้ สถิติทดสอบเอฟสตาร์ของ Marascuilo ซึ่งมีที่มาจากสถิติทดสอบยู (U - test) แรกเริ่มของการปรับปรุงสถิติทดสอบเอฟสตาร์มาจาก James (1951: 324-329) พบปัญหาในการใช้สถิติทดสอบยู ซึ่งประมาณได้ด้วยไคสแควร์ที่ชั้นแห่งความอิสระ K - 1 เมื่อ K คือ ขนาดของกลุ่มตัวอย่างนั้น การใช้สถิติทดสอบนี้จะ accuarte สำหรับกลุ่มตัวอย่างขนาดใหญ่

ดังนั้น James จึงคิดแก้ปัญหากรณีที่จะใช้ในกลุ่มตัวอย่างขนาดเล็ก โดยให้ $h(w_1, W_2, \ldots, w_k, P)$ เป็นฟังก์ชั่นของ $h(w)$

และ $P_r\left[\sum_k w_k(\bar{X}_k - \bar{X}^*)^2 > h(w)\right] = P$ การประมาณค่าวิกฤตของ $\sum_k w_k(\bar{X}_k - \bar{X}^*)^2$ เป็น

$$h(w) = \chi^2\left[1 + \frac{3\chi^2 + (K+1)}{2(K^2-1)} \sum_k \frac{1}{N_k - 1}\left(1 - \frac{w_k}{\sum w_k}\right)^2\right]$$

เมื่อ $\sum_k w_k(\bar{X}_k - \bar{X}^*)^2$ คือค่าของการทดสอบยู

K คือจำนวนขนาดกลุ่มตัวอย่าง

N_k คือจำนวนตัวอย่างในแต่ละกลุ่ม

χ^2 คือค่าไคสแควร์ ที่ชั้นความเป็นอิสระเท่ากับ $K - 1$

แต่ James ก็ไม่สามารถแก้ปัญหาในกรณีกลุ่มตัวอย่างขนาดเล็กได้ดี

ดังนั้นในปีเดียวกันนี้ Welch จึงได้คิดแก้ปัญหาต่อ โดยการ Generate ฟังก์ชั่นโมเมนต์ ของ $\sum_k w_k(\bar{X}_k - \bar{X}^*)^2$ ขึ้นใหม่

โดยที่ $\sum_k w_k(\bar{X}_k - \bar{X}^*)^2$ จะมีการแจกแจงเป็น $\{(K-1) + A/\nu_2\}$ ของการแจกแจงเอฟ

เมื่อ $\nu_1 = K - 1$

$$\nu_2 = \left\{\frac{3}{K^2-1} \sum_k \frac{1}{N_k - 1}\left(1 - \frac{w_k}{\sum w_k}\right)^2\right\}^{-1}$$

และ $$\frac{A}{\nu_2} = \frac{2(K-2)}{K+1} + \sum_k \frac{1}{N_k - 1}\left(1 - \frac{w_k}{\sum w_k}\right)^2$$

Welch จึงเสนอสถิติทดสอบขึ้นใหม่ สำหรับใช้กับกลุ่มตัวอย่างขนาดเล็กในรูปของ v^2 เป็น

$$v^2 = \frac{\sum_{k=1}^{K} w_k (\bar{X}_k - \bar{X}^*)^2 / K - 1}{\left[1 + \frac{2(K-2)}{K^2 - 1} \sum_{k=1}^{K} \frac{1}{N_k - 1} \left(1 - \frac{w_k}{\Sigma w_k}\right)^2\right]}$$

v^2 จะประมาณการแจกแจงด้วยเอฟ ที่ $\nu_1 = K - 1$

$$\text{และ } \nu_2 = \left\{ \frac{3}{K^2 - 1} \sum_k \frac{1}{N_k - 1} \left(1 - \frac{w_k}{\Sigma w_k}\right)^2 \right\}^{-1}$$

ต่อมาในปี ค.ศ. 1971 Marascuilo ได้เสนอสถิติทดสอบเอฟสตาร์ ในการวิเคราะห์ความแปรปรวน เมื่อข้อตกลงเบื้องต้นของความแปรปรวนไม่เป็นจริง โดยตัดส่วนของ v^2 ของ Welch และคำนวณชั้นความอิสระเช่นเดียวกับสถิติทดสอบของ Welch ซึ่งจะประมาณการแจกแจงด้วยเอฟ ค่าของสถิติทดสอบเอฟสตาร์จึงเป็น

$$\sum_{k=1}^{K} w_k (\bar{X}_k - \bar{X}^*)^2 / K - 1$$

ตามที่ Marascuilo เสนอสถิติทดสอบเอฟสตาร์เช่นข้างต้น มาจากเหตุผลดังนี้

$$\text{จาก} \quad F = \frac{MS_B}{MS_w}$$

$$\text{จะได้} \quad F = \frac{1}{K-1} \sum_{k=1}^{K} \frac{N_k (\bar{X}_k - \bar{X})^2}{S_p^2}$$ โดยใช้ S_p^2 แทน MS_w

$$= \frac{1}{K-1} \left[\sum_{k=1}^{K} \frac{N_k}{S_p^2} (\bar{X}_k - \bar{X})^2 \right]$$

$$\text{และ} \quad F^* = \frac{1}{K-1} \left[\sum_{k=1}^{K} \frac{N_k}{S_k^2} (\bar{X}_k - \bar{X}^*)^2 \right]$$

$$= \frac{1}{K-1} \sum_{k=1}^{K} w_k (\bar{X}_k - \bar{X}^*)^2$$

เมื่อ $w_k = \frac{N_k}{S_k^2}$

และ $\bar{X}^* = \frac{\sum_{k=1}^{K} w_k \bar{X}_k}{\sum_{k=1}^{K} w_k}$

ถ้าใช้ S_p^2 แทน S_k^2 จะได้ $\bar{X}^* = \bar{X}$

$$\bar{X}^* = \frac{\sum_{k=1}^{K} (N_k/S_k^2)\ \bar{X}_k}{\sum_{k=1}^{K} N_k/S_k^2}$$

$$= \frac{\sum_{k=1}^{K} (N_k/S_p^2)\ \bar{X}_k}{\sum_{k=1}^{K} N_k/S_p^2}$$

$$= \frac{1/S_p^2 \sum_{k=1}^{K} N_k \bar{X}_k}{1/S_p^2 \sum_{k=1}^{K} N_k}$$

$$= \frac{\sum_{k=1}^{K} N_k \bar{X}_k}{\sum_{k=1}^{K} N_k}$$

$$= \frac{1}{N} \sum_{k=1}^{K} N_k \bar{X}_k$$

$$= \bar{X}$$

3.2.1 ข้อตกลงเบื้องต้น (Marascuilo 1971: 366)

1. ประชากรมีการแจกแจงแบบปกติ

2. กลุ่มตัวอย่างแต่ละกลุ่มได้รับการสุ่มจากประชากร

3. กลุ่มตัวอย่างแต่ละกลุ่มเป็นอิสระต่อกัน

3.2.2 การคำนวณค่าทดสอบ

ถ้าให้ $X_{11}, X_{21}, \ldots, X_{N1}$

$X_{12}, X_{22}, \ldots, X_{N2}$

$X_{13}, X_{23}, \ldots, X_{N3}$

$\vdots$

$X_{1k}, X_{2k}, \ldots, X_{Nk}$ เป็นค่าของตัวแปรสุ่มจากกลุ่มตัวอย่าง K กลุ่ม ที่เป็นอิสระต่อกัน เมื่อลักษณะข้อมูลอยู่ในมาตราอันตรภาคเป็นอย่างน้อย จะคำนวณค่าทดสอบเอฟสตาร์ได้จากสูตร

$$F^* = \frac{1}{K-1} \sum_{k=1}^{K} w_k (\bar{X}_k - \bar{X}^*)^2$$

เมื่อ $w_k = N_k / S_k^2$

และ $$\bar{X}^* = \frac{\sum_{k=1}^{K} w_k \bar{X}_k}{\sum_{k=1}^{K} w_k}$$

K เป็นจำนวนกลุ่มตัวอย่าง

N_k เท่ากับจำนวนค่าสังเกตกลุ่มที่ k

$\bar{X}_k$ เท่ากับค่าเฉลี่ยของค่าสังเกตกลุ่มที่ k

S_k^2 เท่ากับความแปรปรวนของค่าสังเกตกลุ่มที่ k

F^* จะถูกประมาณด้วยการแจกแจงเอฟ ที่ชั้นความเป็นอิสระ $\nu_1 = K - 1$ และ $\nu_2 = \dot{\nu}_2^*$

$$\text{เมื่อ} \quad \nu_2^* = \frac{K^2 - 1}{3\Lambda}$$

$$\text{โดยที่} \quad \Lambda = \sum_{k=1}^{K} \frac{1}{N_k - 1}\left(1 - \frac{w_k}{w}\right)^2$$

$$\text{และ} \quad w = \sum_{k=1}^{K} w_k$$

3.3 <u>สถิติทดสอบยู</u> (U - test)

L.A. Marascuilo เป็นผู้เสนอและเน้นบทบาทการใช้ในปี ค.ศ. 1966 ซึ่งปรากฏในบทความเรื่องการเปรียบเทียบพหุคูณกลุ่มตัวอย่างขนาดใหญ่ (Large Sample Multipel Comparisons) เพื่อแก้ปัญหาในการใช้สถิติทดสอบสำหรับการวิเคราะห์ความแปรปรวน เมื่อข้อตกลงเบื้องต้นของความแปรปรวนประชากรไม่เป็นจริงของสถิติทดสอบ เอฟ โดยที่การทดสอบนี้ใช้การแจกแจงไคสแควร์ ที่ชั้นความเป็นอิสระ เท่ากับ K - 1 เมื่อ K เป็นจำนวนของกลุ่มตัวอย่าง และแสดงได้ดังนี้ (Marascuilo 1977: 32-33)

จากคำจำกัดความของตัวแปรตาม K ตัว ของการแจกแจงแบบไคสแควร์ เซทของ $(Z_1, Z_2, ..., Z_k)$ ซึ่งมี $N(0, 1)$

จะได้ $\chi_\nu^2 = Z_1^2 + Z_2^2 + ... + Z_k^2$ เป็นการแจกแจงแบบไคสแควร์ โดยมีค่าพารามิเตอร์ $\nu = K$ และค่า ν นี้เรียกว่าชั้นความเป็นอิสระ

ในกรณีเฉพาะของคำจำกัดความนี้ คือ ประชากร K ซึ่งมีค่าพารามิเตอร์ $[\theta_1, \theta_2, ..., \theta_K]$ และประมาณค่าโดย $[\hat{\theta}_1, \hat{\theta}_2, ... \hat{\theta}_K]$ โดยตกลงว่า $\hat{\theta}_K$ มีค่าใกล้เคียง $N(\theta_0\,, \sigma^2_{\hat{\theta}_k})$

$$\text{โดยที่} \quad Z_1 = \frac{\hat{\theta}_1 - \theta_0}{\sigma_{\hat{\theta}_1}}, \quad Z_2 = \frac{\hat{\theta}_2 - \theta_0}{\sigma_{\hat{\theta}_2}}, \quad ..., \quad Z_K = \frac{\hat{\theta}_K - \theta_0}{\sigma_{\hat{\theta}_K}}$$

ซึ่ง θ_0 เป็นพารามิเตอร์ที่ทราบค่า และให้

$$\chi_\nu^2 = Z_1^2 + Z_2^2 + ... + Z_K^2 = \sum_{k=1}^{K} \frac{(\theta_k - \theta_0)^2}{\sigma^2_{\hat{\theta}_k}}$$

โดยคำจำกัดความของไคสแควร์ ตัวแปรสุ่มจะแจกแจงเป็นไคสแควร์ด้วย $\nu = K$ แต่ค่า χ^2 จะไม่สามารถคำนวณได้ เพราะว่า θ_0 โดยปกติเราไม่ทราบค่า ถ้า θ_0 ถูกประมาณโดย

$$\theta_0 = \frac{\frac{\hat{\theta}_1}{\sigma^2_{\hat{\theta}_1}} + \frac{\hat{\theta}_2}{\sigma^2_{\hat{\theta}_2}} + \ldots + \frac{\hat{\theta}_w}{\sigma^2_{\hat{\theta}_K}}}{\frac{1}{\sigma^2_{\hat{\theta}_1}} + \frac{1}{\sigma^2_{\hat{\theta}_2}} + \ldots + \frac{1}{\sigma^2_{\hat{\theta}_K}}} = \frac{w_1\hat{\theta}_1 + w_2\hat{\theta}_2 + \ldots + w_K\hat{\theta}_K}{w_1 + w_2\ 1 \ldots + w_K}$$

โดยที่ $w_1 = \frac{1}{\sigma^2_{\hat{\theta}_1}}$, $w_2 = \frac{1}{\sigma^2_{\hat{\theta}_2}}$, ..., $w_K = \frac{1}{\sigma^2_{\hat{\theta}_K}}$

และถ้า χ^2 ถูกกำหนดใหม่ในรูป

$$U = \sum_{k=1}^{K} \frac{(\hat{\theta}_k - \hat{\theta}_0)^2}{\sigma^2_{\hat{\theta}_k}} = \hat{Z}_1^2 + \hat{Z}_2^2 + \ldots + \hat{Z}_K^2$$

จึงแสดงได้ว่า U จะมีการประมาณการแจกแจงเป็น χ^2_ν ด้วย $\nu = K - 1$ เพราะการประมาณค่า θ_0 จากข้อมูล ชั้นความเป็นอิสระจะถูกใช้ไป 1 ค่า U จึงถูกจำกัดทำให้ไม่สามารถเปลี่ยนแปลงได้อย่างเป็นอิสระ สำหรับ $\sigma^2_{\hat{\theta}}$ แต่ละตัวที่ไม่ทราบค่า จะประมาณค่าจากข้อมูลและแทนค่าที่ประมาณนั้นใน U

3.3.1 <u>ข้อตกลงเบื้องต้น</u>

1. ประชากรมีการแจกแจงแบบปกติหรือกลุ่มตัวอย่างมีขนาดใหญ่
2. กลุ่มตัวอย่างได้รับการสุ่มจากประชากร
3. กลุ่มตัวอย่างแต่ละกลุ่มเป็นอิสระต่อกัน

3.3.2 <u>การคำนวณค่าทดสอบ</u>

ถ้าให้ $X_{11}, X_{21}, \ldots, X_{N1}$

$X_{12}, X_{22}, \ldots, X_{N2}$

$X_{13}, X_{23}, \ldots, X_{N3}$

$\vdots$

$X_{1k}, X_{2k}, \ldots, X_{Nk}$ เป็นค่าของตัวแปรสุ่มจากกลุ่มตัวอย่าง K กลุ่มที่เป็นอิสระต่อกัน เมื่อลักษณะข้อมูลอยู่ในมาตราอันตรภาคเป็นอย่างน้อย จะคำนวณค่าทดสอบยู ได้จากสูตร

$$U = \sum_{k=1}^{K} w_k (\bar{X}_k - \bar{X}^*)^2$$

เมื่อ $w_k = N_k/S_k^2$

และ $\bar{X}^* = \sum_{k=1}^{K} w_k \cdot \bar{X}_k / \sum_{k=1}^{K} w_k$

K เป็นจำนวนกลุ่มตัวอย่าง

N_k เท่ากับจำนวนค่าสังเกตกลุ่มที่ k

$\bar{X}_k$ เท่ากับค่าเฉลี่ยของค่าสังเกตกลุ่มที่ k

S_k^2 เท่ากับความแปรปรวนของค่าสังเกตกลุ่มที่ k

การแจกแจง U จะถูกประมาณด้วยไคสแควร์ที่ชั้นความเป็นอิสระเท่ากับ $K - 1$

4. <u>งานวิจัยที่เกี่ยวข้อง</u>

Rogan และ Keselman (1977: 493-498) ได้ศึกษาความแกร่งของสถิติทดสอบเอฟ เมื่อความแปรปรวนของประชากรไม่เท่ากัน สำหรับกลุ่มตัวอย่างขนาดเท่ากัน โดยใช้สัมประสิทธิ์ความแปรผัน ตามที่ Box ได้ศึกษาไว้ในปี 1954 พบว่า แนวโน้มต่อไปนี้ปรากฏมากที่สุด กล่าวคือ ความแตกต่างของความแปรปรวนของประชากรยิ่งมีมากขึ้น ความแตกต่างระหว่างการประมาณค่าโดยวิธีเชิงประจักษ์ (empirical estimate) กับระดับนัยสำคัญ (nominal significance levels) จะมากขึ้นด้วย และยิ่งขนาดของกลุ่มตัวอย่างใหญ่ขึ้นเท่าไร ผลกระทบจากความไม่เท่ากันของความแปรปรวนของประชากรก็น้อยลง แม้กระนั้นก็ดี เมื่อกลุ่มตัวอย่างใหญ่ แต่ความไม่เท่ากันของความแปรปรวนของประชากรใหญ่ด้วยอัตราความคลาดเคลื่อนจากการทดลองจะใหญ่กว่าอัตราความคลาดเคลื่อนที่ระบุอย่างน้อย 2-4 % ซึ่งขัดแย้งกับที่ Box รายงานไว้ในปี 1954 ว่าน่าจะเกิดขึ้นเมื่อความแปรปรวนของประชากรมีขนาดใหญ่มากและกลุ่มตัวอย่างมีขนาดเล็ก นั่นก็คือ ANOVA F - test ไม่แกร่งต่อความไม่เท่ากันของความแปรปรวนของประชากรทั้งหมด เมื่อกลุ่มตัวอย่างมีขนาดเท่ากัน และถ้าจะสรุปในกรณีกลุ่มตัวอย่างเท่ากัน ก็ต้องบอกเงื่อนไขของความไม่เท่ากันของความแปรปรวนด้วย

Tomarken และ Serlin (1986: 90-99) ได้ศึกษาผลที่มีต่อการฝ่าฝืนข้อตกลงเบื้องต้นในด้านความเป็นเอกพันธุ์ของความแปรปรวนของสถิติทดสอบเอฟ โดยศึกษากับกลุ่มตัวอย่างขนาดเท่ากัน คือ ขนาด 12 และ 20 เมื่อกำหนดจำนวนกลุ่มตัวอย่างเป็น 3 กลุ่ม และ 4 กลุ่ม กำหนดความแตกต่างของความแปรปรวนเป็นอัตราส่วน คือ 12:4:1 สำหรับกลุ่มตัวอย่าง 3 กลุ่ม และ 12:6:4:1 สำหรับกลุ่มตัวอย่าง 4 กลุ่ม พบว่า การทดสอบเอฟ ไม่สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ตามที่กำหนด เมื่อกลุ่มตัวอย่างมีขนาด 3 กลุ่ม ทั้งระดับอัตราความคลาดเคลื่อนที่ระบุเท่ากับ .05 และ .01 และเมื่อกลุ่มตัวอย่างมีขนาด 4 กลุ่ม ที่ระดับอัตราความคลาดเคลื่อนที่ระบุเท่ากับ .01 โดยทุกกรณีที่ไม่สามารถควบคุมได้นี้ อัตราความคลาดเคลื่อนประเภทที่ 1 มากกว่าที่กำหนด ซึ่งสอดคล้องกับผลสรุปจากการศึกษาของ Scheffé (1959: 354) ที่พบว่า กรณีกลุ่มตัวอย่างเท่ากัน เมื่อความแปรปรวนของประชากรแตกต่างกัน อัตราความคลาดเคลื่อนประเภทที่ 1 จะมากกว่าอัตราความคลาดเคลื่อนที่ระบุ

Budescu และ Appelbaum (1981: 55-74) ใช้เทคนิคมอนติคาร์โลศึกษาผลของการแปลงข้อมูลเพื่อให้ได้ค่าความแปรปรวนมีเสถียรภาพ (Variance Stabilizing Transformation) ที่มีต่ออำนาจการทดสอบเอฟ โดยจำแนกการแจกแจงของประชากรเป็น 5 แบบ คือ แบบไบโนเมียล แบบปกติที่มีค่ามัชฌิมเลขคณิตและค่าความแปรปรวนเท่ากับการแจกแจงแบบไบโนเมียล แบบพัวซอง แบบปกติที่มีค่ามัชฌิมเลขคณิตและค่าความแปรปรวนเท่ากับการแจกแจงแบบพัวซอง และการแจกแจงปกติ เมื่อความแปรปรวนของประชากรไม่เท่ากัน และกลุ่มตัวอย่างมีขนาด 4 กลุ่มไม่เท่ากัน พบว่า การแปลงข้อมูลไม่มีผลต่ออำนาจการทดสอบเอฟ ที่ระดับนัยสำคัญ .05 และ .01 นอกจากนี้ยังทำให้อำนาจการทดสอบเอฟ มีค่าลดลง โดยเฉพาะในกรณีกลุ่มตัวอย่างขนาดเล็กมาจากประชากรที่มีความแปรปรวนสูง และกลุ่มตัวอย่างขนาดใหญ่มาจากประชากรที่มีความแปรปรวนต่ำ

Brown และ Forsythe (1974: 129-132) ศึกษาความแกร่ง และอำนาจการทดสอบของสถิติทดสอบเอฟ สถิติทดสอบเอฟสตาร์ของ Satterthwaite สถิติทดสอบของ Welch และสถิติทดสอบของ James ทดสอบความเท่ากันของมัชฌิมเลขคณิต เมื่อความแปรปรวนของประชากรเท่ากันและไม่เท่ากัน โดยศึกษากลุ่มตัวอย่างขนาดเล็ก เมื่อกลุ่มตัวอย่างเท่ากัน และมีขนาด 4 และ 11 สำหรับจำนวนกลุ่มตัวอย่าง 4 กลุ่ม กลุ่มตัวอย่างเท่ากันขนาด 4, 6, 11, 16 และ 21 สำหรับจำนวนกลุ่มตัวอย่าง 6 กลุ่ม และกลุ่มตัวอย่างเท่ากันขนาด 21 สำหรับจำนวนกลุ่มตัวอย่าง 8 กลุ่ม กรณีกลุ่มตัวอย่างขนาดไม่เท่ากันศึกษากลุ่มตัวอย่างขนาด (4, 8, 10, 12), (11, 16, 16, 21) และ (4, 6, 6, 8, 10, 12) พบว่า

1. สถิติทดสอบเอฟไม่แกร่งสำหรับความแปรปรวนของประชากรไม่เท่ากัน โดยเฉพาะเมื่อกลุ่มตัวอย่างมีขนาดเล็ก และขนาดตัวอย่างไม่เท่ากัน

2. เมื่อความแปรปรวนของประชากรเท่ากัน สถิติทดสอบเอฟสตาร์ และสถิติทดสอบของ Welch มีอำนาจการทดสอบใกล้เคียงกับสถิติทดสอบเอฟ

3. ในกลุ่มตัวอย่างขนาดเล็กสถิติทดสอบของ James จะปฏิเสธสมมติฐานสูญบ่อยครั้ง

4. เมื่อสุดขีดของค่าเฉลี่ย (extreme mean) มีความแปรปรวนน้อยหรือมีความแปรปรวนมาก สถิติทดสอบของ Welch จะมีอำนาจการทดสอบมากกว่าสถิติทดสอบเอฟสตาร์

Clinch และ Keselman (1982: 207-214) ศึกษาความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบของสถิติทดสอบเอฟ สถิติทดสอบเอฟสตาร์ของ Satterthwaite และสถิติทดสอบของ Welch เมื่อประชากรมีการแจกแจง 3 แบบ คือ แบบปกติ แบบไคสแควร์ (df = 2) และแบบที (df = 5) ความแปรปรวนของประชากรเท่ากันและแตกต่างกัน เมื่อความแปรปรวนของประชากรแตกต่างกันมี 4 กรณี คือ

$$[\sigma_1^2 = 0.7317 \text{ , } \sigma_2^2 = 0.9106 \text{ , } \sigma_3^2 = 1.0894 \text{ , } \sigma_4^2 = 1.2683]$$

$$[\sigma_1^2 = 0.4633 \text{ , } \sigma_2^2 = 0.8211 \text{ , } \sigma_3^2 = 1.1789 \text{ , } \sigma_4^2 = 1.5367]$$

$$[\sigma_1^2 = 0.1950 \text{ , } \sigma_2^2 = 0.7317 \text{ , } \sigma_3^2 = 1.2683 \text{ , } \sigma_4^2 = 1.8050]$$

$$[\sigma_1^2 = 0.1515 \text{ , } \sigma_2^2 = 0.4343 \text{ , } \sigma_3^2 = 0.7172 \text{ , } \sigma_4^2 = 2.6971]$$

ขนาดของกลุ่มตัวอย่างจำนวน 4 กลุ่ม กรณีกลุ่มตัวอย่างขนาดเท่ากันมีขนาด 12 และ 36 กรณีกลุ่มตัวอย่างไม่เท่ากันมีขนาด (6, 10, 14, 18) และ (18, 30, 42, 54) พบว่า

1. เมื่อการแจกแจงของประชากรเป็นแบบปกติ ความแปรปรวนของประชากรแตกต่างกันและกลุ่มตัวอย่างมีขนาดเท่ากัน สถิติทดสอบเอฟจะมีอัตราความคลาดเคลื่อนจากการทดลองใหญ่กว่า α = .05 จำนวน 6 กรณีในทั้งหมด 45 กรณี และในกรณีเมื่อกลุ่มตัวอย่างไม่เท่ากันคู่กับความแปรปรวนไม่เท่ากันในลักษณะ $n_1 < n_2 < n_3 < n_4$ กับ $\sigma_1^2 > \sigma_2^2 > \sigma_3^2 > \sigma_4^2$ สถิติทดสอบเอฟจะไม่สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ตามเกณฑ์ของ Bradley (1978) ในขณะที่สถิติทดสอบ F^* จะสามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ดีในเงื่อนไขนี้

2. เมื่อประชากรมีการแจกแจงแบบที (df = 5) สถิติทดสอบของ Welch และสถิติทดสอบเอฟสตาร์ จะสามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ตามเกณฑ์ที่กำหนดในเงื่อนไข $n_1 < n_2 < n_3 < n_4$ กับ $\sigma_1^2 > \sigma_2^2 > \sigma_3^2 > \sigma_4^2$ ซึ่งสถิติทดสอบเอฟไม่สามารถควบคุมได้

3. อำนาจการทดสอบของสถิติทดสอบใกล้เคียงกัน สำหรับกลุ่มตัวอย่างขนาดเท่ากัน และสถิติทดสอบของ Welch มีอำนาจการทดสอบมากที่สุด ตามด้วยสถิติทดสอบเอฟ และสถิติทดสอบเอฟสตาร์ สำหรับกรณี $n_1 < n_2 < n_3 < n_4$ คู่กับ $\sigma_1^2 > \sigma_2^2 > \sigma_3^2 > \sigma_4^2$ ไม่สามารถเปรียบเทียบกันได้ เพราะว่ามีสถิติทดสอบ เอฟสตาร์เท่านั้น ที่สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ตามเกณฑ์ที่กำหนด