บทที่ 2

ตัวสถิติและผลงานวิจัยที่เกี่ยวข้อง

ตัวสถิติที่ใช้ในการทดสอบการเท่ากันของความแปรปรวนของประชากรมีหลายแบบ ซึ่งได้มีการพัฒนาขึ้นเรื่อย ๆ พร้อมทั้งได้มีผลงานวิจัยศึกษาเปรียบเทียบอำนาจการทดสอบของตัวสถิติต่าง ๆ ซึ่งจะได้กล่าวเกี่ยวกับตัวสถิติที่ศึกษาครั้งนี้พร้อมทั้งผลงานวิจัยที่เกี่ยวข้องพอเป็นสังเขปดังนี้

2.1 ตัวสถิติที่ใช้ในการศึกษา

2.1.1 ตัวสถิติ χ^2 หรือตัวสถิติของบาร์ตเลต (Bartlett's Statistic)

$$\chi^2_{k-1} = \frac{\sum_i^k (n_i - 1)\ \ell n s^2 - \sum_i^k (n_i - 1)\ \ell n s_i^2}{1 + \frac{1}{3(k-1)}\left[\sum_i^k \frac{1}{n_i - 1} - \frac{1}{\sum_i^k (n_i - 1)}\right]}$$

เมื่อ $s_i^2 = \frac{1}{n_i - 1} \sum_j^{n_i} (x_{ij} - x_{i.})^2 \ ; \ i = 1,2, \ldots, k \ , \ j = 1,2,\ldots,n_i$

$$s^2 = \frac{\sum_i^k (n_i - 1)\ s_i^2}{\sum_i^k (n_i - 1)}$$

n_i ขนาดของกลุ่มตัวอย่างที่ i

k จำนวนประชากร

จะปฏิเสธสมมติฐานว่าง (H_o) เมื่อค่าสถิติที่คำนวณได้มากกว่าค่า χ^2 ที่องศาความเป็นอิสระ (degree of freedom) เท่ากับ k - 1

ตัวสถิติของบาร์ตเลต พัฒนามาจากตัวสถิติของนีแมน-เปียร์สัน ใช้เพื่อทดสอบการเท่ากันของความแปรปรวนของประชากร k ประชากร สมมติฐานในการทดสอบคือ H_o : $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2 = \sigma^2$ และ H_A : ความแปรปรวนของประชากรอย่างน้อย 2 ประชากรไม่เท่ากัน Neyman สร้างตัวสถิติทดสอบโดยใช้ อัตราส่วนภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Ratio)

$$\lambda = \frac{L\ (\omega)}{L\ (\Omega)}$$

เมื่อ L (ω) คือ ฟังก์ชั่นภาวะน่าจะเป็น (Likelihood function) ของระวางที่พารามิเตอร์ (parameter space) ภายใต้สมมติฐานว่าง (H_o)

L (Ω) คือ ฟังก์ชั่นภาวะน่าจะเป็นของระวางที่พารามิเตอร์ (parameter space) ภายใต้สมมติฐานว่างและสมมติฐานแย้ง (alternative hypothesis) หรือ H_A

ถ้ากำหนดให้ X_{ij} ; i = 1,2,3, ..., k และ j = 1,2, ..., n_i มีการแจกแจงแบบปกติที่มีค่าเฉลี่ยเท่ากับ μ_i และความแปรปรวน σ_i^2 หรือ $X_{ij} \sim N(\mu_i\ ,\ \sigma_i^2)$ ฟังก์ชั่นความหนาแน่น (probability density function) หรือ pdf. ของ X_{ij} คือ

$$p(x_{ij}) = \frac{1}{\sqrt{2\pi}\ \sigma_i} e^{-(x_{ij} - \mu_i)^2/2\sigma_i^2}$$

ฟังก์ชั่นภาวะน่าจะเป็น คือ

$$L = \prod_{j=1}^{n_i} p\ (x_{1j})\ p\ (x_{2j}) \ldots p(x_{kj})$$

$$= \frac{1}{(2\ \pi)^{\frac{n}{2}}\ \sigma_1^{n_1} \ldots \sigma_k^{n_k}} \exp\left[-\frac{1}{2} \sum_{i=1}^{k} \sum_{i=1}^{n_i} \frac{(x_{ij} - \mu_i)^2}{\sigma_i}\right] \text{--------}(1)$$

เมื่อ $\sum_{i=1}^{k} n_i = n$

$$\ell nL = -\frac{n}{2}\ell n\, 2\pi - n_1 \ell n\, \sigma_1 - \cdots - n_k \ell n\, \sigma_k - \frac{1}{2}\left[\sum_{i=1}^{k} \frac{1}{\sigma_i^2} \sum_{i=1}^{n_i} (x_{ij} - \mu_i)^2\right] \text{-- (2}$$

$$\frac{\partial \ell nL}{\partial \mu_i} = \frac{1}{\sigma_i^2} \sum_{j=1}^{n_i} (x_{ij} - \mu_i) = 0 \qquad \text{----------(3)}$$

$$\frac{\partial \ell nL}{\partial \sigma_i} = \frac{-n_i}{\sigma_i} + \frac{1}{\sigma_i^3} \sum_{j=1}^{n_i} (x_{ij} - \mu_i)^2 = 0 \qquad \text{----------(4)}$$

ดังนั้นตัวประมาณภาวะน่าจะเป็นสูงสุด (maximum likelihood estimators) สำหรับ Ω คือ

$$\hat{\mu}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij} = \bar{x}_{i.} \qquad \text{----------(5)}$$

$$\hat{\sigma}_i^2 = \frac{1}{n_i} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{i.})^2 \qquad \text{---------(6)}$$

แทน (5) และ (6) ใน (1) จะได้

$$L(\hat{\Omega}) = \frac{1}{(2\pi)^{\frac{n}{2}} \hat{\sigma}_1^{n_1} \cdots \hat{\sigma}_k^{n_k}} \exp\left\{-\frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \left[\frac{(x_{ij} - \bar{x}_{i.})^2}{\sum_{j}^{n_i} (x_{ij} - \bar{x}_{i})^2 / n_i}\right]\right\}$$

$$= \frac{1}{(2\pi)^{\frac{n}{2}} \hat{\sigma}_1^{n_1} \cdots \hat{\sigma}_1^{n_k}} \exp\left\{-\frac{1}{2} \sum_{i=1}^{k} \left[\frac{\sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{i.})^2}{\sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{i.})^2 / n_i}\right]\right\}$$

$$L(\hat{\Omega}) = \frac{1}{(2\pi)^{\frac{n}{2}}\ \hat{\sigma}_1^{n_1} \ldots \hat{\sigma}_k^{n_k}} \exp\left(-\frac{1}{2}\sum_{i=1}^{k} n_i\right)$$

$$= \frac{e^{-\frac{n}{2}}}{(2\pi)^{\frac{n}{2}}\ \hat{\sigma}_1^{n_1} \ldots \hat{\sigma}_k^{n_k}} \quad ; \quad \left(\sum_{i=1}^{k} n_i = n\right) \qquad \text{---------(7)}$$

$L(\omega)$ **ภายใต้**สมมติฐานว่าง ซึ่ง μ_i มีค่าแปรเปลี่ยน (vary) แต่ $\sigma_i^2 = \sigma^2$ จะได้

$$L(\omega) = \frac{1}{(2\pi)^{\frac{n}{2}}\ \sigma^{\sum_{i}^{k} n_i}} \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij}-\mu_i)^2\right]\right\}$$

$$L(\omega) = \frac{1}{(2\pi)^{\frac{n}{2}}\ \sigma^n} \exp\left[-\frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij}-\mu_i)^2}{2\sigma^2}\right] \qquad \text{---------(8)}$$

$$\ln L(\omega) = -\frac{n}{2}\ln 2\pi - n\ln\sigma - \frac{1}{2\sigma^2}\left[\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij}-\mu_i)^2\right] \qquad \text{----------(9)}$$

$$\frac{\partial \ln L(\omega)}{\partial \mu_i} = \frac{1}{\sigma^2}\sum_{j=1}^{n_i}(x_{ij}-\mu_i) = 0 \qquad \text{---------(10)}$$

$$\frac{\partial \ln L(\omega)}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij}-\mu_i)^2 = 0 \qquad \text{---------(11)}$$

ตัวประมาณภาวะน่าจะเป็นสูงสุดสำหรับ ω คือ

$$\hat{\mu}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij} = \bar{x}_i \qquad \text{---------(12)}$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_k}(x_{ij}-\bar{x}_{i.})^2 \qquad \text{---------(13)}$$

แทนค่า (6) ใน (13) จะได้

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{k} n_i \, \hat{\sigma}_i^2 \qquad \text{---------(14)}$$

แทนค่า (12) และ (14) ใน (8)

$$L(\hat{\omega}) = \frac{1}{(2\pi)^{\frac{n}{2}} \left(\sum_{1}^{k} n_i \, \hat{\sigma}_i^2 / n\right)^{\frac{n}{2}}} \exp\left[-\frac{1}{2} \frac{\sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{i.})^2}{\sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{i.})^2 / n} \right]$$

$$= \frac{1}{(2\pi)^{\frac{n}{2}} \left(\sum_{i=1}^{k} n_i \hat{\sigma}_i^2 / n\right)^{\frac{n}{2}}} \exp\left(-\frac{n}{2}\right)$$

$$= \frac{e^{-\frac{n}{2}}}{(2\pi)^{\frac{n}{2}} \left(\sum_{i=1}^{k} n_i \, \hat{\sigma}_i^2 / n\right)^{\frac{n}{2}}} \qquad \text{---------(15)}$$

อัตราส่วนภาวะน่าจะเป็นสูงสุด คือ

$$\lambda = \frac{L(\hat{\omega})}{L(\hat{\Omega})}$$

$$= \frac{e^{-\frac{n}{2}}}{(2\pi)^{\frac{n}{2}} \left(\sum_{i=1}^{k} n_i \, \hat{\sigma}_i^2 / n\right)^{\frac{n}{2}}} \cdot \frac{(2\pi)^{\frac{n}{2}} \hat{\sigma}_1^{n_1} \cdots \hat{\sigma}_k^{n_k}}{e^{-\frac{n}{2}}}$$

$$= \frac{\hat{\sigma}_1^{n_1} \cdots \hat{\sigma}_k^{n_k}}{\left(\sum_{i=1}^{k} n_i \hat{\sigma}_i^2 / n\right)^{\frac{n}{2}}}$$

$$\lambda = \frac{\prod_{i=1}^{k} \hat{\sigma}_i^{n_i}}{\hat{\sigma}^n}$$

Ω มีพารามิเตอร์ 2k ตัวคือ μ_i และ σ_i ; i=1, 2, ... ,k และ ω มีพารามิเตอร์ k+1 ตัวคือ μ_i และ σ ซึ่ง Wilks (1938) ได้แสดงให้เห็นว่าถ้ากลุ่มตัวอย่างมีขนาดใหญ่เพียงพอแล้วการแจกแจงของ $-2\ \ell n\ \lambda$ จะมีแนวโน้มเข้าสู่การแจกแจงแบบ χ^2 ที่องศาความเป็นอิสระ r-s เมื่อ r คือจำนวนพารามิเตอร์ภายใต้ Ω และ s คือจำนวนพารามิเตอร์ ภายใต้ ω ดังนั้น

$$-2\ \ell n\ \lambda = -2\ \ell n \frac{\prod_{i=1}^{k} \hat{\sigma}_i^{n_i}}{\hat{\sigma}^n}$$

$$= -\sum_{i=1}^{k} n_i\ \ell n\ \hat{\sigma}_i^2 + n\ \ell n\ \hat{\sigma}^2$$

$$= -\sum_{i=1}^{k} n_i\ \ell n\ \hat{\sigma}_i^2 + \sum_{i=1}^{k} n_i\ \ell n\ \hat{\sigma}^2$$

$$= -\sum_{i=1}^{k} n_i\ \ell n \frac{\hat{\sigma}_i^2}{\hat{\sigma}^2} \sim \chi^2_{k-1} \qquad \text{--------(16)}$$

Bartlett ได้ปรับปรุงตัวสถิติใน (16) โดยการแทนตัวประมาณภาวะน่าจะเป็นสูงสุดที่เอนเอียง(biased maximum likelihood estimators) ด้วยตัวประมาณที่ไม่เอนเอียง (unbiased estimators) ของ variances และเปลี่ยนองศาความเป็นอิสระ n_i เป็น n_i-1

กำหนดให้ $$B = -\sum_{i=1}^{k} (n_i-1)\ \ell n \frac{s_i^2}{s^2}$$

เมื่อ $$s_i^2 = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}(X_{ij} - \bar{X}_{i.})^2$$

$$s^2 = \frac{\sum_{i=1}^{k}(n_i - 1)s_i^2}{\sum_{i=1}^{k}(n_i-1)} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij} - \bar{X}_{i.})^2}{\sum_{i=1}^{k}(n_i-1)}$$

ซึ่งจะทำให้ $B/C \sim \chi^2_{k-1}$ เมื่อ $C = 1 + \frac{1}{3(k-1)}\left[\sum_{i=1}^{k}\frac{1}{n_i-1} - \frac{1}{\sum_{i=1}^{k}(n_i-1)}\right]$

ดังนั้นตัวสถิติของ Bartlett คือ

$$B/C = \frac{1}{C}\left[\sum_{i=1}^{k}(n_i-1)\ln s^2 - \sum_{i=1}^{k}(n_i-1)\ln s_i^2\right]$$

หรือในกรณีที่ใช้ฟังก์ชั่นลอการิทึม (Logarithm function) ฐาน 10 จะได้

$$B/C = \frac{2.3026}{C}\left[\sum_{i=1}^{k}(n_i-1)\log s^2 - \sum_{i=1}^{k}(n_i-1)\log s_i^2\right]$$

2.1.2 ตัวสถิติ W' (modified Levene's statistic)

$$W' = \frac{\sum_{i=1}^{P} n_i(\bar{Z}'_{i.} - \bar{Z}'_{..})^2/P-1)}{\sum_{i=1}^{P}\sum_{j=1}^{n_i}(Z'_{ij} - \bar{Z}'_{i.})^2/\sum_{i=1}^{P}(n_i-1)}$$

เมื่อ $Z'_{ij} = |X_{ij} - X'_i|$; $i = 1, 2, \ldots ,P$ และ $j = 1, 2, \ldots ,n_i$

X'_i คือมัธยฐานของกลุ่มตัวอย่างที่ i

จะปฏิเสธสมมติฐานว่าง เมื่อค่า W' ที่คำนวณได้มากกว่าค่า F จากตารางที่องศาความเป็นอิสระ P-1 และ $\sum_{i}^{P}(n_i-1)$

W' เป็นตัวสถิติที่พัฒนามาจากตัวสถิติของ Levene ซึ่ง Levene ได้สร้างตัวสถิติทดสอบการเท่ากันของความแปรปรวนโดยวิธีการเปรียบเทียบความแปรปรวนของแต่ละกลุ่มแบบเดียวกับการวิเคราะห์ความแปรปรวนทางเดียว (One-way ANOVA)

ถ้า $X_{i1}, \ldots, X_{in_i}$ เป็นตัวอย่างสุ่มของค่าสังเกตสำหรับตัวแปรสุ่ม X_i ซึ่ง $X_{ij} \sim G(\mu_i, \sigma_i^2)$; $i = 1, 2, \ldots, P$ และ $j = 1, 2, \ldots, n_i$ ในการทดสอบการเท่ากันของความแปรปรวนของประชากร P กลุ่ม กำหนด $Z_{ij} = |x_{ij} - \bar{x}_i|$ ให้ U คือ ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองที่เกิดขึ้นระหว่าง P กลุ่ม (mean square error between P sample) และ V คือค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองที่เกิดขึ้นภายในกลุ่ม (mean square error with in samples) Levene สุ่มตัวอย่างโดยวิธีการมอนติคาร์โล เมื่อสมมติฐานคือ $H_o : \sigma_i = b$ และ $H_A : \sigma_i = a_i b$ โดยการกำหนดค่า a_i ต่าง ๆ กันจะได้

$$U(Z, H_A) = \frac{1}{P-1} \sum_{i=1}^{P} \left[a_i \bar{Z}_i - \left(\frac{1}{P} \sum_{i=1}^{P} a_i \bar{Z}_i\right)\right]^2$$

$$V(Z, H_A) = \frac{1}{P} \sum_{i=1}^{P} \sum_{j=1}^{n_i} \left[\frac{a_1^2 (Z_{ij} - \bar{Z}_{i.})^2}{n_i - 1}\right]$$

โดยที่ $F = \dfrac{\text{ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองที่เกิดขึ้นระหว่างกลุ่ม}}{\text{ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองที่เกิดขึ้นภายในกลุ่ม}}$

$$F = \frac{U}{V}$$

ดังนั้นตัวสถิติของ Levene คือ

$$W = \frac{\sum_{i=1}^{P} n_i (\bar{Z}_{i.} - \bar{Z}_{..})^2 / (P-1)}{\sum_{i=1}^{P} \sum_{j=1}^{n_i} (Z_{ij} - \bar{Z}_{i.})^2 / \sum_{i}^{P} (n_i - 1)}$$

เมื่อ $\bar{Z}_{i.} = \sum_{j=1}^{n_i} Z_{ij}/n_i$

$$\bar{Z}_{..} = \sum_{i=1}^{P} \sum_{j=1}^{n_i} Z_{ij}/\Sigma n_{i.}$$

ค่าวิกฤตของ W ใช้ตาราง F ที่องศาความเป็นอิสระ P-1 และ $\sum_{i=1}^{P}(n_i-1)$

Miller (1968: 567-582) ได้แสดงให้เห็นว่าควรประมาณความแปรปรวนของ $\bar{Z}_{i.} = \sum_{j=1}^{n_i} Z_{ij}/n_i$ ด้วย $\sum_{i=1}^{n_i}(Z_{ij} - \bar{Z}_{i.})^2/ n_i(n_i - 1)$ จึงจะถูกต้อง แต่เมื่อการแจกแจงของตัวแปรสุ่มเป็นแบบไม่สมมาตร (asymmetric) มัธยฐานไม่เท่ากับค่าเฉลี่ย ($\bar{X}$) เพราะจำนวนค่าสังเกตที่น้อยกว่า $\bar{X}$ และมากกว่า $\bar{X}$ มีจำนวนไม่เท่ากัน ตัวสถิติของ Levene จึงไม่เป็น asymptotically distribution free ซึ่ง Miller ได้เสนอแนะว่าถ้าเปลี่ยน $Z_{ij} = |x_{ij} - \bar{x}|$ เป็น $Z_{ij} = |x_{ij} - x_i'|$ เมื่อ x_i' คือ ค่ามัธยฐานจากตัวอย่าง (sample median) แล้วตัวสถิติของ Levene เป็น asymptotically distribution free

2.2.3 ตัวสถิติ C หรือ Cadwell's statistic

ตัวสถิติ C คำนวณจากอัตราส่วนระหว่างพิสัยที่มากที่สุดและพิสัยที่น้อยที่สุด

$$C = \frac{\max(r_1, r_2, \ldots, r_k)}{\min(r_1, r_2, \ldots, r_k)}$$

เมื่อ r_i คือพิสัย (range) ของกลุ่มตัวอย่างจากประชากรที่ i

$i = 1, 2, \ldots, k$

ค่าวิกฤตของตัวสถิติ C ใช้ตารางของ Harter (1963; cited by Leslie and Brown 1966: 221-227) จะปฏิเสธสมมติฐานว่าง เมื่อค่า C ที่คำนวณได้มากกว่าค่าจากตาราง

C เป็นตัวสถิติที่ Cadwell พัฒนามาจากตัวสถิติของ Hartley ซึ่ง Hartley's statistic = $\max(S_1^2, S_2^2, \ldots, S_k^2)/\min(S_1^2, S_2^2, \ldots, S_k^2)$ ใช้ทดสอบการเท่ากันของความแปรปรวนของประชากรที่มีการแจกแจงแบบปกติ k ประชากรซึ่งแต่ละกลุ่มตัวอย่างมีขนาดเท่ากัน Harter ได้ปรับปรุงตารางของตัวสถิติ C โดยเพิ่มขนาดของกลุ่มตัวอย่างมากขึ้น ตารางค่าวิกฤตของตัวสถิติ C แสดงเป็นค่าร้อยละ (percentage points) ของอัตราส่วน r_{max}/r_{min} ที่ระดับนัยสำคัญ 5% , 2.5% , 1% และ 0.5% ซึ่งคำนวณจากฟังก์ชันความหนาแน่น ของ r_{max}/r_{min}

ถ้า $C(k, n) = r_{max}/r_{min}$ ---------(1)

เมื่อ k คือ จำนวนประชากรที่นำมาทดสอบ

n คือ ขนาดของตัวอย่าง

จะสามารถหาฟังก์ชันความหนาแน่นของ C(k, n) ด้วยขั้นตอนดังนี้

(1) หาฟังก์ชันความหนาแน่นของพิสัย

(2) หาฟังก์ชันลอคการิทึมของฟังก์ชันในข้อ (1)

(3) หาฟังก์ชันความหนาแน่นของพิสัยในข้อ (2)

(4) หาแอนติลอคการิทึม (antilogarithm) ในข้อ (3)

เนื่องจาก พิสัย = ค่าสูงสุด - ค่าต่ำสุด ฟังก์ชันความหนาแน่นของพิสัยจึงหาได้จากการแจกแจงของตัวแปรสุ่มแบบ Order Statistics โดยฟังก์ชันความหนาแน่นร่วม (joint probability density function) ของ Y_i และ Y_j เมื่อ $Y_i < Y_j$; i, j = 1, 2,...,n เป็น Order Statistics ของค่าสังเกต (X)

ฟังก์ชันความหนาแน่นร่วมคือ

$$g_{i,j}(y_i, y_j) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!}\left[F(y_i)\right]^{i-1}\left[F(y_j)-F(y_i)\right]^{j-i-1}\left[1-F(y_j)\right]^{n-j} f(y_j)\, f(y_i$$

จะได้ $$g(y_1, y_n) = \frac{n!}{(n-2)!}\left[F(y_n)-F(y_1)\right]^{n-2} f(y_1)f(y_n)$$

หา p.d.f ของพิสัยโดยการเปลี่ยนตัวแปรให้ $X = y_n - y_1$, $t = y_1$ จะได้ $X = y_n - t$

$$g(X,t) = n(n-1)\left[F(X+t)-F(t)\right]^{n-2} f(t)f(t+X)$$

หา ฟังก์ชันความหนาแน่นชายขอบ (Marginal probability density function)

$$g(X) = \int_{-\infty}^{\infty} n(n-1)\left[F(X+t)-F(t)\right]^{n-2} f(t)f(t+X)\,dt \qquad \text{---------(2)}$$

ให้ $G_n(F,X)$ คือฟังก์ชันความน่าจะเป็นสะสม (Cumulative probability function) ของพิสัยของตัวอย่างขนาด n ที่มาจากประชากรที่มีฟังก์ชันความน่าจะเป็นสะสม คือ $F(X)$

ดังนั้น $$G_n(F,X) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} n(n-1)\left[F(X+t)-F(t)\right]^{n-2} f(t+X)f(t)\,dt\,dt$$

เนื่องจาก $$\int_{-\infty}^{\infty} n(n-1)\left[F(X+t)-F(t)\right]^{n-2} f(X+t)\,dt = n\left[F(X+t)-F(t)\right]^{n-1}$$

นั่นคือ $$G_n(F,X) = \int_{-\infty}^{\infty} n\left[F(X+t)-F(t)\right]^{n-1} f(t)\,dt$$

ที่ซึ่ง $$f(X) = \frac{d}{dX}F(X)$$

ดังนั้น $$G_n(F,X) = n\int_{-\infty}^{\infty}\left[F(X+t)-F(t)\right]^{n-1} dF(t) \qquad \text{--------(3)}$$

จาก (1) ฟังก์ชันลอการิทึมฐาน e (natural logarithm function) ของ C(k, n) คือ

$$\ell n C(k,n) = \ell n\left(\frac{r_{max}}{r_{min}}\right)$$

$$= \ell n\, r_{max} - \ell n\, r_{min}$$

$$= (\ell n\, r)_{max} - (\ell n\, r)_{min}$$

กำหนดให้ P(k,n;x) แทนฟังก์ชั่นความน่าจะเป็นสะสมของ $\ell n\ C(k,n)$ จะได้

$$\begin{aligned} P(k,\ n\ ;\ X) &= Pr(\ell n\ C(k,n) \leq X) \text{ หรือ } Pr\ (\ell n\ r_{max} - \ell n\ r_{min}) \leq X) \\ &= Pr(C(k,n) \leq e^X) \text{ หรือ } Pr\ (r_{max} - r_{min}) \leq e^X) \\ &= Pr\ (G_n(F\ ;\ e^X) < X) \\ &= G_k\ (G_n(F\ ;\ e^X)\ ;\ X) \qquad \text{--------(4)} \end{aligned}$$

จาก (2),(3)และ (4) จะได้ p.d.f ของ r_{max}/r_{min} คือ

$$P(k,n;X) = kn^k(n-1)\int_{-\infty}^{\infty} dL \left\{\int_{-\infty}^{\infty}\left[F(t+e^L)-F(t)\right]^{n-2} f(t+e^L)\, dF(t)\right\} \cdot$$

$$\left\{\int_{-\infty}^{\infty}\left[\left\{F(t+e^{L+X})-F(t)\right\}^{n-1} - \left\{F(t-e^L)-F(t)\right\}^{n-1}\right] dF(t)\right\}^{k-1}$$

ในปี ค.ศ. 1850 Link (1966 cited by Leslie and Brown:221-227) คำนวณ P (k,n;x) โดยกำหนด (k, n) = (2, 2) ได้

$$P(2,\ 2\ ;\ X) = \frac{4}{\Pi} \arctan(e^X) - 1$$

2.2 <u>ตัวสถิติอื่น</u> ๆ สำหรับการทดสอบการเท่ากันของความแปรปรวน k กลุ่ม

2.2.1 <u>Jackknife Test (J)</u> เป็นตัวสถิติทดสอบที่ Miller (1964: 1594-1605) สร้างขึ้น

$$J = \frac{\sum_{i=1}^{k} n_i(\bar{U}_{i.} - \bar{U}_{..})^2/(k-1)}{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(U_{ij} - \bar{U}_{i.})^2 / \sum_i (n_i - 1)}$$

เมื่อ $U_{ij} = n_i \ln S_i^2-(n_i-1)\ln S_{i(j)}^2$; (i=1 ,...,k ; j=i....n;

$$S_{i(j)}^2 = \frac{1}{(n_i-2)}\sum_{k\neq j}(X_{ik}-\bar{X}_{i(j)})^2$$

$$\bar{X}_{i(j)} = \frac{1}{(n_i-1)}\sum_{k\neq j} X_{ik}$$

ค่าวิกฤตของ J ใช้ตาราง F ที่องศาความเป็นอิสระ k-1 และ $\sum_i (n_i-1)$

2.2.2 <u>Box Test</u> (Layard 1973: 195-198) เป็นตัวสถิติที่ Box สร้างขึ้นโดยการแบ่งกลุ่มตัวอย่างที่ i ออกเป็นกลุ่มย่อย (subsample) J_i กลุ่ม

$$G = \frac{\sum_{i=1}^{k} J_i(\bar{Y}_{i.}-\bar{Y}_{..})^2/(k-1)}{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij}-\bar{Y}_{i.})^2/\sum_{i=1}^{k}(J_i-1)}$$

เมื่อ $Y_{ij} = \ln S_{ij}^2$

S_{ij}^2 คือความแปรปรวนของตัวอย่างย่อย (subsample variance) ที่ j ของกลุ่มตัวอย่างที่ i

ซึ่ง $G \sim \chi_{k-1}^2/k-1$

2.2.3 <u>χ^2 test</u> เป็นตัวสถิติที่เสนอโดย Layard (1973: 195-198) คือเมื่อ n_i มีขนาดใหญ่พอ $(n_i-1)^{\frac{1}{2}}\ln S_i^2 \sim N(\ln\sigma_i^2, \tau_i^2)$ ซึ่ง $\tau_i^2 = 2+\{1-(1/n_i)\}r$ เมื่อ r คือความโด่ง และประมาณความโด่งด้วยวิธีรวมความโด่ง (pool kurtosis) ของทุกกลุ่มตัวอย่าง ตัวสถิติคือ

$$S = \tau^2 S'/\hat{\tau}^2$$

เมื่อ
$$S' = \sum_{i=1}^{k}(n_i-1)\left[\ln S_i^2-\frac{\sum_{i=1}^{k}(n_i-1)\ln S_i^2}{\sum_{i=1}^{k}(n_i-1)}\right]^2/\tau^2$$

$$\hat{\tau}^2 = 2 + [1-(1/\bar{n})]\hat{r}$$

$$\hat{r} = \frac{\left(\sum_{i=1}^{k} n_i\right)\sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij}-\bar{X}_i)^4}{\left\{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij}-\bar{X}_1)^2\right\}^2} - 3$$

X_{ij} คือค่าสังเกตที่ j ของกลุ่มตัวอย่างที่ i

$j = 1, 2, \ldots, n_i$ และ $j = 1, 2, \ldots, k$

ซึ่ง $S_i \sim \chi^2_{k-1}$

2.2.4 Hartley's statistic (Hartley 1950: 308-312, cited by Gartside 1972: 343) เมื่อขนาดของกลุ่มตัวอย่างเท่ากัน

$$H = \frac{\max(S_1^2, S_2^2, \ldots, S_k^2)}{\min(S_1^2, S_2^2, \ldots, S_k^2)}$$

เมื่อ S_i^2 คือความแปรปรวนของกลุ่มตัวอย่างจากประชากรที่ i

ค่าวิกฤตของ H ได้จากตารางที่ปรับปรุงโดย David (1952 cited by Leslie and Brown 1966: 221-227)

2.2.5 Lehmann's Test (Ghosh 1972: 221-233) เมื่อขนาดของกลุ่มตัวอย่างเท่ากัน

$$L = \sum_{j=1}^{k} \nu_j (Z_j-\bar{Z})^2$$

เมื่อ $Z_i = \ln(S_j/\nu_j)$

$$\bar{Z} = \sum_{j=1}^{k} \nu_j Z_j/N$$

$$N = \sum_{j}^{k} \nu_j$$

$$S_j = \sum_{i=1}^{\nu_j+1} (X_{ji}-\bar{X}_j)^2$$

$$\bar{X}_j = \sum_{i=1}^{\nu_j+1} X_{ji}/(\nu_j+1)$$

ν_j คือ ขนาดของกลุ่มตัวอย่างที่ j

การทดลอบจะยอมรับสมมติฐานว่าง (H_o) ที่ระดับนัยสำคัญ α เมื่อ

$$L\left\{(X_{ji})\right\} \leq \ell_{1-\alpha} \quad \text{เมื่อ } \ell_{1-\alpha} \sim \frac{2\nu}{\nu-1}\chi^2_{k-1,\ 1-\alpha}$$

2.2.6 <u>Cochran's statistic</u> (Cochran 1941: 47, cited by Gartside 1972 : 343) เมื่อขนาดของกลุ่มตัวอย่างเท่ากัน

$$Co = \frac{\max(S_1^2,\ S_2^2,\ldots,S_k^2)}{\sum_{i}^{k} S_i^2}$$

เมื่อ S_i^2 คือความแปรปรวนของกลุ่มตัวอย่างจากประชากรที่ i

ตารางค่าวิกฤตของ Cochran's statistic อยู่ในหนังสือ Introduction to Statistical Analysis, 3rd ed. 1969 เขียนโดย Dixon, W.J. และ Massey, F.J.,Jr.

2.3 <u>ผลงานวิจัยที่เกี่ยวข้อง</u>

การศึกษาเกี่ยวกับการทดลอบการเท่ากันของความแปรปรวนเพื่อเปรียบเทียบอำนาจการทดลอบของตัวสถิติเมื่อเกิดการฝ่าฝืนข้อตกลงเบื้องต้นของการใช้ตัวสถิติทดลอบมีดังต่อไปนี้

Pearson (1966: 229-233) ทดสอบการเท่ากันของความแปรปรวนของ 5 ประชากรโดยสุ่มตัวอย่างขนาด 5 จากประชากรที่มีการแจกแจงแบบปกติด้วยวิธีการมอนติคาร์โล ซิมูเลชั่น ซึ่งตั้งสมมติฐานแย้ง (H_A) สำหรับอัตราส่วนความแปรปรวน 8 แบบคือ $1:1:1:\sqrt{2}:\sqrt{2}$, $1:1:1:2:2$, $1:1:1:2\sqrt{2}:2\sqrt{2}$, $1:1:1:4:4$, $1/\sqrt{2}:1:1:1:\sqrt{2}$, $1/\sqrt{2}:1:1:1:2$, $1/2\sqrt{2}:1:1:1:2\sqrt{2}$ และ $1/4:1:1:1:4$ ใช้ตัวสถิติทดสอบ 3 ประเภทคือสถิติทดสอบ M(M-test) หรือการทดสอบนีแมน-เปียร์สัน (Neyman & Pearson Test) สถิติทดสอบของฮาร์ตเลย์ และสถิติทดสอบของแคดเวล Pearson ได้สรุปผลและให้ข้อเสนอแนะไว้ดังนี้คือ

1. อำนาจการทดสอบของตัวสถิติทดสอบฮาร์ตเลย์และสถิติทดสอบแคดเวล ไม่แตกต่างกัน

2. สถิติทดสอบ M มีอำนาจการทดสอบสูงสุด แต่อย่างไรก็ตามถ้าความแปรปรวนของประชากรมีความแตกต่างกันมาก ๆ ควรจะใช้ตัวสถิติทดสอบของฮาร์ตเลย์ หรือตัวสถิติทดสอบของแคดเวล

3. ถ้าผู้วิจัยต้องการทดสอบการเท่ากันของความแปรปรวนโดยใช้กลุ่มตัวอย่างที่มีขนาดเล็กและเท่ากันโดยต้องการผลลัพธ์อย่างเร็วแล้วควรใช้สถิติทดสอบแคดเวล

Gartside (1972: 342-346) ศึกษาวิธีการต่าง ๆ ในการทดสอบเพื่อเปรียบเทียบความแปรปรวนของประชากรที่มีการแจกแจงแบบปกติมาตรฐานโดยใช้ตัวสถิติทดสอบ 8 ประเภท ได้แก่ สถิติทดสอบของบาร์ตเลต (Bartlett's test) สถิติทดสอบของบาร์ตเลตแบบปรับปรุง (The modified Bartlett's test) สถิติทดสอบของ Cochran (Cochran's statistic) สถิติทดสอบของฮาร์ตเลย์ สถิติทดสอบของแคดเวล และสถิติทดสอบแบบลอคอโนวา (Log ANOVA) 3 แบบด้วยวิธีมอนติคาร์โล ซิมูเลชั่น จากผลการวิจัยสรุปได้คือ

1. สถิติทดสอบของบาร์ตเลต มีอำนาจการทดสอบสูงที่สุดเมื่อ

1.1 ใช้กลุ่มตัวอย่างขนาด 16 เท่ากันและตั้งสมมติฐานคือ

H_A : $\sigma_1^2 = 1$, $\sigma_2^2 = c$, ... , $\sigma_k^2 = c^{k-1}$ เมื่อ c คือค่าคงที่ที่กำหนดขึ้น

1.2 ใช้กลุ่มตัวอย่างขนาดไม่เท่ากัน โดยที่ H_A : $\sigma_1^2 = \sigma_2^2 = \ldots\sigma_{k-1}^2 = 1$, $\sigma_k^2 = c$

2. สถิติทดสอบของ Cochran มีอำนาจการทดสอบสูงที่สุดเมื่อขนาดของกลุ่มตัวอย่างเท่ากันคือ 16 กรณีที่ H_A : $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_{k-1}^2 = 1,\ \sigma_k^2 = c$

โดยทั่วไปแล้วจากผลงานวิจัยนี้สรุปได้ว่า ตัวสถิติอื่น ๆ เช่น Bartlett's modified สถิติทดสอบของอาร์ตเลย์ และสถิติทดสอบของแคดเวล ยังคงมีอำนาจการทดสอบอยู่ในระดับ ยกเว้น Log ANOVA ที่มีอำนาจการทดสอบในระดับต่ำ นอกจากนี้ Gartside ยังได้เสนอแนะด้วยว่า เมื่อต้องการทดสอบการเท่ากันของความแปรปรวนโดยที่ประชากรมีการแจกแจงแบบปกติ ควรใช้สถิติทดสอบของบาร์ตเลต และถ้าสงสัยว่าจะมีเพียงกลุ่มเดียวเท่านั้นที่มีความแปรปรวนสูงกว่ากลุ่มอื่น ๆ ควรใช้สถิติทดสอบของ Cochran แต่ถ้าต้องการวิธีการทดสอบที่ทำให้ได้ผลสัพธ์โดยเร็วแล้วควรใช้สถิติทดสอบของอาร์ตเลย์หรือสถิติทดสอบของแคดเวล

สำหรับงานวิจัยที่ศึกษาเกี่ยวกับอำนาจการทดสอบของตัวสถิติที่ใช้ทดสอบการเท่ากันของความแปรปรวนนั้น Games (1972: 887-909) ศึกษาโดยใช้กลุ่มตัวอย่างจาก 3 ประชากรและใช้ตัวสถิติทดสอบคือ สถิติทดสอบของบาร์ตเลต สถิติทดสอบของอาร์ตเลย์ (หรือ F_{max} test) สถิติทดสอบของCochran สถิติทดสอบ L-X^2 และL-A เมื่อใช้กลุ่มตัวอย่างขนาด 6 แต่ในกรณีที่เพิ่มกลุ่มตัวอย่างเป็น 18 ใช้ตัวสถิติของบาร์ตเลตและเคนดอล (Bartlett and Kendall's statistic) สถิติทดสอบ Q (Q-test) สถิติทดสอบของบาร์ตเลต สถิติทดสอบ LEV3 และสถิติทดสอบ LEV2 โดยเปรียบเทียบเมื่อประชากรมีลักษณะการแจกแจง 6 แบบคือ แบบปกติ แบบมีความเบ้ 3 ระดับคือ เบ้น้อย (slight skew) เบ้ปานกลาง (moderate skew) และเบ้มาก (extreme skew) แบบสมมาตรที่มีความโด่งแบบเลปโตเคอร์ติคและแบบยูนิฟอร์ม ผลการวิจัยสรุปได้คือ

1. ในกรณีที่ศึกษากลุ่มตัวอย่างที่มีขนาด 6 ถ้าประชากรมีการแจกแจงแบบปกติและความแปรปรวนของประชากรไม่เท่ากันทั้งหมด สถิติทดสอบของบาร์ตเลตและสถิติทดสอบ F_{max} มีอำนาจการทดสอบสูงที่สุดและถ้าประชากรมีความเบ้เพียงเล็กน้อยก็จะไม่มีผลกระทบต่ออำนาจการทดสอบของตัวสถิติ

2. ในกรณีที่ศึกษากลุ่มตัวอย่างขนาด 18 สถิติทดสอบของบาร์ตเลตและเคนดอล สถิติทดสอบ M และสถิติทดสอบ Q มีอำนาจการทดสอบสูงและไม่แตกต่างกันมาก แต่ถ้าประชากรมีการแจกแจงแบบเบ้ สถิติทดสอบ LEV 3 ซึ่งใช้วิธีเปรียบเทียบความแปรปรวนโดยการแบ่งกลุ่ม

ตัวอย่างเป็นกลุ่มย่อย (subsample) จะสามารถควบคุมความคลาดเคลื่อนชนิดที่ 1 ในการทดสอบได้ดีแต่มีอำนาจการทดสอบต่ำกว่าตัวสถิติอื่น

Game ยังได้เสนอแนะไว้ด้วยว่า ถ้าประชากรมีลักษณะการแจกแจงที่มีความโด่ง (r_2) น้อยกว่า 0.5 ถ้า $r_2 = E\left[\{(x_i - \mu_x)^4/\sigma^4\}-3\right]$ ควรทดสอบการเท่ากันของความแปรปรวนโดยใช้สถิติทดสอบของบาร์ตเลต หรือสถิติทดสอบ F_{max} แต่ถ้าผู้วิจัยไม่ทราบแน่ชัดเกี่ยวกับรูปแบบการแจกแจงของประชากรเป็นที่แน่นอนแล้ว ควรใช้สถิติทดสอบ LEV 3 เพราะถึงแม้ว่าจะเป็นวิธีการทดสอบที่มีอำนาจการทดสอบต่ำกว่าตัวสถิติทดสอบอื่น ๆ ก็อาจจะแก้ไขได้โดยการเพิ่มขนาดของกลุ่มตัวอย่าง แต่ถ้าไม่สามารถเพิ่มขนาดของกลุ่มตัวอย่างได้ ผู้วิจัยอาจเลือกใช้ตัวสถิติทดสอบบาร์ตเลตและเคนดอล เพราะถึงแม้ว่าจะมีอำนาจการทดสอบต่ำกว่าตัวสถิติทดสอบของบาร์ตเลต แต่ก็มีความเชื่อถือได้เมื่อมีการฝ่าฝืนข้อตกลงเบื้องต้นของการทดสอบมากกว่า เมื่อประชากรมีการแจกแจงแบบไม่ปกติ และ Game ยังกล่าวอีกว่า นักวิจัยไม่ควรละเลยการทดสอบการเท่ากันของความแปรปรวนก่อนการวิเคราะห์ความแปรปรวน (ANOVA) เพราะถ้าผู้วิจัยต้องการทดสอบการเปรียบเทียบพหุคูณ (Multiple Comparision test) ที่มีข้อตกลงเบื้องต้นเกี่ยวกับการเท่ากันของความแปรปรวน อาจทำให้ผลสรุปผิดพลาดจากที่ควรเป็นจริงได้

จากการศึกษาความแกร่งของตัวสถิติทดสอบการเท่ากันของความแปรปรวน Layard (1973:195-198) ใช้กลุ่มตัวอย่างจาก 4 ประชากร กรณีที่กลุ่มตัวอย่างมีขนาดเท่ากันทั้ง 4 กลุ่ม คือขนาด 10 และ 25 โดยศึกษาตัวสถิติ 4 ประเภทได้แก่ สถิติทดสอบบาร์ตเลต สถิติทดสอบ χ^2 (χ^2 test) สถิติทดสอบ Box (Box test) และสถิติทดสอบแบบแจคไนฟ์ (Jackknife test) ด้วยวิธีการทดลองแบบมอนติคาร์โล กำหนดรูปแบบการแจกแจงของประชากร 3 รูปแบบ คือ แบบยูนิฟอร์ม แบบปกติ และแบบดับเบิลเอกซ์โปเนนเชียล (Double Exponential distribution) เมื่ออัตราส่วนของความแปรปรวนกำหนดเป็น 1:1:1:1 , 1:1:2:2 , 1:2:3:4 และ 1:1:4:4 สรุปผลได้ว่า สถิติทดสอบ Box เป็นตัวสถิติทดสอบที่มีความแกร่ง แต่สถิติทดสอบบาร์ตเลต มีอำนาจการทดสอบสูงสุดเมื่อประชากรมีการแจกแจงแบบปกติ และเมื่อประชากรมีการแจกแจงแบบไม่ปกติตัวสถิติแต่ละประเภทจะมีอำนาจการทดสอบไม่แตกต่างกันมาก Layard ได้เสนอแนะไว้ว่า ควรเลือกใช้ตัวสถิติทดสอบ χ^2 หรือสถิติทดสอบแจคไนฟ์ ถ้าต้องการตัวสถิติทดสอบที่มีความแกร่งและมีอำนาจการทดสอบอยู่ในระดับที่สามารถยอมรับได้ เพราะถึงแม้ว่าสถิติทดสอบ Box จะมีความเชื่อถือได้เมื่อมีการฝ่าฝืนข้อตกลงเบื้องต้น แต่ก็มีอำนาจการทดสอบต่ำกว่าสถิติทดสอบทั้ง 2 ดังกล่าว