บทที่ 2
ตัวสถิติและผลงานวิจัยที่เกี่ยวข้อง

ในหัวข้อนี้จะกล่าวถึงรายละเอียดเกี่ยวกับวิธีการบูทสแตรป และจะขอแนะนำเกี่ยวกับการแจกแจงที่เกี่ยวข้อง ก่อนที่จะกล่าวถึงการทดสอบดังกล่าวทั้ง 3 ชนิด ซึ่งได้แก่ การทดสอบของไวท์ (White's test) การทดสอบของบรูชและพาแกน (Breusch-Pagan's test) และการทดสอบของสโรเตอร์ (Szroeter's test) รวมทั้งงานวิจัยที่เกี่ยวข้อง ดังต่อไปนี้

2.1 <u>วิธีการบูทสแตรป</u> (Bootstrap Method)

วิธีบูทสแตรป (Bootstrap Method) เป็นวิธีทางนอนพาราเมตริก (Nonparametric) เป็นวิธีการสุ่มตัวอย่างซ้ำแบบใส่คืน (Resampling with Replacement) ซึ่งเสนอโดยแบรดเลย์ เอฟรอน (Bradley Efron : 1979) วิธีบูทสแตรปนี้เป็นวิธีที่มักจะนำมาแก้ปัญหาการที่ไม่สามารถหาค่าประมาณได้ ในกรณีที่ข้อมูลไม่เป็นไปตามข้อตกลงเบื้องต้น (Assumption) แต่ในที่นี้จะใช้วิธีบูทสแตรป ในการปรับปรุงการตรวจสอบปัญหาความแปรปรวนของความคลาดเคลื่อนมีค่าไม่คงที่ แนวคิดที่สำคัญของวิธีบูทสแตรป คือ ตัวอย่างเป็นสิ่งที่เราทราบทั้งหมดเกี่ยวกับประชากร และตัวอย่างแต่ละตัวอย่างจะสามารถอธิบายลักษณะของประชากรด้วยความน่าจะเป็นที่เท่าๆ กัน ซึ่งแนวคิดนี้อาจจะทำให้ได้ข้อสรุปที่ดีเกี่ยวกับลักษณะของประชากร โดยที่อาจจะดีมากกว่าการกำหนดข้อตกลงเบื้องต้นเกี่ยวกับประชากร ซึ่งวิธีบูทสแตรปมีหลักเกณฑ์ดังนี้ คือ เราจะให้ตัวอย่างที่ถูกเก็บรวบรวมมาจากประชากรเปรียบเสมือนประชากร แล้วทำการสุ่มตัวอย่างจากตัวอย่างที่มีอยู่แบบใส่คืน(Resampling with Replacement) ด้วยจำนวนครั้ง (Bootstrap Replications) ที่มากเพียงพอ (ซึ่งควรจะอยู่ในช่วง 50 – 200 ก็เพียงพอที่จะทำให้ได้ตัวประมาณที่ดี) เพื่อสร้างการแจกแจงของตัวสถิติตัวอย่าง (Sampling Distribution) และนำไปใช้ประมาณค่าพารามิเตอร์ที่สนใจ ในหัวข้อนี้จะกล่าวถึงรายละเอียดที่ควรทราบ ซึ่งได้แก่ การสุ่มตัวอย่างแบบใส่คืน ดังนี้

<u>การสุ่มตัวอย่างแบบใส่คืน</u> (Sampling with replacement)

การสุ่มตัวอย่างแบบใส่คืน เป็นการสุ่มตัวอย่างที่ยอมให้มีหน่วยตัวอย่างซ้ำกันได้ นั่นก็คือแต่ละหน่วยตัวอย่างมีโอกาส (Probability)ในการถูกสุ่มเท่ากัน เช่น ถ้ากำหนดให้ขนาดของประชากรหรือจำนวนประชากรเท่ากับ N แล้วความน่าจะเป็นหรือโอกาสที่แต่ละหน่วยจะถูกสุ่มจะเท่ากับ $\frac{1}{N}$

2.2 ตัวสถิติที่ใช้ในการทดสอบ

2.2.1 ตัวสถิติทดสอบของไวท์ (White's Statistics)

วิธีการนี้เป็นวิธีการหนึ่งที่พบบ่อยที่สุด ในการใช้ตรวจสอบว่า ความคลาดเคลื่อนมีความแปรปรวนคงที่หรือไม่ นั่นคือ

H_0 : ความคลาดเคลื่อนมีความแปรปรวนคงที่

หรือ $Var(\varepsilon_i) = \sigma^2$ for all i , $i = 1,2,...,n$

H_1 : ความคลาดเคลื่อนมีความแปรปรวนที่ไม่คงที่ หรือ heteroscedasticity

หรือ $Var(\varepsilon_i) = \sigma_i^2 \neq \sigma^2$ for some i

โดยใช้ความสัมพันธ์ระหว่างกำลังสองของเศษตกค้าง (residual) กับตัวแปรอิสระ (Independent Variable) ซึ่งสำหรับในกรณีที่เป็นตัวแบบการถดถอยเชิงเส้นอย่างง่าย

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_p X_{pi} + \varepsilon_i \quad , \ i = 1,2,...,n \tag{1}$$

เราสามารถแสดงความสัมพันธ์ได้ดังนี้

$$(\hat{\varepsilon}_i)^2 = \alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{1i}^2 + \ldots + \alpha_{2p+1} X_{1i} X_{2i} + \ldots + \nu_i \tag{2}$$

เมื่อ $\hat{\varepsilon}_i$ คือ เศษตกค้างของค่าสังเกตที่ i

X_{ji} คือ ตัวแปรต้นหรือตัวแปรอิสระตัวที่ j ของค่าสังเกตที่ i

$\alpha_0, \alpha_1, \alpha_2$ เป็นพารามิเตอร์ที่ไม่ทราบค่า

คุณสมบัติของตัวสถิติของไวท์เมื่อตัวอย่างมีขนาดจำกัด

สำหรับตัวแบบการถดถอยเชิงพหุ (Multiple regression model)ในรูปทั่วไป คือ

$$\tilde{Y} = X\tilde{\beta} + \tilde{\varepsilon}$$

เมื่อ $\tilde{Y}$ คือ เมตริกซ์ตัวแปรตามขนาด $n \times 1$

X คือ เมตริกซ์ตัวแปรต้นหรือตัวแปรอิสระขนาด $n \times (p+1)$

$\tilde{\beta}$ คือ เมตริกซ์พารามิเตอร์ที่ไม่ทราบค่า

$\tilde{\varepsilon}$ คือ เมตริกซ์ความคลาดเคลื่อน (Error Term) ที่มีค่าเฉลี่ยเป็นเวกเตอร์ศูนย์และความแปรปรวนเท่ากับ $\sigma^2 I$

n คือ ขนาดตัวอย่างที่ได้รับการสุ่มมาจากประชากร

p คือ จำนวนตัวแปรอิสระ

นิยามของ White's Statistics คือ

$$\boxed{W \equiv \left(\frac{1}{n} \sum_{i=1}^{n} (\hat{\varepsilon}_i^2 - \hat{\sigma}^2) \psi_i{}' \right) \hat{D}_n^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} (\hat{\varepsilon}_i^2 - \hat{\sigma}^2) \psi_i \right)} \tag{3}$$

เมื่อ ψ_i เป็นเวกเตอร์ที่แสดงสมาชิกของเมตริกซ์ $X \otimes X$ ที่ไม่รวมค่าคงที่ และมีค่าเฉลี่ยตัวอย่างเท่ากับ $\bar{\psi}$

และ $\hat{D}_i \equiv \frac{1}{n}\sum_{i=1}^{n}\left(e_i^2 - \hat{\sigma}^2\right)^2 \left(\psi_i - \bar{\psi}\right)\left(\psi_i - \bar{\psi}\right)'$

หมายเหตุ กรณีตัวแบบการถดถอยประกอบด้วยตัวแปรอิสระ 1 ตัวและมีค่าสังเกต 2 ค่า

$$X \otimes X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \otimes \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ x_2 \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \end{bmatrix} = \begin{bmatrix} x_1 x_1 \\ x_1 x_2 \\ x_2 x_1 \\ x_2 x_2 \end{bmatrix}$$

White (1980) ได้แสดงว่า W มีการแจกแจงแบบไคสแควร์โดยประมาณ (asymptotically distributed as Chi-square) ที่องศาความเป็นอิสระ (degree of freedom) $\frac{k(k+1)}{2} - 1$ แต่เนื่องจากสมการที่ (3) ค่อนข้างยากในการคำนวณ White จึงแนะนำตัวสถิติที่สมมูลกัน นั่นคือ

$$W \equiv n \cdot R^2$$

เมื่อ R^2 เป็นค่าที่ได้มาจากสมการช่วย (Auxillary equation) ซึ่งในกรณีที่เป็นการถดถอยเชิงเส้นอย่างง่าย ค่า R^2 จะได้มาจากสมการที่ (2) ดังนั้น W จะมีการแจกแจงแบบไคสแควร์ที่องศาความเป็นอิสระเท่ากับ 2

ขั้นตอนการหาค่าของตัวสถิติของไวท์

1) จากตัวแบบการถดถอยเชิงเส้นอย่างง่าย

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_p X_{pi} + \varepsilon_i \quad , \; i = 1,2,\ldots,n$$

ทำการหาค่าเศษตกค้าง ($\hat{\varepsilon}_i$) โดยใช้วิธี OLS

2) หาค่า R^2 จากความสัมพันธ์ระหว่างกำลังสองของเศษตกค้างกับตัวแปรอิสระในสมการที่ (2) นั่นคือ

$$(\hat{\varepsilon}_i)^2 = \alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{1i}^2 + \ldots + \alpha_{2p+1} X_{1i} X_{2i} + \ldots + \nu_i$$

3) คำนวณค่า $W = n \cdot R^2$

4) เปรียบเทียบค่า W กับ $\chi^2_{(2)}$

5) ถ้าค่า $W > \chi^2_{(2)}$ เราจะปฏิเสธสมมติฐานว่าง นั่นคือ ความคลาดเคลื่อนมีความแปรปรวนที่ไม่คงที่

ตัวอย่างวิธีการหาสถิติทดสอบของไวท์

ตารางที่ 2.1 แสดงตัวอย่างข้อมูลที่สร้างขึ้นจากการทดลอง

X_i	Y_i	$\hat{\varepsilon}_i$	$\hat{\varepsilon}_i^2$
2.4578	1.1874	-1.1787	1.3894
0.6493	1.6339	1.1611	1.3482
4.1649	4.1201	-0.0332	0.0011
3.6268	2.8279	-0.7620	0.5807
3.0750	2.3099	-0.7024	0.4933
3.3516	4.2133	0.9116	0.8309
2.3034	2.2473	0.0427	0.0018
4.6961	5.2096	0.5003	0.2503
3.0590	3.4557	0.4602	0.2118
4.7970	5.5533	0.7383	0.5450
3.2640	3.6646	0.4544	0.2065
3.8716	2.5303	-1.3159	1.7317
2.2988	1.9289	0.5092	0.2592
4.2459	3.4240	1.2244	1.4990
2.3610	4.9746	0.7365	0.5425
3.5773	-0.0164	-2.2812	5.2038
2.6399	3.3036	-0.2345	0.0550
1.5538	2.3170	-0.2397	0.0575
2.8644	3.1824	0.3907	0.1526
1.6506	1.1395	-0.3816	0.1456

เนื่องจากจำนวนตัวแปรอิสระเท่ากับ 1 ตัว

ดังนั้น จากสมการ $(\hat{\varepsilon}_i)^2 = \alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{1i}^2 + \nu_i$ หรือ $\hat{\tilde{\varepsilon}}^2 = X_{white}\tilde{\alpha} + \tilde{\nu}$

จะได้ค่า

$$\hat{\alpha} = \begin{bmatrix} \hat{\alpha}_0 \\ \hat{\alpha}_1 \\ \hat{\alpha}_2 \end{bmatrix} = (X'_{white} X_{white})^{-1} X'_{white} \hat{\tilde{\varepsilon}}^2$$

$$= \left(\begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_{11} & x_{12} & \cdots & x_{1n} \\ x_{11}^2 & x_{12}^2 & \cdots & x_{1n}^2 \end{bmatrix} \begin{bmatrix} 1 & x_{11} & x_{11}^2 \\ 1 & x_{12} & x_{12}^2 \\ \vdots & \vdots & \vdots \\ 1 & x_{1n} & x_{1n}^2 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_{11} & x_{12} & \cdots & x_{1n} \\ x_{11}^2 & x_{12}^2 & \cdots & x_{1n}^2 \end{bmatrix} \begin{bmatrix} \varepsilon_1^2 \\ \varepsilon_2^2 \\ \vdots \\ \varepsilon_n^2 \end{bmatrix}$$

$$= \left(\begin{bmatrix} 1 & 1 & \cdots & 1 \\ 2.4578 & 0.6493 & \cdots & 1.6506 \\ 6.0408 & 0.4216 & \cdots & 2.7245 \end{bmatrix} \begin{bmatrix} 1 & 2.4578 & 6.0408 \\ 1 & 0.6493 & 0.4216 \\ \vdots & \vdots & \vdots \\ 1 & 1.6506 & 2.7245 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 2.4578 & 0.6493 & \cdots & 1.6506 \\ 6.0408 & 0.4216 & \cdots & 2.7245 \end{bmatrix} \begin{bmatrix} 1.3894 \\ 1.3482 \\ \vdots \\ 0.1456 \end{bmatrix}$$

$$= \begin{bmatrix} 1.8599 & -1.2504 & 0.1925 \\ -1.2504 & 0.9342 & -0.1537 \\ 0.1925 & -0.1537 & 0.0266 \end{bmatrix} \begin{bmatrix} 15.5059 \\ 50.2918 \\ 179.008 \end{bmatrix}$$

$$= \begin{bmatrix} 0.4097 \\ 0.0740 \\ 0.0138 \end{bmatrix}$$

$$SSR = \sum_{i=1}^{n} (\hat{\alpha}_0 + \hat{\alpha}_1 X_{1i} + \hat{\alpha}_2 X_{1i}^2)\hat{\varepsilon}_i^2 - \frac{\left(\sum_{i=1}^{n} \hat{\varepsilon}_i^2\right)^2}{n} = 12.5492 - 12.0217 = 0.5275$$

$$SST = \sum_{i=1}^{n} (\hat{\varepsilon}_i^2)^2 - \frac{\left(\sum_{i=1}^{n} \hat{\varepsilon}_i^2\right)^2}{n} = 38.2038 - 12.0217 = 26.1821$$

$$R^2 = \frac{SSR}{SST} = \frac{0.5275}{26.1821} = 0.0201$$

ค่าสถิติ $W = n \cdot R^2 = 20 \times 0.0201 = 0.402 < \chi^2_{(1),0.05} = 3.846$

ดังนั้น จึงยอมรับสมมติฐานว่าง นั่นคือ ความแปรปรวนของความคลาดเคลื่อนคงที่ ที่ระดับนัยสำคัญ 0.05

2.2.1 <u>ตัวสถิติทดสอบของบรูชและพาแกน</u> (Breusch-Pagan 's Statistics)

การทดสอบของบรูชและพาแกน เป็นการทดสอบปัญหาความแปรปรวนของความคลาดเคลื่อนไม่คงที่ที่พบบ่อยมากชนิดหนึ่ง เป็นการทดสอบว่าความแปรปรวนของความคลาดเคลื่อนที่ประมาณได้จากตัวแบบการถดถอย ขึ้นอยู่กับค่าของตัวแปรอิสระหรือไม่ โดยมีพื้นฐานจากการทดสอบ LM (Lagrangian Multiplier test) ของ Silvey (1959) และได้ถูกพัฒนาขึ้นโดย Godfrey (1978) และ Breusch และ Pagan (1979) ซึ่งใช้ในการตรวจสอบปัญหาความแปรปรวนของความคลาดเคลื่อนไม่คงที่ในรูปแบบเชิงเส้น ดังนี้

<u>รูปแบบที่ 1</u> $\sigma_i^2 = \alpha_1 + \alpha_2 z_{2i} + ... + \alpha_s z_{si}$; $i = 1,2,...,n$

หรืออาจเขียนได้เป็น $\sigma_i^2 = \tilde{z}_i'\tilde{\alpha}$

<u>รูปแบบที่ 2</u> $\sigma_i^2 = (\alpha_1 + \alpha_2 z_{2i} + ... + \alpha_s z_{si})^2$; $i = 1,2,...,n$

หรืออาจเขียนได้เป็น $\sigma_i^2 = (\tilde{z}_i'\tilde{\alpha})^2$

<u>รูปแบบที่ 3</u> $\sigma_i^2 = \exp(\alpha_1 + \alpha_2 z_{2i} + ... + \alpha_s z_{si})$; $i = 1,2,...,n$

หรืออาจเขียนได้เป็น $\sigma_i^2 = \exp(\tilde{z}_i'\tilde{\alpha})$

โดยที่ $\tilde{z}_i = (1, z_{2i}, z_{3i}, ..., z_{si})'$

z_{ji} อาจจะเป็นตัวแปรอิสระ x_{ji} หรือฟังก์ชั่นของตัวแปรอิสระตัวใดตัวหนึ่งเพียงตัวเดียวเท่านั้น ; $j = 2,3,...,s$

$\tilde{\alpha}$ เป็นเวกเตอร์ของพารามิเตอร์ที่ไม่ทราบค่า ขนาด $s \times 1$ นั่นคือ

$$\tilde{\alpha} = (\alpha_1, \alpha_2, ..., \alpha_s)'$$

สมมติฐานที่ต้องการทดสอบ คือ

$H_0 : \alpha_2 = ... = \alpha_s = 0$

H_1 : มี α_j อย่างน้อย 1 ค่าที่ไม่เท่ากับ 0 ; $j = 2,3,...,s$

จะเห็นว่าภายใต้สมมติฐานว่าง จะได้ว่า $\tilde{z}_i'\tilde{\alpha} = \alpha_1$ ซึ่งเป็นค่าคงที่ นั่นคือ ถ้ายอมรับสมมติฐานว่างจะได้ว่า ความคลาดเคลื่อนมีความแปรปรวนคงที่นั่นเอง

พิจารณาความแปรปรวนของความคลาดเคลื่อนในรูปแบบต่างๆ ที่จะใช้ในการทำงานวิจัยชิ้นนี้ ซึ่งได้แก่

1) <u>รูปแบบการคูณ</u> (Multiplicative model)

$$V(\varepsilon_i) = E(\varepsilon_i^2) = \sigma^2 X_i^r$$

ซึ่งเป็นรูปแบบเฉพาะของ $\sigma_i^2 = \exp(\tilde{z}_i'\tilde{\alpha})$

เมื่อ $\tilde{z}_i = (1, \ln X_i)'$ และ $\tilde{\alpha} = (\ln \sigma^2, r)'$

2) <u>รูปแบบการบวก</u> (Additive model)

$$V(\varepsilon_i) = E(\varepsilon_i^2) = \sigma^2 (1 + \lambda X_i)^2$$

ซึ่งเป็นรูปแบบเฉพาะของ $\sigma_i^2 = (\tilde{z}_i'\tilde{\alpha})^2$

เมื่อ $\tilde{z}_i = (1, X_i)'$, $\tilde{\alpha} = (\sigma, \lambda\sigma)'$

ดังนั้นการทดสอบนี้จึงสามารถนำมาใช้ในงานวิจัยครั้งนี้ได้

<u>ขั้นตอนการหาค่าของตัวสถิติของบรูชและพาแกน</u>

1) จากตัวแบบการถดถอยเชิงเส้นอย่างง่าย

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_p X_{pi} + \varepsilon_i \quad , \; i = 1,2,...,n$$

ให้ประมาณค่าพารามิเตอร์ $\beta_0, \beta_1, \ldots \beta_n$ โดยวิธี OLS เพื่อสร้างสมการถดถอย

2) คำนวณค่าของ $\hat{\varepsilon}_i$ จากข้อ 1)

3) นำค่าที่ได้จากข้อ 2) คำนวณหาค่าประมาณความแปรปรวนของความคลาดเคลื่อนจากสูตร

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^2$$

4) คำนวณค่าเวกเตอร์ $\tilde{g}$ โดยสมาชิกในเวกเตอร์ คำนวณได้จาก

$$g_i = \frac{\hat{\varepsilon}_i^2}{\hat{\sigma}^2} - 1 \quad , \quad i = 1,2,\ldots,n$$

5) สถิติทดสอบของบรูชและพาแกน คือ

$$BP = \frac{1}{2}\tilde{g}'z(z'z)^{-1}z'\tilde{g}$$

เมื่อ z เป็นเมตริกซ์ขนาด $n \times 2$ ซึ่งมีสมาชิกในแถวที่ i คือ

$\tilde{z}_i = (1, \ln X_i)$ สำหรับความแปรปรวนในรูปแบบการคูณ

และ $\tilde{z}_i = (1, X_i)$ สำหรับความแปรปรวนในรูปแบบการบวก

ภายใต้สมมติฐานว่าง ตัวสถิติของบรูชและพาแกนจะมีการแจกแจงเข้าใกล้การแจกแจงแบบไคสแควร์ ที่องศาความเป็นอิสระเท่ากับ 1

<u>หมายเหตุ</u> ถ้า z_i เป็นเวกเตอร์ที่มีขนาด $1 \times p$ แล้ว ตัวสถิติของบรูชและพาแกนจะมีการแจกแจงแบบไคสแควร์ที่องศาความเป็นอิสระ $p-1$

6) จะปฏิเสธสมมติฐานว่าง เมื่อ $BP > \chi^2_{(1),\alpha}$

ตัวอย่างวิธีการหาสถิติทดสอบของบรูชและพาแกน

จากตัวอย่างข้อมูลที่สร้างขึ้นในตารางที่ 2.1 สามารถนำมาตรวจสอบปัญหาความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ได้ โดยใช้ตัวสถิติของบรูชและพาแกนได้ดังนี้

ตารางที่ 2.2 แสดงการหาสถิติทดสอบของบรูชและพาแกนโดยใช้ตัวอย่างข้อมูลที่สร้างขึ้น

n	X_i	Y_i	$\hat{\varepsilon}_i$	$\hat{\varepsilon}_i^2$	$g_i = \frac{\hat{\varepsilon}_i^2}{\hat{\sigma}^2} - 1$
1	2.4578	1.1874	-1.1787	1.3894	0.7921
2	0.6493	1.6339	1.1611	1.3482	0.7389
3	4.1649	4.1201	-0.0332	0.0011	-0.9986
4	3.6268	2.8279	-0.7620	0.5807	-0.2510
5	3.0750	2.3099	-0.7024	0.4933	-0.3637
6	3.3516	4.2133	0.9116	0.8309	0.0717
7	2.3034	2.2473	0.0427	0.0018	-0.9977
8	4.6961	5.2096	0.5003	0.2503	-0.6772
9	3.0590	3.4557	0.4602	0.2118	-0.7268
10	4.7970	5.5533	0.7383	0.5450	-0.2971
11	3.2640	3.6646	0.4544	0.2065	-0.7337
12	3.8716	2.5303	-1.3159	1.7317	1.2336
13	2.2988	1.9289	0.5092	0.2592	-0.6657
14	4.2459	3.4240	1.2244	1.4990	0.9334
15	2.3610	4.9746	0.7365	0.5425	-0.3003
16	3.5773	-0.0164	-2.2812	5.2038	5.7120
17	2.6399	3.3036	-0.2345	0.0550	-0.9291
18	1.5538	2.3170	-0.2397	0.0575	-0.9258
19	2.8644	3.1824	0.3907	0.1526	-0.8032
20	1.6506	1.1395	-0.3816	0.1456	-0.8122

จะได้ค่า $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} \hat{\varepsilon}_i^2 = \mathbf{0.7753}$

ค่าสถิติ $BP = \frac{1}{2}\tilde{g}'z(z'z)^{-1}z'\tilde{g} = \mathbf{0.7125} < \chi^2_{(1),0.05} = \mathbf{3.846}$

ดังนั้น จึงยอมรับสมมติฐานว่าง นั่นคือ ความแปรปรวนของความคลาดเคลื่อนคงที่ ที่ระดับนัยสำคัญ 0.05

2.2.2 ตัวสถิติทดสอบของสโรเตอร์ (Szroeter's Statistics)

ตัวสถิติทดสอบของสโรเตอร์เป็นสถิติทดสอบตัวหนึ่งที่ใช้ในการตรวจสอบปัญหาความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ ซึ่งสามารถใช้ได้แม้ว่าจะไม่ทราบรูปแบบความแปรปรวนก็ตาม การตรวจสอบด้วยตัวสถิติดังกล่าวนี้ จำเป็นที่จะต้องมีการเรียงข้อมูลในลักษณะที่ทำให้ $\sigma_{i-1}^2 \le \sigma_i^2$, $i = 1,2,...,n$ และมีวิธีการคำนวณตัวสถิติดังต่อไปนี้

1) คำนวณหาค่าประมาณความคลาดเคลื่อนโดยวิธี OLS $(\hat{\varepsilon}_i)$

2) หาค่าสถิติทดสอบของสโรเตอร์

$$SZ = n(\tilde{h} - \bar{h}) \Bigg/ \left[2\sum_{i=1}^{n} (h_i - \bar{h})^2 \right]^{\frac{1}{2}}$$

โดยที่ $h_i = i$, $w_i = \hat{\varepsilon}_i^2 \Big/ \sum_{i=1}^{n} \hat{\varepsilon}_i^2$, $\tilde{h} = \sum_{i=1}^{n} w_i h_i$, $\bar{h} = \frac{1}{n} \sum_{i=1}^{n} h_i$

ภายใต้สมมติฐานว่าง $H_0 : \sigma_i^2 = \sigma_{i-1}^2$ (เทียบกับ $H_1 : \sigma_i^2 > \sigma_{i-1}^2$) ตัวสถิติของสโรเตอร์จะมีการแจกแจงเข้าใกล้การแจกแจงปกติที่มีค่าเฉลี่ยเป็น 0 และความแปรปรวนเป็น 1 (Standard Normal distribution) ดังนั้นเกณฑ์ในการตัดสินใจ คือ ที่ระดับนัยสำคัญ α เมื่อ $SZ > z_{\alpha}$ จะปฏิเสธสมมติฐานดังกล่าว

ตัวอย่างวิธีการหาสถิติทดสอบของสโรเตอร์

จากตัวอย่างข้อมูลที่สร้างขึ้นในตารางที่ 2.1 สามารถนำมาตรวจสอบปัญหาความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ได้ โดยใช้ตัวสถิติของสโรเตอร์ได้ดังนี้

ตารางที่ 2.3 แสดงการหาสถิติทดสอบของสโรเตอร์โดยใช้ตัวอย่างข้อมูลที่สร้างขึ้น

n	X_i	Y_i	$\hat{\varepsilon}_i$	$\hat{\varepsilon}_i^2$	w_i	$w_i h_i$
1	2.4578	1.1874	-1.1787	1.3894	0.0896	0.0896
2	0.6493	1.6339	1.1611	1.3482	0.0869	0.1739
3	4.1649	4.1201	-0.0332	0.0011	0.0001	0.0002
4	3.6268	2.8279	-0.7620	0.5807	0.0375	0.1498
5	3.0750	2.3099	-0.7024	0.4933	0.0318	0.1591
6	3.3516	4.2133	0.9116	0.8309	0.0536	0.3215
7	2.3034	2.2473	0.0427	0.0018	0.0001	0.0008
8	4.6961	5.2096	0.5003	0.2503	0.0161	0.1291
9	3.0590	3.4557	0.4602	0.2118	0.0137	0.1229
10	4.7970	5.5533	0.7383	0.5450	0.0351	0.3515
11	3.2640	3.6646	0.4544	0.2065	0.0133	0.1465
12	3.8716	2.5303	-1.3159	1.7317	0.1117	1.3402
13	2.2988	1.9289	0.5092	0.2592	0.0167	0.2173
14	4.2459	3.4240	1.2244	1.4990	0.0967	1.3534
15	2.3610	4.9746	0.7365	0.5425	0.0350	0.5248
16	3.5773	-0.0164	-2.2812	5.2038	0.3356	5.3696
17	2.6399	3.3036	-0.2345	0.0550	0.0035	0.0603
18	1.5538	2.3170	-0.2397	0.0575	0.0037	0.0667
19	2.8644	3.1824	0.3907	0.1526	0.0098	0.1870
20	1.6506	1.1395	-0.3816	0.1456	0.0094	0.1878
รวม				15.5059		$\tilde{h}$ = 10.9521

จะได้ค่า $\bar{h} = \frac{1}{n}\sum_{i=1}^{n} h_i = \frac{1}{20}(1+2+3+...+20) = 10.5$

ค่าสถิติ
$$SZ = n(\tilde{h} - \bar{h}) \Big/ \left[2\sum_{i=1}^{n}(h_i - \bar{h})^2\right]^{\frac{1}{2}}$$
$$= 20(10.9521 - 10.5)/63.3765 = 0.1427$$

เนื่องจาก $SZ = 0.1427 < z_{0.05} = 1.645$ ดังนั้น จึงยอมรับสมมติฐานว่าง นั่นคือ ความคลาดเคลื่อนมีความแปรปรวนคงที่ ที่ระดับนัยสำคัญ 0.05

2.3 การแจกแจงที่เกี่ยวข้อง

2.3.1 การแจกแจงแบบยูนิฟอร์ม (Uniform Distribution)

ในที่นี้จะกล่าวถึง การแจกแจงยูนิฟอร์ม ในกรณีที่เป็นการแจกแจงของตัวแปรสุ่มแบบต่อเนื่อง การแจกแจงยูนิฟอร์มอาจเรียกอีกอย่างหนึ่งว่า การแจกแจงเอกรูป ซึ่งเป็นการแจกแจงที่พบบ่อย โดยเฉพาะอย่างยิ่งในการจำลองข้อมูลทางคอมพิวเตอร์ (Simulation) ไม่ว่าจะเป็นสาขาใดก็ตาม เช่น ทางด้านวิศวกรรม ซึ่งต้องการสร้างการแจกแจงที่มีความน่าจะเป็นในการเกิดเหตุการณ์ใดเหตุการณ์หนึ่งเท่าๆ กัน

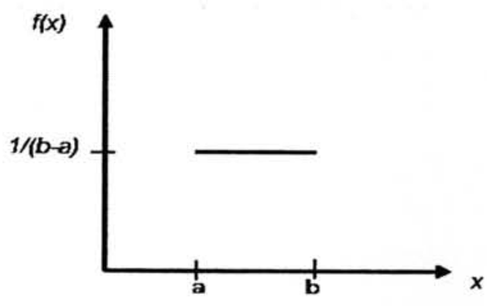
ฟังก์ชั่นความน่าจะเป็นของตัวแปร

ถ้า X เป็นตัวแปรสุ่มที่มีการแจกแจงแบบยูนิฟอร์มในช่วง (a,b) ฟังก์ชั่นหนาแน่นน่าจะเป็น หรือฟังก์ชั่นความน่าจะเป็นของการแจกแจงของ X คือ

$$f(x) = \frac{1}{b-a} \quad ; \quad a < x < b$$

ลักษณะการแจกแจงแบบยูนิฟอร์ม

1. เป็นการแจกแจงแบบต่อเนื่องในช่วงระหว่าง a และ b โดยที่ $-\infty < a < b < \infty$
2. ความน่าจะเป็นในช่วงระหว่าง a และ b จะมีค่าเท่ากัน นั่นคือ เท่ากับ $1/(b-a)$ และในช่วงอื่นๆ ความน่าจะเป็นจะเท่ากับ 0
3. พื้นที่ใต้กราฟระหว่าง a และ b จะเท่ากับ 1



รูปที่ 2.1 แสดงความหนาแน่นน่าจะเป็นของการแจกแจงแบบยูนิฟอร์ม

2.3.2 การแจกแจงแบบปกติ (Normal Distribution)

การแจกแจงปกติเป็นการแจกแจงความน่าจะเป็นแบบต่อเนื่อง (Continuous Probability Distribution) ชนิดหนึ่ง มีชื่อเรียกอีกอย่างหนึ่งว่า การแจกแจงแบบเก๊าส์เซียน (Gaussian Distribution) เป็นการแจกแจงที่มีความสำคัญมาก สำหรับตัวแปรเชิงสุ่มแบบต่อเนื่อง (continuous random variable) เนื่องจากข้อมูลส่วนใหญ่มีการแจกแจงหรือใกล้เคียงปกติ ตัวแปรสุ่มที่มีการแจกแจงปกติ เรียกว่า "ตัวแปรสุ่มปกติ" (Normal random variable)

ความสำคัญของการแจกแจงปกติ

การแจกแจงปกติ เป็นการแจกแจงที่มีความสำคัญมากในการประยุกต์เชิงสถิติ มีเหตุผลหลัก ๆ ได้แก่

1. นักวิทยาศาสตร์สังเกตว่า ตัวแปรสุ่มที่ศึกษาในการทดลองส่วนใหญ่จะประมาณเข้าสู่รูปการแจกแจงแบบปกติ เช่น ความสูงหรือน้ำหนักที่แตกต่างกันของคนในกลุ่มที่มีความคล้ายคลึงกัน หรือ การแจกแจงของความยาวของชิ้นส่วนของเหล็กที่ผลิตในกระบวนการหนึ่งที่มีความแน่นอน
2. การวัดการกระจายรูปต่างๆ สามารถจะประมาณได้ด้วยการแจกแจงปกติโดยอาศัยทฤษฎีบทแนวโน้มเข้าสู่ส่วนกลาง (Central Limit Theorem) นั่นคือ ถ้าสุ่มตัวอย่างจากรูปการแจกแจงหนึ่ง โดยสุ่มตัวอย่างขนาดโต แล้วโดยรูปการแจกแจงของตัวมันเองจะไม่สามารถประมาณได้ด้วยการแจกแจงแบบปกติได้ แต่ถ้าอาศัยทฤษฎีบทแนวโน้มเข้าสู่ส่วนกลาง ฟังก์ชั่นที่สำคัญๆ ของค่าสังเกตในตัวอย่าง จะมีการแจกแจงที่ประมาณได้ด้วยการแจกแจงแบบปกติ โดยเฉพาะสำหรับตัวอย่างสุ่มที่มีขนาดโตจากรูปการแจกแจงใดๆ ที่มีความแปรปรวนจำกัดแน่นอน การแจกแจงของค่าเฉลี่ยของตัวอย่างจะประมาณด้วยการแจกแจงแบบปกติ

ฟังก์ชั่นความน่าจะเป็นของตัวแปร

ถ้า X เป็นตัวแปรสุ่มที่มีการแจกแจงปกติที่มีค่าเฉลี่ย μ และความแปรปรวน σ^2 (โดยที่ $\sigma > 0$) แล้วฟังก์ชั่นความน่าจะเป็น (Probability density function) คือ

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad , \quad -\infty < x < \infty$$

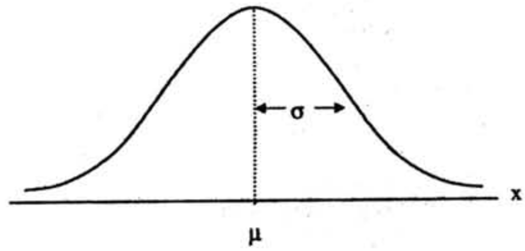
หรืออาจเขียนแทนด้วย $N(x;\mu,\sigma^2)$

เมื่อ $\pi = 3.14159...$ และ $e = 2.71828...$

เรียกการแจกแจงปกติที่มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 1 ว่า “การแจกแจงปกติมาตรฐาน” (Standard Normal Distribution) และเรียกเส้นโค้งดังกล่าวว่า “เส้นโค้งปกติมาตรฐาน” (Standard Normal Curve)

ลักษณะการแจกแจงแบบปกติ

ลักษณะของการแจกแจงแบบปกติ คือ เป็นรูปโค้งสมมาตรคล้ายระฆังคว่ำ (Bell-shaped) ขึ้นอยู่กับพารามิเตอร์ 2 ตัว คือ μ และ σ^2 กราฟจะโด่งหรือแบนราบขึ้นอยู่กับค่าความแปรปรวน σ^2 ถ้าค่าความแปรปรวนมีค่าน้อย กราฟจะโด่ง ถ้าถ้าค่าความแปรปรวนมาก กราฟจะค่อนข้างลาด และเส้นโค้งจะมีจุดยอดที่ $x = \mu$ โดยมีความสูงเท่ากับ $\frac{1}{\sigma\sqrt{2\pi}}$ และสมมาตรที่ $x = \mu$ ดังรูป



รูปที่ 2.2 แสดงเส้นโค้งการแจกแจงปกติ

คุณสมบัติของการแจกแจงปกติ

1. กราฟมีจุดยอดเพียงจุดเดียว
2. ค่าเฉลี่ย มัธยฐาน และฐานนิยมจะเท่ากันและอยู่ที่จุดกึ่งกลาง นั่นคือ อยู่ที่ $x = \mu$ (โค้งสูงที่สุด)
3. ปลายโค้งเข้าใกล้แกน x เมื่อ x มีค่าห่างจาก μ ออกไปทุกที
4. โค้งปกติจะมีลักษณะสมมาตรกับแกนตั้งที่ลากผ่าน μ โดยมีค่าเฉลี่ยเป็นจุดกึ่งกลาง ซึ่งแบ่งพื้นที่ออกเป็น 2 ส่วน โดยที่ครึ่งหนึ่งของพื้นที่ใต้โค้งปกติจะอยู่ทางซ้ายของจุดกึ่งกลาง และอีกครึ่งหนึ่งของพื้นที่ใต้โค้งปกติจะอยู่ทางขวาของจุดกึ่งกลาง
5. พื้นที่ใต้โค้งปกติทั้งหมดเท่ากับ 1

2.3.3 การแจกแจงแบบไคสแควร์ (Chi-square Distribution)

ตัวแปรสุ่มไคสแควร์ คือ ตัวแปรสุ่มที่ได้รับการพัฒนามาจากการแจกแจงปกติ กล่าวคือ ถ้า $x_1, x_2, ..., x_\upsilon$ เป็นตัวแปรสุ่มที่มีการแจกแจงปกติ โดยมีค่าเฉลี่ย μ และความแปรปรวน σ^2 แล้ว จะได้ว่า

$$z = \frac{x - \mu}{\sigma} \sim N(0,1)$$

$$\text{ดังนั้น} \quad z^2 = \left(\frac{x-\mu}{\sigma}\right)^2 \sim \chi^2_{(1)}$$

ถ้าให้ $A = z_1^2 + z_2^2 + ... + z_\upsilon^2$ แล้ว A จะมีการแจกแจงแบบไคสแควร์ ที่องศาความเป็นอิสระเท่ากับ υ

ความสำคัญของการแจกแจงแบบไคสแควร์

การแจกแจงแบบไคสแควร์เป็นการแจกแจงแบบต่อเนื่องที่พบบ่อยในการนำมาประยุกต์ใช้กับข้อมูล ซึ่งเป็นตัวแปรสุ่มที่มีค่าวัดเชิงคุณภาพหรือข้อมูลอยู่ในรูปการแจงนับ เช่น

การศึกษาผลผลิตของโรงงานอาจจำแนกเป็นดีมาก ดี ปานกลาง มีตำหนิเล็กน้อย และใช้ไม่ได้ หรือการศึกษาเกี่ยวกับอาชีพอาจจำแนกเป็นพ่อค้า นักบริหารธุรกิจ และอาจารย์ เป็นต้น

ในปัจจุบันข้อมูลในลักษณะดังกล่าวมีการนำมาใช้มากขึ้นในสาขาต่างๆ เช่น ทางสังคมศาสตร์ ซึ่งการวิเคราะห์ข้อมูลเหล่านี้สามารถทำได้โดยการทดสอบสมมติฐานโดยใช้การแจกแจงแบบไคสแควร์ (Chi-square Test) เช่น การทดสอบภาวะสารูปสนิทดี (goodness of fit test) การทดสอบความเป็นอิสระ (Test of Independence) และการทดสอบภาวะเอกพันธ์ (Test of Homogeneity) เป็นต้น

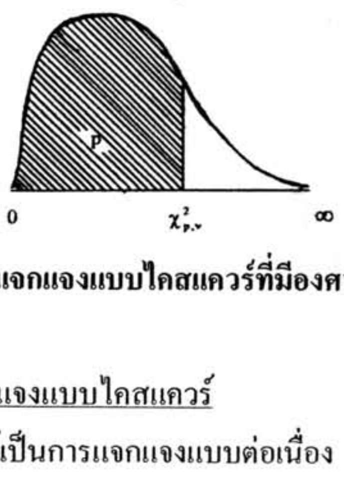
<u>ฟังก์ชั่นความน่าจะเป็นของตัวแปร</u>

การแจกแจงแบบไคสแควร์เป็นกรณีเฉพาะของการแจกแจงแบบแกมม่า เมื่อ $\alpha = \frac{\upsilon}{2}$ และ $\beta = 2$ นั่นคือ ถ้าให้ X เป็นตัวแปรสุ่มที่มีการแจกแจงแบบไคสแควร์ ที่องศาความเป็นอิสระเท่ากับ υ ฟังก์ชั่นความหนาแน่นน่าจะเป็นของ X คือ

$$f(x;\upsilon) = \begin{cases} \dfrac{1}{2^{\frac{\upsilon}{2}}\left(\frac{\upsilon}{2}-1\right)!} e^{-\frac{x}{2}} x^{\frac{\upsilon}{2}-1} & ; \ x > 0 \\ 0 & ; \ x \text{ มีค่าเป็นอย่างอื่น} \end{cases}$$

โดยที่ $e = 2.71828...$

และ ค่าเฉลี่ยเท่ากับ υ ความแปรปรวนเท่ากับ 2υ



รูปที่ 2.3 แสดงลักษณะกราฟของการแจกแจงแบบไคสแควร์ที่มีองศาความเป็นอิสระเท่ากับ υ

<u>คุณสมบัติของการแจกแจงแบบไคสแควร์</u>

1. การแจกแจงแบบไคสแควร์เป็นการแจกแจงแบบต่อเนื่อง
2. ตัวแปร χ^2 มีค่าตั้งแต่ 0 ถึง ∞ (Infinity)
3. โค้งการแจกแจงมีลักษณะเบ้ขวา โดยขึ้นอยู่กับองศาความเป็นอิสระ υ
4. ถ้าองศาความเป็นอิสระมีค่ามาก ($\upsilon \to \infty$) โค้งการแจกแจงแบบไคสแควร์จะคล้ายกับโค้งการแจกแจงปกติ

2.4 งานวิจัยที่เกี่ยวข้อง

สำหรับผลงานวิจัยที่เกี่ยวข้องกับ การทดสอบความแปรปรวนของความคลาดเคลื่อนเป็นค่าคงที่หรือไม่นั้น พบว่าได้มีนักวิจัยทำการศึกษาในหัวข้อนี้มาเป็นจำนวนมาก แต่ในที่นี้จะกล่าวถึงเฉพาะผลงานวิจัยที่สำคัญและเกี่ยวข้องกับงานวิจัยชิ้นนี้เท่านั้น

Jinook Jeong และ Kyoungwoo Lee ได้ศึกษาถึงประสิทธิภาพและอำนาจของการทดสอบของไวท์ในสมการถดถอยโดยใช้วิธีบูทสแตรป และผลปรากฏว่าตัวสถิติบูทสแตรปของไวท์มีอำนาจในการทดสอบมากกว่า White's test โดยเฉพาะอย่างยิ่งในกรณีที่ตัวอย่างมีขนาดเล็ก

Helmut Herwartz ได้ศึกษาเสนอการทดสอบอิทธิพลแบบสุ่ม (random effect) ใน Panel data เมื่อความคลาดเคลื่อนของข้อมูลภาคตัดขวางมีความสัมพันธ์กัน (cross sectional error correlation) โดยใช้การทดสอบของบรูชและพาแกน และได้ทำการหาค่าวิกฤตของสถิติทดสอบดังกล่าวด้วยวิธีบูทสแตรป