ระดับขั้นความเสรีที่เหมาะสมสำหรับการประมาณการแจกแจงไค-สแควร์ด้วยการแจกแจงปกติ

Degrees of freedom for good approximation of Chi-Square distribution by Normal distribution

วราฤทธิ์ พานิชกิจโกศลกุล

ภาควิชาคณิตศาสตร์และสถิติ คณะวิทยาศาสตร์และเทคโนโลยี

มหาวิทยาลัยธรรมศาสตร์ ศูนย์รังสิต ปทุมธานี 12121

ในการอนุมานเกี่ยวกับประชากรนั้นจะอาศัยข้อมูลที่รวบรวมจากกลุ่มตัวอย่างมาทำการศึกษา หรือที่เรียกว่า ข้อมูลตัวอย่าง และนำผลของการศึกษาข้อมูลตัวอย่างนั้นไปใช้ในการอธิบายลักษณะที่สนใจศึกษา หรือที่เรียกว่า พารามิเตอร์ของประชากร การอาศัยข้อมูลเพียงบางส่วนหรือข้อมูลตัวอย่างเพื่อนำไปอธิบายลักษณะของประชากรที่สนใจศึกษานี้เปรียบเสมือนการสรุปผลจากกรณีเฉพาะไปสู่กรณีทั่วไป ซึ่งเรียกว่า การอนุมาน (Inference) ดังนั้นการศึกษาจากข้อมูลตัวอย่างและใช้วิธีการทางสถิติมาทำการหาข้อสรุปเกี่ยวกับประชากรนั้น จึงเรียกกันว่า การอนุมานเชิงสถิติ (Statistical Inference) ซึ่งในการอนุมานอาจเป็นการทำนาย หรือการตัดสินใจเกี่ยวกับลักษณะบางอย่างของประชากรตามที่ผู้วิจัยสนใจ การอนุมานเชิงสถิติต้องอาศัยทฤษฎีการแจกแจงความน่าจะเป็น การที่ได้ทราบการแจกแจงความน่าจะเป็นของตัวอย่างสุ่มจะเป็นประโยชน์อย่างยิ่งต่อการอนุมานเชิงสถิติ

การหาการแจกแจงความน่าจะเป็นของตัวแปรสุ่มที่ต้องการในหลายปัญหาเป็นไปได้ยาก หรือมีรูปแบบที่ซับซ้อนไม่เหมาะสม หรือไม่ง่ายต่อการนำไปใช้ประโยชน์ ตัวแปรสุ่มดังกล่าวมีจำนวนไม่น้อยที่พบว่าการแจกแจงนั้นขึ้นอยู่กับจำนวนเต็มซึ่งอาจเป็นพารามิเตอร์ของการแจกแจง หรือเป็นขนาดตัวอย่าง n กล่าวคือ การแจกแจงที่ขึ้นอยู่กับขนาดตัวอย่าง n นั้นเมื่อ n มีค่ามาก การแจกแจงอาจจะลู่เข้า (Converge) หรือเบนเข้าหาการแจกแจงที่ง่ายกว่า หรือเป็นที่รู้จักกันดีที่สามารถใช้เป็นการแจกแจงโดยประมาณได้ เพราะฉะนั้น การประมาณการแจกแจงความน่าจะเป็นของตัวแปรสุ่มที่ขึ้นอยู่กับจำนวนเต็มจึงเป็นสิ่งที่น่าทำการศึกษา โดยมีจุดสนใจเพื่อให้ทราบถึงค่าของจำนวนเต็มว่าควรมีค่าเป็นเท่าใดจึงจะทำให้การประมาณการแจกแจงความน่าจะเป็นของตัวแปรสุ่มดังกล่าวมีประสิทธิภาพในสถานการณ์ต่างๆ สำหรับการศึกษาครั้งนี้จะทำการศึกษาการประมาณการแจกแจงไค-สแควร์ด้วยการแจกแจงปกติ เพื่อหาค่าระดับขั้นความเสรี (Degrees of freedom) ν ที่น้อยสุดที่ควรใช้ในการประมาณการแจกแจงดังกล่าว

1. การแจกแจงไค-สแควร์

ถ้า X เป็นตัวแปรสุ่มที่มีการแจกแจงปกติ ด้วยค่าเฉลี่ย μ และความแปรปรวน σ^2 จะได้ว่า $Z = \dfrac{X-\mu}{\sigma}$ จะมีการแจกแจงปกติมาตรฐาน หรือ $Z^2 = \left(\dfrac{X-\mu}{\sigma}\right)^2$ จะมีการแจกแจงไค-สแควร์ด้วยระดับขั้นความเสรี เท่ากับ 1

ถ้า $X_1, X_2, \ldots, X_n$ เป็นตัวอย่างสุ่มขนาด n จะได้ว่า $Y = \sum_{i=1}^{n}\left(\dfrac{X_i-\mu}{\sigma}\right)^2$ มีการแจกแจงไค-สแควร์ด้วยระดับขั้นความเสรีเท่ากับ $\nu = n$

โดยทั่วไปแล้วตัวแปรสุ่มต่อเนื่อง Y ที่มีการแจกแจงไค-สแควร์ด้วยระดับขั้นความเสรีเท่ากับ ν ซึ่งเขียนแทนด้วย $Y \sim \chi^2(\nu)$ โดยมีฟังก์ชันความหนาแน่น คือ

$$f(y) = \frac{1}{\Gamma(\nu/2)\,2^{\nu/2}}\, y^{(\nu/2)-1}\, e^{-y/2}, \quad 0 \le y < \infty$$

โดยที่ ν = 1,2,3,... และมีสมบัติดังต่อไปนี้ [1]

ฟังก์ชันก่อกำเนิดโมเมนต์	$(1-2t)^{-\nu/2}$, $t<\frac{1}{2}$
ค่าเฉลี่ย	ν
ความแปรปรวน	2ν
ฐานนิยม	$\nu-2$ เมื่อ $\nu \geq 2$
มัธยฐาน	$\nu-\frac{2}{3}$ เป็นค่าประมาณเมื่อ ν มีค่ามาก
สัมประสิทธิ์ความเบ้	$2^{3/2}\nu^{-1/2}$
สัมประสิทธิ์ความโด่ง	$3+12/\nu$

2. การแจกแจงปกติ

การแจกแจงปกติเป็นการแจกแจงของตัวแปรสุ่มต่อเนื่องที่มีความสำคัญและใช้กันมาก เนื่องจากข้อมูลส่วนใหญ่มีการแจกแจงปกติหรือใกล้เคียงการแจกแจงปกติ

ตัวแปรสุ่มต่อเนื่อง X ที่มีการแจกแจงปกติ ด้วยพารามิเตอร์ μ และ σ^2 ซึ่งเขียนแทนด้วย $X \sim N(\mu,\sigma^2)$ โดยมีฟังก์ชันความหนาแน่น คือ

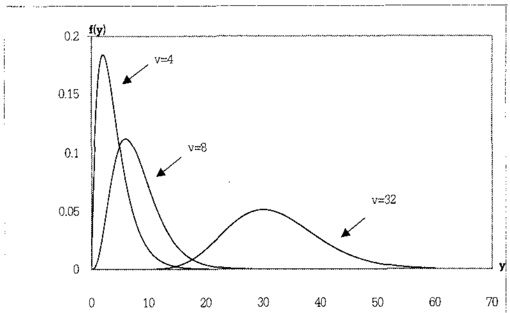
$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad , \ -\infty<x<\infty$$

ซึ่ง $-\infty<\mu<\infty$ และ $0<\sigma^2<\infty$ และมีสมบัติดังต่อไปนี้ [1]

ฟังก์ชันก่อกำเนิดโมเมนต์	$e^{\mu t+\frac{1}{2}\sigma^2 t^2}$
ค่าเฉลี่ย	μ
ความแปรปรวน	σ^2
ฐานนิยม	μ
มัธยฐาน	μ
สัมประสิทธิ์ความเบ้	0
สัมประสิทธิ์ความโด่ง	3

3. การประมาณการแจกแจงไค-สแควร์ด้วยการแจกแจงปกติ

การประมาณการแจกแจงไค-สแควร์ด้วยการแจกแจงปกติ จะขึ้นอยู่กับระดับขั้นความเสรี ν ซึ่งเป็นค่าพารามิเตอร์ของการแจกแจงไค-สแควร์กล่าวคือ เมื่อ ν มีค่าเพิ่มขึ้น การแจกแจงไค-สแควร์จะมีลักษณะใกล้เคียงการแจกแจงปกติมากขึ้น ดังแสดงในรูปที่ 1



รูปที่ 1 ฟังก์ชันความหนาแน่นของการแจกแจงไค-สแควร์ด้วยระดับขั้นความเสรี ν เท่ากับ 4 , 8 และ 32

4. การทดสอบการแจกแจงปกติของ Shapiro-Wilk

การทดสอบการแจกแจงปกติของ Shapiro-Wilk [2] มีสมมุติฐานที่ใช้ทดสอบ คือ

H_0 : ข้อมูลมีการแจกแจงปกติ

H_1 : ข้อมูลไม่มีการแจกแจงปกติ

และมีสถิติที่ใช้ทดสอบ ได้แก่

$$W = \frac{\left\{\sum_{i=1}^{k} a_{n-i+1}(X_{n-i+1}-X_i)\right\}^2}{\sum_{i=1}^{n}(X_i-\bar{X})^2}$$

เมื่อ n คือ ขนาดตัวอย่าง

k คือ จำนวนเต็มที่เล็กที่สุดที่มากกว่า n/2

a_i คือ ค่าสัมประสิทธิ์ซึ่งเปิดจากตาราง Coefficient $\{a_{n-i+1}\}$ for the W test for normality

X_i คือ ข้อมูลตัวอย่างที่เรียงลำดับ (Order sample)

การทดสอบนี้มีเกณฑ์การตัดสินใจคือ

จะปฏิเสธสมมุติฐาน H_0 ถ้าค่าสถิติ W ที่คำนวณได้น้อยกว่าค่าสถิติ W ที่เปิดจากตาราง Percentage points of the Shapiro-Wilk test ที่ขนาดตัวอย่างและระดับนัยสำคัญที่ต้องการ

5. การทดสอบสมมุติฐานระดับนัยสำคัญที่กำหนด

ในการเปรียบเทียบระดับนัยสำคัญที่กำหนด (α) กับระดับนัยสำคัญที่ควรจะเป็น (α_0) ผู้ศึกษาจะทดสอบว่าระดับนัยสำคัญที่กำหนดจะมีค่าน้อยกว่าระดับนัยสำคัญที่ควรจะเป็นอย่างมีนัยสำคัญหรือไม่ ดังนั้นสามารถเขียนสมมุติฐานการทดสอบดังนี้

$$H_0 \ : \ \alpha \geq \alpha_0$$

$$H_1 \ : \ \alpha < \alpha_0$$

จะปฏิเสธสมมุติฐาน H_0 เมื่อ $\dfrac{\hat{\alpha}-\alpha_0}{\sqrt{\dfrac{\alpha_0(1-\alpha_0)}{M}}} \leq Z_{1-\alpha^*}$

หรือ $\hat{\alpha} \leq \alpha_0 + Z_{1-\alpha^*}\sqrt{\dfrac{\alpha_0(1-\alpha_0)}{M}}$

เมื่อ α คือ ระดับนัยสำคัญที่กำหนด

α_0 คือ ระดับนัยสำคัญที่ควรจะเป็น

ซึ่ง $\hat{\alpha} = \dfrac{\text{จำนวนครั้งทั้งหมดที่ข้อมูลไม่มีการแจกแจงปกติ}}{10{,}000}$

α^* คือ ระดับนัยสำคัญของการทดสอบ ในที่นี้กำหนดเท่ากับ 0.05

M คือ จำนวนรอบของการทดลอง ในที่นี้เท่ากับ 10,000 รอบ

ตัวอย่างเช่น เมื่อระดับนัยสำคัญที่กำหนดเท่ากับ 0.01 จะได้ว่า จะปฏิเสธสมมุติฐาน H_0 เมื่อ

$$\hat{\alpha} \leq 0.01 + (-1.645)\sqrt{\frac{0.01(0.99)}{10{,}000}}$$

$$\hat{\alpha} \leq 0.0836$$

6. วิธีการศึกษา

การศึกษาครั้งนี้ต้องการหาค่าระดับขั้นความเสรี ν ที่น้อยสุดที่ควรใช้ในการประมาณการแจกแจงไค-สแควร์ $\chi^2(\nu)$ ด้วยการแจกแจงปกติ $N(\nu, 2\nu)$ วิธีการศึกษามีดังนี้

6.1 กำหนดระดับนัยสำคัญที่ควรจะเป็น (α_0) คือ 0.01 , 0.05 และ 0.10

6.2 สร้างข้อมูลที่มีการแจกแจงไค-สแควร์ ณ พารามิเตอร์ที่กำหนด โดยมีจำนวนซ้ำเท่ากับ 30

6.3 ทดสอบการแจกแจงของข้อมูลที่สร้างขึ้นในข้อ 6.2 ด้วยการทดสอบการแจกแจงปกติของ Shapiro-Wilk

6.4 หาค่าประมาณของระดับนัยสำคัญที่กำหนด ($\hat{\alpha}$)

6.5 เปรียบเทียบระดับนัยสำคัญที่กำหนด (α) น้อยกว่าระดับนัยสำคัญที่ควรจะเป็น (α_0) หรือไม่ โดยใช้การทดสอบสมมุติฐานระดับนัยสำคัญที่กำหนด ณ ระดับนัยสำคัญของการทดสอบ คือ 0.05

6.6 คัดเลือกค่าระดับขั้นความเสรี ν ซึ่งเป็นพารามิเตอร์ของการแจกแจงไค-สแควร์ที่น้อยที่สุดที่ทำให้การประมาณมีความคลาดเคลื่อนอยู่ในระดับที่กำหนด กล่าวคือ

- ถ้าค่า $\alpha_0 = 0.01$ จะยอมรับค่าระดับขั้นความเสรี ν ที่ใช้ในการประมาณค่าเมื่อค่า $\hat{\alpha} \leq 0.00836$

- ถ้าค่า $\alpha_0 = 0.05$ จะยอมรับค่าระดับขั้นความเสรี ν ที่ใช้ในการประมาณค่าเมื่อค่า $\hat{\alpha} \leq 0.04641$

- ถ้าค่า $\alpha_0 = 0.10$ จะยอมรับค่าระดับขั้นความเสรี ν ที่ใช้ในการประมาณค่าเมื่อค่า $\hat{\alpha} \leq 0.09507$

โดยจะกำหนดระดับขั้นความเสรี ν เริ่มต้นเป็นค่าน้อยๆ และเพิ่มขนาดขึ้นจนกระทั่งเป็นไปตามเงื่อนไข

การศึกษาครั้งนี้ใช้เทคนิคการจำลองแบบมอนติคาร์โล (Monte Carlo simulation technique) ทำซ้ำ 10,000 รอบ ซึ่งโปรแกรมที่ใช้ในการศึกษาครั้งนี้เขียนด้วยภาษาฟอร์แทรนเพาวเวอร์สเตชัน โดยใช้เครื่องไมโครคอมพิวเตอร์

7. ผลการศึกษา

ถ้าระดับขั้นความเสรี ν ของการแจกแจงไค-สแควร์มีค่ามากแล้ว สามารถประมาณค่าความน่าจะเป็นแบบไค-สแควร์ด้วยการแจกแจงปกติ โดยใช้พารามิเตอร์ $\mu = \nu$ และ $\sigma^2 = 2\nu$ ในการศึกษาจะเลือกระดับขั้นความเสรี ν ที่มีค่าน้อย และเพิ่มขนาด ν จนกระทั่งทำให้ระดับนัยสำคัญที่กำหนด (α) น้อยกว่าระดับนัยสำคัญที่ควรจะเป็น (α_0) อย่างมีนัยสำคัญ ด้วยความเชื่อมั่นร้อยละ 95 เพราะฉะนั้น ระดับขั้นความเสรี ν ที่แสดงในตารางจึงเป็นค่าน้อยที่สุดในการประมาณการแจกแจงดังกล่าว ดังนั้นถ้ากำหนดระดับขั้นความเสรี ν มีขนาดใหญ่กว่าค่าน้อยที่สุดนี้ การแจกแจงไค-สแควร์จะใกล้เคียงการแจกแจงปกติมากขึ้น สามารถแสดงผลการศึกษาดังตารางที่ 1

ตารางที่ 1 สรุปผลระดับขั้นความเสรี ν ที่ควรใช้ในการประมาณการแจกแจงไค-สแควร์ $\chi^2(\nu)$ ด้วยการแจกแจงปกติ $N(\nu,2\nu)$ ณ α_0 = 0.01 , 0.05 และ 0.10

ระดับนัยสำคัญที่ควรจะเป็น α_0	ν มากกว่าหรือเท่ากับ
0.01	838
0.05	847
0.10	863
ค่าเฉลี่ย	849

สรุปผลการศึกษาดังนี้

1) จากผลการศึกษาพบว่า ค่าเฉลี่ยของระดับขั้นความเสรี ν ทั้ง 3 ระดับนัยสำคัญที่ควรจะเป็น คือ 849

2) ในการประมาณการแจกแจงไค-สแควร์ $\chi^2(\nu)$ ด้วยการแจกแจงปกติ $N(\nu,2\nu)$ จะให้ค่าความถูกต้องมากขึ้นเมื่อระดับขั้นความเสรี ν มีค่ามาก แสดงดังตารางที่ 2

ตารางที่ 2 เปรียบเทียบค่าความน่าจะเป็นสะสมของการแจกแจงไค-สแควร์ค่าที่ถูกต้องและค่าประมาณ ณ ระดับขั้นความเสรี ν = 10 , 50 และ 849

r	$P(\chi^2(10) \leq r)$	
	ค่าที่ถูกต้อง	ค่าประมาณ
2	0.00366	0.03682
4	0.05265	0.08986
6	0.18474	0.18555
8	0.37116	0.32736
10	0.55951	0.50000
r	$P(\chi^2(50) \leq r)$	
	ค่าที่ถูกต้อง	ค่าประมาณ
25	0.00119	0.00621
30	0.01116	0.02275
r	$P(\chi^2(50) \leq r)$	
	ค่าที่ถูกต้อง	ค่าประมาณ
35	0.05318	0.06681
40	0.15677	0.15866
45	0.32621	0.30854
50	0.52660	0.50000
r	$P(\chi^2(849) \leq r)$	
	ค่าที่ถูกต้อง	ค่าประมาณ
700	0.00006	0.00015
720	0.00050	0.00087
750	0.00645	0.00814
800	0.11574	0.11720
820	0.24330	0.24079
849	0.50645	0.50000

8. การประยุกต์ใช้ผลการศึกษา

ผลการศึกษาระดับขั้นความเสรี ν ที่เหมาะสมสำหรับการประมาณการแจกแจงไค-สแควร์ด้วยการแจกแจงปกติ สามารถนำไปประยุกต์ใช้ได้ดังนี้

8.1 การหาค่าความน่าจะเป็นของความแปรปรวนตัวอย่าง

ถ้า S^2 เป็นความแปรปรวนของตัวอย่างสุ่มขนาด n จากประชากรที่มีการแจกแจงปกติ และมีความแปรปรวน σ^2 จะได้ตัวแปรสุ่ม $\chi^2 = \frac{(n-1)S^2}{\sigma^2}$ มีการแจกแจงไค-สแควร์ ด้วยระดับขั้นความเสรี เท่ากับ $\nu = n-1$ [3]

การหาค่าความน่าจะเป็นของความแปรปรวนตัวอย่าง จะขอยกตัวอย่างประกอบดังนี้ สมมติว่าข้อมูลชุดหนึ่งมีการแจกแจงปกติด้วยความแปรปรวน σ^2 เท่ากับ 12.75 สุ่มตัวอย่างมาจำนวน 850 หน่วย พบว่ามีความแปรปรวนตัวอย่าง S^2 เท่ากับ 10.00 ถ้าต้องการหาค่าความน่าจะเป็นที่ความแปรปรวนตัวอย่างจะมีค่ามากกว่า 13.00 หรือ $P(S^2 > 13.00)$ สามารถคำนวณได้ดังนี้

$$P(S^2 > 13.00) = P(\frac{(n-1)S^2}{\sigma^2} > \frac{(850-1)(13.00)}{12.75})$$
$$= P(\chi^2(\nu) > 865.6471)$$
$$= 0.338145$$

เนื่องจากระดับขั้นความเสรี ν มีค่ามาก สามารถนำการแจกแจงปกติมาประมาณการแจกแจงไค-สแควร์ ซึ่งประมาณค่าความน่าจะเป็นได้ดังนี้

$$P(\chi^2 > 865.6471) = P(\frac{\chi^2-\nu}{\sqrt{2\nu}} > \frac{865.6471-849}{\sqrt{(2)(849)}})$$
$$= P(Z > 0.40399)$$
$$= 0.367680$$

8.2 การประมาณค่าแบบช่วงของความแปรปรวนประชากร

ค่าประมาณแบบช่วงของ σ^2 ที่ระดับความเชื่อมั่น $(1-\alpha)100\%$ คือ

$$\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}$$

การประมาณค่าแบบช่วงของความแปรปรวนประชากรจะขอยกตัวอย่างประกอบดังนี้ สมมติว่าข้อมูลชุดหนึ่งมีการแจกแจงปกติ สุ่มตัวอย่างมาจำนวน 850 หน่วย พบว่ามีความแปรปรวนของตัวอย่าง S^2 เท่ากับ 10.00 ถ้าต้องการหาค่าประมาณแบบช่วงของ σ^2 ที่ระดับความเชื่อมั่น 95% สามารถคำนวณได้ดังนี้

$$\frac{(850-1)(10.00)}{\chi^2_{0.025,849}} < \sigma^2 < \frac{(850-1)(10.00)}{\chi^2_{0.975,849}}$$
$$\frac{8,490}{931.6409} < \sigma^2 < \frac{8,490}{770.1468}$$
$$9.1130 < \sigma^2 < 11.0239$$

เนื่องจากระดับขั้นความเสรี ν มีค่ามาก สามารถนำการแจกแจงปกติมาประมาณการแจกแจงไค-สแควร์ ซึ่งประมาณได้ดังนี้

เนื่องจาก $Z \approx \frac{\chi^2-\nu}{\sqrt{2\nu}}$

หรือ $\chi^2 \approx Z\sqrt{2\nu}+\nu$

ดังนั้น $\chi^2_{\alpha/2,\nu} \approx Z_{\alpha/2}\sqrt{2\nu}+\nu$

$$\chi^2_{1-\alpha/2,\nu} \approx Z_{1-\alpha/2}\sqrt{2\nu}+\nu$$

จากข้างต้นค่าประมาณแบบช่วงของ σ^2 ที่ระดับความเชื่อมั่น 95% โดยนำการแจกแจงปกติมาประมาณการแจกแจงไค-สแควร์ประมาณได้ดังนี้

$$\frac{(n-1)S^2}{Z_{\alpha/2}\sqrt{2\nu}+\nu} < \sigma^2 < \frac{(n-1)S^2}{Z_{1-\alpha/2}\sqrt{2\nu}+\nu}$$
$$\frac{(850-1)(10.00)}{1.96\sqrt{(2)(84959)}+849} < \sigma^2 < \frac{(850-1)(10.00)}{-1.96\sqrt{(2)(849)}+849}$$
$$9.1313 < \sigma^2 < 11.0513$$

9. บทสรุป

การประมาณการแจกแจงไค-สแควร์ด้วยการแจกแจงปกติจะให้ค่าใกล้เคียงค่าจริงมากขึ้น หรือค่าที่ถูกต้องมากขึ้นเมื่อระดับขั้นความเสรี ν มีค่าอย่างน้อย 838 , 847 และ 863

เมื่อระดับนัยสำคัญที่ควรจะเป็นเท่ากับ 0.01, 0.05 และ 0.10 ตามลำดับ ด้วยความเชื่อมั่นร้อยละ 95 ผลที่ได้สามารถนำไปประยุกต์ใช้เมื่อสุ่มตัวอย่างขนาด 800 หน่วยขึ้นไป ในการหาค่าความน่าจะเป็นของความแปรปรวนตัวอย่าง และการประมาณค่าแบบช่วงของความแปรปรวนประชากร ในกรณีตัวอย่างมีขนาดใหญ่ทำให้ไม่สะดวกที่จะใช้ค่าจากตารางการแจกแจงไค-สแควร์ ซึ่งส่วนมากหนังสือสถิติทั่วไปมักแสดงระดับขั้นความเสรี ν ตั้งแต่ 1 ถึง 100 เท่านั้น

10. เอกสารอ้างอิง

[1] Merran, E., Nicholas, H., and Brian P., Statistical Distributions., 3 ed., John Wiley and Sons, Inc., New York, 2000.

[2] Roumporn, R., A comparison on power of tests for normality among Chi-square, Shapiro-Wilk W Statistic and Shapiro-Francia W' Statistic., M.S. Thesis, Mahidol University, Bangkok, 156 p, 2000.

[3] Dennis, D. W., William, M., and Richard, L. S.,Mathematical Statistics with Applications., 6 ed.,Duxbury., Pacific Grove, 853 p, 2002.