บทที่ 2

ทฤษฎีที่เกี่ยวข้องกับการวิจัย

การวิจัยครั้งนี้ทำการศึกษาเพื่อเปรียบเทียบอำนาจการทดสอบของตัวสถิติที่ใช้ในการทดสอบเทียบความกลมกลืนสำหรับการแจกแจงแบบปกติ ได้แก่ ตัวสถิติทดสอบ r ตัวสถิติทดสอบ Z_A และตัวสถิติทดสอบ Z_C ภายใต้ลักษณะการแจกแจงของประชากรและสถานการณ์ต่างๆ ดังได้กล่าวในขอบเขตของการวิจัยในบทที่ 1

ในบทนี้จะกล่าวถึงรายละเอียดของตัวสถิติทดสอบแต่ละตัว และรายละเอียดของการแจกแจงที่ใช้ในการวิจัย ดังนี้

ในการทดสอบการแจกแจงของประชากร ให้ $x_1, x_2, ..., x_n$ เป็นตัวอย่างสุ่ม โดยในงานวิจัยนี้การแจกแจงที่ต้องการทดสอบคือ การแจกแจงแบบปกติ ซึ่งโดยทั่วไปแล้วผู้วิเคราะห์จะไม่ทราบค่าพารามิเตอร์ของการแจกแจง จึงต้องทำการประมาณค่าพารามิเตอร์ (μ และ σ^2) ซึ่งในที่นี้ใช้ตัวประมาณไม่เอนเอียง (*Unbiased Estimator*)

ได้ $\bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{n}$ เป็นตัวประมาณไม่เอนเอียงของ μ

และ $s^2 = \dfrac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$ เป็นตัวประมาณไม่เอนเอียงของ σ^2

เพราะฉะนั้น สมมติฐานสำหรับการทดสอบเทียบความกลมกลืนสำหรับการแจกแจงแบบปกติ คือ

H_0 : ประชากรมีการแจกแจงแบบ $N(\bar{x}, s^2)$

H_1 : ประชากรไม่ได้มีการแจกแจงแบบ $N(\bar{x}, s^2)$

เนื่องจากข้อมูลแต่ละชุดให้ค่าเฉลี่ย $(\bar{x})$ และความแปรปรวน (s^2) แตกต่างกัน ทำให้สมมติฐานของการทดสอบเปลี่ยนแปลงไปเรื่อยๆ ดังนั้นเพื่อความสะดวกและเป็นบรรทัดฐานเดียวกันในการวิจัย จึงจะทำการแปลงข้อมูลแต่ละชุดให้เป็นมาตรฐาน ได้ค่าเฉลี่ย $(\bar{x}')$ เท่ากับ 0 และความแปรปรวน $(s^{2'})$ เท่ากับ 1

เพราะฉะนั้น การทดสอบสมมติฐานของการทดสอบข้างต้นจึงเท่ากับการทดสอบสมมติฐานของการทดสอบดังนี้

H_0 : ประชากรมีการแจกแจงแบบ $N(0,1)$

H_1 : ประชากรไม่ได้มีการแจกแจงแบบ $N(0,1)$

2.1 ตัวสถิติทดสอบ *r*

ในปี ค.ศ.1975 *James J. Filliben* ได้เสนอตัวสถิติ *r* โดยใช้สัมประสิทธิ์สหสัมพันธ์ (*Correlation Coefficient*) ระหว่าง *order sample* กับ *order statistic median* จากการแจกแจงปกติ $N(0,1)$ โดย *Filliben* ได้ดัดแปลงตัวสถิติของ *Shapiro* และ *Wilk* ซึ่งศึกษาโดยใช้ *Probability Plot* ระหว่าง $x_{(i)}$ ซึ่งเป็น *order sample* กับ $E(x_{(i)})$ จากการศึกษาของ *Filliben* พบว่าการหาค่าของ $E(x_{(i)})$ มีความยุ่งยากในแง่ของเทคนิคอินทิเกรท *Filliben* จึงดัดแปลงตัวสถิติให้ง่ายขึ้น โดยใช้ *Probability Plot* ระหว่าง $x_{(i)}$ กับ $loc(x_{(i)}) = med(x_{(i)}) = m_{(i)}$ ซึ่ง $m_{(i)}$ จะเป็น *order statistic medians* จากการแจกแจงปกติ $N(0,1)$ ถ้าสมมติฐานว่างถูกต้อง จุดระหว่าง $x_{(i)}$ กับ $m_{(i)}$ จะเข้าใกล้เส้นตรง ดังนั้น *Filliben* ซึ่งใช้สัมประสิทธิ์สหสัมพันธ์ ในการจัดว่าจุดระหว่าง $x_{(i)}$ กับ $m_{(i)}$ จะเข้าใกล้เส้นตรงหรือไม่ ตัวสถิติที่ได้จึงเป็น

$$r = Corr(X, M)$$

$$= \frac{\sum_{i=1}^{n}(x_{(i)} - \bar{x})(m_{(i)} - \bar{m})}{\sqrt{\sum_{i=1}^{n}(x_{(i)} - \bar{x})^2 \sum_{i=1}^{n}(m_{(i)} - \bar{m})^2}}$$

เมื่อ $m_{(i)}$ = *order statistic medians* จาก $N(0,1)$

$$= \Phi^{-1}(M_i)$$

ซึ่ง

$$M_i = \begin{cases} 1 - M_n, & i = 1 \\ \dfrac{i - 0.3175}{n + 0.365}, & i = 2,3,\ldots, n-1 \\ (0.5)^{1/n}, & i = n \end{cases}$$

แต่เนื่องจาก $m_{(i)} = -m_{(n-i+1)}$

เพราะฉะนั้น $\sum_{i=1}^{n} m_{(i)} = 0$ และ $\bar{m} = 0$

$$r = \frac{\sum_{i=1}^{n} (x_{(i)} - \bar{x})\, m_{(i)}}{\sqrt{\sum_{i=1}^{n} (x_{(i)} - \bar{x})^2 \sum_{i=1}^{n} m_{(i)}^2}}$$

$$= \frac{\sum_{i=1}^{n} x_{(i)} m_{(i)} - \bar{x} \sum_{i=1}^{n} m_{(i)}}{\sqrt{\sum_{i=1}^{n} (x_{(i)} - \bar{x})^2 \sum_{i=1}^{n} m_{(i)}^2}}$$

$$= \frac{\sum_{i=1}^{n} x_{(i)} m_{(i)}}{\sqrt{\sum_{i=1}^{n} (x_{(i)} - \bar{x})^2 \sum_{i=1}^{n} m_{(i)}^2}} \qquad \text{.....................(1)}$$

เนื่องจากตัวสถิติทดสอบ r ดังกล่าวข้างต้นเป็นตัวสถิติทดสอบภายใต้สมมติฐานว่าง H_0 : ประชากรมีการแจกแจงแบบ $N(\mu,\sigma^2)$ ซึ่งในงานวิจัยนี้ต้องการทดสอบภายใต้สมมติฐานว่าง H_0 :ประชากรมีการแจกแจงแบบ $N(0,1)$ จึงต้องทำการแปลงข้อมูลให้มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 1 โดยทำการแปลงข้อมูล x_i ให้อยู่ในรูป x_i'

โดยที่ $x_i' = \dfrac{x_i - \bar{x}}{s}$, $\bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{n}$, $s^2 = \dfrac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$

ซึ่งในงานวิจัยนี้จะทำการแปลงข้อมูลที่ได้เพื่อกำหนดสมมติฐานของการทดสอบให้เป็นบรรทัดฐานเดียวกัน เพื่อสะดวกในการเปรียบเทียบตัวสถิติทดสอบ โดยผลจากการแปลงจะทำให้

$$x_{(i)}' = \frac{x_{(i)} - \bar{x}}{s}$$

ดังนั้น ตัวสถิติทดสอบ r คือ

$$r = \frac{\sum_{i=1}^{n} x_{(i)}' m_{(i)}}{\sqrt{\sum_{i=1}^{n} (x_{(i)}' - \bar{x}')^2 \sum_{i=1}^{n} m_{(i)}^2}}$$

เมื่อ n = ขนาดตัวอย่าง

$\bar{x}'$ = 0

$m_{(i)}$ = *order statistic medians* จาก $N(0,1)$

= $\Phi^{-1}(M_i)$ (รายละเอียดของการหาค่าสำหรับการวิจัยนี้แสดงไว้ในโปรแกรมย่อย SUBROUTINE STUD ในภาคผนวก ข)

$$M_i = \begin{cases} 1 - M_n\ , & i = 1 \\ \dfrac{i - 0.3175}{n + 0.365}\ , & i = 2,3,\ldots, n-1 \\ (0.5)^{1/n}\ , & i = n \end{cases}$$

เกณฑ์การตัดสินใจในการทดสอบสมมติฐานที่ระดับนัยสำคัญ α คือ จะปฏิเสธสมมติฐานว่าง H_0: ประชากรมีการแจกแจงแบบ $N(0,1)$ เมื่อค่า r ที่คำนวณได้น้อยกว่าค่าวิกฤตที่เปิดจากตาราง *Percent points of the normal probability plot correlation coefficient r* ในภาคผนวก ก ตามขนาดตัวอย่าง n และระดับนัยสำคัญที่ศึกษา

<u>ตัวอย่างที่ 1</u> การใช้ตัวสถิติทดสอบ r ในการทดสอบเทียบความกลมกลืนสำหรับการแจกแจงแบบปกติ

ข้อมูลตัวอย่าง

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
4	-6	1	8	-7	5	-9	2	-3	0

จากข้อมูลข้างต้น ประมาณค่าของ μ ได้ $\bar{x}$ = -0.5 และประมาณค่าของ σ^2 ได้ s^2 = 5.603

เพราะฉะนั้น สมมติฐานสำหรับการทดสอบเทียบความกลมกลืนสำหรับการแจกแจงแบบปกติ คือ

H_0 : ประชากรมีการแจกแจงแบบ $N(-0.5,5.603)$

H_1 : ประชากรไม่ได้มีการแจกแจงแบบ $N(-0.5,5.603)$

เมื่อทำการแปลงข้อมูล x_i ให้อยู่ในรูป $x'_i = \dfrac{x_i - \bar{x}}{s}$ แล้ว ทำให้สมมติฐานว่างของการทดสอบ H_0 : ประชากรมีการแจกแจงแบบ $N(-0.5,5.603)$ เทียบเท่ากับ H_0 : ประชากรมีการแจกแจงแบบ $N(0,1)$

ดังนั้น สมมติฐานสำหรับการทดสอบเทียบความกลมกลืนสำหรับการแจกแจงแบบปกติเมื่อแปลงข้อมูลเป็นค่ามาตรฐาน คือ

H_0 : ประชากรมีการแจกแจงแบบ $N(0,1)$

H_1 : ประชากรไม่ได้มีการแจกแจงแบบ $N(0,1)$

เรียงลำดับข้อมูลใหม่และคำนวณค่าต่างๆ ตามสูตร r ได้ดังตารางต่อไปนี้

ลำดับที่ i	$x_{(i)}$	$x'_{(i)}$	M_i	$m_{(i)}$	$m^2_{(i)}$	$(x'_{(i)} - \bar{x}')^2$	$x'_{(i)} m_{(i)}$
1	-9	-1.5172	0.0670	-1.4985	2.2455	2.3019	2.2735
2	-7	-1.1602	0.1623	-0.9850	0.9703	1.3461	1.1428
3	-6	-0.9817	0.2588	-0.6470	0.4187	0.9637	0.6351
4	-3	-0.4462	0.3553	-0.3711	0.1377	0.1991	0.1656
5	0	0.0892	0.4518	-0.1211	0.0147	0.0080	-0.0108
6	1	0.2677	0.5482	0.1211	0.0147	0.0717	0.0324
7	2	0.4462	0.6447	0.3711	0.1377	0.1991	0.1656
8	4	0.8032	0.7412	0.6470	0.4187	0.6451	0.5197
9	5	0.9817	0.8377	0.9850	0.9703	0.9637	0.9670
10	8	1.5172	0.9330	1.4985	2.2455	2.3019	2.2735
	$\bar{x} = -0.5$	$\bar{x}' = 0.0$			7.5738	9.0003	8.1644
	$s = 5.603$	$s = 1.0$					

ตัวอย่างการหาค่า $m_{(i)}$ เช่น

$$m_{(1)} = \Phi^{-1}(0.0670)$$

จากตารางการแจกแจงปกติมาตรฐาน พบว่า $m_{(1)}$ อยู่ระหว่าง $Z = -1.49$ และ $Z = -1.50$ เพราะฉะนั้น $m_{(1)} = -1.4985$

$$r = \frac{\sum_{i=1}^{n} x'_{(i)} m_{(i)}}{\sqrt{\sum_{i=1}^{n} (x'_{(i)} - \bar{x}')^2 \sum_{i=1}^{n} m_{(i)}^2}}$$

$$= \frac{8.1644}{\sqrt{(9.0003)(7.5738)}}$$

$$= 0.9889$$

จากค่าสถิติทดสอบ r ที่ได้จากการคำนวณ นำไปเปรียบเทียบกับค่าวิกฤตที่เปิดจากตาราง *Percent points of the normal probability plot correlation coefficient* r ในภาคผนวก ก เมื่อ $n = 10$ และ $\alpha = 0.05$ จะได้ค่า $r = 0.917$ พบว่า ค่าสถิติทดสอบที่คำนวณได้มีค่ามากกว่าค่าวิกฤตที่เปิดจากตาราง ดังนั้นจึงยอมรับ H_0 สรุปได้ว่า ประชากรมีการแจกแจงแบบแบบ $N(0,1)$

2.2 ตัวสถิติทดสอบใหม่

ในปี ค.ศ.2002 *Zhang J.* ได้เสนอตัวสถิติที่ใช้ในการทดสอบเทียบความกลมกลืนขึ้น โดยมีหลักการดังนี้

ให้ $x_1, x_2, ..., x_n$ เป็นตัวอย่างสุ่มขนาด n ซึ่งมีฟังก์ชันการแจกแจงสะสม $F(x)$ มีสถิติลำดับ คือ $x_{(1)}, x_{(2)}, ..., x_{(n)}$ ต้องการทดสอบสมมติฐาน

$H_0 \ : \ F(x) = F_0(x)$ สำหรับทุกค่าของ $x \in (-\infty, \infty)$

$H_1 \ : \ F(x) \neq F_0(x)$ สำหรับบางค่าของ $x \in (-\infty, \infty)$

โดยที่ $F_0(x)$ เป็นฟังก์ชันการแจกแจงแบบต่อเนื่องที่ต้องการทดสอบ โดยใช้ t เป็นจุดแบ่ง x ให้มีการแจกแจงแบบเบอร์นูลลี (*Bernoulli distribution*) และทำการทดสอบหลายๆ ครั้งที่ t เป็นค่าต่างๆ $(-\infty,\infty)$ แล้วทดสอบผลร่วมที่ได้ที่เหมือนกันเทียบกับผลทั้งหมดที่เป็นไปได้โดยมีสมมติฐานที่ทดสอบดังนี้

ให้

$$H_0 = \bigcap_{t \in (-\infty,\infty)} H_{0t}$$

$$H_1 = \bigcup_{t \in (-\infty,\infty)} H_{1t}$$

โดยที่ $H_{0t} : F(t) = F_0(t)$ และ $H_{1t} : F(t) \neq F_0(t)$

ดังนั้นการทดสอบ H_0 กับ H_1 เท่ากับการทดสอบ H_{0t} กับ H_{1t} สำหรับทุก $t \in (-\infty, \infty)$

การทดสอบ H_{0t} และ H_{1t} เมื่อ t คงที่ จะทำให้ตัวอย่างสุ่มแบ่งเป็น 2 ส่วนภายใต้ฟังก์ชันตัวชี้บ่ง $x_{it} = \mathrm{I}(x_i \le t)$ $(i = 1,2,\ldots,n)$ ดังนั้น $P(x_{it} = 1) = F(t)$ และ $P(x_{it} = 0) = 1 - F(t)$ เพราะฉะนั้น $x_{it} \sim Ber\,(F(t))$

ให้ Z_t เป็นตัวสถิติที่ใช้ทดสอบสมมติฐาน H_{0t} กับ H_{1t} จาก *Zhang* (2002) ได้กำหนดตัวสถิติทดสอบ Z ในการทดสอบ $H_0 = \bigcap_{t \in (-\infty,\infty)} H_{0t}$ และ $H_1 = \bigcup_{t \in (-\infty,\infty)} H_{1t}$ ดังนี้

$$Z = \int_{-\infty}^{\infty} Z_t \, dw(t) \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(2)$$

โดยที่ $w(t)$ เป็นฟังก์ชันถ่วงน้ำหนัก (*weight function*)

เมื่อใช้ตัวสถิติทดสอบอัตราส่วนภาวะน่าจะเป็น (*Likelihood Ratio Test Statistic*) หรือตัวสถิติ G_t^2 ในการทดสอบสมมติฐาน H_{0t} และ H_{1t} นั่นคือ Z_t ในการทดสอบนี้คือ G_t^2

โดยตัวสถิติทดสอบอัตราส่วนภาวะน่าจะเป็น (*Likelihood Ratio Test Statistic*) หรือตัวสถิติ G_t^2 หาได้ดังนี้

จาก $x_{it} \sim Ber\,(F(t))$

ดังนั้นจะได้ $f(x) = F(t)^x \, (1 - F(t))^{1-x}$

และ $l(x) = F(t)^{\sum_{i=1}^{n} x_i} \, (1 - F(t))^{n - \sum_{i=1}^{n} x_i}$

เพราะฉะนั้นจะได้ตัวสถิติทดสอบอัตราส่วนภาวะน่าจะเป็น (*Likelihood Ratio Test Statistic*) (G_t^2) คือ

$$G_t^2 = 2\ln\left(\frac{l[F_n(t)]}{l[F_0(t)]}\right)$$

$$= 2\ln\left(\frac{F_n(t)^{\sum_{i=1}^{n} x_i} \, (1 - F_n(t))^{n - \sum_{i=1}^{n} x_i}}{F_0(t)^{\sum_{i=1}^{n} x_i} \, (1 - F_0(t))^{n - \sum_{i=1}^{n} x_i}}\right)$$

$$= \quad 2\ \ln\left(\frac{F_n(t)^{nF_n(t)}\,(1-F_n(t))^{n(1-F_n(t))}}{F_0(t)^{nF_n(t)}\,(1-F_0(t))^{n(1-F_n(t))}}\right)$$

$$= \quad 2n\ \ln\left(\frac{F_n(t)^{F_n(t)}\,(1-F_n(t))^{(1-F_n(t))}}{F_0(t)^{F_n(t)}\,(1-F_0(t))^{(1-F_n(t))}}\right)$$

$$= \quad 2n\ \ln\left(\left(\frac{F_n(t)}{F_0(t)}\right)^{F_n(t)}\left(\frac{1-F_n(t)}{1-F_0(t)}\right)^{1-F_n(t)}\right)$$

$$= \quad 2n\left[F_n(t)\ \ln\left\{\frac{F_n(t)}{F_0(t)}\right\} + \left\{1-F_n(t)\right\}\ln\left\{\frac{1-F_n(t)}{1-F_0(t)}\right\}\right]$$

และ $F_n(t)$ เป็น *Empirical Distribution Function* ของตัวอย่าง $x_1, x_2, \ldots, x_n$
และใช้ฟังก์ชันถ่วงน้ำหนัก (*weight function*) ที่แตกต่างกัน ทำให้ได้ตัวสถิติทดสอบ Z_A และตัวสถิติทดสอบ Z_C ดังนี้

2.2.1 ตัวสถิติทดสอบ Z_A

จากสมการ (2) ให้ $dw(t) = \dfrac{1}{F_n(t)\ \{1 - F_n(t)\}}\,dF_n(t)$ และให้ $G_t^2 = Z_t$ นั่นคือ

$$Z \;=\; \int_{-\infty}^{\infty} Z_t\, dw(t)$$

$$=\; \int_{-\infty}^{\infty} 2n\left[F_n(t)\ \ln\left\{\frac{F_n(t)}{F_0(t)}\right\} + \left\{1-F_n(t)\right\}\ln\left\{\frac{1-F_n(t)}{1-F_0(t)}\right\}\right]\frac{1}{F_n(t)\ \{1 - F_n(t)\}}\,dF_n(t)$$

$$=\; 2n\int_{-\infty}^{\infty}\frac{F_n(t)}{F_n(t)\left\{1 - F_n(t)\right\}}\ln\left\{\frac{F_n(t)}{F_0(t)}\right\}dF_n(t) \;+$$

$$2n\int_{-\infty}^{\infty}\frac{\left\{1-F_n(t)\right\}}{F_n(t)\left\{1 - F_n(t)\right\}}\ln\left\{\frac{1-F_n(t)}{1-F_0(t)}\right\}dF_n(t)$$

$$= 2\sum_{i=1}^{n}\left(\frac{n}{n-i+\frac{1}{2}}\ln\left\{\frac{i-\frac{1}{2}}{nF_0(x_{(i)})}\right\}+\frac{n}{i-\frac{1}{2}}\ln\left[\frac{n-i+\frac{1}{2}}{n\left\{1-F_0(x_{(i)})\right\}}\right]\right)^{1}, \quad F_n(t) = \frac{i-\frac{1}{2}}{n}$$

$$= -\sum_{i=1}^{n}\left[\frac{\ln\left\{F_0(x_{(i)})\right\}}{n-i+\frac{1}{2}}+\frac{\ln\left\{1-F_0(x_{(i)})\right\}}{i-\frac{1}{2}}\right] \qquad \text{..................(3)}$$

เรียก Z ที่ได้จากสมการข้างต้นว่า Z_A

เนื่องจากตัวสถิติทดสอบ Z_A ดังกล่าวข้างต้นเป็นตัวสถิติทดสอบภายใต้สมมติฐานว่าง H_0 :ประชากรมีการแจกแจงแบบ $N(\mu,\sigma^2)$ ซึ่งในงานวิจัยนี้ต้องการทดสอบภายใต้สมมติฐานว่าง H_0 :ประชากรมีการแจกแจงแบบ $N(0,1)$ จึงต้องทำการแปลงข้อมูลให้มีค่าเฉลี่ยเท่ากับ 0 และความแปร ปรวนเท่ากับ 1 โดยทำการแปลงข้อมูล x_i ให้อยู่ในรูป x_i'

โดยที่ $$x_i' = \frac{x_i-\bar{x}}{s}, \qquad \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}, \quad s^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}$$

ซึ่งในงานวิจัยนี้จะทำการแปลงข้อมูลที่ได้เพื่อกำหนดสมมติฐานของการทดสอบให้เป็นบรรทัดฐานเดียวกัน เพื่อสะดวกในการเปรียบเทียบตัวสถิติทดสอบ โดยผลจากการแปลงจะทำให้

$$x_{(i)}' = \frac{x_{(i)}-\bar{x}}{s}$$

ดังนั้น ตัวสถิติทดสอบ Z_A คือ

$$Z_A = -\sum_{i=1}^{n}\left[\frac{\ln\left\{F_0(x_{(i)}')\right\}}{n-i+\frac{1}{2}}+\frac{\ln\left\{1-F_0(x_{(i)}')\right\}}{i-\frac{1}{2}}\right]$$

เมื่อ n = ขนาดตัวอย่าง

$F_0(x_{(i)}')$ = ความน่าจะเป็นสะสมของการแจกแจงแบบปกติมาตรฐาน

$$= \Phi\left(\frac{x_{(i)}'-\bar{x}'}{s'}\right)$$

[1]Zhang, J., "Powerful goodness-of-fit tests based on the likelihood ratio," Journal of the Royal Statistical Society Serie B 64 (2002): 284.

เกณฑ์การตัดสินใจในการทดสอบสมมติฐานที่ระดับนัยสำคัญ α คือ จะปฏิเสธสมมติฐานว่าง H_0 : ประชากรมีการแจกแจงแบบ $N(0,1)$ เมื่อค่า $10Z_A - 32$ ที่คำนวณได้มากกว่าค่าวิกฤตที่เปิดจากตาราง *Percentage points of* $10Z_A - 32$ *for testing normality* ในภาคผนวก ก ตามขนาดตัวอย่าง n และระดับนัยสำคัญที่ศึกษา

<u>ตัวอย่างที่ 2</u> การใช้ตัวสถิติทดสอบ Z_A ในการทดสอบเทียบความกลมกลืนสำหรับการแจกแจงแบบปกติ

ข้อมูลตัวอย่าง

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
4	-6	1	8	-7	5	-9	2	-3	0

จากข้อมูลข้างต้น ประมาณค่าของ μ ได้ $\bar{x}$ = -0.5 และประมาณค่าของ σ^2 ได้ s^2 = 5.603

เพราะฉะนั้น สมมติฐานสำหรับการทดสอบเทียบความกลมกลืนสำหรับการแจกแจงแบบปกติ คือ

H_0 : ประชากรมีการแจกแจงแบบ $N(-0.5,5.603)$

H_1 : ประชากรไม่ได้มีการแจกแจงแบบ $N(-0.5,5.603)$

เมื่อทำการแปลงข้อมูล x_i ให้อยู่ในรูป $x_i' = \dfrac{x_i - \bar{x}}{s}$ แล้ว ทำให้สมมติฐานว่างของการทดสอบ H_0 : ประชากรมีการแจกแจงแบบ $N(-0.5,5.603)$ เทียบเท่ากับ H_0 : ประชากรมีการแจกแจงแบบ $N(0,1)$

ดังนั้น สมมติฐานสำหรับการทดสอบเทียบความกลมกลืนสำหรับการแจกแจงแบบปกติ เมื่อแปลงข้อมูลเป็นค่ามาตรฐาน คือ

H_0 : ประชากรมีการแจกแจงแบบ $N(0,1)$

H_1 : ประชากรไม่ได้มีการแจกแจงแบบ $N(0,1)$

เรียงลำดับข้อมูลใหม่และคำนวณค่าต่างๆ ตามสูตร Z_A ได้ดังตารางต่อไปนี้

ลำดับที่ i	$x_{(i)}$	$x'_{(i)}$	$F_0(x'_{(i)})$	$n-i+\frac{1}{2}$	$\frac{\ln\left\{F_0(x'_{(i)})\right\}}{n-i+\frac{1}{2}}$	$i-\frac{1}{2}$	$\frac{\ln\left\{1-F_0(x'_{(i)})\right\}}{i-\frac{1}{2}}$
1	-9	-1.5172	0.0646	9.5	-0.2884	0.5	-0.1336
2	-7	-1.1602	0.1230	8.5	-0.2465	1.5	-0.0875
3	-6	-0.9817	0.1631	7.5	-0.2418	2.5	-0.0712
4	-3	-0.4462	0.3277	6.5	-0.1716	3.5	-0.1135
5	0	0.0892	0.5356	5.5	-0.1135	4.5	-0.1704
6	1	0.2677	0.6055	4.5	-0.1115	5.5	-0.1691
7	2	0.4462	0.6723	3.5	-0.1135	6.5	-0.1716
8	4	0.8032	0.7891	2.5	-0.0948	7.5	-0.2075
9	5	0.9817	0.8369	1.5	-0.1187	8.5	-0.2133
10	8	1.5172	0.9354	0.5	-0.1336	9.5	-0.2884
	$\bar{x}$ = -0.5	$\bar{x}'$ = 0.0					
	s = 5.603	s = 1.0					

$$Z_A = -\sum_{i=1}^{n}\left[\frac{\ln\left\{F_0(x'_{(i)})\right\}}{n-i+\frac{1}{2}}+\frac{\ln\left\{1-F_0(x'_{(i)})\right\}}{i-\frac{1}{2}}\right]$$

$= -\{(-0.2884-0.1336)+(-0.2465-0.0875)+(-0.2418-0.0712)+$
$(-0.1716-0.1135)+(-0.1135-0.1704)+(-0.1115-0.1691)+$
$(-0.1135-0.1716)+(-0.0948-0.2075)+(-0.1187-0.2133)+$
$(-0.1336-0.2884)\}$

$= -(-3.2600)$

$= 3.26$

$\therefore \quad 10Z_A - 32 = 10(3.26)-32$

$= 0.6$

จากค่า $10Z_A - 32$ ที่ได้จากการคำนวณ นำไปเปรียบเทียบกับค่าวิกฤตที่เปิดจากตารางที่ 2 ในภาคผนวก ก เมื่อ $n = 10$ และ $\alpha = 0.05$ จะได้ค่า $10Z_A - 32 = 2.970$ พบว่า ค่าสถิติทดสอบที่คำนวณได้มีค่าน้อยกว่าค่าวิกฤตที่เปิดจากตาราง *Percentage points of* $10Z_A - 32$ *for testing normality* ดังนั้นจึงยอมรับ H_0 สรุปได้ว่า ประชากรมีการแจกแจงแบบแบบ $N(0,1)$

2.2.2 ตัวสถิติทดสอบ Z_C

จากสมการ (1) ให้ $dw(t) = \dfrac{1}{F_0(t)\,\{1 - F_0(t)\}}\,dF_0(t)$ และให้ $G_t^2 = Z_t$ นั่นคือ

$$Z = \int_{-\infty}^{\infty} Z_t\, dw(t)$$

$$= \int_{-\infty}^{\infty} 2n\left[F_n(t)\ln\left\{\frac{F_n(t)}{F_0(t)}\right\} + \{1 - F_n(t)\}\ln\left\{\frac{1-F_n(t)}{1-F_0(t)}\right\}\right]\frac{1}{F_0(t)\,\{1 - F_0(t)\}}\,dF_0(t)$$

$$= \sum_{i=1}^{n}\left[\ln\left\{\frac{1}{F_0(x_{(i)})} - 1\right\} - b_{i-1} + b_i\right]^2 + C_n^{\ 2}$$

โดยที่ C_n เป็นค่าคงที่

และ $b_i = i\ln\left(\frac{i}{n}\right) + (n-i)\ln\left(1 - \frac{i}{n}\right)$

ทำให้ $b_{i-1} - b_i \approx \ln\left\{\dfrac{n - \frac{1}{2}}{i - \frac{3}{4}} - 1\right\}$

เพราะฉะนั้นจะได้

$$Z = \sum_{i=1}^{n}\left[\ln\left\{\frac{F_0(x_{(i)})^{-1} - 1}{(n - \frac{1}{2})/(i - \frac{3}{4}) - 1}\right\}\right]^2 \quad \text{.......................(4)}$$

เรียก Z ที่ได้จากสมการข้างต้นว่า Z_C

---

[2]Ibid., p. 284.

เนื่องจากตัวสถิติทดสอบ Z_C ดังกล่าวข้างต้นเป็นตัวสถิติทดสอบภายใต้สมมติฐานว่าง H_0:ประชากรมีการแจกแจงแบบ $N(\mu,\sigma^2)$ ซึ่งในงานวิจัยนี้ต้องการทดสอบภายใต้สมมติฐานว่าง H_0:ประชากรมีการแจกแจงแบบ $N(0,1)$ จึงต้องทำการแปลงข้อมูลให้มีค่าเฉลี่ยเท่ากับ 0 และความแปร ปรวนเท่ากับ 1 โดยทำการแปลงข้อมูล x_i ให้อยู่ในรูป x'_i

โดยที่ $$x'_i = \frac{x_i - \bar{x}}{s}, \qquad \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}, \quad s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

ซึ่งในงานวิจัยนี้จะทำการแปลงข้อมูลที่ได้เพื่อกำหนดสมมติฐานของการทดสอบให้เป็นบรรทัดฐานเดียวกัน เพื่อสะดวกในการเปรียบเทียบตัวสถิติทดสอบ โดยผลจากการแปลงจะทำให้

$$x'_{(i)} = \frac{x_{(i)} - \bar{x}}{s}$$

ดังนั้น ตัวสถิติทดสอบ Z_C คือ

$$Z_C = \sum_{i=1}^{n}\left[\ln\left\{\frac{F_0(x'_{(i)})^{-1} - 1}{(n-\frac{1}{2})\big/(i-\frac{3}{4}) - 1}\right\}\right]^2$$

เมื่อ n = ขนาดตัวอย่าง

$F_0(x'_{(i)})$ = ความน่าจะเป็นสะสมของการแจกแจงแบบปกติมาตรฐาน

$$= \Phi\left(\frac{x'_{(i)} - \bar{x}'}{s'}\right)$$

เกณฑ์การตัดสินใจในการทดสอบสมมติฐานที่ระดับนัยสำคัญ α คือ จะปฏิเสธสมมติฐานว่าง H_0: ประชากรมีการแจกแจงแบบ $N(0,1)$ เมื่อค่า Z_C ที่คำนวณได้มากกว่าค่าวิกฤตที่เปิดจากตาราง *Percentage points of* Z_C *for testing normality* ในภาคผนวก ก ตามขนาดตัวอย่าง *n* และระดับนัยสำคัญที่ศึกษา

<u>ตัวอย่างที่ 3</u> การใช้ตัวสถิติทดสอบ Z_C ในการทดสอบเทียบความกลมกลืนสำหรับการแจกแจงแบบปกติ

ข้อมูลตัวอย่าง

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
4	-6	1	8	-7	5	-9	2	-3	0

จากข้อมูลข้างต้น ประมาณค่าของ μ ได้ $\bar{x}$ = -0.5 และประมาณค่าของ σ^2 ได้ s^2 = 5.603

เพราะฉะนั้น สมมติฐานสำหรับการทดสอบเทียบความกลมกลืนสำหรับการแจกแจงแบบปกติ คือ

H_0 : ประชากรมีการแจกแจงแบบ $N(-0.5,5.603)$

H_1 : ประชากรไม่ได้มีการแจกแจงแบบ $N(-0.5,5.603)$

เมื่อทำการแปลงข้อมูล x_i ให้อยู่ในรูป $x_i' = \dfrac{x_i - \bar{x}}{s}$ แล้ว ทำให้สมมติฐานว่างของการทดสอบ H_0 : ประชากรมีการแจกแจงแบบ $N(-0.5,5.603)$ เทียบเท่ากับ H_0 : ประชากรมีการแจกแจงแบบ $N(0,1)$

ดังนั้น สมมติฐานสำหรับการทดสอบเทียบความกลมกลืนสำหรับการแจกแจงแบบปกติเมื่อแปลงข้อมูลเป็นค่ามาตรฐาน คือ

H_0 : ประชากรมีการแจกแจงแบบ $N(0,1)$

H_1 : ประชากรไม่ได้มีการแจกแจงแบบ $N(0,1)$

เรียงลำดับข้อมูลใหม่และคำนวณค่าต่างๆ ตามสูตร Z_C ได้ดังตารางต่อไปนี้

ลำดับที่ i	$x_{(i)}$	$x'_{(i)}$	$F_0(x'_{(i)})$	$F_0(x'_{(i)})^{-1}-1$	$\dfrac{n-\frac{1}{2}}{i-\frac{3}{4}}-1$	$\ln\left\{\dfrac{F_0(x'_{(i)})^{-1}-1}{(n-\frac{1}{2})\big/(i-\frac{3}{4})-1}\right\}$
1	-9	-1.5172	0.0646	14.4767	37.0000	-0.4075
2	-7	-1.1602	0.1230	7.1309	6.6000	0.0336
3	-6	-0.9817	0.1631	5.1302	3.2222	0.2020
4	-3	-0.4462	0.3277	2.0514	1.9231	0.0281
5	0	0.0892	0.5356	0.8672	1.2353	-0.1536
6	1	0.2677	0.6055	0.6514	0.8095	-0.0944
7	2	0.4462	0.6723	0.4875	0.5200	-0.0281
8	4	0.8032	0.7891	0.2673	0.3103	-0.0648
9	5	0.9817	0.8369	0.1949	0.1515	0.1094
10	8	1.5172	0.9354	0.0691	0.0270	0.4075
	$\bar{x}$ = -0.5	$\bar{x}'$ = 0.0				
	s = 5.603	s = 1.0				

$$Z_C = \sum_{i=1}^{n}\left[\ln\left\{\frac{F_0(x'_{(i)})^{-1}-1}{(n-\frac{1}{2})\big/(i-\frac{3}{4})-1}\right\}\right]^2$$

$$= (-0.4075)^2 + (0.0336)^2 + (0.2020)^2 + (0.0281)^2 + (-0.1536)^2 + (-0.0944)^2 + (-0.0281)^2 + (-0.0648)^2 + (0.1094)^2 + (0.4075)^2$$

$$= 0.4244$$

จากค่า Z_C ที่ได้จากการคำนวณ นำไปเปรียบเทียบกับวิกฤตที่เปิดจากตาราง *Percentage points of* Z_C *for testing normality* ในภาคผนวก ก เมื่อ n = 10 และ α = 0.05 จะได้ค่า Z_C = 6.650 พบว่า ค่าสถิติทดสอบที่คำนวณได้มีค่าน้อยกว่าค่าที่ได้จากการเปิดตาราง ดังนั้นจึงยอมรับ H_0 สรุปได้ว่า ประชากรมีการแจกแจงแบบ $N(0,1)$

ในงานวิจัยครั้งนี้สนใจที่จะเปรียบเทียบอำนาจการทดสอบของตัวสถิติที่ใช้ในการทดสอบเทียบความกลมกลืนสำหรับการแจกแจงแบบปกติทั้ง 3 ตัวดังกล่าวข้างต้น เมื่อประชากรมีการแจกแจงแบบต่างๆ ได้แก่ การแจกแจงแบบปกติ การแจกแจงแบบลอกนอร์มอล การแจกแจงแบบเบตา การแจกแจงแบบแกมมา การแจกแจงแบบที และการแจกแจงแบบจอห์นสัน ดังนั้นจะกล่าวถึงการแจกแจงเหล่านี้โดยย่อในหัวข้อต่อไปนี้ โดยแต่ละการแจกแจงจะระบุค่าเฉลี่ย ความแปรปรวนของการแจกแจง สัมประสิทธิ์ความเบ้ *(Coefficient of skewness)* แทนด้วยสัญลักษณ์ γ_1 และสัมประสิทธิ์ความโด่ง (*Coefficient of kurtosis*) แทนด้วยสัญลักษณ์ γ_2 ไว้ด้วย โดยสัมประสิทธิ์ความเบ้ (γ_1) และสัมประสิทธิ์ความโด่ง (γ_2) ของการแจกแจงของ X มีนิยามดังนี้[3]

$$\gamma_1 = \frac{E\left[(X-E(X))^3\right]}{\left(Var(X)\right)^{3/2}}$$

$$\gamma_2 = \frac{E\left[(X-E(X))^4\right]}{\left(Var(X)\right)^{2}}$$

ซึ่งค่าสัมประสิทธิ์ $\gamma_1 = 0$ ถ้าการแจกแจงไม่มีความเบ้ หรือมีความสมมาตร (*symmetry*) $\gamma_1 > 0$ ถ้าการแจกแจงแบ้ขวา หรือเบ้บวก (*positive skew*) และ $\gamma_1 < 0$ ถ้าการแจกแจงแบ้ซ้าย หรือเบ้ลบ *(negative skew)* สำหรับค่าสัมประสิทธิ์ γ_2 จะเปรียบเทียบกับความโด่งของการแจกแจงแบบปกติ ซึ่งการแจกแจงแบบปกติ มีค่า $\gamma_1 = 0$ และ $\gamma_2 = 3$ ถ้า $\gamma_2 = 3$ กล่าวว่าการแจกแจงมีความโด่งปกติ (*mesokurtic*) ถ้า $\gamma_2 > 3$ กล่าวว่าการแจกแจงมีความโด่งมากกว่าปกติ (*leptokurtic*) และถ้า $\gamma_2 < 3$ กล่าวว่าการแจกแจงมีความโด่งน้อยกว่าปกติ (*platykurtic*)

โดยทั่วไปแล้ว เราไม่สามารถทราบค่าจริงของ γ_1 และ γ_2 ดังนั้นจะทำการประมาณค่าสัมประสิทธิ์ความเบ้ (γ_1) และสัมประสิทธิ์ความโด่ง (γ_2) จากตัวประมาณโมเมนต์ ด้วยสูตรดังนี้

---

[3]มานพ วราภักดิ์, การจำลองเบื้องต้น, (กรุงเทพฯ: ศูนย์ผลิตตำราเรียนสถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ, 2547), หน้า 137-138.

จากข้อมูล $x_1, x_2, \ldots, x_n$

$$\hat{\gamma}_1 = \frac{\left[\sum_{i=1}^{n} (x_i - \bar{x})^3 \Big/ n \right]}{\left[\sum_{i=1}^{n} (x_i - \bar{x})^2 \Big/ n \right]^{\frac{3}{2}}}$$

$$\hat{\gamma}_2 = \frac{\left[\sum_{i=1}^{n} (x_i - \bar{x})^4 \Big/ n \right]}{\left[\sum_{i=1}^{n} (x_i - \bar{x})^2 \Big/ n \right]^{2}}$$

2.3 การแจกแจงแบบปกติ (*Normal Distribution*)

เมื่อ X เป็นตัวแปรสุ่มต่อเนื่องที่มีการแจกแจงแบบปกติ ด้วยพารามิเตอร์ μ และ σ^2 เขียนแทนด้วย $X \sim N(\mu,\sigma^2)$ ฟังก์ชันความหนาแน่นของ X จะอยู่ในรูปของ

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}, -\infty < x < \infty, -\infty < \mu < \infty, \sigma^2 > 0$$

เมื่อ μ เป็นพารามิเตอร์แสดงตำแหน่งของการแจกแจง

σ^2 เป็นพารามิเตอร์แสดงขนาดของการแจกแจง

ถ้าตัวแปรสุ่ม X มีการแจกแจงแบบปกติ จะได้ว่า

1. ค่าเฉลี่ยของตัวแปรสุ่ม X

$$E(X) = \mu$$

2. ความแปรปรวนของตัวแปรสุ่ม X

$$Var(X) = \sigma^2$$

3. สัมประสิทธิ์ความเบ้

$$\gamma_1 = 0$$

4. สัมประสิทธิ์ความโด่ง

$$\gamma_2 = 3$$

2.4 การแจกแจงแบบลอกนอร์มอล (*Lognormal Distribution*)

เมื่อ X เป็นตัวแปรสุ่มต่อเนื่องที่มีการแจกแจงแบบลอกนอร์มอล ด้วยพารามิเตอร์ μ และ σ^2 เขียนแทนด้วย $X \sim LN(\mu,\sigma^2)$ ฟังก์ชันความหนาแน่นของ X จะอยู่ในรูปของ

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(\ln x-\mu)^2\right\}, \quad x>0,\ -\infty<\mu<\infty,\ \sigma>0$$

เมื่อ μ เป็นพารามิเตอร์แสดงขนาดของการแจกแจง

σ^2 เป็นพารามิเตอร์แสดงรูปร่างของการแจกแจง

ถ้าตัวแปรสุ่ม X มีการแจกแจงแบบลอกนอร์มอล จะได้ว่า

1. ค่าเฉลี่ยของตัวแปรสุ่ม X

$$E(X) = \exp\left\{\mu+\frac{\sigma^2}{2}\right\}$$

2. ความแปรปรวนของตัวแปรสุ่ม X

$$Var(X) = \exp(2\mu+\sigma^2)[\exp(\sigma^2)-1]$$

3. สัมประสิทธิ์ความเบ้

$$\gamma_1 = (\omega+2)\sqrt{\omega-1}, \qquad \omega=\exp(\sigma^2)$$

4. สัมประสิทธิ์ความโด่ง

$$\gamma_2 = \omega^4+2\omega^3+3\omega^2-3, \qquad \omega=\exp(\sigma^2)$$

2.5 การแจกแจงแบบเบตา (*Beta Distribution*)

เมื่อ X เป็นตัวแปรสุ่มต่อเนื่องที่มีการแจกแจงแบบเบตา ด้วยพารามิเตอร์ α และ β เขียนแทนด้วย $X \sim Be(\alpha,\beta)$ ฟังก์ชันความหนาแน่นของ X จะอยู่ในรูปของ

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\, x^{\alpha-1}(1-x)^{\beta-1}, \qquad 0 \le x \le 1,\ \alpha > 0,\ \beta > 0$$

เมื่อ α และ β เป็นพารามิเตอร์แสดงรูปร่างของการแจกแจง

ถ้าตัวแปรสุ่ม X มีการแจกแจงแบบเบตา จะได้ว่า

1. ค่าเฉลี่ยของตัวแปรสุ่ม X

$$E(X) = \frac{\alpha}{\alpha+\beta}$$

2. ความแปรปรวนของตัวแปรสุ่ม X

$$Var(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

3. สัมประสิทธิ์ความเบ้

$$\gamma_1 = \frac{2(\beta-\alpha)(\alpha+\beta+1)^{1/2}}{(\alpha+\beta+2)(\alpha\beta)^{1/2}}$$

4. สัมประสิทธิ์ความโด่ง

$$\gamma_2 = \frac{3(\alpha+\beta+1)(2(\alpha+\beta)^2+\alpha\beta(\alpha+\beta-6))}{\alpha\beta(\alpha+\beta+2)(\alpha+\beta+3)}$$

2.6 การแจกแจงแบบแกมมา (*Gamma Distribution*)

เมื่อ X เป็นตัวแปรสุ่มต่อเนื่องที่มีการแจกแจงแบบแกมมา ด้วยพารามิเตอร์ α และ λ เขียนแทนด้วย $X \sim G(\alpha,\lambda)$ ฟังก์ชันความหนาแน่นของ X จะอยู่ในรูปของ

$$f(x) = \frac{\lambda^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} \exp\left(-\lambda x\right) , \qquad x \geq 0, \alpha > 0, \lambda > 0$$

เมื่อ α เป็นพารามิเตอร์แสดงรูปร่างของการแจกแจง

λ เป็นพารามิเตอร์แสดงขนาดของการแจกแจง

ถ้าตัวแปรสุ่ม X มีการแจกแจงแบบแกมมา จะได้ว่า

1. ค่าเฉลี่ยของตัวแปรสุ่ม X

$$E(X) = \frac{\alpha}{\lambda}$$

2. ความแปรปรวนของตัวแปรสุ่ม X

$$Var(X) = \frac{\alpha}{\lambda^2}$$

3. สัมประสิทธิ์ความเบ้

$$\gamma_1 = \frac{2}{\sqrt{\alpha}}$$

4. สัมประสิทธิ์ความโด่ง

$$\gamma_2 = 3 + \frac{6}{\alpha}$$

2.7 การแจกแจงแบบที (*t Distribution*)

เมื่อ X เป็นตัวแปรสุ่มต่อเนื่องที่มีการแจกแจงแบบที ด้วยพารามิเตอร์ k เขียนแทนด้วย $X \sim t(k)$ ฟังก์ชันความหนาแน่นของ X จะอยู่ในรูปของ

$$f(x) = \frac{\Gamma\left[\dfrac{k+1}{2}\right]}{\sqrt{\pi k}\,\Gamma\left(\dfrac{k}{2}\right)} \frac{1}{\left(1+\dfrac{x^2}{k}\right)^{(k+1)/2}} , \quad -\infty < x < \infty , k = 1,2,\ldots$$

เมื่อ k เป็นพารามิเตอร์แสดงรูปร่างของการแจกแจง

ถ้าตัวแปรสุ่ม X มีการแจกแจงแบบที จะได้ว่า

1. ค่าเฉลี่ยของตัวแปรสุ่ม X

$$E(X) = 0 , \qquad k > 1$$

2. ความแปรปรวนของตัวแปรสุ่ม X

$$Var(X) = \frac{k}{k-2} , \qquad k > 2$$

3. สัมประสิทธิ์ความเบ้

$$\gamma_1 = 0 , \qquad k > 3$$

4. สัมประสิทธิ์ความโด่ง

$$\gamma_2 = \frac{3(k-2)}{k-4} , \qquad k > 4$$

2.8 การแจกแจงแบบจอห์นสัน (*Johnson Distribution*)[4]

เมื่อ X เป็นตัวแปรสุ่มต่อเนื่องที่มีการแจกแจงแบบจอห์นสัน ด้วยพารามิเตอร์ $\alpha_1, \alpha_2, \gamma$ และ β เขียนแทนด้วย $X \sim J(\alpha_1, \alpha_2, \gamma, \beta)$ ฟังก์ชันความหนาแน่นของ X จะอยู่ในรูปของ

$$f(x) = \frac{\alpha_2}{\sqrt{2\pi}\sqrt{(x-y)^2 + \beta^2}} \exp\left[-\frac{1}{2}\left\{\alpha_1 + \alpha_2 \ln\left(\frac{x-\gamma}{\beta} + \sqrt{\left(\frac{x-\gamma}{\beta}\right)^2 + 1}\right)\right\}^2\right],$$

$$-\infty < x < \infty,\ -\infty < \alpha_1 < \infty,\ \alpha_2 > 0,\ -\infty < \gamma < \infty,\ \beta > 0$$

เมื่อ $\alpha_1, \alpha_2, \gamma$ และ β เป็นพารามิเตอร์ของการแจกแจง

การแจกแจงแบบจอห์นสันมีรูปร่างได้ทั้งแบบเบ้ซ้าย เบ้ขวา และสมมาตร โดยขึ้นอยู่กับการกำหนดค่าของพารามิเตอร์ ซึ่งการแจกแจงจะเป็นแบบเบ้ซ้ายถ้า $\alpha_1 > 0$ สมมาตรถ้า $\alpha_1 = 0$ และเบ้ขวาถ้า $\alpha_1 < 0$

---

[4]เรื่องเดียวกัน, หน้า 204.