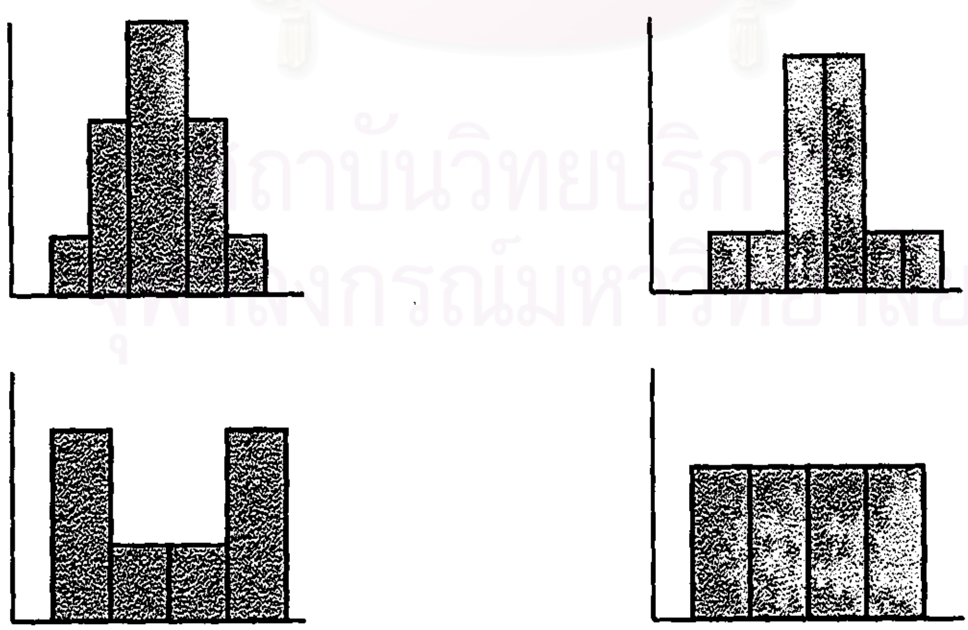
บทที่ 2

ตัวสถิติที่เกี่ยวข้องกับงานวิจัย

วิธีการหนึ่งที่สามารถบอกถึงลักษณะของข้อมูลในเบื้องต้น คือ การนำข้อมูลมาสร้างเป็นภาพ เมื่อเรานำข้อมูลมาสร้างเป็นกราฟหรือฮีสโตแกรม ภาพที่ได้จะสะท้อนให้เห็นรูปแบบการแจกแจงของข้อมูลเหล่านั้น ซึ่งรูปแบบการแจกแจงเป็นคุณสมบัติที่สามารถบอกได้ถึงผลสรุปที่ทำให้สามารถเข้าใจโครงสร้างของข้อมูลได้ แต่การแจกแจงไม่ใช่มีเพียงหนึ่งหรือสองการแจกแจง แต่มีหลายรูปแบบที่เป็นไปได้ เราอาจมองโดยรวมแล้วสามารถแยกได้เป็นสองกลุ่มใหญ่ คือ การแจกแจงที่สมมาตร และการแจกแจงที่ไม่สมมาตร

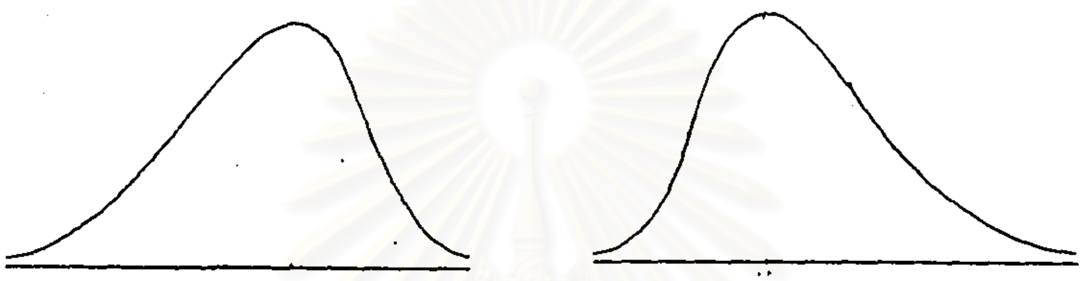
การแจกแจงสมมาตร (Symmetric Distribution) เป็นการแจกแจงที่ข้อมูลมีลักษณะการกระจายทางขวาเหมือนการกระจายทางซ้าย ซึ่งทั้งสองด้านเป็นภาพสะท้อนของกันและกันระหว่างค่าเฉลี่ย หรือ ค่ามัธยฐาน (ถ้ามี) ที่เป็นค่ากลางของข้อมูล การแจกแจงที่สมมาตรนี้มีหลายแบบด้วยกัน ดังเช่นในรูปที่ 2.1 รูปแบบการแจกแจงแบบนี้มีรูปแบบที่สำคัญในทางสถิติ คือ การแจกแจงปกติ (Normal Distribution)



รูปที่ 2.1 รูปแบบการแจกแจงที่สมมาตรแบบต่าง ๆ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

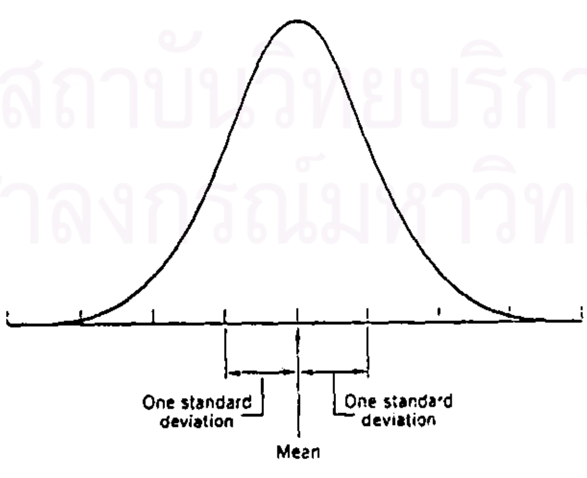
การแจกแจงที่ไม่สมมาตร (Non-symmetric Distribution) เป็นการแจกแจงที่ข้อมูลมีลักษณะการกระจายทางขวาไม่เหมือนกับการกระจายทางซ้าย อาจจะเป็นการที่ข้อมูลส่วนใหญ่อยู่ทางขวา และกระจายไปในทางด้านซ้าย หรือในทางกลับกัน ข้อมูลส่วนใหญ่อยู่ทางซ้ายและกระจายไปในทางด้านขวา รูปแบบการแจกแจงแบบนี้เราอาจเรียกอีกอย่างหนึ่งว่าการแจกแจงแบบเบ้ (Skewed Distribution) ดังแสดงในรูปที่ 2.2



รูปที่ 2.2 รูปแบบการแจกแจงที่ไม่สมมาตรแบบต่าง ๆ

การแจกแจงปกติ (Normal Distribution)

การแจกแจงที่สำคัญและใช้ประโยชน์อย่างมากในทางสถิติ ทั้งทางด้านทฤษฎีและปฏิบัติ คือ การแจกแจงปกติ ซึ่งเป็นการแจกแจงต่อเนื่อง (Continous Distribution) กราฟของการแจกแจงปกติเรียกว่า โค้งปกติ (normal curve) มีลักษณะคล้ายระฆังคว่ำ ดังแสดงในรูปที่ 2.3



รูปที่ 2.3 ลักษณะการแจกแจงปกติ

รูปแบบฟังก์ชันของการแจกแจงปกติ

ตัวแปรสุ่ม x มีการแจกแจงปกติ ซึ่งมีค่าเฉลี่ยเท่ากับ μ และค่าความแปรปรวนเท่ากับ σ^2 ฟังก์ชันความหนาแน่นของความน่าจะเป็น (probability density function) ของ x กำหนดได้ดังนี้

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-1}{2}\frac{(x-\mu)^2}{\sigma^2}}$$

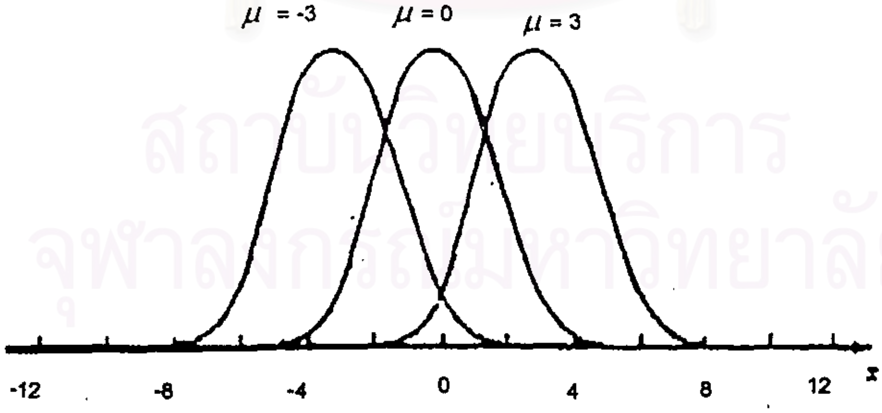
เมื่อ

μ และ σ เป็นค่าพารามิเตอร์ โดยที่ $\sigma > 0$

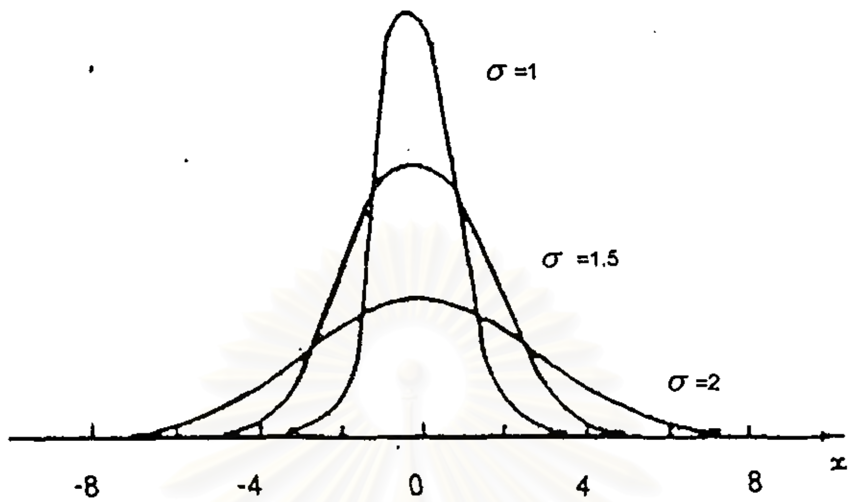
π = 3.14189...

e = 2.71828...

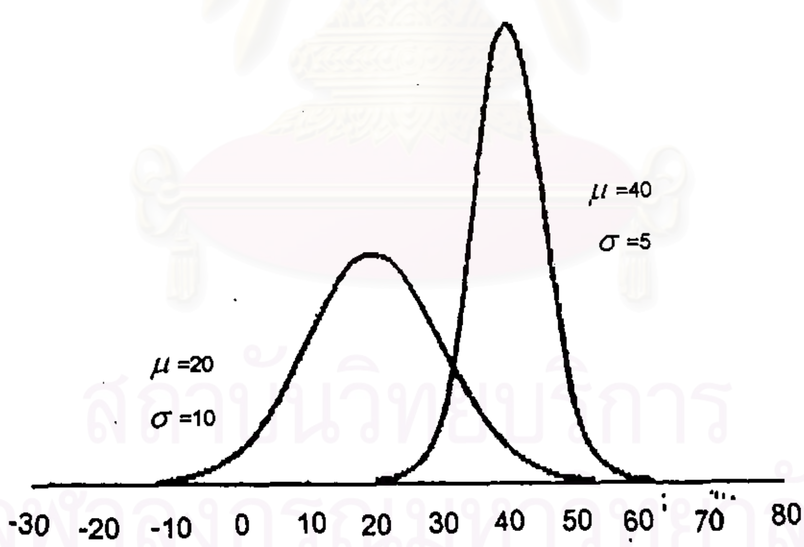
เนื่องจากสมการของการแจกแจงปกติขึ้นอยู่กับค่าพารามิเตอร์ 2 ตัว คือ μ และ σ เราอาจเขียนฟังก์ชันความหนาแน่นของความน่าจะเป็นของ x แทนด้วย $N\left(\mu,\sigma^2\right)$ ซึ่งค่าของ μ และ σ จะเป็นตัวกำหนดตำแหน่งและรูปร่างของโค้งปกติ



รูปที่ 2.4 การแจกแจงปกติที่มีค่าเฉลี่ยแตกต่างกัน
แต่มีส่วนเบี่ยงเบนมาตรฐานเท่ากัน



รูปที่ 2.5 การแจกแจงปกติที่มีค่าเฉลี่ยเท่ากัน
แต่มีส่วนเบี่ยงเบนมาตรฐานไม่เท่ากัน

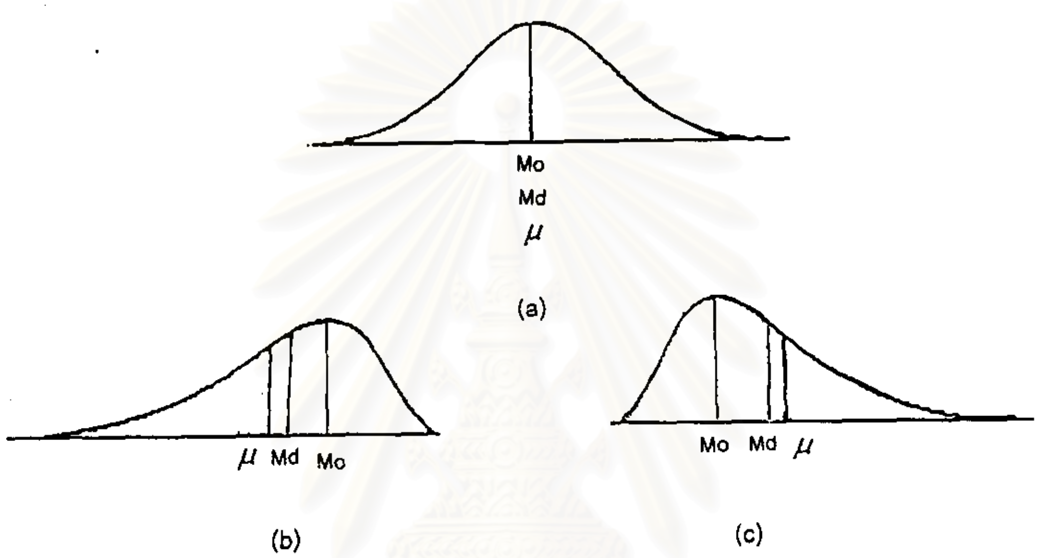


รูปที่ 2.6 การแจกแจงปกติที่มีค่าเฉลี่ย
และส่วนเบี่ยงเบนมาตรฐานแตกต่างกัน

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

การแจกแจงแบบเบ้ (Skewed distribution)

ค่าสถิติเบื้องต้นที่สามารถหาได้จากข้อมูลอย่างง่าย ๆ คือ ค่าเฉลี่ย (Mean) ค่ามัธยฐาน (Median) ค่าฐานนิยม (Mode) ทั้งสามค่ายังสามารถบอกถึงรูปแบบการแจกแจงของข้อมูลได้อย่างคร่าว ๆ อีกด้วย นั่นคือ เมื่อการแจกแจงเป็นแบบสมมาตรจะพบว่า ค่าสถิติทั้งสามค่าจะมีค่าเท่ากัน อยู่ที่จุดเดียวกันของการแจกแจง แต่เมื่อการแจกแจงมีรูปแบบที่ไม่สมมาตร ค่าของสถิติทั้งสามค่าจะแตกต่างกันออกไป ดังรูปที่ 2.7



รูปที่ 2.7 แสดงค่าสถิติ Mean , Median , Mode
ในรูปแบบการแจกแจงแบบต่าง ๆ
a) การแจกแจงแบบสมมาตร
b) การแจกแจงแบบเบ้ซ้าย
c) การแจกแจงแบบเบ้ขวา

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

นอกจากค่าสถิติทั้งสามค่าที่สามารถบอกรูปแบบการแจกแจงได้อย่างคร่าว ๆ แล้ว ยังมีค่าสถิติอีก 2 ค่าที่สามารถบอกรูปแบบการแจกแจงของข้อมูลนั้น ๆ ได้ และสามารถแสดงให้เห็นถึงความใกล้เคียงกับค่าของกรณีปกติ ค่าสถิติทั้ง 2 ค่าคือ ความเบ้และความโด่ง

ความเบ้ (skewness) :

สัมประสิทธิ์ความเบ้ (skewness coefficient) เป็นค่าที่ใช้วัดว่าการแจกแจงนั้น ๆ เป็นการแจกแจงแบบสมมาตรหรือไม่ โดยวิธีที่ใช้ในการวัดค่าความเบ้ คือ วิธีโมเมนต์ (Moment) ซึ่งมีสูตรสำหรับหาค่าวัดสมมาตรหรือสัมประสิทธิ์ความเบ้จากประชากรเป็นดังนี้

$$\alpha_3 = \frac{\mu_3}{\sigma^3} = \frac{E\left((x-\mu)^3\right)}{\left(V(x)\right)^{3/2}}$$

เมื่อ

μ_3 คือ โมเมนต์ศูนย์กลางอันดับที่ 3 ซึ่งมีค่าเท่ากับ $E\left((x-\mu)^3\right)$

σ คือ ส่วนเบี่ยงเบนมาตรฐานของประชากร ซึ่งมีค่าเท่ากับ $\sqrt{V(x)}$

สามารถประมาณค่าวัดสมมาตรหรือสัมประสิทธิ์ความเบ้จากข้อมูลตัวอย่างได้ค่าดังนี้

$$\hat{\alpha}_3 = \frac{m_3}{(m_2)^{3/2}}$$

โดยที่

$$m_3 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^3}{n}$$

$$m_2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}$$

จะได้ว่า

$$\text{ความเบ้} = \frac{\sqrt{n}\sum_{i=1}^{n}(x_i-\bar{x})^3}{\left[\sum_{i=1}^{n}(x_i-\bar{x})^2\right]^{3/2}}$$

ค่าความเบ้จะเข้าใกล้ 0 เมื่อการแจกแจงของข้อมูลเข้าใกล้สมมาตร และจะมีค่าเป็นบวกถ้าเป็นการแจกแจงแบบเบ้ขวา สำหรับการแจกแจงแบบเบ้ซ้ายค่าความเบ้จะมีค่าเป็นลบ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ความโด่ง (kurtosis) :

สัมประสิทธิ์ความโด่ง (kurtosis coefficient) เป็นค่าที่ใช้แสดงว่ามีค่าจำนวนมากหรือจำนวนน้อยที่อยู่ใกล้กึ่งกลางของการแจกแจงของข้อมูลนั้น ๆ โดยวิธีที่ใช้ในการวัดค่าความโด่ง คือ วิธีโมเมนต์ (Moment) ซึ่งมีสูตรสำหรับหาค่าความโด่งเป็นดังนี้

$$\alpha_4 = \frac{\mu_4}{\sigma^4} = \frac{E\left((x-\mu)^4\right)}{(V(x))^2}$$

เมื่อ

μ_4 คือ โมเมนต์ศูนย์กลางอันดับที่ 4 ซึ่งมีค่าเท่ากับ $E\left((x-\mu)^4\right)$

σ คือ ส่วนเบี่ยงเบนมาตรฐานของประชากร ซึ่งมีค่าเท่ากับ $\sqrt{V(x)}$

สามารถประมาณค่าความโด่งจากข้อมูลตัวอย่างได้ค่าดังนี้

$$\hat{\alpha}_4 = \frac{m_4}{(m_2)^2}$$

โดยที่

$$m_4 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^4}{n}$$

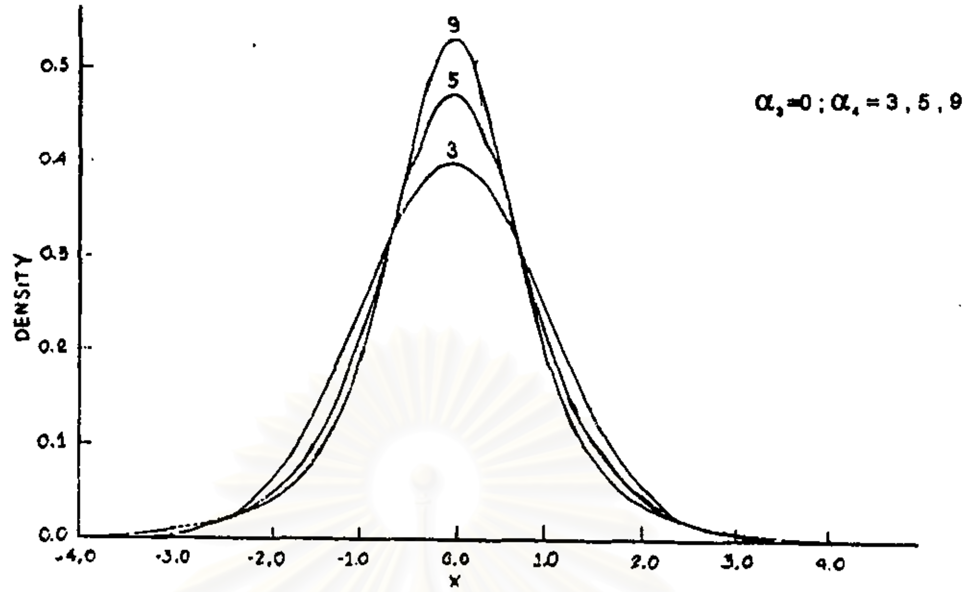
$$m_2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}$$

จะได้ว่า

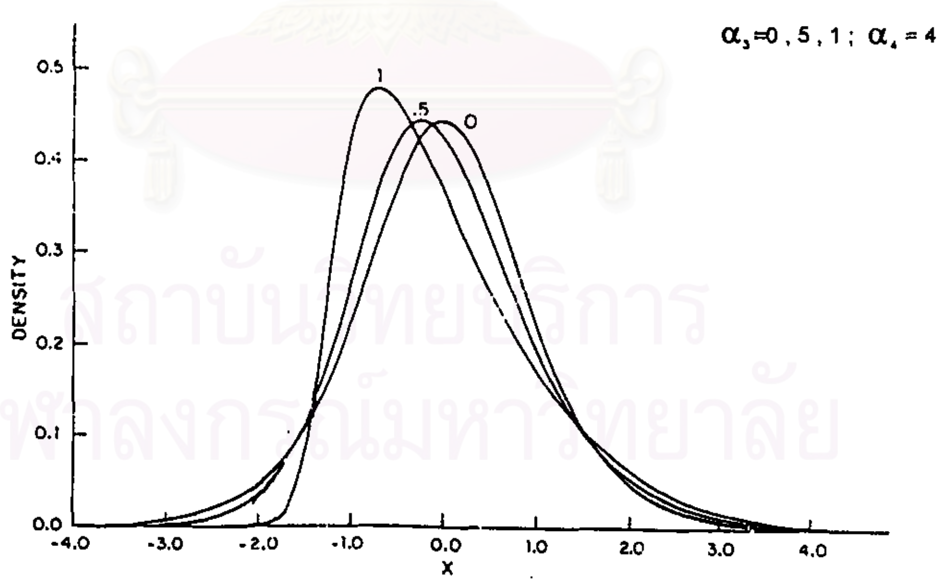
$$\text{ความโด่ง} = \frac{n\sum_{i=1}^{n}(x_i-\bar{x})^4}{\left[\sum_{i=1}^{n}(x_i-\bar{x})^2\right]^2}$$

ถ้าค่าความโด่ง < 3 แสดงว่าการแจกแจงนั้นมีลักษณะแบนราบกว่าการแจกแจงปกติ และถ้าค่าความโด่ง > 3 แสดงว่าการแจกแจงนั้นมีลักษณะโด่งกว่าการแจกแจงปกติ และจะเข้าใกล้ 3 เมื่อการแจกแจงของข้อมูลเข้าใกล้การแจกแจงปกติ

รูปแบบการแจกแจงแบบเบ้ต่าง ๆ โดยใช้ค่าความเบ้และค่าความโด่งเป็นค่าวัดความสมมาตร แสดงการเปรียบเทียบได้ดังรูปที่ 2.8 และ รูปที่ 2.9



รูปที่ 2.8 แสดงรูปแบบการแจกแจงที่มีความโด่ง (α_4) แตกต่างกัน



รูปที่ 2.9 แสดงรูปแบบการแจกแจงที่มีความเบ้ (α_3) แตกต่างกัน

การแจกแจงแลมดาร์ของตูกีร์ (Tukey ' s Lambda Distribution)

สำหรับการแจกแจงความน่าจะเป็นแบบต่อเนื่องในรูปแบบต่างๆ เราสามารถนิยามได้โดยใช้ฟังก์ชันการแจกแจง (distribution function) หรือโดยใช้ฟังก์ชันความหนาแน่น (density function) อย่างไรก็ตามถ้าเราสามารถหาค่าของฟังก์ชันเปอร์เซ็นต์ไทล์ (percentile function) ได้ เราก็สามารถที่จะนิยามได้โดยใช้ฟังก์ชันเปอร์เซ็นต์ไทล์นั้น ๆ ซึ่งการแจกแจงแลมดาร์ของตูกีร์ได้ใช้หลักการของฟังก์ชันเปอร์เซ็นไทล์ในการหารูปแบบการแจกแจงที่เหมาะสมสำหรับข้อมูลชุดหนึ่ง ดังสมการ

$$R(p) = \left[p^{\lambda} - (1-p)^{\lambda}\right] / \lambda \qquad ;\ 0 \le p \le 1$$

ภายหลังต่อมา Ramberg และ Schmeiser ได้วิเคราะห์สมการข้างต้นให้มีลักษณะทั่วไปมากยิ่งขึ้น โดยสมการที่ได้นั้นขึ้นอยู่กับค่าพารามิเตอร์ 4 ค่า ได้สมการคือ

$$X = R(p) = \lambda_1 + \frac{\left[p^{\lambda_3} - (1-p)^{\lambda_4}\right]}{\lambda_2} \quad ; \qquad 0 < p < 1 \qquad (3.1)$$

เมื่อ

λ_1	คือ	พารามิเตอร์กำหนดตำแหน่ง (location parameter)
λ_2	คือ	พารามิเตอร์มาตราส่วน (scale parameter)
λ_3 และ λ_4	คือ	พารามิเตอร์สัณฐาน (shape parameter)

จากฟังก์ชันเปอร์เซ็นต์ไทล์ข้างต้น สามารถหาฟังก์ชันความหนาแน่นได้ คือ

$$f(x) = f[R(p)]$$

$$= \lambda_2\left[\lambda_3 p^{\lambda_3 - 1} + \lambda_4 (1-p)^{\lambda_4 - 1}\right]^{-1} \qquad ; \qquad 0 < p < 1$$

Ramberg และ Schmeiser ได้ใช้สมการโมเมนต์ที่ k ในการหาค่าพารามิเตอร์ทั้ง 4 ค่า และแสดงค่าพารามิเตอร์ทั้ง 4 ค่าไว้ในตารางในค่าระดับความเบ้และความโด่งแตกต่างกัน (ตา-รางแสดงไว้ในภาคผนวก)

จากฟังก์ชันความหนาแน่นข้างต้น สามารถประมาณรูปแบบความหนาแน่นที่เหมาะสมได้เป็นอย่างดี ตัวอย่างเช่น การแจกแจงแบบปกติ เราจะพบว่า สามารถประมาณได้ด้วยพารามิเตอร์ทั้ง 4 ค่า คือ $\lambda_1 = 0$, $\lambda_2 = 0.1975$ และ $\lambda_3 = \lambda_4 = 0.1349$

การแจกแจงแบบไวบูลล์ (Weibull Distribution)

นักฟิสิกส์ชาวสวีเดนชื่อ Waloddi Weibull เป็นผู้แนะนำการแจกแจงนี้เมื่อ ค.ศ. 1939 ซึ่งเป็นการแจกแจงที่เกิดขึ้นเนื่องจากรูปแบบความเป็นจริงโดยทั่วไป สำหรับอายุการใช้งานของเครื่องจักรกลต่าง ๆ โดยมีฟังก์ชันความหนาแน่นเป็นดังนี้

$$f(x)=\begin{cases}\alpha\beta^{-\alpha}x^{\alpha-1}e^{-(\alpha/\beta)^{\alpha}} & ;\ x>0\,,\alpha>0\,,\beta>0\\ 0 & ;\ \text{อื่นๆ}\end{cases}$$

โดยที่

x เป็นค่าของข้อมูลที่มีการแจกแจงแบบไวบูลล์

α เป็นพารามิเตอร์ลักษณะ(shape parameter) ของการแจกแจง

β เป็นพารามิเตอร์มาตราส่วน (scale parameter) ของการแจกแจง

คุณสมบัติของการแจกแจงแบบไวบูลล์

1. โค้งมีลักษณะเปลี่ยนแปลงไปตามพารามิเตอร์
2. ค่าความแปรปรวนและค่าเฉลี่ยของตัวแปรสุ่มไวบูลล์ขึ้นอยู่กับพารามิเตอร์ α และ β โดยที่

$$\text{ค่าเฉลี่ย} = \frac{\beta}{\alpha}\Gamma\left(\frac{1}{\alpha}\right)$$

$$\text{ค่าความแปรปรวน} = \frac{\beta^2}{\alpha}\left\{2\Gamma\left(\frac{2}{\alpha}\right)-\frac{1}{\alpha}\left[\Gamma\left(\frac{1}{\alpha}\right)\right]^2\right\}$$

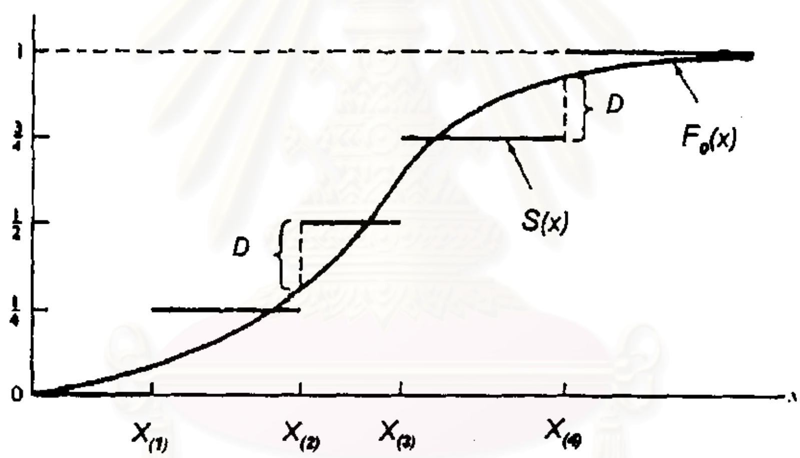
3. การแจกแจงไวบูลล์ที่มีค่า $\alpha=1$ และ β เป็นค่าใด ๆ ที่มากกว่า 0 นั้นจะเป็นการแจกแจงแบบเลขชี้กำลัง (Exponential Distribution) ที่มีพารามิเตอร์เป็น β นั่นคือ

$$Weibull(1,\beta)=exp(\beta)$$

สถิติที่ใช้ในการทดสอบการแจกแจง

การทดสอบว่าข้อมูลที่มีอยู่นั้นมีการแจกแจงแบบใด เรียกว่าการทดสอบเทียบความกลมกลืน (Goodness of Fit Test) โดยเลือกใช้วิธีการทดสอบของ โคลโมโกรอฟ – สมินอฟ (Kolmogorov – Smirnov) ซึ่งเป็นวิธีที่จะหลีกเลี่ยงปัญหาการจัดอันตรภาคชั้น (class interval) ของข้อมูล และเป็นวิธีที่ใช้กันมากในการทดสอบการแจกแจงแบบต่อเนื่อง

การทดสอบแบบนี้เป็นการทดสอบว่าข้อมูลตัวอย่างที่ได้มามีการแจกแจงแบบใดแบบหนึ่ง โดยพิจารณาฟังก์ชันการแจกแจงสะสม (cumulative distribution function : cdf) ของตัวอย่าง เปรียบเทียบกับฟังก์ชันการแจกแจงสะสมตามทฤษฎี (cumulative theoretical distribution function) ของแต่ละการแจกแจง ซึ่งการวิจัยในครั้งนี้ต้องการทดสอบการแจกแจงแบบปกติ



รูปที่ 2.10 กราฟเปรียบเทียบฟังก์ชันการแจกแจงสะสมของตัวอย่าง ($S(x)$) และฟังก์ชันการแจกแจงสะสมตามทฤษฎี ($F_0(x)$)

จากรูปที่ 2.8 จะได้ค่า D ซึ่งเป็นค่าความแตกต่างระหว่างค่า $S(x)$ กับค่า $F_0(x)$ ในแต่ละค่าของ x การทดสอบนี้มีตัวสถิติทดสอบคือ

$$D = \max(D^+, D^-)$$

เมื่อ

$$D^+ = \max[\, {}^{i}\!/_{m} , F(x_i)]$$

$$D^- = \max[\, F(x_i) , {}^{(i-1)}\!/_{m}]$$

โดยที่

$F(x_i)$ คือ ความถี่สะสมที่ได้จากทฤษฎี

i คือ ลำดับที่ของข้อมูล ; $i = 1,2,\ldots,m$

m คือ จำนวนข้อมูลที่นำมาทดสอบ

i/m คือ ความถี่สะสมที่ได้จากตัวอย่างลำดับที่ i

$(i-1)/m$ คือ ความถี่สะสมที่ได้จากตัวอย่างลำดับที่ $i-1$

นำค่า D ที่คำนวณได้ไปเปรียบเทียบกับค่าวิกฤตในตารางโคลโมโกรอฟ – สมินอฟ ถ้าค่า D ที่คำนวณได้มากกว่าค่า D ที่ได้จากตาราง จะสรุปได้ว่า การแจกแจงที่นำมาทดสอบไม่เป็นไปตามสมมติฐานที่ตั้งไว้

การแปลงข้อมูล

Single (1988 : 166) กล่าวไว้ว่า สำหรับกลุ่มของตัวเลข 1 กลุ่ม การแปลงข้อมูล จะหมายถึงการแทนที่ค่าของข้อมูลเดิมที่มีอยู่ด้วยค่าอื่น ซึ่งจุดประสงค์ของการแปลงข้อมูลคือ

1. ใช้ในการแสดงลักษณะของข้อมูล
2. เพื่อที่จะให้การแจกแจงนั้นมีรูปแบบที่สมมาตร

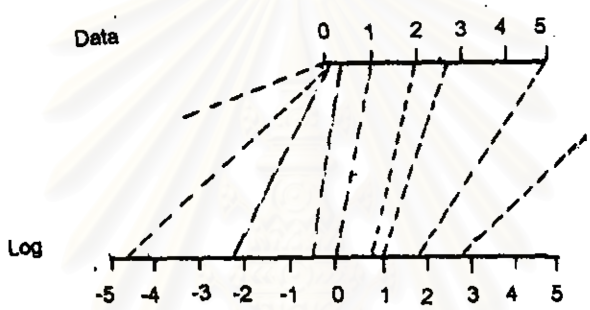
ซึ่งโดยทั่วไป รูปแบบที่สมมาตรที่เราสนใจคือ การแจกแจงแบบปกติ ในการที่จะเปลี่ยนรูปแบบการแจกแจงนั้น การแปลงข้อมูลจะต้องทำมากกว่าการบวกหรือการคูณด้วยค่าคงที่ธรรมดา และจะทำการแปลงข้อมูลเฉพาะที่เป็นค่าบวก โดยส่วนใหญ่แล้ววิธีที่ง่ายที่สุดคือการยกกำลัง เช่นยกกำลัง ½ , ยกกำลัง -1 ซึ่งการยกกำลังด้วยเลขที่เป็นค่าลบนั้นมักจะใช้ในการแปลงข้อมูลจากเล็กให้ใหญ่ขึ้น

วิธีการแปลงโดยใช้ล็อก (Logarithm Transformation) :

เป็นการแปลงข้อมูลโดยใช้ค่าล็อกกาลิทึม โดยอาจเรียกสั้น ๆ ว่า ล็อก สามารถคำนวณได้จากการยึดตัวเลขบวกจำนวนหนึ่งเป็นฐาน (base) ค่าล็อกของข้อมูลเป็นค่าที่เมื่อยกกำลังด้วยค่าฐานแล้วจะได้ค่าของข้อมูลเดิมกลับคืนมา ซึ่งจะแบ่งเป็น 2 ประเภท คือ ล็อกฐาน 10 และล็อกฐาน e เราพบว่า ล็อกฐาน e สามารถหาได้จากล็อกฐาน 10 โดยการคูณ 2.3026...

เข้ากับค่าของล็อกฐาน 10 ซึ่งหมายความว่า เมื่อเราใช้การแปลงทั้งสองแบบนี้จะให้ผลของรูปแบบการแจกแจงที่เหมือนกัน

ความสามารถของการใช้ค่าล็อกในการแปลงข้อมูล : ถ้าค่าของข้อมูลมีค่าน้อยกว่า 0 เมื่อใช้ล็อกในการแปลงข้อมูลจะไม่สามารถทำได้ เพราะค่าล็อกไม่สามารถติดลบได้ สำหรับข้อมูลที่มีค่าอยู่ระหว่าง 0 ถึง 1 เมื่อใช้ล็อกในการแปลงข้อมูลจะได้ค่าที่ติดลบ และมีค่าที่น้อยลงอย่างมากเมื่อค่าของข้อมูลมีค่าเข้าใกล้ 0 ส่วนข้อมูลที่มีค่ามากกว่า 1 นั้นการใช้การแปลงแบบนี้จะทำให้ค่าที่มีค่ามากอยู่นั้นมีค่าลดลงอย่างมาก

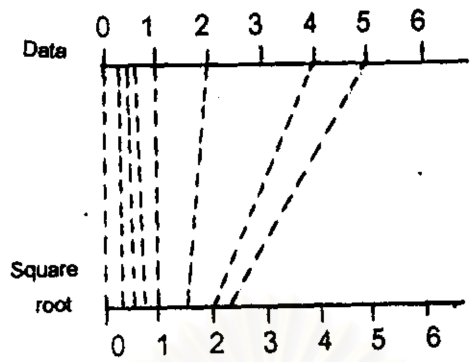


รูปที่ 2.11 แสดงการแปลงโดยใช้ล็อกฐาน e

วิธีการแปลงโดยใช้รากที่สอง (Square Root Transformation) :

เป็นการแปลงโดยใช้ค่ารากที่สอง ซึ่งค่ารากที่สองของข้อมูลเป็นค่าที่ถ้าเรานำค่าของตัวมันเองคูณเข้าไป จะได้ค่าของข้อมูลเดิมกลับคืนมา การหาค่ารากที่สอง จะกระทำได้ด้วยการคิดจากการหารที่ยุ่งยากและซับซ้อน จึงมีผู้ที่เสนอตารางค่ารากที่สอง เพื่อให้ดูง่ายขึ้น

ความสามารถของการใช้ค่ารากที่สองในการแปลงข้อมูล : ถ้าค่าของข้อมูลมีค่าน้อยกว่า 0 จะไม่สามารถใช้การแปลงแบบนี้ได้ สำหรับข้อมูลที่มีค่าอยู่ระหว่าง 0 ถึง 1 นั้นเมื่อใช้การแปลงโดยใช้ค่ารากที่สองแล้วจะทำให้ได้ค่าที่มากขึ้น โดยที่จะมีค่าเข้าใกล้ 1 มากขึ้นเมื่อค่าของข้อมูลเริ่มต้นมีค่าเข้าใกล้ 1 และสำหรับข้อมูลที่มีค่ามากกว่า 1 นั้น จะทำให้ค่าของข้อมูลที่มีค่ามากนั้นมีค่าลดลง



รูปที่ 2.12 แสดงการแปลงโดยใช้ค่ารากที่สอง

วิธีการแปลงโดยใช้การกลับเศษส่วน (Recipocal Transformation) :

เป็นการแปลงข้อมูลโดยใช้วิธีกลับเศษส่วน ซึ่งการกลับเศษส่วนหมายถึงการให้ค่า 1 ถูกหารด้วยค่าของข้อมูล

ความสามารถของการใช้วิธีกลับเศษส่วนในการแปลงข้อมูล : ถ้าค่าของข้อมูลเดิมที่มีอยู่นั้นมีค่าน้อยกว่า 0 แล้ว การแปลงแบบนี้จะทำให้ได้ข้อมูลที่มีค่าเข้าใกล้ 0 มากขึ้นเมื่อค่าของข้อมูลเดิมมีค่าน้อยมาก แต่ถ้าข้อมูลเดิมมีค่ามากกว่า 0 วิธีการแปลงนี้จะทำให้ข้อมูลมีค่าน้อยลงเมื่อข้อมูลเดิมมีค่ามาก

ข้อเสียของการแปลงแบบนี้คือ เมื่อแปลงแล้วลำดับที่ของข้อมูลจะเปลี่ยนไป ถ้าเรียงจากน้อยไปหามากเมื่อเป็นข้อมูลเดิม แต่จะเปลี่ยนเป็นการเรียงจากมากไปหาน้อยเมื่อทำการแปลงข้อมูลด้วยวิธีนี้

ข้อมูล	-2	-1.5	-1	-0.5	0.5	1	1.5	2
การกลับเศษส่วน	-0.5	-0.67	-1	-2	2	1	0.67	0.5

วิธีการแปลงโดยใช้เลขยกกำลัง (Power Transformations) :

เป็นการแปลงด้วยเลขยกกำลังอื่น ๆ เราสามารถที่จะนำเลขจำนวนจริงใด ๆ ที่นอกเหนือจากเลขยกกำลังดังวิธีข้างต้นมายกกำลังเพื่อที่จะทำการแปลงข้อมูลนั้น ๆ ซึ่งจะเห็นได้ว่าการใช้รากที่สองก็คือการยกกำลังด้วยเลข ½ หรือวิธีการกลับเศษส่วนคือการยกกำลังด้วยเลข -1 ดังนั้น เราจึงสามารถที่ใช้เลขยกกำลังตัวอื่น ๆ ในการแปลงข้อมูลได้

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

การแปลงแบบนี้สามารถแบ่งได้เป็น 2 รูปแบบ คือ

1. X^p

2. $\dfrac{X^p - 1}{p}$

เมื่อ p เป็นเลขจำนวนจริงใด ๆ และเมื่อ p มีค่าเท่ากับ 0 เราจะให้การแปลงนี้มีรูปแบบการแปลงแบบล็อกการิทึม รูปแบบการแปลงในแบบแรกจะเป็นการยกกำลังง่าย ๆ แบบทั่วไป ซึ่งภายหลังได้มีการปรับปรุงรูปแบบการแปลงในแบบที่หนึ่งให้เป็นรูปแบบการแปลงในแบบที่สอง โดย BOX & COX (1964) รูปแบบการแปลงในแบบที่สองนี้ถูกพูดถึงในการแก้ปัญหาเกี่ยวกับข้อตกลงเบื้องต้นอื่น ๆ ด้วยนอกเหนือจากเรื่องของการแจกแจง เช่น ข้อตกลงเบื้องต้นเรื่องความแปรปรวน

ความสามารถของการใช้เลขยกกำลังอื่น ๆ ในการแปลงข้อมูล : สำหรับวิธีการแปลงแบบนี้จะใช้เลขยกกำลังที่มีค่าน้อยกว่า 1 กับข้อมูลที่มีการแจกแจงแบบเบ้ขวา และในทางตรงกันข้ามจะใช้เลขยกกำลังที่มีค่ามากกว่า 1 เมื่อข้อมูลมีการแจกแจงแบบเบ้ซ้าย สำหรับการแปลงในรูปแบบที่สอง คือ $\dfrac{X^p - 1}{p}$ ถ้าเลขยกกำลังเป็น 0 ค่าที่แปลงได้จะมีค่าเข้าใกล้ค่าที่ได้จากการแปลงโดยใช้ค่าล็อกฐาน e

รูปแบบของการแปลงข้อมูลทั้งหมดเป็นรูปแบบของเลขยกกำลัง เช่น การยกกำลังด้วย –1 เป็นการแปลงโดยใช้การกลับเศษส่วน การยกกำลังด้วย ½ เป็นการแปลงโดยใช้รากที่สอง ในการวิจัยครั้งนี้แยกรูปแบบการแปลงเหล่านี้ออกจากการแปลงโดยใช้เลขยกกำลัง เนื่องจากรูปแบบต่าง ๆ เป็นรูปแบบเฉพาะ ที่นิยมใช้กันโดยทั่วไป ผู้วิจัยต้องการเปรียบเทียบรูปแบบเหล่านี้กับเลขยกกำลังตัวอื่น ๆ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย