บทที่ 2

ทฤษฎีที่เกี่ยวข้องกับการวิจัย

ในการประมาณค่าแบบช่วง คุณลักษณะของช่วงความเชื่อมั่นที่ผู้วิจัยพึงประสงค์มีสองประการคือ ประการแรกต้องการให้ช่วงที่ประมาณได้คลุมค่าพารามิเตอร์ที่ต้องการประมาณ ด้วยความเชื่อมั่นในระดับที่ผู้วิจัยกำหนด หรือกล่าวว่าต้องเป็นช่วงความเชื่อมั่นที่ให้ค่าระดับความเชื่อมั่นไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด ประการที่สองคือต้องการให้ช่วงที่ประมาณได้เป็นช่วงที่แคบ เพื่อที่จะได้นำไปอธิบายขอบเขตของค่าพารามิเตอร์ได้ดี ดังนั้นในการศึกษาเปรียบเทียบวิธีการประมาณช่วงความเชื่อมั่นสำหรับความแปรปรวนของการแจกแจงแบบปกติทั้ง 3 วิธีในการวิจัยครั้งนี้ จึงพิจารณาค่าระดับความเชื่อมั่น และค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นที่คำนวณได้จากวิธีการประมาณแต่ละวิธี เป็นเกณฑ์ในการตัดสินว่าวิธีการประมาณใดเป็นวิธีการประมาณที่เหมาะสมที่สุด ซึ่งรายละเอียดของการคำนวณค่าระดับความเชื่อมั่น และค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นได้กล่าวไว้ในบทที่ 3 สำหรับในบทนี้จะกล่าวถึงทฤษฎีต่าง ๆ ที่เกี่ยวข้องกับงานวิจัยนี้ ได้แก่ การประมาณค่าพารามิเตอร์แบบช่วง การแจกแจงของค่าความแปรปรวนตัวอย่าง และ รายละเอียดของวิธีการประมาณช่วงความเชื่อมั่นสำหรับความแปรปรวนของการแจกแจงแบบปกติทั้ง 3 วิธี ที่นำมาศึกษาเปรียบเทียบกันคือ วิธีไคสแควร์ วิธีช่วงความเชื่อมั่นที่สั้นที่สุด และวิธีของเบส์ นอกจากนี้จะกล่าวถึง เกณฑ์ที่ใช้ในการเปรียบวิธีการประมาณแบบช่วง การวิเคราะห์ความแปรปรวน ซึ่งในการวิจัยนี้นำมาใช้ในการทดสอบความแตกต่างระหว่างค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นของวิธีการประมาณทั้ง 3 วิธี การแจกแจงแบบปกติ (normal distribution) และการแจกแจงแบบไคสแควร์ (chi-square distribution) เนื่องจากในการประมาณช่วงความเชื่อมั่นสำหรับความแปรปรวนของการแจกแจงแบบปกติ ใช้ค่าความแปรปรวนตัวอย่างเป็นตัวประมาณแบบจุด และการแจกแจงของค่าความแปรปรวนตัวอย่าง มีความสัมพันธ์กับการแจกแจงแบบไคสแควร์ และในตอนท้ายของบทนี้ จะกล่าวถึงผลงานวิจัยที่เกี่ยวข้อง มีรายละเอียดต่าง ๆ ดังนี้

2.1 การประมาณค่าพารามิเตอร์แบบช่วง

ให้ $X_1, X_2, \ldots, X_n$ เป็นตัวอย่างสุ่มจากการแจกแจงซึ่งมี θ เป็นพารามิเตอร์ที่ไม่ทราบค่า

สมมติสามารถหาตัวสถิติ $T_1(X_1,\ldots,X_n)$ และ $T_2(X_1,\ldots,X_n)$ ซึ่งสำหรับค่าจริง θ ใด ๆ

$$Pr(T_1(X_1,\ldots,X_n) < \theta < T_2(X_1,\ldots,X_n)) = 1-\alpha$$

โดยที่ $1-\alpha$ เป็นความน่าจะเป็นคงที่ $(0 < \alpha < 1)$

จากนี้เมื่อทราบค่าของ $X_i = x_i$ $(i = 1,\ldots,n)$ และค่าของ $T_1(X_1,\ldots,X_n)$ และ $T_2(X_1,\ldots,X_n)$ สมมติให้เป็น t_1 และ t_2 ตามลำดับ จะได้ช่วง (t_1,t_2) และเรียกช่วง (t_1,t_2) ว่า "ช่วงความเชื่อมั่น $100(1-\alpha)$ เปอร์เซ็นต์ สำหรับ θ" [$100(1-\alpha)$% confidence interval for θ] หรือกล่าวว่า ค่าจริงของ θ จะตกอยู่ในช่วง (t_1,t_2) ด้วยความเชื่อมั่น $100(1-\alpha)$ เปอร์เซ็นต์ และเรียกค่า t_1 ว่า "ขีดจำกัดล่างของช่วงความเชื่อมั่น" (lower confidence limit) เรียกค่า t_2 ว่า "ขีดจำกัดบนของช่วงความเชื่อมั่น" (upper confidence limit) และเรียกค่า $1-\alpha$ ว่า "สัมประสิทธิ์ความเชื่อมั่น" (confidence coefficient)

2.2 การแจกแจงของค่าความแปรปรวนตัวอย่าง

จากที่กล่าวมาแล้วในหัวข้อที่ 2.1 ว่า เราจะเรียกช่วง (t_1,t_2) ว่าช่วงความเชื่อมั่น $100(1-\alpha)$ เปอร์เซ็นต์สำหรับพารามิเตอร์ θ ก็ต่อเมื่อ t_1 และ t_2 ต้องเป็นค่าที่ทำให้ $Pr(t_1 < \theta < t_2) = 1-\alpha$ และเมื่อใช้ Θ เป็นตัวประมาณแบบจุดสำหรับพารามิเตอร์ θ จะได้ว่า t_1 และ t_2 จะขึ้นอยู่กับ Θ และการแจกแจงความน่าจะเป็นของ Θ ดังนั้นในการประมาณช่วงความเชื่อมั่นสำหรับความแปรปรวนของการแจกแจงแบบปกติ (σ^2) ซึ่งมักใช้ความแปรปรวนตัวอย่าง $S^2 = [\sum_{i=1}^{n}(X_i-\bar{X})^2]/(n-1)$ เป็นตัวประมาณแบบจุดของ σ^2 จึงจะพิจารณาการแจกแจงความน่าจะเป็นของความแปรปรวนตัวอย่าง เพื่อนำไปใช้ประโยชน์ในการประมาณช่วงความ

เชื่อมั่นสำหรับความแปรปรวนของการแจกแจงแบบปกติ ดังนั้นในหัวข้อนี้จะแสดงการหาการแจกแจงความน่าจะเป็นของ S^2 และ $(n-1)S^2/\sigma^2$ ดังต่อไปนี้

เนื่องจาก X_i , i = 1,2,...,n เป็นตัวแปรสุ่มที่มีการแจกแจงเหมือนกันและเป็นอิสระต่อกัน (i.i.d.) ที่มีการแจกแจงแบบ $N(\mu,\sigma^2)$ เพราะฉะนั้น

$$f(x_1,x_2,\ldots,x_n) = \left[\frac{1}{\sqrt{2\pi}\sigma}\right]^n \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right]$$

$$= \left[\frac{1}{\sqrt{2\pi}\sigma}\right]^n \exp\left[-\frac{\sum_{i=1}^{n}(x_i-\bar{X})^2}{2\sigma^2} - \frac{n(\bar{X}-\mu)^2}{2\sigma^2}\right]$$

และ $\bar{X} \sim N(\mu,\sigma^2/n)$

ให้ $Y_1 = \bar{X} = (1/n).(X_1+X_2+\ldots+X_n)$

$Y_2 = X_2$, $Y_3 = X_3$,..., $Y_n = X_n$

ดังนั้น ได้การแปลงผกผัน

$x_1 = ny_1 - y_2 - \ldots - y_n$

$x_2 = y_2$

. .

. .

$x_n = y_n$

และยาโคเบียน (Jacobian) ของการแปลงผกผันคือ J = n และฟังก์ชันความหนาแน่นร่วมของ $Y_1,Y_2,\ldots,Y_n$ คือ

$$f(y_1,y_2,\ldots,y_n) = f_x(ny_1-y_2-\ldots-y_n,y_2,\ldots,y_n)\ .\ |J|$$

$$= n\left[\frac{1}{\sqrt{2\pi}\sigma}\right]^n \exp\left[-\frac{1}{2\sigma^2}\left[(ny_1-y_2-\ldots-y_n-y_1)^2 + \sum_{i=2}^{n}(y_i-y_1)^2 + n(y_1-\mu)^2\right]\right]$$

เนื่องจาก $f(y_1) = \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} \exp\left[-\frac{n}{2\sigma^2}(y_1-\mu)^2\right]$

จะได้ฟังก์ชันความหนาแน่นอย่างมีเงื่อนไขของ $Y_2, Y_3, \ldots, Y_n$ เมื่อกำหนด $Y_1 = y_1$ ดังนี้

$$f(y_2,\ldots,y_n|y_1) = \sqrt{n}\left[\frac{1}{\sqrt{2\pi}\sigma}\right]^{n-1} \exp\left[-\frac{w}{2\sigma^2}\right]$$

เมื่อ $w = (ny_1-y_2-\ldots-y_n-y_1)^2 + \sum_{i=2}^{n}(y_i-y_1)^2$

จาก $(n-1)S^2 = \sum_{i=1}^{n}(X_i-\bar{X})^2$

$$= (nY_1-Y_2-\ldots-Y_n-Y_1)^2 + \sum_{i=2}^{n}(Y_i-Y_1)^2$$

$$= W$$

พิจารณาฟังก์ชันที่ให้โมเมนต์อย่างมีเงื่อนไขของ $(n-1)S^2/\sigma^2 = W/\sigma^2$ เมื่อกำหนด $Y_1 = y_1$ จะได้ว่า

$$E[\exp(tW/\sigma^2)|y_1] = \int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty} \exp(tw/\sigma^2) \,.\, f(y_2,\ldots,y_n|y_1)dy_2\ldots dy_n$$

$$= \int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty} \sqrt{n}\left[\frac{1}{\sqrt{2\pi}\sigma}\right]^{n-1} \exp\left[-\frac{(1-2t)}{2\sigma^2}w\right]dy_2\ldots dy_n$$

$$= \left[\frac{1}{1-2t}\right]^{(n-1)/2} \int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty} \sqrt{n}\left[\frac{1-2t}{2\pi\sigma^2}\right]^{(n-1)/2} .\exp\left[-\frac{(1-2t)w}{2\sigma^2}\right]dy_2\ldots dy_n$$

เมื่อ $1-2t > 0$ หรือ $t < 1/2$ ซึ่งอินทิกรัลสุดท้ายคือ $f(y_2,\ldots,y_n|y_1)$ เมื่อแทน σ^2 ด้วย $\sigma^2/(1-2t)$ และมีค่าเท่ากับ 1

เพราะฉะนั้น $E[\exp(tw/\sigma^2)|y_1] = (1-2t)^{-(n-1)/2}$, $t < 1/2$

ซึ่งเทอมขวามือคือฟังก์ชันที่ให้โมเมนต์ของตัวแปรสุ่ม $\chi^2_{(n-1)}$

จึงกล่าวได้ว่าการแจกแจงของ $(n-1)S^2/\sigma^2$ เมื่อกำหนด $\bar{X} = \bar{x}$ เป็นแบบ $\chi^2_{(n-1)}$ หรือ Gamma$[(n-1)/2,2]$ และจะได้ว่า S^2 มีการแจกแจงแบบ Gamma$[(n-1)/2,2\sigma^2/(n-1)]$ และสามารถเขียนฟังก์ชันความหนาแน่นของ S^2 ได้เป็น

$$f(s^2) = \frac{[(n-1)/2]^{(n-1)/2}[s^2/\sigma^2]^{(n-3)/2} \exp[-(n-1)s^2/2\sigma^2]}{\Gamma[(n-1)/2]\ \sigma^2}$$

โดยที่มีค่าเฉลี่ยและความแปรปรวนดังนี้

$$E(S^2) = \sigma^2$$

$$V(S^2) = 2\sigma^4/(n-1)$$

นอกจากนี้ จะเห็นได้ว่าการแจกแจงอย่างมีเงื่อนไขของ $(n-1)S^2/\sigma^2$ เมื่อกำหนด $\bar{X} = \bar{x}$ ไม่ได้ขึ้นกับ $\bar{x}$ จึงสรุปได้ว่าตัวแปรสุ่ม $\bar{X}$ และ $(n-1)S^2/\sigma^2$ เป็นอิสระต่อกัน และดังนั้น $\bar{X}$ และ S^2 เป็นอิสระต่อกันด้วย

2.3 วิธีการประมาณที่ใช้ในการศึกษา

รายละเอียดที่จะนำเสนอต่อไป เป็นผลงานของนักสถิติในรุ่นก่อน ที่ได้ทำการศึกษาวิธีการประมาณช่วงความเชื่อมั่นสำหรับความแปรปรวนของการแจกแจงแบบปกติ ที่มีรูปแบบของช่วงความเชื่อมั่นอยู่บนพื้นฐานของความแปรปรวนตัวอย่าง (S^2) กล่าวคือ ใช้ความแปรปรวนตัวอย่าง (S^2) เป็นตัวประมาณแบบจุดของความแปรปรวนประชากร (σ^2) 3 วิธี คือวิธีไคสแควร์ วิธีช่วงความเชื่อมั่นที่สั้นที่สุด และวิธีของเบส์ โดยวิธีการประมาณค่าพารามิเตอร์ที่นำมาศึกษานี้สามารถแบ่งแยกได้เป็นสองแนวความคิด ซึ่งในที่นี้จะกล่าวถึงแนวความคิดดังกล่าวพอสังเขปดังนี้

ในทางสถิติแนวความคิดเกี่ยวกับพารามิเตอร์ในการอนุมานเชิงสถิติแบ่งได้เป็นสองกรณี กรณีแรกเรียกว่าสถิติแบบคลาสสิค หรือแบบนอนเบส์เซียน (Classical or NonBayesian Statistics) ส่วนในกรณีที่สอง เรียกว่า สถิติแบบเบส์เซียน (Bayesian Statistics) ซึ่งวิธีการประมาณค่าพารามิเตอร์ที่นำมาเปรียบเทียบในการวิจัยครั้งนี้เป็นการศึกษาทั้งสองแนวความคิด กล่าวคือ การประมาณช่วงความเชื่อมั่นด้วยวิธีไคสแควร์ เป็นสถิติแบบคลาสสิค หรือนอนเบส์เซียน พิจารณาว่า พารามิเตอร์เป็นค่าคงที่ที่ไม่ทราบค่า และต้องการประมาณ ส่วนวิธีช่วงความเชื่อมั่นที่สั้นที่สุด เป็นแนวความคิดของสถิติแบบคลาสสิคเช่นกัน แต่มีการนำวิธีการทางคณิตศาสตร์มาประยุกต์ใช้ เพื่อหาความยาวของช่วงความเชื่อมั่นที่สั้นที่สุด และวิธีสุดท้ายคือวิธีของเบส์ เป็นแนวความคิดของสถิติแบบเบส์เซียน พิจารณาว่าพารามิเตอร์เป็น ตัวแปรสุ่มและมีการแจกแจง

ความน่าจะเป็นของตัวเอง เรียกว่า การแจกแจงแรกเริ่ม (prior distribution) ซึ่งการแรกเริ่มนี้เป็นที่น่าสนใจมาก เนื่องจากในบางกรณีอาจไม่สามารถทราบลักษณะการแจกแจงของพารามิเตอร์ได้ เช่นเหตุการณ์ที่ไม่เคยเกิดขึ้นมาก่อนเลยในอดีต ดังนั้นจึงเกิดปัญหาว่าจะกำหนดการแจกแจงแรกเริ่มอย่างไร ส่วนในกรณีที่เป็นเหตุการณ์ที่เคยเกิดขึ้นมาแล้วในอดีต ก็อาจพอมองเห็นลักษณะการแจกแจงของพารามิเตอร์ และนำมาใช้เป็นการแจกแจงแรกเริ่มได้ สำหรับในการวิจัยครั้งนี้ สนใจศึกษาในเหตุการณ์ที่ไม่เคยเกิดขึ้นเลยในอดีต หรืออาจเคยเกิดขึ้นแต่ผู้วิจัยอาจไม่มีข้อมูลเพียงพอที่จะนำมากำหนดการแจกแจงแรกเริ่มได้ (noninformative prior distribution) ซึ่งจากผลงานของนักสถิติที่นำมาศึกษาเสนอว่า เมื่อไม่สามารถทราบการแจกแจงแรกเริ่มได้ วิธีการที่ควรกระทำคือกำหนดให้การแจกแจงแรกเริ่ม มีการแจกแจงแบบสม่ำเสมอ (uniform distribution) ซึ่งนักสถิติที่ยึดมั่นในแนวความคิดนี้เชื่อว่า การกำหนดการแจกแจงแรกเริ่ม โดยไม่มีกฎเกณฑ์แน่นอนเช่นนี้ จะไม่ก่อให้เกิดปัญหาเกี่ยวกับคุณภาพของการประมาณแต่อย่างใด ทั้งนี้เพราะในการประมาณค่าพารามิเตอร์แบบช่วงในสถิติแบบเบส์เซียน อาศัยการแจกแจงที่มีเงื่อนไข ของพารามิเตอร์เมื่อกำหนดข้อมูลตัวอย่าง หรือที่เรียกว่าการแจกแจงภายหลัง (posterior distribution) เป็นสิ่งสำคัญในการประมาณ กล่าวคือถ้าต้องการประมาณค่าแบบช่วงของพารามิเตอร์ θ ก็คือการหาค่า $Pr(T_1 < \theta < T_2 \mid \underset{\sim}{x}) = 1-\alpha$ ซึ่งจะต้องทราบฟังก์ชันความหนาแน่นแบบมีเงื่อนไขของ Θ เมื่อกำหนด $\underset{\sim}{x}$ หรือที่เรียกว่า ฟังก์ชันความหนาแน่นภายหลัง จึงจะสามารถหาช่วงความเชื่อมั่นดังกล่าวได้ ซึ่งรายละเอียดต่าง ๆ ได้กล่าวไว้ในหัวข้อที่ 2.3.3 สำหรับในหัวข้อต่อไปจะกล่าวถึงรายละเอียดของวิธีการประมาณทั้ง 3 วิธีดังต่อไปนี้

2.3.1 <u>วิธีไคสแควร์</u>

จากที่กล่าวมาแล้วในหัวข้อที่ 2.2 จะได้ว่า rS^2/σ^2 มีการแจกแจงแบบ χ^2 ด้วยองศาอิสระ r ดังนั้น สำหรับค่าคงที่ a และ b

$$\Pr(a < rS^2/\sigma^2 < b) = 1-\alpha$$

หรือ

$$\Pr(rS^2/b < \sigma^2 < rS^2/a) = 1-\alpha$$

การคำนวณช่วงความเชื่อมั่นสำหรับ σ^2 ที่ใช้กันอยู่โดยทั่วไป กำหนดให้

$$\Pr(rS^2/\sigma^2 < a) = \Pr(rS^2/\sigma^2 > b) = \alpha/2$$

หรือ

$$\Pr(\chi^2_r < a) = \Pr(\chi^2_r > b) = \alpha/2$$

ดังนั้น ตัวหารสำหรับช่วงความเชื่อมั่นโดยวิธีไคสแควร์ (a,b) ถูกกำหนดโดย

$$\int_b^{\infty} f(x)\,dx = \alpha/2 \qquad (2.3.1.1)$$

และ

$$\int_a^{b} f(x)\,dx = F(b) - F(a) = 1-\alpha \qquad (2.3.1.2)$$

ให้ a และ b ที่กำหนดโดยสมการ 2.3.1.1 และ 2.3.1.2 แทนด้วย a_c และ b_c ตามลำดับ

ดังนั้นช่วงความเชื่อมั่น $(1-\alpha)100\%$ สำหรับ σ^2 คือ $[rs^2/b_c, rs^2/a_c]$

เมื่อ $f(x)$ คือ ฟังก์ชันความหนาแน่นของการแจกแจง χ^2_r

$$f(x) = \frac{x^{(r/2)-1}\exp(-x/2)}{\Gamma(r/2)\,2^{r/2}}$$

r คือ องศาอิสระ เท่ากับ n-1

s^2 คือ ความแปรปรวนตัวอย่าง

a_c, b_c คือ ตัวหารสำหรับช่วงความเชื่อมั่นโดยวิธีไคสแควร์

โดยที่ a_c เท่ากับ $\chi^2_{r,1-\alpha/2}$

b_c เท่ากับ $\chi^2_{r,\alpha/2}$

2.3.2 วิธีช่วงความเชื่อมั่นที่สั้นที่สุด

ในปี ค.ศ. 1959 เทต (Tate, R. F.) และ เคล็ท (Klett, G. W.) เสนอให้ใช้วิธีการทางคณิตศาสตร์ เพื่อหาความยาวที่สั้นที่สุดของช่วงความเชื่อมั่นในรูปแบบ

$[rs^2/b, rs^2/a]$

ให้ L คือ ความยาวของช่วงความเชื่อมั่น ดังนั้น

$$L = rs^2(1/a - 1/b)$$

โดย วิธีลากรองจ์ (Lagrange Method) หาความยาวที่สั้นที่สุดของ L

ภายใต้เงื่อนไข $\Pr(a < rS^2/\sigma^2 < b) = 1-\alpha$

ดังนั้น ให้ $\phi = rs^2(1/a - 1/b) + k[F(b) - F(a) - (1-\alpha)]$

หาอนุพันธ์ของ ϕ เทียบกับ a,b,k จะได้

$$\partial\phi/\partial a = -rs^2/a^2 - kf(a) = 0$$

$$\partial\phi/\partial b = rs^2/b^2 + kf(b) = 0$$

$$\partial\phi/\partial k = F(b) - F(a) - (1-\alpha) = 0$$

จะได้ว่า ตัวหารสำหรับช่วงความเชื่อมั่นโดยวิธีช่วงความเชื่อมั่นที่สั้นที่สุด(a,b) ถูกกำหนดโดย

$$a^2f(a) = b^2f(b)$$

หรือ

$$\frac{a^{(r+2)/2}e^{(-a/2)} - b^{(r+2)/2}e^{(-b/2)}}{[2^{(r/2)}\Gamma(r/2)]} = 0 \qquad (2.3.2.1)$$

$$F(b) - F(a) = 1-\alpha \qquad (2.3.2.2)$$

ให้ a และ b ที่กำหนดโดย สมการเงื่อนไขที่ 2.3.2.1 และ 2.3.2.2 แทนด้วย a_{ML} และ b_{ML} ตามลำดับ จากนั้นแก้สมการหาค่า a_{ML} และ b_{ML} ด้วยวิธีการวิเคราะห์เชิงตัวเลข (Numerical Analysis) โดยใช้วิธีของนิวตัน - ราฟสัน (Newton - Raphson Method) ซึ่งขั้นตอนต่างๆ ได้แสดงไว้ในบทที่ 3 หัวข้อที่ 3.2.1 ดังนั้น ช่วงความเชื่อมั่น $(1-\alpha)100\%$ สำหรับ σ^2 คือ $[rs^2/b_{ML}, rs^2/a_{ML}]$

เมื่อ
- r คือ องศาอิสระ
- s^2 คือ ความแปรปรวนตัวอย่าง
- a_{ML}, b_{ML} คือ ตัวหารสำหรับช่วงความเชื่อมั่นที่สั้นที่สุด
- k คือ ตัวคูณลากรองจ์ (Lagrange Multiplier)

2.3.3 วิธีของเบส์

จากที่ได้กล่าวมาข้างต้นว่า แนวความคิดเกี่ยวกับพารามิเตอร์ แบ่งออกได้เป็นสองแนวทาง สำหรับในการประมาณช่วงความเชื่อมั่นสำหรับความแปรปรวนของการแจกแจงแบบปกติ โดยวิธีของเบส์ เป็นสถิติแบบเบส์เซียน ซึ่งการศึกษาในแนวทางนี้ถือว่าการแจกแจงแรกเริ่ม (prior distribution) และการแจกแจงภายหลัง (posterior distribution) มีความสำคัญมาก มีรายละเอียดต่าง ๆ ดังต่อไปนี้

เมื่อการแจกแจงของตัวแปรสุ่ม X ขึ้นอยู่กับค่าของพารามิเตอร์ θ ซึ่งโดยทั่วไปจะไม่ทราบค่าจริงของ θ และเป็นไปได้ว่า พารามิเตอร์จะมีการแจกแจงความน่าจะเป็น นั่นคือ θ เป็นค่าของตัวแปรสุ่ม Θ ซึ่งอาจเรียก Θ ว่า พารามิเตอร์สุ่ม (random parameter) และมีปริภูมิพารามิเตอร์(parameter space) R ดังนั้น สำหรับตัวอย่างสุ่ม $X_1,X_2,\ldots,X_n$ จาก X จะมีฟังก์ชันความหนาแน่นร่วมแบบมีเงื่อนไข (conditional joint density function) $h(\underset{\sim}{x} \mid \theta)$ และฟังก์ชันความหนาแน่นร่วม (joint density function) ของ $\underset{\sim}{X}$ และ Θ เขียนในรูปทั่วไปได้เป็น

$$f(\underset{\sim}{x},\theta) = g(\theta)h(\underset{\sim}{x}\mid\theta)$$

ซึ่ง $g(\theta)$ คือฟังก์ชันความหนาแน่นของ Θ

ฟังก์ชันความหนาแน่นเดี่ยว (marginal density function) ของ $\underset{\sim}{X}$ หาได้จาก

$$h(\underset{\sim}{x}) = \int g(\theta)h(\underset{\sim}{x}\mid\theta)\, d\theta$$ เมื่อ Θ เป็นตัวแปรสุ่มที่ต่อเนื่อง

นอกจากนี้ ฟังก์ชันความหนาแน่นแบบมีเงื่อนไขของ Θ เมื่อกำหนด $X_1=x_1$, $X_2=x_2,\ldots,X_n=x_n$ แทนด้วยสัญลักษณ์ $g(\theta\mid\underset{\sim}{x})$ และได้ว่า

$$g(\theta\mid\underset{\sim}{x}) = \frac{g(\theta)h(\underset{\sim}{x}\mid\theta)}{h(\underset{\sim}{x})} \qquad \theta \in R$$

เรียก ฟังก์ชันความหนาแน่น $g(\theta)$ ว่า ฟังก์ชันความหนาแน่นแรกเริ่ม (prior density function) และเรียกฟังก์ชันความหนาแน่นแบบมีเงื่อนไข $g(\theta\mid\underset{\sim}{x})$ ของ Θ เมื่อทราบค่าของ $\underset{\sim}{X}$ ว่า ฟังก์ชันความหนาแน่นภายหลัง (posterior density function) และสามารถเขียนได้

ในรูปแบบ

$$g(\theta|\underset{\sim}{x}) \quad \propto \quad h(\underset{\sim}{x}|\theta)g(\theta)$$

เมื่อกำหนด α จะได้ ช่วงความเชื่อมั่นโดยวิธีของเบส์ (Bayesian confi dence interval) คือ ช่วง(t_1,t_2) ที่ทำให้

$$\int_{t_1}^{t_2} g(\theta|\underset{\sim}{x})d(\theta) = 1-\alpha$$

ในปี ค.ศ.1960 ลีนเลย์ (Lindley, D. V.) อีส (East, D. A.)และ แฮมมิลตัน (Hamilton, P. A.) ได้เสนอตารางเกี่ยวกับการอนุมานค่าความแปรปรวนของการแจกแจงแบบปกติ โดยใช้หลักการของเบส์ โดยทำการศึกษาในกรณีที่ทราบค่าเฉลี่ยของประชากร และผู้วิจัยไม่มีข้อมูลเกี่ยวกับพารามิเตอร์ที่ทำการศึกษา ซึ่งในที่นี้คือ ความแปรปรวนของการแจกแจงแบบปกติ (σ^2) จึงเป็นการยุ่งยากในการกำหนดการแจกแจงแรกเริ่ม เขาได้กล่าวถึงทางเลือกของการแจกแจงแรกเริ่มสำหรับในกรณีที่มีพารามิเตอร์ที่ไม่ทราบค่าเพียงค่าเดียว (single parameter) ว่า เมื่อ Θ เป็นพารามิเตอร์ที่ต้องการศึกษา และ $\phi(\Theta)$ เป็นการแปลงแบบหนึ่งต่อหนึ่ง (one-to-one transformation) ของ θ จะได้ว่าการแจกแจงแรกเริ่มของ Θ เป็นสัดส่วนกับ $|d\phi(\theta)/d\Theta|$ โดยที่ $\phi(\theta)$ มีการแจกแจงแบบสม่ำเสมอ

ในกรณีของการประมาณค่าความแปรปรวนของการแจกแจงแบบปกติ (σ^2) ลีนเลย์ทำการศึกษาในกรณีทราบค่าเฉลี่ยของประชากร (μ) เขาเสนอให้ $\phi(\sigma^2) = \ln\sigma^2$ มีการแจกแจงแบบสม่ำเสมอในช่วง $-\infty$ ถึง ∞ ดังนั้นการแจกแจงแรกเริ่มของ σ^2 เป็นสัดส่วนกับ σ^{-2} กล่าวคือ

$$g(\sigma^2) \quad \propto \quad |d\ln\sigma^2/d\sigma^2| \quad = \quad \sigma^{-2}$$

ดังนั้นการแจกแจงภายหลังของ σ^2 กำหนดโดย

$$g(\sigma^2|\underset{\sim}{x}) \quad \propto \quad h(\underset{\sim}{x}|\sigma^2)g(\sigma^2)$$

$$h(\underset{\sim}{x}|\sigma^2) \quad \propto \quad \sigma^{-n} \;.\; \exp[-\sum_{i=1}^{n}(x_i-\mu)^2/2\sigma^2]$$

ให้ $s^2 = (1/n)\sum_{i=1}^{n}(x_i-\mu)^2$ คือ ความแปรปรวนตัวอย่างของ σ^2

ดังนั้น $$g(\sigma^2|\underset{\sim}{x}) \quad \propto \quad \sigma^{-n} \;.\; \sigma^{-2} \;.\; \exp[-ns^2/2\sigma^2]$$

$$\propto \quad \sigma^{-(n+2)} \cdot \exp[-ns^2/2\sigma^2]$$

$$= \quad k \cdot \sigma^{-(n+2)} \cdot \exp[-ns^2/2\sigma^2]$$

อาศัยคุณสมบัติของฟังก์ชันความหนาแน่น จะได้

$$k \quad = \quad [\Gamma(n/2)]^{-1} \cdot [ns^2/2]^{n/2}$$

ดังนั้น $$g(\sigma^2|\underset{\sim}{x}) = \frac{(ns^2/2)^{n/2} \cdot (\sigma^2)^{-[(n/2)+1]} \cdot \exp[-ns^2/2\sigma^2]}{\Gamma(n/2)} \qquad (2.3.3.1)$$

การประมาณช่วงความเชื่อมั่นสำหรับ σ^2 ลีนเลย์ใช้ช่วงความเชื่อมั่นแบบ H.P.D.[1] (highest posterior density interval) ซึ่งกำหนดโดย

a) $g(\sigma^2_0|\underset{\sim}{x}) = g(\sigma^2_1|\underset{\sim}{x})$

b) $Pr(\sigma^2_0 < \sigma^2 < \sigma^2_1|\underset{\sim}{x}) = 1-\alpha$

แต่เนื่องจากในการประมาณช่วงความเชื่อมั่นแบบ H.P.D. กำหนดให้คำนวณจากการแจกแจงแรกเริ่มที่มีการแจกแจงแบบสม่ำเสมอ ดังนั้นในการประมาณช่วงความเชื่อมั่นสำหรับ σ^2 ลีนเลย์จึงคำนวณโดยวิธีการประมาณช่วงแบบ H.P.D. สำหรับ $\ln\sigma^2$ แทน เนื่องจาก $\ln\sigma^2$ มีการแจกแจงแบบสม่ำเสมอ ซึ่งสามารถทำได้เพราะการประมาณช่วงความเชื่อมั่นแบบ H.P.D. มีคุณสมบัติไม่เปลี่ยนแปลงภายใต้การแปลงเชิงเส้น (invariant under linear transformation)[2] ดังนั้นสำหรับเงื่อนไข a) จึงถูกแทนด้วย $g(\ln\sigma^2_0|\underset{\sim}{x}) = g(\ln\sigma^2_1|\underset{\sim}{x})$ ซึ่งการแจกแจงภายหลังของ $\ln\sigma^2$ หาได้จาก $g(\sigma^2|\underset{\sim}{x})$ ดังนี้

จากสมการที่ 2.3.3.1 โดยการแปลง จะได้

$$g(\ln\sigma^2|\underset{\sim}{x}) = \frac{(ns^2/\sigma^2)^{n/2} \cdot \exp(-ns^2/2\sigma^2)}{[\Gamma(n/2)2^{n/2}]}$$

$$= \frac{\exp[(n/2)(\ln n+\ln s^2-\ln\sigma^2)-(1/2)\exp(\ln n+\ln s^2-\ln\sigma^2)]}{[\Gamma(n/2)2^{n/2}]} \qquad -\infty < \ln\sigma^2 < \infty \qquad (2.3.3.2)$$

---

[1] Box, G.E.P. and Tiao, G.C. Bayesian Inference in Statistical Analysis (Mass: Addison-Wesley, 1973), p. 89.

[2] Ibid., pp. 122-123.

ซึ่งถ้าพิจารณาเปรียบเทียบ สมการที่ 2.3.3.1 กับการแจกแจงแบบอินเวอรท์ไคสแควร์ (χ^{-2}_n distribution) จะสามารถเขียนได้ว่า

$$\sigma^2 \sim (ns^2)\chi^{-2}_n$$

เมื่อ $$f(\chi^{-2}_n) = \frac{(\chi^{-2}_n)^{-[(n/2)+1]}\exp(-1/2\chi^{-2}_n)}{[\Gamma(n/2)2^{n/2}]}, \quad \chi^{-2}>0 \qquad (2.3.3.3)$$

ดังนั้น ในการประมาณช่วงความเชื่อมั่นสำหรับ σ^2 โดยวิธีการของลียเลย์ คือการประมาณช่วงความเชื่อมั่นแบบ H.P.D. ของ $\ln\sigma^2$ ซึ่งในการวิจัยครั้งนี้ เรียกช่วงความเชื่อมั่นชนิดนี้ว่า ช่วงความเชื่อมั่นสำหรับ σ^2 โดยวิธีของเบส์ จึงเกิดจาก

a) $$\frac{a^{n/2}\, e^{-a/2}}{\Gamma(n/2)2^{n/2}} = \frac{b^{n/2}\, e^{-b/2}}{\Gamma(n/2)2^{n/2}}$$

หรือ $$\frac{a^{n/2}\, e^{-a/2} - b^{n/2}\, e^{-b/2}}{\Gamma(n/2)2^{n/2}} = 0 \qquad (2.3.3.4)$$

และ b) $$\Pr(a < \chi^2_n < b) = 1-\alpha$$

$$F(b) - F(a) = 1-\alpha \qquad (2.3.3.5)$$

ให้ a และ b ที่กำหนดโดยสมการที่ 2.3.3.4 และ 2.3.3.5 แทนด้วย a_B และ b_B ตามลำดับ จากนั้นแก้สมการหาค่า a_B และ b_B ด้วยวิธีการวิเคราะห์เชิงตัวเลข (Numerical Analysis) โดยใช้วิธีของนิวตัน - ราฟสัน (Newton - Raphson Method) ซึ่งขั้นตอนต่างๆ ได้แสดงไว้ในบทที่ 3 หัวข้อที่ 3.2.1 ดังนั้นช่วงความเชื่อมั่น $(1-\alpha)100\%$ สำหรับ σ^2 ในกรณีที่ทราบค่าเฉลี่ยประชากร (μ) คือ $[ns^2/b_B, ns^2/a_B]$

สำหรับในการวิจัยครั้งนี้ศึกษาในกรณีที่ไม่ทราบค่าเฉลี่ยของประชากร จะประมาณค่าเฉลี่ยประชากร (μ) ด้วยค่าเฉลี่ยตัวอย่าง $\bar{X}$ และ $s^2 = [\sum_{i=1}^{n}(x_i-\bar{X})^2]/(n-1)$ จากนั้นใช้ค่าตัวหารสำหรับช่วงความเชื่อมั่นโดยวิธีของเบส์ โดยการลดองศาอิสระลง 1 ดังนั้นช่วงความเชื่อมั่น $(1-\alpha)100\%$ สำหรับ σ^2 ที่ใช้ในการวิจัยนี้ คือ $[rs^2/b_B, rs^2/a_B]$

เมื่อ r คือ องศาอิสระ เท่ากับ n-1

s^2 คือ ความแปรปรวนตัวอย่าง

a_B, b_B คือ ตัวหารสำหรับช่วงความเชื่อมั่นโดยวิธีของเบส์

2.4 เกณฑ์ที่ใช้ในการเปรียบเทียบวิธีการประมาณแบบช่วง

การเปรียบเทียบวิธีการประมาณค่าแบบช่วงสำหรับความแปรปรวนของการแจกแจงแบบปกติทั้ง 3 วิธี จะทำการเปรียบเทียบค่าระดับความเชื่อมั่น และค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นที่คำนวณได้จากแต่ละวิธี ที่ระดับความเชื่อมั่น 4 ระดับคือ 90%, 95%, 99% และ 99.5% ในแต่ละสถานการณ์ของการทดลอง ทำการทดลองซ้ำ 2,000 ครั้ง โดยในการเปรียบเทียบค่าระดับความเชื่อมั่น จะเปรียบเทียบค่าระดับความเชื่อมั่นกับค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด ซึ่งในการตรวจสอบว่าวิธีการประมาณใด ให้ค่าระดับความเชื่อมั่นไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดหรือไม่นั้น จะอาศัยการทดสอบสมมติฐานโดยใช้ตัวสถิติ z ดังนี้

$$-z_{\alpha/2} < \frac{\hat{P}-p}{\sqrt{p(1-p)/n}} < z_{\alpha/2}$$

$$p-z_{\alpha/2}\sqrt{p(1-p)/n} < \hat{P} < p+z_{\alpha/2}\sqrt{p(1-p)/n}$$

จะได้ $(p-z_{\alpha/2}\sqrt{p(1-p)/n}\,,\ p+z_{\alpha/2}\sqrt{p(1-p)/n})$

1. ที่ระดับความเชื่อมั่น 90%

   $H_0 : p = 0.90$

   $H_1 : p = 0.90$

   จะได้ $(0.90-1.645\sqrt{0.90(0.10)/2000}\ ,\ 0.90+1.645\sqrt{0.90(0.10)/2000})$

   $(0.8890\ ,\ 0.9110)$

2. ที่ระดับความเชื่อมั่น 95%

   $H_0 : p = 0.95$

   $H_1 : p = 0.95$

   จะได้ $(0.95-1.96\sqrt{0.95(0.05)/2000}\ ,\ 0.95+1.96\sqrt{0.95(0.05)/2000})$

   $(0.9405\ ,\ 0.9596)$

3. ที่ระดับความเชื่อมั่น 99%

   $H_0 : p = 0.99$

   $H_1 : p = 0.99$

จะได้ $(0.99-2.576\sqrt{0.99(0.01)/2000}\ ,\ 0.99+2.576\sqrt{0.99(0.01)/2000})$

$(0.9843\ ,\ 0.9957)$

4. ที่ระดับความเชื่อมั่น 99.5%

$H_o : p = 0.995$

$H_1 : p = 0.995$

จะได้ $(0.995-2.81\sqrt{0.995(0.005)/2000}\ ,\ 0.995+2.81\sqrt{0.995(0.005)/2000})$

$(0.9906\ ,\ 0.9994)$

นั่นคือ ที่ระดับความเชื่อมั่น 90%, 95%, 99% และ 99.5% ถ้าวิธีการประมาณใดให้ค่าระดับความเชื่อมั่นไม่ต่ำกว่าค่า 0.8890, 0.9405, 0.9843 และ 0.9906 ตามลำดับ จะถือว่าวิธีการประมาณนั้นให้ค่าระดับความเชื่อมั่นไม่ต่ำกว่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดในสถานการณ์นั้น ๆ จากนั้นจึงมาพิจารณาเปรียบเทียบค่าความยาวเฉลี่ยของช่วงความเชื่อมั่น ถ้าวิธีการประมาณใดให้ค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นต่ำที่สุด จะถือเป็นวิธีการประมาณที่เหมาะสมที่สุด ทั้งนี้ ในการเปรียบเทียบค่าความยาวเฉลี่ยของช่วงความเชื่อมั่น จะเปรียบเทียบเฉพาะในกรณีที่วิธีการประมาณให้ค่าระดับความเชื่อมั่นไม่ต่ำกว่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดเท่านั้น

2.5 <u>การวิเคราะห์ความแปรปรวน</u> (Analysis of Variance)

ในกรณีที่ผู้วิจัยต้องการเปรียบเทียบความแตกต่างระหว่างค่าเฉลี่ยหลาย ๆ ค่า (3 ค่าขึ้นไป) การทดสอบโดยใช้ตัวสถิติ z หรือ t โดยทำการทดสอบทีละคู่ จะทำให้เกิดความคลาดเคลื่อน วิธีการที่เหมาะสมสำหรับทดสอบค่าเฉลี่ยหลาย ๆ ค่า คือ การวิเคราะห์ความแปรปรวน (Analysis of Variance) ซึ่งในที่นี้จะกล่าวถึง การวิเคราะห์ความแปรปรวนทางเดียว (One-way Analysis of Variance) ในกรณีที่แต่ละประชากร (วิธีการประมาณ) มีจำนวนค่าสังเกตเท่ากัน ซึ่งถูกนำมาใช้ในการทดสอบความแตกต่างระหว่างค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นของวิธีการประมาณทั้ง 3 วิธี มีรายละเอียดดังนี้

การทดสอบความแตกต่างระหว่างค่าเฉลี่ยหลาย ๆ ค่ามีสมมติฐานหลัก และสมมติฐานทางเลือก คือ $H_o : \mu_1 = \mu_2 = \ldots = \mu_6$ และ H_1 : มีค่าเฉลี่ยอย่างน้อยสองค่าที่แตกต่าง

กัน ซึ่งลักษณะของข้อมูล เมื่อ t เป็นจำนวนวิธีการประมาณเท่ากับ 3 และ r เป็นจำนวนค่าสังเกตในแต่ละวิธีการประมาณเท่ากับ 2000 มีลักษณะดังนี้

วิธีการประมาณ

	1	2	....	t	
	y_{11}	y_{21}		y_{t1}	
	y_{12}	y_{22}		y_{t2}	
	.	.		.	
	.	.		.	
	y_{1r}	y_{2r}		y_{tr}	
รวม	$y_{1.}$	$y_{2.}$	$y_{3.}$	$y_{t.}$	$y_{..}$
ค่าเฉลี่ย	$\bar{y}_{1.}$	$\bar{y}_{2.}$	$\bar{y}_{3.}$	$\bar{y}_{t.}$	$\bar{y}_{..}$

โดย y_{ij} แทนค่าความยาวของช่วงความเชื่อมั่นของวิธีการประมาณที่ i ค่าสังเกตที่ j

$y_{i.}$ แทนผลรวมของค่าความยาวของช่วงความเชื่อมั่นของวิธีการประมาณที่ i = $\sum_{j=1}^{r} y_{ij}$

$\bar{y}_{i.}$ เป็นค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นของวิธีการประมาณที่ i = $y_{i.}/r$

$y_{..}$ เป็นผลรวมของทุกค่าความยาวเฉลี่ยของช่วงความเชื่อมั่น = $\sum\sum_{i,j=1}^{t\,r} y_{ij}$

$\bar{y}_{..}$ เป็นค่าเฉลี่ยรวม = $y_{..}/tr$

การวิเคราะห์ความแปรปรวนทางเดียว ทำได้โดยการคำนวณค่าต่าง ๆ ดังนี้

ผลบวกกำลังสองของผลรวม (Total Sum Square) = SST = $\sum\sum_{i,j=1}^{t\,r} y_{ij}^2 - y_{..}^2/tr$

ผลบวกกำลังสองระหว่างประชากร = (Sum Square Between Groups) = SSB

$$= (\sum_{i=1}^{t} y_{i.}^2)/r - y_{..}^2/tr$$

ผลบวกกำลังสองภายในประชากรเดียวกัน (Sum Square Within Groups) = SSW

$$= SST - SSB$$

ค่าเฉลี่ยกำลังสองระหว่างประชากร (Mean Square Between Groups) = MSB

= SSB/t-1

ค่าเฉลี่ยกำลังสองภายในประชากรเดียวกัน (Mean Square Within Groups) = MSW

= SSW/t(r-1)

ตัวสถิติที่ใช้ในการทดสอบคือ F = MSB/MSW จากนั้นเปรียบเทียบค่า F ที่คำนวณได้กับค่า F จากตารางที่ระดับนัยสำคัญ α องศาอิสระเป็น (t-1) กับ t(r-1) โดยจะปฏิเสธสมมติฐานหลัก เมื่อ $F > F_{\alpha(t-1,t(r-1))}$

การวิเคราะห์ความแปรปรวน ในกรณีที่ได้ข้อสรุปว่า ปฏิเสธสมมติฐานหลัก นั่นคือ มีค่าเฉลี่ยอย่างน้อยสองค่าแตกต่างกัน ดังนั้นผู้ทดลองสามารถทำการทดสอบความแตกต่างระหว่างค่าเฉลี่ยของสองประชากรภายหลังการวิเคราะห์ความแปรปรวน เพื่อให้ทราบว่าค่าเฉลี่ยของประชากรคู่ใดบ้างที่แตกต่างกัน โดยทำการทดสอบที่เรียกว่า การเปรียบเทียบเชิงซ้อน (Multiple Comparison) ซึ่งการเปรียบเทียบเชิงซ้อนดังกล่าวสามารถทำได้หลายวิธี เช่น วิธีผลต่างน้อยที่สุด (Least Significant Difference) วิธีของทูกี้(Tukey's W-procedure) วิธีของดันแคน (Dancan's New Multiple Range Test) เป็นต้น สำหรับในการวิจัยนี้จะกล่าวถึงวิธีผลต่างน้อยที่สุด ซึ่งเป็นวิธีที่ง่ายและสะดวกในการคำนวณและถูกนำมาใช้ในการวิจัยครั้งนี้ โดยมีรายละเอียดดังนี้

วิธีผลต่างน้อยที่สุด เป็นการเปรียบเทียบความแตกต่างระหว่างค่าเฉลี่ยของประชากรเป็นคู่ ๆ โดยการกำหนดค่าน้อยที่สุดขึ้นมาค่าหนึ่ง เพื่อเป็นตัวเปรียบเทียบว่าความแตกต่างของค่าเฉลี่ยของประชากรคู่ใด ๆ มีค่าเกินกว่าค่าที่กำหนดจะถือว่า ประชากรคู่นั้นแตกต่างกันอย่างมีนัยสำคัญ ค่าที่กำหนดขึ้นนี้เรียกว่า ค่า lsd มีค่าเท่ากับ

$$lsd = t_{\alpha/2}\sqrt{2MSW/r}$$

โดยที่ $t_{\alpha/2}$ คือ ค่า t จากตาราง t ที่มีองศาอิสระเท่ากับ t(r-1)

r คือ จำนวนค่าสังเกตในแต่ละประชากร

2.6 การแจกแจงแบบปกติ (Normal Distribution)

2.6.1 การแจกแจงแบบปกติ อาจกล่าวได้ว่า เป็นการแจกแจงที่สำคัญที่สุดในวิชาสถิติ และเป็นการแจกแจงซึ่งใช้อธิบายข้อมูลที่เกิดขึ้นโดยทั่วไปทางธรรมชาติ เช่น น้ำหนัก ความสูง เป็นต้น ค้นพบโดย เดอมัวฟ์ (De Movire:1667-1754) เป็นนักคณิตศาสตร์ชาวฝรั่งเศส ต่อมา ลาปลาซ (Lapace:1749-1827) นักคณิตศาสตร์ชาวฝรั่งเศสได้นำมาประยุกต์ใช้ในทางด้านสังคมศาสตร์ และวิทยาศาสตร์ อย่างแพร่หลาย หลังจากนั้นนักคณิตศาสตร์ชาวเยอรมัน คือ เกาส์ (Gauss:1777-1855) ได้ขยายข่ายงานประยุกต์ออกไปอีก โดยนำทฤษฎีนี้ไปศึกษาหาความคลาดเคลื่อน (Error) โดยการวัดซ้ำ ๆ ในกลุ่มที่มีขนาดคงเดิม และพบว่าการแจกแจงที่ได้จะมีลักษณะเป็นโค้งปกติ ดังนั้นการแจกแจงแบบปกติ จึงอาจเรียกอีกอย่างหนึ่งว่า Gaussian Distribution

ฟังก์ชันความหนาแน่น (pdf.) ของการแจกแจงแบบปกติ คือ

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} .\exp[-(x-\mu)^2/2\sigma^2] \; ; \; -\infty < x < \infty$$

โดยที่ $f(x)$ = แทนความสูงของโค้งที่วัดจากแกนนอน ณ จุดใด ๆ ทุกจุด (x)

σ = ความเบี่ยงเบนมาตรฐานของประชากร

σ^2 = ความแปรปรวนของประชากร

μ = ค่าเฉลี่ยของประชากร

π = 3.14159

e = 2.71828

x = ค่าของข้อมูลจากกลุ่มตัวอย่าง

และค่า μ , σ เป็นพารามิเตอร์ที่บอกถึงลักษณะของประชากรว่าประชากรนั้นมีตำแหน่งอยู่ที่ใด และมีการกระจายมากน้อยเพียงใด

2.6.2 การแจกแจงแบบปกติมีลักษณะและคุณสมบัติดังนี้

2.6.2.1 ลักษณะของโค้งเป็นรูประฆังคว่ำ (Bell Shaped)

2.6.2.2 เส้นแบ่งครึ่งโค้งอยู่ที่จุดที่เป็นค่าเฉลี่ยของข้อมูล และเส้นนี้ทำให้เส้นโค้งที่อยู่สองข้าง มีลักษณะสมมาตร (Symmetry)

2.6.2.3 จุดที่เป็นค่าเฉลี่ย มัธยฐาน และฐานนิยมเป็นจุดเดียวกันหรือมีค่าเท่ากัน

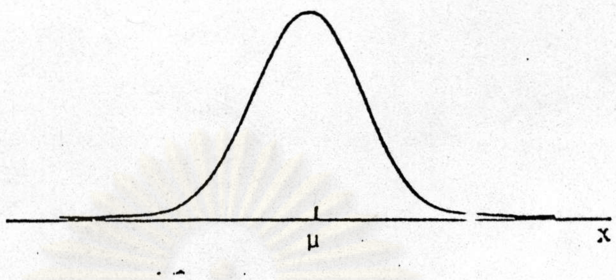
2.6.2.4 มีความโด่ง (kurtosis) ของเส้นโค้ง เท่ากับ 3 ซึ่งเรียกว่า เมโซเคอร์ติค (Mesokurtic) และจุดเปลี่ยนโค้งทั้งสองข้างจะอยู่ ณ. ตรง 1 ส่วนเบี่ยงเบนมาตราฐาน

2.6.2.5 ค่าความเบ้ (Skewness) เท่ากับศูนย์

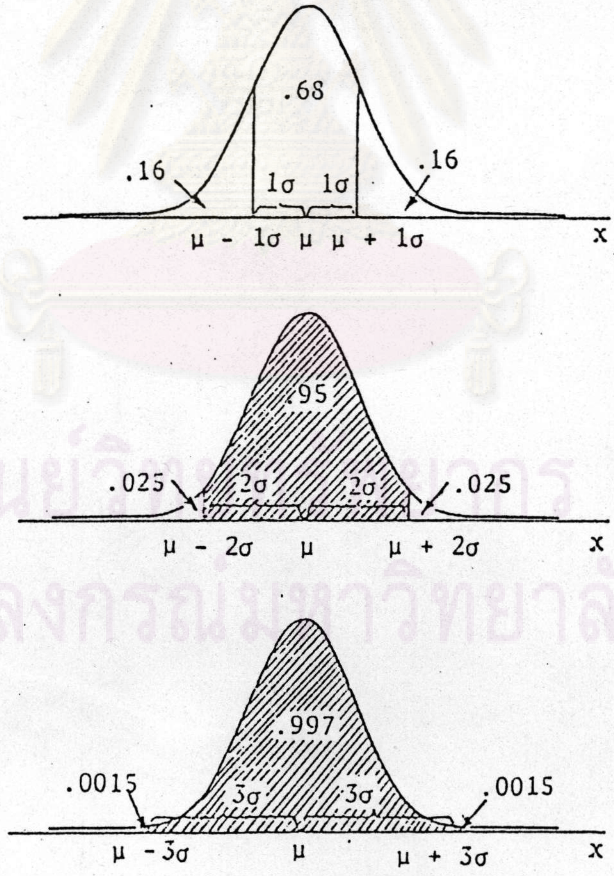
2.6.2.6 ปลายทั้งสองข้างของเส้นโค้งจะค่อย ๆ ลดต่ำลง แต่ไม่จรดกับฐานของโค้งหรือแกนนอน

2.6.2.7 ถ้าลากเส้นตั้งฉากจากแกน x (ซึ่งเป็นแกนนอน) ไปยังเส้นโค้ง โดยที่เส้นดังกล่าวห่างจากจุดเฉลี่ย (μ) ทั้งด้านซ้ายและด้านขวาด้วยระยะหนึ่งเท่า , สองเท่า และสามเท่าของค่าเบี่ยงเบนมาตราฐาน (σ) พื้นที่ที่ปิดกั้นเส้นตั้งฉากกับเส้นโค้งจะเท่ากับ 68% 95% และ 99.7% ของพื้นที่ทั้งหมดตามลำดับ

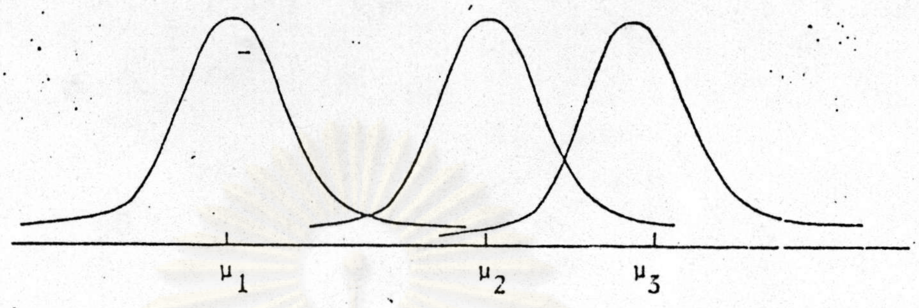
2.6.2.8 พารามิเตอร์ μ และ σ จะเป็นตัวกำหนดตำแหน่งที่ตั้งของเส้นโค้งและความโค้งหรือแบบของโค้งตามลำดับ (ดูรูปที่ 2.1,2.2,2.3,2.4 ประกอบ)



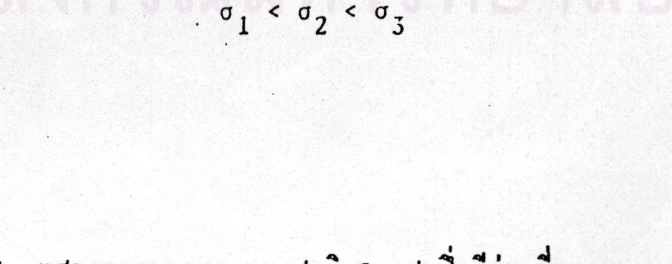
รูปที่ 2.1 แสดงเส้นโค้งการแจกแจงแบบปกติ



รูปที่ 2.2 แสดงพื้นที่ 68% 95% และ 99.7% ของเส้นโค้งปกติ



รูปที่ 2.3 แสดงการแจกแจงแบบปกติ 3 รูป ซึ่งมีค่าเฉลี่ยต่าง ๆ กัน แต่มีค่าเบี่ยงเบนมาตรฐานเท่ากัน



รูปที่ 2.4 แสดงการแจกแจงแบบปกติ 3 รูป ซึ่งมีค่าเบี่ยงเบนมาตรฐานต่างกัน แต่มีค่าเฉลี่ยเท่ากัน

2.7 การแจกแจงไคสแควร์ (Chi-square distribution)

2.7.1 การแจกแจงแบบไคสแควร์ (χ^2 Distribution) เอฟ อาร์ ฮิลเมอร์ท (F. R. Helmert : 1843-1917) นักฟิสิกส์ชาวเยอรมันเป็นผู้ค้นพบเป็นคนแรก เมื่อปี ค.ศ. 1876 ต่อมา คาล เพียสัน (Karl Pearson : 1857-1936) นักสถิติชาวอังกฤษ ได้ทำการพัฒนาและนำมาใช้ประโยชน์ได้อย่างแพร่หลายในปี ค.ศ. 1900

ถ้า $X_1, X_2, \ldots, X_n$ เป็นตัวแปรสุ่มที่มีการแจกแจงเดียวกันและเป็นอิสระต่อกัน (i.i.d.) จากการแจกแจงแบบปกติ โดยมีค่าเฉลี่ยเท่ากับ μ และส่วนเบี่ยงเบนมาตรฐานเท่ากับ σ ถ้าให้ $Z = \dfrac{X-\mu}{\sigma}$

เพราะฉะนั้น $Z \sim N(0,1)$

$$f(z) = \frac{1}{\sqrt{2\pi}} \exp(-z^2/2) , \qquad -\infty < z < \infty$$

ให้ $U = Z^2$

จะได้ $$f(u) = \frac{1}{\sqrt{2\pi}} u^{-1/2} . \exp(-u/2) , \qquad u > 0$$

ซึ่งเป็นฟังก์ชันความหนาแน่นของการแจกแจงแบบ $\chi^2_{(1)}$

ฟังก์ชันที่ให้โมเมนต์ของ U คือ

$$m_U(t) = (1-2t)^{-1/2}$$

ถ้า U_i , $i = 1, 2, \ldots, n$ เป็นตัวแปรสุ่มที่มีการแจกแจงเดียวกันและเป็นอิสระต่อกัน

ฟังก์ชันที่ให้โมเมนต์ของ $W = \sum_{i=1}^{n} U_i$ คือ

$$m_W(t) = [m_U(t)]^n$$
$$= (1-2t)^{-n/2}$$

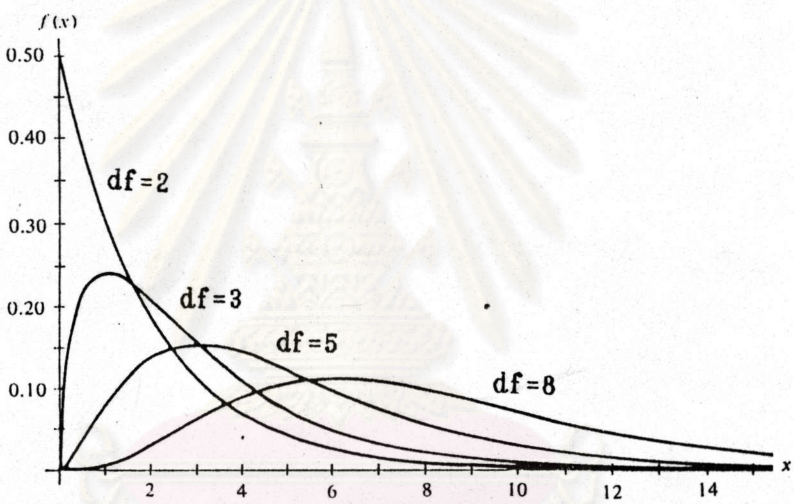
ซึ่งเป็นฟังก์ชันที่ให้โมเมนต์ของการแจกแจงแบบ $\chi^2_{(n)}$

ถ้าประมาณค่าเฉลี่ย μ ด้วย $\bar{X}$ จะได้ว่า $\sum_{i=1}^{n} \dfrac{(X_i-\bar{X})^2}{\sigma^2} = \chi^2_{(n-1)}$

หรือเขียนในรูปความแปรปรวน (s^2) จะได้สูตรดังนี้

$$\chi^2_{(n-1)} = \frac{(n-1)s^2}{\sigma^2} \qquad \text{โดยที่} \qquad s^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{X})^2}{n-1}$$

ลักษณะของการแจกแจงหรือการกระจายของค่าไคสแควร์ จะเปลี่ยนแปลงไปตามขนาดของกลุ่มตัวอย่าง (n) หรือกล่าวอีกนัยหนึ่งว่า การกระจายของไคสแควร์จะแปรตามองศาอิสระสำหรับกลุ่มตัวอย่างกลุ่มเดียว ในที่นี้องศาอิสระเท่ากับ n-1 ดังรูป



รูปที่ 2.5 แสดงการแจกแจงแบบไคสแควร์ ที่องศาอิสระระดับต่าง ๆ

ฟังก์ชันความหนาแน่นของการแจกแจงแบบไคสแควร์คือ

$$f(x) = \frac{1}{2^{n/2}\,\Gamma(n/2)} \cdot x^{(n/2)-1} \cdot e^{-x/2} \quad ; x > 0 \; ; n = 1,2...$$

โดยที่ n คือองศาอิสระ (degree of freedom) ข้อมูลที่มีการแจกแจงแบบไคสแควร์นี้ จะมีลักษณะเบ้ทางขวา กล่าวคือ ข้อมูลค่ากลาง ๆ จะมีจำนวนมากกว่าข้อมูลที่มีค่าสูง เช่น ข้อมูลเกี่ยวกับรายได้ของผู้สำเร็จการศึกษาขั้นปริญญาตรีในเขตกรุงเทพมหานคร เป็นต้น

2.7.2 คุณสมบัติของการแจกแจงแบบไคสแควร์ มีดังนี้

2.7.2.1 ค่าเฉลี่ยของการแจกแจง χ^2 จะเท่ากับองศาอิสระ (degree of freedom) ของมัน ดังนั้นค่าเฉลี่ยของ $\chi^2_n = n$

2.7.2.2 ค่ามัธยฐานของ $\chi^2_n = n-2$ เมื่อ $n > 2$

2.7.2.3 ความเบี่ยงเบนมาตรฐานของ $\chi^2_n = \sqrt{2n}$ หรือมีความแปรปรวน = 2n

2.7.2.4 ค่าความเบ้ (skewness) ของ χ^2_n เป็น $\sqrt{8/n}$ จึงทำให้ลักษณะการกระจายของ χ^2 โดยทั่ว ๆ ไปมักจะมีลักษณะเบ้ทางบวกหรือเบ้ทางขวา (Positivety Skewed)

2.7.2.5 ค่า χ^2 จะมีค่าตั้งแต่ 0 ถึง $+\infty$ คือ มีค่าเป็นบวกทั้งหมด ทั้งนี้เนื่องจาก χ^2 คือค่า Z^2 นั่นเอง

2.7.2.6 ถ้า n มีขนาดใหญ่แล้ว การแจกแจงของ χ^2 จะมีลักษณะเข้าสู่การแจกแจงแบบปกติที่มีค่าเฉลี่ยเป็น n และความเบี่ยงเบนมาตรฐานเป็น $\sqrt{2n}$

2.7.2.7 ถ้า $X_1,...,X_k$ เป็นตัวแปรสุ่มที่มีการแจกแจงเหมือนกันและเป็นอิสระต่อกัน (i.i.d.) ถ้า $X_i \sim \chi^2(n_i)$ จะได้ว่า $\sum_{i=1}^{k} X_i \sim \chi^2(\sum_{i=1}^{k} n_i)$

2.8 ผลงานวิจัยที่เกี่ยวข้อง

สำหรับในหัวข้อนี้จะกล่าวถึงผลงานวิจัยที่เกี่ยวข้อง ในอดีตมีนักสถิติหลายท่านทำการศึกษาเกี่ยวกับ การประมาณช่วงความเชื่อมั่นสำหรับความแปรปรวนของการแจกแจงแบบปกติ ซึ่งในที่นี้จะนำเสนอเฉพาะผลงานวิจัยที่สำคัญพอเป็นสังเขป ดังนี้

ในปี ค.ศ.1959 เทต (Tate, R. F.) และ เคล็ท (Klett, G. W.) นำเสนอตารางแสดงค่าตัวหารสำหรับช่วงความเชื่อมั่น สำหรับความแปรปรวนของการแจกแจงแบบปกติ โดยวิธีช่วงความเชื่อมั่นที่สั้นที่สุด ที่ระดับองศาอิสระ r เท่ากับ 2, 3,..., 29 และที่ระดับสัมประสิทธิ์ความเชื่อมั่น 90%, 95%, 99%, 99.5% และ 99.9% ซึ่งในการศึกษาครั้งนั้นได้มีข้อสงสัยเกี่ยวกับช่วงความเชื่อมั่นชนิดนี้ไว้สองประการคือ ประการแรก ช่วงความเชื่อมั่นที่สั้นที่สุด ต้อง

มีรูปแบบการประมาณแบบช่วงอยู่บนพื้นฐานของค่าความแปรปรวนตัวอย่าง (s^2) เพียงรูปแบบเดียวหรือไม่และประการที่สอง ในบรรดาช่วงความเชื่อมั่นที่มีรูปแบบการประมาณแบบช่วงอยู่บนพื้นฐานของค่าความแปรปรวนตัวอย่างเพียงค่าเดียว ช่วงความเชื่อมั่นที่สั้นที่สุด จะต้องมีรูปแบบเป็น $[rs^2/b, rs^2/a]$ ใช่หรือไม่

ในปี ค.ศ. 1960 ลินเลย์ (Lindley, D. V.) อีส (East, D. A.) และแฮมมิลตัน (Hamilton, P. A.) เสนอตารางเกี่ยวกับการอนุมานค่าความแปรปรวนของการแจกแจงแบบปกติ โดยใช้หลักการของเบส์ ที่ระดับองศาอิสระ r เท่ากับ 1, 2,..., 100 และที่ระดับสัมประสิทธิ์ความเชื่อมั่น 95%, 99% และ 99.9%

ในปี ค.ศ.1972 อาร์เธอร์ โคเฮน (Arthur Cohen) ศึกษาถึงการปรับปรุงช่วงความเชื่อมั่นสำหรับความแปรปรวนของการแจกแจงแบบปกติ โดยเขาได้ศึกษาผลงานของ เทต และ เคล็ท และได้ตอบข้อสงสัยสองประการไว้ดังนี้ ประการแรก ไม่จำเป็นที่ช่วงความเชื่อมั่นที่สั้นที่สุด จะต้องมีรูปแบบการประมาณแบบช่วงอยู่บนพื้นฐานของค่าความแปรปรวนตัวอย่างเพียงค่าเดียวเสมอไป โดยเขาได้เสนอช่วงความเชื่อมั่นสำหรับความแปรปรวนของการแจกแจงแบบปกติอีกรูปแบบหนึ่ง[3] และได้ให้ข้อสรุปว่าช่วงความเชื่อมั่นรูปแบบนี้ให้ค่าระดับความเชื่อมั่นสูงกว่า $1-\alpha$ และมีความยาวของช่วงใกล้เคียงกับ ช่วงความเชื่อมั่นที่สั้นที่สุด ตามที่เทตและเคล็ทเสนอไว้ ประการที่สอง ในบรรดาช่วงความเชื่อมั่นที่รูปแบบการประมาณแบบช่วงอยู่บนพื้นฐานของค่าความแปรปรวนตัวอย่างเพียงค่าเดียว ช่วงความเชื่อมั่นที่สั้นที่สุดคือช่วงความเชื่อมั่นที่อยู่ในรูปแบบ $[rs^2/b, rs^2/a]$ โดยกล่าวว่า ถ้าทราบค่าความแปรปรวนตัวอย่างเพียงค่าเดียว ซึ่งเป็นที่ทราบกันดีว่า rs^2/σ^2 มีการแจกแจงแบบไคสแควร์ ด้วยองศาอิสระ r แล้วจะไม่มีช่วงความเชื่อมั่นในรูปแบบอื่นใด ที่จะให้ค่าระดับความเชื่อมั่นสูงกว่าหรือเท่ากับ $1-\alpha$ และให้ค่าความยาวเฉลี่ยของช่วงเท่ากับหรือสั้นกว่า ช่วงความเชื่อมั่นโดยวิธีช่วงความเชื่อมั่นที่สั้นที่สุด ของเทตและเคล็ท

---

[3] Arthur Cohen. "Improved Confidence Intervals for the Variance of a Normal Distribution," <u>J.Am.Statist.Assoc</u>. 67 (1972): 382-387.

ในปี ค.ศ.1975 โรเจอร์ คริสแมน (Roger Crisman) ศึกษาเกี่ยวกับช่วงความเชื่อมั่นที่สั้นที่สุดสำหรับส่วนเบี่ยงเบนมาตราฐานของการแจกแจงแบบปกติ 3 วิธี คือ วิธีการแรก โดยการถอดรากที่สองของช่วงความเชื่อมั่นสำหรับความแปรปรวนโดยวิธีไคสแควร์ วิธีการที่สอง คือการหารากที่สองของช่วงความเชื่อมั่นสำหรับความแปรปรวน โดยวิธีช่วงความเชื่อมั่นที่สั้นที่สุด ตามวิธีการของ เทตและเคล็ท และวิธีการสุดท้าย ใช้วิธีการของเทตและเคล็ทเช่นกัน เรียกว่า วิธีช่วงความเชื่อมั่นที่สั้นที่สุดสำหรับส่วนเบี่ยงเบนมาตราฐาน กล่าวคือ ใช้วิธีการทางคณิตศาสตร์ เพื่อหาความยาวที่สั้นที่สุดของ ช่วงความเชื่อมั่นสำหรับส่วนเบี่ยงเบนมาตราฐานที่มีรูปแบบ คือ $[\sqrt{(n-1)s^2/b}\ ,\ \sqrt{(n-1)s^2/a}]$ ได้ข้อสรุปว่า วิธีการสุดท้ายจะให้ค่าความยาวของช่วงสั้นที่สุด และจากผลการวิจัยครั้งนั้น โรเจอร์ ได้ตารางแสดงค่าตัวหารสำหรับช่วงความเชื่อมั่นสำหรับส่วนเบี่ยงเบนมาตราฐานโดยวิธีช่วงความเชื่อมั่นที่สั้นที่สุด ที่ระดับองศาอิสระ r เท่ากับ 2, 3,...., 29 และสัมประสิทธิ์ความเชื่อมั่น 80%, 90%, 95%, 99% และ 99.5%

- <u>ความสัมพันธ์ระหว่างงานวิจัยนี้กับผลงานวิจัยที่เกี่ยวข้อง</u>

ทั้งผลงานวิจัยของ เทตและเคล็ท และผลงานวิจัยของลีนเลย์, อีส และ แฮมมิงตัน ต่างก็เสนอตัวหารสำหรับช่วงความเชื่อมั่นสำหรับความแปรปรวนของการแจกแจงแบบปกติ และคาดว่าจะต้องเป็นช่วงความเชื่อมั่นที่เหมาะสมกว่าวิธีไคสแควร์ ซึ่งเป็นวิธีดั้งเดิมที่ใช้อยู่โดยทั่วไป นอกจากนี้เขาทั้งหลายยังกล่าวว่า เมื่อตัวอย่างมีขนาดใหญ่ วิธีการที่เสนอจะให้ผลใกล้เคียงกับวิธีไคสแควร์ แต่ก็ยังไม่ได้มีข้อสรุปที่แน่นอนว่า ณ ระดับขนาดตัวอย่างเท่าใดจึงจะเป็นเช่นนั้น ดังนั้น ในการวิจัยครั้งนี้ จึงทำการศึกษาเปรียบเทียบวิธีการประมาณช่วงความเชื่อมั่นสำหรับความแปรปรวนของการแจกแจงแบบปกติ ซึ่งเป็นผลงานของ เทตและเคล็ท กับ ผลงานของ ลีนเลย์, อีส และ แฮมมิงตัน กับ วิธีไคสแควร์ ที่ระดับขนาดตัวอย่างที่แตกต่างไปจากเดิม และยังศึกษาต่อไปอีกว่า ที่ระดับขนาดตัวอย่างมีค่ามากเท่าใด ที่วิธีการประมาณทั้ง 3 วิธี จะให้ผลใกล้เคียงกันในแง่ของความกว้างของช่วงความเชื่อมั่น