หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 1

บทนำ

1.1 ที่มาและความสำคัญของปัญหา

การศึกษาวิจัยในสาขาต่างๆ โดยทั่วไปจำเป็นต้องอาศัยระเบียบวิธีการทางสถิติในการวิเคราะห์ข้อมูลเพื่อหาข้อสรุปเกี่ยวกับประชากรที่สนใจ ซึ่งข้อสรุปที่ได้ควรมีความน่าเชื่อถือและมีความถูกต้องในระดับที่ยอมรับในเชิงทฤษฎี การให้ได้มาซึ่งข้อสรุปที่ดีนั้น ผู้วิจัยจะต้องมีความรู้ความเข้าใจในเรื่องที่จะศึกษาเป็นอย่างดี และต้องมีความรู้ทางด้านสถิติมากพอสมควร เช่น ทราบถึงข้อจำกัดและความเหมาะสมของวิธีการแต่ละวิธีที่จะนำมาใช้กับข้อมูล ตลอดจนความสอดคล้องของสถิติที่ใช้กับจุดมุ่งหมายของการวิเคราะห์ เป็นต้น เมื่อพิจารณาในแง่ของการอนุมานเชิงสถิติ เราอาจแบ่งวิธีการทางสถิติได้เป็น 2 ประเภท คือ

- สถิติที่ใช้พารามิเตอร์ (Parametric Statistics) หรือสถิติที่ทราบรูปแบบการแจกแจง ผู้วิจัยจะใช้เพื่อวิเคราะห์ข้อมูลเมื่อตัวอย่างสุ่มมาจากประชากรซึ่งทราบว่ามีการแจงแจงอย่างใดอย่างหนึ่ง
- สถิติที่ไม่ใช้พารามิเตอร์ (Nonparametric Statistics)หรือสถิติที่ไม่ทราบรูปแบบการแจกแจง ผู้วิจัยจะใช้วิธีการนี้เมื่อไม่สามารถระบุได้ว่าข้อมูลที่นำมาวิเคราะห์มาจากการแจกแจงแบบใด

การนำสถิติประเภทจำกัดรูปแบบการแจกแจงไปใช้ในการอนุมานเชิงสถิติ จำเป็นต้องทราบข้อสมมติเบื้องต้นบางอย่างเกี่ยวกับรูปแบบการแจกแจง ทั้งนี้เพราะหลักเกณฑ์ในการสร้างสถิติเหล่านี้ ได้สร้างมาจากข้อสมมติเกี่ยวกับการแจกแจงของประชากรและในบางกรณีก็สร้างมาจากหลักเกณฑ์อื่น ๆ อีกด้วย

ในทางปฏิบัตินักวิจัยส่วนใหญ่มักจะไม่มีการทดสอบดูว่าการแจกแจงของข้อมูลที่ศึกษามีลักษณะเป็นไปตามข้อสมมติเบื้องต้นหรือไม่ซึ่งอาจเป็นผลเสียเนื่องจากความถูกต้องและความหมายของผลที่ได้จากการวิเคราะห์โดยวิธีการทางสถิติประเภทจำกัดรูปแบบการแจกแจงที่ขึ้นอยู่กับความถูกต้องของข้อสมมติเบื้องต้นขาดความน่าเชื่อถือและอาจผิดไปจากความเป็นจริงได้ การแจกแจงปกติเป็นการแจงแจงตามทฤษฎี (Theoretical Distribution) การแจกแจงหนึ่งที่มีความสำคัญมากในทางสถิติ และมีการนำไปใช้ประโยชน์มากในการอนุมานเชิงสถิติ

ในปัจจุบันการวิเคราะห์ตัวแปรพหุ (Multivariate Analysis) เป็นวิธีทางสถิติที่ใช้กันอยู่ทั่วไป โดยอาจจะใช้ในรูป การทดสอบสมมติฐาน การวิเคราะห์องค์ประกอบหลัก (Principal Component Analysis) การวิเคราะห์ปัจจัย (Factor Analysis) และการวิเคราะห์จำแนกประเภท (Discriminant Analysis) เป็นต้น ซึ่งการวิเคราะห์บางประเภทมีข้อสมมติเบื้องต้นว่า ประชากรมีการแจกแจงปกติหลายตัวแปร (Multivariate Normal Distribution) จึงจำเป็นต้องมีการตรวจสอบข้อสมมติเบื้องต้นก่อนการทำวิจัยว่าข้อมูลที่จะนำมาวิเคราะห์มีการแจกแจงปกติหลายตัวแปรหรือไม่ เพราะหาก ปรากฏว่าข้อมูลที่ศึกษามีการแจกแจงที่ไม่ใช่การแจกแจงปกติหลายตัวแปร การนำไปใช้ภายใต้ข้อสมมติเบื้องต้นของการแจกแจงปกติหลายตัวแปร ก็จะทำให้ได้ผลสรุปที่ไม่ถูกต้องและขาดความน่าเชื่อถือ

ในการทดสอบว่าข้อมูลมาจากการแจกแจงปกติหลายตัวแปรหรือไม่นั้น มีนักสถิติหลายท่านได้คิดค้นตัวสถิติที่จะใช้ในการทดสอบการแจกแจงปกติหลายตัวแปร อาทิเช่น

ในปี ค.ศ.1991 Aydin Ozturk ได้เสนอตัวสถิติทดสอบ $Q_n^{(k)}$ เพื่อใช้ทดสอบการแจกแจงปกติหลายตัวแปร สำหรับกรณีหนึ่งตัวแปรและหลายตัวแปร โดยวิธีการเปรียบเทียบกราฟของ $Q_n^{(k)}$ ที่ได้จากข้อมูลตัวอย่างกับกราฟที่ได้จากสมมติฐานว่าง การทดสอบดังกล่าวนอกจากแสดงถึงอำนาจการทดสอบแล้วยังทำให้สามารถระบุการแจกแจงที่แท้จริงของข้อมูลตัวอย่างจากกราฟที่ได้

ในปี ค.ศ.1991 K.V.Mardia and J.T.Kent ได้เสนอตัวสถิติทดสอบสำหรับใช้ทดสอบการแจกแจงปกติหลายตัวแปร และทดสอบความเป็นอิสระกันระหว่างตัวแปรโดยพัฒนามาจากสถิติเราว์สคอร์ (Rao's Score Statistic)

ในปี ค.ศ.1992 Govind S. Mudholkar, Maria Mcdermott and Deo Kumar Srivastava ได้พัฒนาตัวสถิติทดสอบการแจกแจงปกติหนึ่งตัวแปรที่เสนอโดย Lin and Mudholkar(1980) เพื่อใช้ทดสอบการแจกแจงปกติหลายตัวแปร และได้ศึกษาอำนาจการทดสอบของตัวสถิติที่นำเสนอโดยใช้เทคนิคมอนติคาร์โลจำลองข้อมูลตัวอย่าง 5,000 ศึกษาที่ขนาดตัวอย่าง 10, 20 และ 50 ที่จำนวนตัวแปรต่างๆ กัน พบว่าตัวสถิติทดสอบที่เสนอสามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ดีเมื่อจำนวนตัวแปรน้อยกว่าหรือเท่ากับ 5 และขนาตัวอย่างเท่ากับ 15 หรือจำนวนตัวแปรเท่ากับ 6 และขนาดตัวอย่างมากกว่าหรือเท่ากับ 15 โดยกำหนดระดับนัยสำคัญ 0.02, 0.15, 0.10, 0.05 และ 0.01 และตัวสถิติที่เสนอมีอำนาจการทดสอบต่ำสุดเมื่อสมมติฐานทางเลือกเป็นการแจกแจง Morgenstem (Normal Marginal) กรณีสองตัวแปร

ในปี ค.ศ.1995 Charles L. Dunn ได้เสนอตัวสถิติทดสอบ Bimodal (B) และตัวสถิติทดสอบ Omnibus (O) เพื่อใช้ในการทดสอบการแจกแจงปกติหลายตัวแปรเมื่อไม่ทราบค่าพารามิเตอร์ และได้ศึกษาเปรียบเทียบอำนาจการทดสอบกับตัวสถิติทดสอบอื่นๆ คือ ตัวสถิติทดสอบ

Rayleigh (R/C) ตัวสถิติทดสอบ Ajne (A) และตัวสถิติทดสอบ Gine' (G) พบว่า ตัวสถิติทดสอบที่เสนอให้อำนาจการทดสอบสูงสุดเมื่อสมมติฐานทางเลือกมีการแจกแจงBimodal และการแจกแจง Omnibus

ในปี ค.ศ.1996 Kanta Naito ได้เสนอตัวสถิติทดสอบการแจกแจงปกติ p ตัวแปร โดยใช้อินทิกรัลถ่วงน้ำหนักของกำลังสองของค่าสัมบูรณ์ของฟังก์ชันแคแรคเตอรีสติค (Weighted Integral of the Squared Modulus of the Studentized Empirical Characteristic Function) และศึกษาเปรียบเทียบอำนาจการทดสอบของตัวสถิติทดสอบที่เสนอกับสถิติทดสอบของ K.Murota and K.Takeuchi , ตัวสถิติทดสอบของ T.Epps and L.B.Pulley, ตัวสถิติทดสอบของ S.S.Shapiro and M.B.Wilk และตัวสถิติทดสอบสัมประสิทธิ์ความโด่งของ Mardia พบว่า ในกรณีที่ทำการทดสอบการแจกแจงปกติหนึ่งตัวแปร ตัวสถิติทดสอบที่เสนอมีประสิทธิภาพใกล้เคียงกับสถิติทดสอบต่าง ๆ ส่วนกรณีที่ทำการทดสอบการแจกแจงปกติสองตัวแปร พบว่าตัวสถิติทดสอบที่เสนอมีประสิทธิภาพมากกว่าหรือเท่ากับตัวสถิติทดสอบสัมประสิทธิ์ความโด่งของ Mardia

ในปี ค.ศ.1999 Takeaki Kariya , Ruey S. Tsay and Nobuhiko Terui ได้เสนอตัวสถิติทดสอบการแจกแจงปกติหลายตัวแปรโดยใช้คุณลักษณะ Hermitian Polynomial ของความเป็นปกติหลายตัวแปรและศึกษาเปรียบเทียบอำนาจการทดสอบของตัวสถิติทดสอบที่เสนอกับตัวสถิติทดสอบ Madia และตัวสถิติทดสอบ Hinich's bispectral ซึ่งศึกษาข้อมูลอนุกรมเวลา พบว่าตัวสถิติทดสอบที่เสนอให้อำนาจการทดสอบสูงสุดเมื่อข้อมูลมาการแจกแจง U(0,1) และการแจกแจง Student-t ที่องศาความเป็นอิสระเท่ากับ 3 ส่วนการแจกแจงอื่น ๆ ให้อำนาจการทดสอบใกล้เคียงกับตัวสถิติทดสอบอื่น ๆ

ในปี ค.ศ.1999 (พ.ศ.2540) เสาวลักษณ์ ฮุนนางกูร ได้ทำวิทยานิพนธ์เรื่อง " การเปรียบเทียบกำลังการทดสอบของแบบทดสอบการแจกแจงปกติหลายตัวแปร " โดยศึกษาเปรียบเทียบอำนาจการทดสอบของตัวสถิติทดสอบการแจกแจงปกติหลายตัวแปรที่พัฒนามาจากสถิติ Rao score ที่เสนอโดย K.V.Mardia and J.T.Kent (T,1991) ตัวสถิติทดสอบที่พัฒนามาจากสถิติ Shapiro-Wilk ที่เสนอโดย Govind S.Mudholkar, Deo Kumar Srivastava and C.Thomas Lin(W_F,1995) และตัวสถิติทดสอบที่เสนอโดย Charles L.Dunn (O,1995) โดยศึกษาทั้งกรณีทราบค่าและไม่ทราบค่าพารามิเตอร์ของการแจกแจงที่มีขนาดตัวอย่าง 20 , 30 และ 50 เมื่อระดับนัยสำคัญ 0.05 และ 0.10 ผลการศึกษาสรุปว่า ตัวสถิติทดสอบการแจกแจงปกติหลายตัวแปรทั้งสามสามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ทั้งกรณีทราบค่าและไม่ทราบค่าพารามิเตอร์ของการแจกแจง ในกรณีไม่ทราบค่าพารามิเตอร์ ตัวสถิติทดสอบ W_F ให้อำนาจการทดสอบสูงสุดเมื่อขนาดตัวอย่างเท่ากับ 20 และตัวสถิติทดสอบ T ให้อำนาจการทดสอบสูงสุดเมื่อขนาดตัว

อย่างเท่ากับ 30 และ 50 ส่วนในกรณีทราบค่าพารามิเตอร์ ตัวสถิติทดสอบ T ให้อำนาจการทดสอบสูงสุดเมื่อขนาดตัวอย่างเท่ากับ 20 และตัวสถิติทดสอบ O ให้อำนาจการทดสอบสูงสุดเมื่อขนาดตัวอย่างเท่ากับ 30 และ 50

จากผลการวิจัยข้างต้นยังไม่มีผู้ใดทำการวิจัยเปรียบเทียบอำนาจการทดสอบของตัวสถิติที่เสนอโดย Kanta Naito และของ Takeaki Kariya, Ruey S. Tsay and Nobuhiko Terui ดังนั้นผู้วิจัยจึงสนใจที่จะศึกษาเปรียบเทียบอำนาจการทดสอบของตัวสถิติทดสอบ 2 ตัวนี้กับตัวสถิติทดสอบของ K.V.Mardia and J.T.Kent ซึ่งเป็นตัวสถิติทดสอบที่มีประสิทธิภาพมากในหลายๆ สถานการณ์ ที่เสาวลักษณ์ ฮุนนางกูรได้ศึกษาไว้แล้ว โดยดูว่า ตัวสถิติทดสอบตัวใดที่จะมีอำนาจการทดสอบสูงที่สุดในสถานการณ์ต่างๆ

1.2 วัตถุประสงค์ของการวิจัย

การวิจัยนี้มีวัตถุประสงค์ ดังนี้

1. เพื่อเปรียบเทียบอำนาจการทดสอบของตัวสถิติทดสอบการแจกแจงปกติหลายตัวแปร กรณีไม่ทราบค่าพารามิเตอร์ของการแจกแจง นั่นคือ ตัวสถิติทดสอบ MK (T) ตัวสถิติทดสอบ N (T_W) และตัวสถิติทดสอบ KTT (W_0)
2. เพื่อใช้เป็นข้อมูลในการตัดสินใจที่จะนำตัวสถิติทดสอบไปใช้ประโยชน์

1.3 สมมติฐานของการวิจัย

การวิจัยนี้มีสมมติฐานของการวิจัย เป็นดังนี้

1. ตัวสถิติทดสอบ N จะให้อำนาจการทดสอบสูงสุดเมื่อขนาดตัวอย่างมีขนาดเล็ก
2. ตัวสถิติทดสอบ MK จะให้อำนาจการทดสอบสูงสุดเมื่อขนาดตัวอย่างมีขนาดใหญ่

1.4 ข้อตกลงเบื้องต้น

การวิจัยนี้มีข้อตกลงเบื้องต้น ดังนี้

1. ศึกษาความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 เพื่อพิจารณาความสามารถในการควบคุมความคลาดเคลื่อนประเภทที่ 1 ของตัวสถิติทดสอบทั้ง 3 ตัว
2. ศึกษาค่าอำนาจการทดสอบของตัวสถิติทดสอบที่สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ โดยพิจารณาเลือกตัวสถิติทดสอบที่ให้อำนาจการทดสอบสูงสุดในแต่ละสถานการณ์

หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

1.5 ขอบเขตของการวิจัย

การวิจัยนี้ทำการศึกษาภายใต้ขอบเขตต่อไปนี้

1. จำนวนตัวแปรที่ศึกษาเท่ากับ 2 และ 3 ตัวแปร
2. การแจกแจงของประชากร 4 การแจกแจง

(1) การแจกแจงปกติหลายตัวแปร (Multivariate Normal Distribution)

กำหนดให้ $\underset{\sim}{X}$ เป็นตัวแปรสุ่มต่อเนื่องที่มีการแจกแจงปกติหลายตัวแปรด้วยพารามิเตอร์ $\underset{\sim}{\mu}$ และ Σ จะได้ฟังก์ชันความหนาแน่นของ $\underset{\sim}{X}$ เป็นดังนี้

$$f(\underset{\sim}{x}) = \frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\underset{\sim}{x}-\underset{\sim}{\mu})' \Sigma^{-1}\left(\underset{\sim}{x}-\underset{\sim}{\mu}\right)\right] \quad ; \quad \underset{\sim}{0} < \underset{\sim}{x} < \infty \ , \ \underset{\sim}{x} = (x_1, x_2, \ldots, x_p)'$$

$$; \quad -\infty < \underset{\sim}{\mu} < \infty \ , \ \Sigma > \underset{\sim}{0}$$

โดยที่ $\underset{\sim}{\mu}$ แทน เวกเตอร์ค่าเฉลี่ย

Σ แทน เมทริกซ์ความแปรปรวนร่วม

p แทน จำนวนตัวแปร

กำหนดค่าพารามิเตอร์ดังนี้

กรณี 2 ตัวแปร ค่าเฉลี่ย $\mu_1 = 5$, $\mu_2 = 10$

ความแปรปรวน $\sigma_1^2 = 4$, $\sigma_2^2 = 9$

ค่าสหสัมพันธ์ระหว่างตัวแปร $\rho_{12} = \rho_{21} = 0.3, 0.6, 0.9$

กรณี 3 ตัวแปร ค่าเฉลี่ย $\mu_1 = 5$, $\mu_2 = 10$, $\mu_3 = 15$

ความแปรปรวน $\sigma_1^2 = 4$, $\sigma_2^2 = 9$, $\sigma_3^2 = 16$

ค่าสหสัมพันธ์ระหว่างตัวแปร ρ_{12} , ρ_{13} , ρ_{23} มีค่าดังตารางต่อไปนี้

กรณีที่	ρ_{12}	ρ_{13}	ρ_{23}
1	0.3	0.3	0.3
2	0.6	0.6	0.6
3	0.9	0.9	0.9
4	0.3	0.3	0.6
5	0.3	0.3	0.9
6	0.6	0.6	0.9
7	0.3	0.6	0.9

เมื่อ ρ_{ij} หมายถึง สหสัมพันธ์ระหว่างตัวแปร X_i กับ X_j และ $\rho_{ij} = \rho_{ji}$; $i, j = 1,2,3$

(2) <u>การแจกแจงล็อกนอร์มอลหลายตัวแปร (Multivariate lognormal Distribution)</u>

กำหนดให้ $\underset{\sim}{X}$ เป็นตัวแปรสุ่มต่อเนื่องที่มีการแจกแจงล็อกนอร์มอลหลายตัวแปร ด้วยพารามิเตอร์ $\underset{\sim}{\mu}$ และ Σ จะได้ฟังก์ชันความหนาแน่นของ $\underset{\sim}{X}$ เป็นดังนี้

$$f(\underset{\sim}{x}) = \frac{1}{\underset{\sim}{x}(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\ln \underset{\sim}{x} - \underset{\sim}{\mu})' \Sigma^{-1}\left(\ln \underset{\sim}{x} - \underset{\sim}{\mu}\right)\right] \quad ; \quad \underset{\sim}{0} < \underset{\sim}{x} < \infty$$

$$; \quad -\infty < \underset{\sim}{\mu} < \infty \;,\; \Sigma > \underset{\sim}{0}$$

โดยที่ $\underset{\sim}{\mu}$ แทน เวกเตอร์ค่าเฉลี่ย

Σ แทน เมตริกซ์ความแปรปรวนร่วม

p แทน จำนวนตัวแปร

กำหนดค่าพารามิเตอร์

ค่าเฉลี่ย $\mu_i = 5$; $i = 1, 2, 3$ และค่าความแปรปรวนดังต่อไปนี้

กรณีที่	จำนวนตัวแปรเท่ากับ 2 (σ_1^2, σ_2^2)	จำนวนตัวแปรเท่ากับ 3 $(\sigma_1^2, \sigma_2^2, \sigma_3^2)$
1	(0.691,0.691)	(0.691,0.691,0.691)
2	(2.600,2.600)	(2.600,2.600,2.600)
3	(5.440,5.440)	(5.440,5.440,5.440)
4	(8.900,8.900)	(8.900,8.900,8.900)
5	(12.70,12.70)	(12.70,12.70,12.70)
6	(0.691,2.600)	(0.691,2.600,5.440)
7	(5.440,8.900)	(5.440,8.900,12.70)
8	(8.900,12.70)	(0.691,2.600,12.70)

ค่าสหสัมพันธ์ระหว่างตัวแปร

กรณี 2 ตัวแปร $\rho_{12} = \rho_{21} = 0.1, 0.3, 0.4, 0.6, 0.7, 0.9$

กรณี 3 ตัวแปร ρ_{12} , ρ_{13} , ρ_{23} มีค่าดังตารางต่อไปนี้

กรณีที่	ρ_{12}	ρ_{13}	ρ_{23}
1	0.1	0.1	0.1
2	0.3	0.3	0.3
3	0.4	0.4	0.4
4	0.6	0.6	0.6
5	0.7	0.7	0.7

กรณีที่	ρ_{12}	ρ_{13}	ρ_{23}
6	0.9	0.9	0.9
7	0.1	0.1	0.4
8	0.1	0.1	0.7
9	0.4	0.4	0.7
10	0.1	0.4	0.7
11	0.3	0.3	0.6
12	0.3	0.3	0.9
13	0.6	0.6	0.9
14	0.3	0.6	0.9

เมื่อ ρ_{ij} หมายถึง สหสัมพันธ์ระหว่างตัวแปร X_i กับ X_j และ $\rho_{ij} = \rho_{ji}$; $i, j = 1,2,3$

(3) <u>การแจกแจงสติวเดนท์-ทีหลายตัวแปร (Multivariate Student-t Distribution)</u>

กำหนดให้ $\underset{\sim}{X}$ เป็นตัวแปรสุ่มต่อเนื่องที่มีการแจกแจงสติวเดนท์-ทีหลายตัวแปรด้วยองศาความเป็นอิสระเท่ากับ υ จะได้ฟังก์ชันความหนาแน่นของ $\underset{\sim}{X}$ เป็นดังนี้

$$f(\underset{\sim}{x}) = \frac{\Gamma\left(\frac{\upsilon+p}{2}\right)}{(\upsilon\pi)^{\frac{p}{2}}\Gamma\left(\frac{\upsilon}{2}\right)|\Sigma|^{\frac{1}{2}}}\left(1+\frac{1}{\upsilon}\underset{\sim}{x}'\Sigma^{-1}\underset{\sim}{x}'\right)^{-\frac{(\upsilon+p)}{2}} \quad ; \quad -\infty < \underset{\sim}{x} < \infty$$

เมื่อ υ และ p แทน องศาความเป็นอิสระและจำนวนตัวแปร ตามลำดับ

กำหนดค่าพารามิเตอร์

องศาความเป็นอิสระเท่ากับ 3, 4, 5

ค่าเฉลี่ย $\mu_i = 0$; $i = 1, 2, 3$ และค่าความแปรปรวนดังต่อไปนี้

องศาความเป็นอิสระ	จำนวนตัวแปรเท่ากับ 2 (σ_1^2, σ_2^2)	จำนวนตัวแปรเท่ากับ 3 $(\sigma_1^2, \sigma_2^2, \sigma_3^2)$
3	(3 , 3)	(3 , 3 , 3)
4	(2 , 2)	(2 , 2 , 2)
5	(1.67 , 1.67)	(1.67 , 1.67 , 1.67)

ค่าสหสัมพันธ์ระหว่างตัวแปร

กรณี 2 ตัวแปร $\rho_{12} = \rho_{21} = 0.1, 0.3, 0.4, 0.6, 0.7, 0.9$

กรณี 3 ตัวแปร ρ_{12} , ρ_{13} , ρ_{23} **มีค่าดังตารางต่อไปนี้**

กรณีที่	ρ_{12}	ρ_{13}	ρ_{23}
1	0.1	0.1	0.1
2	0.3	0.3	0.3
3	0.4	0.4	0.4
4	0.6	0.6	0.6
5	0.7	0.7	0.7
6	0.9	0.9	0.9
7	0.1	0.1	0.4
8	0.1	0.1	0.7
9	0.4	0.4	0.7
10	0.1	0.4	0.7
11	0.3	0.3	0.6
12	0.3	0.3	0.9
13	0.6	0.6	0.9
14	0.3	0.6	0.9

เมื่อ ρ_{ij} **หมายถึง สหสัมพันธ์ระหว่างตัวแปร** X_i **กับ** X_j **และ** $\rho_{ij} = \rho_{ji}$; $i, j = 1,2,3$

(4) <u>การแจกแจงไคสแควร์หลายตัวแปร (Multivariate Chi-square Distribution)</u>

กำหนดให้ $\underset{\sim}{X}$ เป็นตัวแปรสุ่มต่อเนื่องที่มีการแจกแจงไคสแควร์หลายตัวแปรด้วยองศาความเป็นอิสระเท่ากับ υ จะได้ฟังก์ชันความหนาแน่นของ $\underset{\sim}{X}$ เป็นดังนี้

$$f(\underset{\sim}{x}) = \sum_{j=0}^{\infty} \frac{\Gamma\left(\frac{\upsilon}{2}+j\right)}{\Gamma\left(\frac{\upsilon}{2}\right) j!} \rho^{2j}(1-\rho^2)^{\frac{n}{2}} \prod_{i=1}^{2} \left\{ \left(\frac{x_i}{1-\rho^2}\right)^{\frac{\upsilon}{2}+j-1} \exp\left(-\frac{x_i}{1-\rho^2}\right) \frac{1}{\Gamma\left(\frac{\upsilon}{2}+j\right) 2^{\frac{\upsilon}{2}+j}(1-\rho^2)} \right\}$$

$$; \quad \underset{\sim}{x} = (x_1, x_2)' \quad , \quad \underset{\sim}{x} > 0$$

เมื่อ υ และ ρ แทน องศาความเป็นอิสระและสหสัมพันธ์ระหว่างตัวแปร X_1 และ X_2 ตามลำดับ

กำหนดค่าพารามิเตอร์

องศาความเป็นอิสระเท่ากับ 3, 4, 5 จะมีค่าเฉลี่ย ความแปรปรวนและสหสัมพันธ์ระหว่างตัวแปร ดังต่อไปนี้

องศาความเป็นอิสระ	จำนวนตัวแปรเท่ากับ 2			จำนวนตัวแปรเท่ากับ 3		
	(μ_1, μ_2)	(σ_1^2, σ_2^2)	(ρ_{12})	(μ_1, μ_2, μ_3)	$(\sigma_1^2, \sigma_2^2, \sigma_3^2)$	$(\rho_{12}, \rho_{13}, \rho_{23})$
3	(3 , 3)	(6 , 6)	(0.3)	(3 , 3 , 3)	(6 , 6 , 6)	(0.3 , 0.3 , 0.3)
4	(4 , 4)	(8 , 8)	(0.5)	(4 , 4 , 4)	(8 , 8 , 8)	(0.5 , 0.5 ,0.5)
5	(5 , 5)	(10 , 10)	(0.6)	(5 , 5 , 5)	(10 , 10 , 10)	(0.6 , 0.6 ,0.6)

เมื่อ $\rho_{ij} = \rho_{ji} \quad ; \quad i, j = 1,2,3$

3. ขนาดตัวอย่างที่ศึกษาเท่ากับ 20, 30, 40, 50, 60, 70, 80, 90 และ 100

4. ระดับนัยสำคัญของการทดสอบการแจกแจงปกติหลายตัวแปรเท่ากับ 0.05 และ 0.10

5. การประมาณค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 กำหนดจำนวนซ้ำของการตรวจสอบเท่ากับ 2,000 ครั้งในแต่ละสถานการณ์

6. ระดับนัยสำคัญของการทดสอบทวินาม (Binomial Test) ในการทดสอบความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 เท่ากับ 0.05

7.การหาค่าอำนาจการทดสอบของสถิติทดสอบทำซ้ำจำนวน 2,000 ครั้งในแต่ละสถานการณ์

8. การสร้างตารางวิกฤตของตัวสถิติทดสอบ N และตัวสถิติทดสอบ KTT กำหนดจำนวนซ้ำของการคำนวณค่าสถิติทดสอบทั้งสองเท่ากับ 15,000 ครั้งในแต่ละสถานการณ์

9. การวิจัยนี้ทำการทดลองข้อมูลตามสถานการณ์ต่างๆ ด้วยเทคนิคมอนติคาร์โล (Monte Carlo Technique) โดยใช้โปรแกรมภาษาฟอร์แทรน

1.6 ขั้นตอนการดำเนินการวิจัย

การวิจัยนี้มีขั้นตอนการดำเนินการวิจัยเป็นดังนี้

1.6.1 คำนวณค่าประมาณความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1

การคำนวณค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 จะคำนวณภายใต้สมมติฐานว่าง H_0 เป็นจริง ซึ่งมีขั้นตอนดังนี้

(1) กำหนดให้ข้อมูลมีการแจกแจงปกติหลายตัวแปร ด้วยค่าพารามิเตอร์ของการแจกแจงและขนาดตัวอย่างตามที่ระบุในหัวข้อ 1.5

(2) จำลองข้อมูลให้มีลักษณะตามที่กำหนดใน (1)

(3) ประมาณค่าพารามิเตอร์ ในที่นี้คือค่าเฉลี่ยและความแปรปรวน (เนื่องจากการวิจัยครั้งนี้ ศึกษากรณีไม่ทราบค่าพารามิเตอร์)

(4) คำนวณค่าสถิติทดสอบทั้ง 3 ตัว (วิธีการคำนวณค่าสถิติทดสอบทั้ง 3 ตัว แสดงไว้ในบทที่ 2)

(5) เปรียบเทียบค่าสถิติที่คำนวณได้กับค่าวิกฤตของตัวสถิติดังกล่าว เพื่อตัดสินใจว่าจะปฏิเสธหรือยอมรับสมมติฐานว่าง

(6) นับจำนวนครั้งของการปฏิเสธสมมติฐานว่าง

(7) ทำการคำนวณขั้นตอนที่ (2) – (6) ซ้ำจนครบ 2,000 ครั้ง

(8) ความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 จากการทดลอง (ค่าประมาณ) เท่ากับ จำนวนครั้งที่ปฏิเสธสมมติฐานว่างหารด้วยจำนวนครั้งที่ทำการทดลอง (2,000 ครั้ง)

1.6.2 ทดสอบความสามารถในการควบคุมความคลาดเคลื่อนประเภทที่ 1

การทดสอบว่าตัวสถิติแต่ละตัวสามารถควบคุมคลาดเคลื่อนประเภทที่ 1 ได้หรือไม่จะใช้การทดสอบทวินาม ภายใต้ระดับนัยสำคัญ 0.05 โดยตัวสถิติทดสอบที่สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ระดับ 0.05 ได้จะมีค่าประมาณความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 อยู่ในช่วง $[\,0\,,0.0580\,]$ ส่วนตัวสถิติทดสอบที่สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ระดับ 0.10 ได้จะมีค่าประมาณความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 อยู่ในช่วง $[\,0\,,0.1110\,]$

1.6.3 คำนวณค่าอำนาจการทดสอบและเปรียบเทียบค่าที่คำนวณได้

การทดสอบค่าอำนาจการทดสอบของตัวสถิติทดสอบที่สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้จะคำนวณภายใต้สมมติฐานว่าง H_0 ไม่เป็นจริง ซึ่งมีขั้นตอนดังนี้

(1) กำหนดให้ข้อมูลมีการแจกแจงล็อกนอร์มอลหลายตัวแปร การแจกแจงสติวเดนท์-ทีหลายตัวแปร และการแจกแจงไคสแควร์หลายตัวแปร ด้วยค่าพารามิเตอร์ของแต่ละการแจกแจงและขนาดตัวอย่างตามที่ระบุในหัวข้อ 1.5

(2) จำลองข้อมูลให้มีลักษณะตามที่กำหนดใน (1)

(3) ประมาณค่าพารามิเตอร์ ในที่นี้คือค่าเฉลี่ย ความแปรปรวน

(4) คำนวณค่าสถิติทดสอบทั้ง 3 ตัว

(5) เปรียบเทียบค่าสถิติที่คำนวณได้กับค่าวิกฤตของตัวสถิติดังกล่าว เพื่อตัดสินใจว่าจะปฏิเสธหรือยอมรับสมมติฐานว่าง

(6) นับจำนวนครั้งของการปฏิเสธสมมติฐานว่าง

(7) ทำการคำนวณขั้นตอนที่ (2) - (6) ซ้ำจนครบ 2,000 ครั้ง

(8) ค่าอำนาจการทดสอบจากการทดลอง (ค่าประมาณ) เท่ากับ จำนวนครั้งที่ปฏิเสธสมมติฐานว่างหารด้วยจำนวนครั้งที่ทำการทดลอง (2,000 ครั้ง)

(9) ทำซ้ำขั้นตอนที่ (2) – (8) จนครบทุกสถานการณ์ของทุกการแจกแจงตามข้อ (1)

จากนั้นทำการเปรียบเทียบค่าอำนาจการทดสอบที่คำนวณได้ในแต่ละสถานการณ์ โดยที่ตัวสถิติทดสอบที่ให้ค่าอำนาจการทดสอบสูงสุดจะเป็นตัวสถิติทดสอบที่มีประสิทธิภาพในการทดสอบมากที่สุด

1.7 เกณฑ์การตัดสินใจ

การเปรียบเทียบอำนาจการทดสอบของตัวสถิติทดสอบการแจกแจงปกติหลายตัวแปรของตัวสถิติทดสอบที่สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ มีหลักการพิจารณาดังนี้

1. ค่าประมาณความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 หาได้จากสัดส่วนของเหตุการณ์ที่ค่าสถิติตกอยู่ในขอบเขตวิกฤตเมื่อข้อมูลมาจากการแจกแจงปกติหลายตัวแปร และในการทดสอบว่าตัวสถิติทดสอบสามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้หรือไม่ จะใช้การทดสอบทวินาม โดยทำการทดสอบว่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ของตัวสถิติทดสอบ (α) มีค่าไม่เกินความน่าจะเป็นของความคลาดเคลื่อนที่กำหนด (α_0) หรือไม่ ภายใต้ระดับนัยสำคัญ α^* โดยมีรูปแบบการทดสอบเป็นดังนี้

$$H_0 \ : \ \alpha \le \alpha_0$$
$$H_1 \ : \ \alpha > \alpha_0$$

โดยทฤษฎีบทลิมิตเข้าสู่ส่วนกลาง (Central Limit Theorem) จะได้ว่า

$$P\left(\frac{\hat{\alpha}-\alpha_0}{\sqrt{\alpha_0(1-\alpha_0)/n^*}} < Z_{\alpha^*}\right) = 1-\alpha^*$$

หรือ

$$P\left(\hat{\alpha} < \alpha_0 + Z_{\alpha^*}\cdot\sqrt{\frac{\alpha_0(1-\alpha_0)}{n^*}}\right) = 1-\alpha^*$$

ดังนั้นช่วงของการยอมรับความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 คือ

$$\left(0\ ,\ \alpha_0 + Z_{\alpha^*}\cdot\sqrt{\frac{\alpha_0(1-\alpha_0)}{n^*}}\right)$$

โดยที่

α แทน ความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ของตัวสถิติทดสอบ

$\hat{\alpha}$ แทน ค่าประมาณความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ของตัวสถิติทดสอบที่ได้จากการทดลอง

α_0 แทน ความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ที่กำหนด ในที่นี้มี 2 ระดับ คือ 0.05 และ 0.10

α^* แทน ระดับนัยสำคัญของการทดสอบทวินาม ในที่นี้เท่ากับ 0.05

n^* แทน จำนวนซ้ำของการทดลอง เท่ากับ 2,000 รอบ

ตัวสถิติทดสอบจะสามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ ถ้า $\hat{\alpha}$ อยู่ในช่วงของการยอมรับ ดังต่อไปนี้

- กรณีควบคุมความคลาดเคลื่อนประเภทที่ 1 (α_0) เท่ากับ 0.05 ตัวสถิติทดสอบจะสามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ ถ้า $\hat{\alpha}$ อยู่ในช่วง [0, 0.0580]
- กรณีควบคุมความคลาดเคลื่อนประเภทที่ 1 (α_0) เท่ากับ 0.10 ตัวสถิติทดสอบจะสามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ ถ้า $\hat{\alpha}$ อยู่ในช่วง [0, 0.1110]

เมื่อผ่านเกณฑ์การควบคุมค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 แล้วจะทำการเปรียบเทียบค่าอำนาจการทดสอบของตัวสถิติทดสอบต่อไป

2. ค่าอำนาจการทดสอบ จะวัดจากสัดส่วนของจำนวนครั้งที่ปฏิเสธสมมติฐานว่างเมื่อสมมติฐานว่างนั้นไม่จริง

1.8 คำจัดความที่ใช้ในการวิจัย

ความคลาดเคลื่อนประเภทที่ 1 (Type I error) คือ ความคลาดเคลื่อนที่เกิดจากการปฏิเสธสมมติฐานว่าง (H_0) เมื่อสมมติฐานว่างนั้นเป็นจริง และความน่าจะเป็นของการเกิดความคลาดเคลื่อนชนิดนี้มีชื่อเรียกว่า ระดับนัยสำคัญ แทนด้วย α

ความคลาดเคลื่อนประเภทที่ 2 (Type II error) คือ ความคลาดเคลื่อนที่เกิดจากการยอมรับสมมติฐานว่าง (H_0) เมื่อสมมติฐานว่างนั้นไม่จริง และความน่าจะเป็นของการเกิดความคลาดเคลื่อนชนิดนี้ แทนด้วย β

อำนาจการทดสอบ (Power of the test) คือ ความน่าจะเป็นที่จะปฏิเสธสมมติฐานว่างเมื่อสมมติฐานว่างนั้นไม่จริง ซึ่งมีค่าเท่ากับ $1-\beta$ เมื่อ β คือ ความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 2

1.9 ประโยชน์ที่คาดว่าจะได้รับ

ประโยชน์ที่คาดว่าจะได้รับจากการวิจัยครั้งนี้ คือ

1. เพื่อเป็นแนวทางในการเลือกใช้ตัวสถิติทดสอบการแจกแจงปกติหลายตัวแปรที่เหมาะสม
2. เพื่อเป็นแนวทางในการศึกษาอำนาจการทดสอบของตัวสถิติทดสอบการแจกแจงปกติหลายตัวแปรอื่นๆ เพื่อให้ได้ตัวสถิติที่เหมาะสมกับสถานการณ์ต่างๆ ต่อไป